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

Ziming Zhu

This paper uses a new linked sample constructed from full-count census data of 1851–1911 to revise estimates of intergenerational occupational mobility in England. I find that conventional estimates of intergenerational elasticities are attenuated by classical measurement error and 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 ($\beta$) of occupational status from 0.4 to 0.6–0.7. Victorian England was therefore a society of limited social mobility. The long-run evolution and international comparisons of social mobility in England are discussed.

Social mobility—the movement of individuals between social groups between generations or across the lifetime—is a subject that 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. 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, Adams (1931) in *The Epic of America* 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 right to extol nineteenth-century English society as one of openness and low barriers? Or were opportunities few

Ziming Zhu, LSE Fellow, Department of Economic History, London School of Economics and Political Science, London, WC2A 2AE, United Kingdom. E-mail: z.zhu11@lse.ac.uk.

and far between? Using a newly constructed and improved set of linked data featuring between 67,000 and 160,000 father-son pairs 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 inter-generational 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 error causes 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 percent.

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 surname-based measures (Clark and Cummins 2015). Long (2013) was the first to estimate rates of social mobility using linked census data for England.[1] However, the surprisingly high rate of mobility he found may not be a true reflection of the state of nineteenth-century English society. Ward's (2023) research on historical mobility in the United States highlights the issue of measurement error in mobility studies. In addition, there are limitations, not least in terms of Long's (2013) sample size, in his use of a 2 percent sample of the 1851 census, while questions remain about false positives in census linking causing significant attenuation bias, which could lead us to conclude that mobility was far greater than what it was in reality (Bailey et al. 2020; Anbinder et al. 2021). Such concerns are addressed in this paper.

Although this is not the first time English historical mobility has been estimated, this paper makes three important contributions to the literature. First, it provides revised intergenerational elasticities of occupational status for Victorian and Edwardian England after accounting for classical measurement error.[2] Such errors arise because occupations in historical censuses are measured with noise (in the form of data errors or transitory shocks); this leads to attenuation bias in the IGE estimated and an over-estimation of the extent of social mobility. Second, it constructs a high-quality linked sample using the Integrated Census Microdata (I-CeM) complete-count census data, which greatly expands the number of obser-vations that were previously available to Long (2013). Finally, it devises a new method for estimating the rate of, and consequently correcting for (at

---

[1] Long and Ferrie use the same data and linkage methods for Britain (Long and Ferrie 2013). Later works involving the England and Wales censuses likewise only used a sample rather than the full 1851 census (Long and Ferrie 2018; Pérez 2019).

[2] The results are robust to alternative methods of census linkage and different occupational indices. False positives and reweighting do not have a significant impact on my findings.

least partially), false positives in census linking without the prerequisite of possessing a highly reliable, hand-linked reference dataset.

The rest of the paper is organized as follows. The second section reviews the existing literature on historical social mobility. The third section presents the data used and the census linking process and outcomes. The fourth section outlines the methodology, or how social (occupational) mobility is measured in this paper. The results are shown in the fifth section; they represent a significant revision from previous works and highlight the impact of measurement error. The sixth section discusses the implications of these results and makes some comparisons across both time and space. The seventh section concludes.

## SOCIAL MOBILITY IN VICTORIAN ENGLAND

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, economic, and political changes and reforms.

Victorian England was the outcome of one of the most transformative events in economic history—the Industrial Revolution. Yet, even though the "revolution" was well past its most tempestuous stage by 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, with the rise of clerical workers, post offices, and bureaucratic organizations (Thomas 2004). In addition, a number of other social changes were taking place during this period. The country was becoming more urbanized, 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 (un) availability of individual-level sources that include variables that 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 percent, 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 both total and upward mobility due to the age differences between fathers and sons. To illustrate, at the time of marriage, the father, aged 50, is at the peak of his career while the son, aged 25, has only started working. If both father and son are found to have the same occupation on the marriage register, we may mistake it for no mobility when, in fact, the son may have a better occupation than his father when he reaches 50. Moreover, we might overstate the degree of downward mobility if we find the son to be of a lower occupational status than his father at the time of marriage without accounting for the fact that the son has not had the same amount of time to develop his career.

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 the degree of social mobility in England remained largely unchanged from the mid-nineteenth to the twenty-first century. However, rather than characterizing Victorian England as a mobile society, they conclude, based on the high levels of persistence in the socioeconomic status of surnames, that England was and still is a society in which one's own achievements can largely be determined at birth by the virtues of their name.

There are several reasons why the surprisingly high rate of mobility found by Long (2013) 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 percent sample of the 1851 census. This raises issues of representativeness while also increasing the likelihood of Type I errors in linking.[3]

---

[3] 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 percent sample, some non-unique individuals may appear as unique if their duplicates are eliminated by the process of sampling. To demonstrate this possibility, I have tried to link people using a 2 percent sample rather than the full census for the initial year. The results for this are shown in Online Appendix L.

Moreover, Bailey et al. (2020) and Anbinder et al. (2021) both emphasized 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.[4] Bailey et al. were also skeptical of the use of phonetic names in linking algorithms—the strategy that Long (2013) used in his linking.

The issue of classical measurement error is another factor that could lead to significant attenuation bias. There are two potential sources of measurement error. The main source of error is the misreporting of occupations. Inferring socioeconomic status from occupations from historical censuses is subject to measurement error because occupations are sometimes 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 digitizing the data.[5] For example, Ward (2023) exploits the re-enumeration of St. Louis in the United States in 1880 to show that across two censuses conducted on the same population in the same year, over 30 percent of occupations may have been misreported. A second but perhaps less likely source of measurement error would be transitory shocks to a person's status. Occupational status, particularly in the past, could be unstable and transitory, and people could be affected by temporary shocks to their labor market outcomes, which they may recover from a few years later (such as before the next census). Thus, the occupation observed in one census year may not be an accurate reflection of one's true socioeconomic status.

One way of correcting for the attenuation bias caused by measurement error is through an instrumental variable (IV) approach. Solon (1992) demonstrated the effectiveness of this approach in the modern context by instrumenting fathers' incomes with their educational outcomes. However, when there is no second measure of the same person's socioeconomic status available (as often is the case with historical censuses), Ward (2023) proposes that measurement error can also be corrected by instrumenting the father's occupation observed in one census with his occupation observed in another census. This should reduce the attenuation bias caused by measurement error and lead to a significant upward revision of the IGE.

---

[4] For instance, Bailey et al. (2020) estimate that false links could bias IGE downward by 30 percent or more.

[5] For a detailed explanation of the census-taking procedure in Britain between 1851 and 1911, see Online Appendix A. For more details on the changes (or similarity) of occupations reported for fathers across two consecutive censuses, see Online Appendix M.

After accounting for measurement error, Ward (2023) finds that the revised IGE estimates for the United States 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 United States was hence less mobile than modern-day United States. This represents a significant departure from the existing consensus that posits a decline in intergenerational mobility in the United States since the nineteenth century (Long and Ferrie 2013; Song et al. 2020). Therefore, our understanding of British/English occupational mobility since the Victorian era may be open to question too. In addition, past research 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 United States. This could also be subject to amendment if the effects of classical measurement error are different across countries.

## DATA AND CENSUS LINKING

### *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 anonymized information from the British decennial censuses between 1851 and 1911 (except for 1871)—compiled and published by Schürer 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 anonymization. This information is necessary to conduct record linkage.

The censuses of 1851 to 1911 recorded all the vital information that is needed for occupational mobility research, specifically name, age, sex, place of birth, and occupation, with reasonable reliability. This information was then transcribed and enriched by the I-CeM project via a computer program.[6] 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

---

[6] 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 standardizing the data; and adding a number of enriched variables, mainly relating to household structure.

with that published in the *Census Report* by the General Register Office (Higgs et al. 2013).[7]

The most significant undertaking of I-CeM is the standardization 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 standardized to the parish level. Naturally, the automatic coding of this vast number of occupational strings will introduce errors, leading to some occupations being miscoded. Higgs et al. (2013) assert that for at least 95 percent of individuals with an occupation title, the coding is "correct." Other variables, such as marital status and household relationships, have also been standardized, coded, and checked for consistency.

*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 incumbent 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 such that when they are applied to the sample used in the construction of HISCAM, they should have 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 from these countries, is used. Table 1 shows a sample

---

[7] This occurs when the number of people found in a particular place for a given year in the raw I-CeM data is inconsistent with the population total for that said place published in the *Census Report* in that year.

TABLE 1
SAMPLE OF OCCUPATIONS WITH HISCAM SCORES

| OCCODE | Occupation Description | HISCAM |
|---|---|---|
| 84 | Other domestic indoor servants – undefined | 39.90 |
| 196 | Coal miners – hewers, workers at the coal face | 45.59 |
| 765 | General laborers | 46.84 |
| 132 | Railway laborer (not railway contractor's laborer) | 46.84 |
| 181 | Agricultural laborer, farm servant | 47.26 |
| 702 | Sugar planter grower | 50.13 |
| 653 | Tailors (not merchants) – default | 50.81 |
| 723 | Gas works service | 51.08 |
| 11 | Police | 52.38 |
| 287 | Electricians (undefined) | 55.63 |
| 347 | Fitters (ships) | 58.68 |
| 536 | Printers | 60.25 |
| 1 | Post Office – telegraphists, telephone operators | 63.09 |
| 405 | Builders | 63.47 |
| 52 | Schoolmasters and teachers (default) | 67.45 |
| 119 | Commercial or business clerks | 67.91 |
| 120 | Bankers | 88.22 |
| 65 | Civil engineers | 91.20 |
| 5 | MPs, ministers of the Crown & Peers | 99.00 |
| 38 | Barrister, advocate | 99.00 |

*Notes*: "OCCODE" is the numeric code for occupational groupings in the I-CeM Occupational Matrix.
*Sources*: "OCCODE" and "Occupation description" come from I-CeM (Schürer and Higgs 2014, UKDA, SN 7481); "HISCAM" is taken from Lambert et al. (2013).

of some common occupations observed in the census with their respective HISCAM scores.

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 CCC index constructed by Clark, Cummins, and Curtis (2023), 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 to create this index is the same as the one used by Lambert et al. (2013) for HISCAM. Using information from marriages, Clark, Cummins, and Curtis (2023) 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 (MP) and a miner, will have very few social interactions between them (in other words, very few sons of MPs marry daughters of miners), thus they will be given vastly different scores. On the other hand, many marriages occur between

bank clerks' and teachers' sons and daughters, so they are given similar scores. Again, the scores are between 0 and 100, with higher scores representing higher status.[8]

Finally, a prerequisite for calculating the Altham statistics—an alternative way of estimating social mobility employed in this paper and by many others in the literature—is to arrange occupations into a suitable number of social classes in a hierarchical order.[9] This research uses HISCLASS—an international historical social class scheme based on the Historical International Classification of Occupations codes (HISCO) (van Leeuwen et al. 2002; van Leeuwen and Maas 2011). Occupations in HISCLASS are ranked and assorted into 12 classes (with 1 being the highest) based on 4 dimensions: manual and non-manual divisions, skill level, degree of supervisory power, and economic sector. These 12 levels can be condensed into smaller schemes with fewer classes. To make comparisons with previous research easier, a four-class scheme will be used.[10] Table 2 describes each of the 12 classes in HISCLASS and how they can be combined into the 4-class occupational categories, as shown by Antonie et al. (2022).

### Census Linking Procedure

To conduct record linkage across the censuses, this project selects English-born sons 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 error, the sons are linked across every 10-year interval and the fathers are linked across one 10-year interval.[11] This is done for three periods: 1851–1881, 1861–1891, and 1881–1911.

Historical census record linkage is a complicated process due to the lack of a unique identifier like a Social Security Number across datasets.

---

[8] They also construct a different index using an alternative methodology—principal component analysis. Clark, Cummins, and Curtis (2023) find that, reassuringly, HISCAM is very effective at capturing socioeconomic status. All their indices show a strong association with HISCAM.

[9] For more discussion of the Altham statistics, see "Measuring Mobility Using Altham Statistics" in the "Methodology" section.

[10] The same classification scheme was used by Long and Ferrie (2013) and Pérez (2019).

[11] 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.

TABLE 2
HISCLASS LEVELS AND OCCUPATIONAL CATEGORIES

| Levels | HISCLASS Levels' Description | Class | Categories |
|---|---|---|---|
| 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 levels and descriptions are taken from van Leeuwen and Maas (2011); conversion to four-class occupational categories follows Antonie et al. (2022).

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 minimizing these two types of errors. Choosing an algorithm that eliminates as many false positives as possible while still achieving a satisfactory match rate is crucial for automated record linking (Ruggles, Fitch, and Roberts 2018).

This paper adopts a prominent automated census linkage technique developed by Abramitzky, Boustan, and Eriksson (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.[12] The procedure is outlined in Online Appendix B. This paper opts for the more conservative approach in matching, which minimizes 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 is motivated by the findings of Bailey et al. (2020), who reviewed a number of prominent

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

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.[13] They concluded that reducing false matches is more important than generating a higher match rate for improving inferences with linked data, as 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.

Since the use of phonetic names in census linking has come under criticism for the high rate of false positives produced when attempting to link Irish immigrants in the United States across the American censuses (Anbinder et al. 2021), this paper opts for matching using string distances by adopting the Jaro-Winkler version of the ABE methodology. Moreover, 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 in separate matching processes. For example, I match both sons and fathers from 1851 to 1861 and then identify sons who are found to be living with their fathers in both years. Then I can compare if the fathers I matched through census linking in 1861 are the same people as the ones co-residing with the sons in the census. The detailed procedure and results are outlined in Online Appendix C. The benefit of this way of testing for false positives is that unlike the conventional method of benchmarking a linked sample against a high-quality dataset (Bailey et al. 2020; Abramitzky et al. 2020; Anbinder et al. 2021), which is rare to find given the historical nature, a double-linked sample is much more accessible.

There are *a priori* reasons to believe that false matches may be less of an issue with linking British censuses. While the U.S. data lacked detailed birthplace information, such that Abramitzky, Boustan, and Eriksson (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 standardized parish of birth. The latter was also not available to Long (2013), so they were not able to address the issue of some parishes

---

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

having multiple or changing names. Moreover, Anbinder et al. (2021) recognized 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. (2020) 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—younger people are more likely to emigrate, while the rate of mortality increases with age.

However, the impact of a non-representative sample may be less significant than false positives. Bailey et al. (2020) show that reweighting the sample by inverse probability can effectively address the issue of sample selection bias.[14] They also suggest that after removing the incorrect links, reweighting makes little difference. Abramitzky et al. (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.

*Census Linking Outcomes*

Table 3 shows the summary statistics for the baseline and the multiple-link samples 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 percent. 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 percent. This still generates between 68,000 to 160,000 father-son pairs—a huge

---

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

TABLE 3
SUMMARY STATISTICS OF LINKAGE RESULTS, 1851–1911

| | 1851–1881 | | | 1861–1891 | | | 1881–1911 | | |
|---|---|---|---|---|---|---|---|---|---|
| | Population | Baseline | ML | Population | Baseline | ML | Population | Baseline | ML |
| **Characteristics (Son) in 1881/1891/1911** | | | | | | | | | |
| Final age (mean) | 39.68 | 39.44 | 39.12 | 39.65 | 39.43 | 39.53 | 39.67 | 39.48 | 39.40 |
| HISCAM (mean) | 54.38 | 54.14 | 53.95 | 54.42 | 54.33 | 54.40 | 55.29 | 55.31 | 55.43 |
| CCC (mean) | 53.13 | 53.26 | 53.51 | 52.72 | 52.98 | 53.66 | 53.39 | 53.65 | 54.41 |
| First name length (mean) | 6.27 | 6.32 | 6.30 | 7.80 | 7.83 | 7.84 | 7.88 | 7.97 | 7.98 |
| Surname length (mean) | 6.33 | 6.36 | 6.40 | 8.31 | 8.40 | 8.41 | 8.34 | 8.39 | 8.43 |
| Kids (mean) | 2.85 | 3.03 | 3.03 | 2.46 | 2.57 | 2.64 | 2.00 | 2.00 | 1.98 |
| Servants (mean) | 0.22 | 0.22 | 0.22 | 0.19 | 0.20 | 0.21 | 0.13 | 0.13 | 0.13 |
| **Characteristics (Father) in 1851/1861/1881** | | | | | | | | | |
| Initial age (mean) | 40.81 | 41.79 | 41.42 | 40.96 | 41.83 | 42.05 | 40.82 | 41.52 | 41.26 |
| HISCAM (mean) | 53.35 | 52.32 | 52.07 | 53.52 | 52.58 | 52.51 | 54.31 | 53.50 | 53.35 |
| CCC (mean) | 53.38 | 52.25 | 52.61 | 53.06 | 52.08 | 52.61 | 53.13 | 51.97 | 52.41 |
| **Relationship Status (Son) in 1881/1891/1911** | | | | | | | | | |
| Head | 82.75 | 86.14 | 86.03 | 84.63 | 86.99 | 88.08 | 80.73 | 83.43 | 84.56 |
| Son | 4.26 | 5.58 | 6.61 | 4.40 | 5.70 | 5.98 | 5.75 | 7.26 | 8.25 |
| Visitor | 0.63 | 0.40 | 0.38 | 0.56 | 0.37 | 0.30 | 0.71 | 0.49 | 0.40 |
| Lodger | 3.30 | 1.83 | 1.54 | 2.84 | 1.51 | 1.16 | 1.10 | 0.48 | 0.31 |
| Boarder | 2.39 | 1.42 | 1.27 | 2.29 | 1.44 | 1.06 | 4.22 | 2.92 | 2.29 |
| **Marital Status (Son) in 1881/1891/1911** | | | | | | | | | |
| Single | 12.33 | 10.67 | 11.17 | 12.44 | 11.01 | 10.35 | 15.53 | 13.99 | 13.40 |
| Married | 83.57 | 85.81 | 85.50 | 84.08 | 85.94 | 86.75 | 81.46 | 83.54 | 84.33 |
| **Occupational Structure (Son) in 1881/1891/1911** | | | | | | | | | |
| Agriculture | 15.71 | 19.60 | 23.76 | 13.26 | 16.56 | 20.25 | 10.30 | 12.71 | 16.48 |
| Manufacturing | 60.19 | 57.96 | 56.29 | 62.03 | 59.84 | 57.52 | 59.74 | 57.74 | 56.02 |
| Services | 24.10 | 22.44 | 19.95 | 24.71 | 23.61 | 22.23 | 29.96 | 29.55 | 27.50 |

TABLE 3 (CONTINUED)
SUMMARY STATISTICS OF LINKAGE RESULTS, 1851–1911

| | 1851–1881 | | | 1861–1891 | | | 1881–1911 | | |
|---|---|---|---|---|---|---|---|---|---|
| | Population | Baseline | ML | Population | Baseline | ML | Population | Baseline | ML |
| Residential Region (Son) in 1881/1891/1911 | | | | | | | | | |
| London | 15.91 | 11.49 | 7.74 | 15.61 | 9.35 | 6.91 | 12.77 | 7.49 | 4.53 |
| Extra London | 8.90 | 9.62 | 8.06 | 9.95 | 10.20 | 10.79 | 13.41 | 13.27 | 12.56 |
| Lancashire | 13.32 | 10.30 | 8.17 | 13.92 | 11.69 | 8.96 | 13.62 | 12.48 | 11.49 |
| Yorkshire | 12.68 | 13.60 | 15.22 | 12.85 | 14.41 | 15.26 | 13.25 | 14.13 | 14.51 |
| Birth County (Son) | | | | | | | | | |
| Greater London | 18.15 | 15.48 | 10.07 | 19.75 | 13.58 | 12.64 | 22.13 | 15.78 | 12.34 |
| Lancashire | 10.96 | 8.31 | 6.53 | 11.29 | 9.23 | 6.89 | 12.12 | 11.25 | 9.85 |
| Yorkshire | 12.15 | 12.87 | 14.42 | 11.52 | 12.99 | 14.40 | 12.52 | 13.40 | 14.26 |
| Observations (N) | 1,291,487 | 293,889 | 68,329 | 1,445,779 | 311,119 | 86,884 | 2,148,480 | 612,481 | 164,318 |
| Match rate (%) | | 22.76 | 5.29 | | 21.52 | 6.01 | | 28.51 | 7.65 |

*Notes:* "Population" includes all men aged 35–45 in 1881, 1891, and 1911 when comparing with the sons in the linked sample and all men aged 30–55 in 1851, 1861, and 1881 when comparing with the fathers; "ML" refers to the sample where sons are double- or triple-linked and fathers are double-linked; "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 stated otherwise.

*Sources:* Author's analysis of I-CeM (Schürer and Higgs 2014, UKDA, SN 7481) and I-CeM Names and Addresses (Schürer and Higgs 2015, UKDA, SN 7856).

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 suggests that both the baseline and the multiple links samples are very representative of the full population. In terms of occupational status—measured by HISCAM and CCC—and age, both the sons and their fathers show negligible differences from the wider population. The same is true for the sons' first and last name lengths, and the number of kids and servants they have.

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 populations. 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. (2020) demonstrated, these issues can be corrected using inverse probability weights (see Online Appendix D for more detail), and later results will show that reweighting does not change the results significantly.

## METHODOLOGY

### *Calculating Intergenerational Elasticity (IGE)*

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 the father's outcome ($Y_{i,t-1}$):

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

where $\alpha$ is the constant, $\varepsilon_{i,t}$ is a set of random factors, and the coefficient of interest is $\beta$, 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 $\beta$ toward 0:

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

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

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

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 research on historical mobility is 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 ($\varepsilon_{i,t}$ and $\mu_{i,t}$) observed are uncorrelated across different observations:

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

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

Both methods for correcting measurement error (averaging across multiple short-run observations or IV) have been implemented for modern-day studies (for instance, by Altonji and Dunn (1991), Solon (1992), Zimmerman (1992) in the U.S. context, and Dearden, Machin, and Reed (1997), Grawe (2004) for the British context) and more recently, for historical studies by Ward (2023). The instrumental variables approach is shown to work as well, if not better than averaging across three father's observations (Ward 2023). 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.

This may seem to be an unusual use of the IV method, given that the purpose of using the instrument is not for causal identification. However, there is an established tradition of using IV methods to correct for measurement error. Fuller (1987) outlined that where the independent variable $x_t$ is measured with errors, we can correct for the attenuation bias caused by such errors using an instrument $W_t$; a possible choice for $W_t$ is a measurement of $x_t$ obtained by an independent method. Indeed, this is the approach taken by Solon (1992), who used a father's years of education as an instrument for a father's earnings in a single year. Ward (2023) adapts this approach to the nineteenth century by using the father's occupational status measured in a different year as an instrument for the father's occupational status observed in one year.

The validity of such an instrument lies in the fact that it provides additional information for measuring our independent variable, the father's true socioeconomic status. Though this second measure of the father's status may produce additional measurement error, as long as these errors are uncorrelated with each other—a standard assumption in the literature—the IV estimator will remain consistent (Solon 1992; Modalsli and Vosters 2019; Ward 2023).

A potential limitation to this strategy is that the instruments available are often endogenous. In Solon's (1992) case, the father's education may be positively but imperfectly correlated with the son's status, and in this paper, a father's occupation in a second census may also be positively correlated with the son's future occupational status. If this was the case, then the IV estimator will be upward-inconsistent, so the IGE obtained using the IV approach becomes an upper-bound estimate for the true level of father-son association in status, and the OLS estimate becomes a lower-bound since it is downward-inconsistent (Solon 1992; Mitnik 2020).

Another 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 $\beta$ 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 son's 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 is quite similar (see Online Appendix H), so life-cycle effects are not significant enough to cast doubts on the results and their interpretations.

*Measuring Mobility Using Altham Statistics*

Several papers in the literature on social mobility in the nineteenth and twentieth centuries 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 (Long 2013; Long and Ferrie 2013; Pérez 2019; Antonie et al. 2022). The diagonals in the table represent the number or share of sons who do not show mobility— those who held an occupation belonging to the same social class as their fathers at a similar stage in their life cycles. The cells above the diagonals contain the upwardly mobile, and the cells below the diagonals contain 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 upwardly mobile sons 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 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 into Stata 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:

$$d(P,Q) = \left\{ \sum_{i=1}^{r} \sum_{j=1}^{s} \sum_{l=1}^{r} \sum_{m=1}^{s} \left[ \ln\left( \frac{p_{ij}p_{lm}}{p_{im}p_{lj}} \right) - \ln\left( \frac{q_{im}q_{lj}}{q_{ij}q_{lm}} \right) \right]^2 \right\}^{1/2} \tag{6}$$

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

TABLE 4
INTERGENERATIONAL ELASTICITIES OF OCCUPATIONAL STATUS (HISCAM),
1851–1911

| | 1851–1881 | | | 1861–1891 | | | 1881–1911 | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | (1) OLS | (2) OLS | (3) IV | (4) OLS | (5) OLS | (6) IV | (7) OLS | (8) OLS | (9) IV |
| $\beta$ | 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 |

*Notes*: Standard errors in parenthesis; all estimates are statistically significant to p<0.01; "ML" stands for Multiple Links and denotes whether sons and fathers can be linked across multiple censuses.
*Sources*: Author's analysis of I-CeM (Schürer and Higgs 2014, UKDA, SN 7481) and I-CeM Names and Addresses (Schürer and Higgs 2015, UKDA, SN 7856).

words, $d(P,I)$ and $d(Q,I)$ measure 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 error in Altham statistics, Ward (2023) proposes that only those whose fathers are observed to be in the same class more than once should be kept in the sample.

EMPIRICAL RESULTS

*Main Results—IGE Estimates*

Table 4 illustrates the main findings of this paper. The IGE of log occupational status for the baseline sample is shown in Columns (1), (4), and (7) for the periods 1851–1881, 1861–1891, and 1881–1911.[15] The OLS estimates of the $\beta$ for the sample with multiple links (where sons can be linked across multiple censuses) are shown in Columns (2), (5), and (8).[16] Standard errors are shown in parenthesis; all estimates are statistically significant at the 0.01 level. The $\beta$ for the sample with multiple links is slightly higher than the $\beta$ 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.

[15] All results in the main paper and the appendix are replicable using the replication package by Zhu (2023) deposited on OpenICPSR.
[16] Binned scatter plots for the relationship between father's and son's HISCAM scores are shown in Online Appendix E. They demonstrate that the relationship is clearly linear.

TABLE 5
DETAILED IV RESULTS (WITH FIRST STAGE), 1851–1911

| | 1851–1881 | | 1861–1891 | | 1881–1911 | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| | First Stage Father's Log Score in 1851 | Second Stage Son's Log Score in 1881 | First Stage Father's Log Score in 1861 | Second Stage Son's Log Score in 1891 | First Stage Father's Log Score in 1881 | Second Stage Son's Log Score in 1911 |
| Variables | | | | | | |
| Father's log score in 1861 | 0.586*** (0.003) | | | | | |
| Father's log score in 1851 | | 0.679*** (0.007) | | | | |
| Father's log score in 1851 | | | 0.635*** (0.003) | | | |
| Father's log score in 1861 | | | | 0.648*** (0.006) | | |
| Father's log score in 1891 | | | | | 0.646*** (0.002) | |
| Father's log score in 1881 | | | | | | 0.624*** (0.004) |
| Constant | 1.627*** (0.012) | 1.299*** (0.029) | 1.447*** (0.011) | 1.422*** (0.025) | 1.399*** (0.007) | 1.527*** (0.016) |
| N | 65,700 | 65,700 | 83,095 | 83,095 | 159,723 | 159,723 |
| $R^2$ | 0.366 | 0.072 | 0.368 | 0.081 | 0.444 | 0.096 |
| IV F-stat | 37,867 | | 48,393 | | 127,719 | |

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
*Sources*: Author's analysis of I-CeM (Schürer and Higgs 2014, UKDA, SN 7481) and I-CeM Names and Addresses (Schürer and Higgs 2015, UKDA, SN 7856).

More importantly, the results clearly suggest that measurement error 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 (detailed regression output with first-stage results can be found in Table 5). 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 percent. 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 $\beta$ ranges 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 $\beta$ 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 error 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 error 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 6 provides some additional results. When a different occupational score index (CCC score) is applied, there is still a significant extent of attenuation in the $\beta$ estimated using the conventional OLS formula, caused by measurement error. The $\beta$ rises from between 0.52–0.53 to between 0.63–0.71—21 to 34 percent higher—after instrumenting with a second father's observation. Interestingly, the CCC $\beta$ 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 CCC are much higher, and likewise the IV coefficients for the last period, suggest that the CCC index may be a better measure of occupational status for England during this period, though more work is required to attest to 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 $\beta$ 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

TABLE 6
ADDITIONAL ESTIMATES OF IGE

| | OLS | | | IV | | |
|---|---|---|---|---|---|---|
| | $\beta$ | SE | $N$ | $\beta$ | SE | $N$ |
| *1851–1881* | | | | | | |
| Main results | 0.414 | (0.004) | 66,965 | 0.679 | (0.007) | 65,700 |
| Time-adjusted | — | — | — | — | — | — |
| CCC 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 |
| CCC 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 |
| False positive check | — | — | — | — | — | — |
| *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 |
| CCC 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 |

*Notes*: Standard errors in parenthesis; all estimates are statistically significant to p<0.01; all occupations are scored using HISCAM-U2 unless otherwise stated; "Main results" refer to the results shown in Table 4; "Time-adjusted" estimates are produced using the "HISCAM-E" and "HISCAM-L" schemes to score occupations differently for fathers and sons to reflect the changes in socioeconomic status associated with each occupation; "CCC scores" estimates are produced when occupations are scored by the CCC scheme devised by Clark, Cummins, and Curtis (2023); "Weighted" estimates are produced when the linked sample is reweighted according to population characteristics, following the procedure outlined in Online Appendix D; "NYSIIS" estimates are obtained when IGE is estimated using a linked sample produced by the standard ABE algorithm that matches individuals using their phonetic names (NYSIIS) rather than string distances; "False positive check" estimates are produced when individuals who are likely to be false positive matches are dropped from the sample, according to the procedure outlined in Online Appendix C. *Sources*: Author's analysis of I-CeM (Schürer and Higgs 2014, UKDA, SN 7481) and I-CeM Names and Addresses (Schürer and Higgs 2015, UKDA, SN 7856).

same HISCAM-U2 scale. The difference is greater for the 1881–1911 period and significant to the 95 percent confidence interval.

In addition, estimating $\beta$ 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 Online Appendix D. The "NYSIIS" sample is produced using the phonetic name version of the ABE matching algorithm, as outlined in Online Appendix B. Lastly, the "False Positive Check" sample refers to the multiple links sample after
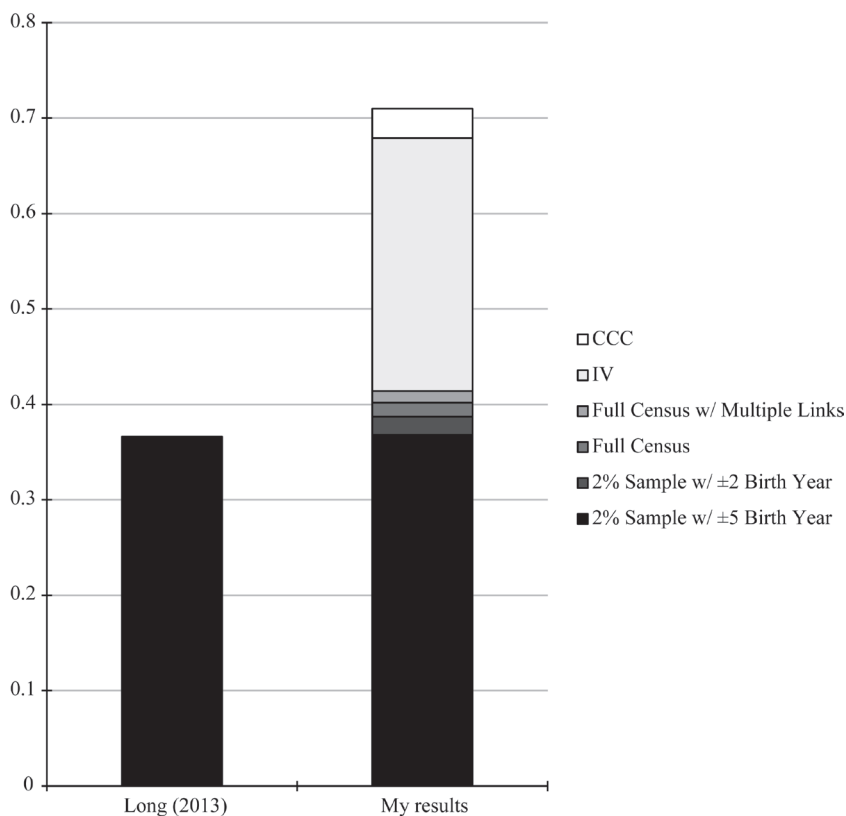
FIGURE 1
COMPARISON OF MY RESULTS FOR 1851–1881 WITH LONG (2013)

*Notes*: "2% Sample w/ ± 5 Birth Year" refers to using a 2 percent sample of the 1851 census and allowing for birth year to differ by at most plus or minus five years—this is the approach taken with census linkage in Long (2013), which I have also replicated in my work; I use the same 2 percent sample that I have created through randomization but with the further restriction of only allowing the birth year to differ by two years to produce the "2% Sample w/ ± 2 Birth Year" estimate; "Full Census" estimate is taken from Table 4, Column (1); "Full Census w/ Multiple Links" is taken from Table 4, Column (2); "IV" estimate is taken from Table 4, Column (3); "CCC" is the estimate obtained using both the IV strategy and the CCC scores instead of HISCAM.
*Sources*: Long (2013) and author's analysis of I-CeM (Schürer and Higgs 2014, UKDA, SN 7481) and I-CeM Names and Addresses (Schürer and Higgs 2015, UKDA, SN 7856).

removing those who were deemed likely to be false positives, using the method discussed in Online Appendix C. As the table highlights, none of these changes affect the results enough to warrant a reconsideration of this paper's findings.

Finally, Figure 1 shows the comparison of my results for the period 1851–1881 with Long (2013), and the difference each change in the data and the methodology makes to the estimates of intergenerational

TABLE 7
SUMMARY OF ALTHAM STATISTICS, 1851–1881

| Estimates | $d(P, I)$ | $G^2$ | $d(Q, I)$ | $G^2$ | $d(P, Q)$ | $G^2$ |
|---|---|---|---|---|---|---|
| Long and Ferrie (2013) | 22.7 | 777 | | | | |
|    vs. 1851–1881 | | | 23.1 | 23951 | 5.6 | 65 |
|    vs. 1851–1881 (corrected) | | | 27.2 | 25140 | 7.4 | 139 |
| Pérez (2019) | 20.8 | 800 | | | | |
|    vs. 1851–1881 | | | 23.1 | 23951 | 7.9 | 54 |
|    vs. 1851–1881 (corrected) | | | 27.2 | 25140 | 10.5 | 98 |

*Notes*: The "corrected" series are estimates that have been corrected for measurement error using Ward's (2023) approach; all estimates are significant at the 99 percent level; $d(P, I)$, $d(Q, I)$, and $d(P, Q)$ all have 9 degrees of freedom.
*Sources*: Unless otherwise stated, all estimates are derived from author's analysis of I-CeM (Schürer and Higgs 2014, UKDA, SN 7481) and I-CeM Names and Addresses (Schürer and Higgs 2015, UKDA, SN 7856); the rest are from Long and Ferrie (2013) and Pérez (2019).

mobility. As expected, the IV strategy contributes to most of the difference between my estimates and those of Long (2013). However, it is evident that other changes in data and methodology, including using the full 1851 census rather than a 2 percent sample, also raise the estimated IGE. This suggests that, among other things, the use of a sample overestimates the true extent of intergenerational mobility, even without correcting for measurement error, which may have been caused by a higher incidence of false positives.[17]

## ALTHAM STATISTICS

Table 7 shows the Altham statistics derived from mobility tables for 1851–1881 before and after correcting for measurement error, using the methodology that Ward (2023) implemented, and how they compare to two existing studies that estimated social mobility using a similar classification scheme but with a 2 percent sample of the 1851 census instead (Long and Ferrie 2013; Pérez 2019). The mobility tables are not shown in the results but can be found in the Online Appendix.

The Altham statistics confirm that the new sample, constructed using the full-count census data, exhibits less mobility than the sample used previously in both Long and Ferrie's (2013) and Pérez's (2019) works. In addition, the impact of attenuation bias from classical measurement error is also confirmed by comparing the distance from perfect mobility before and after correcting for measurement error in the sample—the corrected sample is further away from the matrix of complete independence between rows and columns as expected.

---

[17] See Online Appendix L for more details on false positives caused by the use of a sample.

There are several issues with estimating intergenerational mobility via Altham statistics. As mobility tables are constructed based on just a handful of classes of occupations, a lot of within-class mobility could be missed. In addition, it also does not distinguish between large and small moves across two categories—there is "no difference" between a son with a father who is a farmer becoming a banker or a clerk. On the other hand, the IGE is computed using HISCAM scores, which not only better captures the difference in socioeconomic status associated with occupations belonging to the same broad social class, but also the difference in how large and small each move across the boundary is. Moreover, the method of correcting for measurement error implemented here removes all sons with fathers who have an unstable occupational status from the sample. This could potentially bias the results. Hence, the preferred method of choice for estimating mobility in this paper is the IGE.

Nevertheless, the overall message from this paper is clear: intergenerational mobility in the nineteenth and early twentieth centuries is at odds with the optimistic depiction of Victorian society as one of openness and opportunity.

## DISCUSSION

This paper considerably challenges previous estimates of IGE of occupational status and entails a substantial revision of the perceived wisdom on Victorian social mobility. Table 8 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 (2023) on the United States.

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, Machin, and Reed's (1997) calculations for 1958. If we compare the IV results, however, the decline in father-son association becomes weaker—from 0.68 in 1851–1881 to 0.61 in 1881–1911 and between 0.56 to 0.59 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.[18] Thus, it appears that intergenerational mobility

---

[18] Their name-based estimates are derived using the latent-factor model, which also accounts for issues of measurement error. Results are taken from Table 7 in Clark and Cummins (2015).

TABLE 8
IGE ESTIMATES FOR ENGLAND AND THE UNITED STATES

| Country and Period | OLS | IV | Other (See Notes) |
|---|---|---|---|
| *England* | | | |
| 1851–1881 | 0.41 | 0.68 | — |
| 1861–1891 | 0.41 | 0.65 | — |
| 1881–1911 | 0.41 | 0.62 | — |
| 1851–1881 (Long 2013) | 0.37 | — | — |
| 1881–1901 (Long 2013) | 0.31 | — | — |
| 1972 (Long 2013) | 0.33 | — | — |
| 1958 (Dearden, Machin, and Reed 1997) | 0.22 | 0.59 | — |
| 1958 (Grawe 2004) | — | 0.58 | — |
| 1888–1917 (Clark and Cummins 2015) | — | — | 0.81 |
| 1918–1959 (Clark and Cummins 2015) | — | — | 0.69 |
| 1960–1987 (Clark and Cummins 2015) | — | — | 0.74 |
| *United States (Ward 2023)* | | | |
| 1850 | 0.49 | 0.73 | 0.81 |
| 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 |

*Notes*: Unless otherwise stated, all estimates for England are my own work; Long's (2013) estimates are based on imputed earnings from occupations; Clark and Cummins (2015) results are name-based estimates; Clark and Cummins (2015) split their 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 table due to large standard errors; "Other" estimates from Ward's (2023) work on the United States are IV estimates after accounting for racial (Black and White) differences in intergenerational mobility.
*Sources*: My estimates come from my own analysis of I-CeM (Schürer and Higgs 2014, UKDA, SN 7481) and I-CeM Names and Addresses (Schürer and Higgs 2015, UKDA, SN 7856); Long (2013); Clark and Cummins (2015); Dearden, Machin, and Reed (1997); Grawe (2004). All estimates for the United States are taken from Ward (2023).

increased at a slow rate in England from the nineteenth to the twentieth century.

On the other hand, there are also reasons to suspect that my estimates do not capture the full extent of father-son association in socioeconomic status in the past. Whereas Dearden, Machin, and Reed (1997) and Grawe (2004) had information on the net weekly wages of sons, daughters, and fathers from the 1958 National Children Development Survey, the censuses of 1851 to 1911 only provide occupations. While the IV approach helps to reduce the measurement error associated with inferring status from occupations, it does not address the measurement error from assigning scores to occupations. In addition, improvements 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 may be issues in comparing occupational mobility in the past with the present day, in the absence of data on occupational earnings, 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 could be between 0.61 and 0.68 (or at least as high as such), and the "true" figure may be even higher. At the turn of the century, therefore, England was much closer to a society of profound immobility 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 nineteenth- and early-twentieth-century United States from Ward (2023) are just as high as those for Victorian and Edwardian England.[19] In addition, censuses in England and Wales tend to be more detailed in their description of occupations, such as distinguishing between farmers and agricultural laborers, which not only makes them very useful for social mobility studies but also means that there could be more measurement error in the U.S. censuses arising from the inaccurate or coarse reporting of occupations. Such measurement error could still persist despite the use of the IV approach. Thus, there may be even more attenuation bias present in estimates of American historical intergenerational mobility. This undermines the notion that there was something "exceptional" about American social mobility in the nineteenth century, as Long and Ferrie (2013) claimed.

CONCLUSION

Using an improved set of linked data of between 67,000 and 160,000 father-son pairs constructed 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 error causes significant attenuation bias to estimates of social mobility; correcting for it could

---

[19] One caveat here is that Ward (2023) uses Song et al.'s (2020) literacy-based occupational scores, whereas HISCAM scores are created from social interactions. This might warrant some caution when comparing the coefficients for the United Kingdom and the United States.

raise the IGE obtained from 0.4 to almost 0.7. The results are robust to alternative methods of census linkage and 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, even if Victorian mobility was not particularly remarkable. The surprising fact about English social mobility 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 United States suggests that classical measurement error can have a significant impact on estimates of intergenerational mobility through attenuation bias. After using similar methods to account for measurement error, the intergenerational elasticities of occupational status in England do not appear to be radically different from that of the United States. Therefore, nineteenth-century societies on both sides of the Atlantic were equally immobile, with fathers and sons—in terms of their occupational status—very much alike.

## REFERENCES

Abramitzky, Ran, Leah Boustan, and Katherine Eriksson. "A Nation of Immigrants: Assimilation and Economic Outcomes in the Age of Mass Migration." *Journal of Political Economy* 122, no. 3 (2014): 467–506.

———. "To the New World and Back Again: Return Migrants in the Age of Mass Migration." *Industrial and Labor Relations Review* 72, no. 2 (2019): 300–22.

Abramitzky, Ran, Leah Boustan, Katherine Eriksson, James Feigenbaum, and Santiago Pérez. "Automated Linking of Historical Data." NBER Working Paper No. 25825, Cambridge, MA, June 2020.

Adams, James T. *The Epic of America*. Boston: Little, Brown, and Co, 1931.

Altham, Patricia M. E. "The Measurement of Association of Rows and Columns for an r x s Contingency Table." *Journal of the Royal Statistical Society. Series B (Methodological)* 32, no. 1 (1970): 63–73.

Altham, Patricia M. E, and Joseph P. Ferrie. "Comparing Contingency Tables Tools for Analyzing Data from Two Groups Cross-Classified by Two Characteristics." *Historical Methods* 40, no. 1 (2007): 3–16.

Altonji, Joseph, and Thomas A. Dunn. "Relationships Among the Family Incomes and Labor Market Outcomes of Relatives." NBER Working Paper No. 3724, Cambridge, MA, June 1991.

Anbinder, Tyler, Dylan Connor, Cormac Ó Gráda, and Simone Wegge. "The Problem of False Positives in Automated Census Linking: Evidence from Nineteenth-Century New York's Irish Immigrants." CAGE Working Paper No. 568, Coventry, UK, June 2021.

Antonie, Luiza, Kris Inwood, Chris Minns, and Fraser Summerfield. "Intergenerational Mobility in a Mid-Atlantic Economy: Canada, 1871–1901." *Journal of Economic History* 82, no. 4 (2022): 1003–29.

Bailey, Martha J., Connor Cole, Morgan Henderson, and Catherine Massey. "How Well Do Automated Linking Methods Perform? Lessons from US Historical Data." *Journal of Economic Literature* 58, no. 4 (2020): 997–1044.

Baines, Dudley, and Robert Woods. "Population and Regional Development." In *The Cambridge Economic History of Modern Britain Volume II: Economic Maturity, 1860–1939*, edited by Roderick Floud and Paul Johnson, 25–55. Cambridge: Cambridge University Press, 2004.

Becker, Gary S., and Nigel Tomes. "Human Capital and the Rise and Fall of Families." *Journal of Labor Economics* 4, no. 3 (1986): S1–S39.

Bogart, Dan, Xuesheng You, Eduard J. Alvarez-Palau, Max Satchell, and Leigh Shaw-Taylor. "Railways, Divergence, and Structural Change in 19th Century England and Wales." *Journal of Urban Economics* 128 (2022): 1–23.

Clark, Gregory, and Neil Cummins. "Intergenerational Wealth Mobility in England, 1858–2012: Surnames and Social Mobility." *Economic Journal* 125, no. 582 (2015): 61–85.

Clark, Gregory, Neil Cummins, and Matthew Curtis. "Measuring Mobility: Intergenerational Status Mobility across Time and Space." CEPR Discussion Paper No. 17788, London, UK, January 2023.

Dearden, Lorraine, Stephen Machin, and Howard Reed. "Intergenerational Mobility in Britain." *Economic Journal* 107, no. 440 (1997): 47–66.

Delger, Henk, and Jan Kok. "Bridegrooms and Biases: A Critical Look at the Study of Intergenerational Mobility on the Basis of Marriage Certificates." *Historical Methods* 31, no. 3 (1998): 113–21.

Fuller, Wayne A. *Measurement Error Models*. New York: John Wiley & Sons, 1987.

Grawe, Nathan D. "Intergenerational Mobility for Whom? The Experience of High- and Low-Earning Sons in International Perspective." In *Generational Income Mobility in North America and Europe*, edited by Miles Corak, 58–88. Cambridge: Cambridge University Press, 2004.

Haider, Steven, and Gary Solon. "Life-Cycle Variation in the Association between Current and Lifetime Earnings." *American Economic Review* 96, no. 4 (2006): 1308–20.

Higgs, Edward, Christine Jones, Kevin Schürer, and Amanda Wilkinson. "Integrated Census Microdata (I-CeM) Guide." Online document, University of Essex, Department of History, September 2013. Download from https://www.essex.ac.uk/-/media/documents/research/icem_guide.pdf?la=en.

Lambert, Paul S., Richard L. Zijdeman, Marco H. D. van Leeuwen, Ineke Maas, and Kenneth Prandy. "The Construction of HISCAM: A Stratification Scale Based on Social Interactions for Historical Comparative Research." *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 46, no. 2 (2013): 77–89.

Long, Jason. "The Surprising Social Mobility of Victorian Britain." *European Review of Economic History* 17, no. 1 (2013): 1–23.

Long, Jason, and Joseph Ferrie. "Intergenerational Occupational Mobility in Great Britain and the United States Since 1850." *American Economic Review* 103, no. 4 (2013): 1109–37.

———. "Grandfathers Matter(ed): Occupational Mobility across Three Generations in the US and Britain, 1850–1911." *Economic Journal* 128, no. 612 (2018): F422–F445.

Mazumder, Bhashkar. "Fortunate Sons: New Estimates of Intergenerational Mobility in the United States Using Social Security Earnings Data." *Review of Economics and Statistics* 87, no. 2 (2005): 235–55.

Miles, Andrew. "How Open was Nineteenth-Century British Society? Social Mobility and Equality of Opportunity, 1839–1914." In *Building European Society: Occupational Change and Social Mobility in Europe, 1840–1940*, edited by Andrew Miles and David Vincent, 18–39. Manchester: Manchester University Press, 1993.

———. *Social Mobility in Nineteenth- and Early-Twentieth Century England*. New York: St. Martin's Press, 1999.

Mitch, David. "'Inequalities Which Every One May Remove': Occupational Recruitment, Endogamy, and the Homogeneity of Social Origins in Victorian England." In *Building European Society: Occupational Change and Social Mobility in Europe, 1840–1940*, edited by Andrew Miles and David Vincent, 140–64. Manchester: Manchester University Press, 1993.

———. "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, no. 1 (2005): 26–38.

Mitnik, Pablo A. "Intergenerational Income Elasticities, Instrumental Variable Estimation, and Bracketing Strategies." *Sociological Methodology* 50 (2020): 1–46.

Modalsli, Jørgen Heibø, and Kelly Vosters. "Spillover Bias in Multigenerational Income Regressions." Statistics Norway, Research Department Discussion Paper No. 897, Oslo, Norway, 2019.

Pérez, Santiago. "Intergenerational Occupational Mobility across Three Continents." *Journal of Economic History* 79, no. 2 (2019): 383–416.

Ruggles, Steven, Catherine A. Fitch, and Evan Roberts. "Historical Census Record Linkage." *Annual Review of Sociology* 44, no. 1 (2018): 19–37.

Schürer, Kevin, Edward Higgs, and FINDMYPAST LIMITED. *Integrated Census Microdata (I-CeM), 1851–1911*. Distributed by UK Data Service, Colchester, UK, 4 April 2014. Available from SN 7481, http://doi.org/10.5255/UKDA-SN-7481-2.

———. *Integrated Census Microdata (I-CeM) Names and Addresses, 1851–1911: Special License Access*, 2nd Edition. Distributed by UK Data Service, Colchester, UK, 21 December 2015. Available from SN 7856, http://doi.org/10.5255/UKDA-SN-7856-2.

Smiles, Samuel. *Self-Help with Illustrations of Conduct and Perseverance*. Boston: Ticknor and Fields, 1863.

Solon, Gary. "Intergenerational Income Mobility in the United States." *American Economic Review* 82, no. 3 (1992): 393–408.

Song, Xi, Catherine G. Massey, Karen A. Rolf, Joseph P. Ferrie, Jonathan L. Rothbaum, and Yu Xie. "Long-Term Decline in Intergenerational Mobility in the United States since the 1850s." *Proceedings of the National Academy of Sciences* 117, no. 1 (2020): 251–58.

Thomas, Mark. "The Service Sector." In *The Cambridge Economic History of Modern Britain Volume II: Economic Maturity, 1860–1939*, edited by Roderick Floud and Paul Johnson, 99–132. Cambridge: Cambridge University Press, 2004.

Van Leeuwen, Marco H. D., and Ineke Maas. *HISCLASS: A Historical International Social Class Scheme*. Leuven: Leuven University Press, 2011.

Van Leeuwen, Marco H. D., Ineke Maas, Andrew Miles, and Sören Edvinsson. *HISCO: Historical International Standard Classification of Occupations*. Leuven: Leuven University Press, 2002.

Ward, Zachary. "Intergenerational Mobility in American History: Accounting for Race and Measurement Error." *American Economic Review* 113, no. 12 (2023): 3213–48.

Zhu, Ziming. "Like Father Like Son? Intergenerational Immobility in England, 1851–1911." Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2023-11-24. https://doi.org/10.3886/E195292V1

Zimmerman, David J. "Regression Toward Mediocrity in Economic Stature." *American Economic Review* 82, no. 3 (1992): 409–29.