Economic History Working Papers No: 349

Like Father Like Son? Intergenerational immobility in England, 1851-1911

Ziming Zhu, LSE

December 2022

Economic History Department, London School of Economics and Political Science,
Houghton Street, London, WC2A 2AE, London, UK. T: +44 (0) 20 7955 7084.

Like Father Like Son? Intergenerational Immobility in England, 1851-1911 *

Ziming Zhu, LSE †

Keywords: social mobility, intergenerational mobility, nineteenth century, England

JEL Codes: J62, N33

Abstract

This paper uses a linked sample of between 67,000 and 160,000 father-son pairs in 1851-1911 to provide revised estimates of intergenerational occupational mobility in England. After correcting for classical measurement errors using instrumental variables, I find that conventional estimates of intergenerational elasticities could severely underestimate the extent of father-son association in socioeconomic status. Instrumenting one measure of the father's outcome with a second measure of the father's outcome raises the intergenerational elasticities (β) of occupational status from 0.4 to 0.6-0.7. Victorian England was therefore a society of limited social mobility. The implications of my results for long-run evolution and international comparisons of social mobility in England are discussed.

*Sincere gratitude to my supervisors Neil Cummins and Chris Minns for their dedication and assistance. I also thank Gregory Clark, Matthew Curtis, Zachary Ward, Patrick Wallis, Jane Humphries, Richard Breen, and participants at the European Historical Economics Society Annual Conference, the British Society of Population Studies Annual Conference, and the LSE Graduate Seminar for their insightful comments and advice. All errors are my own.
†Ziming Zhu, PhD Candidate, Department of Economic History, z.zhu11@lse.ac.uk.

1 Introduction

Social mobility – the movement of individuals between social groups between generations or across the lifetime – is a subject which has fascinated the minds of scholars and the common people. Commentators of the past believed strongly that people can elevate themselves from humble beginnings to the upper echelons of society through sheer efforts. Samuel Smiles (1863) expounded the prospect of social advancement in nineteenth-century Britain in his work *Self-Help*, a book central to the ideology of Victorian liberalism. Across the Atlantic, James Truslow Adams (1931) coined the pursuit of upward mobility as the 'American Dream', a timeless expression of aspiration and optimism that is still spoken of enthusiastically to the present day.

Were Victorian liberals were mistaken in their exaltation of nineteenth-century English society as one of openness and low barriers? Or were opportunities few and far between? Using a newly constructed and improved set of linked data from the full-count England and Wales decennial censuses, this paper estimates the intergenerational elasticity (IGE) of occupational status in England between 1851 and 1911, following the Becker-Tomes model of intergenerational transmission of human capital (Becker and Tomes, 1986). The results show that, contrary to the findings of some earlier works, social mobility was rather limited during the Victorian (and Edwardian) era. Measurement errors cause significant attenuation bias to estimates of social mobility; correcting for it could raise the IGE obtained from 0.4 to 0.6-0.7, or as much as 64 per cent.

This paper thus extends the existing literature on Victorian social mobility. Most previous studies have relied on marriage registers (Miles, 1993, 1999; Mitch, 1993, 2005) or on surname-based measures (Clark and Cummins, 2015). Long (2013) was the first and to this day the only attempt to estimate rates of social mobility using linked census data for England. However, the surprisingly high rate of mobility found by Long may not be a true reflection of the state of nineteenth-century English society. Firstly, his sample size was restricted by the use of a 2 per cent sample of the 1851 census. This raises issues of representativeness while also increases the likelihood of Type I errors in linking. Moreover, Bailey et al. (2019) and Anbinder et al. (2021) both emphasised the issue of false positives when linking census records, which could cause significant attenuation bias, leading us to conclude that mobility was far greater than what it was in reality. Long also did not attempt to address the issue of measurement errors, which has been highlighted by Ward (2021) in his work on historical mobility in the US.

Although this is not the first time English historical mobility has been estimated, this paper is first to have used the Integrated Census Microdata (I-CeM) complete-count census data, thus greatly expanding the number of observations. Moreover, it provides revised intergenerational elasticities of occupational status after accounting for both classical measurement errors and false positives in census linking. The results are robust to alternative methods of census linkage and to different occupational indices. False positives and reweighting do not have a significant impact on my findings.

The rest of the paper is organised as follows. Section 2 reviews the existing literature on historical social mobility. Section 3 presents the data used and the census linking process and outcomes. Section 4 outlines the methodology, or how social (occupational) mobility is measured in this paper. The results are shown in Section 5; they represent a significant revision from previous works and highlight the impact of measurement errors. Section 6 discusses the implications of these results and makes some comparisons across both time and space. Section 7 concludes.

2 Literature Review

The reign of Queen Victoria is commonly associated with the ascent of Britain as the most dominant Great Power in the world. Through economic and military power and coercion, Britain acquired its 'empire on which the sun never sets'; the nineteenth century witnessed the pinnacle of British imperialism. Domestically, far removed from Britain's exploits in global affairs, it was also a period of social changes, disturbances, and reforms.

Victorian England was the outcome of one of the most transformative events in economic history – the Industrial Revolution. Yet, despite the 'revolution' ending nominally in 1830, the process of structural change carried on. Between 1851 and 1911, the share of employment in agriculture more than halved while the service sector continued to expand rapidly (Thomas, 2004), with the rise of clerical workers, post offices, and bureaucratic organisations. In addition, a number of other social changes were taking place during this period. The country was becoming more urbanised, better connected (with developments in transport and communication infrastructure), and more migratory (Baines and Woods, 2004; Bogart et al., 2022). The passage of the Married Women's Property Act in 1882 ended the law of coverture, enabling married women to own properties legally, while the 1870 Education Act made schooling compulsory. Therefore, it is easy to see why one might be interested in the extent of social mobility during the Victorian (and Edwardian) era.

Research on historical social mobility is often confined by the (in)availability of individual-level sources that include variables which convey one's social status. In the absence of reliable information on income, occupations are often the preferred measure of status. Miles (1993, 1999) studied over 10,000 marriage registers between 1839 and 1914 and found that the share of sons in a different occupational class to their fathers was only 38 per cent, thereby concluding that Britain during this period was 'profoundly unequal'. His findings are corroborated by Mitch (1993, 2005), who finds similar levels of mobility in his sample. However, Delger and Kok (1998) argue that marriage registers underestimate total mobility while also biases mobility estimates in a downward direction due to the age differences between fathers and sons.

Long (2013) overcame the weaknesses of marriage registers by linking fathers and sons from the 1851, 1881, and 1901 censuses. His results confirm the inadequacies of estimating mobility from marriage registers. He found that Victorian society was much more mobile than previously thought, and almost as mobile as late-twentieth century Britain; this appears to reaffirm the beliefs of Victorian liberal observers like Smiles. This finding is at odds with the estimates derived from alternative methods and sources. Clark and Cummins (2015), using surname-based estimates of wealth mobility, found that social mobility did stagnate over the long run, but they also show that nineteenth-century England was far from an open society.

However, the surprisingly high rate of mobility found by Long may not be a true reflection of the state of nineteenth-century English society. Firstly, his sample size (12,516 father-son pairs for 1851-81, and 4,071 for 1881-1901) was restricted by the use of a 2 per cent sample of the 1851 census. This raises issues of representativeness while also increases the likelihood of Type I errors in linking.[1] Moreover, Bailey et al. (2019) and Anbinder et al. (2021) both emphasised the issue of false positives, which could cause significant attenuation bias, leading us to conclude that mobility was far greater than what it was in reality.[2] Bailey et al. were also sceptical of the use of phonetic names in linking algorithms – the strategy which Long (2013) used in his linking.

[1]Automated census linking often entails the removal of individuals that do not have a unique combination of name, age, and birthplace, since the algorithm cannot distinguish which is the correct match. By using a 2 per cent sample, some non-unique individuals may appear as unique if their duplicates are eliminated by the process of sampling.

[2]For instance, Bailey et al. (2019, p. 3) estimate that false links could bias IGE downward by up to 20 per cent.

The issue of classical measurement error is another factor which could lead to significant attenuation bias. Inferring socioeconomic status from occupations from historical censuses is subject to measurement errors because occupations sometimes are misreported by the head of household who filled out the censuses, or by census enumerators who transcribed the census returns onto the enumerator's book; they could also be miscoded during the process of digitising the data.[3] Occupational status, particularly in the past, could be unstable and transitory, and people could be affected by temporary shocks to their labour market outcomes which they may recover from a few years later (i.e. before the next census). Thus, the occupation observed in one census year may not be an accurate reflection of one's true socioeconomic status. Ward (2021) proposes that by averaging father's occupations across different censuses or by instrumenting one father's occupation with another, measurement errors can be accounted for. This should reduce the attenuation bias caused by measurement errors and lead to a significant upward revision of the IGE.

After accounting for measurement errors, Ward finds that the revised IGE estimates for the US between 1850 and 1940 increased from between 0.36-0.49 to between 0.53-0.71. He concludes that nineteenth- and early-twentieth century US was hence less mobile than modern day US. This represents a significant departure from the existing consensus that posits a decline in intergenerational mobility in the US since the nineteenth century (Long and Ferrie, 2013). Therefore, our understanding of British/English occupational mobility since the Victorian era may be open to scepticism too. In addition, past research in comparing rates of historical social mobility between countries, such as that of Long and Ferrie (2013) and Pérez (2019), found Britain to be much less mobile than the US. This could also be subject to amendment if the effects of classical measurement errors are different across countries.

3 Data

3.1 The Census and I-CeM

This research uses two sources of data. The first is the Integrated Census Microdata (I-CeM) – a database containing all the anonymised information from the British decennial censuses between 1851 and 1911 (except for 1871) – compiled and published by Schürer

[3]For a detailed explanation of the census-taking procedure in Britain between 1851 and 1911, see Appendix A

and Higgs (2014). The second is the I-CeM Names and Addresses database (Schürer and Higgs, 2015), which contains data on the names and addresses of the individuals in the main I-CeM database that have been removed by the process of anonymisation. This information is necessary to conduct record linkage.

The censuses of 1851 to 1911 recorded all the vital information that are needed for occupational mobility research, specifically name, age, sex, place of birth, and occupation, with reasonably reliability. This information was then transcribed and enriched by the I-CeM project via a computer programme.[4] This automatic processing, aside from achieving practical efficiency, ensured that decisions concerning the validity of the underlying data source have been applied consistently across the entire database. Of course, this process cannot be perfect. For example, it is not possible to reconcile all the geographical information in the database with that published in the Census Report (Higgs et al., 2013).

The most significant undertaking of I-CeM is the standardisation of raw textual strings. There were over 7.3 million unique strings for occupations and over 6.7 million for birthplace information, which had to be processed and coded into numeric occupation codes. This enables the use of the I-CeM database for this study, since occupations have been coded into a manageable range of categories, while birth places have been standardised to the parish level. Naturally, the automatic coding of this vast number of occupational strings will introduce errors, leading to some occupations being mis-coded. Higgs et al. (2013) assert that for at least 95 per cent of individuals with an occupation title the coding is 'correct'. Other variables, such as marital status and household relationships, have also been standardised, coded, and checked for consistency.

3.2 Measuring Occupational Status

In order to measure the association and transmission of socioeconomic status from fathers to sons, occupations must first be assigned a score that reflects their positions in society. One way of doing this is to assign scores based on the Historical Cambridge Social Interaction and Stratification Scale (HISCAM). This scale was constructed by Lambert et al. (2013) using patterns of intergenerational occupational connections, by exploiting data on social connections – such as marriage, friendship, or parent-child relationships – between the in-

[4]This involved: reconciling the data with the Census Reports; reformatting the input data; performing a number of consistency checks on the data and altering the data accordingly; reformatting and standardising the data; adding a number of enriched variables, mainly relating to household structure.

cumbent of occupations. The main assumption here is that people with similar social status will interact more often. Based on their methodology, they assign a score between 0 and 100 to each occupation, with higher scores indicating a higher social status. The scores are then rescaled to a mean of 50 and a standard deviation of 10.

The data used to construct HISCAM cover the period between 1800 and 1938 and originate from seven countries – Belgium, Britain, Canada, France, Germany, the Netherlands, and Sweden. Different variations of the HISCAM scale have been created depending on the subset of the data used. For this paper, the 'hiscam_u2' scale, which is generated using only male records, is used. To ensure that the occupational mobility (or immobility) observed is not simply a product of the way occupations are scored by HISCAM, an alternative system of scoring occupations will be used. The one chosen here is the FOE-RCII index constructed by Clark et al. (2022), using a set of 1.7 million marriage registers in England between 1837 and 1940. In comparison, Lambert et al. (2013) had information from 990,000 marriages, of which only around 51,000 came from Britain between 1800 and 1938.

The methodology applied create this index is the same as the one used by Lambert et al. (2013) for HISCAM. Using information from marriages, Clark et al. (2022) calculate how closely the holders of each occupation are associated with each other by social connections, such as marriages. Occupations that are far apart in terms of social connections, such as a Member of Parliament and a miner, will have very few social interactions between them (i.e. very few sons of MPs marrying daughters of miners), thus they will be given vastly different scores. Again, the scores are between 0 and 100, with higher scores representing higher status.[5]

3.3 Census Linking Procedure

To conduct record linkage across the censuses, this project selects English-born sons who are aged 5 to 15 with fathers aged 30 to 55 at the start and tracks them across a 30-year period. Two linked samples are then produced. For the baseline sample, the sons are matched once at the end of the period when they are aged 35 to 45. For the multiple links (ML) sample, which is used to correct for measurement errors, the sons are linked across every 10-year interval and the fathers are linked across one 10-year interval. This is done for three periods:

[5]They also construct a different index using an alternative methodology – principal component analysis. Clark et al. find that, reassuringly, HISCAM is very effective at capturing socioeconomic status. All their indices show a strong association with HISCAM.

1851-1881, 1861-1891, and 1881-1911.[6]

Historical census record linkage is a complicated process, due to the lack of a unique identifier like a Social Security Number across datasets. Matching relies heavily on intransient information such as name, birth year, and birthplace. Both the reporting and recording of this limited set of characteristics can be inconsistent. This creates the potential for false matches (Type I errors) and missed matches (Type II errors), and there is a trade-off between minimising these two types of errors. Choosing an algorithm that eliminate as many false positives as possible while still achieving a satisfactory match rate is crucial for automated record linking (Ruggles et al., 2018).

This paper adopts a prominent automated census linkage technique developed by Abramitzky et al. (2014, 2019) – henceforth ABE – which matches individuals over time by names (and their Jaro-Winkler string distances), places of birth (in this case parish), and inferred birth year from age.[7] The procedure is outlined in Appendix B. This paper opts for the more conservative approach in matching, which minimises false positives at the expense of a smaller sample (fewer Type I errors, more Type II errors).

The adoption of a more conservative approach to linking adheres to the findings and recommendation made by Bailey et al. (2019), who reviewed a number of prominent automated linkage methodologies (including ABE). They compared the intergenerational mobility elasticity estimates derived from algorithm-linked samples of two pairs of high-quality datasets to the estimate derived from hand-linked samples and a synthetic 'ground truth' sample created by the authors.[8] They concluded that reducing false matches is more important than generating a higher match rate for improving inferences with linked data, evidenced by the extent of attenuation of the mobility estimates caused by the errors. Although different linking methods produce different samples, eliminating false matches renders estimates from different algorithms statistically indistinguishable.

The ABE methodology (the NYSIIS version) has also come under criticism for the high

[6]To take the 1881-1911 period as an example, sons would be linked between 1881 and 1891, 1881 and 1901, and 1881 and 1911, while fathers would be linked between 1881 and 1891. Similar process follows for 1851-1881, except sons would not be linked between 1851 and 1871 since the 1871 data is not available. For 1861-1891, fathers are linked between 1851 and 1861 instead.

[7]Initially, matching in the ABE algorithm was based on phonetic names (NYSIIS). This was used in Abramitzky et al. (2014). The matching procedure for ABE-NYSIIS is described in Appendix B and carried out for robustness tests. The JW version of ABE is taken from Abramitzky et al. (2019).

[8]The ground truth sample was built with deliberate alterations by the authors to mimic errors in recording, transcribing, and digitising the data, which ensures complete certainty about correct and incorrect links. The synthetic data yields very similar results to the hand-linked records.

rate of false positives produced when attempting to link Irish immigrants in the US across the American censuses (Anbinder et al., 2021). To ensure that the results obtained in this paper are not significantly impacted by false matches, I have devised a method for estimating the rate of Type I errors and used this to construct a more conservative 'true' sample for robustness tests. The test for false positives exploits the fact that sons and fathers are matched across multiple census years and is outlined Appendix C.

There are a priori reasons to believe that false matches may be less of an issue with linking British censuses. While the US data lacked detailed birthplace information, such that Abramitzky et al. (2014, 2019) could only match people based on the state of birth (equivalent to county level for England), the I-CeM database allows matching based on standardised parish of birth. The latter was also not available to Long (2013), so they were not able to address the issue of some parishes having multiple or changing names. Moreover, as Anbinder et al. (2021) recognised that matching Irish people may produce a higher rate of false positives due to a higher incidence of common names. Therefore, the likelihood of Type I error from the use of the ABE algorithm in linking the British censuses should be even lower.

Another issue with census linking is the representativeness of the linked data. Bailey et al. (2019) contend that linking, whether by hand or by machine, cannot produce a fully representative sample. This is because individuals are required to be 'unique' by name, age, and birthplace, which necessarily means that it will be easier to match people with rarer and/or longer names. This may inadvertently introduce bias into the sample if people with these names systematically differ from people with common names. Moreover, people with higher levels of education may be easier to link since they can report their information more accurately and more consistently over time. The match rate may also vary with age, as the incidence of emigration and mortality differs between the young and the old.

However, the impact of a non-representative sample may be less significant than false positives. Bailey et al. (2019) shows that reweighting the sample by inverse probability can effectively address the issue of sample selection bias.[9] They also suggest that after removing the incorrect links, reweighting makes little difference. Abramitzky, Boustan, Eriksson, James and Pérez (2020) also state that coefficient estimates and parameters of interest derived from different samples, weighted or otherwise, produced by the different algorithms they tested are very similar and do not change the interpretation.

[9] For robustness test, I follow their advice on reweighting the sample using inverse probability. The procedure is described in Appendix D.

3.4 Census Linking Results

Table 1 shows the linkage results for the periods 1851-1881, 1861-1891, and 1881-1911. For the baseline samples, between 290,000 and 610,000 father-son pairs have been successfully matched, which translates to a match rate of 21 to 29 per cent. Upon restricting the sample to sons who can be matched across every census in the 30-year period with fathers who can be matched across a 10-year interval, the match rate decreases to between 5 to 8 per cent. This still generates between 68,000 to 160,000 father-son pairs – a huge improvement on the sample size of Long (2013), who had only 12,516 father-son pairs for 1851-81 and 4,071 pairs for 1881-1901.

A comparison of the key socioeconomic indicators suggest that both the baseline and the multiple links samples are very representative of the full population, as shown by Figure 1. The numbers are indexed against the population with the population set at 0. In terms of occupational status – measured by HISCAM and RCII – and age both the sons and their fathers show negligible difference to the wider population. The same is true for the sons' first and last name lengths, and the number of kids and servants they have.

Table 1: Census linkage results for 1851-1911

	Population	Linked (Baseline)	Multiple Links
1851-1881			
N	1,291,487	293,889	68,329
Match Rate (%)		22.76	5.29
1861-1891			
N	1,445,779	311,119	86,884
Match Rate (%)		21.52	6.01
1881-1911			
N	2,148,480	612,481	164,318
Match Rate (%)		28.51	7.65

Sources: author's analysis of I-CeM (UKDA, SN 7481) and I-CeM Names and Addresses (UKDA, SN7856). Notes: match rates are lower-bound estimates, calculated using the following formula (*Size of Linked Sample*) / (*Population of Potential Matches in* 1851/1861/1881).

Other variables, such as household relationship status, marital status, occupational structure, and geographical distribution, are also presented. It may be worth noting that in terms of the geographical distribution of the linked sample, both by county of birth and by registration district of residence, matching tends to be biased against dense, urban regions such as London and Lancashire. This is to be expected since it is more difficult to find 'unique' individuals in parishes with denser population. As a result, the linked sample also tends to be more agricultural, especially for the more restrictive sample with multiple links. As Bailey et al. (2019) demonstrated, these issues can be corrected using inverse probability weights (see Appendix E for more detail), and Section V will show that reweighting does not change the results significantly.

4 Methodology

A standard approach in estimating intergenerational mobility in the social mobility literature, particularly for the modern era, is to calculate the IGE of any measure of socioeconomic status by regressing the log of son's outcome ($Y_{i,t}$) on the log of father's outcome ($Y_{i,t-1}$):

$$Y_{i,t} = \alpha + \beta Y_{i,t-1} + \epsilon_{i,t} \tag{1}$$

Where α is the constant, $\epsilon_{i,t}$ is a set of random factors, and the coefficient of interest is β, which is the IGE estimate. A perfectly mobile society will have an IGE of 0, indicating no association between the father's outcome and the son's outcome. Conversely, a very immobile society will have an IGE of close to 1.

The socioeconomic outcome of an individual observed in a given year consists of a permanent component and an uncorrelated transitory component. As such, our occupation-based measures of status may be noisy, so the occupational status of the father observed in a single year may deviate from his permanent status, which attenuates β towards 0:

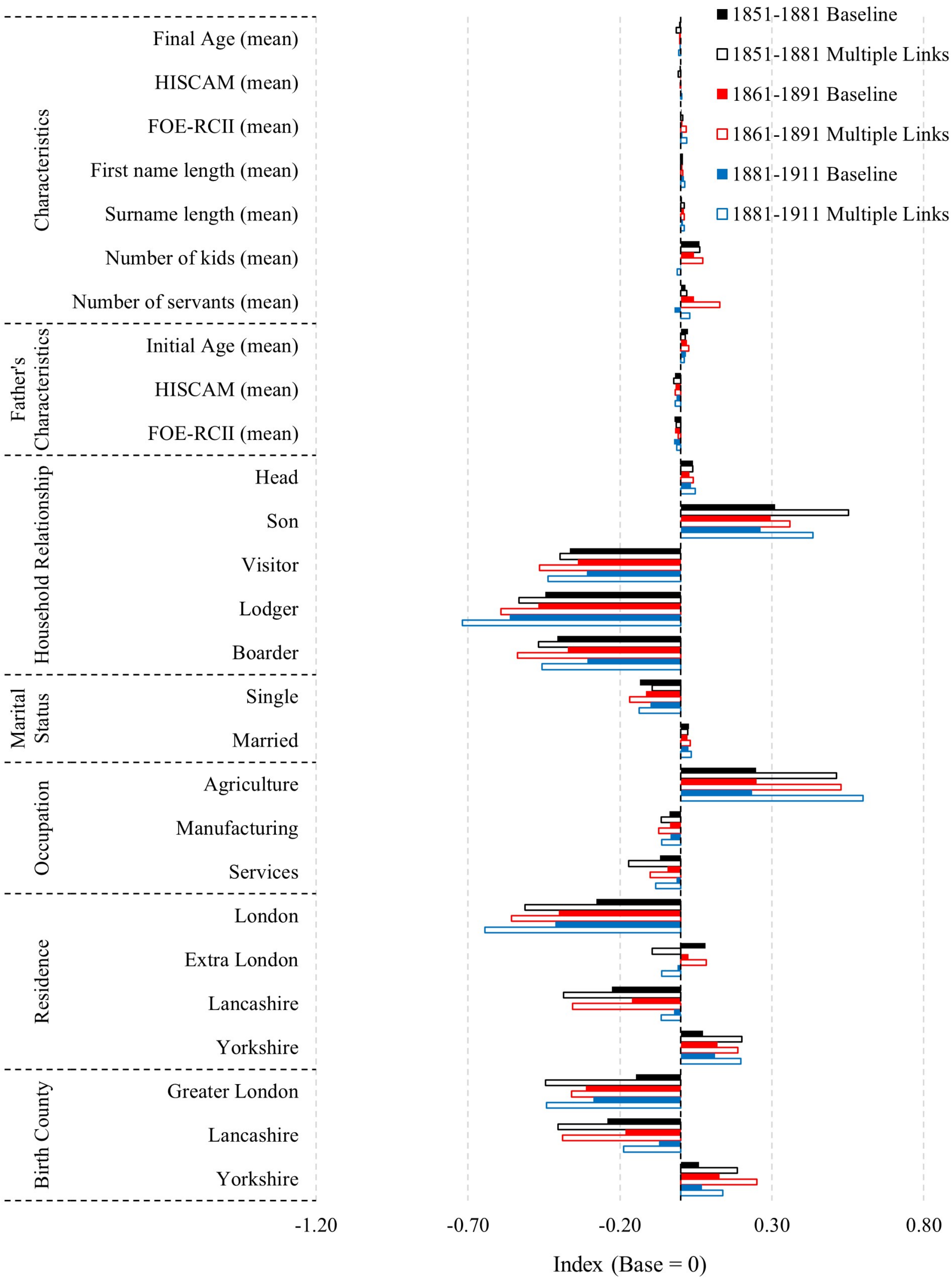
$$Y_{i,t-1} = y_{i,t-1} + u_{i,t-1} \tag{2}$$

To address the issue of classical measurement errors, one method is to average multiple observations of the father's status by T times:

$$plim\widehat{\beta_{avg}} = \beta \frac{var(y_{i,t-1})}{var(y_{i,t-1}) + \frac{var(u_{i,t-1})}{T}} \tag{3}$$

This reduces the attenuation bias caused by errors-in-variables. Modern-day mobility studies often use an average of incomes from many years – a classic example being Mazumder (2005) who averaged fathers' earnings as many as 16 times – but historical mobility research is

Figure 1: Sample representativeness, 1851-1911



Sources: author's analysis of I-CeM (UKDA, SN 7481) and I-CeM Names and Addresses (UKDA, SN7856).Notes: calculated based on raw numbers shown in Appendix E, Table E1, E2, E3. 'Extra London' refers to the regions of Middlesex, Kent, Essex, and Surrey that are not included in 'London'. 'Greater London' refers to the entire regions of Middlesex, Kent, Essex, and Surrey. 'Yorkshire' includes all Ridings of Yorkshire.

limited by data availability and the costs of linking censuses. Though the costs have fallen in recent years with the advent of big data and automated census linking, it is still difficult to obtain more than three observations of occupational status (over time) for a single individual as the census was taken only once per decade. More observations also mean greater sample attrition.

A second method is to instrument the father's outcome with a second measure of the father's outcome ($Z_{i,t-1}$), assuming that the transitory components of the occupational statuses ($\epsilon_{i,t}$ and $\mu_{i,t}$) observed are uncorrelated across different observations:

$$Y_{i,t} = \beta_0 + \beta_1 \widehat{Y_{i,t-1}} + \epsilon_{i,t} \tag{4}$$

$$\widehat{Y_{i,t-1}} = \pi_0 + \pi_1 Z_{i,t-1} + \mu_{i,t} \tag{5}$$

Both methods have been implemented for modern-day studies (for instance by Altonji and Dunn (1991); Solon (1992); Zimmerman (1992) in the US context, and Dearden et al. (1997); Grawe (2004) for the British context) and more recently, for historical studies by Ward (2021). The instrumental variables (IV) approach is shown to work as well if not better than averaging across three father's observations. To carry out the IV method, this paper instruments the father's occupation at the start of each of the three periods (1851-1881, 1861-1891, and 1881-1911) with the father's occupation observed in another census, 10 years apart.

One concern with the IV approach is that life-cycle variations in socioeconomic status could have an impact on the IGE estimated. Haider and Solon (2006) show that attenuation or amplification bias to β could occur if the incomes of sons are observed at younger or older ages; this can be mitigated by measuring status at mid-life – around early 40s (Haider and Solon, 2006; Modalsli and Vosters, 2019). This falls within the middle of the age range (35 to 45) from which the sons' occupational status is taken in this paper. Moreover, additional checks show that the IGE estimated using the occupational status of sons observed at different census years quite similar (Appendix I), so life-cycle effects are not significant enough to cast doubts on the results and their interpretations.

5 Results

Table 2 illustrates the main findings of this paper. The IGE of log occupational status for the baseline sample are shown in columns 1, 4, and 7 for the periods 1851-1881, 1861-1891, and 1881-1911. The OLS estimate of the β for the sample with multiple links are shown in columns 2, 5, and 8. Standard errors are shown in parenthesis; all estimates are statistically significant at the 0.01 level. The β for the sample with multiple links is slightly higher than the β for the baseline sample across all periods. This may indicate that linking sons across multiple years, rather than just once across the 30-year interval, reduces the likelihood of false positives and hence the attenuation bias associated with false matches, though the difference is not huge.

Table 2: Intergenerational elasticities of occupational status (HISCAM), 1851-1911

	1851-1881			1861-1891			1881-1911		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	OLS	OLS	IV	OLS	OLS	IV	OLS	OLS	IV
β	0.402	0.414	0.679	0.384	0.405	0.648	0.391	0.408	0.624
	(0.002)	(0.004)	(0.007)	(0.002)	(0.004)	(0.006)	(0.001)	(0.003)	(0.004)
ML	NO	YES	YES	NO	YES	YES	NO	YES	YES
N	257,844	66,965	65,700	267,089	84,097	83,095	597,517	161,568	159,723

Sources: author's analysis of I-CeM (UKDA, SN 7481) and I-CeM Names and Addresses (UKDA, SN7856). Notes: standard errors in parenthesis; all estimates are statistically significant to $p < 0.01$; ML stands for 'Multiple Links' - whether the sons have been double- or triple-linked.

More importantly, the results clearly suggest that measurement errors associated with occupational status cause significant downward bias in historical mobility estimates. Columns 3, 6, and 9 show the estimates of IGE after instrumenting one father's occupation with a second father's occupation. After accounting for errors-in-variables through the instrumental variable approach, the association between the father's and son's occupational status increases from around 0.41 to between 0.62 and 0.68 – an increase of 53 to 64 per cent. This is a considerable revision on previous estimates by Long (2013), whose estimates of IGE of occupational earnings stood between 0.26 and 0.37 for the periods 1851-1881 and 1881-1901. It is important to note too that even without using the IV approach, the extent of mobility is lower than what Long had estimated, as the OLS β range from 0.38 to 0.41.

Part of the discrepancy may be explained by the differences in the linked sample. My sample, which is much larger in size, may have been more representative and less prone to Type I errors, which would explain the higher β estimated vis-à-vis Long (2013). Most of the differences, however, came from using the instrumental variable approach. This reinforces the concerns over the attenuation bias caused by measurement errors in many existing estimates of social mobility – they could be overestimating mobility by twice as much, if not more.

While the OLS estimates show no changes in the rate of occupational mobility over time, the IV estimates suggest that England was becoming gradually more mobile over the course of the nineteenth century. This might be explained by the effects of measurement errors weakening over time as occupations become more stable, and people become more adept at reporting their personal information. Nevertheless, the decline is quite modest in magnitude.

Table 3 provides some additional results. When a different occupational score index (RCII scores) is applied, there is still a significant extent of attenuation in the β estimated using the conventional OLS formula, caused by measurement errors. The β rises from between 0.52-0.53 to between 0.63-0.71 – 21 to 34 per cent higher – after instrumenting with a second father's observation. Interestingly, the RCII β obtained using the IV approach is akin to the one for HISCAM, except for the 1881-1911 period, which might be expected given that both indices are constructed using similar methods. The fact that the OLS coefficients for RCII are much higher, and likewise the IV coefficients for the last period, suggest that the RCII index may be a better measure of occupational status for England during this period, though more work is required to attest this. Regardless, the results confirm that there is a sizeable reduction in the degree of openness versus earlier estimates of intergenerational mobility.

Allowing occupational scores to vary over time to adjust for the changes in the socioeconomic status associated with each occupation also makes a modest improvement to the β estimated. HISCAM provides two alternative scales constructed using historical records from different periods: HISCAM-E for an early period of 1800 to 1890, and HISCAM-L for a later period of 1890 to 1938 (Lambert et al., 2013). The 'Time-Adjusted' OLS and IV estimates for 1861-1891 and 1881-1911 are produced when sons' occupations are scored using the HISCAM-L scale and fathers' occupations are scored using the HISCAM-E scale. Both estimates are higher than when fathers' and sons' occupations are scored using the same HISCAM-U2 scale. The difference is greater for the 1881-1911 period and significant to the 95 per cent confidence interval.

In addition, estimating β using different samples constructed for robustness checks produced very similar results. The 'Weighted' sample refers to the multiple links sample with inverse probability weights assigned according to the procedure outlined in Appendix D.

Table 3: Additional estimates of IGE

	OLS			IV		
	β	SE	N	β	SE	N
1851-1881						
Main Results	0.414	(0.004)	66,965	0.679	(0.007)	65,700
RCII Scores	0.529	(0.004)	66,854	0.710	(0.005)	65,559
Weighted	0.411	(0.009)	66,965	0.655	(0.014)	65,700
NYSIIS	0.415	(0.004)	69,036	0.678	(0.007)	67,684
False Positive Check	0.410	(0.005)	59,256	0.669	(0.008)	58,163
1861-1881						
Main Results	0.405	(0.004)	84,097	0.648	(0.006)	83,095
Time-Adjusted	0.417	(0.004)	84,097	0.655	(0.007)	83,095
RCII Scores	0.520	(0.003)	83,908	0.630	(0.004)	82,862
Weighted	0.396	(0.007)	84,097	0.632	(0.012)	83,095
NYSIIS	0.401	(0.004)	87,844	0.649	(0.006)	86,745
1881-1911						
Main Results	0.408	(0.003)	161,568	0.624	(0.004)	159,723
Time-Adjusted	0.427	(0.003)	161,568	0.645	(0.004)	159,723
RCII Scores	0.530	(0.002)	161,015	0.691	(0.003)	159,029
Weighted	0.397	(0.005)	161,568	0.611	(0.007)	159,723
NYSIIS	0.406	(0.003)	162,447	0.623	(0.004)	160,575
False Positive Check	0.404	(0.003)	142,086	0.622	(0.004)	140,464

Sources: author's analysis of I-CeM (UKDA, SN 7481) and I-CeM Names and Addresses (UKDA, SN7856). Notes: all estimates are statistically significant to $p < 0.01$; all occupations are scored using HISCAM-U2 except where RCII Scores are stated and Time-Adjusted estimates where HISCAM-E (for fathers' occupations) and HISCAM-L (for sons' occupations) scores were used.

The 'NYSIIS' sample is produced using the phonetic name version of the ABE matching algorithm, as outlined in Appendix B. Lastly, the 'False Positive Check' sample refers to the multiple links sample after removing those who were deemed likely to be false positives, using the method discussed in Appendix C. As the figure highlights, none of these changes affect the results enough to warrant a reconsideration of this paper's findings.

6 Discussions

This paper considerably challenges previous estimates of IGE of occupational status and entail a substantial revision of the perceived wisdom on Victorian social mobility. Table 4 compares the results from this paper to some of the other estimates in the literature, both within the context of England and with the work of Ward (2021) on the US.

The first thing to note is that my OLS estimates for the entire period of 1851-1911 suggest that Long (2013) overestimated the extent of social mobility between 1851-1901. It also shows that there was an increase in mobility between the Victorian and Edwardian eras and the late-twentieth century – based on Long's computation for 1972 and Dearden et al. (1997)'s calculations for 1958. If we compare the IV results, however, the decline in father-son association becomes a lot milder – from 0.679 in 1851-1881 to 0.613 in 1881-1911 and between 0.558 to 0.594 in 1958. My results are in line with the lower-bound estimates of intergenerational wealth elasticities of around 0.64 (not shown in the table), but lower than the upper-bound estimates, found by Clark and Cummins (2015) using probated wealth at death for those dying between 1888 and 1917.[10]

Though it is possible that social mobility increased very slowly between the nineteenth and twentieth centuries, there are also reasons to suspect that my estimates are not capturing the full extent of father-son association in socioeconomic status. Whereas Dearden et al. (1997) and Grawe (2004) had information on net weekly wages of sons, daughters, and fathers from the 1958 National Children Development Survey, the censuses of 1851 to 1911 only provide occupations. And while the IV approach helps to reduce the measurement errors associated with inferring status from occupations, it does not address the measurement errors from assigning scores to occupations. In addition, improvement could also be made to this process by allowing the scores to change according to regional and temporal variations to reflect the rise and fall of certain occupations.

Even though there is not enough information to confidently conclude whether occupational mobility in England increased over time or not, the results still challenge the view that the Victorians lived in an open and mobile society. New estimates suggest that father-son association between 1851 and 1911 was at least between 0.613 and 0.679 and the true figure may be even higher. At the turn of the century, therefore, England was much closer to a society of profound inequalities than one of surprising opportunities.

Finally, my results also speak to the international comparisons of historical mobility. After applying the IV approach, nineteenth-century England does not seem to be exhibiting radically different rates of mobility. Except for the birth cohorts between 1870 and 1900, where there is a dip in father-son association before rising back up again, the IGE estimates for

[10] Their name-based estimates are derived using the latent-factor model, which also accounts for issues of measurement errors.

Table 4: Comparison of IGE estimates between England and the US

Country and Period	OLS	IV	Other	Notes
England				
1851-1881 (Zhu, 2022)	0.41	0.68		
1861-1891 "	0.41	0.65		
1881-1911 "	0.41	0.62		
1851-1881 (Long, 2013)	0.37			
1881-1901 "	0.31			
1972 OMS "	0.33			Oxford Mobility Study data
1958 (Dearden et al., 1997)	0.22	0.59		
1958 (Grawe, 2004)		0.58		
1888-1917 (C&C, 2015)			0.81	Name-based estimates
1918-1959 "			0.69	"
1960-1987 "			0.74	"
USA				
1850 (Ward, 2021)	0.49	0.73	0.81	β in 'Other' accounts for race
1860 "	0.41	0.64	0.77	"
1870 "	0.36	0.55	0.71	"
1880 "		0.42	0.61	"
1890 "		0.49	0.62	"
1900 "	0.39	0.57	0.68	"
1910 "	0.42	0.61	0.70	"

Sources: my estimates come from my own analysis of I-CeM (UKDA, SN 7481) and I-CeM Names and Addresses (UKDA, SN7856); Long (2013); Clark and Cummins (2015); Dearden et al. (1997); Grawe (2004); Ward (2021). Notes: the OMS series are taken from Long (2013)'s calculations based on imputed earnings from occupations. Clark and Cummins (2015) split the sample into 'rich', 'prosperous', 'rich or prosperous', and 'poor', and estimated the IGE for each of these groups, but only the highest estimates are used here, while estimates for the 'poor' group have been excluded in this graph due to large standard errors.

nineteenth- and early-twentieth-century US from Ward (2021) are just as high as those for Victorian and Edwardian England.[11] This undermines the notion that there was something 'exceptional' about American social mobility in the nineteenth century, as Long and Ferrie (2013) had claimed.

[11] One caveat here is that Ward (2021) uses Song et al. (2020)'s literacy-based occupational scores, whereas HISCAM scores are created from social interactions. This might warrant some caution when comparing the coefficients for UK and US.

7 Conclusion

Using a newly constructed and improved set of linked data from the full-count England and Wales decennial censuses, this paper revises the estimates for occupational mobility in England between 1851 and 1911. The results show that, contrary to the findings of some earlier works, social mobility was rather limited during the Victorian (and Edwardian) era. Measurement errors cause significant attenuation bias to estimates of social mobility; correcting for it could raise the IGE obtained by as much as 64 per cent. The results are robust to alternative methods of census linkage and to different occupational indices. False positives and reweighting do not have a significant impact on my findings.

These new estimates represent a significant divergence from the views of those who held Victorian social mobility in a positive light. Victorian liberals were certainly mistaken in their exaltation of nineteenth-century English society as one of openness and low barriers. Opportunities, it would seem, were few and far between. From a long-run perspective, occupational mobility may have increased over time. Yet, if that is indeed what was happening (since we do not have evidence strong enough to stake a claim), it only did so slowly and gradually. From this standpoint, Long (2013) may have been right to be surprised by the extent of social mobility in England. Though what was surprising was not that Victorian social mobility was high – because it was not; it was the seemingly slow and perhaps non-existent increase in intergenerational mobility over the course of a century in which so many social, economic, and political transformations had taken place.

Finally, comparing the revised estimates for England with the revised estimates for the US suggest that scholars of historical mobility ought to beware of the impact of measurement errors and attenuation bias on their own estimates. After using similar methods to account for classical measurement errors, the intergenerational elasticities of occupational status in England do not appear to be radically different to that of the US. This raises questions on whether there was an 'American exceptionalism' in the rate of social mobility in nineteenth-century US. Researchers should also take into consideration of how the quality of their sources might affect comparisons of mobility across countries, as places with better quality sources may be 'penalised' with higher estimates of father-son association in status simply as a result of weaker attenuation bias.

References

Primary Sources

Schürer, K. and Higgs, E. (2014), 'Integrated census microdata (I-CeM), 1851-1911 (SN7481)', Distributed by UK Data Service. http://doi.org/10.5255/UKDA-SN-7481-2.

Schürer, K. and Higgs, E. (2015), 'Integrated census microdata (I-CeM) names and addresses, 1851-1911: Special license access (SN7856)', Distributed by UK Data Service. http://doi.org/10.5255/UKDA-SN-7856-2.

Secondary Sources

Abramitzky, R., Boustan, L. and Eriksson, K. (2019), 'To the new world and back again: Return migrants in the age of mass migration', *ILR Review* **72**, 300–322.

Abramitzky, R., Boustan, L., Eriksson, K., James, F. and Pérez, S. (2020), 'Automated linking of historical data', pp. 1–43.

Abramitzky, R., Boustan, L. P. and Eriksson, K. (2014), 'A nation of immigrants: Assimilation and economic outcomes in the age of mass migration', *Journal of Political Economy* **122**, 467–506.

Abramitzky, R., Mill, R. and Pérez, S. (2020), 'Linking individuals across historical sources: A fully automated approach', *Historical Methods* **53**, 94–111.

Adams, J. T. (1931), *The Epic of America*, Little, Brown, and Co.

Altham, P. M. E. (1970), 'The measurement of association of rows and columns for an r × s contingency table', *Journal of the Royal Statistical Society. Series B (Methodological)* **32**, 63–73.

Altham, P. M. and Ferrie, J. P. (2007), 'Comparing contingency tables: Tools for analyzing data from two groups cross-classified by two characteristics', *Historical Methods* **40**, 3–16.

Altonji, J. G. and Dunn, T. A. (1991), 'Relationships among the family incomes and labor market outcomes of relatives'.

Anbinder, T., Cormac, D. C., Gráda, and Wegge, S. (2021), 'The problem of false positives in automated census linking: Evidence from nineteenth-century new york's irish immigrants', pp. 1–55.

Antonie, L., Inwood, K., Minns, C. and Summerfield, F. (2022), 'Intergenerational mobility in a mid-atlantic economy: Canada, 1871–1901', *Journal of Economic History* pp. 1–27.

Arkell, T. (1994), ‘Identifying the census enumerators – cornwall in 1851’, *Local Population Studies* **53**, 70–75.

Bailey, M., Cole, C., Henderson, M. and Massey, C. (2019), ‘How well do automated linking methods perform? lessons from u.s. historical data’, pp. 1–65.

Baines, D. and Woods, R. (2004), Population and regional development, *in* R. Floud and P. Johnson, eds, ‘The Cambridge Economic History of Modern Britain Volume II: Economic Maturity, 1860-1939’, Cambridge University Press, pp. 25–55.

Becker, G. S. and Tomes, N. (1986), ‘Human capital and the rise and fall of families’, *Journal of Labor Economics* **4**, S1–S39.

Black, S. E. and Devereux, P. J. (2010), ‘Recent developments in intergenerational mobility’, pp. 1–90.

Bogart, D., You, X., Alvarez-Palau, E. J., Satchell, M. and Shaw-Taylor, L. (2022), ‘Railways, divergence, and structural change in 19th century england and wales’, *Journal of Urban Economics* **128**.

Clark, G. and Cummins, N. (2015), ‘Intergenerational wealth mobility in england, 1858-2012: Surnames and social mobility’, *Economic Journal* **125**, 61–85.

Clark, G., Cummins, N. and Curtis, M. (2022), ‘The mismeasure of man: Why intergenerational occupational mobility is much lower than conventionally measured, england, 1800-2021’, pp. 1–35.

Dearden, L., Machin, S. and Reed, H. (1997), ‘Intergenerational mobility in britain’, *Economic Journal* **107**, 47–66.

Delger, H. and Kok, J. (1998), ‘Bridegrooms and biases: A critical look at the study of intergenerational mobility on the basis of marriage certificates’, *Historical Methods* **31**, 113–121.

Erikson, R. and Goldthorpe, J. H. (1993), *The Constant Flux: A Study of Class Mobility in Industrial Societies*, Clarendon Press.

Feigenbaum, J. J. (2018), ‘Multiple measures of historical intergenerational mobility: Iowa 1915 to 1940’, *Economic Journal* **128**, F446–F481.

Ferrie, J. P. (1996), ‘A new sample of males linked from the public use microdata sample of the 1850 u.s. federal census of population to the 1860 u.s. federal census manuscript schedules’, *Historical Methods* **29**, 141–156.

Grawe, N. D. (2004), Intergenerational mobility for whom? the experience of high- and low-earning sons in international perspective, *in* M. Corak, ed., ‘Generational Income Mobility in North America and Europe’, Cambridge University Press, pp. 58–88.

Haider, S. and Solon, G. (2006), ‘Life-cycle variation in the association between current and lifetime earnings’, *American Economic Review* **96**, 1308–1320.

Higgs, E. (1989), *Making Sense of the Census Revisited: The Manuscript Returns for England and Wales, 1801-1901*, HMSO.

Higgs, E. (2005), *Making Sense of the Census Revisited: Census Records for England and Wales, 1801-1901 – a Handbook for Historical Researchers*, The National Archives and Institute of Historical Research.

Higgs, E., Jones, C., Schürer, K. and Wilkinson, A. (2013), 'Integrated census microdata (i-cem) guide'.

Lambert, P. S., Zijdeman, R. L., Leeuwen, M. H. V., Maas, I. and Prandy, K. (2013), 'The construction of hiscam: A stratification scale based on social interactions for historical comparative research', *Historical Methods* **46**, 77–89.

Leeuwen, M. H. D. V. and Maas, I. (2011), *HISCLASS: A Historical International Social Class Scheme*, Leuven University Press.

Leeuwen, M. H. D. V., Maas, I., Miles, A. and Edvinsson, S. (2002), *HISCO: Historical International Standard Classification of Occupations*, Leuven University Press.

Long, J. (2013), 'The surprising social mobility of victorian britain', *European Review of Economic History* **17**, 1–23.

Long, J. and Ferrie, J. (2013), 'Intergenerational occupational mobility in gb and us since 1850', *American Economic Review* **103**, 1109–1137.

Mazumder, B. (2005), 'Fortunate sons: New estimates of intergenerational mobility in the united states using social security earnings data', *Review of Economics and Statistics* **87**, 235–255.

Miles, A. (1993), How open was nineteenth-century british society? social mobility and equality of opportunity, 1839-1914, *in* A. Miles and D. Vincent, eds, 'Building European Society: Occupational Change and Social Mobility in Europe, 1840-1940', Manchester University Press, pp. 18–39.

Miles, A. (1999), *Social Mobility in Nineteenth- and Early-Twentieth Century England*, St. Martin's Press.

Mitch, D. (1993), "inequalities which every one may remove": Occupational recruitment, endogamy, and the homogeneity of social origins in victorian england, *in* A. Miles and D. Vincent, eds, 'Building European Society: Occupational Change and Social Mobility in Europe, 1840-1940', Manchester University Press, pp. 140–164.

Mitch, D. (2005), 'Literacy and occupational mobility in rural versus urban victorian england: Evidence from the linked marriage register and census records for birmingham and norfolk, 1851 and 1881', *Historical Methods* **38**, 26–38.

Modalsli, J. H. and Vosters, K. (2019), 'Spillover bias in multigenerational income regressions standard'.

Pérez, S. (2019), 'Intergenerational occupational mobility across three continents', *Journal of Economic History* **79**, 383–416.

Ruggles, S., Fitch, C. A. and Roberts, E. (2018), 'Annual review of sociology historical census record linkage', *Annual Review of Sociology* **44**, 19–37.

Smiles, S. (1863), *Self-Help with Illustrations of Conduct and Perseverance*, Ticknor and Fields.

Solon, G. (1992), 'Intergenerational income mobility in the united states', *American Economic Review* **82**, 393–408.

Song, X., Massey, C. G., Rolf, K. A., Ferrie, J. P., Rothbaum, J. L. and Xie, Y. (2020), 'Long-term decline in intergenerational mobility in the united states since the 1850s', *Proceedings of the National Academy of Sciences* **117**, 251–258.

Thomas, M. (2004), The service sector, *in* R. Floud and P. Johnson, eds, 'The Cambridge Economic History of Modern Britain Volume II: Economic Maturity, 1860-1939', Cambridge University Press, pp. 99–132.

Tillott, P. (1972), Sources of inaccuracy in the 1851 and 1861 censuses, *in* E. A. Wrigley, ed., 'Nineteenth-Century Society: Essays in the Use of Quantitative Methods for the Study of Social History', Cambridge University Press, pp. 82–133.

Tillott, P. M. (1968), 'The analysis of census returns', *Local Historian* **8**, 2–9.

Ward, Z. (2021), 'Intergenerational mobility in american history: Accounting for race and measurement error', pp. 1–46.

Zimmerman, D. J. (1992), 'Regression toward mediocrity in economic stature', *American Economic Review* **82**, 409–429.

Appendix

A England and Wales Census, 1851-1911

Nineteenth- and early-twentieth-century censuses are an invaluable source of quantitative information into the lives of people living in Victorian and Edwardian England, and an alternative primary resource for the study of occupational mobility in the past. The act of census taking began in 1801, although it was not until 1841 that names and details of individuals were collected, and information on birth places and occupations remained limited until the 1851 census (Higgs, 1989). An awareness of the procedures involved in census taking from 1851 onwards may be required to understand the limits and reliability of the information obtained from the census returns.

A simple explanation of how the census was taken is as follows. The country was first divided into enumeration districts, each containing roughly 200 households and one enumerator. The enumerators delivered a 'household schedule' and written instructions to each household on the night of the census – normally in March or April to avoid the distortions caused by seasonal movements in the summer by some sections of the population – which had to be filled out and returned by the household head. On collection day, the enumerators would collect and check the schedules, and help the household heads to complete the schedule if they could not do so. Up until 1911, the enumerators would then standardise and copy the information onto the Census Enumerator's Book (CEB). Both the schedules and the books were submitted for checking to the district registrars before they were sent to the Census Office, where they were checked again by the clerks. The household returns were then destroyed. For the 1911 census, the original schedules were used for the tabulation of statistics, so there was no standardisation of the raw data by the enumerators (Higgs, 2005).

One concern that scholars may have with the use of nineteenth-century censuses for historical research is the quality of census enumerators. Enumerators were hired on a temporary basis by local registrars, and anyone can be hired as long as they satisfied the basic requirements (Higgs et al., 2013).[12] In urban areas, the enumerators were often local government officers and schoolteachers, but in the countryside the registrars may have had to depend on the

[12] The requirements for an enumerator were: a person of intelligence and activity; able to read, write, and have some arithmetic knowledge; able to undertake the requisite physical exertion involved; must not be younger than 18 or older than 65; must be temperate, orderly, and respectable, conduct himself with strict propriety, and have the goodwill on the inhabitants of his district (women were allowed to become enumerators after 1891).

farmers and their kin (Arkell, 1994). Unsurprisingly, there is a lot of variation in the abilities of enumerators – they differed in their ability to read and write, and in their ability to comprehend lengthy instructions given to them by the registrars (Tillott, 1968). Fortunately, the enumerators generally appear to be of a satisfactory standard. In an area sampled by Tillott (1972), only six of the ninety enumerators showed evidence of unsuitability for their task. This may be especially true for the towns, where enumerators were more likely to be men of clerkly habits employed in occupations that require a certain degree of literacy.

Another source of inaccuracies may come from the householders who inadvertently give out the wrong information, mostly due to ignorance or ambiguity in the instructions. Insofar as people's intentions to answer the questions truthfully were concerned, there is little evidence to suggest that this is a huge issue (Tillott, 1972). With information on name, sex, occupation, and birthplace, there is generally little room for falsification, though inconsistencies may occur as a result of spelling variants with names, ambiguous definitions and instructions given to the recording of occupations, and geographical ignorance (Tillott, 1972; Higgs, 2005). In cases where the householder was illiterate, the enumerators were responsible for filling the schedules. The proportion of schedules filled out by enumerators varied widely across regions – for example, in the six enumeration districts of Great Missenden in Buckinghamshire, this proportion ranged from 5.3 to 64.7 per cent (Higgs et al., 2013). Thus, there may be cases where the wrong information was recorded due to miscommunications between the enumerator and the household. With the introduction of compulsory education after 1870, one would expect the ability to read and fill the schedules improved for both the householder and the enumerator.

B ABE Census Linking Algorithm

The ABE algorithm matches individuals over time by names (string distances or phonetic names), places of birth (in this case parish), and inferred birth year from age (Abramitzky et al., 2020). Matching via string distances is the preferred method in this paper. The procedure for both string distances and phonetic names versions are as follows.

Using Jaro-Winkler string distances

1. The raw strings for first and last names in dataset A (i.e., all men in 1851) and dataset B (i.e., all men in 1881) are cleaned, which removes non-alphabetic characters and

accounts for shortened names such as 'Ben' for Benjamin and spelling variants.

2. The data is then split into smaller blocks by initial letters of first and last names, age, and birthplace. The string distances of all names within plus and minus 5 years of reported age between dataset A and B are calculated, and only pairs of individuals in A and B with string distances of less than 0.1 in both first and last names are kept.

3. There are three potential outcomes in the matching procedure:

   (a) No potential match could be found for a given individual in dataset A, so this observation is dropped from the data.

   (b) There may be only one potential match for an individual in dataset A, and the corresponding match in dataset B has no other potential matches in dataset A. This is determined to be a successful match.

   (c) In cases where they are more than one potential match by name in dataset B, the individual (let us call him B1) closest in inferred birth year to the observation in dataset A is matched only if the second closest observation in B is more than 2 years apart in reported age to B1.

4. To minimise Type I errors, this paper adopts the conservative approach where matches are also required to be unique within a 5-year band (plus or minus 2 years in age) and to differ in reported age by no more than 2 years.

Using NYSIIS Phonetic Names

1. The raw strings for first and last names are cleaned.

2. Names are then converted into their phonetic names using the New York State Identification and Intelligence System (NYSIIS) Code.

3. The sample from the initial year is restricted to those who are unique by first and last name, age, and parish of birth, since it is impossible to distinguish between which non-unique individuals should be linked to the potential match.

4. Following from this, matches can be identified based on their vital information through an iterative procedure:

   (a) If a unique match – same name, birth year, and birth parish – is found, the individual is 'matched'.

(b) If there are multiple matches for the same birth year, the observation is discarded.

(c) If no matches are found for the same birth year, the process is expanded to matching within a one-year band (older or younger), and then within a two-year band around the inferred birth year. Again, only unique matches are accepted.

5. To reduce the likelihood of false positives, matches are required to have unique names within a five-year band (plus or minus two years) around the birth year.

C Estimating False Positive Rate

The procedure for estimating the rate of false positives is as follows. Taking the 1881-1911 sample as an example, I first select sons whose relationship status as reported in the census is 'son' in both 1881 and 1891, indicating that they are living with their families in both years. I then check if the fathers they are living with in 1891 are the same individuals that I identified when I linked their fathers from the 1881 to the 1891 census. This is a valid test because fathers and sons are linked across census years independently. I can then calculate the percentage (γ) of sons whose actual fathers they are living with in 1861 are different to the fathers that I linked.

It is important to note here that this is only an upper-bound estimate of the false positive rate (α) associated with the linking algorithm. This is because the linkage process entails running the algorithm twice – once for matching sons from 1881 to 1891, and once for matching fathers from 1881 to 1891. Thus, γ is an outcome of these four scenarios:

- $P(E_s = 1 E_f = 1) = \alpha * \alpha$, where $E_s = 1$ denotes a Type I error in the linkage of sons and $E_f = 1$ denotes a type I error in the linkage of fathers.
- $P(E_s = 1 E_f = 0) = \alpha * (1 - \alpha)$, meaning fathers are correctly matched between 1881 and 1891 but sons are false matches.
- $P(E_s = 0 E_f = 1) = (1 - \alpha) * \alpha$, meaning that sons are matched correctly between 1881 and 1891 but fathers are false matches.
- An unknown percentage x that represents the share of false positives eliminated by the requirement for sons to have a match in every census year within the 30-year interval. In other words, sons who can be falsely matched between 1881 and 1911 but not between 1881 and 1891 or between 1881 and 1901.

Combining these scenarios produce the following equation:

$$\gamma = 2\alpha - \alpha^2 - x \tag{C1}$$

Solving the quadratic would reveal the true rate of false positive rate associated with the linkage algorithm:

$$2\alpha - \alpha^2 - x - \gamma = 0 \tag{C2}$$

However, since we do not know the exact value of x, we can only derive a lower-bound estimate of the false positive rate by assuming $x = 0$.

Table C1 shows the upper and lower bound estimates for the rate of false positives. For both 1851-1881 and 1881-1911, the false positive rate lies between 8 and 17 per cent. This compares quite favourably to the performance of various prominent linkage algorithms when linking US censuses. For instance, Bailey et al. (2019) found that the most conservative version of ABE-NYSIIS produces a false positive rate of 17 to 23 per cent; Ferrie (1996), when using exact names, produces a false positive rate of 20 to 23 per cent; Feigenbaum (2018) produces a false positive rate of 16 to 29 per cent. Only the Expectation-Maximisation (EM) algorithm constructed by Abramitzky, Mill and Pérez (2020) performs to a similar or better standard - false positive rate of 10 to 15 per cent. Evidently, the availability of more precise birthplace information makes a huge difference to how well automated census linking performs.

Table C1: False positive rate of census linkage

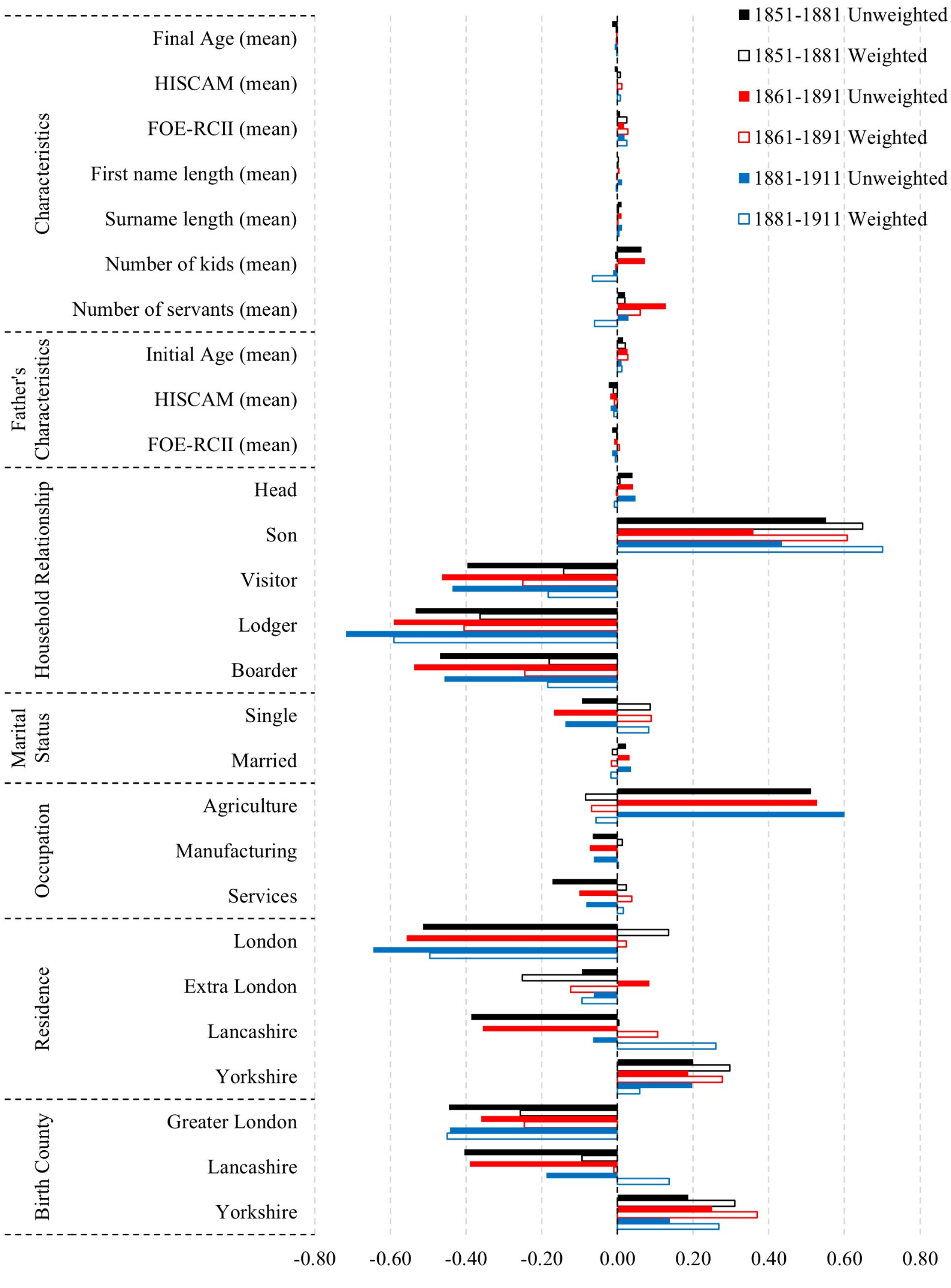
	1851-1881		1881-1911	
	Numbers	Percentage	Numbers	Percentage
Correct Match	37,965	82.84	99,641	83.37
Wrong Match (Sample False Positive)	7,865	17.16	19,869	16.63
Process False Positive (Lower-Bound)		8.98		8.69
Total	45,830		119,510	

Sources: author's analysis of I-CeM (UKDA, SN 7481) and I-CeM Names and Addresses (UKDA, SN7856).

D Inverse Probability Weight

Figure D1 shows the representativeness of the weighted sample in comparison to the unweighted sample. The weighted sample is more representative in almost all variables except

Figure D1: Sample representativeness (weighted vs. unweighted), 1851-1911



Sources: author's analysis of I-CeM (UKDA, SN 7481) and I-CeM Names and Addresses (UKDA, SN7856).Notes: calculated based on raw numbers shown in Appendix E, Table E4, E5, E6. 'Extra London' refers to the regions of Middlesex, Kent, Essex, and Surrey that are not included in 'London'. 'Greater London' refers to the entire regions of Middlesex, Kent, Essex, and Surrey. 'Yorkshire' includes all Ridings of Yorkshire.

in over-representing 'son' in relationship status and Yorkshire in birth and residence counties.

To address the issue of non-representative sample, I ran probit regressions of linkage outcomes (a dummy variable with value of 1 if the observation has been successfully linked) for sons on first name length, last name length, combined name length, first name commonness, last name commonness, age and its quadratic term, total male population in the parish and county of residence in the final census year, and occupational sector defined by the HISCO major groups (0 to 9). Name commonness is defined as the share of people aged 5 to 15 with the same name living in the same parish in 1851 for the 1851-1881 sample, 1861 for the 1861-1891 sample, and 1881 for the 1881-1911 sample. I then assign inverse probability weights based on the following equation:

$$Weight = \frac{1 - P_i(L_i = 1|X_i)}{P_i(L_i = 1|X_i) \cdot q(1-q)} \tag{D1}$$

Where $P_i(L_i = 1|X_i)$ denotes the probability of being linked, and q is the share of people linked.

E Sample Representativeness Table

Table E1, E2, and E3 shows the summary statistics for comparing the sample representativeness of the baseline and the multiple links samples for 1851-1881, 1861-1891, and 1881-1911. These are then indexed against the population to create the graph seen in Figure 1. Table E4, E5, and E6 shows the comparison between the unweighted multiple links sample with the weighted sample. The results were used to construct Figure D1 in Appendix D.

Table E1: Representativeness results, 1851-1881

	Population	Baseline	ML
Characteristics (Son) in 1881			
Final Age (mean)	39.68	39.44	39.12
HISCAM (mean)	54.38	54.14	53.95
FOE-RCII (mean)	53.13	53.26	53.51
First name length (mean)	6.27	6.32	6.30
Surname length (mean)	6.33	6.36	6.40
Kids (mean)	2.85	3.03	3.03
Servants (mean)	0.22	0.22	0.22
Characteristics (Father) in 1851			
Initial Age (mean)	40.81	41.79	41.42
HISCAM (mean)	53.35	52.32	52.07
FOE-RCII (mean)	53.38	52.25	52.61
Relationship Status (Son) in 1881			
Head	82.75	86.14	86.03
Son	4.26	5.58	6.61
Visitor	0.63	0.40	0.38
Lodger	3.30	1.83	1.54
Boarder	2.39	1.42	1.27
Marital Status (Son) in 1881			
Single	12.33	10.67	11.17
Married	83.57	85.81	85.50
Occupational Structure (Son) in 1881			
Agriculture	15.71	19.60	23.76
Manufacturing	60.19	57.96	56.29
Services	24.10	22.44	19.95
Residential Region (Son) in 1881			
London	15.91	11.49	7.74
Extra London	8.90	9.62	8.06
Lancashire	13.32	10.30	8.17
Yorkshire	12.68	13.60	15.22
Birth County (Son)			
Greater London	18.15	15.48	10.07
Lancashire	10.96	8.31	6.53
Yorkshire	12.15	12.87	14.42
Observations (N)	1,291,487	293,889	68,329
Match Rate (%)		22.76	5.29

Sources: author's analysis of I-CeM (UKDA, SN 7481) and I-CeM Names and Addresses (UKDA, SN7856). Notes: 'Population' includes all men aged 35-45 in 1881 when comparing with the sons in the linked sample and all men aged 30-55 in 1851 when comparing with the fathers. 'Manufacturing' in Occupational Structure also includes Mining and Transport sectors. 'Extra London' refers to the regions of Middlesex, Kent, Essex, and Surrey that are not included in 'London'. 'Greater London' refers to the entire regions of Middlesex, Kent, Essex, and Surrey. 'Yorkshire' includes all Ridings of Yorkshire. All numbers are in percentages unless they are means.

Table E2: Representativeness results, 1861-1891

	Population	Baseline	ML
Characteristics (Son) in 1891			
Final Age (mean)	39.65	39.43	39.53
HISCAM (mean)	54.42	54.33	54.40
FOE-RCII (mean)	52.72	52.98	53.66
First name length (mean)	7.80	7.83	7.84
Surname length (mean)	8.31	8.40	8.41
Kids (mean)	2.46	2.57	2.64
Servants (mean)	0.19	0.20	0.21
Characteristics (Father) in 1861			
Initial Age (mean)	40.96	41.83	42.05
HISCAM (mean)	53.52	52.58	52.51
FOE-RCII (mean)	53.06	52.08	52.61
Relationship Status (Son) in 1891			
Head	84.63	86.99	88.08
Son	4.40	5.70	5.98
Visitor	0.56	0.37	0.30
Lodger	2.84	1.51	1.16
Boarder	2.29	1.44	1.06
Marital Status (Son) in 1891			
Single	12.44	11.01	10.35
Married	84.08	85.94	86.75
Occupational Structure (Son) in 1891			
Agriculture	13.26	16.56	20.25
Manufacturing	62.03	59.84	57.52
Services	24.71	23.61	22.23
Residential Region (Son) in 1891			
London	15.61	9.35	6.91
Extra London	9.95	10.20	10.79
Lancashire	13.92	11.69	8.96
Yorkshire	12.85	14.41	15.26
Birth County (Son)			
Greater London	19.75	13.58	12.64
Lancashire	11.29	9.23	6.89
Yorkshire	11.52	12.99	14.40
Observations (N)	1,445,779	311,119	86,884
Match Rate (%)		21.52	6.01

Sources: author's analysis of I-CeM (UKDA, SN 7481) and I-CeM Names and Addresses (UKDA, SN7856). Notes: 'Population' includes all men aged 35-45 in 1891 when comparing with the sons in the linked sample and all men aged 30-55 in 1861 when comparing with the fathers. 'Manufacturing' in Occupational Structure also includes Mining and Transport sectors. 'Extra London' refers to the regions of Middlesex, Kent, Essex, and Surrey that are not included in 'London'. 'Greater London' refers to the entire regions of Middlesex, Kent, Essex, and Surrey. 'Yorkshire' includes all Ridings of Yorkshire. All numbers are in percentages unless they are means.

Table E3: Representativeness results, 1881-1911

	Population	Baseline	ML
Characteristics (Son) in 1911			
Final Age (mean)	39.67	39.48	39.40
HISCAM (mean)	55.29	55.31	55.43
FOE-RCII (mean)	53.39	53.65	54.41
First name length (mean)	7.88	7.97	7.98
Surname length (mean)	8.34	8.39	8.43
Kids (mean)	2.00	2.00	1.98
Servants (mean)	0.13	0.13	0.13
Characteristics (Father) in 1881			
Initial Age (mean)	40.82	41.52	41.26
HISCAM (mean)	54.31	53.50	53.35
FOE-RCII (mean)	53.13	51.97	52.41
Relationship Status (Son) in 1911			
Head	80.73	83.43	84.56
Son	5.75	7.26	8.25
Visitor	0.71	0.49	0.40
Lodger	1.10	0.48	0.31
Boarder	4.22	2.92	2.29
Marital Status (Son) in 1911			
Single	15.53	13.99	13.40
Married	81.46	83.54	84.33
Occupational Structure (Son) in 1911			
Agriculture	10.30	12.71	16.48
Manufacturing	59.74	57.74	56.02
Services	29.96	29.55	27.50
Residential Region (Son) in 1911			
London	12.77	7.49	4.53
Extra London	13.41	13.27	12.56
Lancashire	24.05	23.52	22.51
Yorkshire	2.74	3.05	3.28
Birth County (Son)			
Greater London	22.13	15.78	12.34
Lancashire	12.12	11.25	9.85
Yorkshire	12.52	13.40	14.26
Observations (N)	2,148,480	612,481	164,318
Match Rate (%)		28.51	7.65

Sources: author's analysis of I-CeM (UKDA, SN 7481) and I-CeM Names and Addresses (UKDA, SN7856). Notes: 'Population' includes all men aged 35-45 in 1911 when comparing with the sons in the linked sample and all men aged 30-55 in 1881 when comparing with the fathers. 'Manufacturing' in Occupational Structure also includes Mining and Transport sectors. 'Extra London' refers to the regions of Middlesex, Kent, Essex, and Surrey that are not included in 'London'. 'Greater London' refers to the entire regions of Middlesex, Kent, Essex, and Surrey. 'Yorkshire' includes all Ridings of Yorkshire. All numbers are in percentages unless they are means.

Table E4: Representativeness results (weighted vs. unweighted), 1851-1881

	Population	Unweighted	Weighted
Characteristics (Son) in 1881			
Final Age (mean)	39.68	39.12	39.58
HISCAM (mean)	54.38	53.95	54.81
FOE-RCII (mean)	53.13	53.51	54.43
First name length (mean)	6.27	6.30	6.28
Surname length (mean)	6.33	6.40	6.35
Kids (mean)	2.85	3.03	2.84
Servants (mean)	0.22	0.22	0.22
Characteristics (Father) in 1851			
Initial Age (mean)	40.81	41.42	41.64
HISCAM (mean)	53.35	52.07	52.77
FOE-RCII (mean)	53.38	52.61	53.36
Relationship Status (Son) in 1881			
Head	82.75	86.03	83.27
Son	4.26	6.61	7.02
Visitor	0.63	0.38	0.54
Lodger	3.30	1.54	2.10
Boarder	2.39	1.27	1.96
Marital Status (Son) in 1881			
Single	12.33	11.17	13.40
Married	83.57	85.50	82.48
Occupational Structure (Son) in 1881			
Agriculture	15.71	23.76	14.39
Manufacturing	60.19	56.29	60.94
Services	24.10	19.95	24.67
Residential Region (Son) in 1881			
London	15.91	7.74	18.06
Extra London	8.90	8.06	6.66
Lancashire	13.32	8.17	13.37
Yorkshire	12.68	15.22	16.44
Birth County (Son)			
Greater London	18.15	10.07	13.48
Lancashire	10.96	6.53	9.94
Yorkshire	12.15	14.42	15.92
Observations (N)	1,291,487	68,329	68,329
Match Rate (%)		5.29	5.29

Sources: author's analysis of I-CeM (UKDA, SN 7481) and I-CeM Names and Addresses (UKDA, SN7856). Notes: 'Population' includes all men aged 35-45 in 1881 when comparing with the sons in the linked sample and all men aged 30-55 in 1851 when comparing with the fathers. 'Manufacturing' in Occupational Structure also includes Mining and Transport sectors. 'Extra London' refers to the regions of Middlesex, Kent, Essex, and Surrey that are not included in 'London'. 'Greater London' refers to the entire regions of Middlesex, Kent, Essex, and Surrey. 'Yorkshire' includes all Ridings of Yorkshire. All numbers are in percentages unless they are means.

Table E5: Representativeness results (weighted vs. unweighted), 1861-1891

	Population	Unweighted	Weighted
Characteristics (Son) in 1891			
Final Age (mean)	39.65	39.53	39.54
HISCAM (mean)	54.42	54.40	55.02
FOE-RCII (mean)	52.72	53.66	54.18
First name length (mean)	7.80	7.84	7.79
Surname length (mean)	8.31	8.41	8.32
Kids (mean)	2.46	2.64	2.45
Servants (mean)	0.19	0.21	0.20
Characteristics (Father) in 1861			
Initial Age (mean)	40.96	42.05	42.07
HISCAM (mean)	53.52	52.51	53.11
FOE-RCII (mean)	53.06	52.61	53.31
Relationship Status (Son) in 1891			
Head	84.63	88.08	84.43
Son	4.40	5.98	7.07
Visitor	0.56	0.30	0.42
Lodger	2.84	1.16	1.69
Boarder	2.29	1.06	1.73
Marital Status (Son) in 1891			
Single	12.44	10.35	13.55
Married	84.08	86.75	82.75
Occupational Structure (Son) in 1891			
Agriculture	13.26	20.25	12.35
Manufacturing	62.03	57.52	62.01
Services	24.71	22.23	25.64
Residential Region (Son) in 1891			
London	15.61	6.91	15.98
Extra London	9.95	10.79	8.72
Lancashire	13.92	8.96	15.39
Yorkshire	12.85	15.26	16.41
Birth County (Son)			
Greater London	19.75	12.64	14.88
Lancashire	11.29	6.89	11.19
Yorkshire	11.52	14.40	15.77
Observations (N)	1,445,779	86,884	86,884
Match Rate (%)		6.01	6.01

Sources: author's analysis of I-CeM (UKDA, SN 7481) and I-CeM Names and Addresses (UKDA, SN7856). Notes: 'Population' includes all men aged 35-45 in 1891 when comparing with the sons in the linked sample and all men aged 30-55 in 1861 when comparing with the fathers. 'Manufacturing' in Occupational Structure also includes Mining and Transport sectors. 'Extra London' refers to the regions of Middlesex, Kent, Essex, and Surrey that are not included in 'London'. 'Greater London' refers to the entire regions of Middlesex, Kent, Essex, and Surrey. 'Yorkshire' includes all Ridings of Yorkshire. All numbers are in percentages unless they are means.

Table E6: Representativeness results (weighted vs. unweighted), 1881-1911

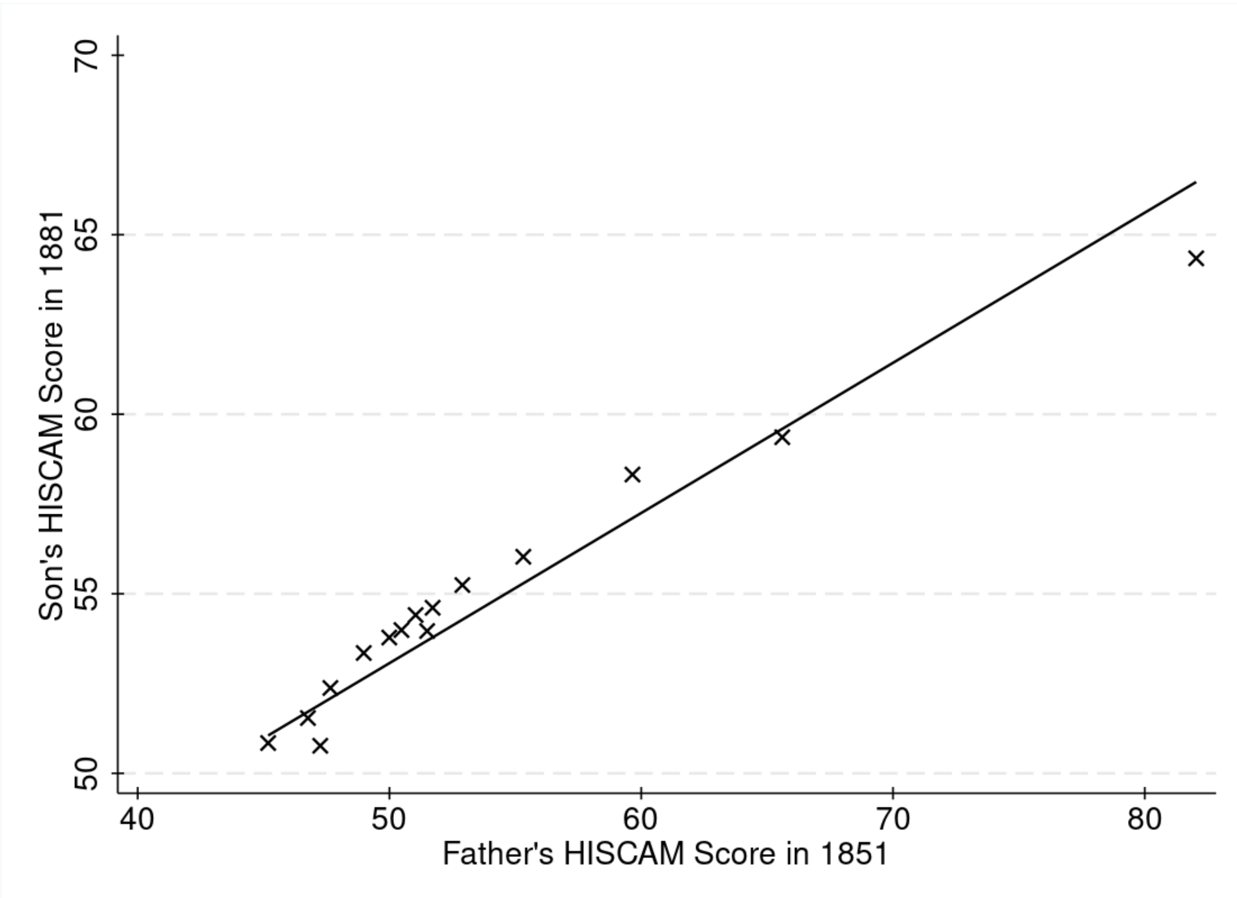
	Population	Unweighted	Weighted
Characteristics (Son) in 1911			
Final Age (mean)	39.67	39.40	39.60
HISCAM (mean)	55.29	55.43	55.74
FOE-RCII (mean)	53.39	54.41	54.67
First name length (mean)	7.88	7.98	7.86
Surname length (mean)	8.34	8.43	8.36
Kids (mean)	2.00	1.98	1.87
Servants (mean)	0.13	0.13	0.12
Characteristics (Father) in 1881			
Initial Age (mean)	40.82	41.26	41.28
HISCAM (mean)	54.31	53.35	53.82
FOE-RCII (mean)	53.13	52.41	52.85
Relationship Status (Son) in 1911			
Head	80.73	84.56	80.11
Son	5.75	8.25	9.78
Visitor	0.71	0.40	0.58
Lodger	1.10	0.31	0.45
Boarder	4.22	2.29	3.44
Marital Status (Son) in 1911			
Single	15.53	13.40	16.82
Married	81.46	84.33	80.00
Occupational Structure (Son) in 1911			
Agriculture	10.30	16.48	9.71
Manufacturing	59.74	56.02	59.86
Services	29.96	27.50	30.43
Residential Region (Son) in 1911			
London	12.77	4.53	6.43
Extra London	13.41	12.56	12.16
Lancashire	24.05	22.51	30.31
Yorkshire	2.74	3.28	2.90
Birth County (Son)			
Greater London	22.13	12.34	12.16
Lancashire	12.12	9.85	13.77
Yorkshire	12.52	14.26	15.88
Observations (N)	2,148,480	164,318	164,318
Match Rate (%)		7.65	7.65

Sources: author's analysis of I-CeM (UKDA, SN 7481) and I-CeM Names and Addresses (UKDA, SN7856). Notes: 'Population' includes all men aged 35-45 in 1911 when comparing with the sons in the linked sample and all men aged 30-55 in 1881 when comparing with the fathers. 'Manufacturing' in Occupational Structure also includes Mining and Transport sectors. 'Extra London' refers to the regions of Middlesex, Kent, Essex, and Surrey that are not included in 'London'. 'Greater London' refers to the entire regions of Middlesex, Kent, Essex, and Surrey. 'Yorkshire' includes all Ridings of Yorkshire. All numbers are in percentages unless they are means.

F Binscatter Plots

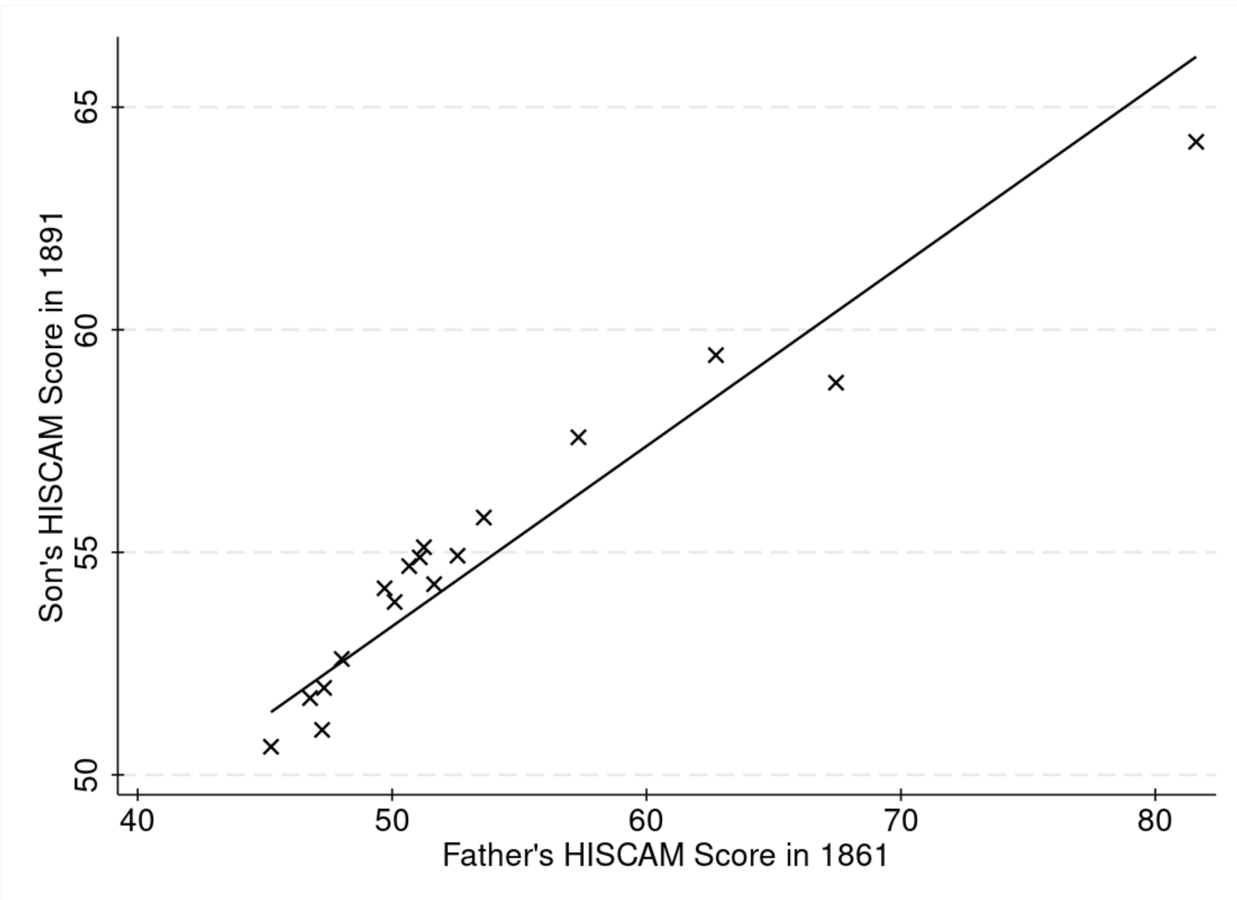
Figures F1, F2, and F3 show the binscatter plots for 1851-1881, 1861-1891, and 1881-1911. The relationship between fathers' and sons' outcomes is clearly linear.

Figure F1: Binscatter plot for 1851-1881



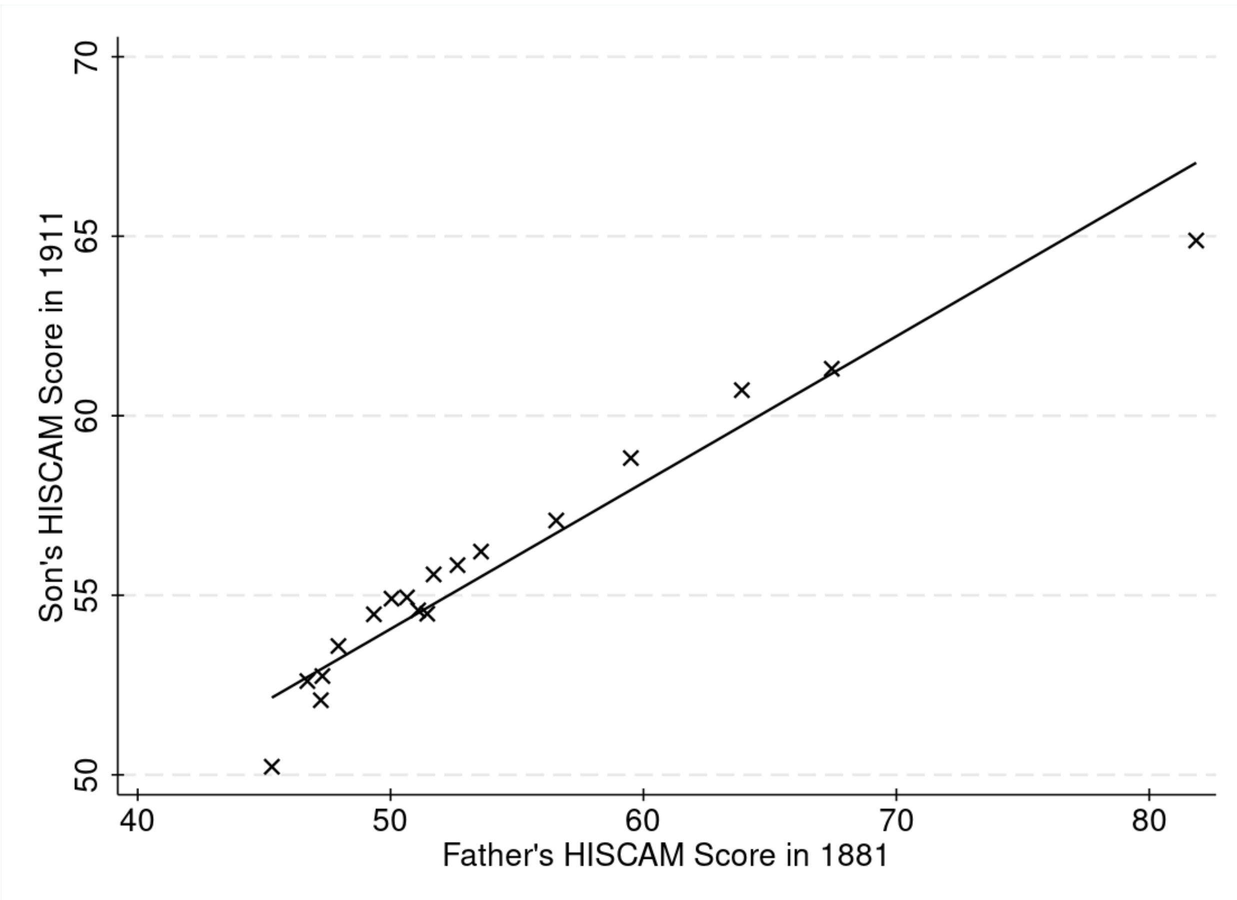
Sources: author's analysis of I-CeM (UKDA, SN 7481) and I-CeM Names and Addresses (UKDA, SN7856).

Figure F2: Binscatter plot for 1861-1891



Sources: author's analysis of I-CeM (UKDA, SN 7481) and I-CeM Names and Addresses (UKDA, SN7856).

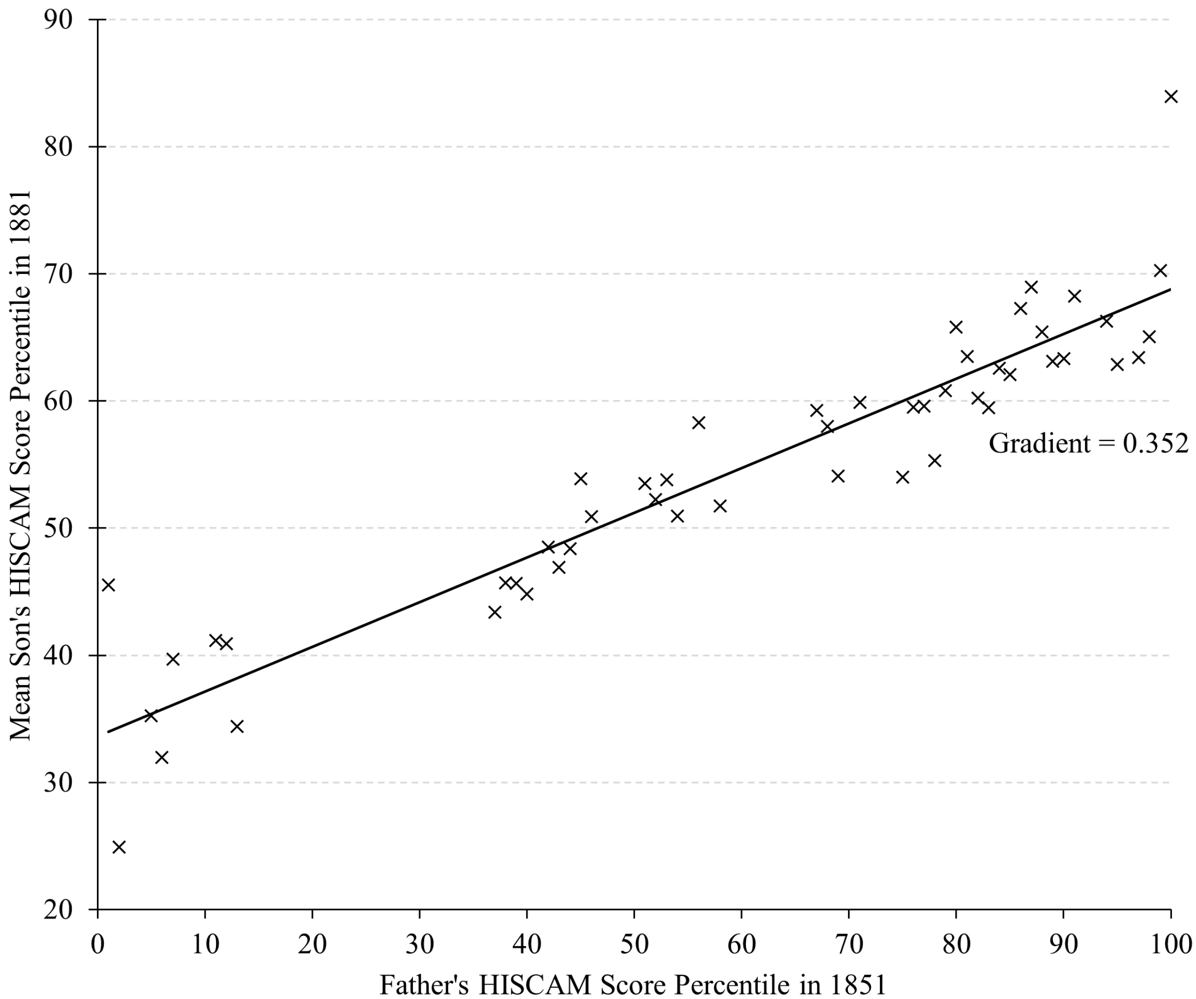
Figure F3: Binscatter plot for 1881-1911



Sources: author's analysis of I-CeM (UKDA, SN 7481) and I-CeM Names and Addresses (UKDA, SN7856).

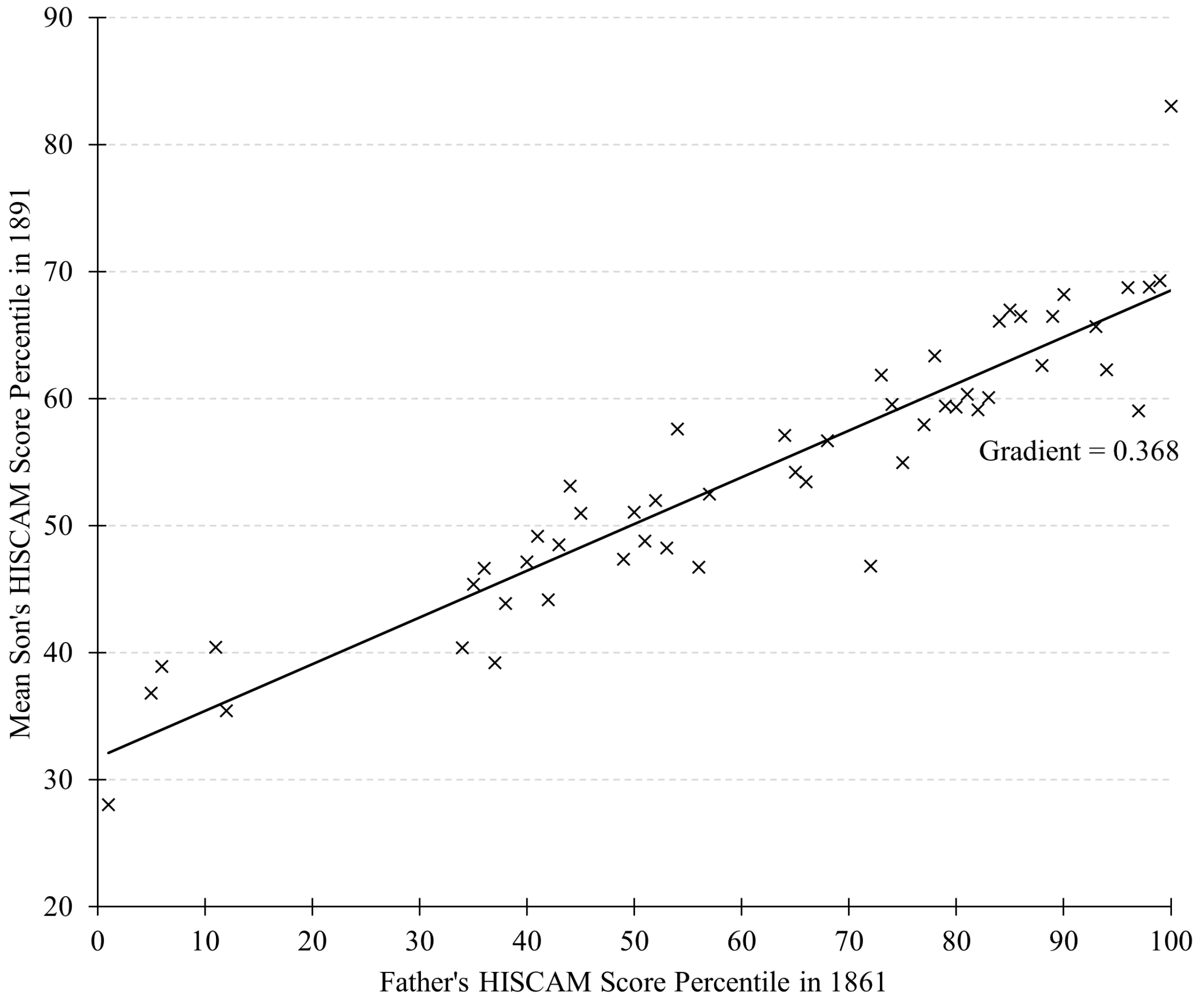
G Rank-Rank Correlations

Figure G1: Father-son rank-rank correlation, 1851-1881



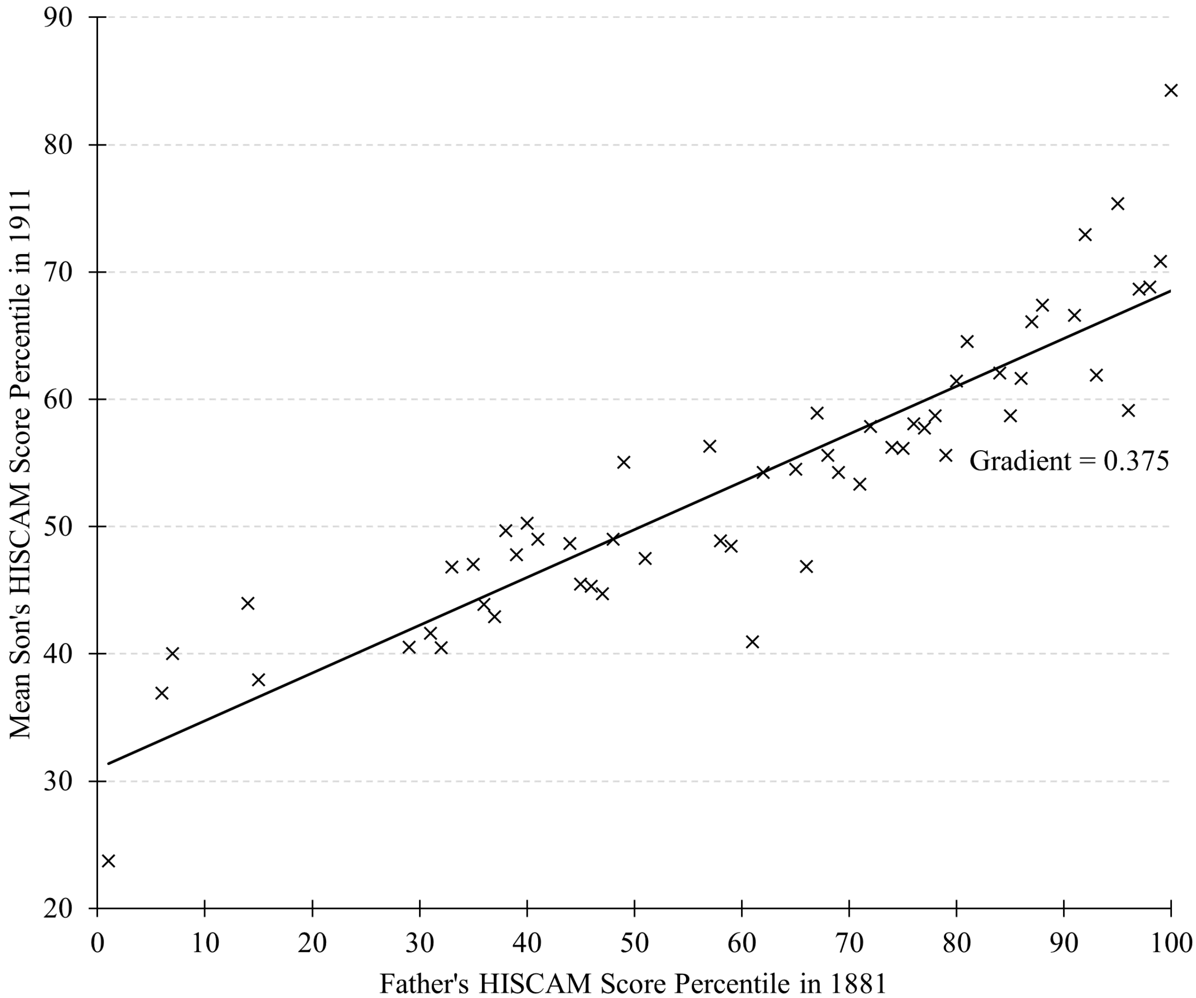
Sources: author's analysis of I-CeM (UKDA, SN 7481) and I-CeM Names and Addresses (UKDA, SN7856).

Figure G2: Father-son rank-rank correlation, 1861-1891



Sources: author's analysis of I-CeM (UKDA, SN 7481) and I-CeM Names and Addresses (UKDA, SN7856).

Figure G3: Father-son rank-rank correlation, 1881-1911



Sources: author's analysis of I-CeM (UKDA, SN 7481) and I-CeM Names and Addresses (UKDA, SN7856).

H Computing Correlations from IGE

In the literature on intergenerational earnings mobility, a standard alternative to IGE (β) is the intergenerational correlation (ρ), which can be calculated by multiplying β with the ratio of the children's and parents' standard deviation (σ) of log earnings (Black and Devereux, 2010):

$$\rho = \beta(\sigma_1/\sigma_0) \tag{H1}$$

Table H1 shows the intergenerational correlations in occupational status calculated using the same formula.

Table H1: Intergenerational Occupational Correlations Computed from Elasticities

	1851-1881		1861-1891		1881-1911	
	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	OLS	OLS	OLS	OLS	OLS
Elasticity	0.402	0.414	0.384	0.405	0.391	0.408
Correlations	0.456	0.480	0.431	0.460	0.432	0.457
Multiple Links	NO	YES	NO	YES	NO	YES
N	257,844	66,965	267,089	84,097	597,517	161,568

Sources: author's analysis of I-CeM (UKDA, SN 7481) and I-CeM Names and Addresses (UKDA, SN7856).

I Life-Cycle Effects on IGE

Additional checks were conducted to determine whether life-cycle effects had an impact on the IGE estimated. The first set of checks – 'age controls' – involve running the OLS and IV regressions with the sons' and fathers' age and their square terms as controls. This had virtually no impact on the size of β. The same was true after narrowing down the age range of the fathers, which meant using only father-son pairs where the fathers were aged 35 to 45 at the start of the period (hence in the similar age range as the sons when their occupational statuses are taken).

Since sons are linked across multiple censuses, it is also possible to estimate an IGE for each stage of their occupational trajectory. Taking the 1881-1911 sample as an example, the 'early-career' β is estimated based on the sons' occupational scores 10 years after their first census year (i.e. 1891); the 'mid-career' β is estimated 20 years after their first census year;

the 'peak' β is the benchmark chosen for this paper – 30 years after their first census year, when the sons are aged 35 to 45.

Table I1: Life-Cycle Effects on β

	OLS			IV		
	β	SE	N	β	SE	N
1851-1881						
Age Controls	0.414	(0.004)	66,965	0.679	(0.007)	65,700
Narrower Father Age Range	0.411	(0.006)	38,317	0.669	(0.010)	37,611
Early-Career	0.392	(0.004)	60,512	0.716	(0.007)	59,526
Mid-Career	No Data – 1871 Census not available					
Peak	0.379	(0.005)	60,512	0.652	(0.008)	59.526
1861-1881						
Age Controls	0.405	(0.004)	84,097	0.647	(0.006)	83,095
Narrower Father Age Range	0.405	(0.005)	47,549	0.659	(0.008)	47,067
Early-Career	No Data – 1871 Census not available					
Mid-Career	0.424	(0.004)	83,163	0.677	(0.006)	82,181
Peak	0.402	(0.004)	83,163	0.646	(0.006)	82,181
1881-1911						
Age Controls	0.408	(0.003)	161,568	0.624	(0.004)	159,723
Narrower Father Age Range	0.406	(0.003)	92,768	0.626	(0.005)	91,988
Early-Career	0.411	(0.002)	151,864	0.666	(0.004)	150,219
Mid-Career	0.385	(0.003)	151,864	0.608	(0.004)	150,219
Peak	0.378	(0.003)	151,864	0.604	(0.004)	150,219

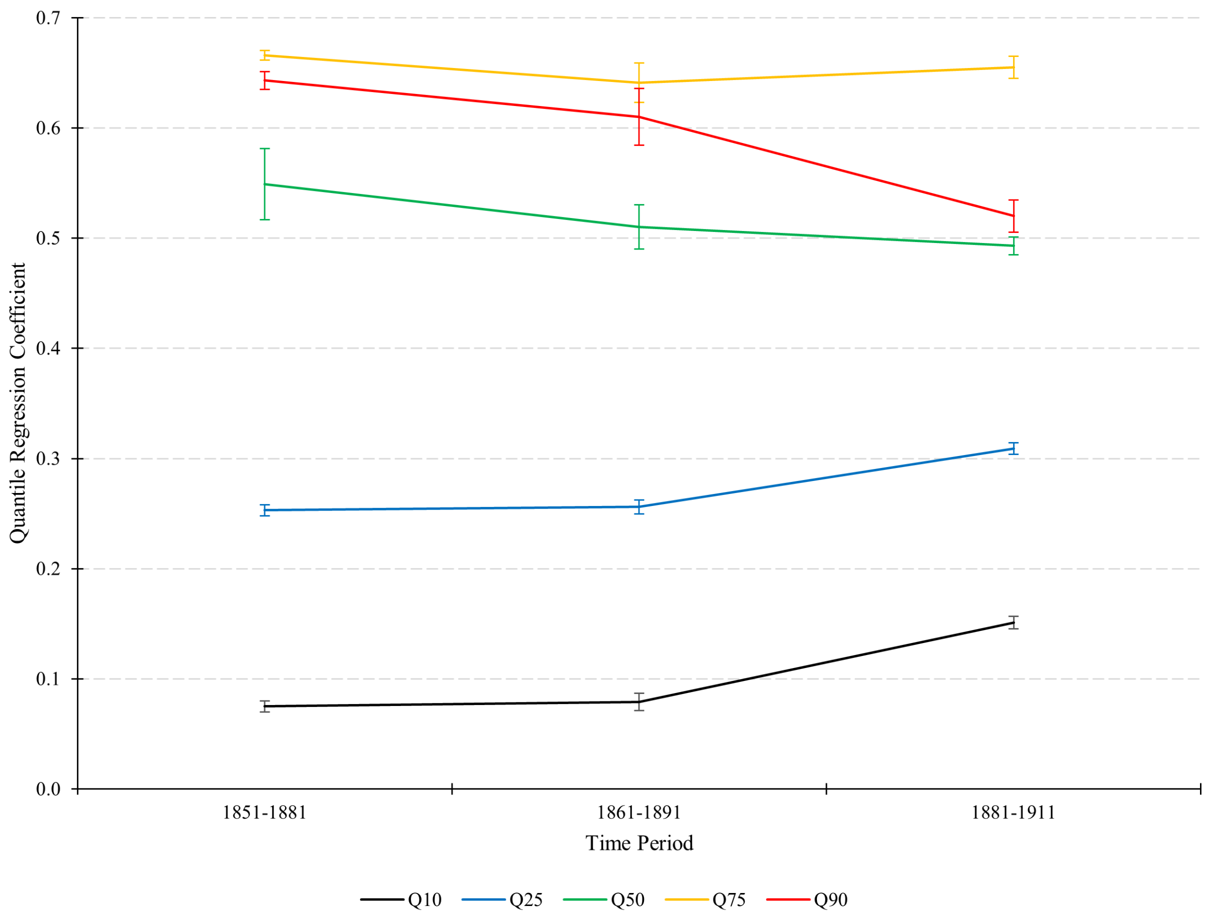
Sources: author's analysis of I-CeM (UKDA, SN 7481) and I-CeM Names and Addresses (UKDA, SN7856).

The results in Table I1 suggests that there may be some modest life-cycle effects depending on the sons' age. Existing findings on life-cycle bias in intergenerational (permanent) income elasticities suggest that using annual incomes from sons at younger ages will lead to an attenuation bias on β while using annual incomes from sons at older ages will lead to an amplification bias (Haider and Solon, 2006). On the contrary, my results suggest that for occupational status, there may be an amplification bias from using the occupations of sons at younger ages, if we take the β estimated at around age 40 as the true level. In any case, any life-cycle bias observed here appear to be modest and there is no indication that my preferred estimates are under-estimating intergenerational mobility due to life-cycle effects.

J Quantile Regression Results

Figure J1 shows the father-son association in occupational status from quantile regressions at the tenth, twenty-fifth, fiftieth, seventy-fifth, and ninetieth percentiles. The results sug-

Figure J1: Quantile regression results, 1851-1911



Sources: author's analysis of I-CeM (UKDA, SN 7481) and I-CeM Names and Addresses (UKDA, SN7856).

gest that the transmission of status is stronger between high-status fathers and sons than between their low-status counterparts. This may be explained by the fact that high-status families have more resources and avenues to protect the socioeconomic status of their future generations.

K Simulation of Minimum IGE

Table K1 shows the mean β and standard errors from 1,000 OLS regressions on samples of randomly matched fathers and sons. For each period, the samples used are the same pool of fathers and sons as the ML linked sample but with random matching of fathers and sons. A total of 1,000 random samples were constructed for each period using this method. The mean β is therefore the minimum level of father-son association possible, and is very close to zero.

Table K1: IGE estimates from randomly matching fathers and sons

	1851-1881	1861-1891	1881-1911
β	0.000086	0.000012	0.000092
SE	(0.004561)	(0.003968)	(0.002825)
N	66,965	84,097	161,568

Sources: author's analysis of I-CeM (UKDA, SN 7481) and I-CeM Names and Addresses (UKDA, SN7856).

L Altham Statistics

A large proportion of the literature on nineteenth- and twentieth-century social mobility relies on an entirely different approach, based on the construction of mobility tables – a two-way contingency table plotting the father's social class against the son's social class (Erikson and Goldthorpe, 1993; Long, 2013; Long and Ferrie, 2013; Pérez, 2019; Antonie et al., 2022). Table L1 shows a simple demonstration of a mobility table. The cells coloured grey contain all the immobile sons – those who held an occupation belonging to the same social class as their fathers at a similar stage in their life cycles. The cells coloured in green contain the upwardly mobile, and the cells in red are the downwardly mobile. Mobility rates can be calculated by aggregating all individuals with the same mobility pattern. For instance, the rate of upward mobility is simply the percentage of all sons in the green cells as a share of the total number of father-son pairs.

However, simply comparing the mobility rates between different mobility tables is not enough to inform us whether one society is more mobile than another. This is because raw mobility rates are affected by the marginal frequencies of the two tables. Thus, it cannot distinguish

Table L1: Example of mobility table

Son's Class in 1881	Father's Social Class in 1851		
	Upper	Middle	Lower
Upper			
Middle			
Lower			

Sources: author's own construction.

whether differences in mobility are caused by the different distributions of occupations in the two mobility regimes, or by the differences in the strength of association between fathers' and sons' outcomes.

One measure that could account for differences in the marginal frequencies between two tables and quantify relative mobility is the Altham statistic, devised by Altham (1970) and coded by Altham and Ferrie (2007). For two tables P and Q with r rows and s columns, the Altham statistic sums the squares of the differences between the natural logarithms of the cross-product ratios in the two tables:

$$d(P,Q) = \left\{ \sum_{i=1}^{r} \sum_{j=1}^{s} \sum_{l=1}^{r} \sum_{m=1}^{s} \left[ln\left(\frac{p_{ij}p_{lm}}{p_{im}p_{lj}}\right) - ln\left(\frac{q_{im}p_{lj}}{q_{ij}q_{lm}}\right)\right]^2 \right\}^{\frac{1}{2}} \tag{L1}$$

Tables with very similar mobility patterns will produce a $d(P,Q)$ value of close to 0, and a very large value if the two tables are very different. The likelihood ratio G^2 statistic with $(r-1)(s-1)$ degrees of freedom is used to establish statistical significance and whether we can accept that $d(P,Q) \neq 0$.

To see which table is more mobile, the same procedure is carried out again to estimate $d(P,I)$ and $d(Q,I)$ where table I is just a matrix of ones, representing complete independence of rows and columns. In other words, $d(P,I)$ and $d(Q,I)$ measures the distance of tables P and Q from perfect mobility. If $d(P,I) > d(Q,I)$ and $d(P,Q) > 0$, relative mobility is greater in table Q than in table P. To correct for measurement errors in Altham statistic, Ward (2021) proposes that only those whose fathers are observed to be in the same class more than once should be kept in the sample.

Finally, to enable the construction of mobility tables, occupations must be arranged into a suitable number of social classes in a hierarchical order. This research uses HISCLASS – an international historical social class scheme based on the Historical International Classification of Occupations codes (HISCO) (Leeuwen et al., 2002; Leeuwen and Maas, 2011). Occupations in HISCLASS are ranked and assorted into twelve classes (with one being the highest) based on four dimensions: manual and non-manual divisions, skill level, degree of supervisory power, and economic sector. These twelve levels can be condensed into smaller schemes with fewer classes. To make comparisons with previous research easier, a four-class scheme will be used. Table L2 describes each of the twelve classes in HISCLASS and how they can be combined into the four-class version.

Table L2: Conversion of HISCLASS Categories to Four-Class Scheme

HISCLASS Description		Four-Class Scheme	
1	Higher managers	W	White-collar
2	Higher professionals	W	White-collar
3	Lower managers	W	White-collar
4	Lower professionals, and clerical and sales personnel	W	White-collar
5	Lower clerical and sales personnel	W	White-collar
6	Foremen	S	Skilled and semi-skilled manual
7	Medium skilled workers	S	Skilled and semi-skilled manual
8	Farmers and fishermen	F	Farming
9	Lower skilled workers	S	Skilled and semi-skilled manual
10	Lower skilled farm workers	U	Unskilled
11	Unskilled workers	U	Unskilled
12	Unskilled farm workers	U	Unskilled

Sources: HISCLASS categories are taken from Leeuwen and Maas (2011); Conversion to four-class scheme follows Antonie et al. (2022).

Table L3 shows the Altham statistics derived from mobility tables for two period, 1851-1881 (Table L4) and 1881-1911 (Table L5). It also includes the results after correcting for measurement errors by including only sons with fathers who are found in the same occupational class in two consecutive census years. Their respective mobility tables are Table L6 and L7.

The Altham statistics confirm that the new sample, constructed using the full-count census data, exhibits less mobility than the sample used previously in Long (2013) and Long and Ferrie (2013)'s works. In addition, the impact of attenuation bias from classical measurement errors is also confirmed by comparing the distance from perfect mobility before and after correcting for measurement errors in the sample – the corrected sample is further away from the matrix of complete independence between rows and columns as expected. However, whereas the IGE estimates show a decline in father-son association over the period, the

Table L3: Summary of Altham statistics, 1851-1911

Time Period	$d(P,I)$	G^2	$d(Q,I)$	G^2	$d(P,Q)$	G^2
1851-1881	23.1	23951***				
1851-1881 (Corrected)			27.2	25140***	4.3	299***
1881-1911			24.5	45627***	4.9	371***
1881-1911 (Corrected)			29.4	50512***	5.1	783***
1851-1881 (L & F, 2013)			20.8	800***		

Sources: author's analysis of I-CeM (UKDA, SN 7481) and I-CeM Names and Addresses (UKDA, SN7856) and Long and Ferrie (2013). Notes: the 'Corrected' series are estimates which have been corrected for measurement errors using Ward (2021)'s approach. L & F refers to Long and Ferrie (2013).

Altham statistics indicate that social mobility was decreasing between 1851-1881 and 1881-1911, though again, the difference is quite small.

Part of the reasons why the IGE estimates and the Altham statistics disagree on the trend over time is due to the differences in the underlying measures of occupational status. As mobility tables are constructed based on just four classes of occupations, a lot of within-class mobility could be missed out. On the other hand, the IGE is computed using HISCAM scores, which better captures the differences in socioeconomic status associated with occupations belonging to the same sector.

Based on the results, we tentatively conclude that over the course of the Victorian and Edwardian era, social/occupational classes may be becoming more rigid (fewer people moving across big occupational boundaries, i.e. skilled workers to white-collar) but within-class intergenerational occupational mobility may have increased. Yet, it remains to be said, the magnitude of these developments in social mobility was undoubtedly small, and it does not in any way alter the overall message from this paper; intergenerational mobility in nineteenth- and early-twentieth century is at odds with the optimistic depiction of Victorian society as one of openness and opportunity.

Table L4: Intergenerational Mobility Rates, 1851-1881

Son's Class 1881	Father's Class in 1851 W	S	F	U	Total	*N*
W	45.39	16.81	17.97	11.73	22.97	12,291
S	37.85	69.55	19.01	32.99	39.85	32,598
F	4.19	2.11	40.59	2.20	12.27	3,952
U	12.57	11.53	22.43	53.08	24.90	18,124
Total	100.00	100.00	100.00	100.00		
N	7,501	30,086	6,139	23,239		66,965

Sources: author's analysis of I-CeM (UKDA, SN 7481) and I-CeM Names and Addresses (UKDA, SN7856).

Table L5: Intergenerational Mobility Rates, 1881-1911

Son's Class 1911	Father's Class in 1881 W	S	F	U	Total	*N*
W	53.76	21.51	20.02	15.68	27.74	40,417
S	33.67	64.98	18.95	36.36	38.49	79,544
F	2.33	1.04	38.43	2.36	11.04	6,137
U	10.24	12.46	22.60	45.61	22.73	35,470
Total	100.00	100.00	100.00	100.00		
N	26,153	80,855	9,420	45,140		161,568

Sources: author's analysis of I-CeM (UKDA, SN 7481) and I-CeM Names and Addresses (UKDA, SN7856).

Table L6: Intergenerational Mobility Rates (Corrected), 1851-1881

Son's Class 1881	Father's Class in 1851 W	S	F	U	Total	*N*
W	55.23	15.49	16.85	10.70	24.57	9,206
S	31.79	72.41	15.20	30.44	37.46	25,776
F	3.02	1.58	45.96	1.55	13.03	3,147
U	9.96	10.52	21.99	57.31	24.94	14,963
Total	100.00	100.00	100.00	100.00		
N	4,568	24,599	5,067	18,858		53,092

Sources: author's analysis of I-CeM (UKDA, SN 7481) and I-CeM Names and Addresses (UKDA, SN7856).

Table L7: Intergenerational Mobility Rates (Corrected), 1881-1911

Son's Class 1911	Father's Class in 1881				Total	*N*
	W	S	F	U		
W	61.35	20.48	17.02	14.51	28.34	31,795
S	29.25	67.73	13.19	33.69	35.96	63,844
F	1.51	0.77	46.26	1.72	12.56	4,803
U	7.89	11.02	23.53	50.08	23.13	28,355
Total	100.00	100.00	100.00	100.00		
N	19,045	66,983	7,331	35,438		128,797

Sources: author's analysis of I-CeM (UKDA, SN 7481) and I-CeM Names and Addresses (UKDA, SN7856).