Rags to Rags, Riches to Riches: Essays on Occupational Mobility in England, 1851–1911

Ziming Zhu[1]

PhD in Economic History

London School of Economics

January 2025

[1]Supervisors: Prof. Neil Cummins, Prof. Chris Minns

Abstract

This thesis uses census-linked datasets, created from digitised full-count decennial censuses of England and Wales between 1851 and 1911, to estimate occupational mobility in Victorian and Edwardian England. The thesis is divided into three papers.

Existing research in social mobility predominantly focus on the intergenerational aspect while largely ignoring another important channel for mobility — mobility over the life course. Using a linked sample of over 400,000 men, the first paper estimates the levels of life-course (intragenerational) mobility in England between 1851–1881 and 1881–1911. By regressing current occupational ranks on initial occupational ranks, this paper finds an intragenerational rank-rank correlation of between 0.61 to 0.68 over a 10-year period, and between 0.50 to 0.57 over a 30-year period. Low occupational mobility was mostly driven by the primary and secondary sectors, but new occupations in services and professions also appear to be relatively secure. Life-course mobility was limited for the Victorians and experienced by only a small minority working in tertiary sectors. England during this period appears to be far from an open society.

The second paper uses a different and likewise newly constructed 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 error 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 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.

The third paper explores spatial variations in intergenerational mobility in England at the end of the nineteenth century, using a census-linked dataset of between 160,000 to 600,000 father-son pairs observed between 1881 and 1911. The results show that there is already a north-south divide in terms of intergenerational mobility in late-Victorian England, using rank-based measures of relative mobility and absolute mobility. In addition, mobility patterns exhibit clear differences depending on migration history and origins. Migrants from the North are much more mobile than those that remained in the North and experience significant gains in occupational ranks from migration, while the same pattern is not observed for Southern migrants and non-migrants. The advantages of north-south migration hold even after accounting for selective migration using household fixed effects. Finally, there is also evidence that there was a 'Great Gatsby curve' in late-Victorian England, as places of higher occupational inequality were also places of lower social mobility.

Declaration

I, Ziming Zhu, certify that the thesis I have presented for examination for the MPhil/PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it). The copyright of this thesis rests with the author. Quotation from it is permitted, provided that full acknowledgement is made. In accordance with the Regulations, I have deposited an electronic copy of it in LSE Theses Online held by the British Library of Political and Economic Science and have granted permission for my thesis to be made available for public reference. Otherwise, this thesis may not be reproduced without my prior written consent. I warrant that this authorisation does not, to the best of my belief, infringe the rights of any third party. I declare that my thesis consists of 38,188 words.

Acknowledgements

Upon writing the final word of this thesis, the culmination of four long, sometimes challenging, but in the end tremendously rewarding years of doctoral pursuit, I can — and I hope you will agree with me — look back on this process with a lot of pride and satisfaction. It goes without saying that social mobility is a topic that I am deeply interested in. Its importance cannot be understated, for what becomes of us and our lives at the end of the journey should not be primarily determined at the starting point by the virtues of our birth. Whilst it is imperative that we are aware of the state of our societies today, I believe there are also valuable lessons to be learnt from history. Thus, I have set out four years ago to discover just a tiny fraction of what the past was like, as the first step to hopefully many more fruitful enquiries. However, I would not be here today without the indispensable help and support from everyone around me, and I would like to thank some of them here.

Firstly, I must thank my supervisors Neil Cummins and Chris Minns for guiding me through the PhD with their knowledge and utmost dedication, for which I will be forever beholden to. Since meeting Chris four years ago, he has always been incredibly attentive to my needs and steadfast in making sure that I received all the comments, advice, and help he could provide on both my doctoral research and my wider scholarly endeavours in a timely manner. My relationship with Neil, on the other hand, began seven years ago when I first started working for him as a research assistant as an undergraduate student. His work and his approach to research, particularly his fondness for big data and the willingness to embrace the methodological frontiers, has never stopped being an inspiration for me.

Outside of academia, I have always enjoyed our conversations at the pub and at various conferences and facing him in the annual departmental football match, and I look forward to the impending celebration of the completion of my (and his many other students') PhD at his home.

I am deeply grateful for the marvelous company afforded by my fellow PhDs at the Department of Economic History. Many have crossed paths with me throughout the course of these four years, and I unfortunately cannot name all of them here. Nonetheless, I must give a special shoutout to my cohort, Ruoran Cheng, Zane Jennings, Mikhail Kolosov, Victor Perez Sanchez, and Sheila Pugh. Victor, in particular, has truly been a great friend as well as a great colleague. I will always appreciate not just the numerous feedback he gave me on my work, but also our exchanges of gifts from our travels, and our enjoyment of the Dutch national anthem at the end of the F1 race weekends together. I must also mention a few others who I am pleased to have shared the journey with: Hillary Vipond, the best colleague I could have wished for, whose passion for research is only rivaled by her love of Go; Aurelius Noble, a fellow student of Neil and roommate in Prato; Hanzhi Deng, who began his PhD at LSE just as I was about to finish my Bachelor's here, and was there to welcome me when I returned as a PhD.

There have been others whose comments and suggestions have shaped my work and changed it for the better. I thank Patrick Wallis and Jane Humphries, who examined my work as part of the Graduate Review Committee during my second year, and the Doctoral Programme Director, Sara Horrell, for arranging my GRC. I had the pleasure of teaching EH209 classes for Sara, an invaluable experience that gave me an insight into being an educator as well as a scholar. Further thanks to Eric Schneider for his advice on my work and my LSE Fellow interview, and Mohamed Saleh for helping me navigate through the job market process. I would like to extend my appreciation to the rest of the department and especially the professional support staff for all the help they have given me, and to the LSE itself. My PhD would not have been possible without the funding provided by the LSE

Studentship, and I have now spent seven years at my department for my BSc and my PhD; it truly holds a special place in my heart. Beyond LSE, I thank Gregory Clark, who has been immensely helpful in providing me with great feedback over conversations and conferences, and my two MPhil supervisors at Cambridge, Joe Day and Amy Erickson, who taught me so much about research and helped me complete my dissertation through the challenging Covid year. Many thanks to all the participants at the Economic History Society, Economic History Association, European Historical Economics Society annual conferences, the 2023 World Cliometric Conference, the Historical Economic Demography workshops, the LSE and Oxford graduate seminars, and all other various conferences, workshops, and seminars that I have had the honour to present at, for their interests in my work and their feedback.

Last but not least, I would like to express my sincere and heartfelt gratitude to those closest to me. My father who, through hard-work and perseverance, brought our family from our small but stunningly picturesque hometown, Guilin, in China to London, the multicultural, metropolitan capital of the UK. My mother, who gave her all in maintaining and managing our household meticulously, making sure that we had the most comfortable home to return to. It is they, and the rest of my family, who cared and continue to care for me with the utmost patience and diligence, that made me who I am today. I want to thank my beloved partner Nina, with whom I have shared many moments of joy and excitement, and who has always been there for me in times of need without fail. Meeting her during the course of my PhD was the best unanticipated outcome of the last four years that I could have ever imagined. Finally, I am ever so grateful to my close friends outside academia for their friendship and support. In particular, I want to pay tribute to my dearest friend Ravi, a genuinely kind and wonderful friend to me for the past decade and a promising young historian, whose life has sadly been turned upside down by his struggle with Long Covid. I dedicate this thesis to my family, for their selfless love and sacrifice, and to Ravi, for his unwavering friendship and his courage in his ongoing fight against this debilitating illness.

Contents

List of Figures

List of Tables

Chapter 1

Introduction: Occupational Mobility in Historical Perspective

> “*Her father was a country gentleman down in your part of the world, and was a brewer. I don't know why it should be a crack thing to be a brewer; but it is indisputable that while you cannot possibly be genteel and bake, you may be as genteel as never was and brew. You see it every day.*”
>
> — Herbert Pocket in *Great Expectations* (Dickens, 1861)

In *Great Expectations*, arguably the pinnacle of Charles Dickens' creations, the Victorian novelist predicates the story on a number of recurrent themes, chief amongst which is social status and mobility. The novel is a coming-of-age story about the protagonist named Pip, an orphan from the marshes of Kent, and his endeavour to become a gentleman in London upon being bequeathed a fortune by a generous benefactor. The story of Pip is an aspiration of every average Victorian and a fictional realisation of their desire for upward social mobility. His words also exposed a central tenet of Victorian society, that the method with which one strove for a living — in other words their occupation — determined one's social status.

There are two ways in which social mobility occurs. Through one's own efforts and fortunes, a person may improve their standing in society during the course of their lifetime, or they may accumulate enough resources to invest in the next generation such that their children can reach for a place higher than their own. The former is called intragenerational, or life course mobility, and the latter is intergenerational mobility, and both are examples of upward mobility. Of course, the opposite could also happen to those less fortunate; a fall in social hierarchy constitutes downward social mobility.

Social status, which reflects one's position in the social order, can be defined by a number of things. Today, we may perceive a person's social standing by their educational history (having a university degree or above), their income and wealth (monetary or possessions), or their publicity (i.e. being a celebrity). For many times and places, occupation was the best and often the only measure of social status in the past still available to researchers today, and widely used by sociologists (Blau and Duncan, 1967; Erikson and Goldthorpe, 1993; Breen and Jonsson, 2005) and economic historians (Long, 2013; Long and Ferrie, 2013; Collins and Wanamaker, 2014; Clark et al., 2023; Ward, 2023). It is an overarching exhibition of a person's past, present, and future, for it contains information about one's education and training, income, and, as the conversation between Herbert and Pip demonstrated, their social standing in the society. Methodologically, there is also an argument that occupation is a more stable indicator of status that is more resilient to measurement error Jarvis and Song (2017), though new evidence from Ward (2023) and Chapter 3 show that measurement error still persists in occupational data.

This thesis examines social mobility in Victorian and Edwardian England, from 1851 to 1911, from the perspective of occupational mobility. The aim is to reconstruct the archetypal occupational trajectory or experience of the average Victorian man to see how easy or difficult for one to transcend the past accomplishments of themselves or their fathers and climb up the social ladder via their efforts and achievements — their occupational attainment — in the labour market, and how precarious it was for those less auspicious to see their status decline and fortunes wane through occupational change and degradation. It will show that the story of Pip, for most in Victorian England at the time, was nothing more than a story or a dream; many could not realise their great expectations despite their utmost endeavours, nor were they truly haunted by any lingering spectre of downward mobility and a fall from grace. England, in the time of Victoria and Edward VII, was a place of striking social persistence.

1.1 Existing Literature on Social Mobility

There is an established and illustrious tradition of research into social mobility in social sciences. Sociologists have long been preoccupied with changes in occupations, which are also intertwined with the concept of class, either across lifetimes or between generations, with the work of Blau and Duncan (1967) on the American occupational structure being particularly influential in putting occupations at the centre of sociological enquiries into social stratification and mobility. Much of the debates among scholars in this field were dominated by the desire to document the levels of social fluidity in various countries in order to ascertain the relationship between industrialisation and social mobility and whether all industrial societies are as open as each other (Lipset and Zetterberg, 1959; Featherman et al., 1975; Treiman, 1970; Erikson and Goldthorpe, 1993). Despite the abundance of illuminating theoretical expositions and quantitative analyses on this subject, the list of sociological enquiries on social mobility grows ever larger today (Blanden, 2013).

At the same time, economists were also seeking to shed some light on economic mobility around the world. Using the model proposed by Becker and Tomes (1986), they explored the extent to which status — in the form of income, wealth, education and so on — can be transmitted across generations in various countries (Solon, 1999; Black and Devereux, 2011; Jäntti and Jenkins, 2015; Deutscher and Mazumder, 2023). Many of these efforts have been directed at the United States (Solon, 1992; Zimmerman, 1992; Ferrie, 1996; Chetty et al., 2014; Song et al., 2020), but we now have a depth of knowledge about the persistence in social status elsewhere, such as in Britain (Dearden et al., 1997; Blanden et al., 2004; Grawe, 2004). Though the methods and variables of interests may not always align between economists and sociologists, the underlying motivations remain the same — to document the ease or hardship to which an individual could engender movements along the social hierarchy either within their own lifetime or for their offspring.

The levels of social mobility can have several important ramifications on society. Piketty (1995) theorised that social mobility can shape political attitudes and support for redistribution, an argument that has been put forward to explain the relative lack of desire for redistributive policies in the United States vis-à-vis Britain and in the North of England versus the South (Long and Ferrie, 2013; Boberg-Fazlic and Sharp, 2018), which has now been empirically tested by Alesina et al. (2018). Generally, people are more likely to accept higher levels of inequality if there is a high degree of social mobility (Alesina and Angeletos, 2005); and simultaneously, the outcomes of an individual living in an unequal and immobile society may be far different from that of an individual in an unequal but mobile society (Solon, 1999). Inequality and social mobility may also be deeply interconnected — in a highly immobile society, those at the top of the echelons of society are more likely to remain at the top (and those at the bottom stay at the bottom). There is now cross-country evidence to demonstrate that this may be the case: the work of Corak (2013) shows that countries with low social mobility tend to be highly unequal, a phenomenon now dubbed as the 'Great Gatsby curve'.

While we currently possess detailed knowledge of the levels of social mobility, measured in a number of ways, in modern societies — and to some extent the correlates and causes of barriers to greater mobility (see Chetty and Hendren (2018*b*,*a*)) — we know far less about social mobility in the historical context, especially before the twentieth century. Being able to correctly identify the degree of openness in past societies serves a vital purpose. Knowing the levels of social mobility in the past allows us to picture the long-term trends in social mobility so as to determine whether there was a change in social mobility, which is the first step towards identifying the factors that may have caused this change. Latest evidence from Nybom and Stuhler (2024) shows that past events can influence future mobility and reveals the importance of viewing social mobility from a long-term perspective.

Victorian England serves as an ideal location and time period for studying historical mobility. It had already underwent the Industrial Revolution by the second half of the nineteenth century, and the pace of structural changes mellowed but was not halted. Between 1851 and 1911, the share of agricultural workers continued to decline as the service sector grew (Thomas, 2004). A number of social changes also took place, such as the passing of the Married Women's Property Act which ended the laws of coverture, or the introduction of compulsory education with the Elementary Education Act. In addition, there is a great amount of regional diversity in England, as a result of specialisation over the course of industrialisation (Geary and Stark, 2018), which provides a scope for exploring the differences in social mobility between regions. Finally, there is also a practical concern — the introduction of national censuses in nineteenth-century England enables the undertaking of a large-scale, micro-level, quantitative enquiry into the levels of social mobility in this period.

There have been various attempts to (quantitatively) document the extent of social mobility in nineteenth-century England, with differing conclusions. The earliest large scale quantitative studies relied on marriage registers, a convenient source of information on the occupations of grooms and their fathers; the findings from these depict Victorian England as a society of limited intergenerational mobility (Mitch, 1993; Miles, 1993, 1999). Others highlight the potential bias introduced by life-cycle effects stemming from measuring the son's and father's status at different points in their life cycle (Delger and Kok, 1998; Long, 2013; Haider and Solon, 2006). Long (2013)'s work, using census data, suggests that Victorian social mobility has been underestimated as a result.

Yet, the debate is far from over. Clark and Cummins (2015)'s surname-based estimates of long-run wealth mobility indicate a strong persistence of elite surnames at the top from the mid-nineteenth century to the present. Recent publication by Ward (2023) emphasises the issue of measurement error in studies of historical mobility using occupations (alongside issues of racial disparity in social mobility), which can mislead us to think that people in the past were more socially mobile than us, at least in the case of the United States. This already poses a challenge to research which suggests that intergenerational mobility was greater in the past than today (Song et al., 2020), and has implications for cross-country comparisons of social mobility that suggests the United States was more mobile than elsewhere (Long and Ferrie, 2013; Pérez, 2019), given the potential difference in the quality of sources.

Another recent development in the wider literature on social mobility is the awareness that not all places are equally mobile, even within the same country. Chetty et al. (2014)'s seminal paper on the geography of economic mobility in present-day United States reveal remarkable differences in intergenerational income mobility at the regional level. A number of similar attempts at measuring and comparing the social mobility of sub-national regions for a myriad of countries followed (Heidrich, 2017; Berger, 2018; Güell et al., 2018; Connor and Storper, 2020; Corak, 2020; Deutscher and Mazumder, 2020; Eriksen and Munk, 2020; Acciari et al., 2022; Bütikofer et al., 2022; Tan, 2023), including Britain (Friedman and Macmillan, 2017; Rohenkohl, 2019; Bell et al., 2023; Carneiro et al., 2022; Breen and In, 2024), using a variety of sources of methods. It suggests that regional disparity in social mobility is far from uncommon. None of these, however, look deep into history, and thus we are left uncertain whether regional divides in social mobility are a modern phenomenon of the contemporary, or a longstanding legacy of the past. Knowing the origins of regional divergence in social mobility is represents a significant step toward uncovering the causes of social persistence.

Finally, most of the literature discussed thus far have a glaring omission — they focus on only one of the two channels of social mobility. Intragenerational, or life course mobility, have received far less attention, especially in the historical context. While some have found suggestive evidence of a decline in career mobility using autobiographies (Miles, 1999) or employee registers (Savage, 1993), there is no large-scale, representative studies dedicated to the quantification of intragenerational mobility for England in the past. This is surprising and problematic, given that the extent of social mobility is also associated with the ease or difficulty of moving up and falling down the career ladder (Jarvis and Song, 2017; Cheng and Song, 2019). Its impact further extends to other important issues, such as wage inequality (Kambourov and Manovskii, 2009; Mouw and Kalleberg, 2010*a*,*b*).

This thesis contributes to the enquiry of historical social mobility by addressing the existing gaps in the literature in the format of three standalone papers. Chapter 2 is the first paper, which deals with the lack of large-scale empirical estimates of life course mobility in England. Using linked census data containing over 400,000 observations constructed from digitised decennial censuses of England and Wales from 1851 to 1911, intragenerational mobility is measured using a range of quantitative methods. The preferred method is a rank-rank regression framework, adapted from the same commonly used measure in studies of intergenerational mobility (Dahl and Deleire, 2008; Chetty et al., 2014), where I regress future occupational outcomes on past occupational status to measure the extent of association between the two. The results suggest that occupational mobility over the life course was limited for the majority of Victorian men. However, a small but non-negligible minority of people did experience important changes or upheaval over their working lives. Upward mobility prevailed over downward mobility on the whole, but there was a process of regression to the mean occurring as well. The findings remain consistent upon adopting an alternative method of analysing occupational mobility based on occupational classes. Occupational mobility, at least in intragenerational terms, remained at a similar level (if not decreasing) during the entire period of 1851–1911.

Chapter 3, which is comprised of the second paper, revises the existing estimates of intergenerational mobility in Victorian England 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. It estimates the intergenerational elasticity (IGE) of occupational status in England between 1851 and 1911, following the Becker and Tomes (1986) model of intergenerational transmission of human capital. 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 per cent. Neither reweighting the sample nor eliminating potential false positives change the interpretation significantly.

Chapter 4 contains the final paper, which focuses on the regional variations in intergenerational occupational mobility in England from 1881 to 1911. It examines the north-south divide from the perspective of occupational mobility, and its interaction with internal migration. I show that by the end of the nineteenth century, there is already evidence of a North-South divide in England in terms of social mobility, using rank-based measures of relative and absolute mobility. Those who grew up in the North found it much harder to move up the occupational strata compared to their counterparts growing up in the South. Additionally, I explore cross-sectionally the interplay between inequality and spatial variation in intergenerational mobility. The results show that areas with higher inequality tend to have lower social mobility — affirming the existence of the 'Great Gatsby curve' in Victorian England. Moreover, the second part of the empirical analysis focuses on the relationship between migration and intergenerational mobility, and finds that migrants moving away from the North experienced greater mobility than those who remained; this is particularly true for those moving to the South of England. In contrast, Southern migrants were not more mobile than Southern stayers — in fact, those moving to the North were less mobile than those who stayed. The results hold even when comparing between brothers who move and stay, a common approach in the migration literature to eliminate issues of selective migration (Abramitzky et al., 2012; Collins and Wanamaker, 2014; Ward, 2019).

Chapter 5 offers some concluding remarks to the dissertation, overviewing the results and the overall picture of Victorian social mobility in England, and some suggestions for avenues of future research which the results of this work elicited. First, the large differences in the IGE obtained before and after correcting for measurement error suggests that there is scope of more research into the impact of measurement error on existing and future estimates of social mobility, past or present. Second, the existence of regional divides in social mobility highlights that the same set of 'rules' can produce different sets of outcomes and implies there is a need for studies of the causes of social mobility, perhaps at the sub-national level. Lastly, future research on social mobility should also explore the socioeconomic impact of social mobility and further strengthen the need for policymakers to design policies that promote greater social openness.

Chapter 2

Climbing the Ladder: Life-Course Mobility in England, 1851–1911

2.1 Introduction

Social mobility is the process by which individuals move between social categories — often (and for the purpose of this paper) defined by occupational status — either between generations (intergenerational) or over the course of a lifetime (intragenerational). It is the subject of a long-standing preoccupation in research across multiple disciplines, which has precipitated a large number of studies documenting intergenerational mobility in various countries, past and present. Yet, the other crucial component of social mobility — intragenerational mobility — has received much less attention from scholars in this field.

Much of the early enquiries into intragenerational mobility focus on earnings mobility and sought to identify any systematic relationship between income mobility and inequality (Kopczuk et al., 2010), with some cross-national comparisons (most prominently between the United States and Germany (Burkhauser and Poupore, 1997; Gulgun Bayaz-Oturk and Couch, 2014)), as well as establish long-term trends in income mobility over the life course (Berman, 2022).[1] However, intragenerational occupational mobility is deeply intertwined

[1]There is also a considerable interest in the relationship between inequality and intergenerational mobility,

with wage inequality too. Research has suggested that the increase in wage inequality in the United States from the late-twentieth century onwards can be explained by an increase in life-course occupational mobility (Kambourov and Manovskii, 2009); occupational mobility are also important for explaining wage inequalities between gender (Fuller, 2008) and across nations (DiPrete and McManus, 1996). Therefore, studying occupational changes across careers can help inform the causes of income inequality in a society.

Aside from its implications for wage inequality, there are also other reasons for studying occupational mobility over the life course, as shown by a number of recent investigations. Jarvis and Song (2017) finds that rising intragenerational mobility in the United States since 1970 prevented a decline in intergenerational mobility, in a period of increasing inequality. Thus, in the absence of greater mobility over the life course, intergenerational mobility would have declined if inequality has increased. In addition, intragenerational mobility is also deeply relevant to research on labour market behaviours (Kalleberg and Mouw, 2018). Conceptually, intragenerational mobility is a mixed blessing: mobility indicates the possibility of upward progression in the career ladder and (or) transition into a more desirable occupation, but too much mobility introduces instability and uncertainty that may undermine productivity (Jarvis and Song, 2017).

Given the importance of intragenerational mobility, one would imagine that studying the life-course occupational mobility of people in the past could yield important results for understanding the long-run evolution of inequality and social mobility, which has been meticulously documented by an ever-growing literature (Piketty and Saez, 2014; Clark and Cummins, 2015; Song et al., 2020; Clark, 2023). Yet, there is a lack of large-scale empirical research into historical life-course mobility. Much of the existing quantification of intragenerational mobility, whether defined by income or occupation, do not go beyond the second half of the twentieth century and focus predominantly on the United States (Jäntti and Jenkins (2015) provides a good review of research on intragenerational income mobility). However, if

and research purports the existence of the 'Great Gatsby curve' — places with higher inequality also tend to have lower intergenerational mobility (Corak, 2013) we are interested in the historical development of inequality or social mobility, it is vital that we estimate life-course occupational mobility as well as intergenerational, since changing to a better job was a key mechanism for upward mobility.

This paper estimates life-course mobility for men in Victorian and Edwardian England, using linked census data containing over 400,000 observations constructed from digitised decennial censuses of England and Wales from 1851 to 1911. Mobility is estimated in a rank-rank regression framework, adapted from the same commonly used measure in studies of intergenerational mobility (Dahl and Deleire, 2008; Chetty et al., 2014), where I regress future occupational outcomes on past occupational status to measure the extent of association between the two. The results suggest that occupational mobility over the life course was limited for the majority of Victorian men — large improvements or deteriorations in their occupational ranks were unlikely from the start to the end of their careers. However, a small but non-negligible minority of people did experience important changes or upheaval over their working lives. Upward mobility prevailed over downward mobility on the whole, but there was a process of regression to the mean occurring as well. The findings remain consistent upon adopting an alternative method of analysing occupational mobility based on occupational classes. Occupational mobility, at least in intragenerational terms, remained at a similar level (if not decreasing) during the entire period of 1851–1911.

The sectoral analyses of occupational mobility suggest that the low degree of life-course mobility can be attributed to the primary and secondary sector occupations being highly stable and secure. Those working in the tertiary sector, in contrast, are much more likely to change their occupations or exit the industry altogether. However, by the end of the Victorian era, there was the rise of new occupations in the services and professions sector which were also unlikely to engender much intragenerational mobility through occupational changes. This may be an indication that life-course mobility through occupational changes will be increasingly uncommon as the twentieth century takes over.

The results of this paper contribute to the existing literature on historical social mobility by providing an estimate of life-course mobility in the past, derived from a dataset superior in size and quality. Previous efforts to quantify life-course mobility in Britain relied on small samples (no larger than 8,000 observations) due to data constraints (Mitch, 2005; Long, 2013).[2] Others used a range of innovative sources, such as autobiographies (Miles, 1999), employee records of firms (Savage, 1993), or interviews (Vincent, 1993) to tell an illustrating story about the work-life of various sub-populations of the past, which are undoubtedly interesting and informative, but cannot be representative of the entire population. This paper makes use of the newly digitised full-count censuses to construct a much larger sample of between 132,000 to 236,000 people to produce more precise and representative estimates of intragenerational occupational mobility for Victorian England.

A second contribution of this paper is methodological. While there are attempts to observe and estimate life-course mobility, as previously noted, there is a lack of a commonly accepted measure such as the intergenerational elasticities (IGE) used in studies of intergenerational mobility (reviews of measures of intergenerational mobility include Black and Devereux (2011); Deutscher and Mazumder (2023)). This paper adapts the method of rank-rank correlation popularised by Chetty et al. (2014) in their seminal work on intergenerational mobility in present-day United States to outline the use of an 'intragenerational rank-rank correlation' in estimating life-course mobility. Summarising the relationship between past and present (or future) occupations in this way allows for lucid interpretations and enables comparisons across time and space.

The rest of the paper is organised as follows. Section 2.2 surveys the existing debates in historical social mobility and empirical investigations into intragenerational mobility. Section 2.3 discusses the data and census linkage. Section 2.4 outlines how mobility over the life course is measured in this paper. The results obtained using the various methods are then shown in Section 2.5, and Section 2.6 explores occupational mobility in further depths by showing which kinds of occupations are associated with more or less life-course mobility.

[2]Intragenerational mobility was also only a sidenote in Long (2013), whose main aim was to estimate intergenerational mobility using linked census data.

Finally, Section 2.7 concludes.

2.2 Related Literature

Much of the existing efforts to quantify social mobility of past societies, whether by economists and economic historians or sociologists, have focused on the concept of intergenerational mobility, often between fathers and sons, while overlooking the extent of change or continuity within one's career. Recent developments in this line of enquiry in economic history highlight the issue of classical measurement error in earlier studies (such as Long (2013); Long and Ferrie (2013); Song et al. (2020) for Britain and the United States), which led to an overestimation of the extent of social mobility in the past (Ward, 2023; Zhu, 2024). In particular, Zhu (2024) shows that after correcting for measurement error, intergenerational mobility in Victorian England was much lower than previously thought, and close to the estimates for post-war Britain which have made similar corrections for measurement error (Dearden et al., 1997; Blanden et al., 2004).[3] This corroborates with Clark and Cummins (2015)'s view of consistently low wealth mobility since the mid-nineteenth century.

At the same time, sociologists have also been interested in the social mobility of the past in order to construct the long-run trend to settle the debate on the industrialism thesis. Exponents of this theory argue that industrialisation precipitated an increase in the application of rationality and efficiency to decision-making in all aspects of social life and, with it, occupational selection shifted away from ascription by social origins and towards meritocracy, often in the form of formal education (Blau and Duncan, 1967; Treiman, 1970). A counter-theory by Featherman et al. (1975), building on Lipset and Zetterberg (1959), posits that relative mobility (the chances of obtaining a particular occupational class) is the same across all industrial societies, past and present, which has since been confirmed by Erikson and Goldthorpe (1993)'s seminal survey of social mobility in nine European countries in the

[3] The issues of classical measurement error in research on intergenerational mobility has long been recognised and discussed extensively in both Ward (2023) and Zhu (2024).

twentieth century. Turning to Britain specifically, Goldthorpe (1987) asserts that barriers to greater social mobility remained persistent in the post-war period and its stagnation was likely a long-running phenomenon. This appears to be consistent with the aforementioned findings of British or English intergenerational mobility in economic history.

While the levels of intergenerational mobility in societies, past and present, have been extensively surveyed, contemporary studies of intragenerational mobility have helped inform the significance of research into mobility over the life course. Kambourov and Manovskii (2008, 2009) document rising (intragenerational) occupational mobility in the United States in the second half of the twentieth century and model the implications of such increases on wage inequality; they find that the model accounts for 90 per cent of the observed increase in wage inequality between the 1970s to the 1990s. Mouw and Kalleberg (2010*a*,*b*) also identified the central role of occupations for studying wage inequality in late-twentieth-century United States, but argue that mobility between occupations is less important than differential wage growth within firms in explaining rising wage inequality. The importance of mobility in the workplace can likewise be observed in comparions between countries (DiPrete and McManus, 1996) and genders (Fuller, 2008).

On the other hand, intragenerational mobility has significant implications for intergenerational mobility too. Jarvis and Song (2017) show that rising intragenerational mobility may have mitigated a decline in intergenerational mobility as inequality rose in the United States. Cheng and Song (2019) also demonstrate that fathers and sons shared similar life-course mobility and trajectories, suggesting that intergenerational mobility can be viewed as the transmission of life experiences from parents to children. These studies emphasise both the theoretical and empirical importance of life-course mobility in social mobility research and the connections between intragenerational and intergenerational processes of social mobility. In addition, there are methodological consequences too — the findings of significant intragenerational mobility in occupations suggests that research on intergenerational occupational mobility are not free from issues related to measurement error (Jarvis and Song, 2017; Ward, 2023; Zhu, 2024).

The main impediment to estimating intragenerational mobility has been data constraints. Much of the existing research on social mobility in nineteenth-century Britain often relied on marriage registers (Miles, 1993, 1999), which only offer a snapshot of an individual's career. Therefore, they do not offer sufficient evidence on how one's career progressed.

Some have circumvented this constraint by utilising autobiographies, interviews, or case studies of particular organisations. For example, Miles (1999) analysed 479 autobiographical accounts and found that the extent of 'class undulation' between the first and the last occupation recorded declined from just above 40 per cent to around 25 per cent between writers born in 1723–1815 and those born in 1816–64, followed by a stagnation for those born in 1865–1914.[4] Savage (1993) conducted a detailed case study of career mobility in Lloyds Bank between 1880 and 1950, and found that there was a decline in promotion chances over time, which led to a rise in clerical unrest and middle class formation, until new practices were introduced in the 1920s to restore promotion opportunities. Finally, Vincent (1993) employed oral history to map the changes in career pathways in early-twentieth-century Britain. These results, whilst incredibly important and informative, cannot claim to be representative, and therefore cannot reliably indicate the levels and trends of intragenerational mobility in Victorian and Edwardian England.

Recent developments in the digitisation of censuses have enabled scholars to use the England and Wales decennial censuses, which contain a rich array of individual-level information, for studying social mobility in the past. Long (2013) was able to exploit this newly available resource by employing census linkage techniques (explained in section 2.3) to reconstruct a person's life course by connecting different points in their lives observed in the censuses. He was able to estimate that 43.6 per cent of his sample changed social class over a 30 year period between 1851 and 1881. However, the main purpose of his paper was to estimate

[4] Humphries (2010) also used autobigraphies to study occupational mobility of child labourers during the Industrial Revolution, but she was focusing more on the mobility between the father's occupation and the son's first job.

father-son mobility rather than life-course mobility; the latter was just a side note used to demonstrate the pitfalls of using marriage registers to estimate intergenerational mobility. Furthermore, his sample size is still restricted to around 7,700 individuals due to the use of a two per cent sample, instead of the complete-count version, of the 1851 census.

Overall, despite the theoretical and empirical importance of life-course mobility for inequality and a more complete understanding of social mobility, there is yet to be a high quality quantitative study of life-course mobility in the past. This paper will use linked sample created from the full-count decennial censuses to estimate intragenerational mobility between 1851 and 1911, putting life-course mobility at the centre of the enquiry.

2.3 Data and Census Linking

2.3.1 Raw Data

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 and Higgs (2014). The second is the I-CeM Names and Addresses database from 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.[5] This automatic processing, aside from achieving practical efficiency, ensured that decisions concerning the validity of the underlying data

[5]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.

source have been applied consistently across the entire database (Higgs et al., 2013). 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*.

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 the coding of occupations is at least 95 per cent accurate. Other variables, such as marital status and household relationships, have also been standardised, coded, and checked for consistency (Higgs et al., 2013).

2.3.2 Census Linkage

To conduct record linkage across the censuses, this project selects English men who are at the early stages of their occupational trajectory (ages 20–30) in the initial year and tracks their occupational status at each decadal interval, until they reach the latter stages of their career (ages 50–60), so that their peak career status can be captured. This is done for two periods: 1851 to 1881 and 1881 to 1911.

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 can strike the right balance between eliminating false matches and achieving a high 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. The procedure is outlined in Appendix A.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. (2020), 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.[6] 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 (Bailey et al., 2020).

A comparison of the matching results obtained using ABE's algorithm show that by adopting the more conservative approach, the rates of Type I errors are reduced from 24 to 33 per cent, depending on the dataset, to 17 to 23 per cent (Bailey et al., 2020). By their own estimation, Abramitzky, Boustan, Eriksson, James and Pérez (2020) argues that the more conservative version of their linking algorithm produces a very low false positive rate — as low as 2.87 per cent in the case of population-to-population matching (which this paper is attempting), although this may be an underestimation.[7]

[6]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. For a more detailed explanation of how the synthetic ground truth data was constructed, see Bailey et al. (2020)

[7]To test their method, individuals from the entire 1910 and 1920 United States censuses were linked using the ABE algorithm. The more conservative approach linked 8 million people, of which 2.78 million overlapped

The ABE methodology has also 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). To examine the extent and impact of potential false matches, this paper follows the advice of Anbinder et al. (2021) in utilising spousal information to produce a more conservative 'true' sample. This 'true' sample is made up of men who were married from the start of the period, and whose wives throughout the period are present in the household, have the same name, and whose expected age does not differ by more than two years. The results produced using this more conservative approach do not appear to be significantly different.[8]

There are *a priori* reasons to believe that false matches may be less of an issue with linking British censuses. While the United States 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. (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

with people for whom a match was observed in the Family Tree data, which was produced by users of the Family Tree platform. The Family Tree data is treated as the 'ground truth' because users have personal motivation to find their ancestors, as well as private information that may help them to establish a true link. They then calculated the false positive rate based on the percentage of links that did not match the Family Tree data out of all the 2.78 million overlapped links (Abramitzky, Boustan, Eriksson, James and Pérez, 2020).

[8]For more details on the estimation of false positive rates, see Appendix A.C.

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. (2020) shows that reweighting the linked sample according to the population characteristics can move the estimates in line with the true figure. Yet, 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. Moreover, a non-representative linked sample can be reconciled with the population by using inverse propensity weighting to account for the likelihood of being linked, as demonstrated in Bailey et al. (2020). Results for both the unweighted and weighted samples will be reported. Appendix A.D describes the weighting procedure in more detail.

2.3.3 Linked Sample

Using the aforementioned ABE algorithm, this paper has linked 150,220 men, who were aged 20 to 30 in the starting year, between 1851 and 1881, and 260,065 men between 1881 and 1911 (a total of over 410,000 men altogether). To check the representativeness of the linked sample, Table 2.1 compares some characteristics of this group with the full sample — all English men, aged 20 to 30, living in England in 1851 and 1881. The match rate, which is calculated as the percentage of observations in the linked sample over the full sample, is at least 11.7 per cent — slightly lower than the 17 per cent that Abramitzky, Boustan, Eriksson, James and Pérez (2020) achieved. This can be explained by at least two factors. Firstly, they linked men from a later period (1910 and 1920 United States), when mortality would most likely have been lower. However, even in mid-nineteenth century Britain the expected mortality of men in their 20s was low enough — less than 5 per cent — that the changes in mortality would not have made a huge difference to the match rate (Woods and Hinde, 1987).

The more important reason which can account for the lower match rate is that this paper has linked individuals over a longer time frame (30 years instead of 10) and across more than two censuses (three for 1851–1881, and four for 1881–1911). For 1851–1881, the 1851 census was linked to both the 1861 and the 1881 censuses, and only those who were successfully linked in both cases were included in the sample (thus excluding those who can only be linked between 1851 and 1881 but not between 1851 and 1861). The same was true for the 1881–1911 sample, except there were three sets of links — 1881–1891, 1881–1901, and 1881–1911. This may lead to a lower match rate in various ways. For example, those who emigrated in the middle of the period and then returned at the end would be excluded. Moreover, if there were errors in recording or transcribing the data — if someone had reported the wrong age or place of birth, or if the census enumerator had misread someone's name — in one of the census years, then these people would be missed out. However, this may also minimise the rate of type I errors, as the requirement on multiple successful links would reduce the likelihood of a false match caused by transcription errors.

The results from the comparison show that based on some attributes the linked sample was representative of the full sample. The mean ages of the samples are very similar across the board — just under 25 in all cases. The average socioeconomic and occupational status, measured by HISCAM (Historical Cambridge Social Interaction and Stratification) and CCC-RCII, two stratification scales for scoring occupations from Lambert et al. (2013) and Clark et al. (2023), also show very little deviation between the full and the linked samples.[9] Interestingly, despite the concern that matching will be biased towards longer names, the differences in name length are small. In terms of the number of kids and the number of

[9]More explanations on occupation scores in Section 2.4.1.

Table 2.1: Representativeness Results in 1851 (for 1851–1881) and 1881 (for 1881–1911)

	1851–1881			1881–1911		
	Population	Linked	Weighted	Population	Linked	Weighted
Socioeconomic and Personal Characteristics (Mean)						
Age	24.88	24.89	24.66	24.80	24.73	24.56
HISCAM	52.23	52.01	52.74	53.77	53.92	54.29
CCC	51.14	51.38	52.43	51.95	52.70	53.25
HISCAM Rank	48.09	46.96	50.32	49.20	49.24	50.59
CCC Rank	47.49	45.89	49.64	48.75	48.69	50.32
First Name Length	5.83	5.86	5.84	6.55	6.66	6.60
Surname Length	6.29	6.35	6.32	6.34	6.40	6.34
Number of Kids	0.48	0.56	0.52	0.68	0.77	0.71
Number of Servants	0.19	0.23	0.22	0.14	0.16	0.16
Household Status in 1851/1881 (%)						
Head	39.71	44.67	42.64	40.32	46.23	43.28
Son	27.55	32.47	33.96	30.55	33.62	35.76
Visitor	3.30	1.88	2.10	1.11	0.75	0.79
Lodger	8.76	4.97	5.19	6.38	4.29	4.46
Boarder	0.09	0.06	0.07	3.02	4.26	4.59
Marital Status in 1851/1881 (%)						
Single	55.07	51.35	53.54	55.56	51.08	53.99
Married	42.92	47.23	45.03	43.48	48.24	45.36
Occupational Structure (%)						
Agriculture	24.68	31.57	25.28	13.24	18.82	14.32
Manufacturing	57.98	54.52	57.59	62.86	59.18	61.88
Services	17.34	13.91	17.13	23.91	22.00	23.80
Residence (%)						
London	14.29	6.47	11.72	16.36	6.95	9.30
Extra London	6.82	8.04	7.28	8.36	9.35	8.58
Lancashire	11.59	7.24	11.37	14.37	11.45	16.92
Yorkshire West Riding	8.83	9.91	12.04	9.66	10.60	12.39
Birth County (%)						
London+	16.67	13.15	16.59	18.93	12.26	12.75
Lancashire	9.79	6.79	10.62	12.42	9.88	14.37
Yorkshire	11.61	13.52	15.14	12.09	13.53	15.03
N	1,283,845	142,530	142,530	1,967,438	250,785	250,785
Match Rate (%)		11.10			12.75	

Notes: Figures are for the initial year (1851 and 1881). Those without occupations have been excluded when measuring the mean HISCAM score and the occupational structure. 'Extra London' refers to the counties Middlesex, Kent, Essex, and Surrey. 'London+' refers to Middlesex, Kent, Essex, and Surrey (there is no 'London' in the standardised county of birth code, only 'City of London', which is a tiny part of London with no one recorded as being born there). 'Yorkshire' in birth county includes all Ridings of Yorkshire. *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).

servants, there does not seem to be a huge difference either.

In other aspects, however, there appears to be differences between the full sample and the linked sample that are important to note. The birth county variables show the geographical distribution of the birthplaces of the full and the linked samples. The figures for the most populous regions — 'London +' (all of Essex, Kent, Middlesex, and Surrey, which includes regions that are not traditionally considered part of London), Lancashire, and Yorkshire — are shown. There are important differences in the geographical distribution of the two samples, with the proportion born in London and Lancashire being much lower in the linked sample than the full sample.

The differences in the distribution of birthplaces of the full and the linked sample can be reconciled by the fact that the requirement on a unique combination of name, age, and parish of birth penalises those who were born in more populous places. Thus, in regions like London or the large cities of Liverpool and Manchester in Lancashire, where parishes have a higher population density, such unique combinations are harder to come by.

The fact that the matching process may introduce some bias against those born in dense, urban areas can also explain the variations in some other characteristics between the full sample and the linked sample. For instance, the census region variables capture the geographical distribution of the sample on the day of the 1851 and 1881 census. As is the case with birthplaces, the proportion of the linked sample living in denser, more urbanised regions like London and Lancashire is much lower than the proportion found in the full sample.

The occupational structure estimates are constructed using the Historical International Classification of Occupations (HISCO) code, constructed by van Leeuwen et al. (2002), provided in the data. HISCO is a classification scheme for occupations in the nineteenth and early-twentieth centuries built upon the 1968 version of the International Classification of Occupations (ISCO-68) developed by the International Labour Organisation. The HISCO scheme has coded the 1,000 most frequent male and female occupational titles in eight European countries, including Britain, between 1692 and 1971 (but mostly from the nineteenth century) into five-digit codes. These occupations (and their codes) are grouped into nine 'major' groups, demarcated by the initial number of these five-digit codes. HISCO codes beginning with 6 (major group 6) represent agricultural occupations; major groups 7, 8, and 9 are those working in mining and manufacturing, construction, and transport; the rest can be classified as service workers (van Leeuwen et al., 2002).

As Table 2.1 shows, the linked sample is slightly biased towards agricultural occupations, despite the similarity in the socioeconomic status (measured by HISCAM scores) between the full and the linked sample. This may again be explained by the fact that the linking algorithm penalises those born in dense, urban settlements, who would likely have taken up jobs in manufacturing or services in their adult life.

The summary statistics for the re-weighted sample are also shown in the table. In terms of representativeness, the re-weighted sample resembles the population much more closely in occupational structure, and modestly so in household status, marital status. For residence and birth county, the re-weighted sample is more representative than the unweighted one for London and Lancashire but over-weights Yorkshire. There are very little difference in representativeness between the two samples in socioeconomic and personal characteristics.

2.4 Measures of Occupational Mobility

2.4.1 Assigning Scores to Historical Occupations

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 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 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, is used. Table 2.2 shows a sample of some common occupations observed in the census with their respective HISCAM scores.

Table 2.2: 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 (UKDA, SN 7481); 'HISCAM' is taken from Lambert et al. (2013).

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 et al. (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 create this index is the same as the one used by Lambert et al. (2013) for HISCAM. Using information from marriages, Clark et al. (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 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. 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.[10]

2.4.2 Measuring Mobility: Rank-Rank Correlation

Central to the study of intergenerational mobility is Becker and Tomes (1986)'s intergenerational elasticity (IGE), which measures the transmission of status as an intergenerational process between parent and child in a simple framework:

$$Y_{i,t} = \alpha + \beta Y_{i,t-1} + \epsilon_{i,t} \tag{2.1}$$

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

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

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.

Dahl and Deleire (2008), building on the Becker and Tomes (1986) model, pioneered an alternative measure of intergenerational mobility which involved placing parents and children into their respective cohorts' percentile ranks (based on any indicator of socioeconomic status, such as income), then estimate the coefficient in a rank-rank regression between fathers' ($R_{i,t-1}$) and sons' ranks ($R_{i,t}$) — known as the rank-rank correlation or slope:

$$R_{i,t} = \alpha + \rho R_{i,t-1} + \epsilon_{i,t} \tag{2.2}$$

This measure has gained increasing popularity since Chetty et al. (2014) and now used in a number of research (Corak, 2020; Deutscher and Mazumder, 2020; Acciari et al., 2022).[11]

While there is no method of estimating life-course mobility that is as widely accepted and adopted as the IGE (and its variants), some have used the Spearman's rank correlation between the income of an individual at time t and $t + 1$ as a measure of intragenerational mobility. This method is quite comparable to the rank-rank regression used in intergenerational mobility research — when the ranks are established based on the population of interest, the coefficient ρ from the rank-rank regression is the same as the Spearman correlation (Deutscher and Mazumder, 2023).[12] Therefore, this paper will adapt and implement the rank-rank regression in the context of life-course mobility:

$$R_{i,t} = \alpha + \rho R_{i,t-1} + \epsilon_{i,t} \tag{2.3}$$

Where $R_{i,t}$ is the rank of individual i's current occupation in his cohort's national distribution at time t, and $R_{i,t-1}$ is the rank of the same individual's past occupation at time $t - 1$, and

[11] Deutscher and Mazumder (2023) discusses the difference between rank-rank correlation and slope.

[12] Chetverikov and Wilhelm (2024) argues that this holds only when there are no covariates and the distributions are continuous. While the former is the case here, the latter is not so — due to some occupations being very frequent and therefore tied in ranks, some percentiles are larger than one per cent. Appendix A.E shows that the results are similar using Spearman correlation.

ρ is the rank-rank correlation of intragenerational mobility.

The advantages of rank-rank correlation, in the context of intergenerational mobility, have already been extensively discussed in Chetty et al. (2014); Nybom and Stuhler (2017); Deutscher and Mazumder (2023), and some of these are relevant for life-course mobility too. Most crucially, rank-rank correlation can account for the impact of changes in the occupational structure, such as the expansion of certain occupations. It also has the advantage of the ease of interpretation and scope for comparison across time and space. To measure short-term and long-term correlation, I will measure the correlation between two occupations 10 and 30 years apart.

2.4.3 Measuring Mobility: Transition Matrices

An alternative way to classify occupations, and one which has been commonly used in historical and sociological enquiry on social mobility, is to assign a rank to each occupation that reflects its social position according to various dimensions – for instance skill level — using a classification scheme. This research uses HISCLASS — an international historical social class scheme based on HISCO. 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 (Leeuwen and Maas, 2010). These twelve levels can be condensed into a smaller, five-class scheme for comparison with previous estimates from Long (2013). Table 2 describes each of the seven classes in HISCLASS and how they can be combined into the five-class version.

Using this scheme, occupational mobility can be demonstrated with a mobility table — a two-way table with class 'origin' (e.g. occupational class at age 20) on one side and class 'destination' (e.g. occupational class at age 50) on the other. Mobility rates are shown in the table, which can reveal the extent of both downward and upward mobility (and immobility) over the period. A simple measure of occupational mobility that can be derived from the table is the percentage of people who ended up in a different class category than the one they

Table 2.3: Conversion of HISCLASS Categories to Armstrong (1972)'s Scheme

Levels	HISCLASS Description	Class	Armstrong (1972) Scheme
1	Higher managers	I	Professional
2	Higher professionals	I	Professional
3	Lower managers	II	Intermediate
4	Lower professionals	II	Intermediate
5	Lower clerical and sales personnel	II	Intermediate
6	Foremen	III	Skilled
7	Medium skilled workers	III	Skilled
8	Farmers and fishermen	III	Skilled
9	Lower skilled workers	IV	Semi-skilled
10	Lower skilled farm workers	V	Unskilled
11	Unskilled workers	V	Unskilled
12	Unskilled farm workers	V	Unskilled

Sources: HISCLASS categories are taken from van Leeuwen and Maas (2011); definitions of the five classes follow the Armstrong (1972) scheme used in Long (2013), and conversion to the scheme is based on the author's judgement.

started with. Constructing a mobility table and calculating absolute mobility rates enables a comparison with results from other studies.

2.5 Empirical Results

2.5.1 Rank-Rank Correlation

Table 2.4 shows the short-term (10 years) life-course mobility across a range of specifications for men in Victorian England for two separate period: 1851–1861 and 1881–1891. The 10-year rank-rank correlation between previous and current occupation ranges from 0.63 to 0.61 in 1851–1881; in other words, a 1 per cent increase in the percentile rank of 1851 occupation is associated with a 0.6 per cent increase in the percentile rank of 1861 occupation. For 1881–1891, the intragenerational rank-rank correlation ranges from 0.66 to 0.68. Therefore, short-term life-course mobility seems to be very constant during the Victorian era, with little evidence of any increase or decrease.

Moreover, table 2.5 shows that long-term (30 years) life-course mobility was equally stable

Table 2.4: 10-Year Intragenerational Rank-Rank Correlation

	1851–1861				1881–1891			
	HISCAM		CCC		HISCAM		CCC	
Rank-Rank ρ	0.63***	0.61***	0.62***	0.61***	0.66***	0.66***	0.68***	0.68***
	(0.002)	(0.003)	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
Weighted	NO	YES	NO	YES	NO	YES	NO	YES
N	131,512	131,512	131,263	131,263	237,439	237,439	236,276	236,276
R^2	0.413	0.397	0.416	0.410	0.447	0.437	0.468	0.464

Notes: robust standard errors in parenthesis; *** 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).

between 1851 and 1911. For 1851–1881, the 30-year intragenerational rank-rank correlation ranges from 0.50 to 0.52 and for 1881–1911, the rank-rank correlation ranges from 0.55 to 0.57. The long-term correlation is weaker than the short-term correlation, as expected, since there is more time for workers to climb up (or fall down) the occupational ladder.

Table 2.5: 30-Year Intragenerational Rank-Rank Correlation

	1851–1881				1881–1911			
	HISCAM		CCC		HISCAM		CCC	
Rank-Rank ρ	0.52***	0.51***	0.50***	0.50***	0.55***	0.55***	0.57***	0.57***
	(0.002)	(0.003)	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
Weighted	NO	YES	NO	YES	NO	YES	NO	YES
N	131,512	131,512	131,263	131,263	237,439	237,439	236,276	236,276
R^2	0.298	0.284	0.286	0.281	0.307	0.301	0.335	0.331

Notes: robust standard errors in parenthesis; *** 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).

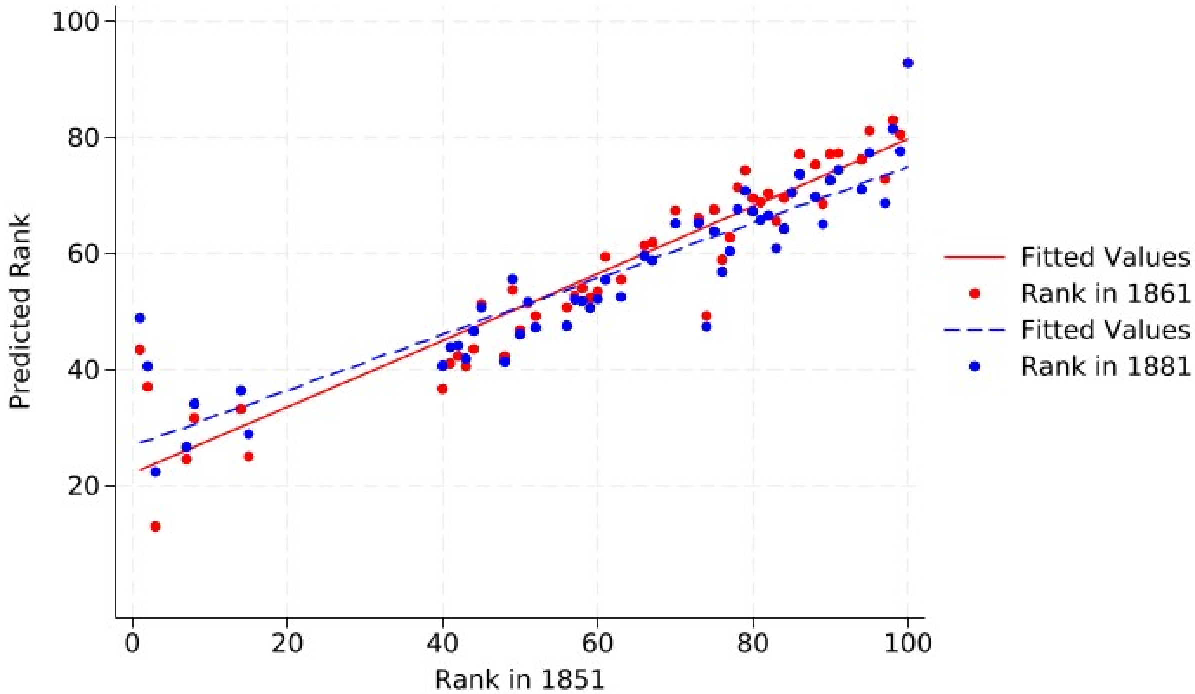
Overall, the short-term and long-term rank-rank correlation suggests that life-course mobility remained at a relatively stable level across the entire (long-)Victorian era. There is no evidence of a substantial rise or decline in the intragenerational correlation in occupational status. Interestingly, the levels of long-term life-course mobility are comparable to the levels of intergnerational mobility — the rank-rank correlation in father-son status for 1881–1911 was 0.54 (Zhu, 2025). Thus, the rank (or status) a Victorian man obtained at the end of his occupational career seems to be as strongly persistent to the rank of his father as it is to the rank he belonged to at the beginning of his career.

Alternative measures of life-course mobility, such as Spearman's rank correlation (Appendix A.E or the intragenerational elasticities (Appendix A.F), show a similar degree of association between past and present occupations over both the short and long run, and indicate that there was no clear change in life-course mobility over the Victorian period. The association of the two occupations, measured in HISCAM ranks or log HISCAM scores, observed over a 10-year interval stands at between 0.61 and 0.70 for the two periods of 1851–1861 and 1881–1891, which is in a similar range of 0.62 to 0.68 observed in the United States between 1840 to 1910 by Ward (2023). Thus, the rate of intragenerational mobility in nineteenth-century England was not exceptionally low or different to those elsewhere.

Figure 2.1 and 2.2 shows the binned scatterplot of the 10-year and 30-year intragenerational rank-rank correlation for the two periods. The figures demonstrate that occupational trajectories are much more unstable at the top and bottom of the spectrum. There was a large degree of regression to the mean — those who started at the bottom tend to rise sharply, while those who started at the top tend to fall.

The decline in the scores of high-ranking groups may perhaps be less surprising than it appears, given that the propensity for downward mobility is much greater than the propensity for upward mobility, as those at the very top of the occupational scale has very little room to move up but can easily move down. Same can be said for those at the bottom — there is not much scope for downward mobility and plenty for upward mobility. However, the results also imply that, even for those starting in the bottom rungs of the occupational ladder, mobility was possible, while there is no perfect strategy for those starting at the top to guard against downward mobility.

Figure 2.1: Binned Scatterplot of Intragenerational Rank-Rank Correlation, 1851–1881

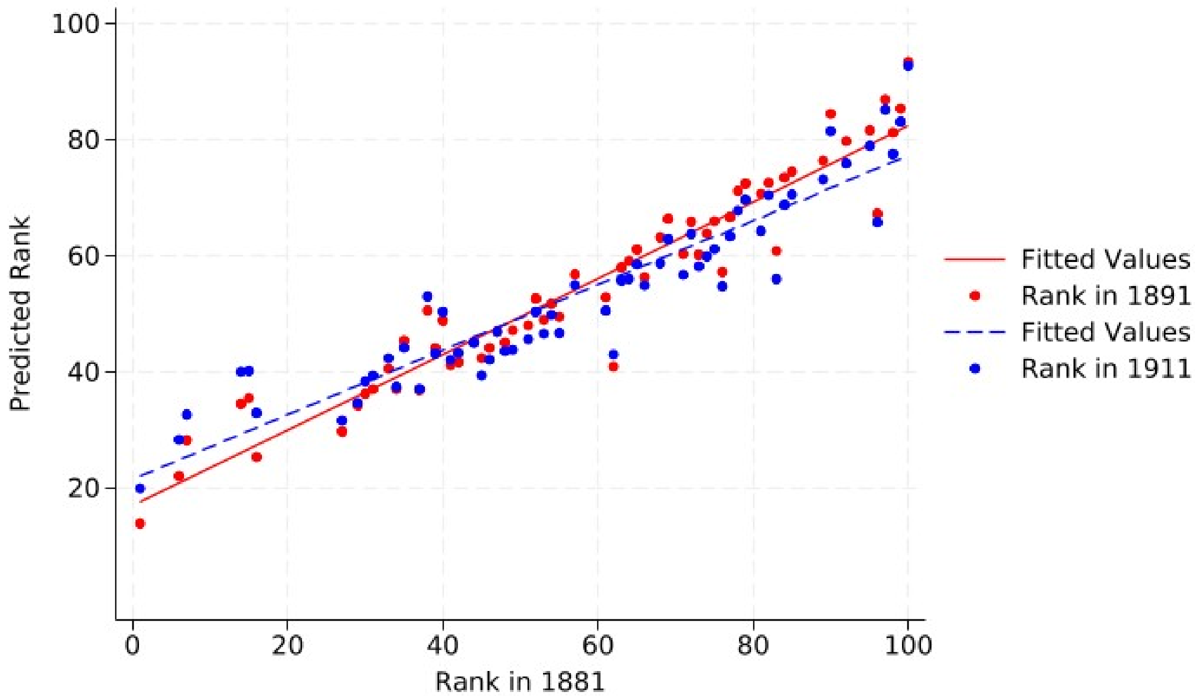


Notes: binned scatterplot based on regression estimated in column 1 of table 2.4 and 2.5.
Sources: author's analysis of I-CeM (Schürer and Higgs (2014), UKDA, SN 7481) and I-CeM Names and Addresses (Schürer and Higgs (2015), UKDA, SN 7856).

2.5.2 Class Mobility Rates

Table 2.6 compares the mobility rates across 10 and 30 years for the periods 1851–1881 and 1881–1911. The results show that around 30 per cent of people move between occupational classes across a 10 year interval. Upward mobility predominates over downward mobility, as the share of people moving upward is between 18 to 20 per cent, while the share of people moving downward is between 10 to 12 per cent. The same is true when we observe occupational mobility across 30 year periods. In both periods, upward mobility across the 30-year interval is around 23 to 26 per cent, while downward mobility is around 11 to 17 per cent. Such findings are in line with expectation, since we would expect people to move up the ladder as they progress in their careers, so upward mobility should be more common

Figure 2.2: Binned Scatterplot of Intragenerational Rank-Rank Correlation, 1881–1911



Notes: binned scatterplot based on regression estimated in column 5 of table 2.4 and 2.5.
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).

than downward mobility.

The comparison of mobility rates also suggest that life-course mobility remained stable across the Victorian period. The three types of mobility — upward, downward, and total — remained at similar levels going from 1851–1881 to 1881–1911. This paper also finds similar levels of intragenerational upward mobility to Long (2013) for 1851–1881, but less downward mobility.

Looking at the share of people who are occupationally mobile over the life course, shown in the transition matrices in Appendix A.G, the most mobile group of people in 1851–1881 over both short- and long-term appear to be those in Class II — the lower professionals and services workers. On the other hand, the most mobile group of people in 1881–1911 are those in Class V — the unskilled workers at the bottom of the ladder.

Table 2.6: Mobility Rates, 1851–1911

Period	Upward	Downward	Total Mobility
10-Year Mobility Rates			
1851–1861 (Unweighted)	19.1	9.7	28.8
1851–1861 Weighted)	19.8	10.8	30.6
1881–1891 (Unweighted)	17.7	12.0	29.7
1881–1891 (Weighted)	17.7	12.9	30.5
30-Year Mobility Rates			
1851–1881 (Unweighted)	25.5	11.4	36.9
1851–1881 (Weighted)	26.2	13.0	39.3
1881–1911 (Unweighted)	23.9	15.0	38.9
1881–1911 (Weighted)	23.3	16.5	39.8
1851–1881 (Long, 2013)	25.4	18.2	43.6

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).

2.6 Who were the Most Mobile?

To understand the profiles of those who experienced occupational mobility versus those who did not, we link the census-reported occupations to the PST (Primary, Secondary, Tertiary) system devised by Wrigley (2010*b*) to classify occupations by their sectors. Each occupation is defined by a four-digit code in the PST system, with each digit representing the four levels of sectoral disaggregation: sector, group, section, and occupation. For instance, someone with the code 1-1-1-1 is in the 'primary' sector, 'agriculture' group, 'farming' section, with the occupation of 'farmer'. Table 2.7 shows the number of groups and sections in each of sectors, with the PST-1 level digit shown. There are 128 groups and 418 sections in total (125 and 412 if the sectorally unspecified and those without occupations are excluded). This section will analyse occupational mobility at the PST-1 (sector) and PST-2 (group) level.

The first step is to examine which sector has the highest intragenerational occupational mobility. Figure 2.3 shows the breakdown of occupational mobility (and immobility) by changes across occupations and sectors for 1851 to 1861. The black sections represent the proportion of people from each sector who kept the same occupation across the two censuses. The dark grey sections show the percentage of people who held two occupations with the

Table 2.7: Number of Groups and Sections in PST

Sector	Group	Section
1. Primary	7	18
2. Secondary	37	154
3. Tertiary (Dealers)	30	102
4. Tertiary (Sellers)	28	63
5. Tertiary (Services and Professions)	16	64
6. Tertiary (Transport and Communications)	7	11
90. Sectorally Unspecific Occupations	0	0
99. Without Occupations or Unstated	3	6
Total	128	418

Sources: Wrigley (2010*b*).

same PST-1 and PST-2 codes, while the light grey sections are the percentages of those with occupations from the same sector (PST-1) but different groups (PST-2).

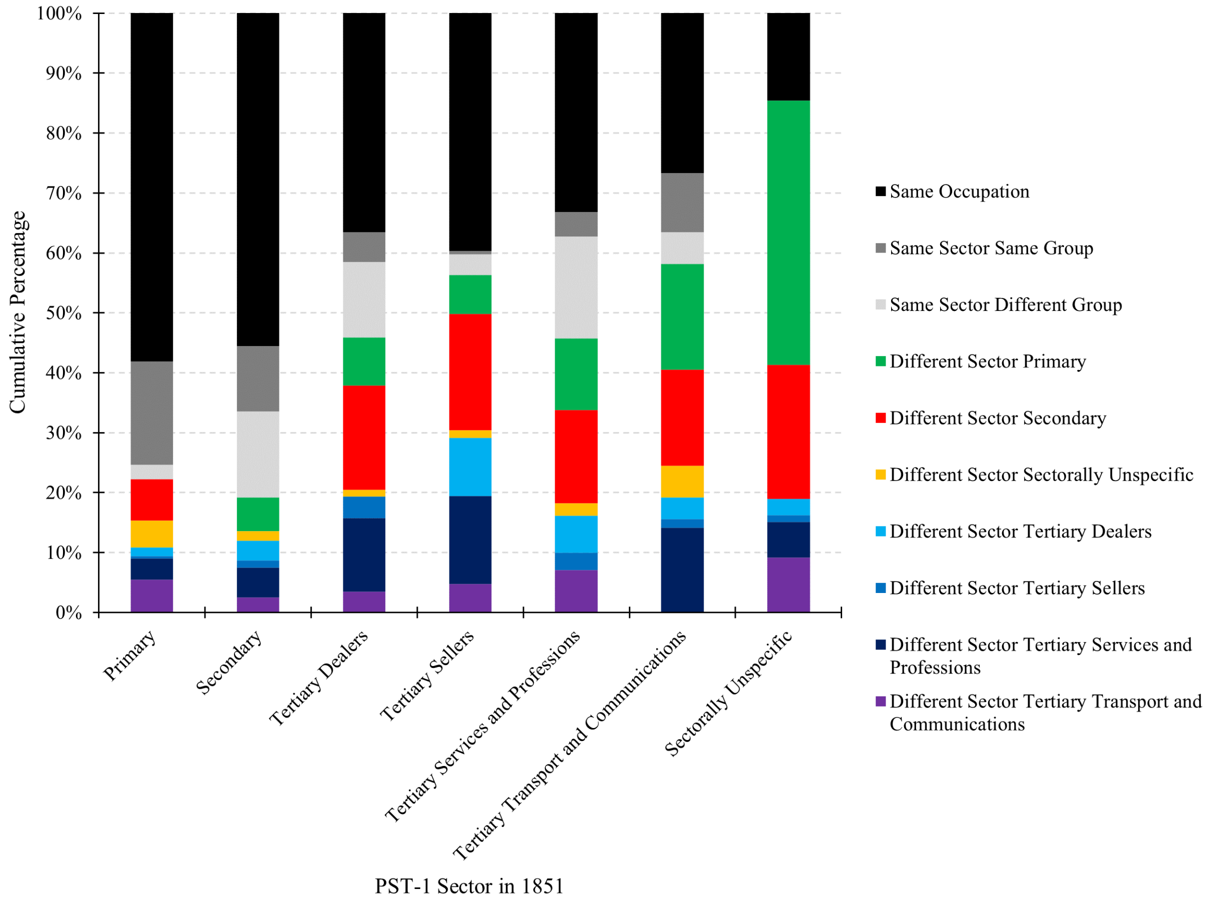
The results in the figure suggest that occupational mobility is very low in the primary and secondary sectors, with over 50 per cent of people having the same occupations over 10 years, and circa 80 per cent staying in the same sector. The tertiary sectors, in contrast, are much more mobile, with 50 per cent or more of people switching to a different sector after 10 years.[13] The turnover of workers in transport and communications occupations appear to be especially high, with less than 30 per cent staying in the same occupation in the span of a decade.[14]

Moreover, when we examine the sectoral changes in occupations, represented by the various coloured sections, we see that people who exited their sectors tend to end up in the secondary sector or tertiary sector, and are less inclined to move into the primary sector. In particular, among the tertiary sector occupations, services and professions appear to have

[13]Even though they are relatively more mobile, at least 75 per cent or more of those employed as dealers, sellers, or in the services and professions, still stayed in the tertiary sector as a whole.

[14]The 'sectorally unspecific' sector is a special case. These are made up of one occupation only — the general labourers. The low rates of people reporting the same occupations over two censuses may reflect real occupational changes or a more precise reporting of occupations leading to them being assigned a sector. For instance, over 70 per cent of the former general labourers in the primary sector are agricultural labourers. They could have changed their jobs from being urban labourers to farm labourers, or they simply reported their jobs more accurately in 1861 than in 1851. Thus, it would be difficult to draw a conclusion on the occupational mobility of general labourers from the available information.

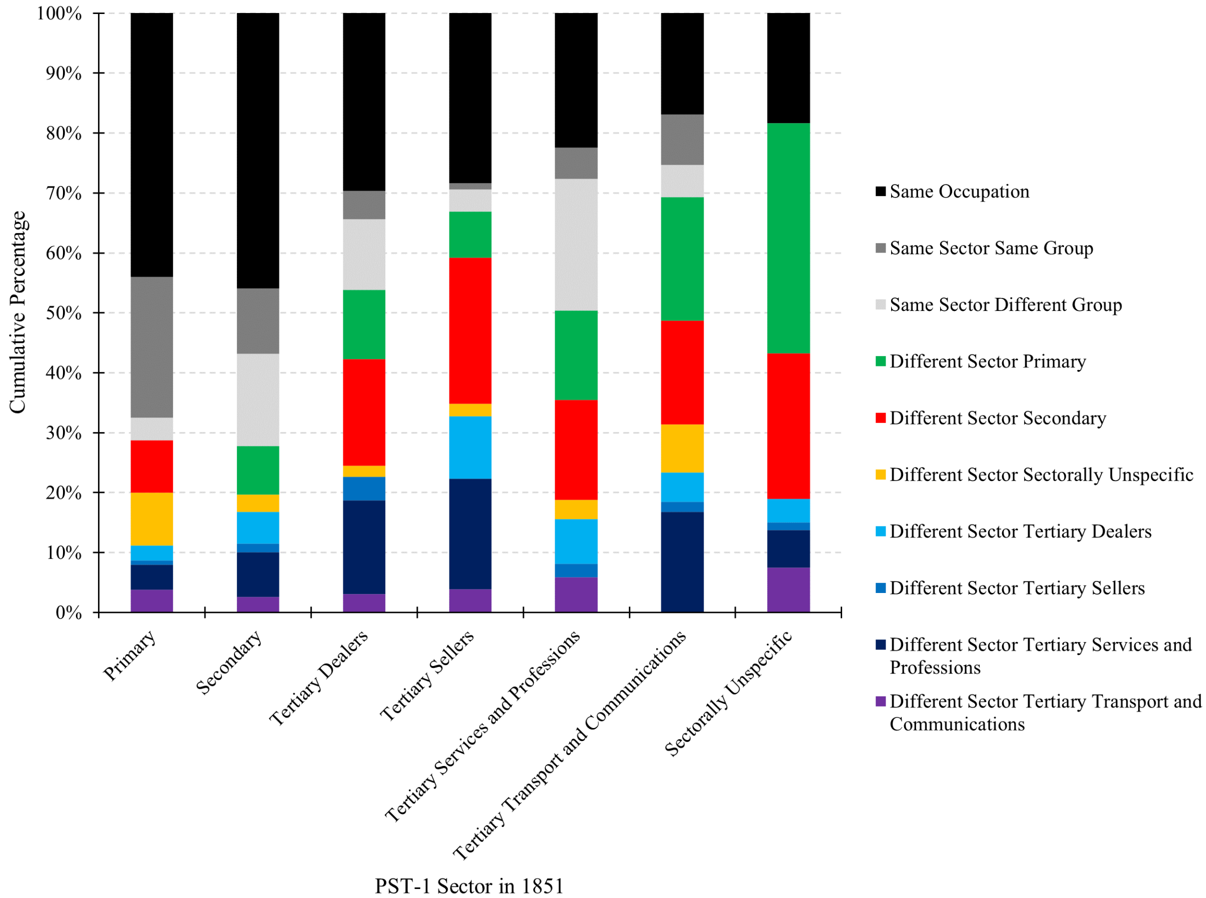
Figure 2.3: Occupational Changes, 1851–1861



Notes: individuals with the same occupational codes in the I-CeM data are classed as 'same occupations'; individuals with different occupational codes but the same PST-1 and PST-2 code are classed as 'same sector same group' while those with the same PST-1 but different PST-2 code are classed as 'same sector different group'; all the rest have different occupations and different PST-1 (therefore also PST-2) codes. *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); raw numbers shown in Appendix A.H, table A.H.1; PST classification based on Wrigley (2010*b*).

been growing the most with the largest intakes. The trends in occupational mobility therefore follow the evolution in England's occupational structure since the Industrial Revolution. Between 1710 and 1871, England and Wales witnessed the continued rise in secondary sector employments from 37 per cent in 1710 to 46 per cent in 1871 and the explosion of the tertiary sector from 12 per cent to 28 per cent, at the expense of primary sector's share of employments which halved during the same period (Shaw-Taylor and Wrigley, 2014).

Figure 2.4: Occupational Changes, 1851–1881

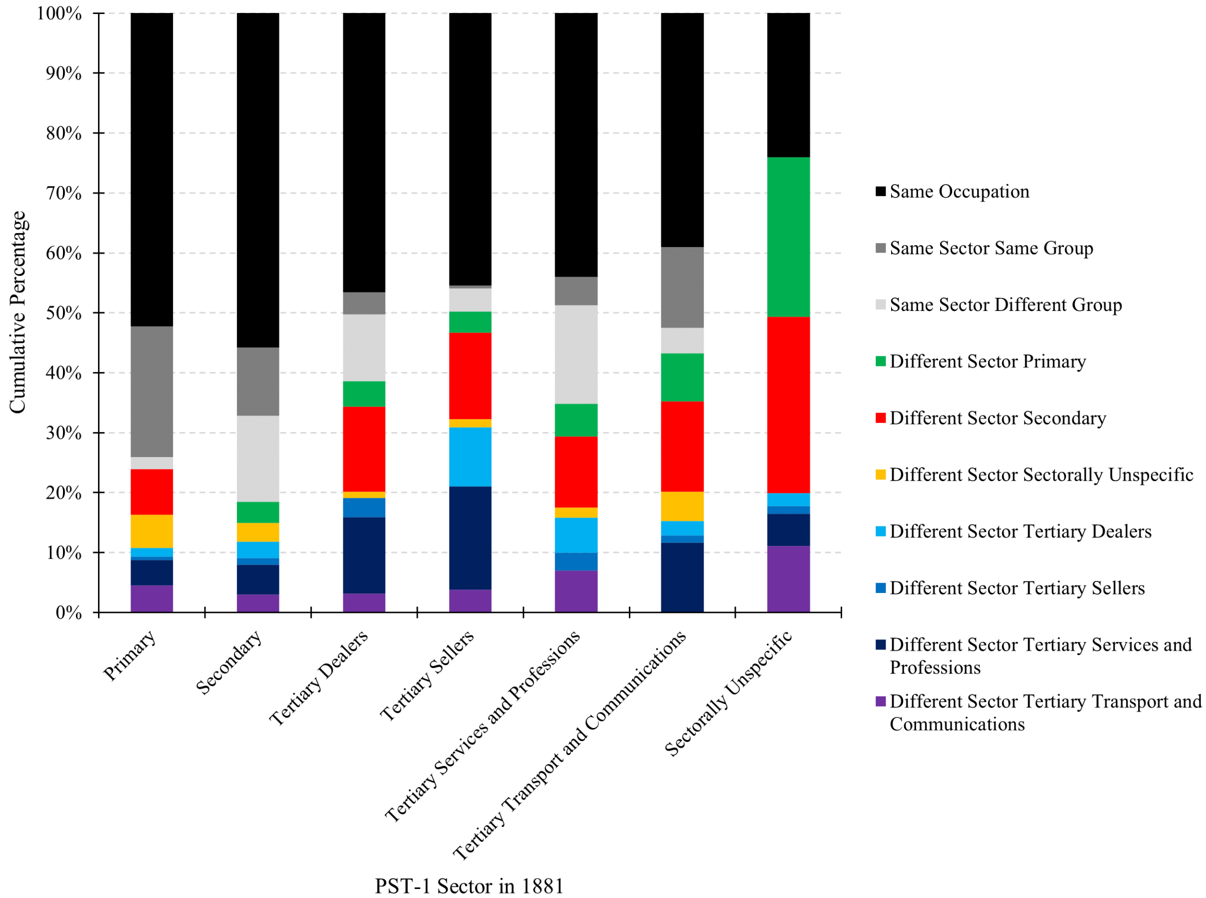


Notes: individuals with the same occupational codes in the I-CeM data are classed as 'same occupations'; individuals with different occupational codes but the same PST-1 and PST-2 code are classed as 'same sector same group' while those with the same PST-1 but different PST-2 code are classed as 'same sector different group'; all the rest have different occupations and different PST-1 (therefore also PST-2) codes. *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); raw numbers shown in Appendix A.H, table A.H.2; PST classification based on Wrigley (2010*b*).

Nevertheless, the high proportion of people who are occupationally stable in the primary and secondary sector reflect the slower pace of structural change in the nineteenth century.

Figure 2.4 shows the sectoral changes across a 30-year period, from 1851 to 1881. Much of the trends are similar, albeit more pronounced, to the sectoral changes from 1851 to 1861. The primary and secondary sectors remain the most stable sectors, with almost 30 per cent moving to a different sector. Among the tertiary sectors, sellers and transport

Figure 2.5: Occupational Changes, 1881–1891



Notes: individuals with the same occupational codes in the I-CeM data are classed as 'same occupations'; individuals with different occupational codes but the same PST-1 and PST-2 code are classed as 'same sector same group' while those with the same PST-1 but different PST-2 code are classed as 'same sector different group'; all the rest have different occupations and different PST-1 (therefore also PST-2) codes. *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); raw numbers shown in Appendix A.H, table A.H.3; PST classification based on Wrigley (2010*b*).

and communications are the most mobile. Once again, a large proportion of people exiting their initial occupation's sector tend to end up in the secondary sector or in services and professions.

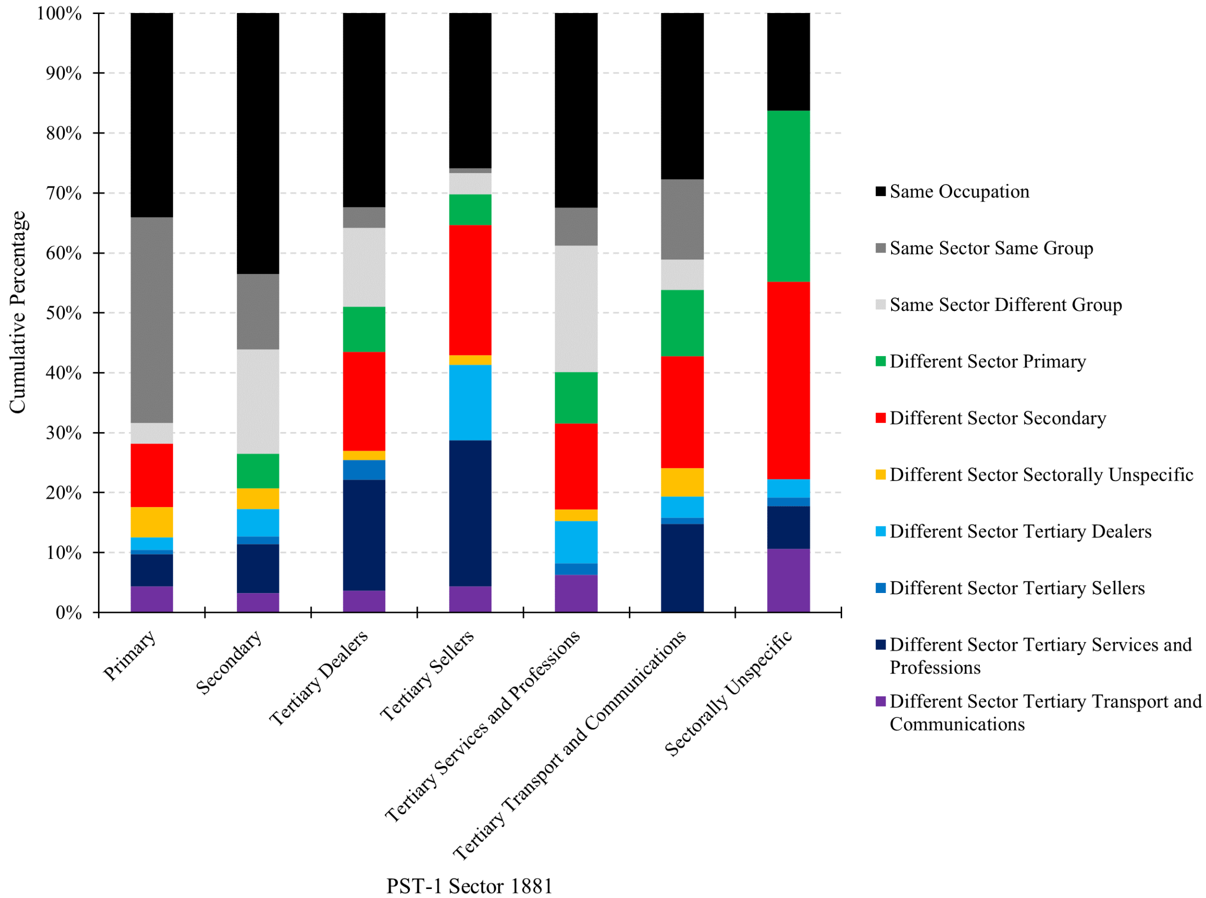
For the period of 1881–1911, there are some similarities to be observed in both the 10-year and 30-year sectoral changes compared to the earlier period. The changes from 1881 to 1891, shown in figure 2.5, suggest that the primary and secondary sectors remain the most occupationally stable sectors. Although the share of individuals remaining in the same occupation is much more similar across the different sectors, the share of people remaining in the same sector is much higher in the primary and secondary sectors. The rate of exit is much higher in the tertiary sectors and, akin to the earlier period, people who are intragenerationally mobile tend to enter into the secondary sector or into services and professions. As the agriculture continues its gradual decline, the share of people entering into this sector is even smaller than the 1851–1881 period.

Finally, figure 2.6 demonstrates the occupational changes across sectors over the 30-year period from 1881 to 1911. The results show that over the longer time frame, much fewer people remain in the same occupations they held 30 years ago — only between 30 to 40 per cent. However, there is still a tendency to remain in the same sectors in some cases, such as in the primary and secondary sectors, and to a lesser extent in services and professions. The sellers are the most occupational unstable, with 70 per cent exiting the sector by the latter stages of their careers, followed by transport and communications and the dealers. Those who exit their initial sectors, as we have seen in previous cases, tend to favour the secondary sector and the services and professions sector.

Having established the sectoral trends in intragenerational occupational mobility, the second part of this analysis is to unpack in greater details the most and least mobile group of occupations. Some restrictions are imposed: to exclude occupational groups which have too few numbers, only PST-2 groups that account for at least 0.1 per cent of the linked sample are included (meaning at least 132 individuals for 1851–1881 and 238 for 1881–1911), and people who began their careers with occupations that do not have a PST-2 group or with sectorally unspecified occupations are also excluded. This leaves us with 56 eligible groups for 1851–1881 and 65 groups for 1881–1911; they account for at least 93 per cent of the linked sample so the proportion of people excluded is not significant.

Table 2.8 shows the top 10 most and least mobile occupational groups between 1851 and 1861, measured by the proportion of incumbents of each group in 1851 who stayed in the

Figure 2.6: Occupational Changes, 1881–1911



Notes: individuals with the same occupational codes in the I-CeM data are classed as 'same occupations'; individuals with different occupational codes but the same PST-1 and PST-2 code are classed as 'same sector same group' while those with the same PST-1 but different PST-2 code are classed as 'same sector different group'; all the rest have different occupations and different PST-1 (therefore also PST-2) codes. *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); raw numbers shown in Appendix A.H, table A.H.4; PST classification based on Wrigley (2010*b*).

same PST sector by 1861. For this period, the most stable occupations are concentrated in the secondary sector, which produced 9 out of 10 occupational groups with the highest shares remaining in the same sector, alongside mining and quarrying occupations from the primary sector. Occupational stability also took different forms among these groups. For instance, over three quarters of those working in 'instrument making' or 'footwear' remained in the same occupation between 1851 and 1861, which signals a high level of occupational security but also very low level of mobility. On the other hand, just over a third of workers in 'non-ferrous metal manufacture and products' stay in the same occupation; this number increases to half if we include those who changed into a different occupation within the same occupational group (non-ferrous metal manufacture and products), and to 83 per cent if including all secondary sector occupations. Predictably, occupational stability is reflected by low levels of changes in the average ranks of workers from each occupational group, as the comparison of average HISCAM ranks between 1851 and 1861 shows.

In terms of the most mobile occupational groups, it is immediately evident that they are concentrated in the tertiary sectors. Occupational groups belonging to traders, sellers, services and professions, and transport and communications can all be found among the top 10 of the most mobile groups. Particularly striking is the high exit rates from high-ranking occupational groups, such as 'fuel dealers', coupled with a large decline in average rank in 1861. It suggests that there is a trade-off between high status and occupational stability or security in these occupations, with large risks of downward mobility.

A similar trend can be seen across the 30-year time interval for 1851–1881 from Table 2.9. Again, the most stable occupations are overwhelmingly found in the secondary sector, although two groups — 'professions' and 'national government service' — from the services and professions sector are also present. These two high-ranking occupational groups therefore offer the attractive combination of high status and security. In terms of the most mobile groups, there is a complete overlap between the 10 groups in 1851–1861 and in 1851–1881. It appears that those working in the tertiary sector tend to be the most mobile (and least stable) over the short- and long-term.

In addition, the role of domestic service in the process of leaving home can be observed as well. Entering into domestic service was often the avenue to leaving home in Victorian England for young women (Day, 2018), but there is also a possibility that this applied to a significant minority of young men too. Since men tended to leave their parents' home at a later age than women — the majority around the age of 21 (Day, 2018) — it would explain

Table 2.8: Top 10 Most and Least Mobile PST-2 Groups, 1851–1861

		Percentage of Same			Average Rank		
PST Code	Group (PST-2 Level) Description	Occupation	PST Group	PST Sector	1851	1861	*N*
Top 10 Same Sector							
2-70	Road transport vehicles	62.75	65.19	89.76	79.99	73.37	1396
2-52	Instrument making	75.77	76.74	89.66	88.67	83.12	619
2-80	Building and construction	67.73	77.87	87.05	59.80	58.86	10949
2-61	Iron and steel manufacture and products	64.24	71.43	86.10	58.37	56.27	5092
2-15	Footwear	78.59	78.76	84.34	52.39	49.50	6003
2-62	Non-ferrous metal manufacture and products	37.82	50.69	83.40	64.43	57.76	1018
1-20	Mining and quarrying	55.07	78.21	82.88	11.01	17.07	5457
2-50	Precious metals and jewelry	43.33	55.76	82.73	65.64	67.92	330
2-25	Wood industries	56.96	64.51	82.66	65.35	60.65	2860
2-71	Boat and ship building	56.52	68.16	80.82	79.27	74.45	782
Top 10 Different Sector							
4-90	Small traders	26.92	27.78	29.91	59.12	63.15	234
5-25	Domestic service	14.69	17.82	30.20	5.43	44.83	2076
3-58	Fuel dealers	25.13	25.13	32.62	91.00	65.11	187
6-5	Rail transport	19.23	32.53	37.00	44.45	52.04	1165
6-1	Road transport (animal power)	23.94	34.21	37.43	57.32	45.20	2669
3-85	Dealers in minor products	7.43	10.14	37.84	90.87	72.63	148
3-10	Dealers in clothing and clothing accessories	24.79	25.63	38.11	91.00	66.21	593
4-41	Sellers of printed products	33.92	34.41	39.07	97.74	80.32	622
5-1	Food, drink and accommodation services	17.05	21.59	40.00	81.78	67.90	440
5-10	Storage	20.05	24.01	42.22	49.00	65.21	379

Notes: first digit in PST code denotes the sector: 1 = Primary, 2 = Secondary, 3 = Tertiary Dealers, 4 = Tertiary Sellers, 5 = Tertiary Services and Professions, 6 = Tertiary Transport and Communications; percentage of Same is cumulative from the most specific (at the occupational level) to the least (at the sectoral level).
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); assignment of occupations to PST classification based on Wrigley (2010*b*).

Table 2.9: Top 10 Most and Least Mobile PST-2 Groups, 1851–1881

		Percentage of Same			Average Rank		
PST Code	Group	Occupation	PST Group	PST Sector	1851	1881	*N*
Top 10 Same Sector							
2-52	Instrument making	70.60	71.41	85.46	88.67	78.63	619
2-70	Road transport vehicles	52.51	55.95	85.24	79.99	69.62	1396
2-80	Building and construction	59.40	74.18	83.09	59.80	57.80	10949
2-50	Precious metals and jewelry	36.97	50.91	82.73	65.64	67.80	330
5-35	Professions	51.14	60.55	78.41	97.16	88.97	2010
5-42	National government service	38.97	45.59	77.21	99.20	90.31	136
2-61	Iron and steel manufacture and products	51.77	61.31	76.22	58.37	54.15	5092
2-71	Boat and ship building	49.74	60.49	75.45	79.27	71.24	782
2-62	Non-ferrous metal manufacture and products	28.29	43.91	75.15	64.43	56.76	1018
2-35	Furnishing	44.44	46.53	75.00	43.99	45.81	288
Top 10 Different Sector							
4-41	Sellers of printed products	5.63	6.75	12.86	97.74	77.33	622
6-1	Road transport (animal power)	12.44	19.48	22.29	57.32	45.07	2669
3-58	Fuel dealers	18.18	18.18	23.53	91.00	62.41	187
5-25	Domestic service	7.47	11.32	27.70	5.43	49.60	2076
3-85	Dealers in minor products	4.05	5.41	27.70	90.87	67.95	148
4-90	Small traders	23.50	25.21	29.91	59.12	60.29	234
6-5	Rail transport	11.16	25.92	30.04	44.45	53.27	1165
3-10	Dealers in clothing and clothing accessories	14.33	15.18	32.88	91.00	65.29	593
5-1	Food, drink and accommodation services	10.00	18.86	32.95	81.78	63.32	440
5-10	Storage	10.82	12.66	34.04	49.00	68.01	379

Notes: first digit in PST code denotes the sector: 1 = Primary, 2 = Secondary, 3 = Tertiary Dealers, 4 = Tertiary Sellers, 5 = Tertiary Services and Professions, 6 = Tertiary Transport and Communications; percentage of Same is cumulative from the most specific (at the occupational level) to the least (at the sectoral level).
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); assignment of occupations to PST classification based on Wrigley (2010*b*).

the prominence of domestic service occupations as one of the most occupationally mobile groups. The results suggest that young men entered into domestic service as their first job to support themselves and their newfound independence, and exited the industry later on in their careers. This is reflected by the low share of domestic servants remaining in the same occupation and same group in 1861 and even lower by 1881. The large increase in average ranks over 10 and 30 years is also a consequence of both the low status of domestic servants and the nature of domestic service being a temporary first job for most rather than a lifelong vocation.

Towards the end of the Victorian era, there remains the tendency for secondary sector workers to be occupationally stable (or immobile). Table 2.10 demonstrates that between 1881 and 1891, the vast majority of the 10 most stable occupational groups are from the secondary sector, mixed with 2 from services and professions. The most mobile occupational groups, once again, is comprised of occupations from the tertiary sectors of dealers, sellers, services and professions, and transport and communications.

Whilst the group of workers most likely to generate life-course mobility, over the 30 year interval from 1881 to 1911 (table 2.11), in terms of shifting to a new job in a new sector remains largely consistent with those observed in 1881–1891 as well as 1851–1881, there are some changes in the composition of stable occupational groups. Here, we see that more and more occupational groups from services and professions emerge as intragenerationally secure, such that the top 10 most stable groups are now split evenly between the secondary sector and the tertiary services and professions sector. Such change reflects the growth of the tertiary sector, in particular the more 'modern' professional, financial, governmental, and media occupations, as opposed to the more 'traditional' occupations such as domestic service, which were relatively stagnant if not in decline (Thomas, 2004).

Lastly, figure 2.7 details the ranks of all 56 occupational groups in both 1861 and 1881, with the size of the 'bubbles' weighted by the numbers employed in these groups. The graph clearly demonstrates that the top right corner, which contain occupational groups that had

Table 2.10: Top 10 Most and Least Mobile PST-2 Groups, 1881–1891

		Percentage of Same			Average Rank		
PST Code	Group	Occupation	PST Group	PST Sector	1881	1891	N
Top 10 Same Sector							
2-70	Road transport vehicles	64.68	70.28	90.33	71.22	66.23	2503
2-62	Non-ferrous metal manufacture and products	41.07	59.77	86.76	49.95	48.48	2145
2-41	Printing	50.35	75.39	86.73	76.73	72.29	2540
2-67	Gas equipment	32.46	42.54	86.69	44.00	52.22	496
2-80	Building and construction	69.05	78.64	86.68	49.42	50.54	24853
5-35	Professions	69.24	74.09	86.48	94.01	88.27	5747
5-16	Media	53.06	71.94	86.39	93.18	89.50	360
2-15	Footwear	80.68	81.34	86.32	43.15	42.46	5532
2-52	Instrument making	70.81	71.38	85.20	81.86	76.97	1216
2-31	Industries producing products from fibres	70.37	70.37	84.07	41.32	39.81	540
Top 10 Different Sector							
6-2	Road transport (motorised)	1.29	1.29	13.14	53.47	53.81	388
5-50	Armed forces	10.01	15.34	31.63	40.67	45.91	1369
3-85	Dealers in minor products	9.97	10.93	33.76	84.09	63.95	311
4-41	Sellers of printed products	36.84	37.58	40.77	97.12	82.02	1474
6-4	Sea transport	26.76	36.12	43.48	70.25	52.25	1794
5-10	Storage	24.23	26.51	44.13	38.00	57.85	1226
6-1	Road transport (animal power)	29.30	40.16	44.29	40.21	38.86	5884
5-25	Domestic service	29.32	33.24	46.41	21.29	40.33	3646
3-58	Fuel dealers	42.14	42.14	46.95	85.02	67.24	541
4-10	Sellers of clothing and clothing accessories	47.47	47.47	48.25	79.00	72.28	257

Notes: first digit in PST code denotes the sector: 1 = Primary, 2 = Secondary, 3 = Tertiary Dealers, 4 = Tertiary Sellers, 5 = Tertiary Services and Professions, 6 = Tertiary Transport and Communications; percentage of Same is cumulative from the most specific (at the occupational level) to the least (at the sectoral level).
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); assignment of occupations to PST classification based on Wrigley (2010*b*).

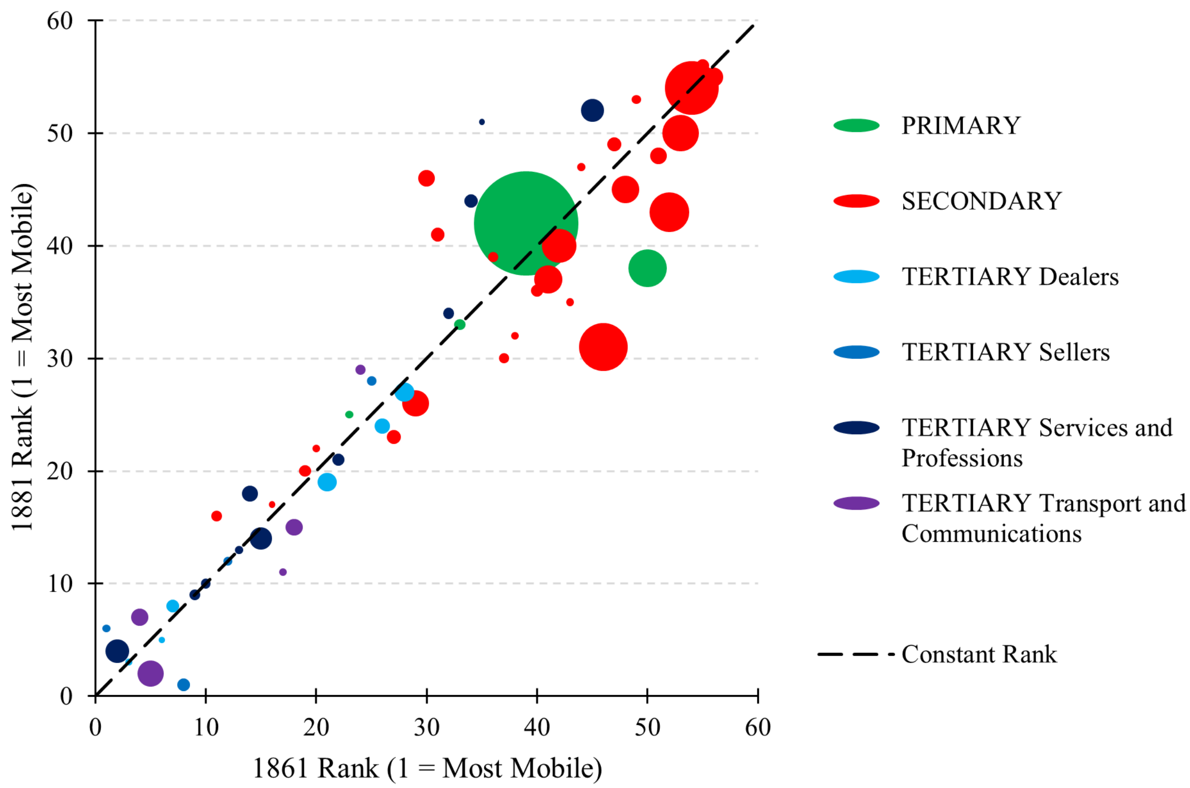
Table 2.11: Top 10 Most and Least Mobile PST-2 Groups, 1881–1911

		Percentage of Same			Average Rank		
PST Code	Group	Occupation	PST Group	PST Sector	1881	1911	*N*
Top 10 Same Sector							
5-35	Professions	61.37	67.69	84.81	94.01	86.79	5747
2-70	Road transport vehicles	48.54	56.41	83.06	71.22	62.03	2503
2-80	Building and construction	58.85	73.08	82.88	49.42	49.62	24853
5-16	Media	44.17	61.11	81.94	93.18	86.32	360
2-62	Non-ferrous metal manufacture and products	29.37	47.69	80.00	49.95	48.31	2145
2-41	Printing	39.45	67.20	79.37	76.73	69.77	2540
5-30	Financial services and professions	39.78	59.38	79.03	95.99	87.18	2265
5-36	Professional support	26.16	28.21	78.03	94.28	85.66	1953
5-42	National government service	31.03	46.26	77.87	99.04	87.75	348
2-67	Gas equipment	19.96	26.21	77.42	44.00	51.60	496
Top 10 Different Sector							
6-2	Road transport (motorised)	0.52	0.52	11.08	53.47	51.60	388
4-41	Sellers of printed products	7.06	8.55	12.35	97.12	78.28	1474
3-85	Dealers in minor products	6.11	7.72	21.54	84.09	62.30	311
4-10	Sellers of clothing and clothing accessories	21.40	21.79	22.96	79.00	67.34	257
5-50	Armed forces	1.97	4.09	24.40	40.67	47.10	1369
4-40	Sellers of paper products	24.10	24.10	31.29	79.00	78.21	278
6-1	Road transport (animal power)	17.59	26.70	31.41	40.21	40.33	5884
4-90	Small traders	25.81	26.17	31.41	54.16	58.92	554
3-58	Fuel dealers	28.10	28.10	33.83	85.02	63.25	541
6-4	Sea transport	17.28	28.65	36.90	70.25	50.16	1794

Notes: first digit in PST code denotes the sector: 1 = Primary, 2 = Secondary, 3 = Tertiary Dealers, 4 = Tertiary Sellers, 5 = Tertiary Services and Professions, 6 = Tertiary Transport and Communications; percentage of Same is cumulative from the most specific (at the occupational level) to the least (at the sectoral level).
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); assignment of occupations to PST classification based on Wrigley (2010*b*).

the highest share of workers remain in their original (1851) occupational sector in both 1861 and 1881, is populated by mainly primary and secondary sector occupations, particularly by groups that accounted for a large share of the occupational structure. By contrast, the bottom left corner, which contain the most mobile occupational groups, are all tertiary sector occupations.

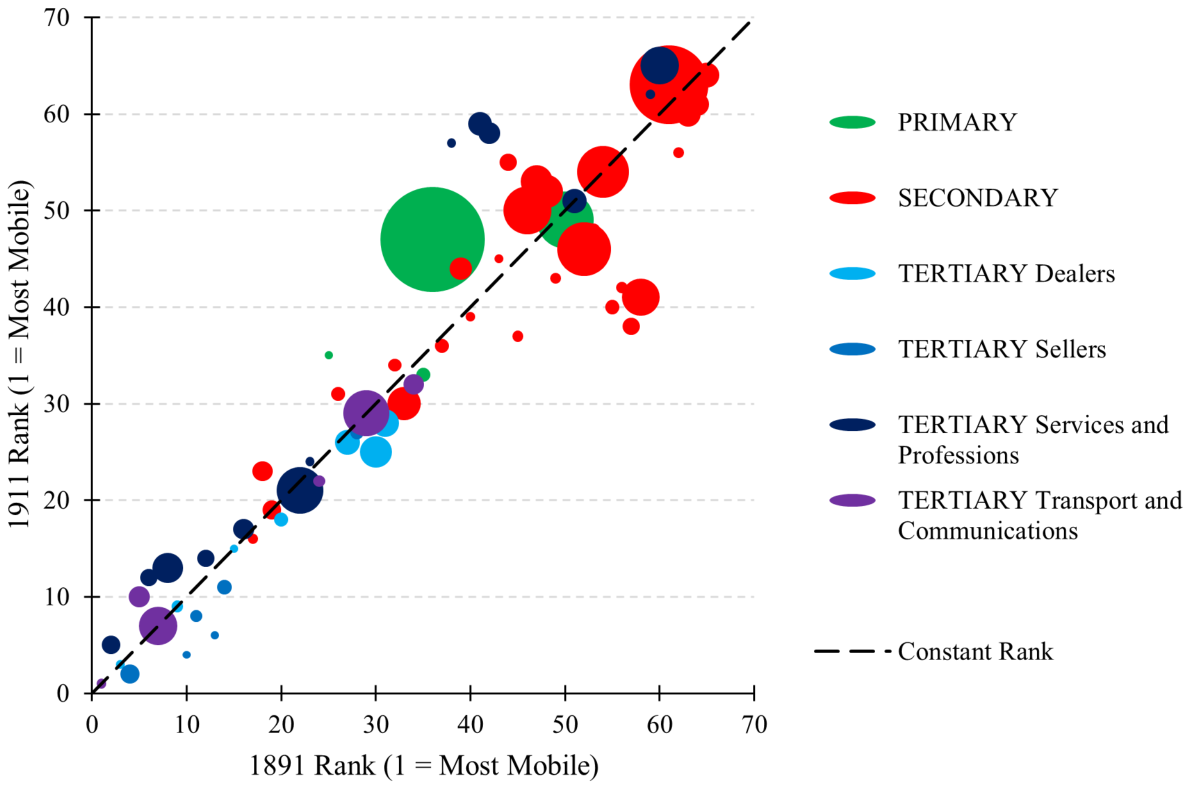
Figure 2.7: PST-2 Group Ranks, 1851–1881



Notes: the x-axis denotes the rank of each PST group in the order of most to least mobile in 1861, and the y-axis denotes rank of each PST group in the order of most to least mobile in 1881; more details on the rank of each occupational group can be found in Appendix A.I, table A.I.
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).

Figure 2.8 shows the same analysis for the 1881–1911 period. Some changes are evident upon comparing with the graph for 1851–1881. The size of the bubbles for tertiary sector occupational groups are on the whole larger for the later period, which indicate the rising share of tertiary sector occupations in the sample. While the least mobile occupational groups are still dominated by the primary and secondary sector, and the most mobile by the tertiary sector, there is a much larger presence of the tertiary services and professions occupations in the top right corner. This is due to the emergence of new occupational groups from this sector as stable and low mobility occupations, which table 2.11 had already demonstrated.

Figure 2.8: PST-2 Group Ranks, 1881–1911



Notes: the x-axis denotes the rank of each PST group in the order of most to least mobile in 1891, and the y-axis denotes rank of each PST group in the order of most to least mobile in 1911; more details on the rank of each occupational group can be found in Appendix A.I, table A.I.
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).

Overall, the analysis of sectoral occupational mobility presents a number of stylised facts on life-course mobility in Victorian England. The limited extent of short- and long-term intragenerational mobility between 1851 and 1911 can to attributed to the highly stable primary and secondary sector occupations, which still employed the majority of the population at this time. Nonetheless, a significant and growing minority of those employed in the tertiary sector did experience a remarkable degree of intragenerational mobility, as evidenced by their tendency to take up occupations outside of their original sector. In particular, a number of occupations involved in dealing and selling appeared to have generated a significant degree of downward mobility, representing a trade-off between high initial status and occupational security. On the other hand, people who started out as domestic servants experienced sizeable upward mobility on average.

The emergence and proliferation of more modern occupations in services and professions, which were high status but also equally secure as many secondary sector occupations, may have counteracted any positive effects on occupational mobility from the growth of the tertiary sector over the second half of the nineteenth century. This may be a signal that life-course mobility through occupational changes is being phased out. As occupations become increasingly secure and careers more stable over the course of the twentieth century, occupational mobility no longer remains the predominant means of achieving social mobility, which instead operates through other channels such as education or income mobility.

2.7 Conclusion

In conclusion, this paper estimates life-course occupational mobility using a number of methods and show that intragenerational mobility was possible but limited in England between 1851 and 1911. The intragenerational rank-rank correlation — our preferred measure — show that the correlation between past and current occupational rank is between 0.61-0.68 over the short-run (10 years) and between 0.50-0.57 over the long-run (30 years). In other words, increasing the percentile rank of one's starting occupational rank by 10 ranks implies an increase of as much as 5.7 ranks by the end of his career. The size of the intragenerational rank-rank correlation is similar to the father-son correlation in occupational ranks documented for the same period in other studies. Estimates of class mobility suggest that around 30 per cent of people experienced class mobility over a decadal interval, and less than 40 per cent over the course of their entire careers. The results therefore depict a picture of Victorian England where a great majority of men do not experience significant occupational mobility across their careers.

Analyses of occupational changes at the more disaggregated sectoral and group levels reveal a number of interesting stylised facts about the nature of occupational mobility in this period. The primary and secondary sectors accounts for the lack of occupational immobility over the life course, as people employed in these sectors tend to be occupationally secure — evidenced by the high shares that remain in the same occupation, group, or sector — with stable average ranks. The tertiary sector is the main driver of life-course mobility, as most of the sectoral changes in occupations occur when people move out of occupations in trading, selling, services and professions, and transport and communications. Despite this, there is a growing presence of new occupations in services and professions which are secure and therefore less likely to precipitate occupational mobility by the end of the nineteenth century.

Given that intergenerational mobility was also limited for Victorian England, as other studies have shown, the lack of life course mobility observed indicates that England at the end of the nineteenth century was a society of marked social persistence. The prospect for social mobility through occupational changes was not realistic for the large majority of men in England during this period. While this is may not be entirely detrimental, as low mobility also implies stability and security, which may be particularly attractive for those with high social status, it does suggest that those with aspirations of rising from the bottom to the top encountered barriers much greater than what was anticipated.

Chapter 3

Like Father, Like Son: Intergenerational Immobility in England, 1851–1911

3.1 Introduction

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 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 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 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 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 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 (2023)'s 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 (2013)'s sample size, in his use of a 2 per cent 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 overestimation of the extent of social mobility. Second, it constructs a high-quality linked sample using the Integrated Census

[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 to different occupational indices. False positives and reweighting do not have a significant impact on my findings.

Microdata (I-CeM) complete-count census data, which greatly expands the number of observations that were previously available to Long (2013). Finally, it devises a new method for estimating the rate of, and consequently correcting for (at 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 organised as follows. Section 3.2 reviews the existing literature on historical social mobility. Section 3.3 presents the data used and the census linking process and outcomes. Section 3.4 outlines the methodology, or how social (occupational) mobility is measured in this paper. The results are shown in Section 3.5; they represent a significant revision from previous works and highlight the impact of measurement errors. Section 3.6 discusses the implications of these results and makes some comparisons across both time and space. Section 3.7 concludes.

3.2 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 (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 (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 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 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 characterising 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 per cent sample of the 1851 census. This raises issues of representativeness while also increasing the likelihood of Type I errors in linking.[3] Moreover, Bailey et al. (2020) 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.[4] Bailey et al. were also sceptical 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 digitising the data.[5] For example, Ward (2023) exploits the

[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 per cent sample, some non-unique individuals may appear as unique if their duplicates are eliminated by the process of sampling.

[4]For instance, Bailey et al. (2020, p. 3) estimate that false links could bias IGE downward by up to 20 per cent.

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

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 per cent 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.

After accounting for measurement errors, 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). Therefore, our understanding of British/English occupational mobility since the Victorian era may be open to scepticism 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

utive censuses, see Appendix B.N.

to be much less mobile than the United States. This could also be subject to amendment if the effects of classical measurement errors are different across countries.

3.3 Data and Census Linking

3.3.1 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 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 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 programme.[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 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 standardisation of raw textual strings. There were over 7.3 million unique strings for occupations and over 6.7 million for birthplace

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

[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.

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.

3.3.2 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 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, is used. Table 3.1 shows a sample of some common occupations observed in the census with their respective HISCAM scores.

Table 3.1: 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 (UKDA, SN 7481); 'HISCAM' is taken from Lambert et al. (2013).

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 et al. (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 et al. (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 (i.e. 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 4-class scheme will be used.[10] Table 2 describes each of the twelve classes in HISCLASS and how they can be combined into the four-class occupational categories, as shown by Antonie et al. (2022).

3.3.3 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 errors, the sons are linked across every 10-year interval and

[8] They also construct a different index using an alternative methodology — principal component analysis. Clark et al. (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 Section 3.4.2
[10] The same classification scheme was used by Long and Ferrie (2013) and Pérez (2019)

Table 3.2: Conversion of HISCLASS Categories to Four-Class Scheme

Levels	HISCLASS Description	Class	Categories
1	Higher managers	W	White-collar
2	Higher professionals	W	White-collar
3	Lower managers	W	White-collar
4	Lower professionals	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 van Leeuwen and Maas (2011); Conversion to four-class scheme follows Antonie et al. (2022).

the fathers are linked across one 10-year interval. This is done for three periods: 1851–1881, 1861–1891, and 1881–1911.[11]

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),

[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.

and inferred birth year from age.[12] The procedure is outlined in Appendix B.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 is motivated by the findings of Bailey et al. (2020), 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.[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

[12] 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.B and carried out for robustness tests. The Jaro-Winkler version of ABE is taken from Abramitzky et al. (2019).

[13] 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.

1861 are the same people as the ones co-residing with the sons in the census. The detailed procedure and results are outlined in Appendix B.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, Boustan, Eriksson, James and Pérez, 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 United States 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, 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. (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, 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.

3.3.4 Census Linking Results

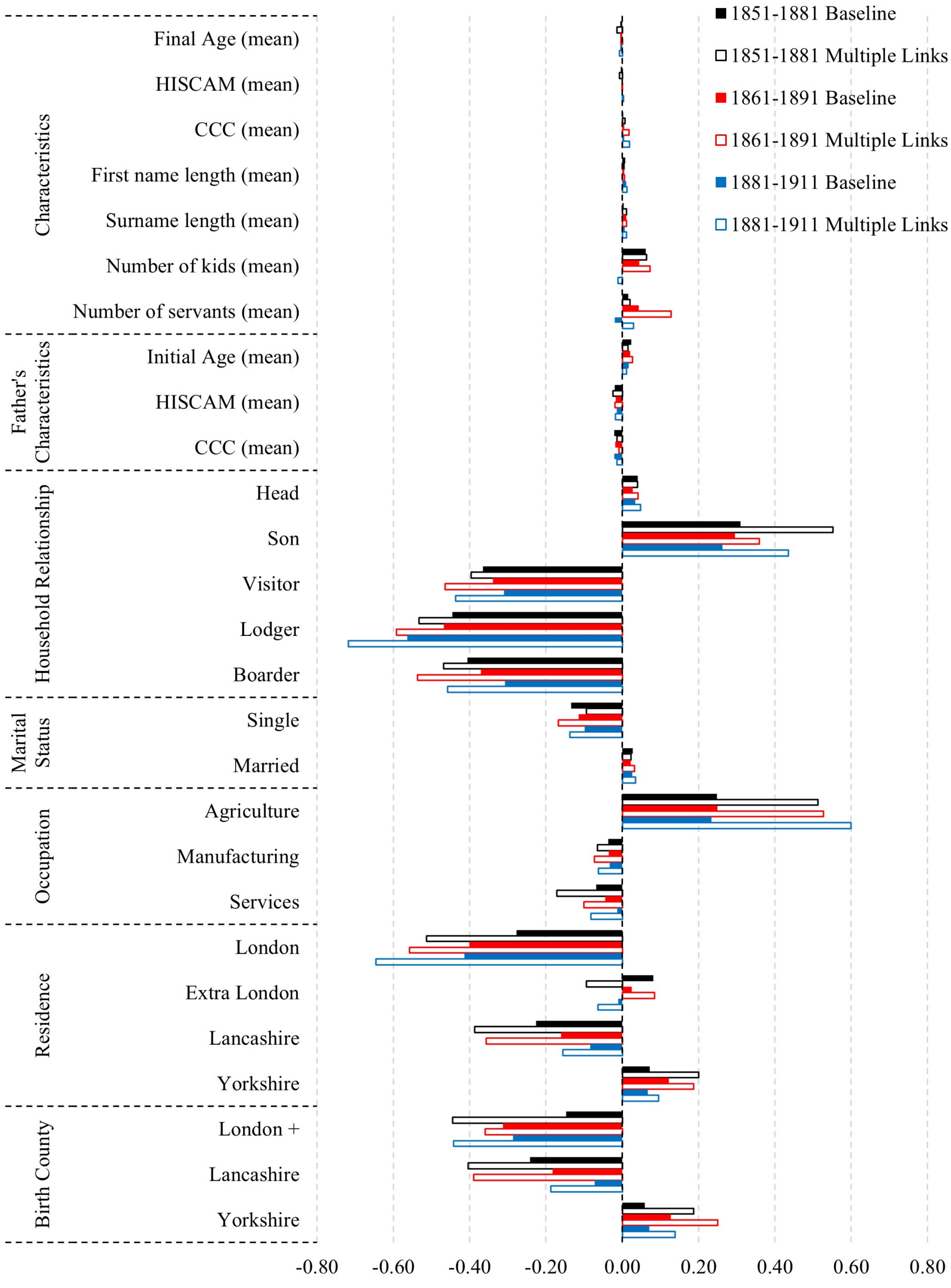
Table 3.3 shows the linkage results for the periods 1851–1881, 1861–1891, and 1881–1911. For the baseline samples, between 290,000 and 610,000 father-son pairs have been successfully matched, which translates to a match rate of 21 to 29 per cent. Upon restricting the sample to sons who can be matched 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 3.1. The numbers are indexed against the population with the population set at 0. 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

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

Figure 3.1: Sample Representativeness, 1851–1911



Notes: calculated based on Appendix B.E, Table B.E.1, B.E.2, B.E.3. 'Extra London' refers to the counties Middlesex, Kent, Essex, and Surrey. 'London+' refers to Middlesex, Kent, Essex, and Surrey (there is no 'London' in the standardised county of birth code, only 'City of London', which is a tiny part of London with no one recorded as being born there). 'Yorkshire' includes all Ridings of Yorkshire. *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).

Table 3.3: 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 (Schürer and Higgs (2014), UKDA, SN 7481) and I-CeM Names and Addresses (Schürer and Higgs (2015), UKDA, SN 7856). Notes: match rates are lower-bound estimates, calculated using the following formula (*Size of Linked Sample*) / (*Population of Potential Matches in* 1851/1861/1881).

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 Appendix B.E for more detail), and Section V will show that reweighting does not change the results significantly.

3.4 Methodology

3.4.1 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 ($Y_{i,t-1}$):

$$Y_{i,t} = \alpha + \beta Y_{i,t-1} + \epsilon_{i,t} \tag{3.1}$$

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 toward 0:

$$Y_{i,t-1} = y_{i,t-1} + u_{i,t-1} \tag{3.2}$$

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

$$plim\widehat{\beta_{avg}} = \beta \frac{var(y_{i,t-1})}{var(y_{i,t-1}) + \frac{var(u_{i,t-1})}{T}} \tag{3.3}$$

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 research 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 ($\epsilon_{i,t}$ and $\mu_{i,t}$) observed are uncorrelated across different observations:

$$Y_{i,t} = \beta_0 + \beta_1 \widehat{Y_{i,t-1}} + \epsilon_{i,t} \tag{3.4}$$

$$\widehat{Y_{i,t-1}} = \pi_0 + \pi_1 Z_{i,t-1} + \mu_{i,t} \tag{3.5}$$

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 United States' context, and Dearden et al. (1997); Grawe (2004) for the British context) and more recently, for historical studies by Ward (2023). 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.

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 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 (1992)'s 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 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 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 (Appendix B.I), so life-cycle effects are not significant enough to cast doubts on the results and their interpretations.

3.4.2 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 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:

$$d(P,Q) = \left\{ \sum_{i=1}^{r} \sum_{j=1}^{s} \sum_{l=1}^{r} \sum_{m=1}^{s} \left[ln \left(\frac{p_{ij} p_{lm}}{p_{im} p_{lj}} \right) - ln \left(\frac{q_{im} p_{lj}}{q_{ij} q_{lm}} \right) \right]^2 \right\}^{\frac{1}{2}} \tag{3.6}$$

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 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 errors in Altham statistic, 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.

3.5 Empirical Results

3.5.1 Main Results — IGE Estimates

Table 3.4 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. The OLS estimate of the β for the sample with multiple links (where sons can be linked across multiple censuses) are shown in Columns (2), (5), and (8).[15] 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 3.4: 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

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. *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).

[15] Binned scatter plots for the relationship between father's and son's HISCAM scores are shown in Appendix B.F. They demonstrate that the relationship is clearly linear.

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 (detailed regression output with first-stage results can be found in Table 3.5). 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 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 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 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 3.6 provides some additional results. When a different occupational score index (CCC score) is applied, there is still a significant extent of attenuation in the 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 CCC β obtained using the IV approach is akin

Table 3.5: IV Estimates of IGE, 1851–1911

	1851–1881		1861–1891		1881–1911	
	(1)	(2)	(3)	(4)	(5)	(6)
	1st Stage	2nd Stage	1st Stage	2nd Stage	1st Stage	2nd Stage
Variables	Father 1851	Son 1881	Father 1861	Son 1891	Father 1881	Son 1911
Father 1861	0.586***					
	(0.003)					
Father 1851		0.679***				
		(0.007)				
Father 1851			0.635***			
			(0.003)			
Father 1861				0.648***		
				(0.006)		
Father 1891					0.646***	
					(0.002)	
Father 1881						0.624***
						(0.004)
Constant	1.627***	1.299***	1.447***	1.422***	1.399***	1.527***
	(0.012)	(0.029)	(0.011)	(0.025)	(0.007)	(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	

Notes: standard errors in parentheses; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$; all variables are log HISCAM scores of occupations. *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).

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 β estimated. HISCAM provides two alternative scales constructed using historical records from

Table 3.6: 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
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
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 3.4; '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 et al. (2023); 'Weighted' estimates are produced when the linked sample is reweighted according to population characteristics, following the procedure outlined in Appendix B.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 Appendix B.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).

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 B.D. The 'NYSIIS' sample is produced using the phonetic name version of the ABE matching algorithm, as outlined in Appendix B.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 B.C. As the table highlights, none of these changes affect the results enough to warrant a reconsideration of this paper's findings.

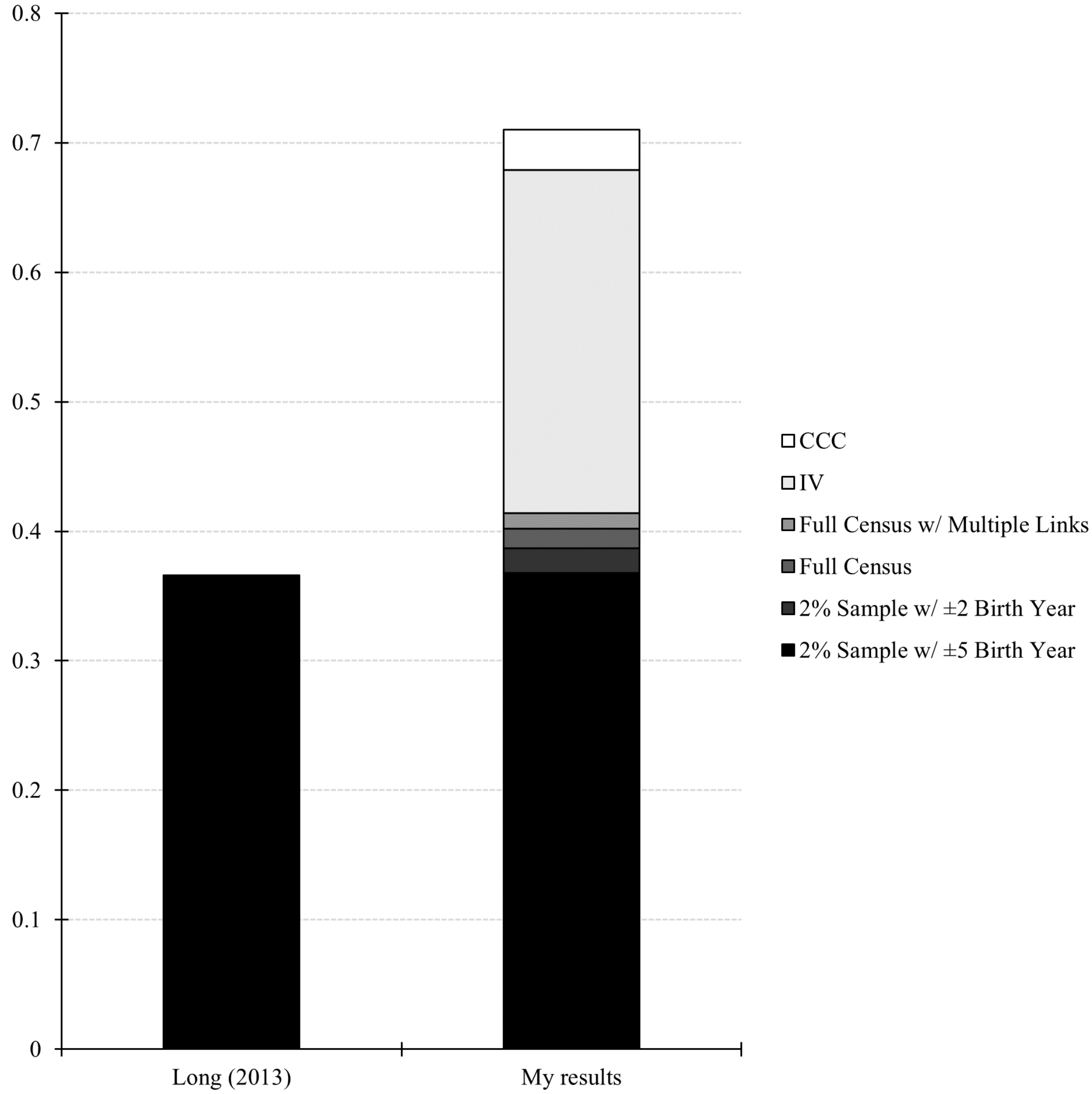
Finally, Figure 3.2 shows the comparison of my results for the period 1851–1881 with Long (2013), and the difference each change in the data and the methodology makes to the estimates of intergenerational mobility. As expected, the IV strategy contributes to most of the difference between my estimates 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 per cent 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.[16]

3.5.2 Altham Statistics

Table 3.7 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 per cent sample of the 1851 census instead (Long, 2013; Pérez, 2019). The mobility tables are not shown in the results but can be found in

[16]See Appendix B.M for more details on false positives caused by the use of a sample.

Figure 3.2: Comparison of Results for 1851–1881 with Long (2013)



Notes: '2% Sample w/ ± 5 Birth Year' refers to using a 2 per cent 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 per cent 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 3.4, Column 1; 'Full Census w/ Multiple Links' is taken from Table 3.4, Column 2; 'IV' estimate is taken from Table 3.4, Column 3; 'CCC' is the estimate obtained using both the IV strategy and the CCC scores instead of HISCAM taken from Table 3.6. *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).

Appendix B.L.

The Altham statistics confirm that the new sample, constructed using the full-count

Table 3.7: Summary of Altham Statistics, 1851–1911

Estimates	*d(P, I)*	*G2*	*d(Q, I)*	*G2*	d(P, Q)	G2
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 which have been corrected for measurement error using Ward (2023)'s approach; all estimates are significant at the 99 per cent 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).

census data, exhibits less mobility than the sample used previously in Long and Ferrie (2013) and Pérez (2019)'s 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.

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.

3.6 Discussion

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 3.8 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 et al. (1997)'s 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.[17] Thus, it appears that intergenerational mobility 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 et al. (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 errors associated with inferring status from occupations, it does not address the measurement errors from assigning scores to occupations. In addition, improvements could also be made

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

Table 3.8: Comparison of IGE Estimates between 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 et al., 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 (2013)'s 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 (2023)'s 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 et al. (1997); Grawe (2004); All estimates for the United States are taken from Ward (2023).

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.[18] 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 United States 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, Long and Ferrie (2013) claimed.

3.7 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 raise the IGE obtained from 0.4 to

[18] One caveat here is that Ward (2023) 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 England and United States.

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 suggest 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.

Chapter 4

Grim Up North? Regional Intergenerational Mobility across England, 1881–1911

4.1 Introduction

Social science has long been interested in the subject of social mobility. Immense efforts have been devoted to document intergenerational mobility — the degree of persistence in socioeconomic outcomes from parents to children — in various countries in the present and the past.[1] Such knowledge is crucial for understanding the equality of opportunities in a society, which has important public policy implications; it is not just useful for emphasising the need for policies to improve the life chances of those underprivileged at birth, but also helps to inform the efficacy of certain policies in different settings. For instance, Piketty (1995) theorised that beliefs and experiences regarding social mobility can influence indi-

[1] There are a number of well-known reviews of the theories, methods, and empirical studies of intergenerational mobility in present-day societies, such as Solon (1999); Black and Devereux (2011); Jäntti and Jenkins (2015); Deutscher and Mazumder (2023). Reviews of historical mobility research are more scarce, but Leeuwen and Maas (2010) provides a survey of some important contributions and developments in sociology and economics/economic history. Other important studies of social mobility in Britain are referenced in this paper, though the list may not be exhaustive.

viduals' preferences for redistributive policies, a relationship which has now been empirically attested by Alesina et al. (2018).

Recently, spatial variations in intergenerational mobility emerged as an area of active research among economists and social scientists. While cross-country differences in the persistence of socioeconomic status have been extensively explored (for example Corak (2006)), there is now a rapidly growing interest in regional differences in intergenerational mobility within a country. Chetty et al. (2014)'s influential work on the geography of economic mobility in the United States demonstrated that the equality of opportunities can vary greatly from one place to another, emphasising the prospect for place-based policies to tackle the intergenerational mobility problem.

England is an ideal location for studying the geography of intergenerational mobility. While the unification of Anglo-Saxon kingdoms into a single polity dates back to over a millennium ago, regional identities remain stronger than ever still. Deep-rooted divisions persist across the country today in all aspects — social to economic, political to cultural. The most prominent geographical divide which continues to afflict England is the 'North-South divide' (Bachtler, 2004). There is a well-known contradistinction in factor endowments between the North and the South, particularly the abundance of coal in the North that played a crucial role in the Industrial Revolution, that led to industrialisation in the North and the Midlands, while the South remained oriented towards agriculture and services (Hudson, 2009; Allen, 2009; Wrigley, 2010*a*). Economic indicators, such as GDP per capita, show that regional disparities, while easing somewhat during the second half of the nineteenth century, was a constant feature of the English economy for a century and half, marked by the dominance of London and the South East (Geary and Stark, 2018).

However, regional inequality is as much a people issue as a place issue. Research on intergenerational mobility in Britain, such as Bell et al. (2023), show that London and the South East exhibit much higher rates of mobility than other regions today. Revealing the long-term evolution of social mobility across different regions is key to both understanding the historical roots of social persistence and conceiving the solution. Addressing the regional gap in the equality of opportunities may constitute a significant step towards alleviating regional inequality — both are issues pertinent to the U.K. government today.[2]

This paper explores historical regional inequalities in the UK from an intergenerational (occupational) mobility perspective, using a set of linked data created from the full-count England and Wales censuses between 1881 and 1911. I show that by the end of the nineteenth century, there is already evidence of a North-South divide in England in terms of social mobility, using rank-based measures of relative and absolute mobility. Those who grew up in the North found it much harder to move up the occupational strata compared to their counterparts growing up in the South. Additionally, I explore cross-sectionally the interplay between inequality and spatial variation in intergenerational mobility. The results show that areas with higher inequality tend to have lower social mobility — affirming the existence of the 'Great Gatsby curve' in Victorian England.

The second part of the empirical analysis focuses on the relationship between migration and intergenerational mobility, and finds that migrants moving away from the North experienced greater mobility than those who remained; this is particularly true for those moving to the South of England. In contrast, Southern migrants were not more occupationally mobile than Southern stayers — in fact, those moving to the North were less (occupationally) mobile than those who stayed. Migrants from the North appeared to have been positively selected according to their fathers' occupational ranks, whereas there seems to have been very little or no selection by the same metric among Southern migrants. However, we can address the selection issue (at least partly) by exploring the differences between brothers who move and stay, a common approach in the migration literature (Abramitzky et al., 2012; Collins and Wanamaker, 2014; Ward, 2019). Comparing the returns to migration calculated using the full linked sample versus those calculated using the brothers sample reveal that at least two-

[2]The government has instituted the 'Social Mobility Commission' to monitor and promote social mobility (Social Mobility Commission, 2023) and the 'Levelling Up' initiative to improve access to opportunities across the country (Department for Levelling Up, Housing & Communities, 2022).

thirds of the gains in occupational ranks from north-south migration can be attributed to selective migration. Nevertheless, the within-brother analysis confirm that migrants moving from the North to the South still experienced sizeable gains from migration in their occupational ranks and in intergenerational mobility compared to their brothers that stayed in the North even after accounting for between-family selection of migrants, but there were no gains from migrating northwards from the South.

The findings of this paper contribute to the burgeoning list of studies seeking to unpack regional variations in intergenerational mobility. The geography of American economic mobility is very well-documented (Chetty et al., 2014; Berger, 2018; Connor and Storper, 2020; Tan, 2023), but others have found that elsewhere in the world, substantial differences in social mobility could arise across regions within the same polity — in Canada (Corak, 2020), Australia (Deutscher and Mazumder, 2020), Italy (Güell et al., 2018; Acciari et al., 2022), Scandinavia (Heidrich, 2017; Eriksen and Munk, 2020; Bütikofer et al., 2022), and of course Britain (Friedman and Macmillan, 2017; Rohenkohl, 2019; Bell et al., 2023; Carneiro et al., 2022; Breen and In, 2024). While historical mobility in Victorian England has been researched (Long, 2013; Long and Ferrie, 2013; Pérez, 2019; Zhu, 2024), this paper is the first to measure intergenerational mobility at the regional (sub-national) level using a large linked dataset of fathers and sons created from the complete-count censuses of England and Wales.[3] The large differences in social mobility uncovered are longstanding and persist to today — possibly a symptom of the North's decline over the twentieth century.

A second contribution of this paper is differentiating the mobility of migrants and stayers, shedding some light into the issues of selective migration and returns to migration in the context of the North-South divide in England. This connects with another vast array of

[3] Long (2013); Long and Ferrie (2013); Pérez (2019) used a 2% sample of the census in 1851 to create their linked dataset. The drawbacks of using a sample for census linking and the potential attenuation bias it may cause have been discussed in Zhu (2024). Boberg-Fazlic and Sharp (2018), who studied North and South differences in pre-industrial mobility using parish reconstitution data, which is very limited in size and covers only select parts of the country. Clark and Cummins (2015)'s study of wealth mobility in England and Wales covers this period as well, using a different methodology (surname-based measures), though their estimates are also only at the national level.

research looking at gains in social (economic) mobility to migration. The literature on the relationship between geographic and social mobility in economic history has extensively surveyed international migration, particularly in the context of the Age of Mass Migration in the United States (Abramitzky and Boustan (2017) provides a good review of the literature). However, there are also a number of studies focusing on domestic migration as a facilitator of social mobility in the past (Long, 2005; Collins and Wanamaker, 2014; Ward, 2019), which sociologists have also explored for modern-day Britain (Friedman and Macmillan, 2017; Buscha et al., 2021; Yu and Athey, 2023; Yu, 2023).

Finally, this paper speaks to the wider interests in understanding regional inequality. Most of the works done on this topic has sought to estimate regional disparities in economic output, productivity, earnings, or unemployment (Hunt, 1986; Blackaby and Murphy, 1995; Gibbons et al., 2010; Geary and Stark, 2018). This paper views the question of regional inequality from a different perspective — social mobility and equality of opportunities. It builds on the aforementioned works on regional social mobility in Britain by looking further into the past. This may inform us whether the higher rates of occupational mobility for London and the South East today (Bell et al., 2023) is a recent development engendered by post-war deindustrialisation, or a longstanding phenomenon that dates back to before England's industrial decline.

The rest of the paper is organised as follows. Section 4.2 reviews the existing literature on geography of intergenerational mobility. Section 4.3 presents the data used and the geographical unit of analysis. Section 4.4 outlines how intergenerational mobility, both relative and absolute, is measured in this paper, as well as definitions of other variables of interest. The geography of intergenerational mobility in late-Victorian England is shown in Section 4.5; they demonstrate, among other things, that there is a North-South divide in mobility. Section 4.6 compares the experience of migrants and stayers in the North versus their counterparts in the South. Section 4.7 discusses the implications of these results. Section 4.8 concludes.

4.2 Geography of Social Mobility

Economic research into social mobility in Britain, until recently, has been preoccupied with estimating national rates of intergenerational mobility (Dearden et al., 1997; Blanden et al., 2004; Grawe, 2004; Gregg et al., 2017). The emergence of new studies mapping the geography of British social mobility has enlightened us that marked differences in intergenerational mobility exist between regions in late-twentieth-century England. Whether in terms of income (Rohenkohl, 2019; Carneiro et al., 2022) or occupational status (Bell et al., 2023), London and the South East performs exceptionally well and appears to be the driving force for social mobility (though Friedman and Macmillan (2017) finds that upward mobility in Inner London was surprisingly low if international migrants are included). On the other hand, Breen and In (2024) find that aside from certain 'hot and cold spots' of high and low mobility, geographic variations in intergenerational mobility seem much more muted in Britain compared to the United States.

While we now have a fairly clear picture of what the geography of social mobility looks like in modern-day Britain, doing this historically is more challenging due to the lack of data. The only exception is Boberg-Fazlic and Sharp (2018), who used parish data to estimate intergenerational mobility before 1850. They found that contrary to the present day, the North was a more mobile place than the South. However, there are limitations with their data. Its geographical coverage only extends to 26 parishes in England, with just under 3,000 father-son pairs in total. The observations for the North come from 3 parishes only, and close to 90 per cent are from a single parish, Gainsborough. This means there are huge representativeness issues entailed. There is also the issue of their inability to track migrants — the latter of which is particularly important given the likely role of geographic mobility in promoting social mobility and the potential interconnection between the two kinds of mobility.

The advent of the census and its digitisation provides a huge statistical underpinning for estimating the rates of intergenerational mobility across England in the late-nineteenth century. Earlier works have already demonstrated the usefulness of digitised census data for the study of Victorian social mobility (Long, 2013; Long and Ferrie, 2013, 2018; Pérez, 2019; Zhu, 2024). Specifically, Zhu (2024) was able to link the full-count censuses of 1851 to 1911 to produce a sample of up to 160,000 father-son pairs to re-estimate the extent of intergenerational occupational mobility. He finds that the intergenerational elasticity (IGE) in occupational status increases from 0.4 to almost 0.7 after accounting for attenuation bias caused by classical measurement error — an issue that has widely been recognised in studies of modern-day social mobility (Altonji and Dunn, 1991; Solon, 1992; Zimmerman, 1992; Dearden et al., 1997; Blanden et al., 2004; Grawe, 2004; Mazumder, 2005; Gregg et al., 2017) but seldom addressed in research on historical mobility — using an instrumental variable (IV) approach proposed by Ward (2023).

The emergence of this large body of literature looking at geography and social mobility enables research into the determinants of social mobility and potential solutions to address unequal access to opportunities in the labour market. The stark contrast in the fortunes of people living under the same set of 'rules' do seem to suggest that, in the words of Chetty et al. (2014), 'intergenerational mobility is a local problem'.

Alongside discovering regions of high and low mobility within a country, these studies have also identified several socioeconomic indicators that are strongly correlated with intergenerational mobility. For instance, the positive relationship between high levels of inequality and high intergenerational elasticity (IGE), dubbed the 'Great Gatsby curve', has been well-established at the cross-country level (Corak, 2013), but there is now evidence to suggest that this may even be true within a single unified state with a federalist or a centralised government. (Chetty et al., 2014; Heidrich, 2017; Güell et al., 2018; Corak, 2020; Acciari et al., 2022). Other correlates, which differ from country to country, also exist. Uncovering the characteristics of socially mobile regions is only possible by observing intergenerational mobility at the sub-national level. Education seems to be a particularly important factor in the case of Britain (Bell et al., 2023; Carneiro et al., 2022).

Another avenue which may bring further insight to the question of spatial differences in intergenerational mobility is by looking at internal migration, which Breen and In (2024) posit was a key factor for people escaping the constraints on social mobility associated with growing up in less favourable places. There is a vast spade of literature looking at different aspects of migration. Particularly pertinent to the issues discussed in this paper is firstly, the role of selective migration. Borjas (1987)'s influential paper modifies the Roy (1951) model and posits a theory of migrant selection in the context of international migration. He argues that positive selection occurs when the sending country has greater income equality than the receiving country, signalling higher returns to human capital at the latter, while negative selection occurs when the reverse is true.

On the other hand, Lee (1966), considering migration more broadly, puts forward a number of hypotheses regarding (among other facets of migration) the characteristics of migrants. Chief among which are that positive selection tends to occur when migrants are responding to pull factors at the destination, while negative selection dominates when migrants are responding to push factors at the origin.

Finally, the literature on the returns to migration in economic history suggests that geographic mobility can be an effective channel for social mobility too.[4] Long (2005) finds substantial gains in socioeconomic status from rural-urban migration in Victorian Britain. Across the Atlantic, internal migration seems to have played a large, positive role in economic mobility that may have outweighed the effects of education (Ward, 2019), and have led to racial convergence in economic status (Collins and Wanamaker, 2014).

Despite the advances made in the study of Victorian mobility using census data, there is yet to be an attempt to explore the regional variations in the father-son association of occupational status. Thus, a vital contribution could be made to our understanding of late-nineteenth-century Britain by unveiling the geography of economic mobility in England as the Victorian Era draws its curtains to a close. Combining this examination with a study of

[4]Studies of returns to migration in the context of international migration are in abundance and will not be discussed here. A useful survey for this literature is Abramitzky and Boustan (2017)

migration further demonstrates the advantages of certain regions in providing its residents with more opportunities, and the relationship between geographic and social mobility.

4.3 Data

4.3.1 I-CeM and Census Linking

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 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.

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 in 1881 and tracks them across a 30-year period to 1911. Two linked samples are then produced. For the triple-linked sample, the sons are linked between all censuses from 1881 to 1911 and their fathers are also be linked from 1881 to 1891. This sample will be used for results at the county level, which include an instrumental variable strategy for reducing attenuation bias by correcting for classical measurement error (outlined in section 4.4.1), and for comparing the occupational mobility of migrants and stayers. For the baseline sample, the sons are matched just once from when they are first observed in 1881 to when they are last observed in 1911. This method is less restrictive and thus produces a larger sample. The triple-linked sample will be used for results at the more granular registration sub-district level (more on this in section 4.3.2) which requires a higher number of observations.

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. The linkage method and the linked sample are both drawn from the author's previous work (Zhu, 2024) where the ABE algorithms are described in greater details.

Table 4.1 shows the linkage results. For the baseline samples, 610,000 father-son pairs have been successfully matched, which translates to a match rate of 29 per cent without accounting for expected attrition of the sample due to mortality and emigration. 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 8 per cent. This still generates 160,000 father-son pairs.

A comparison of the key socioeconomic indicators suggest that both the baseline and the multiple links samples are very representative of the full population. In terms of occupational status —- measured by HISCAM (a stratification scale that assigns scores between 0 and 100 to occupations from low to high status, explained further in section 4.3.3) — 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.

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.

Table 4.1: 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
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
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	13.62	12.48	11.49
Yorkshire	13.25	14.13	14.51
Birth County (Son)			
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

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 counties Middlesex, Kent, Essex, and Surrey. 'London+' refers to Middlesex, Kent, Essex, and Surrey (there is no 'London' in the standardised county of birth code, only 'City of London', which is a tiny part of London with no one recorded as being born there). 'Yorkshire' includes all Ridings of Yorkshire. All numbers are in percentages unless they are means. *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).

4.3.2 Geographical Divisions in England

Any exploration of the spatial variations in intergenerational mobility necessitates the division of the data along geographical lines. The most natural way to achieve this is to divide the sample by the boundaries of English counties. There were 43 counties in England in 1881, thus we can divide our data into 43 sub-samples based the county in which the fathers and sons lived in 1881, when the latter were still in their childhood. The data also provides additional levels of disaggregation (from large to small) — registration district, registration sub-district, and parish. Since counties are large and the areas within are not homogeneous, this paper also uses the baseline sample to map the intergenerational mobility of late-Victorian England at the registration sub-district (RSD) level (approximately 2,000 RSDs in England), in order to further identify regions of high and low social mobility and uncover any potentially interesting patterns that are hidden within a county. Table C.A.1 in Appendix C.A shows the number of RSDs per county.

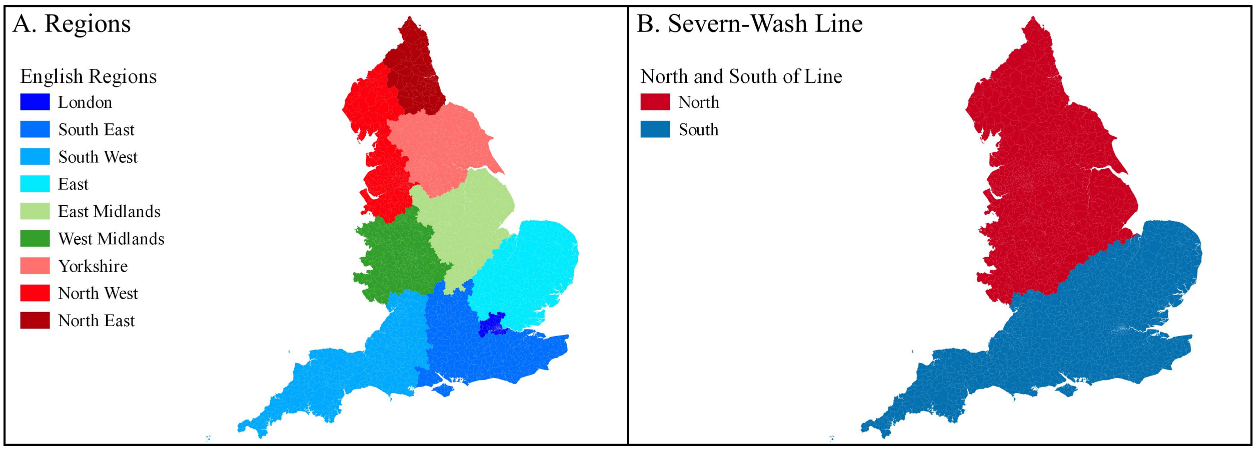
Using their place of residence in 1881 for region of origin mirrors the approach of Chetty et al. (2014), who assigned their sample of children to the location of parents when the child was claimed as a dependent, regardless of where they lived as an adult. It also matches the work of Bell et al. (2023) closely, who used the children's place of birth instead. The choice of assigning regions of origin based on where people lived in their childhood rather than where they settled as adults is further motivated by recent research which find significant childhood exposure effects to growing up in better neighbourhoods (Chetty and Hendren, 2018*a*,*b*).

One final clarification to make is the definition of 'North' and 'South'. While the decision to call oneself a 'Northerner' or 'Southerner' can be influenced by culture and identity as much as geography, there are a number of commonly accepted ways with which people historically and presently have used to divide England into North and South.[5] The most basic division

[5]Scholars, such as Bambra et al. (2023), have also used these methods to separate England into North and South when studying regional inequalities.

separates England into two halves along the 'Severn-Wash Line' — an imaginary line drawn across England, and splitting it into two halves, from the Wash in East Anglia to the Severn Estuary in Bristol. Panel B of Figure 4.1 shows the division of England by the Severn-Wash Line.

Figure 4.1: Regions and Severn-Wash Line



Sources: these maps are created using the 1881 Registration Sub-District GIS shapefile from Day (2016).; Panel A is created by assigning regions to counties based on NUTS level 1 from European Commission (2022); Panel B is created by assigning counties that have the majority of their area north of the Severn-Wash Line illustrated in Bambra et al. (2023) 'North' (and the rest 'South'), and all areas within the same county are given the same label even though some parts of it may lie just south of the line.

Another way is to divide England into nine conventionally recognised regions often used for administrative and statistical purposes by the UK government, as well the European Union when constructing their Nomenclature of Territorial Units for Statistics (NUTS) (European Commission, 2022). These nine regions are London, South East, South West, East of England, East Midlands, West Midlands, Yorkshire and the Humber, North East, and North West. This paper also adopts the same geographical division using historical county boundaries, where the North can be defined as all counties in the North East, the North West, and Yorkshire and the Humber; the list of constituent counties in each region can be found in Table C.A.1 in Appendix C.A. Panel A of Figure 4.1 outlines the different regions of England according to this approach. The Midlands are coloured light and dark green, and all regions in different shades of blue below the Midlands are the 'South', and all regions in different shades of red above the Midlands are the 'North'.

4.3.3 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 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 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. Methodologically, the creation of HISCAM is similar to CAMSIS (Cambridge Social Interaction and Stratification Scale) used by Bell et al. (2023) to score occupations in late-twentieth century Britain.

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. Table 4.2 shows a sample of some common occupations observed in the census with their respective HISCAM scores.

Table 4.2: 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 (UKDA, SN 7481); 'HISCAM' is taken from Lambert et al. (2013).

4.4 Methodology

4.4.1 Measuring Relative Mobility

The commonly asked question in the literature on social mobility — and one that relates to the issue of relative mobility — is 'to what extent are children's outcomes predicted by their parents' achievements?' (Black and Devereux, 2011). The standard approach to answering this question follows the Becker-Tomes model (Becker and Tomes, 1986) by calculating the intergenerational elasticity (IGE) of any measure of socioeconomic status (in this case occupational status) by regressing the outcome Y of the son of family i in generation t on the outcome the father ($Y_{i,t-1}$):

$$Y_{i,t} = \alpha + \beta Y_{i,t-1} + \epsilon_{i,t} \tag{4.1}$$

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.

An alternative and commonly used variant of the IGE in the literature on economic mobility (and particularly the geography of economic mobility), pioneered by Dahl and Deleire (2008) and popularised by Chetty et al. (2014), is to assort the fathers and sons' socioeconomic outcomes (in the case of this paper, the HISCAM score of their occupations) into their respective cohorts' percentile ranks (at the national level rather than within their region, consistent with the literature), then estimate the association between fathers' ($R_{i,t-1}$) and sons' ranks ($R_{i,t}$) — known as the rank-rank slope:

$$R_{i,t} = \alpha + \beta R_{i,t-1} + \epsilon_{i,t} \tag{4.2}$$

The rank-rank slope is similar to the IGE in many respects, but is less influenced by changes in inequality between generations (Deutscher and Mazumder, 2023). Proponents of the rank-rank slope finds it to be a more robust measure of intergenerational mobility across specifications and sample choices (Dahl and Deleire, 2008; Chetty et al., 2014), and more resistant to life-cycle bias (Nybom and Stuhler, 2017). In line with existing studies in the literature, such as Chetty et al. (2014) and Bell et al. (2023), this paper will use the rank-rank slope as the preferred measure of relative mobility, but with one important difference — correcting for measurement error.

The socioeconomic outcome of an individual observed in a given year consists of a permanent component and an uncorrelated transitory component. As such, measures of status based on a single snapshot of a person's life may be noisy, so the occupational status of the father observed in a single census year may deviate from his permanent status. Elasticity measures (such as IGE) therefore suffer from classical measurement error in the dependent variable, which attenuates β towards 0 (Solon, 1992; Mazumder, 2005; Nybom and Stuhler, 2017):

$$Y_{i,t-1} = y_{i,t-1} + u_{i,t-1} \tag{4.3}$$

In studies of modern-day earnings mobility, one solution would be to average across multiple observations of the father's earnings (Mazumder, 2005). However, this strategy is less effective for looking at historical mobility because people are observed in sources of occupational status, such as the census, much less frequently (once per decade for censuses). Thus, it might only be possible to take an average of at most occupations from three censuses (Mazumder (2005) had information on earnings from between 7 to 15 years). Moreover, using an occupation observed at a young age may also introduce life-cycle bias (Nybom and Stuhler, 2017).

To address this issue, one can 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:

$$Y_{i,t} = \beta_0 + \beta_1 \widehat{Y_{i,t-1}} + \epsilon_{i,t} \tag{4.4}$$

$$\widehat{Y_{i,t-1}} = \pi_0 + \pi_1 Z_{i,t-1} + \mu_{i,t} \tag{4.5}$$

This approach has featured in a number of studies on modern-day intergenerational earnings elasticity (Altonji and Dunn, 1991; Solon, 1992; Zimmerman, 1992; Dearden et al., 1997; Grawe, 2004), often using education as an instrument for income. Zhu (2024) estimated the IGE for Victorian England after accounting for classical measurement error by instrumenting father's log occupational score with the log score of the occupation of the same father observed in the next census, a strategy that was first proposed by Ward (2023) in the context of nineteenth-century United States.[6] This approach has been successful in reducing attenuation bias in IGE measured from occupations observed in historical censuses in the case of England and the United States.

This paper will adopt the same approach to correcting for measurement error, with one caveat. Though attenuation bias may be weaker in rank slopes than IGE, transforming measures of status into percentile ranks introduce non-classical measurement error to both the dependent and independent variables, and standard methods for reducing attenuation bias may not apply (Nybom and Stuhler, 2017).[7] Nybom and Stuhler (2017) proposes a method for correcting this bias using a generalised errors-in-variables model that works in a similar way to the IV strategy; both methods produced similar estimates in Ward (2023).[8]

Therefore, this paper will estimate the rank-rank slopes for measuring relative mobility for each county using the triple-linked sample, with the IV estimates being the preferred estimates. The baseline sample will also be used for estimating mobility at the RSD-level (though no correction for measurement error is possible here), which will help unpack the spatial patterns in intergenerational mobility at a more granular level and will enable us to explore the relationship between intergenerational mobility and inequality. Appendix C.B will discuss Nybom and Stuhler (2017)'s method for reducing attenuation bias in rank-rank slopes in greater details and provide the estimates obtained using their method.

4.4.2 Measuring Absolute Mobility

Another measure of interest in estimating intergenerational mobility is the degree of absolute mobility, which is reflected by the expected outcomes of children conditional on the socioeconomic status of the parents. This paper adopts the measure of 'absolute upward mobility' put forward by Chetty et al. (2014), adapted to the context of occupational (rather than

[6]For more details on the IV approach, see Zhu (2024).

[7]Nybom and Stuhler (2017) argue that errors in rank are negatively correlated to true ranks, as top (bottom) ranks cannot be overstated (understated), so there is non-classical measurement error.

[8]Note also that rank-rank slopes after correcting for measurement error via either strategy are much higher than before, thus rank slopes in itself are not sufficient for reducing attenuation bias.

income) mobility — the mean rank (in the national occupational distribution) of sons whose fathers were situated at the 25^{th} percentile of the national occupational distribution of their generation (a measure also used by Bell et al. (2023) to estimate upward occupational mobility in post-war England and Wales).[9] This can be computed simply using the parameters estimated from equation 4.2, and substituting $R_{i,t-1}$ for 25:

$$\widehat{R}^{25} = \widehat{\alpha} + \widehat{\beta} \cdot 25 \tag{4.6}$$

Large differences in $\widehat{\alpha}$ and $\widehat{\beta}$ across regions will therefore result in substantial variations in chances of upward mobility for those at the lower end of the socioeconomic stratum.

4.4.3 Definitions of Other Variables

Measures of Inequality. While the literature has opted for income-based measures of inequality, such as the Gini coefficient, when establishing the existence of the 'Great Gatsby curve' (Corak, 2013), this paper has to rely on other measures of inequality instead, since income or wealth data covering the entire population is not available for this period in Britain. As such, this paper will observe the relationship between occupational inequality and intergenerational mobility at the registration sub-district level, with two different measures of occupational inequality. The first measure is the standard deviation of occupational ranks of all the fathers in the (baseline) linked sample living in each RSD in 1881. Higher standard deviations indicate higher inequality. The second measure is the share of fathers within each RSD who have ranks between the 25^{th} and 75^{th} percentile of the national distribution (within the linked sample), an alternative measure to Gini used by Chetty et al. (2014). A higher share of the this middling group implies a less polarised socioeconomic structure and thus lower inequality.[10]

[9]Since this paper uses a linked sample rather than the full population census, 'national' in this case means the full linked sample aggregated (i.e. not divided into sub-national samples) rather than the population.

[10]RSD inequality is calculated using the baseline linked sample but could also be calculated using the full census. Appendix C.E discusses this in greater detail; the results hold when using population-level

4.4.4 Returns to Migration

As outlined in Abramitzky et al. (2012), a 'naive' interpretation of the returns to migration can be obtained by estimating β in the following approach:

$$Y_i = \alpha + \beta Migrant_i + \gamma X_i + \epsilon_i \tag{4.7}$$

Where $Y_{i,t}$ is the outcome of interest for individual i, $Migrant_i$ is a dummy variable of being a migrant, X_i are observable characteristics to be controlled, and ϵ_i is the individual error term representing unobserved differences between individuals that may influence one's decision to migrate.

A common strategy in migration studies is to identify siblings (in this case brothers) with different migrant status, and compare migrant brothers with non-migrant brothers. Running an OLS regression of equation 4.7 with household fixed effects will thus control for aspects of family background that affect both the tendency to migrate and labour market outcomes:

$$Y_{i,j} = \alpha + \beta Migrant_{i,j} + \gamma X_{i,j} + \eta_j + \nu_{i,j} \tag{4.8}$$

Here, the individual error term ϵ_i is decomposed into η_j, which is the component that is shared between the brothers and addressed by the household fixed effects, and $\nu_{i,j}$ is the component of the error term that is unique to each individual.

Since we are interested in not only the returns to migration for people moving from the North to the South (and vice versa), but also the intergenerational aspect — i.e. how much migration across the north-south border influences the persistence of status from fathers to sons — we can adapt equation 4.8 into the following:

$$Rank_{i,j,t} = \alpha + \beta_1 Migrant_{i,j,t} + \beta_2 Migrant_{i,j,t} \times Rank_{j,t-1} + \gamma X_{i,j,t} + \eta_{j,t} + \nu_{i,j,t} \tag{4.9}$$

inequality.

Where $Rank_{i,j,t}$ is the percentile occupational rank of son i from family j in generation t and $Rank_{j,t-1}$ is the father's rank. The coefficient on the interaction term $Migrant_{i,j,t} \times Rank_{j,t-1}$ measures the difference in the father-son rank-rank slopes between the stayers and the migrants.

4.5 The North South Divide in Opportunities

4.5.1 National Level

Before we explore the spatial variations in intergenerational mobility across late-Victorian England, this section will first present the rank-rank slopes in father and son's occupational status at the national level. Only the OLS and IV results using the baseline and triple-linked sample will be shown here, since Zhu (2024) already discussed many recent issues most pertinent to the social mobility literature, such as measurement error and attenuation bias, census linkage methods (false positives), and re-weighting of the linked sample (to match population characteristics).

Table 4.3: National Relative and Absolute Mobility in England, 1881–1911

Linked Sample	(1) $\widehat{\beta}$ (OLS)	(2) $\widehat{\beta}$ (IV)	(3) $\widehat{R}^{25}$ (OLS)	(4) $\widehat{R}^{25}$ (IV)	(5) N
Baseline	0.364		40.77		597,517
	(0.001)		(0.071)		
Triple-Linked	0.367	0.537	40.895	36.905	159,723
	(0.002)	(0.003)	(0.136)	(0.183)	

Notes: robust standard errors in parenthesis; all coefficients significant at the 1% level; rank-rank slopes, measuring relative mobility, shown in columns 1 and 2; expected rank of sons born to fathers belonging to the 25^{th} percentile rank, measuring absolute mobility, shown in columns 3 and 4; the standard errors in columns 3 and 4 are standard errors of the constant estimated in the rank-rank regression; number of observations indicate the number of father-son pairs, shown in column 5.
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).

The national level results are illustrated in Table 4.3. For the whole late-Victorian England, the father-son correlation in occupational ranks stands at 0.54 (column 2) after accounting for the attenuation bias caused by measurement error, rising from around 0.36 (column 1) prior to applying the IV strategy. Meanwhile, the absolute upward mobility that a son born to a father who was positioned at the 25$^{\text{th}}$ percentile rank in 1881 can expect to achieve by 1911 is just below rank 41 (column 3) according to the OLS estimates and almost rank 37 (column 4) if we use the IV estimates instead. These results serve as a useful benchmark for interpreting the sub-national results in section 4.5.2 and 4.5.3.

4.5.2 County Level

To map the geography of intergenerational mobility in Victorian England, the triple-linked sample is first divided into 43 sub-samples representing each county, then the rank-rank slope for son i growing up in county c is estimated for each sub-sample separately[11] using equation 4.2 in section 4.4.1 such that:

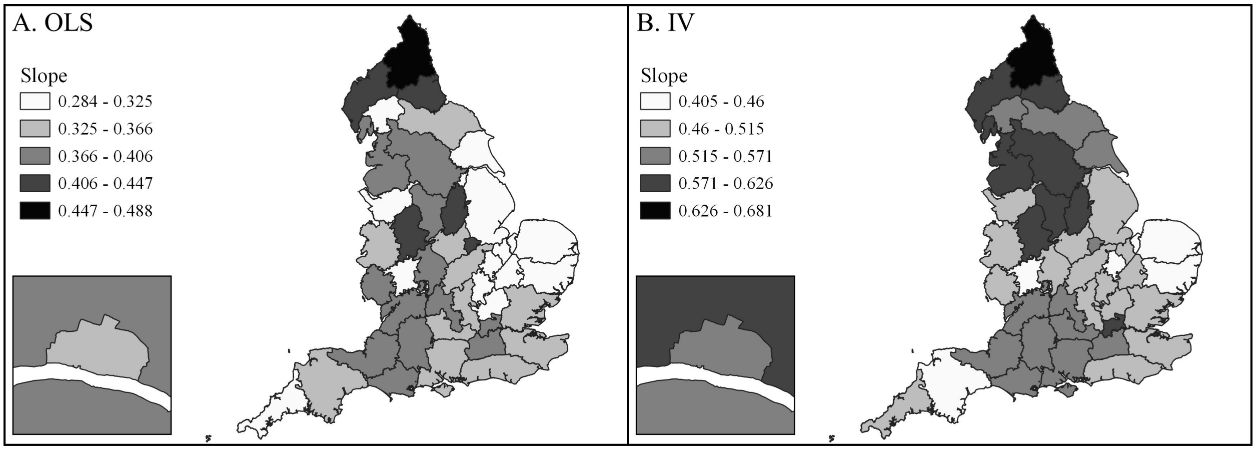
$$SonRank_{i,c} = \alpha_c + \beta_c FatherRank_{i,c} + \epsilon_{i,c} \tag{4.10}$$

Figure 4.2 shows the county-level estimates of relative mobility in rank-rank slope for 1881–1911. The counties are divided into five groups of equal intervals by their mobility estimates. Darker areas represent lower mobility. The map represents the 1851 borders of ancient English counties, obtained from Satchell et al. (2023); London is included but is surrounded by Middlesex and shown separately in the bottom left corner.

Panel A and B shows the rank-rank slopes estimated using OLS and IV. The OLS estimates do not seem to show a clear geographical pattern in intergenerational mobility across England, except that the counties closest to the Scottish border are very immobile. However, once we correct for measurement error using the IV strategy, the North-South divide becomes

[11] Running the regression individually for each sub-sample is the standard approach in the literature, but we can also estimate it without dividing the sample regionally via a fixed effects regression with an interaction term between father's rank and the county of origin: $SonRank_i = \alpha + \beta FatherRank_i + \gamma FatherRank_i \times County_i + County_i + \epsilon_i$. This will produce the same estimates as the preferred methodology. The results using this method are reported in appendix C.C.

Figure 4.2: County-Level Relative Mobility (OLS and IV), 1881–1911

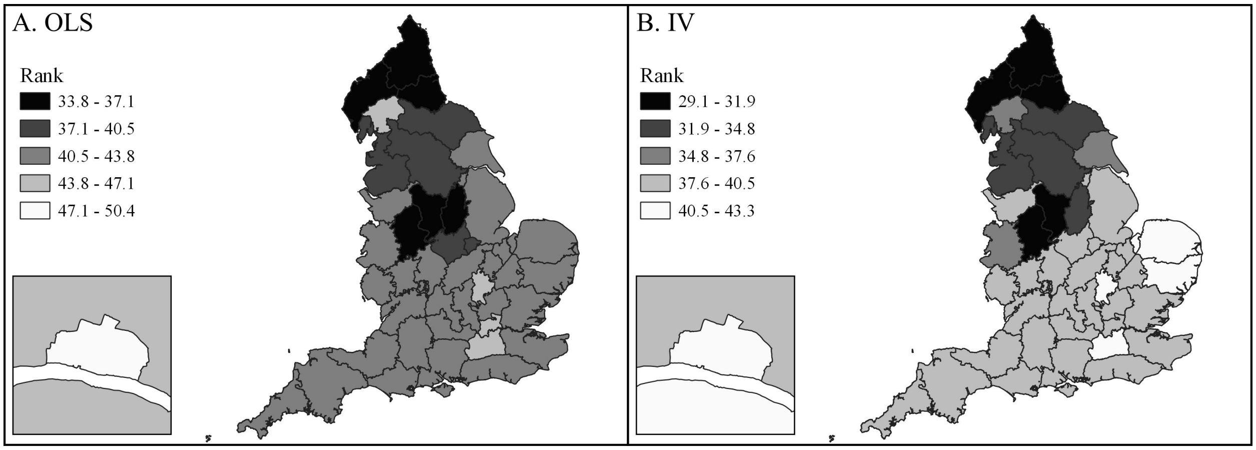


Notes: all coefficients are statistically significant at the 1% level.
Sources: GIS shapefile marks out 1851 borders of ancient counties of England, obtained from Satchell et al. (2023); results produced from author's analysis of I-CeM (Schürer and Higgs (2014), UKDA, SN 7481) and I-CeM Names and Addresses (Schürer and Higgs (2015), UKDA, SN 7856).

clearer. Many counties north of the Severn-Wash Line exhibit very substantial persistence in occupational status from fathers to sons. In the IV results in panel D, Northumberland has the highest rank-rank slope of 0.68, while three counties in the Midlands, the West Riding of Yorkshire, Lancashire, Durham, and Cumberland all have rank slopes of higher than 0.57; the only other county that was similarly immobile in the South was Middlesex. Mobility in these counties are all lower than the national average.

Another interesting feature of the findings on relative mobility is that London and the South East do not exhibit exceedingly high degrees of intergenerational mobility. For example, London has a rank-rank slope of 0.34 (OLS) and 0.52 (IV), while Middlesex has a rank-rank slope of 0.37 (OLS) and 0.60 (IV). The South East counties, while far from being the most immobile, are also not found in the most mobile group. Thus, most of the people who grew up in later-Victorian London and the South East were certainly more mobile than their counterparts from the North (and more mobile than the average person in England in their cohort), but they — including the Londoners — still show a moderate degree of father-son persistence in occupational status. In contrast, East Anglia consistently demonstrates

Figure 4.3: County-Level Absolute Mobility (OLS and IV), 1881–1911



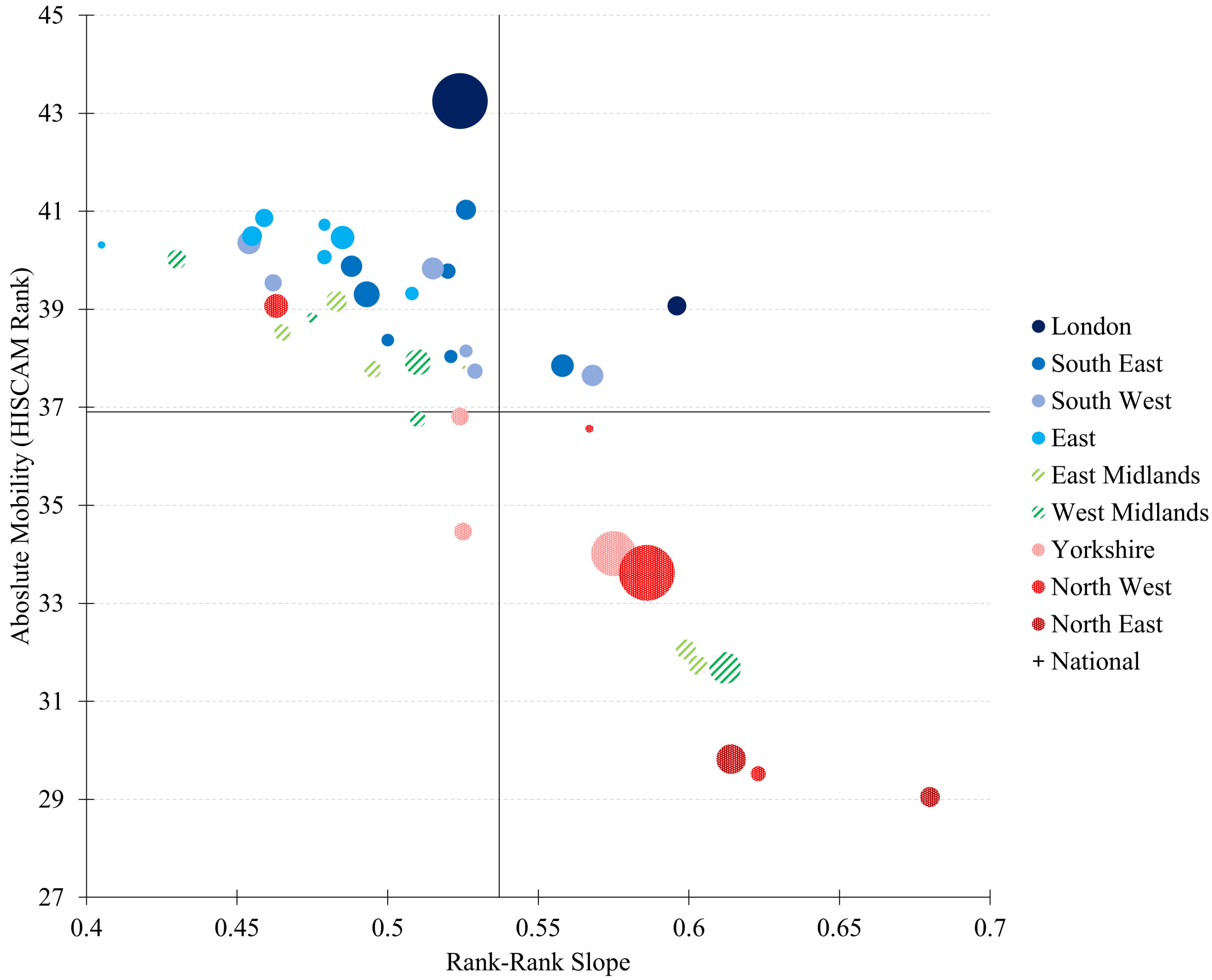
Notes: computed by multiplying the coefficients obtained in Figure 4.2 Panel C and D by 25, as outlined in Section 4.4.2.
Sources: GIS shapefile marks out 1851 borders of ancient counties of England, obtained from Satchell et al. (2023); results produced from author's analysis of I-CeM (Schürer and Higgs (2014), UKDA, SN 7481) and I-CeM Names and Addresses (Schürer and Higgs (2015), UKDA, SN 7856).

high relative mobility across both sets of estimates.

Figure 4.3 demonstrates the absolute mobility of English counties, measured by the expected average rank of sons born to fathers in the 25^{th} percentile of occupational status. Unlike relative mobility, there is a clear North-South divide in absolute mobility in England. Counties in the North are decidedly less mobile than those in the South, particularly in the IV results in Panel B. The three northern counties of Cumberland, Northumberland, and Durham have a $\widehat{R}^{25}$ of around 35 in OLS results and around 30 in IV results. These are all below the national average in absolute mobility shown in table 4.5.1 in the previous section. A clear difference can be observed between most of the counties on the opposite ends of the Severn-Wash Line. Moreover, London shows up as one of the most optimal locations for absolute upward mobility when measured in OLS and IV rank-rank slopes, with a $\widehat{R}^{25}$ of 50 and 43 in panel A and B respectively — much higher than the national average too.

Figure 4.4 demonstrates the clear divide in mobility patterns between the North and the South. The national averages, estimated in section 4.5.1, are represented by the horizontal and vertical black lines. Counties belonging to the North represented by red dotted bubbles

Figure 4.4: County Absolute Mobility by Relative Mobility (IV), 1881–1911



Notes: size of circle represents the number of male children aged 5-15 residing in each county in 1881 in the full-count census; full list of counties belonging to each regions in table C.A.1 in appendix C.A.
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).

— the North West, the North East, and Yorkshire — are mostly located on the bottom right quadrant (aside from the North and East Ridings of Yorkshire in the bottom left and Cheshire in the top left), which suggests that they have low relative and absolute mobility. On the other hand, counties in the South, represented by blue solid circles, are mostly located towards the top left corner, where relative and absolute mobility is high. The mobility patterns of counties in the Midlands (green diagonal-patterned circles) are spread out across the spectrum. The size of the circles represent the size of total male population aged 5 to 15 in each county in the 1881 census.[12]

The county-level findings, therefore, point to the existence of a North-South divide in mobility regimes across England, with the North being a region of social persistence and the South a nucleus of mobility. These findings are consistent with Bell et al. (2023), who found the highest absolute upward occupational mobility in London (and the South East) for the post-war period. It confirms that the pre-eminence of London as the place likeliest to engender a 'rags to riches' story has been a constant feature of English intergenerational mobility since at least the end of the nineteenth century.

4.5.3 Registration Sub-District Level

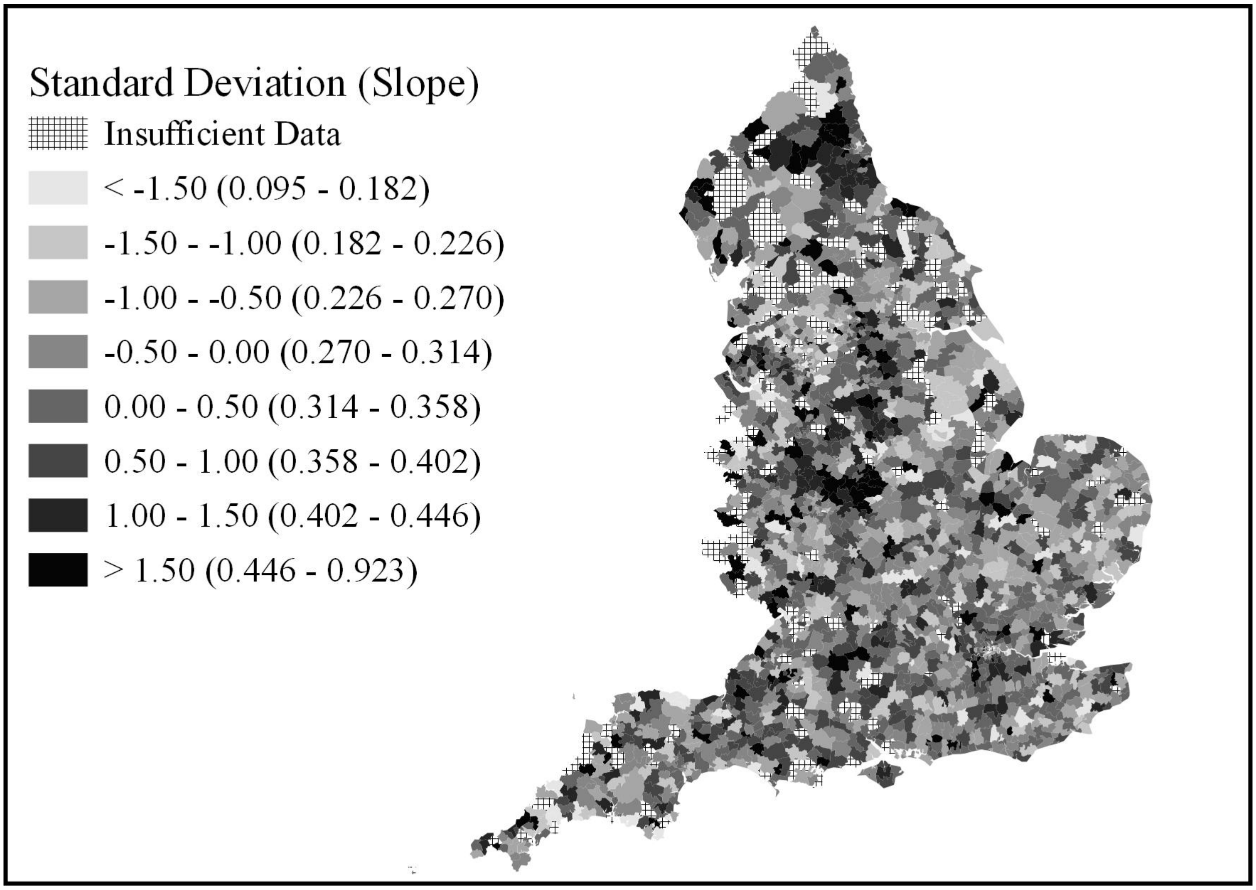
To delve deeper into the intricacies of the sub-national pattern of occupational mobility, the sample is further divided into registration sub-districts. Figure 4.5 shows the mobility pattern of relative mobility in England when divided into RSDs. The RSDs are split into multiple groups depending on how many standard deviations their estimates deviate from the mean (both standard deviations and the mean are calculated by pooling together all the estimates); the numbers in brackets show the exact estimates of the rank-rank slope. Less mobile regions are shown in darker colours. RSDs with insufficient data to produce statistically significant results at the 10 per cent level are shaded in cross-shaped patterns.

The pattern of rank-rank slopes shows a divide between regions north and south of the Severn-Wash line. The least mobile RSDs are concentrated in clusters the Midlands and the North East, in RSDs with a high share of the sample. In contrast, the low mobility RSDs in the South are dispersed across the region, and not in densely populated areas.

Figure 4.6 shows the pattern in absolute mobility across English RSDs. Here, the North-South divide in mobility regimes is even more evident. The least upwardly mobile RSDs are

[12]Total population in the census is used rather than the population in the linked sample to address the issue that linking people from more densely populated places are more difficult due to the restrictions on the uniqueness of name, age, and parish of birth.

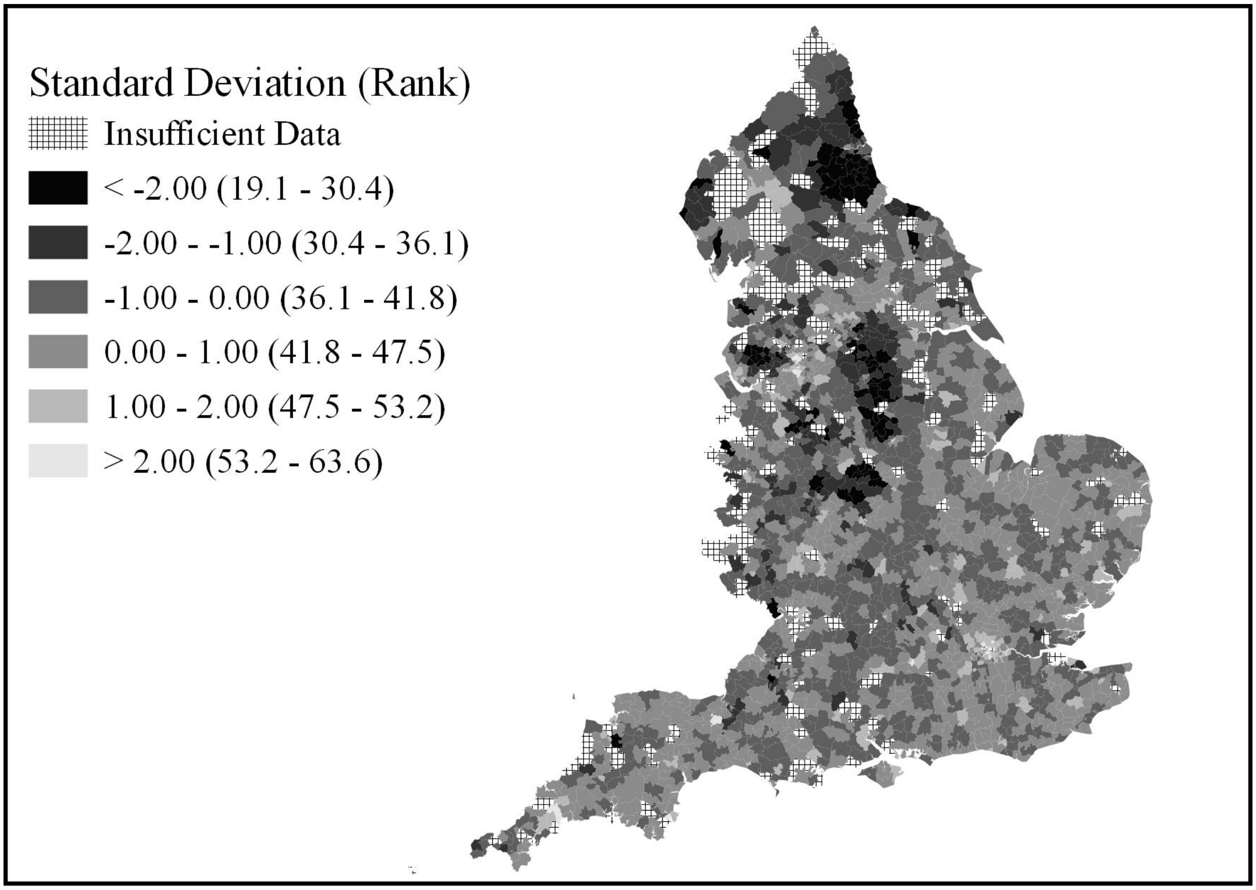
Figure 4.5: RSD-Level Relative Mobility, 1881–1911



Notes: all coefficients are statistically significant at at least the 10% level; RSDs are split into multiple groups depending on how many standard deviations their estimates deviate from the mean (standard deviations and mean calculated by pooling together all the estimates); the numbers in brackets show the exact estimates of the rank-rank slope.
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). The GIS Shapefiles for 1881 Registration Sub-District comes from Day (2016).

all concentrated in the Northern half of England. Particularly, there are several clusters of low mobility in Staffordshire and Derbyshire in the Midlands, the urban areas of Liverpool and Manchester in Lancashire, Cumberland in the North West, and Durham in the North East. In contrast, there are very few low mobility RSDs in the Southern half of England. The evidence shown here echoes the findings of Breen and In (2024) in that this paper also finds areas of remarkable social persistence next to regions that are broadly similar to each other (and to the rest of the country). [13]

Figure 4.6: RSD-Level Absolute Mobility, 1881–1911



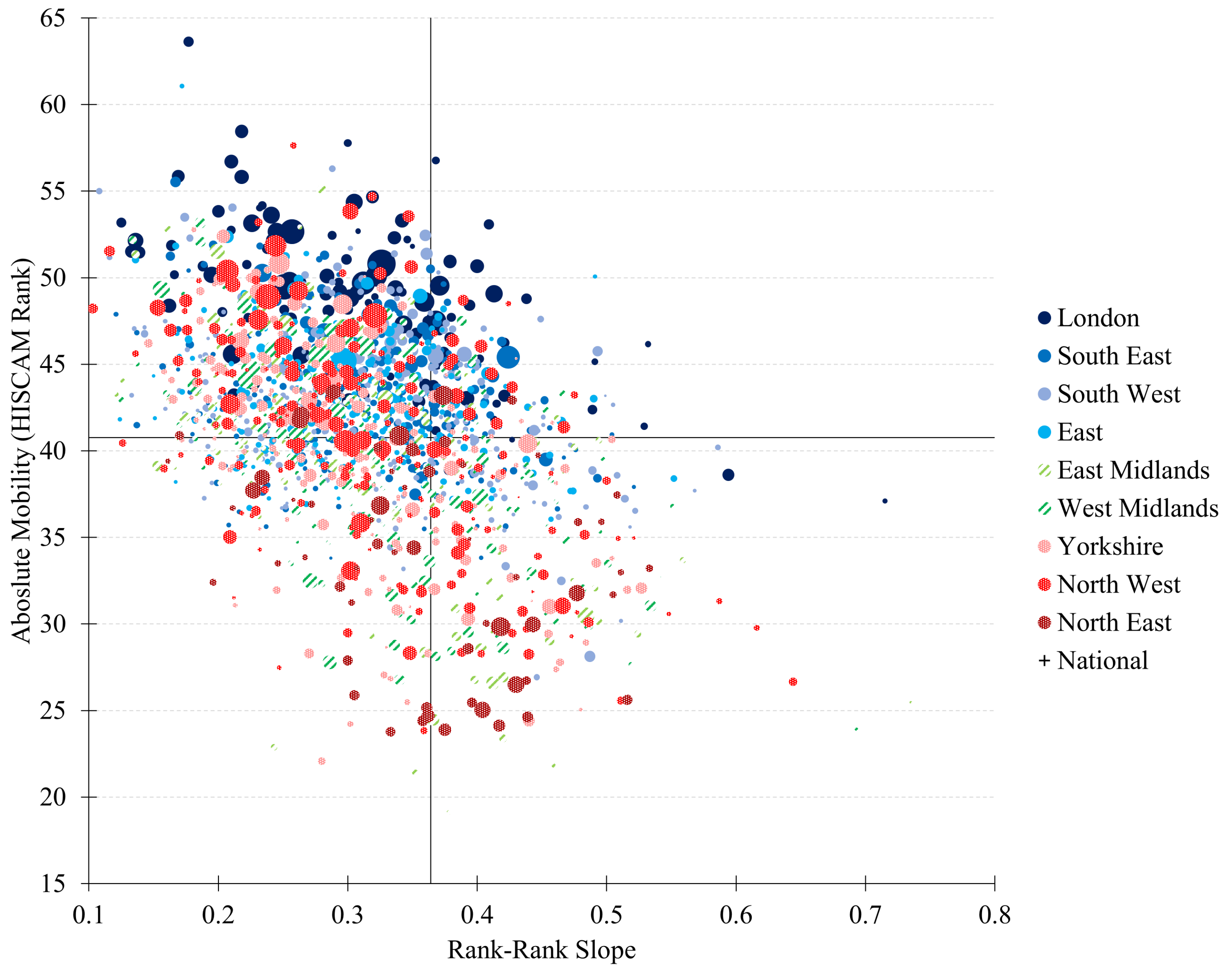
Notes: all coefficients are statistically significant at at least the 10% level; RSDs are split into multiple groups depending on how many standard deviations their estimates deviate from the mean (standard deviations and mean calculated by pooling together all the estimates); the numbers in brackets show the exact estimates of the expected rank.
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). The GIS Shapefiles for 1881 Registration Sub-District comes from Day (2016).

The mobility pattern at the RSD-level, depicted in figure 4.7, shows the north-south divide in a more nuanced fashion. The sub-districts in the South are concentrated in the top left quadrant and near the national average, indicating that they are often more mobile in relative and absolute terms than the average or at least equally so. On the other hand, the

[13]While this resembles the distribution of mining activity within England, the results are not driven by miners alone. Map of mining density in 1911 shown in figure C.G.2 and mobility regimes without miners shown in figure C.G.3 in appendix C.G.

Midlands and the North have RSDs spread out across the spectrum, so while there are indeed areas in the North that are equally as mobile as areas in the South, across the whole the region becomes less mobile than the South on average. The difference is greater in absolute mobility, as there are very few sub-districts in the South that are below the national average line but plenty of sub-districts in the Midlands and the North. This reflects the earlier figures of the map of relative and absolute mobility at the RSD-level, where the distinction between the North and the South is clearer in terms of absolute mobility.

Figure 4.7: RSD Absolute Mobility by Relative Mobility (OLS), 1881–1911



Notes: size of circle represents the number of male children aged 5-15 residing in each county in 1881 in the full-count census; full list of counties belonging to each regions in table C.A.1 in appendix C.A.
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).

Finally, the RSD results enable us to test the relationship between inequality and social mobility — in other words, was there a 'Great Gatsby curve' in late-Victorian England? Table 4.4 provides evidence that within England, there was also a strong relationship between higher inequality and lower social mobility. A one per cent increase in the standard deviation of the father's rank is associated with a 0.4 per cent increase in the rank-rank slope (less relative mobility) and a 0.3 per cent decrease in the expected rank of sons with fathers in the 25$^{\text{th}}$ percentile (less absolute mobility). Using an alternative measure — the fraction of fathers in rank 25 to 75 — yields the same relationship, as RSDs with higher shares of fathers in this middling group (lower inequality) have higher relative and absolute mobility, although the results are statistically insignificant for relative mobility.

Table 4.4: Higher Inequality Correlated with Lower Relative and Absolute Mobility

(1) Measure of Inequality (Log)	(2) Measure of Mobility (Log)	(3) β	(4) N
Std. Dev. of Father's Rank	Rank-Rank Slope	0.43*** (0.119)	1792
Percent of Fathers in Rank 25-75	Rank-Rank Slope	-0.09 (0.063)	1792
Std. Dev. of Father's Rank	Absolute Mobility (Rank)	-0.28*** (0.102)	1792
Percent of Fathers in Rank 25-75	Absolute Mobility (Rank)	0.15*** (0.035)	1792

Notes: *** p<0.01, ** p<0.05, * p<0.1; log-transformed independent variables in column 1 and log-transformed dependent variables in column 2; standard errors clustered at the county-level, shown in parenthesis in column 3; number of observations in column 4 represent the number of RSDs used in this regression.
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).

4.6 Migration and the North-South Divide

4.6.1 Baseline Results: Comparing Stayers and Migrants

A crude way of comparing the intergenerational mobility of migrants and stayers is to follow the same method used in the previous section — sub-divide the sample into stayers and migrants and then calculate the rank-rank slope for each sub-sample separately.

Table 4.5 demonstrates the differences in relative mobility between stayers and migrants. The sample is separated into two groups — one with 'Northern Origins' and one with 'Southern Origins'; these are defined by their places of residence in 1881. 'North Stayers' are defined as people who resided in the 'North' (the three Northern regions of North East, North West, and Yorkshire and the Humber) in 1881 and 1911. 'North Movers' are defined as people who resided in the 'North' in 1881 but not in 1911. 'North to South' defined as people who resided in the 'North' in 1881 and in the 'South' (the four regions of East of England, London, South East, and South West) in 1911. 'North of Severn-Wash Stayers' are defined as people who lived in counties north of the Severn-Wash Line in 1881 and 1911 while 'North of Severn-Wash Movers' are people who lived in counties north of the line in 1881 and south of the line in 1911. Same rules apply for those with 'Southern Origins'.

For those with 'Northern Origins', there is a clear difference in the relative mobility of those who stayed in the North versus those who moved away from the North, whether calculated using the OLS or the IV method. The 'North Stayers', which refers to people who resided in the Northern counties in the North East, the North West, and Yorkshire in both 1881 and 1911, consistently show the strongest association in father-son status compared to those who moved away ('North Movers'), with rank-rank slopes of 0.38 (OLS) and 0.58 (IV). Among those who move, the most occupationally mobile are those who move to the Southern counties ('North to South') in the regions of London, the South East, the South West, and the East. They have rank-rank slopes of 0.26 (OLS) and 0.36 (IV). The difference remains if we use a less precise definition of North and South — the Severn-Wash Line. Once again,

Table 4.5: Relative Mobility of Stayers and Migrants, 1881–1911

	OLS			IV		
	β	SE	N	β	SE	N
Northern Origins						
North Stayers	0.380	0.004	50,095	0.577	0.007	49,484
North Movers	0.352	0.018	2,772	0.479	0.026	2,727
North to South	0.263	0.025	1,344	0.362	0.036	1,316
North of Severn-Wash Stayers	0.376	0.003	82,677	0.562	0.005	81,726
North of Severn-Wash Movers	0.283	0.015	3,296	0.390	0.020	3,250
Southern Origins						
South Stayers	0.351	0.003	67,018	0.507	0.005	66,253
South Movers	0.378	0.014	3,819	0.509	0.020	3,775
South to Nouth	0.431	0.023	1,487	0.587	0.032	1,473
South of Severn-Wash Stayers	0.351	0.003	71,634	0.505	0.005	70,831
South of Severn-Wash Movers	0.369	0.014	3,961	0.499	0.020	3,916

Notes: 'North' defined as all counties in the regions of North East, North West, and Yorkshire and Humber, and 'South' defined as all counties in East of England, London, South East, and South West (see Table C.A.1 in Appendix C.A); 'North Stayers' defined as people who resided in the 'North' in 1881 and 1911; 'North Movers' defined as people who resided in the 'North' in 1881 but not in 1911; 'North to South' defined as people who resided in the 'North' in 1881 and in the 'South' in 1911; 'North of Severn-Wash Stayers' are defined as people who lived in counties north of the Severn-Wash Line in 1881 and 1911 while 'North of Severn-Wash Movers' are people who lived in counties north of the line in 1881 and south of the line in 1911 (Panel B of Figure 4.1 shows the counties north and south of the Severn-Wash Line); same rules apply for those with 'Southern Origins'.
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).

people who stayed North of the line exhibit less intergenerational mobility than those who moved to the South. The difference between movers and stayers, regardless of the definitions of North and South, is greater when the IV correlates are used instead of the OLS estimates, which suggests that the interaction between migration and intergenerational mobility may be obscured by measurement error.

On the other hand, the experience of migrants with 'Southern Origins' is very different. Migrating away from the South did not seem to lead to greater relative mobility. In terms of the OLS estimates, the father-son association in occupational status is higher for those who are migrants (rank-rank slope of 0.38) than those who remained (rank-rank slope of 0.35), especially among those moving to the North (rank-rank slope of 0.43). This is true too for the IV estimates, but the difference is smaller and the coefficients are more alike. The 'South

Stayers' have the lowest IV rank-rank slope (0.51), slightly lower than those of the 'South Movers', and 'South to North' migrants have the highest rank-rank slope (0.59). Anyhow, it is clear that Southern migrants did not enjoy greater intergenerational mobility than their peers who stayed, contrary to Northern migrants. The conclusion holds if the Severn-Wash Line is used to define North and South instead.

Table 4.6: Absolute Mobility of Stayers and Migrants, 1881–1911

	OLS		IV	
	Rank	*N*	Rank	*N*
Northern Origins				
North Stayers	38.4	50,095	33.2	49,484
North Movers	51.5	2,772	46.5	2,727
North to South	60.7	1,344	56.3	1,316
North of Severn-Wash Stayers	38.3	82,677	33.9	81,726
North of Severn-Wash Movers	58.1	3,296	54.2	3,250
Southern Origins				
South Stayers	43.1	67,018	39.5	66,253
South Movers	46.3	3,819	43.1	3,775
South to North	43.4	1,487	39.4	1,473
South of Severn-Wash Stayers	42.9	71,634	39.4	70,831
South of Severn-Wash Movers	46.2	3,961	43.1	3,916

Notes: 'North' defined as all counties in the regions of North East, North West, and Yorkshire and Humber, and 'South' defined as all counties in East of England, London, South East, and South West (see Table C.A.1 in Appendix C.A); 'North Stayers' defined as people who resided in the 'North' in 1881 and 1911; 'North Movers' defined as people who resided in the 'North' in 1881 but not in 1911; 'North to South' defined as people who resided in the 'North' in 1881 and in the 'South' in 1911; 'North of Severn-Wash Stayers' are defined as people who lived in counties north of the Severn-Wash Line in 1881 and 1911 while 'North of Severn-Wash Movers' are people who lived in counties north of the line in 1881 and south of the line in 1911 (Panel B of Figure 4.1 shows the counties north and south of the Severn-Wash Line); same rules apply for those with 'Southern Origins'.
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).

The story with absolute mobility is similar. Table 4.6 shows that those who stayed in the North on average experienced much less upward mobility than those who moved, and migrants to the South were extremely upwardly mobile. The 'North Stayers' have the lowest OLS and IV $\widehat{R}^{25}$ (38 and 33), while the $\widehat{R}^{25}$ for 'North Movers' are much higher (52 and 47) and highest for 'North to South migrants' (61 and 56). In contrast, there is very little difference in terms of absolute mobility for Southerners regardless of whether they migrated.

Upward mobility is only slightly higher among Southern migrants (with an OLS $\widehat{R}^{25}$ of 46 and an IV $\widehat{R}^{25}$ of 43) compared to non-migrants (OLS and IV $\widehat{R}^{25}$ of 43 and 40), and this advantage for migrants disappears if we focus on migrants to the North only (OLS and IV $\widehat{R}^{25}$ of 43 and 39).

A more formal way to demonstrate the extent of the differences in intergenerational mobility between these groups is to combine them into one sample, and running the analysis in a single regression, while also adding in controls for some observable characteristics:

$$\begin{aligned} Rank_{i,t} =& \alpha + \beta_1 Rank_{i,t-1} + \beta_2 NorthSouth_{i,t} \times Rank_{i,t-1} + \\ & \beta_3 SouthStayer_{i,t} \times Rank_{i,t-1} + \beta_4 SouthNorth_{i,t} \times Rank_{i,t-1} + \\ & \beta_5 NorthSouth_{i,t} + \beta_6 SouthStayer_{i,t} + \beta_7 SouthNorth_{i,t} + \gamma X_{i,t} + \epsilon_{i,t} \end{aligned} \tag{4.11}$$

Where $Rank_{i,t}$ is the occupational rank of son i from generation t and $Rank_{i,t-1}$ is the rank of the father of son i from generation *t-1.* The dummy variables $NorthSouth$, $SouthStayer$, $SouthNorth$ indicate if the son is a migrant from the North to the South, a stayer in the South, or a migrant from the South to the North. The reference group is the North Stayers — people who resided in the North in both 1881 and 1911. Thus, β_1 measures the father-son rank-rank slope for North Stayers and the constant α is used to compute the expected rank of sons born to fathers in the 25$^{\text{th}}$ percentile (R^{25}). The coefficients on the interaction terms between the dummy variables and the father's rank thus measures the difference between the rank-rank slopes of the other three groups vis-à-vis β_1 of the North Stayers, and coefficients on the dummy variables measure the difference in their constants from α. $X_{i,t}$ is are the son's age and age-squared in 1881.

Table 4.7 demonstrates the regression results obtained from estimating equation 4.11 using the IV method. As expected, the results (row 2-4 in column 1) show that there are statistically significant differences in the rank-rank slopes of North-South Migrants and South Stayers from the North Stayers, with both groups exhibiting more relative mobility — significantly more in the case of North-South Migrants. On the other hand, South-North

Table 4.7: Intergenerational Mobility for Migrants and Stayers

	Dep. Var. = Son's Rank in 1911	(1)	(2)
	INDEPENDENT VARIABLES	β	R^{25}
(1)	Father's Rank in 1881 (North Stayer)	0.577***	
		(0.007)	
(2)	North-South Migrant X Father's Rank in 1881	-0.214***	
		(0.040)	
(3)	South Stayer X Father's Rank in 1881	-0.070***	
		(0.008)	
(4)	South-North Migrant X Father's Rank in 1881	0.0108	
		(0.032)	
(5)	Constant (North Stayer)	19.34***	33.8
		(0.889)	
(6)	North-South Migrant	28.44***	56.9
		(2.938)	
(7)	South Stayer	8.07***	40.1
		(0.463)	
(8)	South-North Migrant	6.00***	40.0
		(1.910)	
(9)	Age	-0.042	
		(0.180)	
(10)	Age2	0.006	
		(0.009)	

Notes: Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1; Father's Rank in 1881 instrumented by Father's Rank in 1891; column 1 are coefficients estimated from the regression; row 1 and 5 are the coefficients for the reference group 'North Stayers'; row 2-4 (6-8) are differences to row 1 (5) of the reference group; number of father-son pairs (N), where sons resided either in the North or South in 1881 and 1911 (no Midlands) = 122,239.
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).

Migrants show no difference to the North Stayers in relative mobility. Combining these with the coefficients from the dummy variables and the constant (row 5-8 in column 1) we can calculate differences in absolute mobility too (column 2). Here, the results show that North Stayers have the lowest absolute mobility and North-South Migrants the highest; South Stayers and South-North Migrants have the same level of absolute mobility that is higher than the North Stayers but much lower than North-South Migrants.

However, the issue with this comparison is that migrants appeared to have been positively selected from the North but migrant selection was much weaker for those from the

South. Table 4.8 shows that fathers of Northern migrants had much higher ranks than fathers of Northern stayers, whereas fathers of southern migrants and fathers were much more similar in status, though there are no substantial differences in the age profiles (in terms of their own and their fathers'). Positive selection of migrants means that there might be an overestimation of the returns to migration (Abramitzky et al., 2012).

Table 4.8: T-Test of Stayer and Migrant Means in 1881

	North			South		
	Stayer	Migrant	Diff.	Stayer	Migrant	Diff.
Father's Rank	51.37	63.31	-11.94***	47.78	49.81	-2.02***
	(0.13)	(0.54)	(0.57)	(0.11)	(0.50)	(0.49)
Father's Age	40.96	40.94	0.02	41.47	41.07	0.39***
	(0.03)	(0.12)	(0.14)	(0.02)	(0.10)	(3.74)
Own Age	9.40	9.23	0.17**	9.43	9.26	0.17***
	(0.01)	(0.06)	(2.80)	(0.01)	(0.05)	(3.40)
N	50,512	2,819		67,841	3,866	

Notes: Standard errors in parentheses; *** p<0.001, ** p<0.01, * p<0.05.
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).

4.6.2 Within-Brothers Analysis

To understand if the differences in relative and absolute mobility are driven by selective migration alone, or if there are certain factors that engendered greater social mobility in the South, we turn to the within-brother analysis using the method discussed in Section 4.4.4. We focus on migration from the North to the South and vice versa (thus excluding the Midlands), since the results on the spatial variations in intergenerational mobility suggest that there are substantial differences between the North and the South, while the Midlands vary across the spectrum; it also rules out cases where people may have simply moved across the border.

Firstly, table 4.9 illustrate the returns to migration before and after accounting for between-household differences. Column 1 shows that a 'naive' calculation of the returns to migration (following the specification outlined in equation 4.7), using the full triple-linked sample, yields a huge gain for migrants moving from the North to the South. This naive returns to migration implies that, in a pool of children whose families all resided in the North in 1881, migrating to the South by 1911 leads to an improvement of 23.7 ranks on average relative to their counterparts that remained in the North.

However, as shown earlier, North-South migration was strongly selective and much of these gains could potentially be explained by differences in family resources or underlying differences in human capital. Moving from the North to the South incurred large economic costs, which may have precluded children of poorer families — indeed, the t-test results in table 4.8 show that fathers of migrants had much higher status than fathers of stayers. As well as enabling this expensive long-distance internal migration, these migrants may also have accrued higher human capital due to them coming from wealthier, higher-status families. Therefore, it might be the case that had these North-South migrants stayed in the North instead, they still would have significantly outperformed their peers.

If we restrict the sample to households that have produced at least two male children (column 2), the gains from migration remains roughly the same. However, if we include only households that have produced at least one son who migrated away from the North and at least one who stayed (column 3), the gains are reduced to just 7.7 ranks higher even without adding the family fixed effects. This suggests that approximately two-thirds of the returns to migration can be explained by between-household differences of migrants and stayers. In other words, by removing households who are unable to produce children who eventually migrate and thus focus only on households that had enough resources to raise two or more sons who were equally capable (at least theoretically) of migrating to the South (but did not all follow through with migration), we can eliminate 16 of the 23.7 ranks in the native returns to migration. A comparison of the results between columns 1 and 2 suggest that there is no selection effects between families that have produced only one son versus families that have produced more than one son.

Table 4.9: Returns to Migration (Baseline vs. Fixed Effects)

Dep. Var. =	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$Rank_{Son}$	Northern Origins				Southern Origins			
Migrant	23.65***	24.10***	7.69***	6.82***	3.61***	2.93*	0.44	0.69
	(0.71)	(1.40)	(2.25)	(2.01)	(0.80)	(1.51)	(2.18)	(1.81)
Constant	48.55***	44.84***	53.88***	43.39***	51.95***	49.11***	38.08***	37.17***
	(1.39)	(2.50)	(13.02)	(15.11)	(1.10)	(1.94)	(12.16)	(12.04)
Age Controls	YES	YES	YES	YES	YES	YES	YES	YES
Only Include Brothers	NO	YES	YES	YES	NO	YES	YES	YES
At Least One Brother Moves/Stays	NO	NO	YES	YES	NO	NO	YES	YES
Family Fixed Effects	NO	NO	NO	YES	NO	NO	NO	YES
Observations	51,690	16,596	570	570	69,041	22,476	719	719
Households				267				334
R^2	0.016	0.013	0.022	0.051	0.000	0.000	0.003	0.007

Notes: Robust and Clustered (columns 4 and 8) standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1; columns 1-4 show results for those who lived in the North in 1881; columns 5-8 show results for those who lived in the South in 1881; columns 2 and 6 include households who have male siblings (at least 2 male children); columns 3-4 and 7-8 include only households who have male siblings of which 1 stayed and 1 migrated; Migrant status defined as North to South or South to North (excludes Midlands); reference group is the North (South) Stayer in column 1-4 (5-8); controls include age and age squared; number of observations (N) represent each father-son pairs.
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).

Adding the family fixed effects, as specified by equation 4.8, further reduces the gains in the sons' occupational ranks to 6.8 (column 4). This implies that up to 12 per cent of the gains to migration can be explained by between-household differences within this very select group of families that have produced at least one son who migrated away from the North but also at least one son that stayed. On the whole, the results indicate that a naive calculation of the returns to migration leads to a large increase in occupational ranks for North to South migration that is overstated by processes of selective migration. Nonetheless, controlling for selection of migrants through family fixed effects does not fully eliminate the gains from migration; there is a substantial gain in occupational status to be realised by moving from the North to the South.

On the other hand, for those moving away from the South to the North, the returns to migration are much lower even before accounting for between-household differences — a gain of just 3.6 ranks as shown in column 5 (unsurprising given that Southern migrants are only very slightly positively selected compared to Southern stayers). This gain remains at a similar level if we include only families that have produced two or more sons but without restrictions on whether they migrated or not (column 6). The gains from migration are essentially wiped out when using the brothers sample where one migrates and one stays and becomes statistically insignificant (column 7), and adding family fixed effects do not change the results in any meaningful way (column 8). This suggests that while migrating from the North to the South can yield substantial occupational gains relative to the stayers in the North, even after accounting for selective migration, people who migrate away from the South are no more advantageous than those that stay in the South. This shows that migration can have different impact depending on the origin of migrants.

In addition, table 4.10 explores the differences in intergenerational mobility between brothers who stay and move, following the specification set out by equation 4.9. Column 1 shows that those who move from the North to the South experience substantial relative mobility, with a much weaker father-son rank-rank slope compared to their brothers who stay in the North, while absolute mobility is much higher too (as the coefficient on the 'Migrant' dummy variable shows). In comparison, there is neither an advantage in relative mobility nor in absolute mobility for migrants from the South over their non-migrant brothers in the South (column 2).[14]

Table 4.10: Intergenerational Mobility Differences between Brothers

Dep. Var. = Son's Rank in 1911	(1)	(2)
VARIABLES	Northern Origins	Southern Origins
Migrant	24.86***	-1.390
	(7.437)	(5.110)
Instrumented Father's Rank in 1881 × Migrant	-0.267***	0.0430
	(0.101)	(0.0841)
Constant	45.84***	37.93***
	(15.22)	(12.25)
Controls	YES	YES
Brothers Sample	YES	YES
Family Fixed Effects	YES	YES
Observations	556	707
Number of Households	261	328

Notes: Clustered (by household) standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1; Father's Rank in 1881 instrumented by Father's Rank in 1891; column 1 shows results for brothers who lived in the North in 1881; columns 2 show results for brothers who lived in the South in 1881; Migrant status defined as North to South or South to North (excludes Midlands); reference group is the North (South) Stayer in column 1 (2); controls include age and age squared; number of observations (N) represent each father-son pairs.
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).

While there may still be unobserved characteristics that are idiosyncratic to individuals that influence migration, which cannot be accounted for by household fixed effects, the evidence suggests strongly that the North was a place where relative and absolute mobility in the labour market is more precluded than the South. Controlling for selective migration, migrants moving away from the North to the South still do much better than their counter-

[14]Both the results on the returns to migration and on intergenerational mobility differences hold even after controlling for the rural and urban character of the parish in 1881 and 1911 (table C.F.1 and table C.F.2 in appendix C.F), thus the differences in returns to migration cannot have been driven by rural-urban migration dynamics.

parts in the North, whereas moving away from the South to the North do not lead to any gains in occupational ranks or intergenerational mobility. Internal migration, as Breen and In (2024) conjectured, allowed people to escape the constraints on social mobility imposed by growing up in places of high social persistence.

4.7 Discussion

The results of this paper illustrate the extensive variations in rates of intergenerational mobility along geographical bounds in England. It extends the current literature on one aspect of the North-South divide in England — the difference in social mobility between regions. While it has been established that regional disparities in mobility can be found in post-war England (Rohenkohl, 2019; Bell et al., 2023; Carneiro et al., 2022), this paper extends into the late nineteenth century and finds that the North-South divide in intergenerational mobility is longstanding. The northerners in England have experienced lower relative and absolute mobility since at least the Victorian period. Therefore, while the North developed into the industrial centre of the British Empire during the Industrial Revolution, becoming more and more prosperous (and converging with the leading regions (Geary and Stark, 2018)), its people were becoming less and less mobile.

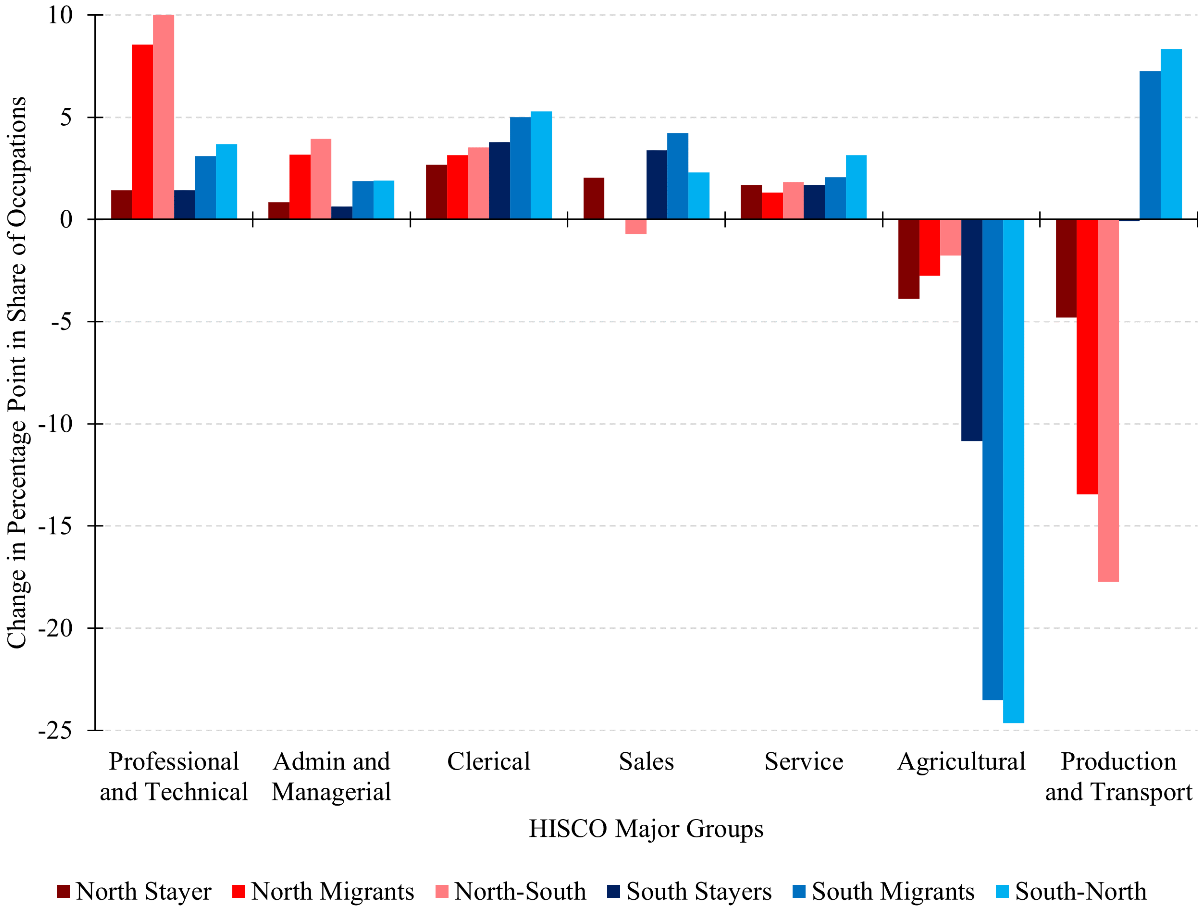
Although this paper is purely descriptive and cannot offer a causal claim as to why the North is less mobile than the South, the evidence provided suggest that the socioeconomic conditions of the North played a role in perpetuating the North-South divide. The analysis on migrants show that migration from the North to the South was positively selected, indicating that those with higher human capital 'voted with their feet' in favour of the South. This lends support to the findings of a selective outmigration of talents from the North over the long run in Clark and Cummins (2018). Moreover, even after controlling for selective migration, there still seems to be advantages for migrating southwards in facilitating intergenerational mobility, which suggests that there were longstanding differences in the economy and the labour markets of the North and South.

This paper also finds that there was a strong relationship between higher inequality and lower intergenerational mobility, showing that there was a 'Great Gatsby curve' in Victorian England, which suggests that the North had more occupational inequality than the South. However, the fact that positively selected migrants from the North migrated to the more equal South seems to be at odds with the Borjas (1987) theory of selective migration. This may be explained by the fact that the labour markets of the two regions were different and opportunities for certain sectors that were available in one region may be scarce in the other. Thus, inequality at the aggregate level may not reflect the premium for certain skills or human capital.

A conjecture is that the North's specialisation in manufacturing and mining activities meant that only specific types of skills and abilities were desired and rewarded. Subsequently, those who possessed different kinds of human capital and were thus more likely to work in sectors and occupations that were different to their fathers' (who are likely employed in the traditional 'Northern' industries) had to seek opportunities elsewhere in the country. This is evident in figure 4.8, which shows that Northern migrants were moving out of production and transport sectors, and into professional and technical occupations, at much higher rate than those who stayed. Southern migrants appeared to have been moving out of agriculture and into production and transport. The difference in absolute mobility between northern and southern stayers also suggest that labour markets in the South were more likely to foster upward mobility for those at the bottom than those in the North. Thus, the selection of migrants seems to be driven simply by strong pull factors in the South, consistent with the hypothesis expounded by Lee (1966).

Regardless of the causes, persistent disparity in social mobility between the North and the South is a problem. Social mobility is often considered as a measure of the 'equality of opportunity' (Corak, 2020) and deeply intertwined with the processes and outcomes of inequality; a high degree of mobility can help offset the negative impact of an individual

Figure 4.8: Changes in Occupational Structure from Fathers to Sons, 1881–1911



Notes: calculated from the difference between the percentage share of occupations in each HISCO major group (outlined in van Leeuwen et al. (2002)), held by sons in the triple-linked sample in 1911, and the percentage share of occupations in each major group held by their fathers in 1881.
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).

living in an unequal society (Solon, 1999). In the same vein, regional inequality could also be alleviated through encouraging greater social mobility in the North. However, this paper has shown that the North has been suffering from lower intergenerational mobility vis-à-vis London and the South East since at least the end of the nineteenth century. Thus, rather than reducing the gap, the mechanisms of intergenerational mobility has only exacerbated the North-South divide in England. This suggests that there is more to be done in terms of policy intervention in the North to improve access to opportunities and diminish social persistence, and there are possible inhibitory factors toward social mobility present in the North that caused this North-South divide in intergenerational mobility. This points to the need for further research in this topic.

4.8 Conclusion

Using 160,000 to 600,000 census-linked father-son pairs, this paper explores the geography of English intergenerational occupational mobility between 1881 and 1911. The findings indicate that by the late-Victorian period, there was already a North-South divide in social mobility in England. In terms of relative mobility, the rank-rank slopes demonstrate that social persistence was greater in the North than in the South. The association between fathers' and sons' ranks increases from 0.4 to above 0.6 when comparing the most mobile county to the least mobile county in England. On top of this, the North-South divide was even more marked in terms of absolute mobility, measured by the average expected rank of sons born to fathers belonging to the bottom 25$^{\text{th}}$ percentile of occupational rankings. The expected rank of said son from London is 43, which is 14 ranks higher than those from Northumberland. These results suggest that between by the end of the Victorian era, there was a already a substantial disparity in social mobility between the North and the South, with the North falling behind the South in intergenerational mobility. The post-war dominance of London and the South East as the engine of mobility was therefore longstanding and entrenched in history.

In addition, within-region inequality seems to have been associated with higher levels of social persistence, as areas with higher occupational inequalities also tended to be areas with lower relative and absolute mobility. This indicates that the 'Great Gatsby curve' existed not only in the context of cross-country comparisons of intergenerational mobility today but also in the setting of Victorian England.

Furthermore, we also detect significant differences in absolute occupational ranks and in intergenerational mobility between migrants and stayers from the North that are not observed in the experiences of their counterparts from the South. While over two-thirds of the gains in occupational ranks from north-south migration can be attributed to between-family selection of migrants, moving from the North to the South still yields a significant advantage of over 8 percentile ranks when comparing between brothers that moved and stayed; the movers also experience greater intergenerational mobility compared to the stayers within the same families. Such advantages are not conferred on migrants moving from the South to the North. The contrasting fate of people coming from a shared origin, even after accounting for selective migration, indicate that socioeconomic conditions in the North could have been responsible in creating labour markets that were more stagnant and less likely to generate mobility compared to those in the South.

The persistent North-South divide in social mobility is detrimental to the North's attempt at catching up to the South and to the political stability and unity of the country. There are still much about the determinants of social mobility and the scope for policy intervention to be learnt. Such investigations and knowledge are essential if we want to narrow the gap in social mobility and restore the North to its former glory.

Chapter 5

Conclusion: From Rags to Rags, Riches to Riches

This thesis demonstrates that England between 1851 and 1911 was a society of marked social persistence and limited occupational mobility, both intragenerationally and intergenerationally. For many Victorians and Edwardians, their final occupational attainment is likely to be similar to the status they held at the beginning of their career and to the occupational status of their father. The extent of occupational mobility was not uniform for the entire population, and significant variations exist. There appears to be a strong north-south divide in social mobility; places in the North tend generate less social mobility than places in the South. The experience of migrants further reinforces this divide. Thus, some subset of the wider population did exhibit a much larger degree of social fluidity than the rest of the country. However, the larger majority knew no 'rags to riches' story that Pip had achieved in Dickens' creation, nor the reverse; for the most part of their lives they were engaged in a languishing struggle of labouring, a protected livelihood of comfort, or an extravagant lifestyle of decadence.

The papers outlined in the earlier chapters contribute and progress our knowledge of social mobility in England's past by addressing the gaps in the existing literature and the shortcomings of earlier studies. Previously, the lack of reliable estimates of life course mobility in Victorian England hides the true character of social mobility in this period of time. This thesis, using a new and improved, high-quality linked census dataset, presents evidence that intragenerational occupational mobility was not unbounded, and any major movements along the occupational ladder were not common. In doing so, the paper also makes a methodological contribution by introducing the intragenerational rank-rank correlation as a way to summarise the extent of correlation between past and present occupational status. The two largest sectors in the nineteenth-century English economy, the primary and secondary sectors, may have been responsible for the low degree of life course mobility, as workers in these two sectors appear to be very occupationally stable. The more mobile tertiary sector was growing, but the new occupations — associated with the rise of modern services and professions — that sprung up were much less inclined to generate occupational mobility than the older tertiary occupations.

In addition, the new estimates for intergenerational mobility calculated in the second paper reveal substantial attenuation bias, induced by measurement error, in the results of earlier studies. Measurement error has long been recognised to be a cause for concern in the economic literature on income mobility, but has received far less attention in research on historical social mobility. The findings of this thesis suggest that even the use of occupations, often seen as a more stable indicator of status than income in a single year, is not free from measurement error. This has led to an overestimation of the degree of social mobility in Victorian England, and the new results revise our understanding of openness and equality of opportunities at this point in history. It shows nineteenth-century England to be a society of marked social persistence, yet not much more so than the United States at the same time. Neither of them, it would seem, were the 'land of opportunity'.

Furthermore, there is now a greater awareness of the regional diversity in people's experience with social mobility, though until now this was largely restricted to the temporal confines of the twentieth century. The third paper represents a significant breakthrough in the literature on this front, venturing further back into the late nineteenth century to uncover the past of spatial variations in intergenerational mobility. It finds that the north-south divide, which has imprinted on the English economy of the present day, was already evident in the late Victorian period from the lens of intergenerational occupational mobility. People who grew up in the North at the end of the nineteenth century were much less mobile than those who grew up in the South. On top of this, the geographic bifurcation in intergenerational mobility was reinforced by migration, as migrants moving from the north to the south were much more socially mobile than those who stayed in the north, yet for those from the south, there was no difference (in rates of social mobility) between the stayers and the migrants. The findings show that the north-south divide has long characterised the English economy and that regional disparity in intergenerational mobility today was a legacy of the past stemming back at least to the late-Victorian era.

The conclusions raise some additional questions which, though presently unable to be answered in this thesis, are useful and perhaps pivotal for future research agenda. First, the pertinent issue of measurement error, which has been demonstrated here and elsewhere, in studies of social mobility for both the contemporary and the historical. Given that sources of information on status, be it income, education, or occupation, vary in quality across time and space, it makes comparisons much more fraught with difficulties and inaccuracies. We may be mistakenly concluding that a place or a time period had greater social mobility not because it was true, but because the quality of the data is poorer. Thus, more research is required into the issue of measurement error and their impact on existing and future estimates of social mobility.

Secondly, the fact that the same set of 'rules' — political, economic, judicial — can lead to vastly different sets of outcomes for the people living in different regions of the same country, both in today's times and in the past, beckons the following logical question of why. While the existence of a substantial divide in social mobility between the North and the South have been confirmed, more work is to be done on establishing the causes of lower mobility in the former. A number of social, economic, or political variables have been touted as correlates or causes; the paper has confirmed the relationship with inequality and conjectured the role of local labour markets (occupational structure), and others have touted the role of education. The next step would be to establish firmer evidence on the potential causal relationships of these factors, and many lessons can still be drawn from history and from studying social mobility at the sub-national level.

Lastly, another avenue for future research may be to investigate the socioeconomic impact of greater social mobility, which would also emphasise the necessity of policies that promote openness. Many have detailed the various correlates of social mobility, such as inequality, and different types of mobility may have complementary effects on each other, as demonstrated by this thesis the case of migration and social mobility. However, most of these cannot yet demonstrate the causal relationship or direction between these variables. There are some progress being made on this front, and economists have documented that beliefs regarding social mobility can affect preferences for redistributive policies. Much more could be done on the subject of this, and looking at past events may provide us with the certainty of everything that could happen had already taken place. Economic history, therefore, may be a promising way forward for research on social mobility.

Appendix A

Paper 1 — Appendix

A.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).[1] 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 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 incon-

[1]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).

sistencies 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.

A.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, Boustan, Eriksson, James and Pérez, 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 — Preferred Linkage Method

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 — Alternative Linkage Strategy for Additional Results

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.

A.C Estimating False Positive Rate

Table A.C.1 shows the results of using spousal information to estimate the rate of false positives. The results suggest that the rate of false positives range from 36 to 26 per cent in 1851–1881 and 29 to 21 per cent in 1881–1911. These are upper-bound estimates, since the possibility of spousal information changing due to divorce and remarriage has not been accounted for.

Table A.C.1: Estimating False Positives Using Spousal Information

	1851–1881		1881–1911	
	N	Percentage	*N*	Percentage
Exact name	34,067	73.68	67,875	78.55
Exact name and Strict Age Range	29,520	63.85	61,734	71.45
Exact name and Loose Age Range	32,397	70.07	65,829	76.19
NYSIIS and Strict Age Range	30,562	66.10	63,935	73.99
NYSIIS and Loose Age Range	33,553	72.57	68,182	78.91
Total	46,234		86,405	

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).

To check the robustness of my results against false positives, the rate of life-course occupational mobility is estimated using only individuals belonging to the 'exact name and strict age range' category — the most conservative criteria. Table A.C.2 shows the intragenerational rank-rank correlation estimated using only individuals that satisfy the restrictions imposed. The size of the coefficients are very similar to the results in the main paper, and there is no change in the overall trends. Thus, the results are not biased in any significant ways by false positives.

Table A.C.2: Intragenerational Rank-Rank Correlation Corrected for False Positives

	1851–1861		1851–1881		1881–1891		1881–1911	
	HISCAM	CCC	HISCAM	CCC	HISCAM	CCC	HISCAM	CCC
Rank-Rank ρ	0.66***	0.67***	0.55***	0.54***	0.68***	0.70***	0.54***	0.58***
	(0.005)	(0.004)	(0.005)	(0.005)	(0.003)	(0.003)	(0.004)	(0.004)
N	28,408	28,357	28,408	28,357	60,008	59,697	60,008	59,697
R^2	0.454	0.491	0.323	0.337	0.462	0.507	0.297	0.348

Notes: robust standard errors in parenthesis; *** 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).

A.D Inverse Propensity Weight

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 the total linkable population in 1851 or 1881 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 (1881 or 1911), 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 and 1881 for the 1881–1911 sample. I then assign inverse propensity weights based on the following equation:

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

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

A.E Spearman's Rank Correlation

Table A.E.1 shows the intragenerational Spearman's rank correlation across 10 and 30 years for the periods 1851–1881 and 1881–1911. One difference between the Spearman's rank correlation and the rank-rank regression results shown in the main paper is that the ranks computed in this calculation are not percentile ranks but raw ranks instead. However, the correlation are very similar to those obtained with the rank-rank regression method, confirming that there was a strong correlation between past and present occupational status, and that the strength of this relationship neither increased nor declined substantially during the entire Victorian era.

Table A.E.1: Spearman's Rank Correlation

	1851–1881		1881–1911	
	10-year	30-year	10-year	30-year
Rank Correlation	0.622***	0.525***	0.667***	0.554***
N	131,512	131,512	237,439	237,439

Notes: *** 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).

A.F Intragenerational Elasticities

Table A.F.1 and A.F.2 show the short-term and long-term intragenerational elasticities, calculated by regressing the log occupational scores of current occupation on the log occupational scores of previous occupation. The elasticities are slightly higher than the intragenerational rank-rank correlation but the trend remains the same — there is a limited extent of life-course mobility which remains stable over the entire Victorian era.

Table A.F.1: 10-year Intragenerational Elasticities

	1851–1861				1881–1891			
	HISCAM		CCC		HISCAM		CCC	
Elasticities	0.65***	0.64***	0.75***	0.74***	0.70***	0.70***	0.78***	0.78***
	(0.004)	(0.005)	(0.005)	(0.005)	(0.002)	(0.003)	(0.003)	(0.003)
Weighted	NO	YES	NO	YES	NO	YES	NO	YES
N	131,512	131,512	131,263	131,263	237,439	237,439	236,276	236,276
R^2	0.361	0.359	0.523	0.519	0.453	0.448	0.597	0.596

Notes: robust standard errors in parenthesis; *** 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).

Table A.F.2: 30-year Intragenerational Elasticities

	1851–1881				1881–1911			
	HISCAM		CCC		HISCAM		CCC	
Elasticities	0.60***	0.59***	0.60***	0.59***	0.63***	0.63***	0.66***	0.65***
	(0.004)	(0.005)	(0.006)	(0.006)	(0.003)	(0.003)	(0.003)	(0.004)
Weighted	NO	YES	NO	YES	NO	YES	NO	YES
N	131,512	131,512	131,263	131,263	237,439	237,439	236,276	236,276
R^2	0.273	0.270	0.362	0.357	0.333	0.330	0.456	0.451

Notes: robust standard errors in parenthesis; *** 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).

A.G Transition Matrices

This section presents all the transition matrices used to calculate the mobility rates in Table 2.6 in Section 2.5.2. The weighted results are calculated using the sample reweighted using the strategy outlined in Appendix A.D. The diagonal cells are the percentages of the occupationally immobile from each class. The cells above the diagonals are the shares of men who experienced upward mobility, and those below the diagonals represent the rates of downward mobility.

Table A.G.1: Transition Matrix, 1851–1861

Class	1851					Total	N
1861	I	II	III	IV	V		
I	65.09	6.15	1.31	1.70	0.75	3.00	3,942
II	17.51	60.71	7.45	7.93	6.48	12.16	15,994
III	7.02	13.29	78.75	10.72	11.00	30.43	40,015
IV	6.88	11.57	7.87	68.11	11.78	22.20	29,195
V	3.51	8.27	4.63	11.54	70.00	32.21	42,366
Total	100.00	100.00	100.00	100.00	100.00	100.00	131,512
N	2,879	11,795	37,596	27,230	52,012	131,512	

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).

Table A.G.2: Transition Matrix, 1851–1881

Class	1851					Total	N
1881	I	II	III	IV	V		
I	62.52	9.02	2.57	2.84	1.33	4.02	5,293
II	19.94	54.93	10.71	13.62	9.35	14.94	19,648
III	7.36	15.79	71.69	13.75	15.77	31.16	40,973
IV	6.50	11.39	9.32	55.10	10.69	19.46	25,596
V	3.68	8.87	5.72	14.70	62.86	30.42	40,002
Total	100.00	100.00	100.00	100.00	100.00	100.00	131,512
N	2,879	11,795	37,596	27,230	52,012	131,512	

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).

Table A.G.3: Transition Matrix, 1881–1891

Class	1881					Total	N
1891	I	II	III	IV	V		
I	70.19	6.16	1.20	1.48	0.85	4.31	10,227
II	16.84	70.16	6.50	8.02	7.57	17.48	41,497
III	3.84	8.22	76.78	9.67	11.55	28.25	67,085
IV	5.83	8.83	10.44	68.77	14.69	25.95	61,615
V	3.30	6.63	5.08	12.05	65.35	24.01	57,016
Total	100.00	100.00	100.00	100.00	100.00	100.00	237,440
N	8,124	37,073	65,383	59,920	66,940	237,440	

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).

Table A.G.4: Transition Matrix, 1881–1911

Class	1881					Total	*N*
1911	I	II	III	IV	V		
I	66.15	8.42	2.25	2.43	1.35	5.19	12,321
II	20.11	63.81	10.93	13.20	11.17	20.14	47,823
III	4.51	9.56	66.11	10.71	15.53	26.93	63,944
IV	5.67	10.46	13.94	58.44	15.54	24.79	58,867
V	3.56	7.76	6.77	15.23	56.42	22.95	54,485
Total	100.00	100.00	100.00	100.00	100.00	100.00	237,440
N	8,124	37,073	65,383	59,920	66,940	237,440	

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).

Table A.G.5: Transition Matrix, 1851–1861 (Weighted)

Class	1851					Total	*N*
1861	I	II	III	IV	V		
I	63.76	6.72	1.41	1.77	0.91	3.42	4,499
II	18.63	61.04	7.71	8.41	7.42	13.61	17,901
III	6.50	11.97	77.37	10.75	11.35	30.93	40,673
IV	7.52	12.58	8.93	68.34	14.32	25.07	32,971
V	3.58	7.69	4.58	10.73	66.00	26.97	35,468
Total	100.00	100.00	100.00	100.00	100.00	100.00	131,512
N	3,249	13,729	39,388	30,949	44,197	131,512	

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).

Table A.G.6: Transition Matrix, 1851–1881 (Weighted)

Class	1851					Total	*N*
1881	I	II	III	IV	V		
I	61.02	10.40	2.87	3.15	1.75	4.78	6,289
II	20.96	55.11	10.98	14.07	10.64	16.45	21,628
III	6.31	13.11	69.40	12.77	15.31	30.46	40,061
IV	7.47	12.87	10.97	55.30	13.79	22.46	29,541
V	4.24	8.51	5.78	14.70	58.52	25.85	33,994
Total	100.00	100.00	100.00	100.00	100.00	100.00	131,512
N	3,249	13,729	39,388	30,949	44,197	131,512	

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).

Table A.G.7: Transition Matrix, 1881–1891 (Weighted)

Class	1881					Total	*N*
1891	I	II	III	IV	V		
I	69.49	6.42	1.24	1.53	1.03	4.52	10,743
II	17.38	69.33	6.35	7.92	8.3	17.79	42,241
III	3.63	8.01	75.98	9.66	11.28	28.39	67,410
IV	6.29	9.67	11.23	69.23	17.23	28.31	67,217
V	3.21	6.57	5.2	11.67	62.16	20.99	49,828
Total	100	100	100	100	100	100	237,440
N	8,425	38,235	67,369	65,695	57,716	237,440	

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).

Table A.G.8: Transition Matrix, 1881–1911 (Weighted)

Class	1881					Total	*N*
1911	I	II	III	IV	V		
I	65.21	8.74	2.33	2.44	1.57	5.44	12,921
II	20.17	62.17	10.45	12.7	11.81	20.07	47,664
III	4.34	9	65.01	10.14	14.29	26.33	62,519
IV	6.45	11.91	15.06	59.28	18.74	27.38	65,004
V	3.83	8.17	7.15	15.43	53.59	20.78	49,333
Total	100	100	100	100	100	100	237,440
N	8,425	38,235	67,369	65,695	57,716	237,440	

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).

A.H Occupational Mobility by Sectors (Numbers)

Table A.H.1: Occupational Mobility (PST), 1851–1861

1851 Sector (Right)	1	2	3	4	5	6	90
Same Occupation	27672	30797	2015	793	3346	1509	747
Same Sector							
Same Group	8190	6091	272	11	410	560	0
Different Group	1141	7921	696	69	1711	300	0
Different Sector							
1	0	3121	440	130	1199	1000	2263
2	3285	0	962	388	1573	905	1145
3	713	1820	0	195	621	203	139
4	201	674	196	0	289	81	57
5	1655	2764	678	292	0	803	304
6	2617	1393	192	96	716	0	472
90	2133	896	62	25	208	301	0
Total	47607	55477	5513	1999	10073	5662	5127

Notes: 1 = Primary, 2 = Secondary, 3 = Tertiary Dealers, 4 = Tertiary Sellers, 5 = Tertiary Services and Professions, 6 = Tertiary Transport and Communications, 90 = Sectorally Unspecified Occupations. *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); assignment of occupations to PST classification based on Wrigley (2010*b*).

Table A.H.2: Occupational Mobility (PST), 1851–1881

1851 Sector (Right)	1	2	3	4	5	6	90
Same Occupation	20959	25484	1635	567	2257	955	942
Same Sector							
Same Group	11158	6038	259	20	531	478	0
Different Group	1801	8554	651	75	2207	303	0
Different Sector							
1	0	4470	635	153	1504	1170	1967
2	4182	0	984	487	1680	980	1245
3	1208	2938	0	208	749	275	200
4	336	792	215	0	229	94	69
5	1959	4135	864	368	0	952	318
6	1824	1444	169	78	591	0	385
90	4180	1621	101	42	325	455	0
Total	47607	55476	5513	1998	10073	5662	5126

Notes: 1 = Primary, 2 = Secondary, 3 = Tertiary Dealers, 4 = Tertiary Sellers, 5 = Tertiary Services and Professions, 6 = Tertiary Transport and Communications, 90 = Sectorally Unspecified Occupations. *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); assignment of occupations to PST classification based on Wrigley (2010*b*).

Table A.H.3: Occupational Mobility (PST), 1881–1891

1881 Sector (Right)	1	2	3	4	5	6	90
Same Occupation	30209	55544	6115	2348	13795	7366	2753
Same Sector							
Same Group	12618	11337	493	25	1498	2537	0
Different Group	1159	14219	1465	198	5152	795	0
Different Sector							
1	0	3588	551	182	1701	1514	3045
2	4426	0	1863	747	3723	2849	3359
3	863	2677	0	513	1821	461	255
4	320	1134	414	0	935	218	146
5	2442	4923	1683	889	0	2201	611
6	2630	3015	410	196	2202	0	1269
90	3172	3088	147	69	534	913	0
Total	57839	99525	13141	5167	31361	18854	11438

Notes: 1 = Primary, 2 = Secondary, 3 = Tertiary Dealers, 4 = Tertiary Sellers, 5 = Tertiary Services and Professions, 6 = Tertiary Transport and Communications, 90 = Sectorally Unspecified Occupations. *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); assignment of occupations to PST classification based on Wrigley (2010*b*).

Table A.H.4: Occupational Mobility (PST), 1881–1911

1881 Sector (Right)	1	2	3	4	5	6	90
Same Occupation	19682	43285	4254	1336	10170	5235	1864
Same Sector							
Same Group	19868	12586	458	41	1997	2510	0
Different Group	1995	17269	1728	185	6609	958	0
Different Sector							
1	0	5800	984	265	2698	2091	3265
2	6124	0	2177	1122	4494	3520	3765
3	1214	4550	0	652	2212	659	348
4	440	1287	429	0	612	206	165
5	3099	8163	2430	1260	0	2781	815
6	2523	3187	483	225	1974	0	1216
90	2894	3398	197	81	592	894	0
Total	57839	99525	13140	5167	31358	18854	11438

Notes: 1 = Primary, 2 = Secondary, 3 = Tertiary Dealers, 4 = Tertiary Sellers, 5 = Tertiary Services and Professions, 6 = Tertiary Transport and Communications, 90 = Sectorally Unspecified Occupations. *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); assignment of occupations to PST classification based on Wrigley (2010*b*).

A.I Occupational Group Mobility Ranks

Table A.I.1: Rankings of Occupational Groups, 1851–1881

Group	PST	1861 Rank	1881 Rank	*N*
Forestry	1-3	23	25	240
Fishing	1-4	33	33	474
Agriculture	1-1	39	42	41409
Mining and quarrying	1-20	50	38	5457
Public Works	2-81	11	16	446
Stone and mineral processing industries	2-76	16	17	184
Drink industries	2-2	19	20	531
Chemical, soap, adhesives, manufacture	2-55	20	22	251
Brick and tile manufacture	2-75	27	23	744
Food industries	2-1	29	26	2690
Industries using leather, bone etc.	2-30	30	46	1022
Printing	2-41	31	41	735
Earthenware, pottery manufacture	2-45	36	39	396
Paper industries	2-40	37	30	394
Minor manufactures and trades	2-85	38	32	217
Industries producing products from fibres	2-31	40	36	572
Machines and tools, making and operation	2-65	41	37	3043
Clothing	2-10	42	40	4482
Glass industries	2-46	43	35	238
Furnishing	2-35	44	47	288
Textiles	2-20	46	31	8825
Boat and ship building	2-71	47	49	782
Wood industries	2-25	48	45	2860

Precious metals and jewelry	2-50	49	53	330
Non-ferrous metal manufacture and products	2-62	51	48	1018
Footwear	2-15	52	43	6003
Iron and steel manufacture and products	2-61	53	50	5092
Building and construction	2-80	54	54	10949
Instrument making	2-52	55	56	619
Road transport vehicles	2-70	56	55	1396
Fuel dealers	3-58	3	3	187
Dealers in minor products	3-85	6	5	148
Dealers in clothing and clothing accessories	3-10	7	8	593
Dealers in drink	3-2	21	19	1364
Dealers in textiles and products	3-20	26	24	914
Dealers in food	3-1	28	27	1473
Small traders	4-90	1	6	234
Sellers of printed products	4-41	8	1	622
Sellers of food	4-1	12	12	302
Sellers of chemical products	4-55	25	28	354
Domestic service	5-25	2	4	2076
Food, drink and accommodation services	5-1	9	9	440
Storage	5-10	10	10	379
Armed forces	5-50	13	13	264
Miscellaneous service industries	5-20	14	18	975
Commercial and administrative services	5-31	15	14	1910
Local government service	5-41	22	21	540
Financial services and professions	5-30	32	34	474
Professional support	5-36	34	44	671
National government service	5-42	35	51	136

Professions	5-35	45	52	2010
Rail transport	6-5	4	7	1165
Road transport (animal power)	6-1	5	2	2669
Communications	6-50	17	11	229
Sea transport	6-4	18	15	1107
Inland navigation	6-3	24	29	369

Notes: first digit in PST denotes sector: 1 = Primary, 2 = Secondary, 3 = Tertiary Dealers, 4 = Tertiary Sellers, 5 = Tertiary Services and Professions, 6 = Tertiary Transport and Communications; rank 1 = most mobile (lowest share with the same sector)

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); assignment of occupations to PST classification based on Wrigley (2010*b*).

Table A.I.2: Rankings of Occupational Groups, 1881–1911

Group	PST	1891 Rank	1911 Rank	*N*
Forestry	1-3	25	35	274
Fishing	1-4	35	33	838
Agriculture	1-1	36	47	43789
Mining and quarrying	1-20	50	49	12899
Stone and mineral processing industries	2-76	17	16	456
Public Works	2-81	18	23	1650
Drink industries	2-2	19	19	1460
Brick and tile manufacture	2-75	21	20	1405
Chemical, soap, adhesives, manufacture	2-55	26	31	802
Fuel industries	2-58	32	34	676
Food industries	2-1	33	30	4498
Paper industries	2-40	37	36	786

Industries using leather, bone etc.	2-30	39	44	2016
Rail transport vehicles	2-72	40	39	385
Minor manufactures and trades	2-85	43	45	342
Boat and ship building	2-71	44	55	1180
Glass industries	2-46	45	37	516
Machines and tools, making and operation	2-65	46	50	9153
Clothing	2-10	47	53	4210
Wood industries	2-25	48	52	4420
Furnishing	2-35	49	43	452
Textiles	2-20	52	46	11703
Earthenware, pottery manufacture	2-45	53	48	636
Iron and steel manufacture and products	2-61	54	54	10621
Precious metals and jewelry	2-50	55	40	818
Industries producing products from fibres	2-31	56	42	540
Instrument making	2-52	57	38	1216
Footwear	2-15	58	41	5532
Building and construction	2-80	61	63	24853
Gas equipment	2-67	62	56	496
Printing	2-41	63	60	2540
Non-ferrous metal manufacture and products	2-62	64	61	2145
Road transport vehicles	2-70	65	64	2503
Dealers in minor products	3-85	3	3	311
Fuel dealers	3-58	9	9	541
Dealers in wood and wood products	3-25	15	15	287
Dealers in clothing and clothing accessories	3-10	20	18	814
Dealers in textiles and products	3-20	27	26	2554
Dealers in drink	3-2	30	25	4054

Dealers in food	3-1	31	28	3050
Sellers of printed products	4-41	4	2	1474
Sellers of clothing and clothing accessories	4-10	10	4	257
Small traders	4-90	11	8	554
Sellers of paper products	4-40	13	6	278
Sellers of food	4-1	14	11	865
Sellers of chemical products	4-55	28	27	692
Armed forces	5-50	2	5	1369
Storage	5-10	6	12	1226
Domestic service	5-25	8	13	3646
Food, drink and accommodation services	5-1	12	14	1190
Miscellaneous service industries	5-20	16	17	1751
Commercial and administrative services	5-31	22	21	8688
Entertainment	5-15	23	24	327
National government service	5-42	38	57	348
Financial services and professions	5-30	41	59	2265
Professional support	5-36	42	58	1953
Local government service	5-41	51	51	2350
Media	5-16	59	62	360
Professions	5-35	60	65	5747
Road transport (motorised)	6-2	1	1	388
Sea transport	6-4	5	10	1794
Road transport (animal power)	6-1	7	7	5884
Inland navigation	6-3	24	22	547
Rail transport	6-5	29	29	8567
Communications	6-50	34	32	1635

Notes: first digit in PST denotes sector: 1 = Primary, 2 = Secondary, 3 = Tertiary Dealers, 4 = Tertiary

Sellers, 5 = Tertiary Services and Professions, 6 = Tertiary Transport and Communications; rank 1 = most mobile (lowest share with the same sector)

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); assignment of occupations to PST classification based on Wrigley (2010*b*).

Appendix B

Paper 2 — Appendix

B.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).[1] 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 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 incon-

[1]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).

sistencies 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.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, Boustan, Eriksson, James and Pérez, 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 — Preferred Linkage Method

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 — Alternative Linkage Strategy for Additional Results

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.

B.C Estimating False Positive Rate

When estimating the rates of intergenerational mobility from linked census data, it has now become commonplace to emphasise the importance of reducing Type I errors (false positives) since a sample with a high share of false positives may attenuate the IGE estimated and therefore overstate the extent of social mobility.

The standard procedure for calculating the rate of false positives associated with a census linking process in the literature is to benchmark the linked dataset with a high-quality reference dataset. For instance, Bailey et al. (2020)'s extensive review of some of the most widely known census linkage algorithms used three different reference datasets, two of which are hand-linked samples and one is a 'ground truth' sample with some noise added in to mimic errors in historical data. Abramitzky, Boustan, Eriksson, James and Pérez (2020) also reviewed their own linkage methods, where they compared their linkages to hand-linked family tree data. Such hand-linked samples are rare to find, and one can only make assumptions about their reliability.

Though there are reference samples that are not hand-linked, those are equally difficult to obtain. Anbinder et al. (2021)'s survey on matching Irish immigrants in the United States used the Emigrant Industrial Savings Bank records, which contain information about customers that are much more detailed than those found in a census. Massey (2017) circumvents the issue of finding high-quality historical datasets by using modern data instead, where she can guarantee the reliability of her reference sample through unique identifiers (i.e.such as Social Security Number). She then conducts record linkage on the same datasets and compares the linkage results with the true links. These exercises are incredibly useful for showing us the potential pitfalls of automated census linkage, but they cannot be replicated in a different historical context (such as for the purpose of estimating false positives in this paper).

In the absence of a high-quality reference dataset, I have devised a method of checking the rate of Type I errors associated with the census linkage process using double-linked samples instead. 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:

$$\gamma = 2\alpha - \alpha^2 - x \tag{B.C.1}$$

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

$$2\alpha - \alpha^2 - x - \gamma = 0 \tag{B.C.2}$$

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 B.C.1 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 the American censuses. For instance, Bailey et al. (2020) 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 B.C.1: 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	

Notes: 'UB' = Upper Bound; 'LB' = Lower Bound. *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).

B.D Inverse Probability Weight

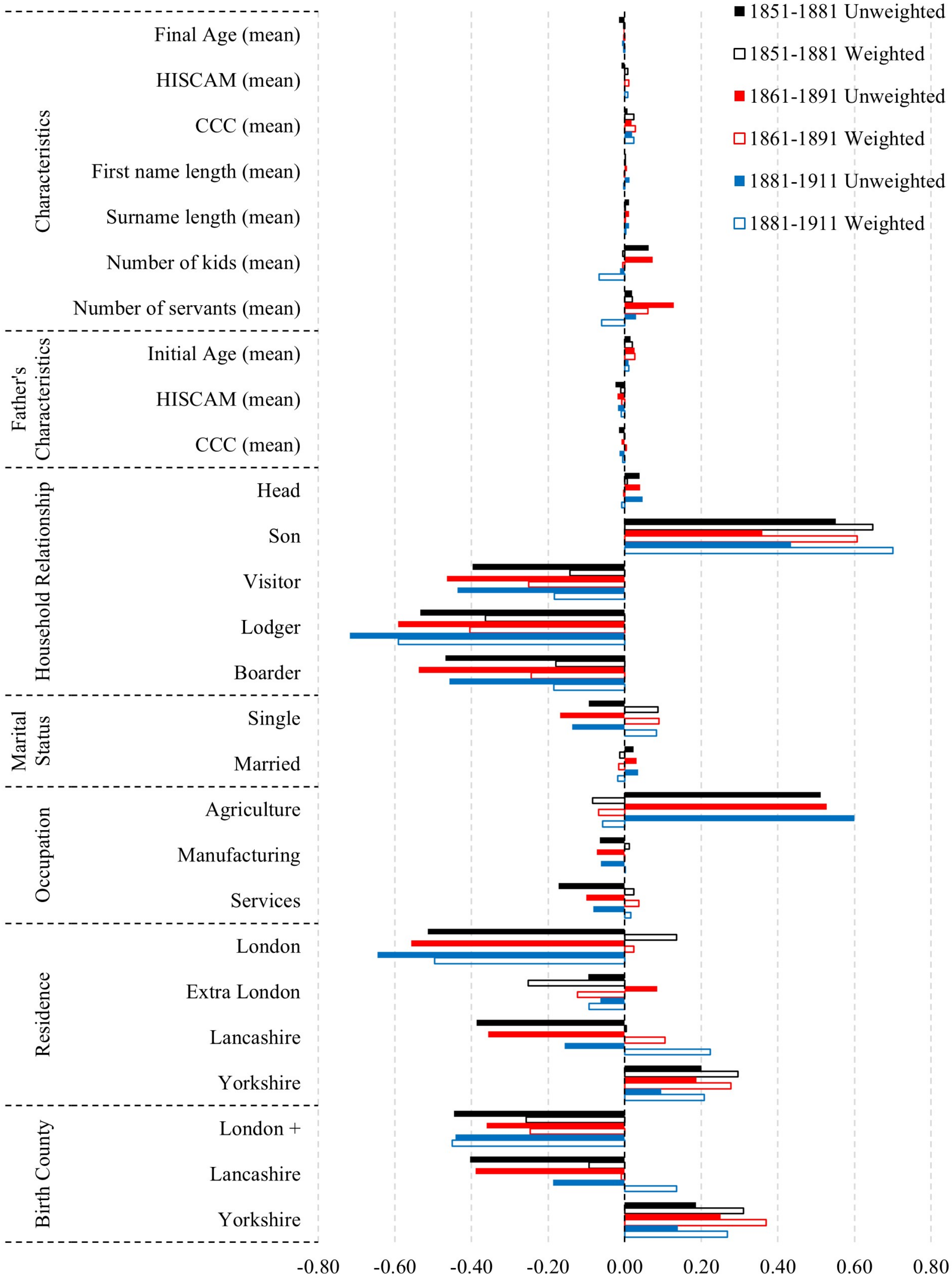
Figure B.D.1 shows the representativeness of the weighted sample in comparison to the unweighted sample. The weighted sample is more representative in almost all variables except in over-representing 'son' in relationship status and Yorkshire in birth and residence counties.

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) on the total linkable population of '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:

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

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

Figure B.D.1: Sample Representativeness (Weighted vs. Unweighted), 1851–1911



Notes: calculated based on Appendix B.E, Table B.E.4, B.E.5, B.E.6. 'Extra London' refers to the counties Middlesex, Kent, Essex, and Surrey. 'London+' refers to Middlesex, Kent, Essex, and Surrey (there is no 'London' in the standardised county of birth code, only 'City of London', which is a tiny part of London with no one recorded as being born there). 'Yorkshire' includes all Ridings of Yorkshire. *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).

B.E Sample Representativeness Table

Table B.E.1, B.E.2, and B.E.3 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 3.1. Table B.E.4, B.E.5, and B.E.6 shows the comparison between the unweighted multiple links sample with the weighted sample. The results were used to construct Figure B.D.1 in Appendix B.D.

Table B.E.1: 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-CCC (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-CCC (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)			
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

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 counties Middlesex, Kent, Essex, and Surrey. 'London+' refers to Middlesex, Kent, Essex, and Surrey (there is no 'London' in the standardised county of birth code, only 'City of London', which is a tiny part of London with no one recorded as being born there). 'Yorkshire' includes all Ridings of Yorkshire. All numbers are in percentages unless they are means. *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).

Table B.E.2: 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-CCC (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-CCC (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)			
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

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 counties Middlesex, Kent, Essex, and Surrey. 'London+' refers to Middlesex, Kent, Essex, and Surrey (there is no 'London' in the standardised county of birth code, only 'City of London', which is a tiny part of London with no one recorded as being born there). 'Yorkshire' includes all Ridings of Yorkshire. All numbers are in percentages unless they are means. *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).

Table B.E.3: 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-CCC (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-CCC (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	13.62	12.48	11.49
Yorkshire	13.25	14.13	14.51
Birth County (Son)			
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

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 counties Middlesex, Kent, Essex, and Surrey. 'London+' refers to Middlesex, Kent, Essex, and Surrey (there is no 'London' in the standardised county of birth code, only 'City of London', which is a tiny part of London with no one recorded as being born there). 'Yorkshire' includes all Ridings of Yorkshire. All numbers are in percentages unless they are means. *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).

Table B.E.4: 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-CCC (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-CCC (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)			
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

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 counties Middlesex, Kent, Essex, and Surrey. 'London+' refers to Middlesex, Kent, Essex, and Surrey (there is no 'London' in the standardised county of birth code, only 'City of London', which is a tiny part of London with no one recorded as being born there). 'Yorkshire' includes all Ridings of Yorkshire. All numbers are in percentages unless they are means. *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).

Table B.E.5: 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-CCC (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-CCC (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)			
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

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 counties Middlesex, Kent, Essex, and Surrey. 'London+' refers to Middlesex, Kent, Essex, and Surrey (there is no 'London' in the standardised county of birth code, only 'City of London', which is a tiny part of London with no one recorded as being born there). 'Yorkshire' includes all Ridings of Yorkshire. All numbers are in percentages unless they are means. *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).

Table B.E.6: 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-CCC (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-CCC (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	13.62	11.49	16.66
Yorkshire	13.25	14.51	16.01
Birth County (Son)			
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

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 counties Middlesex, Kent, Essex, and Surrey. 'London+' refers to Middlesex, Kent, Essex, and Surrey (there is no 'London' in the standardised county of birth code, only 'City of London', which is a tiny part of London with no one recorded as being born there). 'Yorkshire' includes all Ridings of Yorkshire. All numbers are in percentages unless they are means. *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).

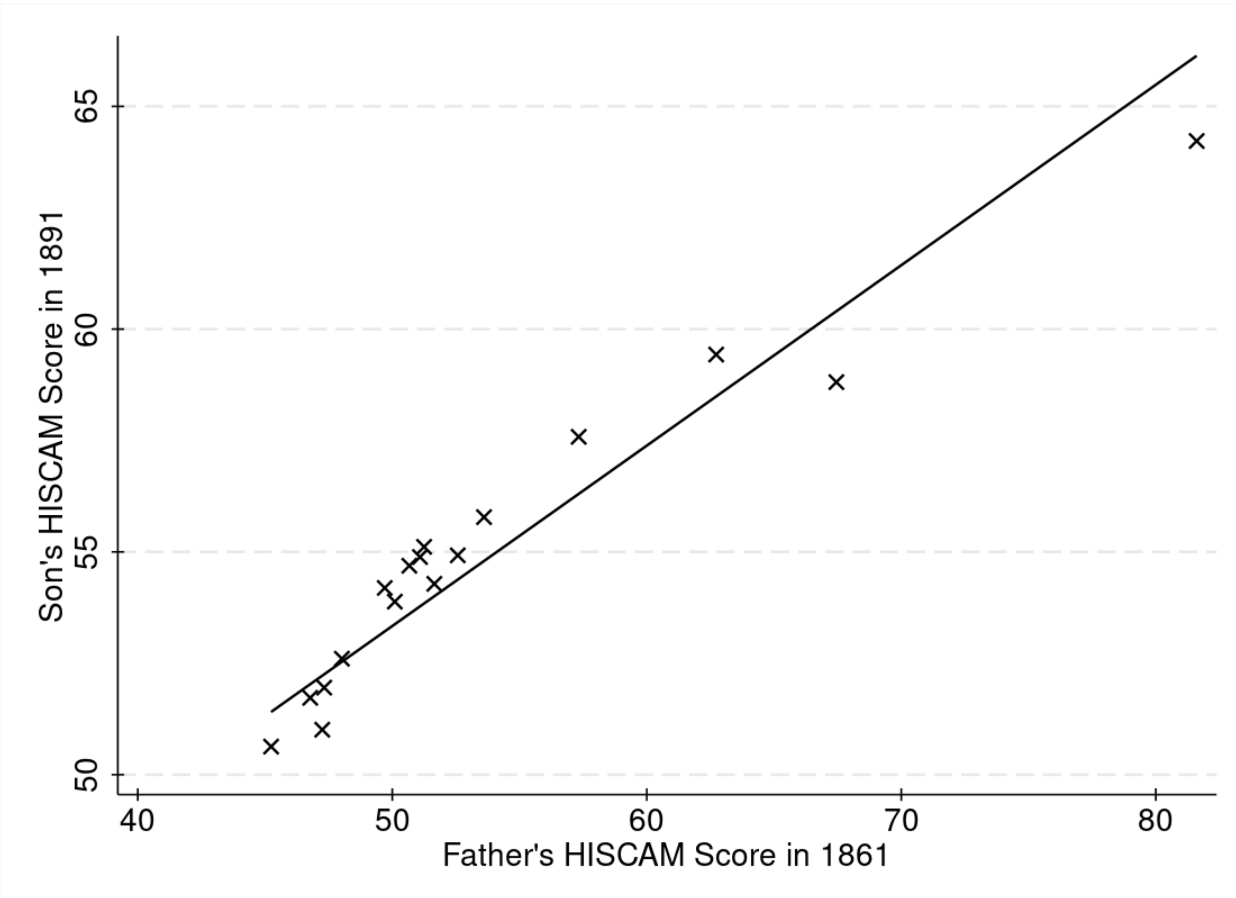
B.F Binned Scatter Plots

Figures B.F.1, B.F.2, and B.F.3 show the binscatter plots for 1851–1881, 1861–1891, and 1881–1911. The relationship between fathers' and sons' outcomes is clearly linear.

Figure B.F.1: Binned Scatter Plot for 1851–1881

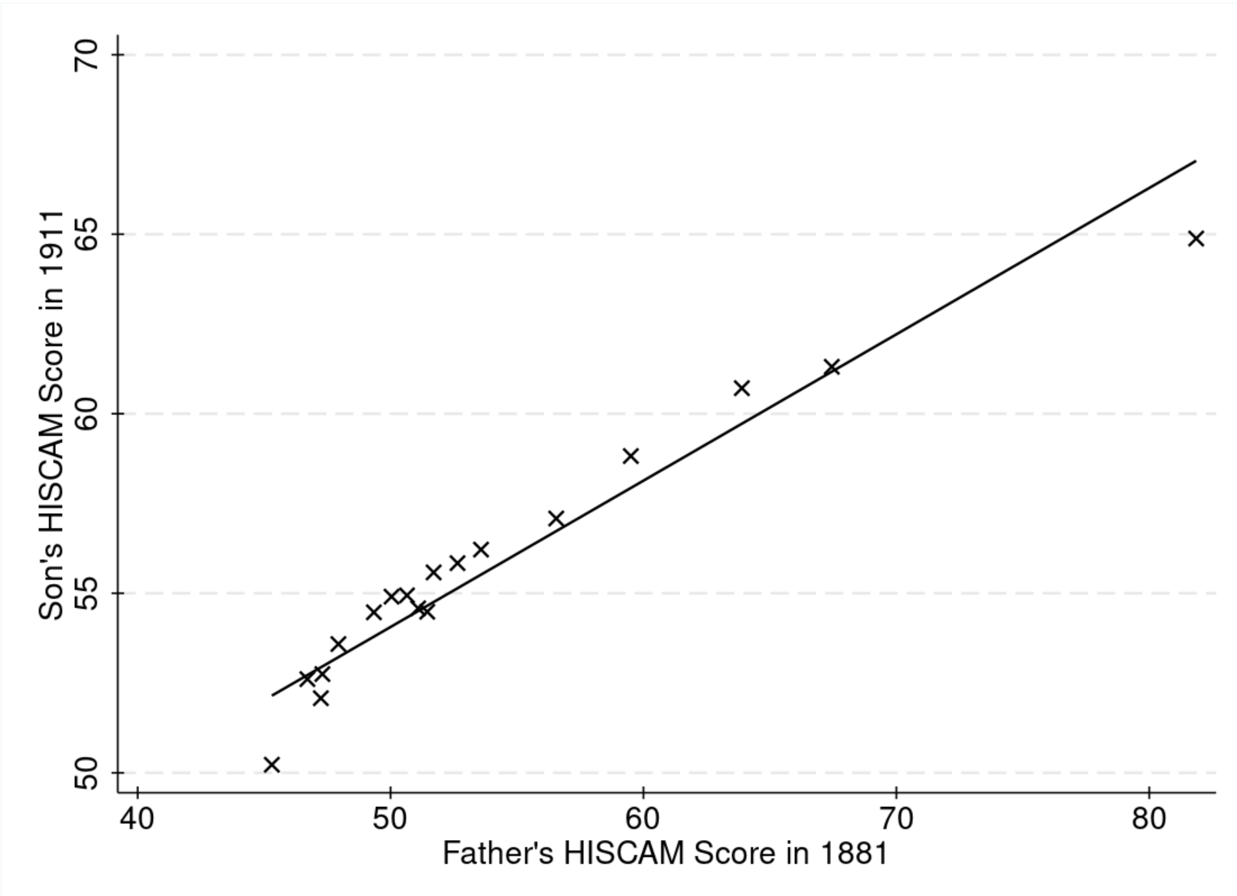
Sources: author's analysis of I-CeM (Schürer and Higgs (2014), UKDA, SN 7481) and I-CeM Names and Addresses (Schürer and Higgs (2015), UKDA, SN 7856).

Figure B.F.2: Binned Scatter Plot for 1861–1891



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).

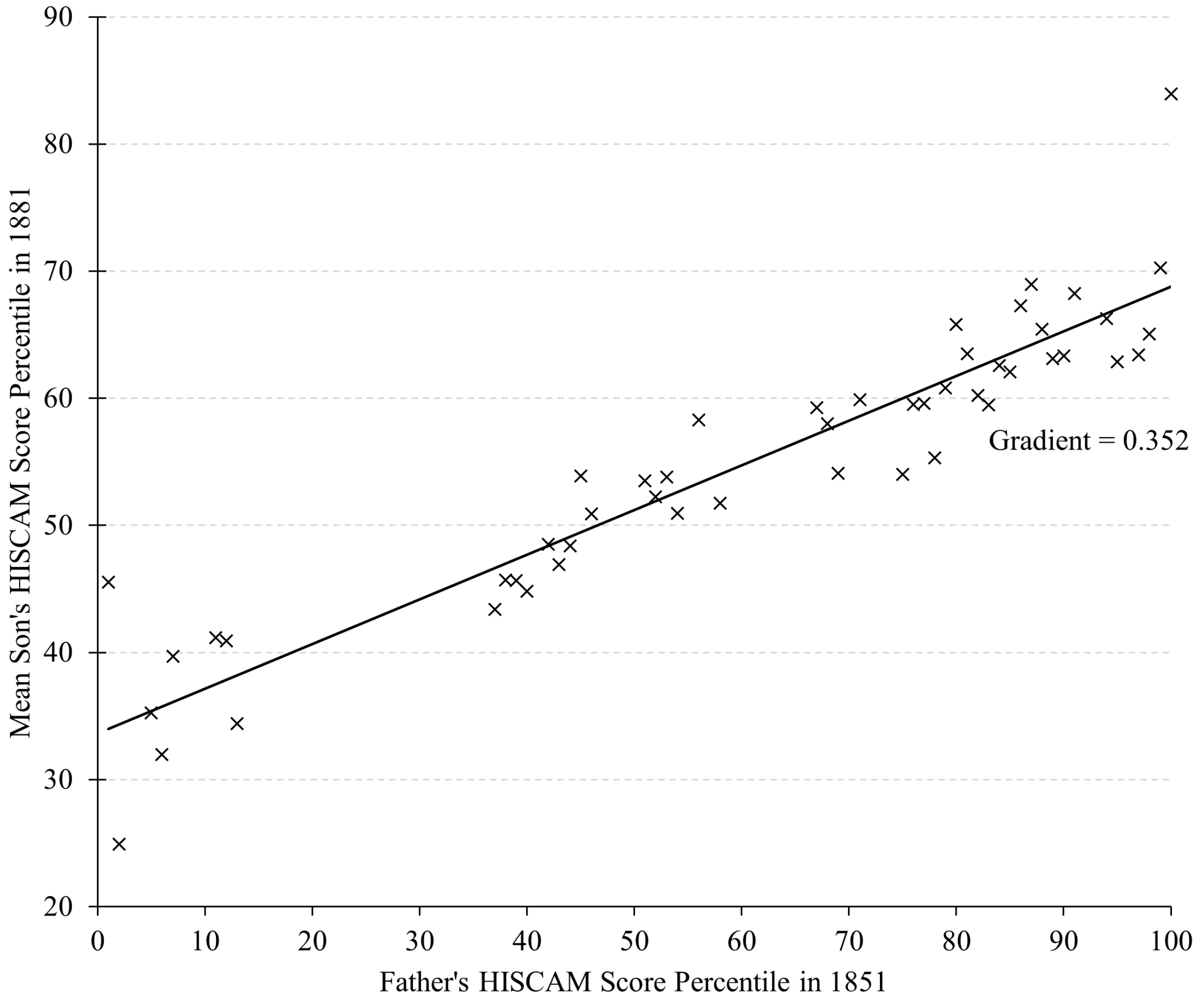
Figure B.F.3: Binned Scatter Plot for 1881–1911



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).

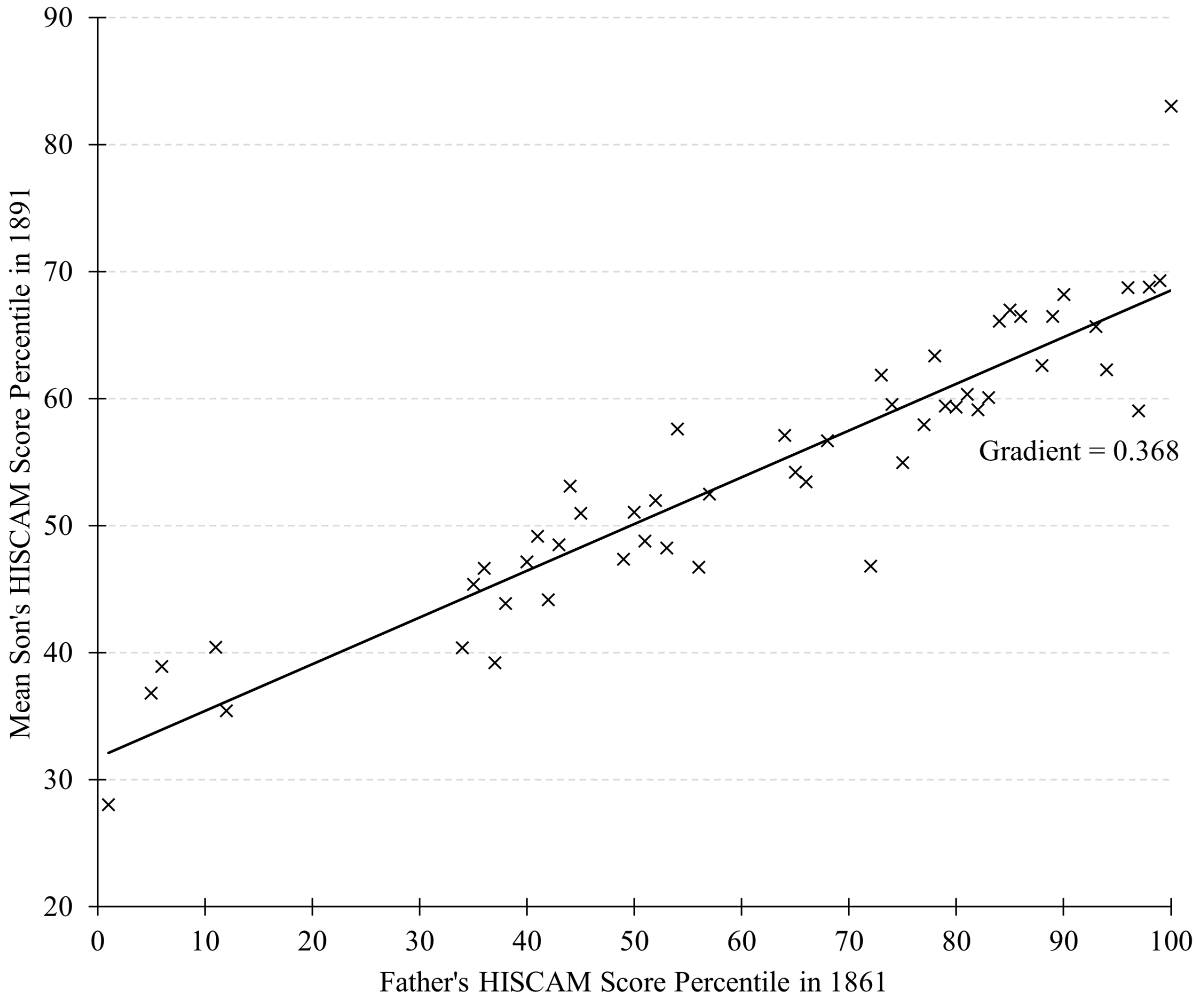
B.G Rank-Rank Correlations

Figure B.G.1: Father-Son Rank-Rank Correlation, 1851–1881



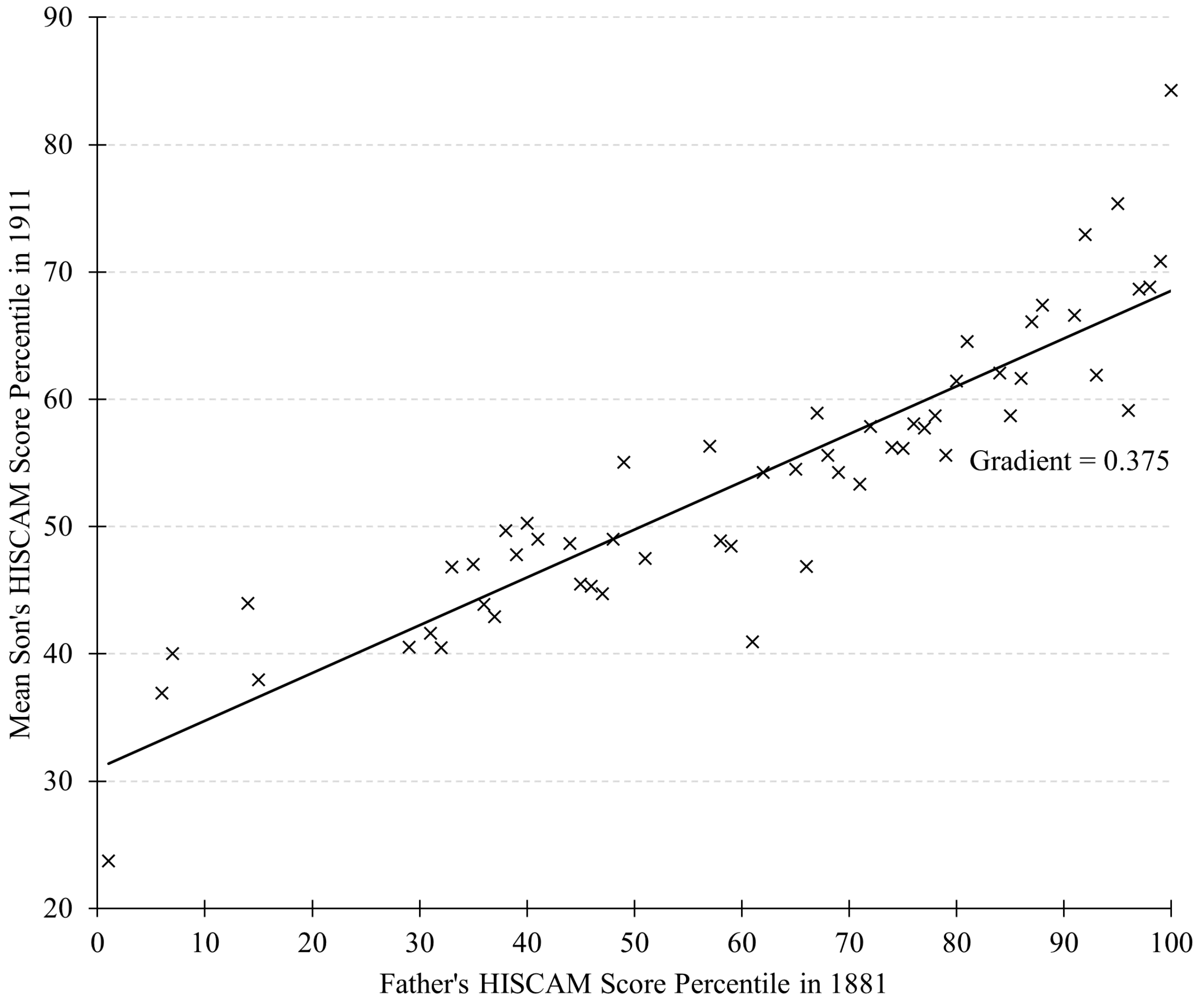
Sources: author's analysis of I-CeM (Schürer and Higgs (2014), UKDA, SN 7481) and I-CeM Names and Addresses (Schürer and Higgs (2015), UKDA, SN 7856).

Figure B.G.2: Father-Son Rank-Rank Correlation, 1861–1891



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).

Figure B.G.3: Father-Son Rank-Rank Correlation, 1881–1911



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).

B.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, 2011):

$$\rho = \beta \frac{\sigma_1}{\sigma_0} \tag{B.H.1}$$

Table B.H.1 shows the intergenerational correlations in occupational status calculated using the same formula.

Table B.H.1: Intergenerational Occupational Correlations Computed from Elasticities

	1851–1881		1861–1891		1881–1911	
	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 (Schürer and Higgs (2014), UKDA, SN 7481) and I-CeM Names and Addresses (Schürer and Higgs (2015), UKDA, SN 7856).

B.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 B.I.1: Life-Cycle Effects on IGE

	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

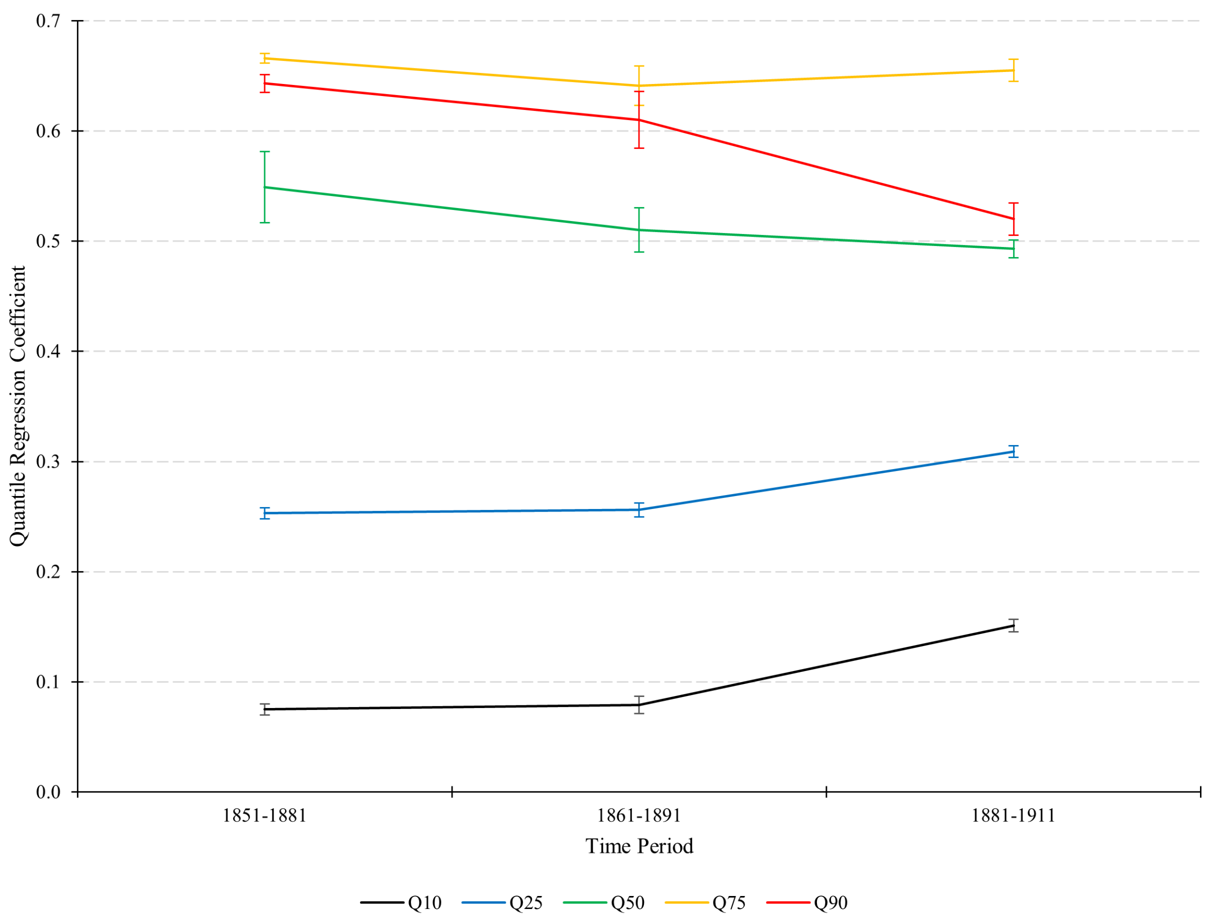
Sources: author's analysis of I-CeM (Schürer and Higgs (2014), UKDA, SN 7481) and I-CeM Names and Addresses (Schürer and Higgs (2015), UKDA, SN 7856).

The results in Table B.I.1 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.

B.J Quantile Regression Results

Figure B.J.1 shows the father-son association in occupational status from quantile regressions at the tenth, twenty-fifth, fiftieth, seventy-fifth, and ninetieth percentiles. The results

Figure B.J.1: Quantile Regression Results, 1851–1911



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).

suggest 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.

B.K Simulation of Minimum IGE

Table B.K.1 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 B.K.1: 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 (Schürer and Higgs (2014), UKDA, SN 7481) and I-CeM Names and Addresses (Schürer and Higgs (2015), UKDA, SN 7856).

B.L Mobility Tables

Table B.L.1: Intergenerational Mobility Rates, 1851–1881

Son's Class 1881	Father's Class in 1851				Total	*N*
	W	S	F	U		
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

Notes: W = white collar; S = skilled and semi-skilled; F = farmer; U = unskilled. *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).

Table B.L.2: Intergenerational Mobility Rates, 1881–1911

Son's Class 1911	Father's Class in 1881				Total	*N*
	W	S	F	U		
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

Notes: W = white collar; S = skilled and semi-skilled; F = farmer; U = unskilled. *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).

Table B.L.3: Intergenerational Mobility Rates (Corrected), 1851–1881

Son's Class 1881	Father's Class in 1851				Total	N
	W	S	F	U		
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

Notes: W = white collar; S = skilled and semi-skilled; F = farmer; U = unskilled. *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).

Table B.L.4: 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

Notes: W = white collar; S = skilled and semi-skilled; F = farmer; U = unskilled. *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).

B.M False Positives Arising from Using a Sample for Census Linkage

A crucial part of census linkage is the restriction on uniqueness of matches —- in order for person A from the first census to be matched to person B from the second census, there can be no other potential matches for both person A and B. The issue with using a 2 per cent sample in the 1851 census is that we can only ensure that there are no other potential matches for person A in the 1881 census, but we cannot be sure that there are no other potential matches for person B in the 1851 census because 98 per cent of the 1851 census has been cut off. To demonstrate how this may affect our linkage, I have generated my own two per cent sample of the 1851 census and tried to match people from this sample to the full 1881 census. This can then be checked for false positives by seeing how many of these matches can be found when linkage is conducted using the full census instead — if the use of a two per cent sample does not generate higher number of false positives, then we would expect all matches to be found in the full linked sample. Table B.M.1 shows the results.

Table B.M.1: Test for Additional False Positives from Using 2% Sample, 1851–1881

	Matched with 2% Sample	Found in Full Linked Sample	Rate
±2 Years	6245	5395	86.39
±5 Years	6557	1431	21.82

Notes: results for '±2 Years' are produced by allowing matches to differ by up to 2 years in inferred birth year while requiring matches to be unique within the same age band, and results for '±2 Years' are produced by allowing matches to differ by up to 5 years in inferred birth year while requiring matches to be unique within the same age band; 'Matched with 2% Sample' refers to the linked sample obtained while using a 2% sample of the 1851 census; the same matching algorithm is then run with the use of the full 1851 census, and individuals who have been successfully matched while using the 2% sample and who can also be found in the linked sample with the full census are shown in the column 'Found in Full Linked Sample'; the 'Rate' is calculated as Matched with 2% Sample / Found in Full Linked Sample * 100. *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).

The results clearly suggest that using a two per cent sample can lead to additional false positives than using the full census. When linkage is conducted using a stricter requirement of only allowing birth years to differ by at most two years (the approach taken in this paper), only 86 per cent of the successful matches from the two per cent sample can be found in the full linked sample — in other words, potentially 14 per cent of matches could be additional false positives. When linkage is conducted using a more relaxed requirement of allowing birth years to differ by five years (as Long (2013) did), the additional rate of false positives caused by the use of the 2 per cent sample is much higher. This is because in the second linkage specification, the restriction on uniqueness is widened — matches have to be unique within a five-year band. Therefore, the removal of the other 98 per cent of the population from the census means that more people would be incorrectly identified as ‘unique’.[2]

B.N Changes in Father’s HISCAM Score over 10-Year Period

Table B.N.1 shows how the father’s occupational status may change over a 10-year period. Despite the relatively short amount of time, only around 60 per cent of fathers’ HISCAM scores stay constant. The majority of the changes in HISCAM scores are relatively small in scale (less than 10). These changes could have plausibly been caused by either a misreporting or miscoding of occupations, or by temporary shocks to the father’s occupational status. However, there is still a rather sizeable amount of changes (slightly above 10 per cent of the sample) which are larger in magnitude (20 scores or above) that do occur in just 10 years. These are perhaps more likely to have been caused by data errors rather than actual shocks to a person’s status. Unfortunately, it is not possible to definitely conclude whether these changes occur because of transitory shocks to occupational status or errors in the reporting, recording, or coding or data, but the results should suggest that these changes are at least symmetrical — they are just as likely to move up as they are to move down.

[2]This is not to say that allowing birth years to differ by five years would entail a significantly higher rate of false positives in itself, only that it would cause a significantly higher number of false positives when linkage is conducted using a sample rather than the full census.

Table B.N.1: Changes in Father's HISCAM Score across 10-Year Period

HISCAM Change	1851	1861	1881
-60	0.03	0.02	0.02
-50	0.11	0.10	0.09
-40	0.38	0.36	0.41
-30	1.20	1.01	1.21
-20	2.85	2.87	3.23
-10	11.43	12.38	12.96
0	63.56	59.96	61.42
10	14.49	16.80	14.08
20	3.69	4.04	4.12
30	1.54	1.70	1.70
40	0.54	0.57	0.62
50	0.15	0.18	0.11
60	0.03	0.02	0.01

Notes: for 1851 and 1861, changes are calculated as (father's score in 1861 − father's score in 1851), and for 1881 it is (father's score in 1891 − father's score in 1881); all figures have been rounded down to the nearest decile if they are lower than 0 and rounded up to the nearest decile if they are above 0; if there are no changes in HISCAM scores, they are kept as 0. *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).

Appendix C

Paper 3 — Appendix

C.A Regions, Counties, and RSDs

Table C.A.1 shows the distribution of registration sub-districts within each county, as well as the region assigned to each county. In total, there are 1,994 RSDs within England in 1881, though approximately 200 of them are too sparsely populated to create a linked sample that is large enough to produce estimates that are statistically significant at the 10 per cent level.

C.B Alternative Method for Correcting Measurement Error in Rank Correlation

Nybom and Stuhler (2017) stipulate that the relationship between the observed rank ($\tilde{y}$) of an individual and their actual rank ($\tilde{y}^*$) is:

$$\tilde{y} = \alpha + \lambda \tilde{y}^* + \tilde{w} \tag{C.B.1}$$

The attenuation factor from measurement error in father and son's ranks are λ_{f} and λ_{s}; λ is less than or equal to one due to percentile rank transformation (the closer it is to one,

Table C.A.1: Number of RSDs by County, 1881

County (Region)	N	County (Region)	N
Bedfordshire (East)	20	London (London)	132
Berkshire (South East)	33	Middlesex (London)	23
Buckinghamshire (South East)	24	Norfolk (East)	61
Cambridgeshire (East)	29	Northamptonshire (East Midlands)	32
Cheshire (North West)	44	Northumberland (North East)	33
Cornwall (South West)	55	Nottinghamshire (East Midlands)	41
Cumberland (North West)	28	Oxfordshire (South East)	22
Derbyshire (East Midlands)	30	Rutland (East Midlands)	4
Devon (South West)	88	Shropshire (West Midlands)	51
Dorset (South West)	34	Somerset (South West)	68
Durham (North East)	33	Staffordshire (West Midlands)	59
Essex (East)	59	Suffolk (East)	47
Gloucestershire (South West)	53	Surrey (South East)	28
Hampshire (South East)	57	Sussex (South East)	58
Herefordshire (West Midlands)	24	Warwickshire (West Midlands)	41
Hertfordshire (East)	28	Westmorland (North West)	10
Huntingdonshire (East)	10	Wiltshire (South West)	40
Kent (South East)	64	Worcestershire (West Midlands)	34
Lancashire (North West)	171	Yorkshire East Riding (Yorkshire)	42
Leicestershire (East Midlands)	26	Yorkshire North Riding (Yorkshire)	51
Lincolnshire (East Midlands)	57	Yorkshire West Riding (Yorkshire)	150
		Total (England)	**1994**

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); regions assigned to counties based on NUTS level 1 from European Commission (2022).

the smaller the measurement error is). Therefore, the observed rank-rank slope (R_{observed}), without correcting for measurement error, is equal to:

$$R_{\text{observed}} = \lambda_{\text{f}} \lambda_{\text{s}} R_{\text{true}} \tag{C.B.2}$$

While regressing one percentile rank ($\tilde{y}_1$) on another observation of the same individual's percentile rank ($\tilde{y}_2$) yields, as long as the errors $\tilde{w}_1$ and $\tilde{w}_2$ are uncorrelated:

$$\frac{Cov(\tilde{y}_1, \tilde{y}_2)}{Var(\tilde{y}_2)} = \frac{Cov(\lambda_{\text{y}} \tilde{y}_1{}^* + \tilde{w}_1, \lambda_{\text{y}} \tilde{y}_2{}^* + \tilde{w}_2)}{Var(\tilde{y}_2)} = {\lambda_{\text{y}}}^2 + \frac{Cov(\tilde{w}_1, \tilde{w}_2)}{1/12} = {\lambda_{\text{y}}}^2 \tag{C.B.3}$$

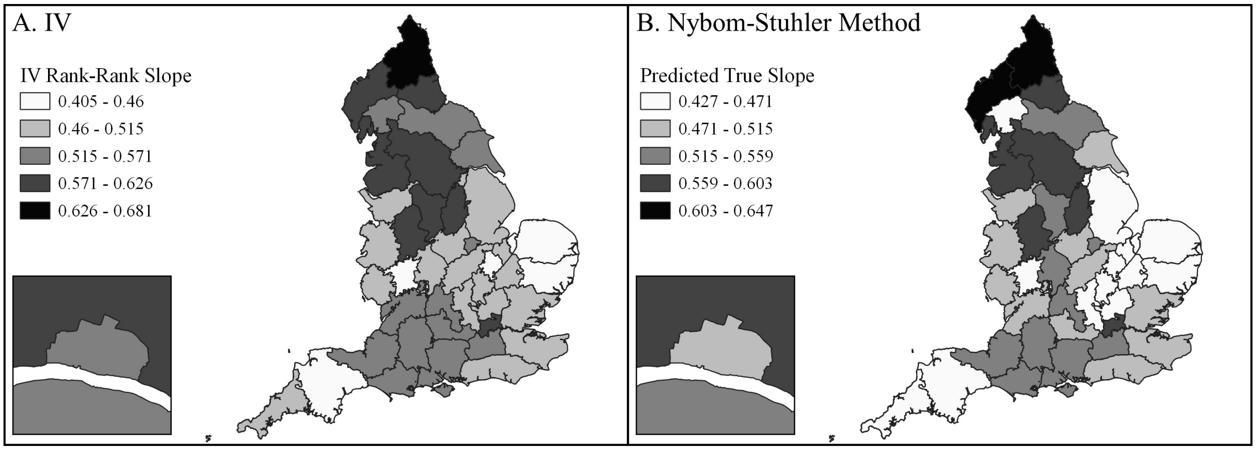
Using this formula, we can derive R_{true} from R_{observed} by estimating ${\lambda_{\text{y}}}^2$ from two observations of the fathers' and two observations of the sons' ranks, and taking the square root of the coefficient as λ_{f} and λ_{s}. Ideally, one would use two observations most adjacent to each other (i.e. from consecutive years), which is possible with income data (Nybom and Stuhler, 2017). Unfortunately, the census is only taken once per decade so the closest we have is two observations within ten years. This should not be an issue though, since estimates of ${\lambda_{\text{y}}}^2$ should be fairly stable over the ages of early-30s to 60 (Nybom and Stuhler, 2017).

A second issue here is that though this method can correct for measurement error in percentile ranks, it does not deal with life-cycle bias. This is relevant in the context of this paper because although there are observations of the sons' ranks in 1901 as well as 1911, using the former may introduce life-cycle bias to the estimates since some of them may be younger than 30 in 1901 and yet to reach their prime occupational status. Thus, the solution here would be to divide R_{observed} by the ${\lambda_{\text{y}}}^2$ obtained from the father's side, assuming that $\lambda_{\text{f}} = \lambda_{\text{s}}$.[1] This is a reasonable assumption since intragenerational mobility appears to be relatively stable for most of the Victorian era (Zhu, 2025).

Figure C.B.1 compares the rank-rank slopes obtained from the IV strategy with the predicted true estimates of the rank-rank slopes using Nybom and Stuhler (2017)'s method.

[1] An assumption also made by Ward (2023)

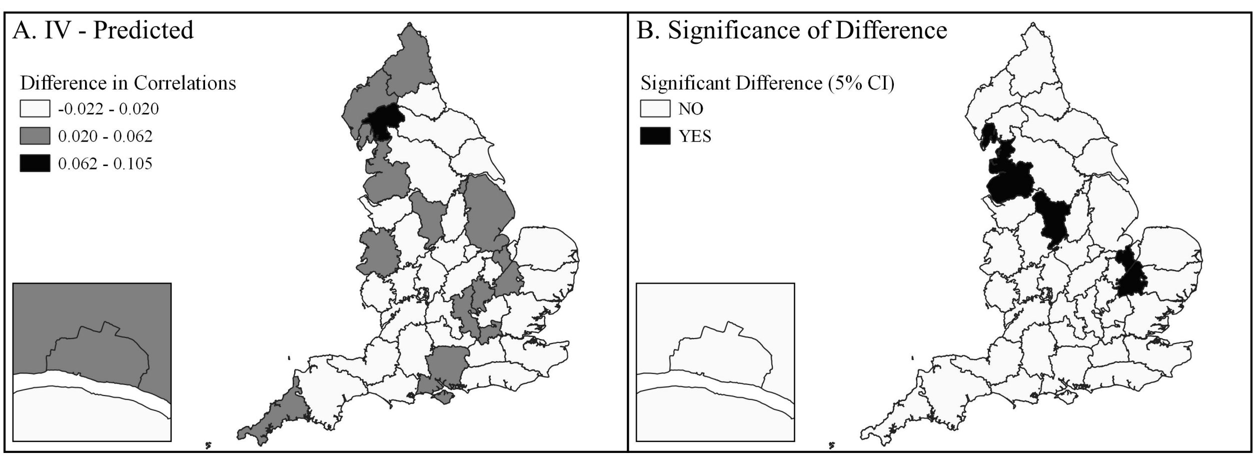
Figure C.B.1: Rank-Rank Slope Results (IV vs. Predicted True Estimates)



Notes: all coefficients are statistically significant at the 1% level.
Sources: GIS shapefile marks out 1851 borders of ancient counties of England, obtained from Satchell et al. (2023); results produced 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 results show that on the whole, the range of the slopes are very similar — the IV slopes (Panel A) range from 0.41 to 0.68 while the predicted true slopes (Panel B) range from 0.43 to 0.65. While there are differences in the estimates for some counties — particularly Westmorland — the overall pattern remains largely similar.

Figure C.B.2: Difference between IV Estimates and Predicted True Estimates)



Sources: GIS shapefile marks out 1851 borders of ancient counties of England, obtained from Satchell et al. (2023); results produced 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).

Panel A Fgiure C.B.2 shows the difference between the two estimates. For the majority of the counties, there are no substantial difference between the two (those in white), while for some the IV estimates are slightly larger than the predicted values (those in grey). Only Westmorland show a substantial difference. However, once we take into account of the standard errors and calculate the five per cent confidence interval of the IV estimates, we see that only four counties show significant difference between the two estimates (Panel B). Hence, the evidence provided here suggest that the IV strategy works just as well as the Nybom and Stuhler (2017) method for addressing the attenuation bias caused by measurement error in rank-rank slope.

C.C Fixed Effects Regression County-Level Results

Table C.C.1 shows the results from using the entire triple-linked sample to run the fixed effects regression:

$$SonRank_i = \alpha + \beta FatherRank_i + \gamma FatherRank_i County_i + County_i + \epsilon_i \qquad \text{(C.C.1)}$$

London is used as the baseline. All other estimates are the differences to London.

Table C.C.1: Fixed Effects Regression Results (Base = London)

	OLS		IV	
VARIABLES	Slope	Constant	Slope	Constant
London (Baseline)	**0.343*** **	**41.80*** **	**0.524*** **	**30.19*** **
	(0.0165)	**(1.203)**	**(0.0255)**	**(1.751)**
Bedfordshire	-0.0373	-5.484***	-0.0453	-1.434
	(0.0248)	(1.546)	(0.0367)	(2.176)
Buckinghamshire	-0.0178	-8.533***	-0.0239	-4.318*
	(0.0253)	(1.581)	(0.0376)	(2.223)
Berkshire	0.0198	-7.807***	-0.00410	-3.415
	(0.0245)	(1.582)	(0.0362)	(2.197)
Cambridgeshire	-0.0231	-7.517***	-0.0152	-3.591*
	(0.0252)	(1.530)	(0.0369)	(2.154)
Cheshire	-0.0217	-6.483***	-0.0610*	-2.678
	(0.0228)	(1.554)	(0.0351)	(2.256)
Cornwall	-0.0526**	-5.469***	-0.0617	-2.214
	(0.0255)	(1.642)	(0.0383)	(2.331)
Cumberland	0.0914***	-18.62***	0.0996**	-16.23***
	(0.0348)	(2.127)	(0.0504)	(2.915)
Derbyshire	0.0421*	-15.65***	0.0795**	-13.54***
	(0.0225)	(1.462)	(0.0340)	(2.072)
Devon	-0.0109	-6.661***	-0.0697**	-1.202
	(0.0213)	(1.439)	(0.0319)	(2.053)
Dorset	0.0236	-9.505***	0.00284	-5.194**
	(0.0266)	(1.647)	(0.0394)	(2.319)
Durham	0.0890***	-18.76***	0.0902***	-15.71***
	(0.0215)	(1.446)	(0.0323)	(2.038)
East Riding of Yorkshire	-0.0268	-6.348***	0.00113	-6.530***
	(0.0236)	(1.625)	(0.0377)	(2.443)
Essex	0.00662	-7.506***	-0.0382	-1.892
	(0.0204)	(1.370)	(0.0305)	(1.947)
Gloucestershire	0.0326	-7.769***	-0.00879	-3.258
	(0.0235)	(1.567)	(0.0345)	(2.194)
Hampshire	0.00914	-7.332***	0.0346	-6.264***
	(0.0229)	(1.525)	(0.0352)	(2.217)
Herefordshire	0.0446	-10.95***	-0.0478	-3.291
	(0.0454)	(2.574)	(0.0581)	(3.171)
Hertfordshire	-0.0260	-6.349***	-0.0444	-2.130
	(0.0240)	(1.552)	(0.0360)	(2.201)
Huntingdonshire	-0.0591	-6.843***	-0.119**	0.00629
	(0.0375)	(2.063)	(0.0543)	(2.870)
Kent	0.000894	-7.403***	-0.0305	-3.217*
	(0.0197)	(1.359)	(0.0297)	(1.946)

Lancashire	0.0357**	-11.90***	0.0625**	-11.22***
	(0.0182)	(1.293)	(0.0281)	(1.878)
Leicestershire	0.00406	-10.58***	-0.0286	-4.791**
	(0.0222)	(1.408)	(0.0332)	(1.983)
Lincolnshire	-0.0581***	-5.541***	-0.0410	-3.109
	(0.0211)	(1.408)	(0.0332)	(2.058)
Middlesex	0.0266	-4.799**	0.0733*	-6.094**
	(0.0272)	(1.896)	(0.0429)	(2.773)
Northumberland	0.145***	-20.08***	0.157***	-18.11***
	(0.0258)	(1.702)	(0.0370)	(2.324)
Norfolk	-0.0240	-6.668***	-0.0678**	-1.103
	(0.0210)	(1.382)	(0.0312)	(1.962)
North Riding of Yorkshire	0.0106	-11.61***	0.00176	-8.860***
	(0.0262)	(1.684)	(0.0404)	(2.457)
Northamptonshire	-0.0142	-8.990***	-0.0578*	-3.340
	(0.0226)	(1.431)	(0.0342)	(2.056)
Nottinghamshire	0.0837***	-17.72***	0.0753**	-13.11***
	(0.0224)	(1.431)	(0.0333)	(2.017)
Oxfordshire	0.0323	-10.41***	-0.00233	-5.199**
	(0.0253)	(1.565)	(0.0376)	(2.213)
Rutland	0.0892*	-13.41***	-0.000149	-5.452
	(0.0491)	(2.737)	(0.0629)	(3.339)
Shropshire	-0.0118	-9.625***	-0.0145	-6.162**
	(0.0294)	(1.776)	(0.0440)	(2.525)
Suffolk	-0.0274	-6.215***	-0.0649**	-0.824
	(0.0207)	(1.374)	(0.0309)	(1.956)
Somerset	0.0471**	-9.778***	0.0440	-6.741***
	(0.0221)	(1.460)	(0.0329)	(2.076)
Surrey	0.0399*	-6.287***	0.00268	-2.372
	(0.0235)	(1.605)	(0.0350)	(2.271)
Sussex	-0.00519	-6.837***	-0.0351	-2.527
	(0.0205)	(1.392)	(0.0309)	(1.988)
Staffordshire	0.0666***	-16.17***	0.0883***	-13.81***
	(0.0201)	(1.364)	(0.0307)	(1.961)
Warwickshire	0.0382*	-10.03***	-0.0141	-5.011**
	(0.0211)	(1.458)	(0.0313)	(2.041)
Westmorland	-0.0584	-4.330	0.0430	-7.827
	(0.0493)	(3.064)	(0.0845)	(4.916)
Wiltshire	0.0293	-10.04***	0.00473	-5.639**
	(0.0242)	(1.541)	(0.0370)	(2.223)
Worcestershire	-0.0588**	-5.633***	-0.0939**	-0.938
	(0.0245)	(1.582)	(0.0375)	(2.268)
West Riding of Yorkshire	0.0417**	-12.25***	0.0510*	-10.57***
	(0.0182)	(1.290)	(0.0281)	(1.873)

Notes: robust 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).

C.D Inference for Ranks

Mogstad et al. (2024) argue that there may be considerable uncertainty concerning the rank of each population when making a comparison between different places according to the value of some feature of said population, such as ranking neighbourhoods by the degree of intergenerational mobility as Chetty et al. (2014) did. Consequently, they have produced a method for constructing simultaneous confidence intervals for the ranks of all populations.

Figure C.D.1 shows the ranking of all English counties by their rank-rank slopes, with rank one being the least mobile, using the IV results from figure 4.2 from the main results. Figure C.D.2 shows the ranking of counties by absolute mobility, with rank one being the most mobile. As evidence in both figures, the confidence intervals are quite large for most counties. While we can infer that counties such as Northumberland, Durham, Staffordshire, and Derbyshire are clearly much less mobile, the rest are much more difficult to distinguish.

Since it is not possible to expand the number of observations while keeping the same IV method, the alternative is to amalgamate the counties into regions, which should reduce the confidence intervals. Figure C.D.3 shows the ranking of regions by relative mobility estimated using the IV method. Figure C.D.2 shows the ranking of regions by absolute mobility. Here, the divides are much clearer.

For relative mobility, it is evident that the North East is the least mobile region, and the East is the most mobile. While it is not possible to separate the rest into individual ranks, we can at least assort them into three groups, with Yorkshire and North West being the group that is the least mobile (after the North East), and South West and South East being the most mobile (after the East).

For absolute mobility, we see an even clearer divide between North and South. The most mobile group of regions are the Southern regions, with London, the East, South East, and South West sharing ranks 1–4, followed by East and West Midlands, North West, and Yorkshire sharing ranks 5–8, and the North East again being the least mobile.

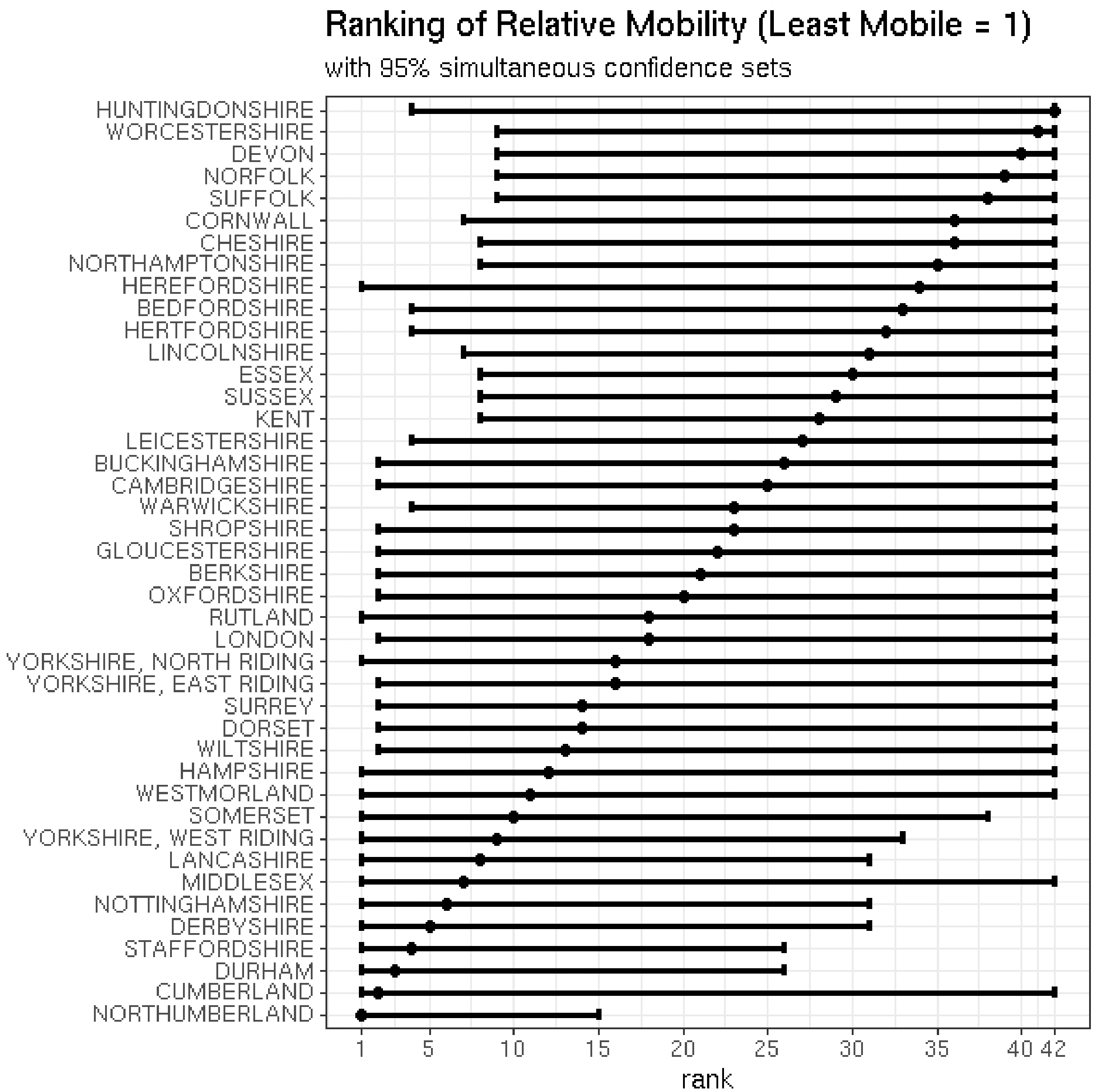
At the same time, we can also reduce the size of the confidence intervals by using the OLS results from the Baseline sample, which have a much larger sample size. Figure C.D.5 and C.D.6 show the simultaneous confidence sets for the rankings of relative and absolute mobility by counties constructed using the OLS estimates produced from the baseline sample. It is now possible to observe more counties forming into smaller groups by the degree of relative and absolute mobility.

For relative mobility, the northern counties of Northumberland and Durham have clearly the lowest relative mobility, and they, alongside Nottinghamshire, Derbyshire, and Cumberland, are the only possible candidates for the least mobile county in England. Nevertheless, it is still difficult to make inferences with certainty regarding the most mobile county (in terms of relative mobility) in England.

In terms of absolute mobility, the difference is clearer. Durham is the least mobile county, followed by possibly Northumberland, Nottinghamshire, Cumberland, Derbyshire, or Staffordshire. Lancashire and the West and North Ridings of Yorkshire also are shown to be quite immobile in terms of absolute mobility. The most mobile county is likeliest to be London (given the large confidence intervals for Westmorland), though what follows London is still unclear.

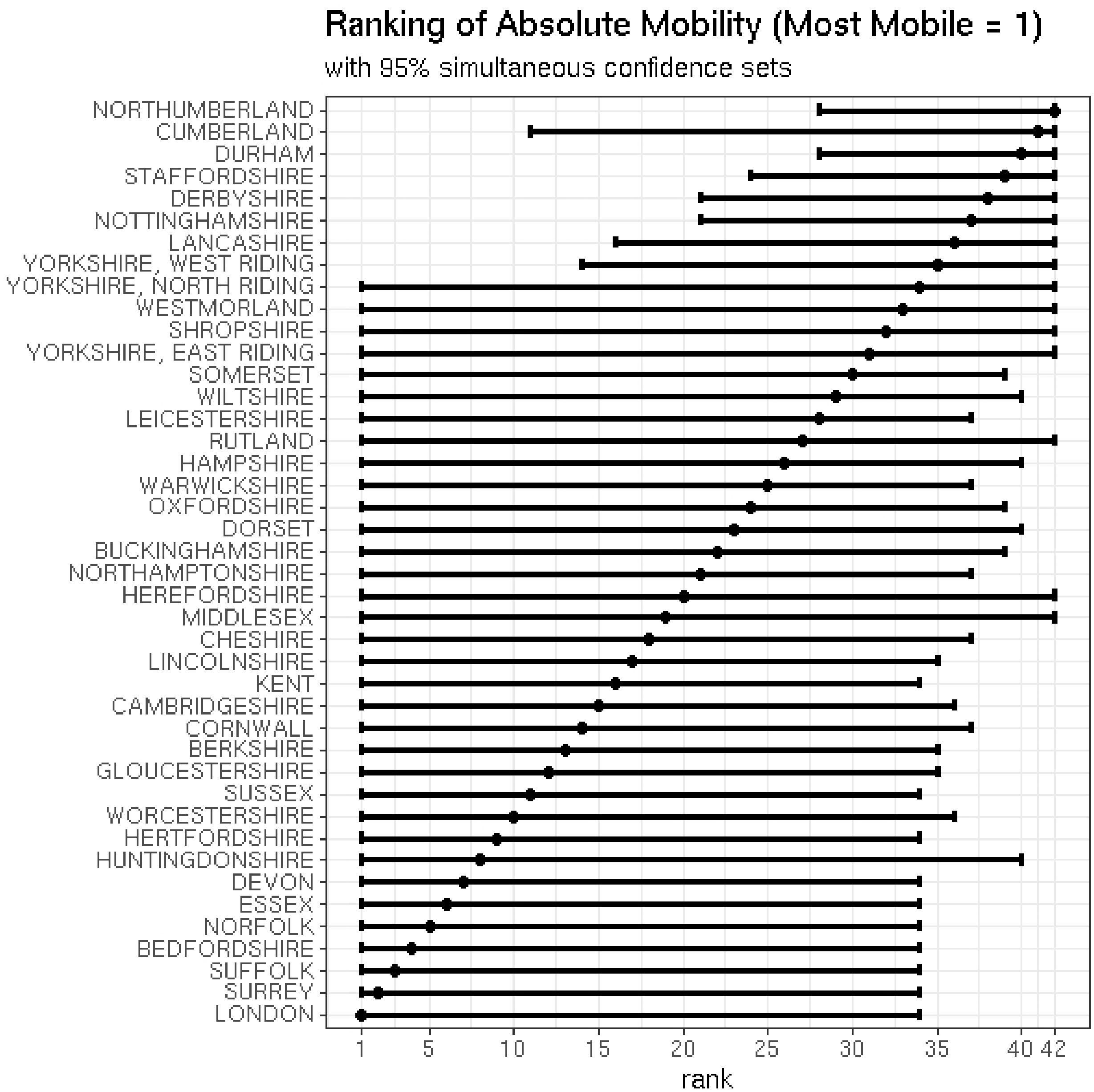
However, if we aggregate these counties into regions again, the rankings become more precise. For relative mobility, portrayed in figure C.D.7, the North East is the least mobile region, followed by the Midlands, North West, and Yorkshire. The most mobile regions are either London and the East, followed by the South East or South West. For absolute mobility (figure C.D.8), London is the most mobile and the North East is the least mobile. The other southern regions share ranks 2–4, while the rest are between ranks 5–8. The results therefore clearly demonstrates the divide between the North and the South, even if the exact ranking is uncertain.

Figure C.D.1: Ranking of Counties by Relative Mobility (IV)



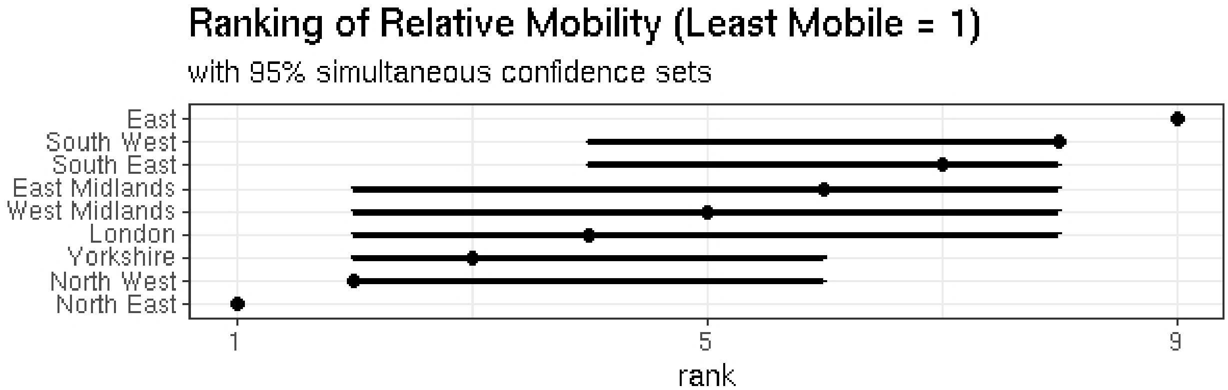
Sources: author's analysis of I-CeM (Schürer and Higgs (2014), UKDA, SN 7481) and I-CeM Names and Addresses (Schürer and Higgs (2015), UKDA, SN 7856). The 'CS Ranks' R package used to produce the results can be accessed at `https://cran.r-project.org/web/packages/csranks/index.html` and is based on the theory in Mogstad et al. (2024).

Figure C.D.2: Ranking of Counties by Absolute Mobility (IV)



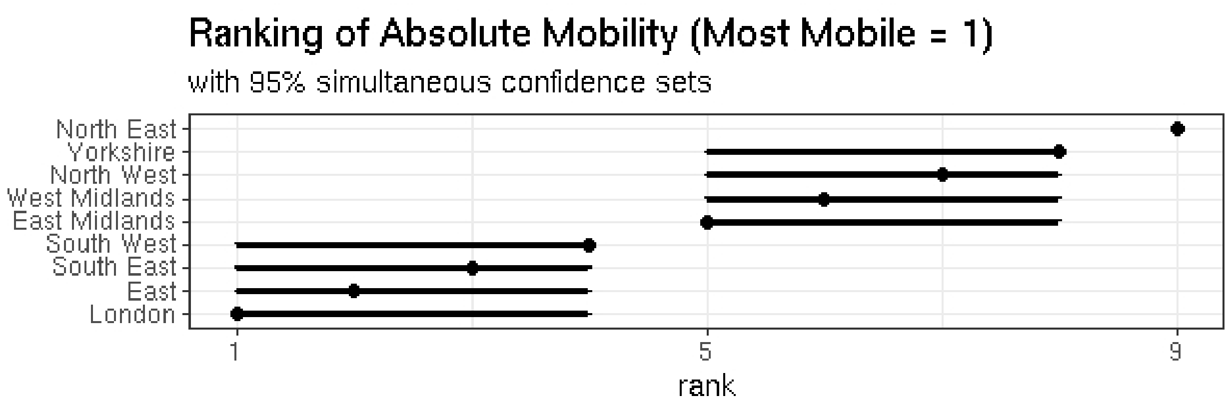
Sources: author's analysis of I-CeM (Schürer and Higgs (2014), UKDA, SN 7481) and I-CeM Names and Addresses (Schürer and Higgs (2015), UKDA, SN 7856). The 'CS Ranks' R package used to produce the results can be accessed at `https://cran.r-project.org/web/packages/csranks/index.html` and is based on the theory in Mogstad et al. (2024).

Figure C.D.3: Ranking of Regions by Relative Mobility (IV)



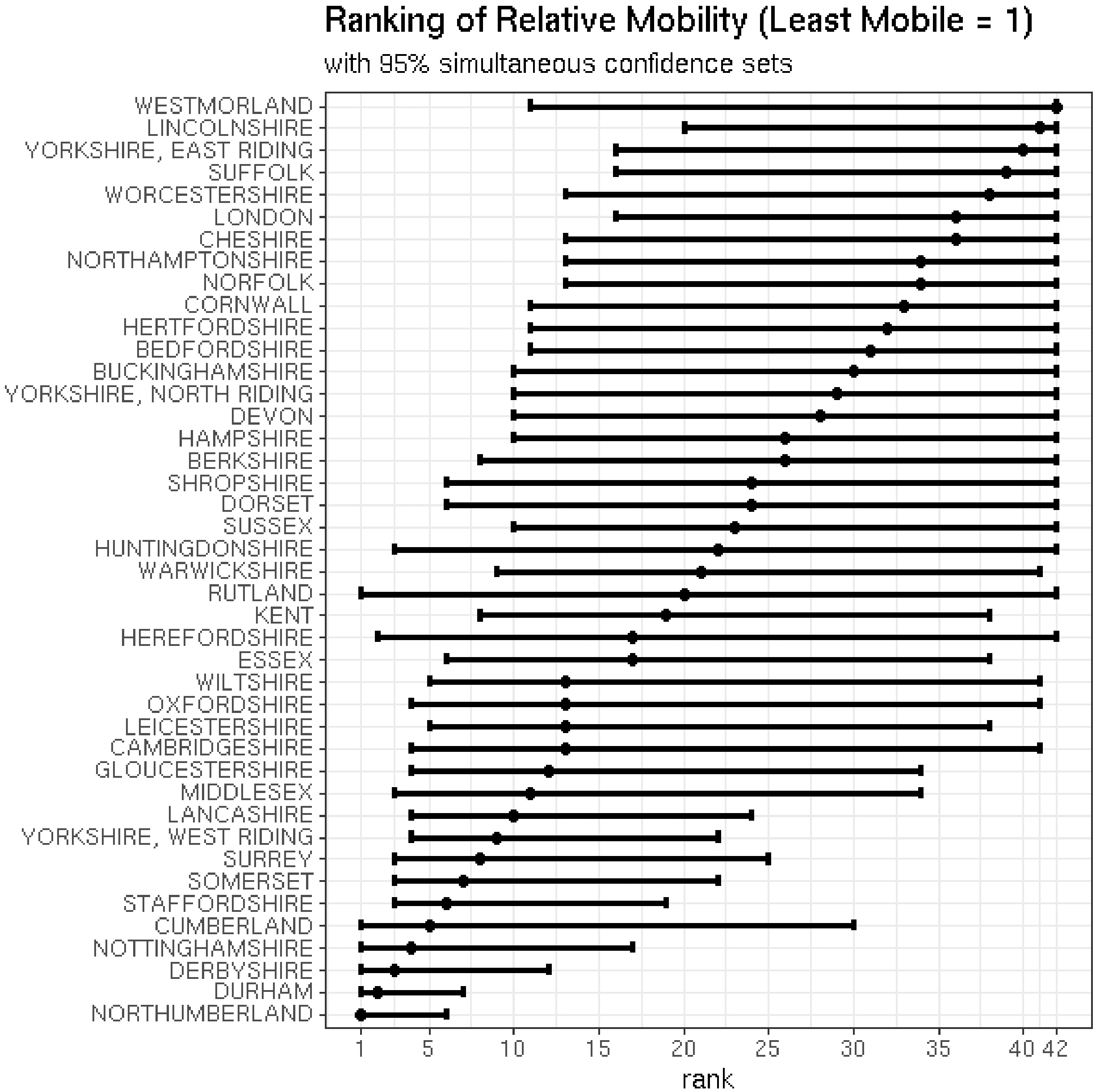
Sources: author's analysis of I-CeM (Schürer and Higgs (2014), UKDA, SN 7481) and I-CeM Names and Addresses (Schürer and Higgs (2015), UKDA, SN 7856). The 'CS Ranks' R package used to produce the results can be accessed at `https://cran.r-project.org/web/packages/csranks/index.html` and is based on the theory in Mogstad et al. (2024).

Figure C.D.4: Ranking of Regions by Absolute Mobility (IV)



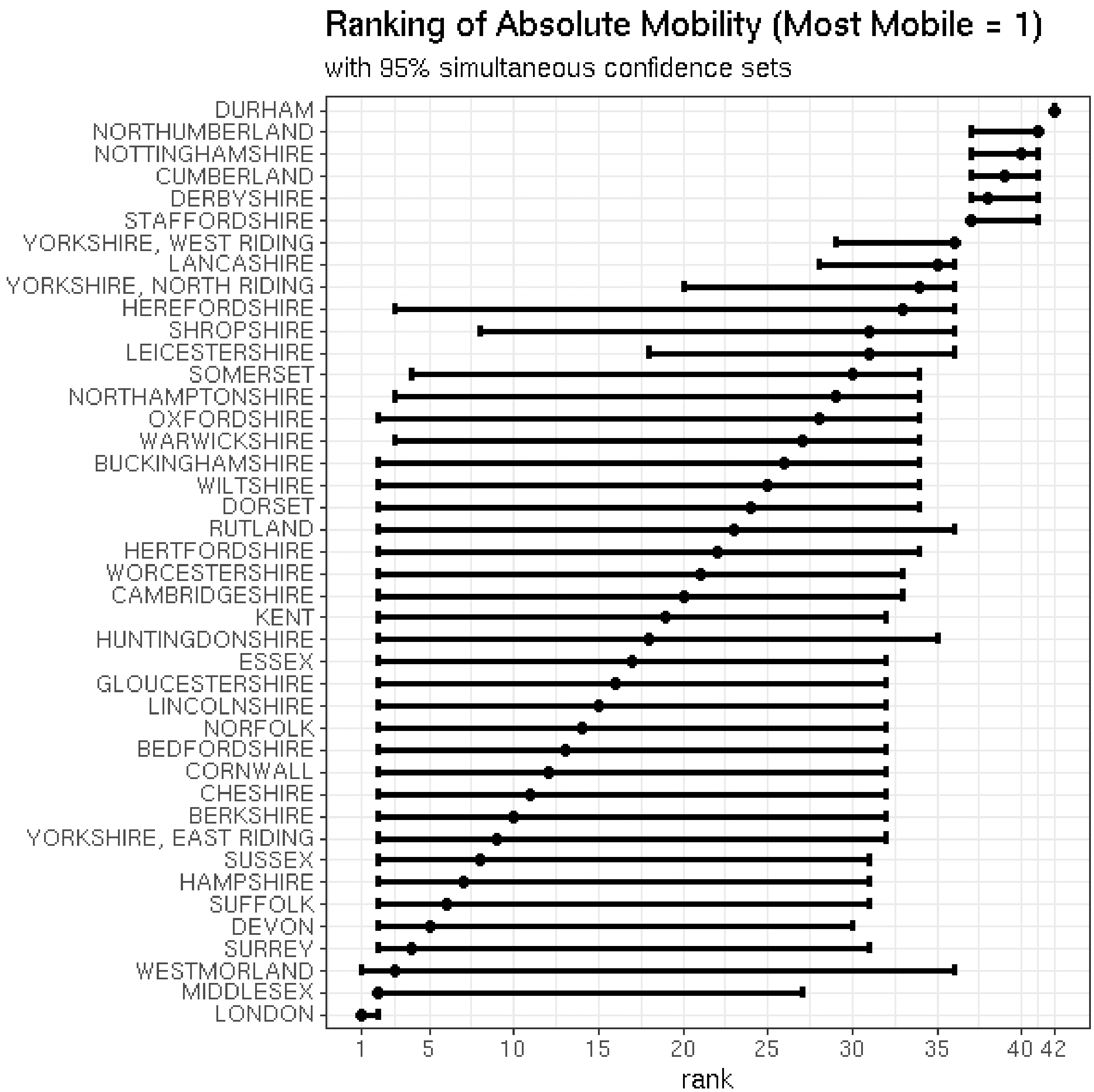
Sources: author's analysis of I-CeM (Schürer and Higgs (2014), UKDA, SN 7481) and I-CeM Names and Addresses (Schürer and Higgs (2015), UKDA, SN 7856). The 'CS Ranks' R package used to produce the results can be accessed at `https://cran.r-project.org/web/packages/csranks/index.html` and is based on the theory in Mogstad et al. (2024).

Figure C.D.5: Ranking of Counties by Relative Mobility (Baseline)



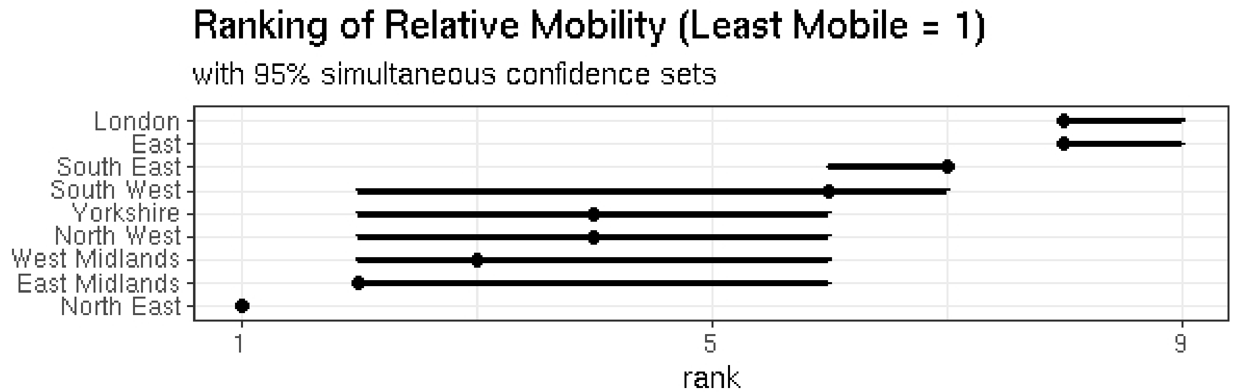
Sources: author's analysis of I-CeM (Schürer and Higgs (2014), UKDA, SN 7481) and I-CeM Names and Addresses (Schürer and Higgs (2015), UKDA, SN 7856). The 'CS Ranks' R package used to produce the results can be accessed at `https://cran.r-project.org/web/packages/csranks/index.html` and is based on the theory in Mogstad et al. (2024).

Figure C.D.6: Ranking of Counties by Absolute Mobility (Baseline)



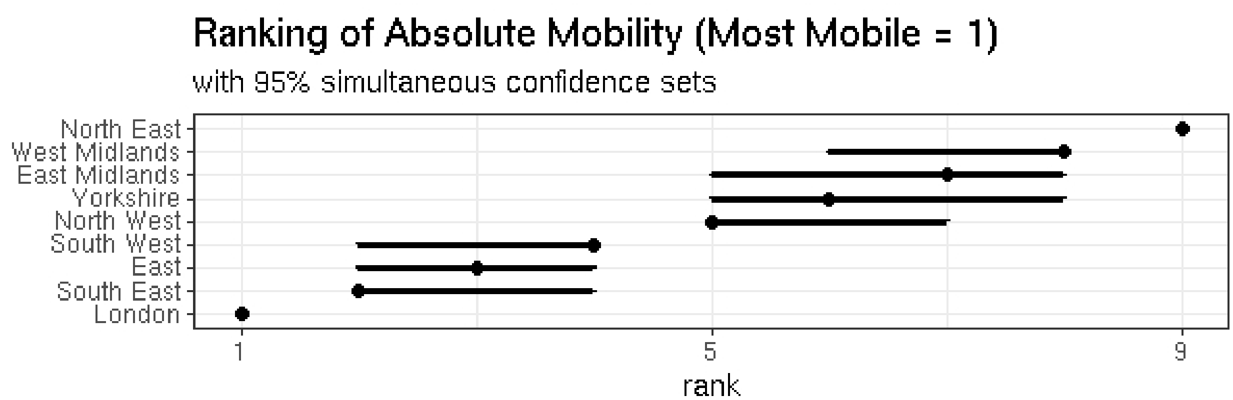
Sources: author's analysis of I-CeM (Schürer and Higgs (2014), UKDA, SN 7481) and I-CeM Names and Addresses (Schürer and Higgs (2015), UKDA, SN 7856). The 'CS Ranks' R package used to produce the results can be accessed at `https://cran.r-project.org/web/packages/csranks/index.html` and is based on the theory in Mogstad et al. (2024).

Figure C.D.7: Ranking of Regions by Relative Mobility (Baseline)



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). The 'CS Ranks' R package used to produce the results can be accessed at `https://cran.r-project.org/web/packages/csranks/index.html` and is based on the theory in Mogstad et al. (2024).

Figure C.D.8: Ranking of Regions by Absolute Mobility (Baseline)



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). The 'CS Ranks' R package used to produce the results can be accessed at `https://cran.r-project.org/web/packages/csranks/index.html` and is based on the theory in Mogstad et al. (2024).

C.E Great Gatsby Curve with Population-Level Inequality Data

An alternative measure of the level of inequality within a RSD is to use the full-count census of 1881 instead, and calculate the two measures of inequality using the total population of working age male (defined as age 15-64) who reported an occupation in the census. The two measures of inequality thus becomes the standard devation of occupational ranks among working men in a RSD, and the percentage share of occupations held by working men between rank 25 and 75 within a RSD. Table C.E.1 presents the results. While the strengths of the relationship between our measures of inequality and mobility differ with regards to the results in the main paper (table 4.4), the direction does not — they all suggest that more unequal places have lower levels of relative and absolute mobility.

Table C.E.1: Great Gatsby Curve (Inequality Calculated Using Full Census)

(1) Measure of Inequality (Log)	(2) Measure of Mobility (Log)	(3) β	(4) N
Std. Dev. of Occupational Ranks	Relative Mobility (Slope)	0.70*** (0.108)	1792
Percent of Rank 25-75	Relative Mobility (Slope)	-0.16** (0.063)	1792
Std. Dev. of Occupational Ranks	Absolute Mobility (Rank)	-0.05 (0.115)	1792
Percent of Rank 25-75	Absolute Mobility (Rank)	0.19*** (0.040)	1792

Notes: *** p<0.01, ** p<0.05, * p<0.1; log-transformed independent variables in column 1 and log-transformed dependent variables in column 2; standard errors clustered at the county-level, shown in parenthesis in column 3; number of observations in column 4 represent the number of RSDs used in this regression.

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).

C.F Returns to Migration with Rural-Urban Controls

A concern regarding the results on migration, particularly the large gains from migration southwards and the intergenerational mobility differences between North stayers and North-South migrants, may be that the findings are driven by rural-urban migration dynamics (i.e. the North-South migrants were more mobile because they were migrating from rural areas in the North to urban centres in the South).[2] Conversely, we would also need to make sure that the lack of returns to migration for migrating northwards was not due to Southern migrants moving from the cities to the countryside. This section shows the results of table 4.9 and 4.10 but with additional controls for rural-urban and urban-rural migration.

The classification of rural and urban parishes in 1881 and 1911 follows Smith and Bennett (2017). Using the Law-Robson framework for urban classification (Law, 1967; Robson, 1973), they assigned each parish a level between 1 and 4, with 1 being urban and 4 being rural, and levels 2 and 3 are parishes in transition to becoming urbanised. I then create a dummy variable for rural-urban migration which equals 1 if someone lived in a more urban parish in 1911 than they did in 1881, and vice versa for urban-rural migration. Table C.F.1 shows the returns to migration after adding controls for rural-urban and urban-rural migration. As the results demonstrate, the returns to migration for North-South migrants were not driven by rural and urban dynamics, and the size of the gains in occupational ranks are very similar to those shown in the main paper (table 4.9. At the same time, the low (or zero) returns to migration for Southern migrants to the North hold.

For intergenerational mobility, table C.F.2 shows that again, the results are very similar with or without controlling for rural-urban and urban-rural migration. Just as table 4.10 in the main results suggest, Northern migrants have higher relative mobility compared to Northern stayers within the same family, while there are no differences between Southern migrants and stayers. Therefore, rural-urban or urban-rural migration cannot explain the differences in mobility outcomes for stayers and migrants from the North and the South.

[2]For example, Long (2005) found positive effects for rural-urban migration in Victorian England.

Table C.F.1: Returns to Migration with Rural-Urban Controls (Baseline vs. Fixed Effects)

Dep. Var. =	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$Rank_{Son}$	Northern Origins				Southern Origins			
Migrant	23.93***	24.55***	8.13***	7.27***	3.68***	2.74*	2.14	-0.01
	(0.72)	(1.41)	(2.30)	(2.18)	(0.80)	(1.52)	(2.23)	(1.94)
Constant	49.61***	45.61***	54.19***	44.36***	52.00***	48.82***	40.89***	35.26***
	(1.39)	(2.50)	(13.01)	(15.15)	(1.11)	(1.95)	(12.14)	(12.25)
Age Controls	YES	YES	YES	YES	YES	YES	YES	YES
Rural-Urban Migration Controls	YES	YES	YES	YES	YES	YES	YES	YES
Only Include Brothers	NO	YES	YES	YES	NO	YES	YES	YES
At Least One Brother Moves/Stays	NO	NO	YES	YES	NO	NO	YES	YES
Family Fixed Effects	NO	NO	NO	YES	NO	NO	NO	YES
Observations	51,690	16,596	570	570	69,041	22,476	719	719
Households				267				334
R^2	0.016	0.016	0.025	0.056	0.000	0.000	0.013	0.010

Notes: 'North' defined as all counties in the regions of North East, North West, and Yorkshire and Humber, and 'South' defined as all counties in East of England, London, South East, and South West (see Table C.A.1 in Appendix C.A); 'North Stayers' defined as people who resided in the 'North' in 1881 and 1911; 'North Movers' defined as people who resided in the 'North' in 1881 but not in 1911; 'North to South' defined as people who resided in the 'North' in 1881 and in the 'South' in 1911; 'North of Severn-Wash Stayers' are defined as people who lived in counties north of the Severn-Wash Line in 1881 and 1911 while 'North of Severn-Wash Movers' are people who lived in counties north of the line in 1881 and south of the line in 1911 (Panel B of Figure 4.1 in the main paper shows the counties north and south of the Severn-Wash Line); same rules apply for those with 'Southern Origins'.
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), with urban classification of parishes from Smith and Bennett (2017).

Table C.F.2: Intergenerational Mobility of Brothers with Rural-Urban Controls

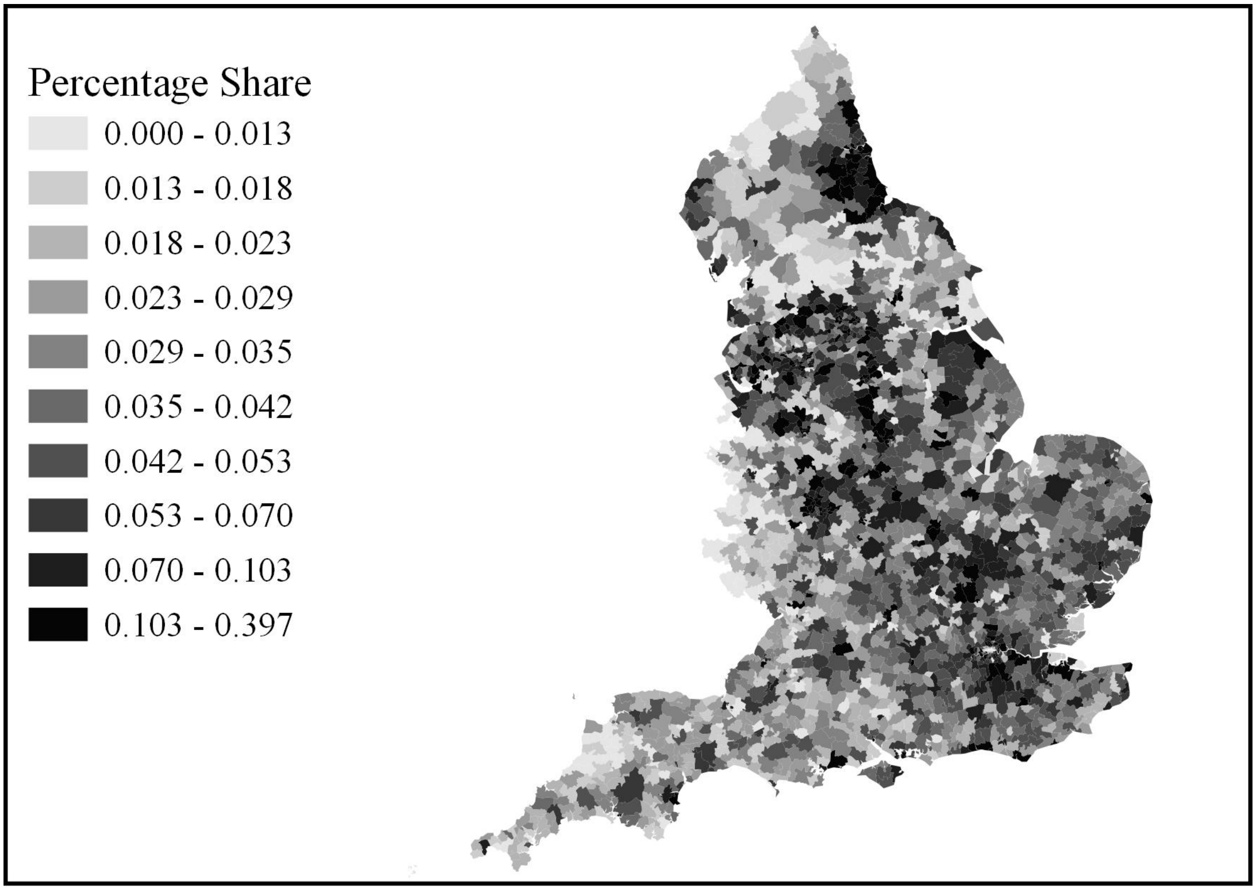
Dep. Var. = Son's Rank in 1911 VARIABLES	(1) Northern Origins	(2) Southern Origins
Migrant	25.12***	-2.534
	(7.443)	(5.370)
Instrumented Father's Rank in 1881 × Migrant	-0.263***	0.0509
	(0.100)	(0.0855)
Constant	47.15***	36.26***
	(15.27)	(12.40)
Age Controls	YES	YES
Rural-Urban Migration Controls	YES	YES
Brothers Sample	YES	YES
Family Fixed Effects	YES	YES
Observations	556	707
Number of Households	261	328

Notes: Clustered (by household) standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1; Father's Rank in 1881 instrumented by Father's Rank in 1891; column 1 shows results for brothers who lived in the North in 1881; columns 2 show results for brothers who lived in the South in 1881; Migrant status defined as North to South or South to North (excludes Midlands); reference group is the North (South) Stayer in column 1 (2); controls include age, age squared, a dummy variable for moving to a more urban parish, and a dummy variable for moving to a more rural parish; number of observations (N) represent each father-son pairs.

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), with urban classification of parishes from Smith and Bennett (2017).

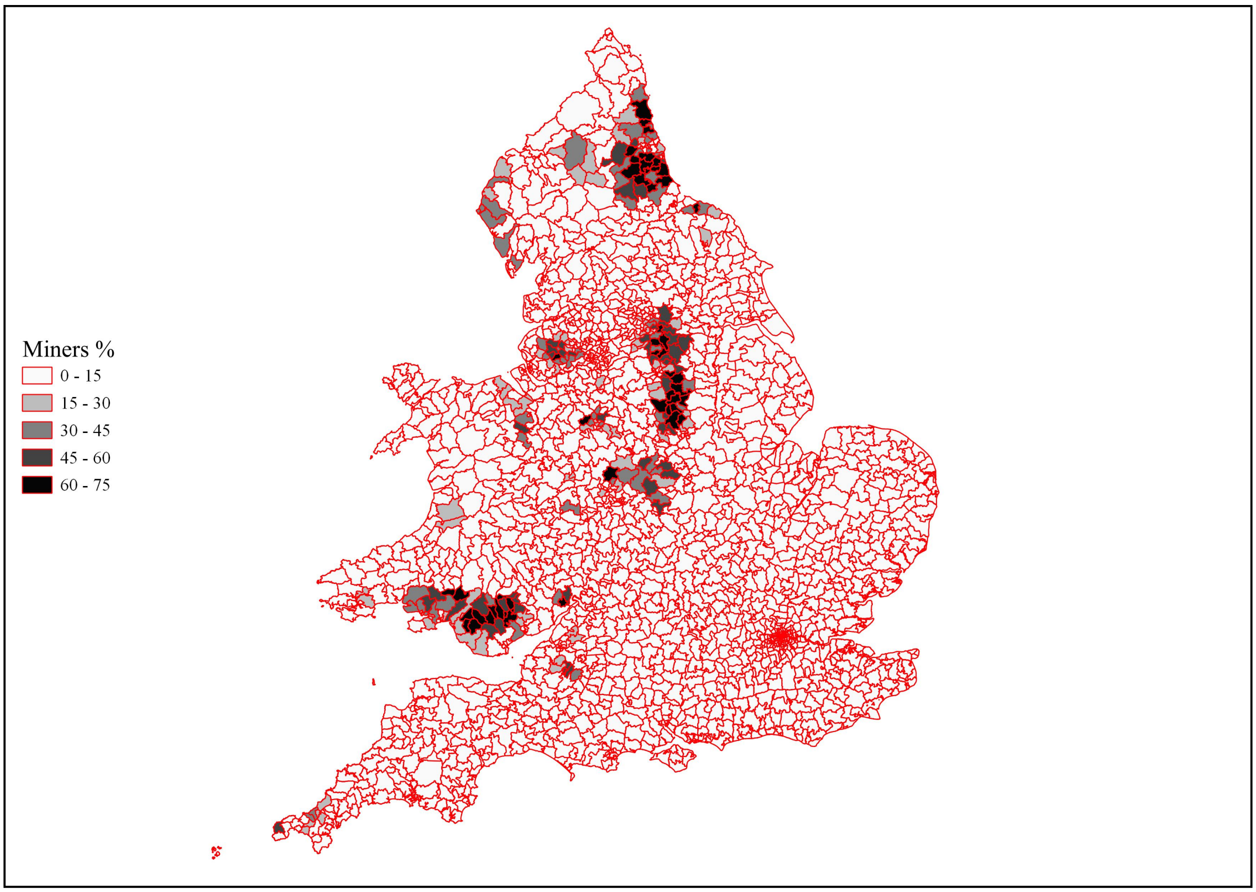
C.G Additional Results

Figure C.G.1: Sample Distribution across RSDs



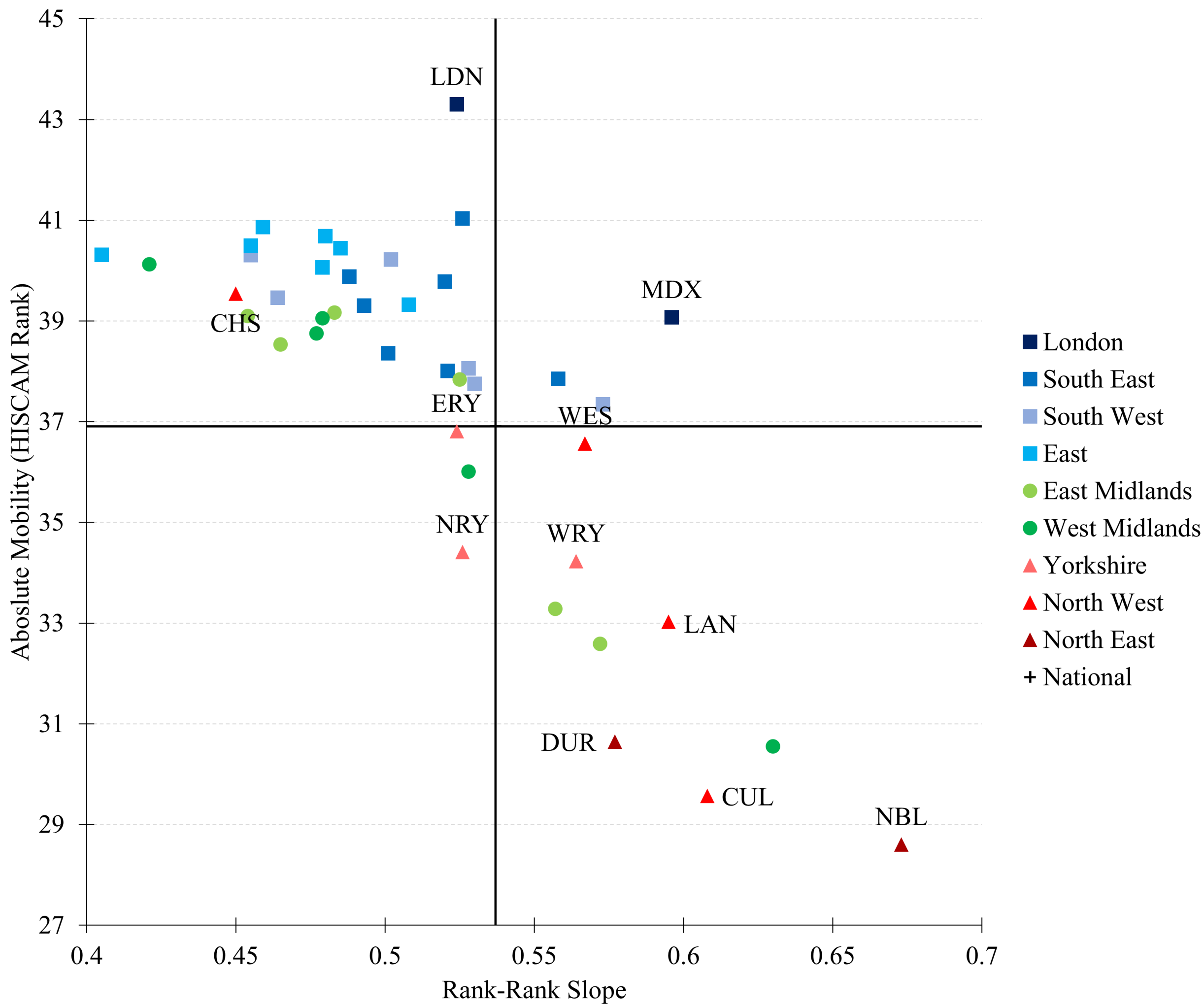
Sources: author's analysis of I-CeM (Schürer and Higgs (2014), UKDA, SN 7481) and I-CeM Names and Addresses (Schürer and Higgs (2015), UKDA, SN 7856). The GIS Shapefiles for 1881 Registration Sub-District comes from Day (2016).

Figure C.G.2: Density of Coal Miners, 1911



Sources: data and GIS shapefiles from Day (2016).

Figure C.G.3: County Mobility Pattern Without Miners, 1881–1911



Notes: full list of counties belonging to each regions in table C.A.1 in appendix C.A.
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).

Bibliography

Abramitzky, R. and Boustan, L. (2017), ‘Immigration in American Economic History’, *Journal of Economic Literature* **55**(4), 1311–45.

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, Working Paper 25825, National Bureau of Economic Research, Cambridge, Massachusetts.

Abramitzky, R., Boustan, L. P. and Eriksson, K. (2012), ‘Europe’s Tired, Poor, Huddled Masses: Self-Selection and Economic Outcomes in the Age of Mass Migration’, *American Economic Review* **102**(5), 1832–56.

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.

Acciari, P., Polo, A. and Violante, G. L. (2022), ‘And Yet It Moves: Intergenerational Mobility in Italy’, *American Economic Journal: Applied Economics* **14**, 118–163.

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

Alesina, A. and Angeletos, G.-M. (2005), ‘Fairness and Redistribution’, *American Economic Review* **95**(4), 960—-980.

Alesina, A., Stantcheva, S. and Teso, E. (2018), ‘Intergenerational Mobility and Preferences for Redistribution’, *American Economic Review* **108**(2), 521–54.

Allen, R. C. (2009), *The British Industrial Revolution in Global Perspective*, Cambridge University Press, Cambridge.

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, Working Paper 3724, National Bureau of Economic Research, Cambridge, Massachusetts.

Anbinder, T., Connor, D., Gráda, C. O. and Wegge, S. (2021), The Problem of False Positives in Automated Census Linking: Evidence from Nineteenth-Century New York's Irish Immigrants, Working Paper 568, CAGE Research Centre, Warwick.

Antonie, L., Inwood, K., Minns, C. and Summerfield, F. (2022), 'Intergenerational Mobility in a Mid-Atlantic Economy: Canada, 1871–1901', *Journal of Economic History* **82**, 1–27.

Arkell, T. (1994), 'Identifying the Census Enumerators – Cornwall in 1851', *Local Population Studies* **53**, 70–75.

Armstrong, W. (1972), The Use of Information about Occupation, *in* E. Wrigley, ed., 'Nineteenth-Century Society', Cambridge University Press, pp. 191–310.

Bachtler, J. (2004), Regional Disparities in the United Kingdom, *in* H. Karl and P. Rollet, eds, 'Employment and Regional Development Policy: Market Efficiency Versus Policy Intervention', Akademie für Raumforschung und Landesplanung, Hanover, pp. 36–49.

Bailey, M. J., Cole, C., Henderson, M. and Massey, C. (2020), 'How Well Do Automated Linking Methods Perform? Lessons from US Historical Data', *Journal of Economic Literature* **58**(4), 997–1044.

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, Cambridge, pp. 25–55.

Bambra, C., Munford, L., Khavandi, S. and Bennett, N. (2023), *Northern Exposure: COVID-19 and Regional Inequalities in Health and Wealth*, Policy Press, Bristol.

Becker, G. S. and Tomes, N. (1986), 'Human Capital and the Rise and Fall of Families', *Journal of Labor Economics* **4**, S1–S39.

Bell, B., Blundell, J. and Machin, S. (2023), 'Where is the Land of Hope and Glory? The Geography of Intergenerational Mobility in England and Wales', *Scandinavian Journal of Economics* **125**, 73–106.

Berger, T. (2018), 'Places of Persistence: Slavery and the Geography of Intergenerational Mobility in the United States', *Demography* **55**, 1547–1565.

Berman, Y. (2022), 'Absolute Intragenerational Mobility in the United States, 1962–2014', *Journal of Economic Inequality* **20**, 587–609.

Black, S. E. and Devereux, P. J. (2011), Recent Developments in Intergenerational Mobility, *in* D. Card and O. Ashenfelter, eds, 'Handbook of Labor Economics', Vol. 4, Elsevier, pp. 1487–1541.

Blackaby, D. H. and Murphy, P. D. (1995), 'Earnings, Unemployment and Britain's North-South Divide: Real or Imaginary?', *Oxford Bulletin of Economics and Statistics* **57**, 487–512.

Blanden, J. (2013), 'Cross-national Rankings of Intergenerational Mobility: A Ccomparison of Approaches from Economics and Sociology', *Journal of Economic Surveys* **27**(1), 38–73.

Blanden, J., Goodman, A., Gregg, P. and Machin, S. (2004), Changes in Intergenerational Mobility in Britain, *in* M. Corak, ed., 'Generational Income Mobility in North America and Europe', Cambridge University Press, p. 122–146.

Blau, P. M. and Duncan, O. D. (1967), *The American Occupational Structure*, Wiley.

Boberg-Fazlic, N. and Sharp, P. (2018), 'North and South: Long-Run Social Mobility in England and Attitudes Toward Welfare', *Cliometrica* **12**(2), 251–276.

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**.

Borjas, G. J. (1987), 'Self-Selection and the Earnings of Immigrants', *American Economic Review* **77**(4), 531–553.

Breen, R. and In, J. (2024), 'Regional Variation in Intergenerational Social Mobility in Britain', *British Journal of Sociology* **n/a**(n/a), 1–17.

Breen, R. and Jonsson, J. O. (2005), 'Inequality of Opportunity in Comparative Perspective: Recent Research on Educational Attainment and Social Mobility', *Annual Review of Sociology* **31**, 223–243.

Burkhauser, R. V. and Poupore, J. G. (1997), 'A Cross-National Comparison of Permanent Inequality in the United States and Germany', *Review of Economics and Statistics* **79**(1), 10–17.

Buscha, F., Gorman, E. and Sturgis, P. (2021), 'Spatial and Social Mobility in England and Wales: A Sub-National Analysis of Differences and Trends over Time', *British Journal of Sociology* **72**(5), 1378–1393.

Bütikofer, A., Dalla-Zuanna, A. and Salvanes, K. G. (2022), 'Breaking the Links: Natural Resource Booms and Intergenerational Mobility', *Review of Economics and Statistics* pp. 1–45.

Carneiro, P., Cattan, S., Dearden, L., Erve, L. V. D., Krutikova, S. and Macmillan, L. (2022), Intergenerational Income Mobility in England and the Importance of Education, Working paper 22/23, lnstitute for Fiscal Studies.

Cheng, S. and Song, X. (2019), 'Linked Lives, Linked Trajectories: Intergenerational Association of Intragenerational Income Mobility', *American Sociological Review* **84**(6), 1037–1068.

Chetty, R. and Hendren, N. (2018*a*), 'The Impacts of Neighborhoods on Intergenerational Mobility I: Childhood Exposure Effects', *Quarterly Journal of Economics* **133**, 1107–1162.

Chetty, R. and Hendren, N. (2018*b*), 'The Impacts of Neighborhoods on Intergenerational Mobility II: County-Level Estimates', *Quarterly Journal of Economics* **133**, 1163–1228.

Chetty, R., Hendren, N., Kline, P. and Saez, E. (2014), 'Where is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States', *Quarterly Journal of Economics* **129**, 1553–1624.

Chetverikov, D. and Wilhelm, D. (2024), Inference for Rank-Rank Regressions, cemmap working paper cwp11/24, Cemmap.

Clark, G. (2023), 'The Inheritance of Social Status: England, 1600 to 2022', *Proceedings of the National Academy of Sciences* **120**(27), 1–8.

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

Clark, G. and Cummins, N. (2018), The Big Sort: Selective Migration and the Decline of Northern England, 1780–2018, Discussion Paper 13023, Centre for Economic Policy Research, London.

Clark, G., Cummins, N. and Curtis, M. (2023), Measuring Mobility: Intergenerational Status Mobility Across Time and Space, Discussion Paper 17788, Centre for Economic Policy Research, London.

Collins, W. J. and Wanamaker, M. H. (2014), 'Selection and Economic Gains in the Great Migration of African Americans: New Evidence from Linked Census Data', *American Economic Journal: Applied Economics* **6**(1), 220–52.

Connor, D. S. and Storper, M. (2020), 'The Changing Geography of Social Mobility in the United States', *Proceedings of the National Academy of Sciences* **117**.

Corak, M. (2006), Do Poor Children Become Poor Adults? Lessons from a Cross Country Comparison of Generational Earnings Mobility, IZA Discussion Papers 1993, Institute of Labor Economics (IZA).

Corak, M. (2013), 'Income Inequality, Equality of Opportunity, and Intergenerational Mobility', *The Journal of Economic Perspectives* **27**(3), 79–102.

Corak, M. (2020), 'The Canadian Geography of Intergenerational Income Mobility', *Economic Journal* **130**, 2134–2174.

Dahl, M. and Deleire, T. (2008), The Association between Children's Earnings and Fathers' Lifetime Earnings: Estimates Using Administrative Data, Discussion Paper 1342–08, Institute for Research on Poverty, Madison, Wisconsin.

Day, J. (2018), 'Leaving Home in 19th Century England and Wales: A Spatial Analysis', *Demographic Research* **39**(4), 95–135.

Day, J. D. (2016), 'Registration sub-district boundaries for England and Wales 1851-1911', This dataset was created by the 'Atlas of Victorian Fertility Decline' project (PI: A.M. Reid) with funding from the ESRC (ES/L015463/1). The Day dataset has been created using Satchell, A.E.M., Kitson, P.M.K., Newton, G.H., Shaw-Taylor, L., and Wrigley E.A., 1851 England and Wales census parishes, townships and places (2016). The Satchell et al dataset is an enhanced version of Burton, N, Westwood J., and Carter P., GIS of the ancient parishes of England and Wales, 1500-1850. Colchester, Essex: UK Data Archive (May 2004), SN 4828 (in the UK data archive catalogue the dataset is now credited to H.R., Southall and N. Burton but the original citation is retained in the Satchell et al. reference due to uncertainty that the current version available may not be identical to the one used by Satchell et al.). SN 4828 is a GIS version of Kain, R.J.P., and Oliver, R.R., Historic parishes of England and Wales: An electronic map of boundaries before 1850 with a gazetteer and metadata. Colchester, Essex: UK Data Archive, May, 2001. SN 4348.

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.

Department for Levelling Up, Housing & Communities (2022), Levelling Up the United Kingdom, White paper, Department for Levelling Up, Housing & Communities.

Deutscher, N. and Mazumder, B. (2020), 'Intergenerational Mobility across Australia and the Stability of Regional Estimates', *Labour Economics* **66**.

Deutscher, N. and Mazumder, B. (2023), 'Measuring Intergenerational Income Mobility: A Synthesis of Approaches', *Journal of Economic Literature* **61**(3), 988–1036.

Dickens, C. (1861), *Great Expectations*, Chapman and Hall, London.

DiPrete, T. A. and McManus, P. A. (1996), 'Institutions, Technical Change, and Diverging Life Chances: Earnings Mobility in the United States and Germany', *American Journal of Sociology* **102**(1), 34–79.

Eriksen, J. and Munk, M. D. (2020), 'The Geography of Intergenerational Mobility — Danish evidence', *Economics Letters* **189**.

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

European Commission (2022), Statistical Regions in the European Union and Partner Countries NUTS and Statistical Regions 2021 — Re-Edition 2022, Manuals and guidelines, Eurostat.

Featherman, D. L., Lancaster, F. J. and Hauser, R. M. (1975), 'Assumptions of Social Mobility Research in the U.S.: The Case of Occupational Status', *Social Science Research* **4**, 329–360.

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.

Friedman, S. and Macmillan, L. (2017), 'Is London Really the Engine-Room? Migration, Opportunity Hoarding and Regional Social Mobility in the UK', *National Institute Economic Review* **240**, R58–R72.

Fuller, S. (2008), 'Job Mobility and Wage Trajectories for Men and Women in the United States', *American Sociological Review* **73**(1), 158–183.

Fuller, W. A. (1987), *Measurement Error Models*, John Wiley & Sons, New York.

Geary, F. and Stark, T. (2018), 150 Years of Regional GDP: United Kingdom and Ireland, *in* J. R. Rosés and N. Wolf, eds, 'Economic Development of Europe's Regions : A Quantitative History Since 1900', Taylor & Francis Group, Florence, pp. 330–362.

Gibbons, S., Overman, H. G. and Pelkonen, P. (2010), Wage Disparities in Britain: People or Place?, SERC Discussion Paper 60, Centre for Economic Performance, London.

Goldthorpe, J. H. (1987), *Social Mobility and Class Structure in Modern Britain*, 2nd edn, Clarendon Press, Oxford.

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, Cambridge, pp. 58–88.

Gregg, P., Macmillan, L. and Vittori, C. (2017), 'Moving Towards Estimating Sons' Lifetime Intergenerational Economic Mobility in the UK', *Oxford Bulletin of Economics and Statistics* **79**(1), 79–100.

Gulgun Bayaz-Oturk, R. V. B. and Couch, K. A. (2014), 'Consolidating the Evidence on Income Mobility in the Western States of Germany and the United States from 1984 to 2006', *Economic Inquiry* **52**(1), 431–443.

Güell, M., Pellizzari, M., Pica, G. and Rodríguez Mora, J. V. (2018), 'Correlating Social Mobility and Economic Outcomes', *The Economic Journal* **128**(612), F353–F403.

Haider, S. and Solon, G. (2006), 'Life-Cycle Variation in the Association between Current and Lifetime Earnings', *American Economic Review* **96**, 1308–1320.

Heidrich, S. (2017), 'Intergenerational Mobility in Sweden: A Regional Perspective', *Journal of Population Economics* **30**, 1241–1280.

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

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, London.

Higgs, E., Jones, C., Schürer, K. and Wilkinson, A. (2013), Integrated Census Microdata (I-CeM) Guide, User guide, University of Essex, Colchester.

Hudson, P. (2009), *The Industrial Revolution*, Bloomsbury Publishing Plc, London.

Humphries, J. (2010), *Childhood and Child Labour in the British Industrial Revolution*, Cambridge University Press, Cambridge.

Hunt, E. H. (1986), 'Industrialization and Regional Inequality: Wages in Britain, 1760–1914', *Journal of Economic History* **46**, 935–966.

Jarvis, B. F. and Song, X. (2017), 'Rising Intragenerational Occupational Mobility in the United States, 1969 to 2011', *American Sociological Review* **82**(3), 568–599.

Jäntti, M. and Jenkins, S. P. (2015), Chapter 10 — Income Mobility, *in* A. B. Atkinson and F. Bourguignon, eds, 'Handbook of Income Distribution', Vol. 2, Elsevier, pp. 807–935.

Kalleberg, A. L. and Mouw, T. (2018), 'Occupations, Organizations, and Intragenerational Career Mobility', *Annual Review of Sociology* **44**, 283–303.

Kambourov, G. and Manovskii, I. (2008), 'Rising Occupational and Industry Mobility in the United States: 1968–97', *International Economic Review* **49**(1), 41–79.

Kambourov, G. and Manovskii, I. (2009), 'Occupational Mobility and Wage Inequality', *Review of Economic Studies* **76**(2), 731–759.

Kopczuk, W., Saez, E. and Song, J. (2010), 'Earnings Inequality and Mobility in the United States: Evidence from Social Security Data Since 1937', *Quarterly Journal of Economics* **125**(1), 91–128.

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.

Law, C. M. (1967), 'The Growth of Urban Population in England and Wales, 1801–1911', *Transactions of the Institute of British Geographers* (41), 125–143.

Lee, E. S. (1966), 'A Theory of Migration', *Demography* **3**(1), 47–57.

Leeuwen, M. H. V. and Maas, I. (2010), 'Historical Studies of Social Mobility and Stratification', *Annual Review of Sociology* **36**, 429–451.

Lipset, S. M. and Zetterberg, H. (1959), Social Mobility in Industrial Societies, *in* S. M. Lipset and R. Bendix, eds, 'Social Mobility in Industrial Society', University of California Press, Berkeley.

Long, J. (2005), 'Rural-Urban Migration and Socioeconomic Mobility in Victorian Britain', *Journal of Economic History* **65**, 1–35.

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.

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

Massey, C. G. (2017), 'Playing with Matches: An Assessment of Accuracy in Linked Historical Data', *Historical Methods: A Journal of Quantitative and Interdisciplinary History* **50**(3), 129–143.

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, Manchester, pp. 18–39.

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

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, Manchester, 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.

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

Modalsli, J. H. and Vosters, K. (2019), Spillover Bias in Multigenerational Income Regressions Standard, Discussion Paper 897, Statistics Norway, Research Department, Oslo.

Mogstad, M., Romano, J. P., Shaikh, A. M. and Wilhelm, D. (2024), 'Inference for Ranks with Applications to Mobility across Neighbourhoods and Academic Achievement across Countries', *Review of Economic Studies* **91**(1), 476–518.

Mouw, T. and Kalleberg, A. L. (2010*a*), 'Do Changes in Job Mobility Explain the Growth of Wage Inequality among Men in the United States, 1977–2005?', *Social Forces* **88**(5), 2053–2077.

Mouw, T. and Kalleberg, A. L. (2010*b*), 'Occupations and the Structure of Wage Inequality in the United States, 1980s to 2000s', *American Sociological Review* **75**(3), 402–431.

Nybom, M. and Stuhler, J. (2017), 'Biases in Standard Measures of Intergenerational Income Dependence', *Journal of Human Resources* **52**, 800–825.

Nybom, M. and Stuhler, J. (2024), 'Interpreting Trends in Intergenerational Mobility', *Journal of Political Economy* **132**(8), 2531–2570.

Piketty, T. (1995), 'Social Mobility and Redistributive Politics', *Quarterly Journal of Economics* **110**(3), 551–584.

Piketty, T. and Saez, E. (2014), 'Inequality in the Long Run', *Science* **344**(6186), 838–843.

Pérez, S. (2019), 'Intergenerational Occupational Mobility across Three Continents', *Journal of Economic History* **79**, 383–416.

Robson, B. (1973), *Urban Growth: An Approach*, Methuen, London.

Rohenkohl, B. (2019), Intergenerational Income Mobility in the UK: New Evidence using the BHPS and Understanding Society, Sheffield Economic Research Paper Series 2019017, University of Sheffield, Sheffield.

Roy, A. D. (1951), 'Some Thoughts on the Distribution of Earnings', *Oxford Economic Papers* **3**(2), 135–146.

Ruggles, S., Fitch, C. A. and Roberts, E. (2018), 'Historical Census Record Linkage', *Annual Review of Sociology* **44**, 19–37.

Satchell, M., Shaw-Taylor, L., Wrigley, E., Kitson, P. and Newton, G. (2023), '1851 England and Wales Ancient Counties', Distributed by UK Data Service. http://doi.org/10.5255/UKDA-SN-852942.

Savage, M. (1993), Career Mobility and Class Formation: British Banking Workers and the Lower Middle Class, *in* A. Miles and D. Vincent, eds, 'Building European Society: Occupational Change and Social Mobility in Europe, 1840–1940', Manchester University Press, Manchester, pp. 196–216.

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.

Shaw-Taylor, L. and Wrigley, E. A. (2014), Occupational Structure and Population Change, *in* R. Floud, J. Humphries and P. Johnson, eds, 'The Cambridge Economic History of Modern Britain', Cambridge University Press, pp. 53–88.

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

Smith, H. and Bennett, R. (2017), 'Urban-Rural Classification using Census data, 1851-1911', Apollo — University of Cambridge Repository. https://doi.org/10.17863/CAM.15763.

Social Mobility Commission (2023), State of the Nation 2023 — People and Places, Policy paper, Social Mobility Commission.

Solon, G. (1992), 'Intergenerational Income Mobility in the United States', *American Economic Review* **82**, 393–408.

Solon, G. (1999), Chapter 29 — Intergenerational Mobility in the Labor Market, *in* O. C. Ashenfelter and D. Card, eds, 'Handbook of Labor Economics', Vol. 3, Elsevier, pp. 1761–1800.

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.

Tan, H. R. (2023), 'A Different Land of Opportunity: The Geography of Intergenerational Mobility in the Early Twentieth-Century United States', *Journal of Labor Economics* **41**, 77–102.

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, Cambridge, 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, London, pp. 82–133.

Tillott, P. M. (1968), 'The Analysis of Census Returns', *Local Historian* **8**, 2–9.

Treiman, D. J. (1970), 'Industrialization and Social Stratification', *Sociological Inquiry* **40**, 207–234.

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

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

Vincent, D. (1993), Mobility, Bureaucracy and Careers in Early-Twentieth-Century Britain, *in* A. Miles and D. Vincent, eds, 'Building European Society: Occupational Change and Social Mobility in Europe, 1840–1940', Manchester University Press, Manchester, pp. 18–39.

Ward, Z. (2019), Internal Migration, Education and Upward Rank Mobility:Evidence from American History, CEH Discussion Papers 04, Centre for Economic History, Research School of Economics, Australian National University.

Ward, Z. (2023), 'Intergenerational Mobility in American History: Accounting for Race and Measurement Error', *American Economic Review* **113**, 3213–3248.

Woods, R. and Hinde, P. A. (1987), 'Mortality in Victorian England: Models and Patterns', *Journal of Interdisciplinary History* **18**(1), 27–54.

Wrigley, E. (2010*a*), *Energy and the English Industrial Revolution*, Cambridge University Press, Cambridge.

Wrigley, E. A. (2010*b*), The PST System of Classifying Occupations, Technical report, Cambridge Group for the History of Population and Social Structure, Cambridge.

Yu, Y. (2023), 'Spatial and Social Mobility in England: The Persistence of the 'Escalator' Effect for the Millennial Generation', *Population, Space and Place* **29**(3), 1–13.

Yu, Y. and Athey, T. (2023), 'Comparing the Social and Spatial Mobility across UK Regions —- Evidence from the 1958 and 1970 Birth Cohorts', *Research in Social Stratification and Mobility* **83**(100767), 1–11.

Zhu, Z. (2024), 'Like Father Like Son? Intergenerational Immobility in England, 1851–1911', *Journal of Economic History* **84**(4), 1143–1173.

Zhu, Z. (2025), Rags to Rags, Riches to Riches: Essays on Occupational Mobility in England, 1851–1911, Phd thesis, London School of Economics, London.

Zimmerman, D. J. (1992), 'Regression Toward Mediocrity in Economic Stature', *American Economic Review* **82**, 409–429.