



THE LONDON SCHOOL  
OF ECONOMICS AND  
POLITICAL SCIENCE ■

Economic History Working Papers

No: 369

---

# Ethnic Wealth Inequality in England and Wales, 1858-2018

Neil Cummins, LSE

August 2024

# Ethnic Wealth Inequality in England and Wales, 1858-2018

Neil Cummins\*

August 19, 2024

## Abstract

Using surnames from the universe of death and wealth-at-death records in England and Wales, from 1858 to 2018, I document the emergence of a modern ethnic wealth gradient. Historically, Non-British ethnicities have average wealth 2-5 times that of the English. However, this premium has decreased over the 20th century. By 1980, non-British ethnicities have no advantage over the British. However, this masks considerable heterogeneity within the non-British ethnicity group. Europeans typically die significantly richer than the English whereas the Pakistani and Swedish die significantly poorer. Some groups always have lower wealth. The Irish, have wealth around 50% of the average English throughout. Surprisingly, the most egalitarian measure of wealth is representation within the top 1%. Most ethnicities have an equal, or greater, representation in the top 1% than the English, 1980-1992. Despite large differences in average wealth between ethnicities, the vast majority of variation, 97.5% is between individuals.

JEL: N00, N33, N34

## 1 Introduction

In the past decade, the study of inequalities of wealth and income, on a global scale with an emphasis on history, have become hugely influential, igniting much academic and public discourse. This recent wave was initiated by Piketty (2014), and has been complemented by Milanovic (2016), Scheidel (2018), and Alfani (2023), amongst many others. There is insatiable interest for explaining inequalities. However despite this, the relationship of ethnic background to economic outcomes, has been relatively ignored by the literature. This is because there is very little systematically collected data that exist which allow us to document ethnic wealth inequalities nor describe their development over time. This paper is an attempt to rectify this, for England and Wales, from 1858 to 2018.

The ethnic composition of a country is a function of its migration and colonial history. Historically, the British Isles were first populated by successive waves of migrant populations, including the people known to us as Celts, Anglo-Saxons and Vikings. The Norman Conquest of 1066 imposed a foreign, French speaking aristocracy.<sup>1</sup> The urban colossus of London attracted economic migrants from all over the World, many escaping persecution (for example the Huguenots and Jews) and later the arrival of masses of Germans, Italians and others seeking fortune from the World's economic capital. The largest migrant group from outside the British

---

\*Economic history, LSE and CEPR. Email: n.j.cummins@lse.ac.uk, neilcummins.com. Thanks to Elisa Jácome and to the participants of the “Explorations in Economic History Special Issue Workshop on Wealth and Income Inequality around the World” at the Federal Reserve Bank of Chicago on October 6-7, 2023. The analysis code and underlying data are available at <https://www.openicpsr.org/openicpsr/project/208142/version/V1/view>. This paper is forthcoming at *Explorations in Economic History*.

<sup>1</sup>This elite still outperforms the English average in terms of educational attainment and wealth, a millennium after the conquest (Clark and Cummins (2014)).

mainland has always been the Irish. The modern era of migration began with reconstruction following the second World War. In particular large inflows of people from former British colonies (and subsequently commonwealth nations), such as India, Pakistan and Jamaica established, and added to existing, communities. Migration from Europe and as always from Scotland and Ireland continued. Rates of migration are historically far higher after 1950 than before.

Contemporary data reveal substantial ethnic group inequities in income and wealth in England. For example, the Pakistani, Black, Bangladeshi and Indian ethnic groups have household wealth somewhere between 11% to 70% that of the White British ethnic group.<sup>2</sup> Despite this, the empirical characterization of economic inequality over the 20th century has almost exclusively focused upon top income and wealth shares (e.g. Atkinson and Harrison (1974); Piketty (2014)). For England, we know that the top 1% held over 73% of all wealth in 1900, and that this share declined to under 20% by 1992 (Cummins (2019b)). However we know nothing about history of wealth inequality stratified by ethnic origin.

The wealth data I use comes from a complete digitization of the *Principal Probate Registry (PPR) Calendars*, 1858-1992. This source records the estate value of all those dying over a nominal threshold of wealth (£10 in 1900, £5,000 in 1990). The *PPR Calendars* represent the most comprehensive, population-wide source of consistently collected wealth-at-death estimates for England over the 20th century. All of the entries record the full name and surname of the wealth holder.

I assign a probable ethnicity to all surnames in these records using the Onomap name classification software.<sup>3</sup> This assignment is based upon global data such as telephone directories and electoral rolls. Network analysis reveals naming clusters that map onto known ethnicities (Mateos et al. (2011)). Onomap applied to the wealth data I have collected, provides a categorization of surnames into almost 140 distinct ethnicities. This is far greater than the 11 ethnicities typically reported by the ONS (Office for National Statistics (2019b, 2020)). The assignment is not perfect and the interpretation must acknowledge that surnames are nominal, changeable ‘labels’ that will only imperfectly predict ethnicity. However, crosschecks of the Onomap assignment with ‘self identified’ ethnicity suggest a high degree of concordance (Kandt and Longley (2018)).

To the wealth data I then add the records of all these dying with wealth below the probate threshold. I have constructed a 100% sample of the Death register index for England and Wales, from 1838 to 2007. These death records also give the full name, as well as the age of the deceased. I can then also infer ethnicity from the surnames in the death records, which allows me to measure the proportion probated of ethnicities, as well as average wealth, 1858-1992. The proportion probated captures the proportion of an ethnicity who have wealth above a certain minimum probate threshold. This was £10 in 1900, and was raised periodically over the 20th century to £5,000 today.<sup>4</sup> For the period 1996 to 2018, I have collected the index to the probate registers (however the estate values are not reported during this period), so I can also calculate a probate rate for this more recent period.

This paper has 5 main results.

I first document the ethnic composition of England and Wales from 1838 to 2007. Complete birth and death registers show that Non-British ethnicities are a stable 3% of all births from, 1838-1950. After 1950, this proportion rises to 25% by 2007. The death registers show lower proportions as many post 1950 migrants are still alive.

Second, I document the history of wealth inequality by ethnicity. From 1858 to 1980, Non-British ethnicities are on average an elite group with wealth 200 to 500% that of the English.

---

<sup>2</sup>These estimates are for 2018 (Office for National Statistics (2020)). I reproduce the central findings in appendix figure A20. In 2019 the Office for National Statistics (ONS) presented the “first analysis” of ethnicity pay gaps using data from the *Annual Population Survey*. Large inequities are clear; “on average, employees from the Chinese ethnic group earned 30.9% more than White British employees; while employees from the Bangladeshi ethnic group, on average, earned 20.2% less than White British employees.” (Office for National Statistics (2019b)).

<sup>3</sup><http://www.onomap.org/>. Onomap was kindly provided to me by Paul Longley and Oliver O’Brien (Department of Geography, University College London).

<sup>4</sup>However, for the period after 2000, the minimum probate threshold level was very loosely applied, and was often at the discretion of the deceased’s bank. See appendix section A for a discussion of this.

This is reflected both in the proportion with significant wealth (to merit probate), and in average wealth. However, post 1980, this wealth differential disappears, and wealth is equalized between the English, and those with non British ethnicities. This equality masks significant variation. The Polish, Pakistani, Black Caribbean, Indian Hindi, Bangladeshi, and Swedish have significantly lower wealth than the English. The German, Dutch, Norwegian and names of Jewish origin are significantly richer than the English. This modern ethnic wealth gradient emerged only after 1980.

Thirdly, I show that the most ethnically egalitarian component of the English wealth structure is membership to the top 1% of wealth holders. Here most ethnicities have a greater probability of appearing than the English, 1858 to 1992.

Fourthly, I show that some ethnic groups always have lower wealth. The Bangladeshi group, for example, after reporting higher average wealth 1858-1913, have lower average wealth 1914 to 1992. The largest non British ethnic group, the Irish, have consistently lower wealth, and lower representation in the top 1%, throughout.

For robustness I present ethnic wealth estimates using an alternative surname classifier. Here I use the distribution of places of birth for 36m individuals from the 1911 census of England and Wales. This provides a classification of surnames, by ethnicity, from a century ago. Contemporary classifiers, such as Onomap, may misclassify certain families who integrate so well into the English, for example, that their naming patterns are indistinguishable from that of the English. There is evidence that this effect is present. The estimates of Irish wealth are significantly higher using the 1911 method than Onomap. But the relative patterns remain consistent. All other ethnicities examined in this way are entirely consistent with Onomap.

Finally, I present a Theil Index decomposition of inequality into an individual component and a between ethnic group component. Despite the large differences in the proportions probated and in the levels of average wealth between ethnic groups, the vast majority of variation, 97.5% is between individuals.

The descriptive patterns presented here are completely novel and present the first attempt to establish an empirical base for the history of ethnic inequality in England. They will provide a basis for an informed discussion of the role of migration, the selectivity of migration, human capital and discrimination in the distribution of wealth over the past century and a half. The modern ethnic wealth gradient is of relatively recent emergence and understanding the forces that created it are of first order importance for both academic economics, and society as a whole.

This is the first study of the history of ethnic wealth inequality in England and Wales.<sup>5</sup> For the US Chetty et al. (2019) have recently analyzed ethnic differences in economic opportunity and Kuhn et al. (2020) have examined ethnic inequality in income and wealth. There does not exist any comparable research of this type for England and Wales.

Within the broader literature (e.g. Piketty (2014); Milanovic (2016); Scheidel (2018); Alfani (2023)), this paper underlines the importance of considering the categorical dimension. Simple tabulations of top percentile shares, and other summary measures of the distribution such as the Gini coefficient, are unable to consider ethnicity. Here I show that there are large differences in wealth by ethnic background that are completely invisible in such summary measures. Further, during the period when wealth inequality was massively declining in England, as measured by the wealth-share of the top 1% (Atkinson and Harrison (1978); Cummins (2021)), ethnic inequalities in wealth were rising significantly. Thus individual, and ethnic, inequality moved in opposite directions over the 20th century. This is a surprising and unexpected result given the equalizing economic growth and tax policies of the post war period, to the 1980s which are argued by Piketty (2014) to have reduced all inequality. Further, the finding that the top 1% is more ethnically equal than the rest of the distribution is also surprising and unexpected given

---

<sup>5</sup>There is a broad and deep economics literature on migration. For the UK I collect here some of the most relevant to the current analysis, empirical results; on the effects of immigration on local labour markets; Manacorda et al. (2012), on fiscal contribution: Dustmann and Frattini (2014), on housing prices; Sá (2014), on identity; Battu and Zenou (2010) and Casey and Dustmann (2010), on attitudes towards migrant groups Dustmann and Preston (2001, 2006). There is also a literature on ethnic inequality in healthcare; on use of health services; Morris et al. (2005). Internationally, “wealth is understudied in the ethnic and migration studies literature” Vallejo and Keister (2019, p.2).

the high degree of wealth persistence documented for the English by Clark and Cummins (2014, 2015a). The dynamics of ethnic wealth inequality, its change over time, and understanding the causal forces behind these differences are crucial for effective social policies. The present study gives a preliminary picture of the set of stylized facts that we can build upon in that journey.

The data underlying the analysis are presented in section 2, the methodology in section 3 which describes in detail the process and accuracy of the surname ethnicity assignment, and the construction of the three wealth measures. Section 4 presents the results: ‘A Quantitative History of the Ethnic Composition of England’ (section 4.1), ‘Who Owns England?’ (section 4.2). The results on ‘Ethnic Wealth Inequality’ are presented in three parts corresponding to the wealth measure used. Firstly, the probate rate, 1858-2018 (section 4.3.1), average wealth 1858-1992 (section 4.3.2) and then representation in the top 1% of wealth holders 1858-1992 (section 4.3.3). Section 5 reports results using the alternative, 1911 Census based, ethnic classification of surnames. I present a decomposition of inequality using Theil’s Entropy measure in section 6, a discussion of the results and their interpretation in section 7, while section 8 concludes.

## 2 Data

This paper describes the history of ethnic wealth inequality using the universe of individual, nominal, death and probate records from England and Wales, 1858 to 1992, and 1996 to 2018. I use 16 million individual probate records combined with 86 million individual death records, 1838 to 1992. This is complemented with a separate dataset that reports for 8 million individual decedents whether or not they held significant wealth, for the period 1996 to 2018.

### 2.1 Wealth

I use a 100% transcription of all individual level wealth-at-death records from the the *Principal Probate Registry (PPR) Calendar* entries, 1858-1992. The PPR calendar records all decedents in England and Wales with wealth above the probate threshold (currently £5,000).<sup>6</sup>

The evidence from existing studies strongly support the credibility of the PPR Calendar wealth data.<sup>7</sup> Those valuations are directly employed here. However, after 1980 there was a change in the system for valuing probates in the PPR Calendars. As opposed to an exact valuation, which was the practice 1892-1979, a large proportion of valuations appear as banded estimations. These are £25,000, £40,000, £70,000, £100,000, £115,000 and £125,000, with each entry listed as “Not Exceeding” the banded amount. Inspecting the distribution of wealth between these bands in 1980, before they were applied in 1981, it is evident that they are either skewed, or uniform, and in all cases non-normal. Therefore I extracted the characteristics of the 1980 wealth distribution between these later applied bands, and then used those distributional characteristics to randomly assign a wealth to every banded observation post 1980. This was done for 1,521,608 observations out of a total of 2,506,371, 1981-1992 inclusive.<sup>8</sup>

Wealth is taken as the probated estate value at death. There is no doubt that the relationship of this *declared* wealth-at-death to lifetime wealth is changing over time. One reason for this are the large changes in taxes due on wealth-at-death over the sample period. Testators, while living, can rearrange portfolios so that this tax can be minimized. Trusts, inter-vivos bequests,

---

<sup>6</sup>The probate threshold during the period 1858-1900 was £10, 1901-1931: £50, 1932-1964: £100, 1965-1974: £500, 1975-1984: £1,500, 1984 onwards: £5,000 Cummins (2021, table 1). Existing research directly using the individual probate valuations includes Wedgwood (1928), Harbury (1962), Perkin (1978), Rubinstein (1977a,b, 1981) Nicholas (1999), Rothery (2007), Turner (2010), Clark and Cummins (2015a) and (2015b), Cummins (2021, 2022). The original printed volumes were digitized and algorithmically parsed and formed into a database suitable for economic analysis, in a process described in Cummins (2021), who investigates in depth the quality of the resulting data.

<sup>7</sup>The wealth-share estimates of the top percentiles from the PPR Calendars match closely existing estimates from Atkinson and Harrison (1978); Atkinson et al. (1989); Atkinson (2013) and Alvaredo et al. (2018), who use aggregated returns reported by the Inland Revenue. The PPR Calendar data also matches well with estimates of aggregate non-pension wealth, as reported by Blake and Orszag (1999). See appendix figure A21 for a reproduction of some of these comparisons over time, from Cummins, 2021.

<sup>8</sup>More detail on this is given in appendix section F.

charitable donations and moving wealth offshore are all factors that will lead to *declared* wealth-at-death being different to lifetime wealth. Cummins (2022) estimates that such hidden wealth accounts for about 35% of total elite wealth. The assumption employed here is that hiding wealth is more or less equal between ethnicities.

The PPR Calendar data was supplemented by a database of the number of deaths and the number of probates, by surname, 1996-2018. Every probate over this period is listed at <https://probatesearch.service.gov.uk/#calendar>.<sup>9</sup>

The threshold estate value above which probate was legally required has been £5,000 from 1984 to today, 2020. In recent years however, the *de facto* reality is that financial institutions have exercised discretion in releasing monies to relatives and beneficiaries from the bank accounts of the recently deceased. In 2020, banks apply their own discretion upon which accounts need probate and which don't. The value they apply as their probate limit could range from £5,000 to £50,000.<sup>10</sup>

It is not clear from existing academic literature or the archives of official Govt. websites advising on probate (<https://webarchive.nationalarchives.gov.uk>) when exactly the nominal probate went from being a flat £5,000 across all institutions, to a discretionary amount that varies in the range £5-£50 thousand, and is institution specific. In 2007-8 (see Atkinson et al. (2017, F8) and as late as 2010 (See Karagiannaki (2015, p.187)), there is evidence that the £5,000 probate threshold was generally applied.<sup>11</sup>

Before 1994, at least, and probably until at least 2010, the assumption that the non-probated estates were worth precisely less than £5,000 appears to be well justified. However, for post-2010, and in particular more recent years, this assumption is not reasonable. Therefore, we can only interpret the probate rate 1996-2018 as being an indicator of wealth that was significant enough for the asset holders (e.g the banks or building societies) to demand an act of probate before transferring the monies. As this could be anywhere between £5-50 thousand, the probate rate after 2010 can only be interpreted as a measure of significant wealth, and not wealth above the legal probate threshold.

Wealth in joint ownership, such as a joint bank account, or housing, was not subject to probate. As everyone has to die, wealth in joint ownership, between a husband and wife for example, should appear in the probate records at some stage. However, across ethnicities, the mapping of probated wealth to 'true' wealth could be distorted by varying levels of inter vivos bequests, associated cultural norms, marriage patterns, family composition and gender dynamics. Unfortunately, these potential biases are impossible to address in this present analysis.

The PPR Calendar data, 1858-1992, is complemented by surname-level data on the number of probates and deaths, from 1996-2018. This was downloaded from <https://probatesearch.service.gov.uk/#calendar>, and the number of deaths per surname was taken from the vital

---

<sup>9</sup>It was necessary to enter an exact surname on the webpage to return the count of that surname for a given year. From a 100% sample of the 1881 census ((Schurer and Woollard, 2000)) and the 100% samples of births, marriages and deaths, 1838-2007, and the probate Calendar 1892-1992, a master-list of 3,535,375 surnames was created. Of these surnames many were mistakes so a second list was created filtering the master list by the criteria that the name appeared at least 5 or more times in the death registers, 1983-2007. This resulted in 92,812 surnames which were searched individually for every year 1996 to 2020, a total of 2,320,300 searches for each of the 25 years. Each surname from this master-list was entered into <https://probatesearch.service.gov.uk/#calendar> and the count recorded (GOV.UK (2018)).

<sup>10</sup>The current official Government advice on probate states "Contact each asset holder (for example a bank or mortgage company) to find out if you'll need probate to get access to their assets. Every organization has its own rules." GOV.UK (2020). A list of the institution specific probate limits are reported here: <https://www.co-oplegalservices.co.uk/media-centre/articles-may-aug-2018/bank-limits-for-probate/>. A news article from 1994 states "Although the Act does not specifically apply to banks and to building societies, they usually apply their discretion in a similar way, and will normally only pay out above the pounds 5,000 limit with a grant of probate." <https://www.independent.co.uk/news/business/why-the-bereaved-must-wait-rules-governing-the-release-of-money-when-a-person-dies-can-cause-1420519.html>. A 2017 blog post by a probate professional [www.todayswillsandprobate.co.uk/guest-writers/obtaining-up-to-50k-without-grant-probate/](http://www.todayswillsandprobate.co.uk/guest-writers/obtaining-up-to-50k-without-grant-probate/) discusses the change.

<sup>11</sup>Atkinson et al. (2017) state "We have been told by Her Majesty's Revenue & Customs (HMRC) that the 'small estate' category probably accounts for the large majority of estates that do not go through probate" (p.F8).

register data (described below) from 1996 to 2007, and extrapolated to 2018. This is described in more detail in appendix section A.

### 2.1.1 Selection

Selection into observation of wealth depends on a set of characteristics that could also vary by ethnicity. Rates of return migration, the socio-economic selectivity of that return migration, cultural norms about remittances and legal knowledge of the probate process, as well as age at death, will all impart some “bias” on the relative wealth patterns reported here. The registry data is comprehensive, universal and unique. However, it lacks the fine grain resolution to tease apart the relative roles of the factors just listed in variations in ethnic patterns of wealth. But in a companion paper, Cummins and Gráda (2022), we assess these relative roles for the Irish in England, and find our results robust. The analysis presented here is meant as an overview of the ethnic wealth inequality landscape.

## 2.2 Vital Registers of Births, Marriages and Deaths, 1837-2007

From the 1st of July 1837 a National Civil Registration system has been in place in England and Wales. The internet age has led to the mass digitization of these records by various groups interested in family history and they have posted this information online. I compiled a database of 125,005,217 births 47,082,406 marriages, and 85,932,666 deaths, from 1837 to 2007, for England and Wales by downloading the individual index entries from two such websites: freebmd.com (1837-1980) and familysearch.org (1980-2007).<sup>12</sup> Figure A22, reported in the appendix, by year for each vital series a comparison of the numbers collected versus that recorded by the official records (from Office for National Statistics (2021).)

In all cases the harvested counts closely match that expected from official statistics for the vast majority of years between 1837 and 2007. The exceptions are the sharp drops in numbers harvested in the 1970s for births and marriages; this is because the underlying website (freebmd.com) was incomplete for those years when the data was collected.

## 3 Empirical Methodology

This analysis rests upon the claim that surnames are hereditary cultural labels that reliably indicate ethnicity.

### What are ‘Surnames’?

Surnames are hereditary cultural labels that typically denote an individual’s membership of a genetically related family.<sup>13</sup> In England, surnames originated amongst the aristocracy, in particular the Norman elite who often denoted themselves and their progeny by their ancestral lands in France. By 1538, when parishes were ordered by Henry Tudor, via Thomas Cromwell, “to keep a register of weddings, christenings, and burials”<sup>14</sup>, surnames appear as universal across all social ranks. Today, an individual cannot get a passport, a visa, register a birth, marriage or death, without also providing the mandatory surname.

Since at least 1538, children typically inherit their surname from their father, or in rarer cases, their mother.<sup>15</sup> Traditionally, women change their surname upon marriage to the surname of the new husband. Thus English surnames generally mark clusters of genetically related men and this transmission of surnames from father to son resembles the genetic inheritance of the Y-chromosome. The more unusual the surname, the more likely holders are close genetic relatives:

<sup>12</sup>The collected marriage counts were divided by 2 because there were two index entries for every one marriage.

<sup>13</sup>Surnames can also be referred to as ‘last-names’ or ‘family names’.

<sup>14</sup>*Letters and Papers, Foreign and Domestic of the Reign of Henry VIII*, Vol. 13, #281 available at <https://www.british-history.ac.uk/>. To see the universality of surnames in England by 1538, inspect the transcriptions of the 1538 parish registers at <https://www.familysearch.org/>.

<sup>15</sup>For example, where a mother is unmarried, children tend to inherit her surname.

this probability is also dependent on the number of founders of a surname, the incidence of non-paternities and genetic drift. (King et al., 2006; King and Jobling, 2009)).<sup>16</sup>

## Surnames can indicate ethnicity

Globally, surnames originated in a variety of historical and geographic contexts. Due to their hereditary nature, they can serve as markers of ethnicity, in that they link subsequent generations to an originator. Members of an ethnic group typically share an identity based upon geography, shared history, language and community. Ethnicity is thus multidimensional, and is non-exclusive in that one person could identify with multiple ethnicities. (This is particularly relevant in modern Britain.) What exactly constitutes ethnicity in any point in time is *socially constructed*, and thus fluid. I employ the term loosely, with the goal of documenting previously invisible inequalities in wealth. Overviews of the use of surnames to infer ethnicity, in the social sciences and genetics, are given in Mateos (2007); Mateos et al. (2011).

## Are Surnames a reliable indicator of ethnicity?

In practical terms, surnames are non-random nominal labels. They are self-declared by their holders and are technically fungible. In England, contrary to wide-held belief, anyone at anytime can change their name.<sup>17</sup> In practice however, people generally preserve the surname they are attributed at birth throughout their life, or until marriage in the case of women. Due to their heritable nature, surnames can also give an indication of an individuals ethnic heritage.<sup>18</sup>

I use the Onomap ethnicity name classifier software to assign ethnicity to all the surnames from the probate and vital registers.<sup>19</sup> Onomap has been developed by Paul Longley and numerous collaborators at UCL. Mateos et al. (2011) describe the methodology behind such classifiers. Using sources such as telephone directories and electoral registers from nearly all countries in the World, a network analysis connects personal names to each other (nodes) by forename-surname pairs (edges). They demonstrate how such constructed ‘naming networks’ spontaneously produce clusters that map onto known ethnocultural groups.<sup>20</sup>

Kandt and Longley (2018) test the accuracy of the Onomap ethnicity classifier for 51m census records for England and Wales from 2011. They compare the Onomap ethnicity prediction with that self-reported by the same individuals. The positivity rate, for each of the 11 census ethnicity groups, by sex, is reported in appendix figure A23 (which is simply an exact reproduction of figure 4 from Kandt and Longley (2018)). The Onomap classifier not perfect. Accuracy rates vary from 85% for ‘White British’ to about 0% for ‘Black Caribbean. For ‘Black African’ the accuracy is about 40%. For other groups, the performance of the tool is better; such as ‘White Irish’ (about 50%), Indian (60%), Pakistani (80%), Bangladeshi (60%) and Chinese (80%).<sup>21</sup>

Given that any surname based ethnic classifier will never categorize individuals with 100% accuracy, these numbers are promising.<sup>22</sup> Where accuracy rates are above 85%, it is plausible

---

<sup>16</sup>Modern forensics has famously solved decades old ‘cold-cases’ exploiting the new art of genetic genealogy ((Gymrek et al., 2013))

<sup>17</sup>A ‘deed poll’ is required to prove the name change for the purpose of official documents. This currently costs £42.44 <https://www.gov.uk/change-name-deed-poll>. Name changes after marriage do not require a deed poll.

<sup>18</sup>For a review of the use of surname based ethnicity classification in demography see Mateos (2007). The use of surname distributions is more widespread in genetics research; for example, see Lasker (1985); Piazza et al. (1987); Jobling (2001); King et al. (2006); King and Jobling (2009). In health, see Smith et al. (2017).

<sup>19</sup>Kindly provided to me by Paul Longley and Oliver O’Brien (Both Department of Geography, University College London).

<sup>20</sup>Mateos et al. (2011) provide a nice analogy that I quote here: Drawing a parallel with amazon.com’s recommendation service; “people who bought this book also bought...” we could say that “people who bear this surname often choose these forenames”. Pursuing this analogy, just like book titles at amazon.com have automatically been clustered into genres using purchasing behavior in a network representation we propose to cluster surnames into cultural, ethnic and linguistic groups of forenaming preference in a similar fashion using population registers” (p.2).

<sup>21</sup>Lakha et al. (2011) also test the performance of the Onomap classifier in Scotland and conclude that “Onomap offers an effective methodology for identifying population groups in both health-related and educational datasets”.

<sup>22</sup>For example, many Irish surnames are of Scottish, English and Welsh origin, yet their holders would declare



that an Onomap selected sample is representative of the underlying ethnic group. Where accuracy rates are lower, the potential that the Onomap sample is unrepresentative of the underlying ethnic group is higher. For accuracy rates of 0-10% it is very hard to claim representability. How concerning are these varying levels of accuracy for the analysis employed here? On one hand, it can be argued that if the classifier is only imperfectly assigning ethnicity, then any observed inequalities could be interpreted as an underestimate of the true status; if Onomap was simply assigning categories randomly we should observe no ethnic inequality.

Alternatively, one could imagine that as ethnic groups assimilate into the English population, many change, or ‘Anglicize’ their name. This social process is well known and has been observed for Huguenot, Irish, German, Greek and Jewish names in recent English history.

Another source of classification error would be where an the group becomes so assimilated that the Onomap methodology cannot distinguish an ethnically distinct cluster. A concern would be that this happens differentially, by status, for ethnic groups. This could happen where one ascendant subset of an immigrant group marries into the English population and those their names become associated with the English and not their own ethnic group of origin. One example of this phenomenon is the surname “D’Arcy”, which originated in Northern France, but Onomap classifies as “English”. This process could be particularly important for immigrant groups that have had longer to a assimilate, such as the Irish and Huguenot.<sup>23</sup> This would result in a biased estimation of the persistence of ethnic group inequality.

For example, the “Irish” in England is interpretable as those *Irish with distinctively Irish names*. An ascendant subset of Irish who assimilate into the English population may now be indistinguishable in naming networks from the English.

Thus I supplement the Onomap ethnic classifier with my own bespoke historical classifier. Here I have to use ‘country’ in place of ‘ethnicity’. I use 36 million de-anonymized individual records from the special access version of the 1911 census of England and Wales, to examine the distribution of country of birth for the over 500,000 surnames (Schurer and Higgs (2021)). As we do not observe the global distribution of surnames in 1911 but rather the distribution *within* England, we cannot simply assign the most frequent country-of-birth to a surname. Based on the observed distribution of a set of well known national surnames, such as Churchill (English), Murphy (Irish), Ferrari (Italian) and Becker (German), I apply a simple set of rules. First, all surnames are assigned to the country that registers the highest proportion of births of a surname. Where there is a country other than England or Wales which accounts for 5% or over of the births of that surname, I update the ethnic origin to that country. See appendix section E, and also a related paper Cummins and Gráda (2022).<sup>24</sup>

The automated assignment of ethnic origin to individuals via surnames has the advantage of being applicable to existing databases across across many countries, and centuries. Thus there is the possibility of extensive international and historical comparisons. Due to the inconsistency of ethnic classifications in officially collected data between countries and over time this is currently not possible (Farkas (2017)). One advantage of the Onomap classification scheme, relative to the 1911 classifier is resolution. As opposed to the 11 census ethnic group categories typically reported (for example in Office for National Statistics (2019b) and Office for National Statistics (2020)), Onomap assigns surnames to about 138 distinct ethnic groups in the probate data.<sup>25</sup>

---

themselves as Irish. For groups such as the Black Caribbean, as many surnames were originally adopted by slaves after their, very often, British surnamed masters.

<sup>23</sup>The author notes that his own surname, ‘Cummins’, is classified as English, despite the frequency of the name being higher in Ireland (<https://forebears.io/surnames/cummins>).

<sup>24</sup>Appendix section H reports an analysis of surname level wealth and classification differences between Onomap and the 1911 census method. For example, do some richer (or poorer) families within some ethnicities become “English” between 1911 and the more recent Onomap observation? Some statistically significant correlations are evident but they are not in any consistent direction nor of magnitude.

<sup>25</sup>Table A3 reported in the appendix, lists the top 30 most numerous Onomap types, by population, for UK in 2004. Around 2% of deaths are classified as “Unknown” and 6% of probates are classified as “Unknown”. For the purposes of classifying percentile shares it was essential to preserve these wealth observations. However, the construction of the PPR Calendar data results in a higher error rate assigning ethnicity of a surname. The analysis assumes that this error is random, there is no reason to think otherwise. The “Unknown” group as a share of deaths is reported in figure 2b. Because this group is unidentified and because of the differential error

It is crucial to clarify that the ethnic classification systems employed here do not identify recent migrants. Rather they identify (1.) clusters of surnames connected through (first name) naming patterns, and (2.) surnames that in 1911 were associated with non English (or Welsh) countries of birth.

### 3.1 The Wealth of Foreign Nationals

A potential difficulty in using the PPR calendars to understand ethnic wealth inequality is the fact that the Calendars will also record foreign wealth holders. Wealthy Indians, Irish and Africans could live abroad but hold assets in London, for example. Fortunately the PPR calendars report the place of death of all decedence. I took the last two words of every address entry, 1892-1992, and cross referenced these words with current and historical names of countries, and with names of Irish towns.<sup>26</sup>

Appendix figure A24 reports this proportion 1892-1992. Surprising the PPR calendars record less foreign wealth holders over time, the proportion declines from about 3% in 1900, to about 1% by 1980. Despite this low proportion, I exclude these foreign nationals from the ethnic wealth calculations executed here. Thus this analysis speaks to those ethnic minorities living in, and dying in, England and Wales.

### 3.2 Wealth Calculations

I analyze ethnic wealth inequality through three measures: 1. the probate rate, 2. average wealth, and 3. the representation of an ethnicity in the top 1% of wealth-holders.<sup>27</sup> To do this I combine the PPR and death registers together to construct a ‘synthetic’ individual level dataset of *all* adult deaths in England 1858 to 1992. In other words I append to the individual level death data extra rows each representing individuals who die below the wealth threshold. This is done by year of death, and by ethnicity, by comparing for each the number of those probated, with the number dying. From 1996 to 2018, I do not have specific estate valuations but only whether a decedent was probated or not. Thus only the probate rate, by ethnicity, is available for that most recent period.

The construction of these individual level data allow the easy calculation of ethnic group averages, and facilitates the application of simple regression models. It also allows an easy calculation of the top wealth percentile threshold, and weighs the surname ethnicity wealth estimates correctly. Finally it allows the easy calculation of standard errors for any ethnic group wealth ‘effects’.

The PPR Calendar records all decedents who have wealth above the probate threshold. As Alvaredo et al. (2017, p.F9) and Cummins (2021), I treat these non-probated estates as reporting ‘insignificant’ wealth. The number of adults who die with some level of wealth below the probate threshold, is calculated for ethnicity  $e$  as

$$N_{np}^e = N_{20}^e - N_p^e \tag{1}$$

where  $N_{np}$  is the number not probated,  $N_p$  is the number probated, and  $N_{20}$  is the number of adult deaths where age at death is greater, or equal, to 20 years, and is calculated from the death registers.<sup>28</sup> For each of these non-probated deaths, I generate one observation, indicating

---

in assigning ethnicity between the PPR Calendar and the death index, the wealth characteristics of this groups are not analyzed. Northern Irish names are assigned as Irish in this paper. Note that particular care had to be taken when assigning Irish names die to the use of the O’ and Mc/Mac prefixes.

<sup>26</sup>Inspecting by eye the top 5,000 strings by frequency led to this choice. I only was able to examine addresses after 1892, as the structure of the data before 1892 meant I only had name and probate year, and probate valuation extracted from the original images.

<sup>27</sup>I do not analyze median wealth as the median wealth of adults dying in England is actually below the probate threshold, a point discussed in Cummins (2021).

<sup>28</sup>As age at death is only recorded in the death registers beginning in 1866, I simply used the average number of adult deaths registered by ethnicity over 1866 to 1875 and used that value for each of the years 1858 to 1865 inclusive. More generally, age-specific mortality rates, calculated by ethnicity, could be used to map the wealth

ethnicity, that is appended to the individual PPR wealth database. I then assign to these non-probated observations an *inferred* wealth. This inferred wealth is set at 50% of the average level of wealth observed in the PPR calendar which was *below* the probate threshold of a given year. As with Cummins (2019b), this follows the the standard method used by official agencies such as HM Revenue and Customs (Turner (2010, p.628-9)). This inference assumes that wealth for an unprobated decedent does not vary with ethnicity. In the absence of other information, this is taken as the best working assumption.<sup>29</sup>

The probate rate ( $pr$ ) is then simply calculated as the simple mean of a probated categorical dummy ( $D_p$ ):

$$pr^e = \frac{N_p^e}{N_{20}^e} = \bar{D}_p^e \quad (2)$$

I can calculate the probate rate by ethnicity and year from 1858 to 1992, and from 1996 to 2018.

Average wealth ( $\bar{w}^e$ ) is

$$\bar{w}^e = \frac{\sum w_p^e + \sum w_{np}^e}{N_{20}^e} \quad (3)$$

where  $w_p$  and  $w_{np}$  represent probated and non-probated wealth. Due to the construction of the synthetic individual level dataset, it is straightforward to calculate average wealth grouped by ethnicity and year here.

Finally, representation within the top 1% is calculated as the mean of a dummy variable for having wealth above or equal to the 99th percentile, calculated across all adult deaths, in a given year.

The final sample size for the synthetic PPR-death register data is 71,668,665, and 12,486,026 for whether an individual is probated 1996 to 2018.

## 4 Results

Table 1 presents the summary statistics for the synthetic PPR-death data, 1858-1992, and table 2 does the same for the probate data 1996 to 2018. For the 1858-1992 period; real wealth varies from zero to £1.3 Billion (in 2015 pounds). John Reeves Ellerman (1862-1933) is the richest decedent (Cummins (2021)). Mean wealth is about £24,000, and median wealth is £930. 22% of adult deaths (over 20 years of age at death) have wealth to merit probate. For the 1996-2018 period the probate rate is 46%.

Section C, reported in the appendix, presents the most common surnames by ‘major’ ethnicity (British, Irish, and Other), the frequency of that surname (as measured by the number of deaths), the number of probates, the probate rate, and median probated wealth. These surname lists can be inspected as a sanity check on the groups ethnic wealth estimates presented later in this section.

### 4.1 A Quantitative History of the Ethnic Composition of England

The ethnic composition of Births and deaths is estimated from the micro-data for every year 1838 to 2007, and presented in figures 1 and 2. About 70% of the surnames that show up in

---

distribution of those dying to those living via the mortality-multiplier method, widely used in the literature (e.g. Alvaredo et al. (2018)). The difficulty is that we do not fully observe the number of living people by ethnicity, as many are migrants (and thus not recorded at birth), and many return to the sending county before death (return migration, so not observable at death). The latter issue, and the effect of age on wealth, is examined in a related article Cummins and Gráda (2022).

<sup>29</sup>One could use the probated wealth distribution to infer the shape of the non-probated (and therefore unobserved) wealth distribution by extrapolating based upon a Pareto, or log-normal, distribution. However, the present analysis focuses on exploring ethnic inequality at different moments of the distribution. The probate rate is thus treated as a separate (but related) measure to average wealth.

Table 1: Summary Statistics, PPR-Death Data, 1858-1992

Statistic	N	Mean	St. Dev.	Min	Median	Max
Death Year	71,668,665	1,926.51	39.15	1,858	1,927	1,992
$D_{Probated}$	71,668,665	.22	.41	0	0	1
Nominal Wealth	71,547,662	3.45	36.89	0.00	.03	94,117.96
Real Wealth	71,547,662	24.40	400.35	0.00	.93	1,257,372.00
$D_{Top1\%}$	71,547,662	.01	.10	0.00	0.00	1.00

Notes: Real wealth is in thousands of 2015 pounds.

Table 2: Summary Statistics, Ethnic Probate Rate Data, 1996-2018

Statistic	N	Mean	St. Dev.	Min	Max
Death Year	12,486,026	2,007.12	6.60	1,996	2,018
$D_{Probated}$	12,486,026	0.46	0.28	0	1

Notes: Only simple probate data available 1996-2018

the Birth registers from England and Wales 1838 to 1950 are classified as English. 20% are either Scottish, Welsh or Irish. About 5% or less are non British or Irish, until about 1950. The biggest change in the ethnic composition of England and Wales is the rise of the non-British and Irish ethnicities as a share of all births after 1950. From about 5% in 1950, these ethnicities account for 25% of all births in 2007.

This has yet to fully show up in the death data as most of this immigration cohort are still alive. Thus the birth registers work better than the death records, as contemporary tracker of the ethnic composition of England and Wales, By share of births the largest non-British or Irish ethnicity over the period 1838 to 2007 are the Pakistani group, followed by the Sikh, Italian, Jewish and German groups.

## 4.2 Who Owns England?

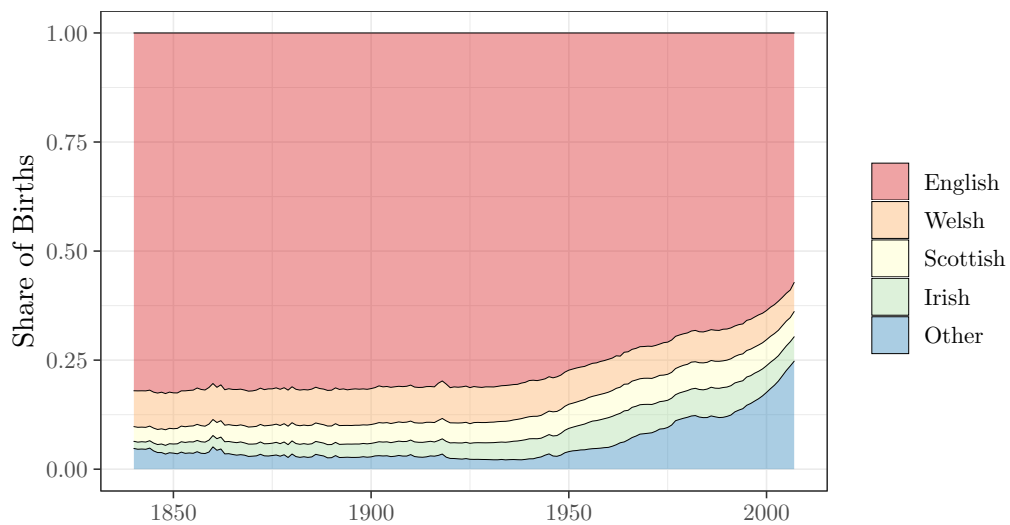
The share of English wealth that is owned by those with “English” surnames is consistently around 75% 1858 to 1992. Figure 3 presents the ethnic ownership of capital, as measured by the sum of probated wealth for the four major ethnic groups of the British Isles, and then separately for the 20 most numerous “Other” ethnic groups (by number of deaths, 1858-92).<sup>30</sup> Over the period 1858-1992, on average: the Welsh own about 6% of wealth, as do the Scottish. The Irish own about 1.7%.<sup>31</sup>

The “Other” category comprises a diverse set of ethnicities. Figure 3b plots the wealth share of these ethnicities, ordered by wealth share (largest to smallest).<sup>32</sup> They account for a small share of English wealth, around 2-3%, over the period. Sharp declines, as indicated by reversals in the time trend of the share of these ethnicities in English wealth ownership are evident during both the First and Second World Wars. The largest share is owned by surnames of Jewish and German origin, and the least by the Bangladeshi ethnic group.

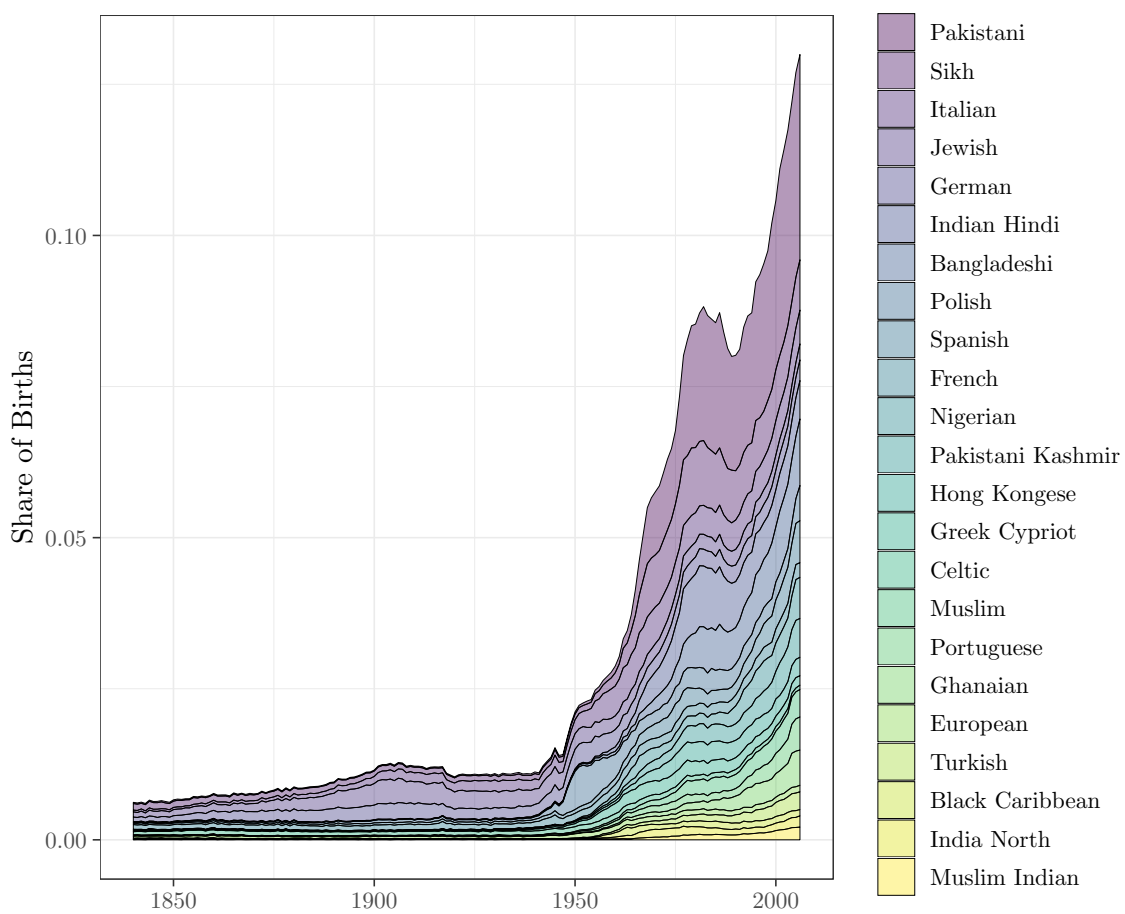
<sup>30</sup> The estimates oscillate from year to year due to the inherent sensitivity of wealth share estimates to super rich outliers so I have applied a rolling mean of 5 years centered on every annual observation.

<sup>31</sup>The kink at 1974 in figure 3a corresponds to changes in the probate threshold (as reported in table A4).

<sup>32</sup>This set of ethnicity excludes those “Other” categories that are not classified by Onomap, who own about 6-8% of English wealth.

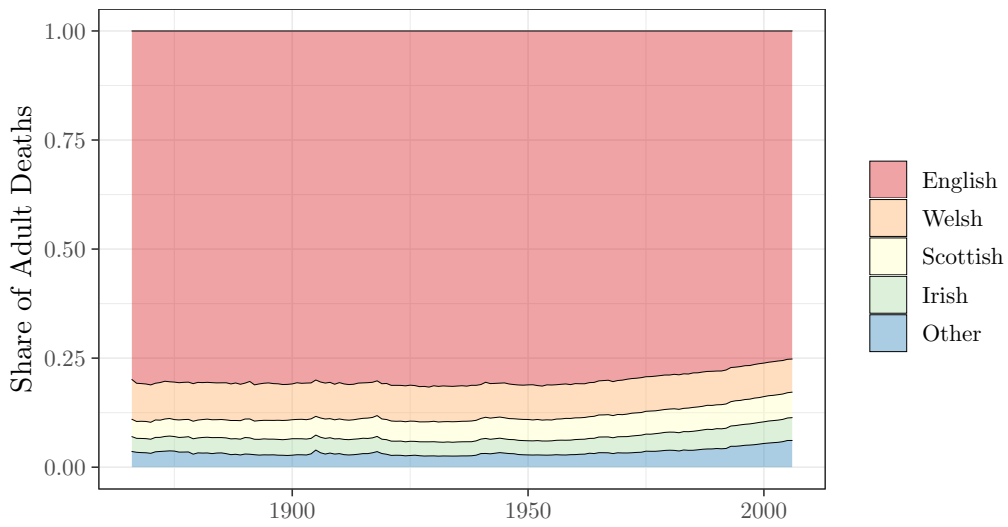


(a) British, Irish and Other

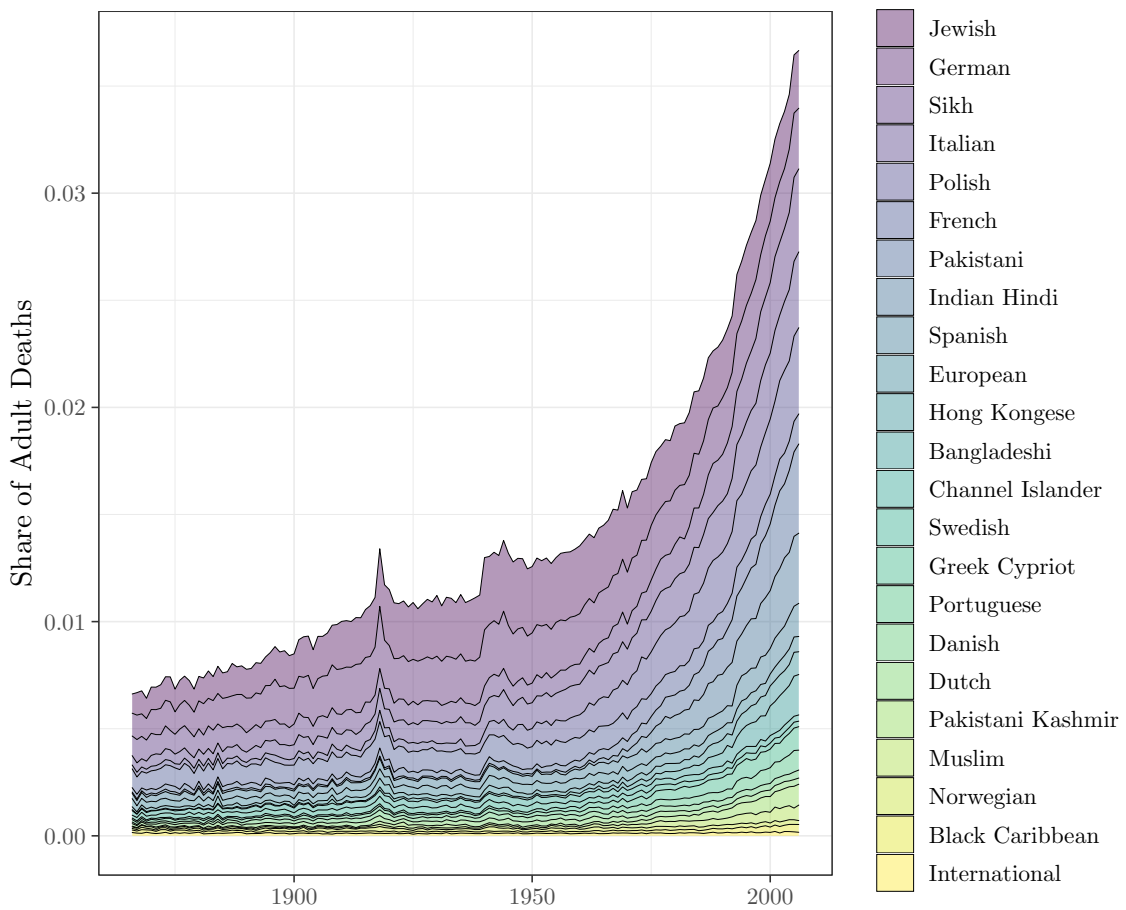


(b) Others

Figure 1: The Ethnic Composition of England and Wales, Birth Records 1838-2007  
 Source: Birth Register Index of All Births in England and Wales, 1838-2007.  $N = 127,760,704$ . Subfigure B is ordered by share of births over the period (largest to smallest).

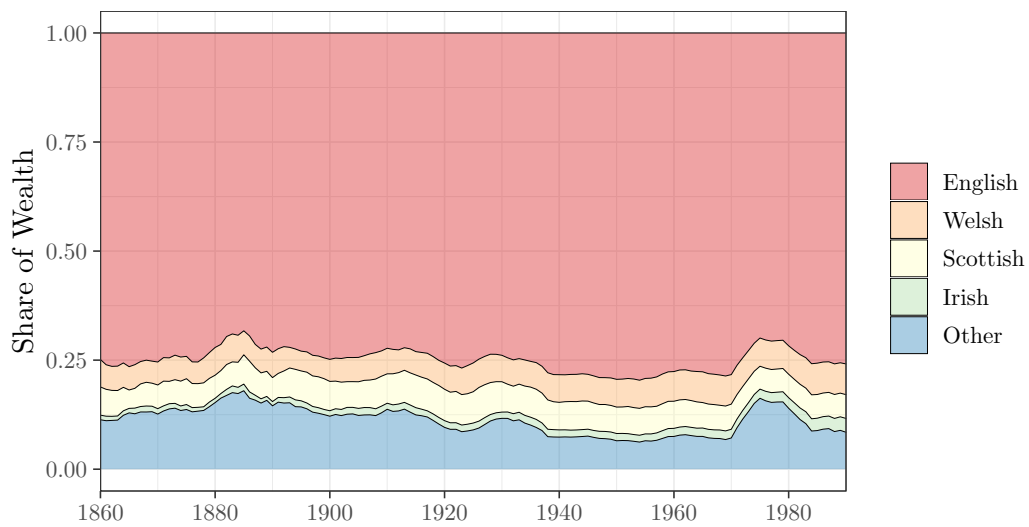


(a) British, Irish and Other

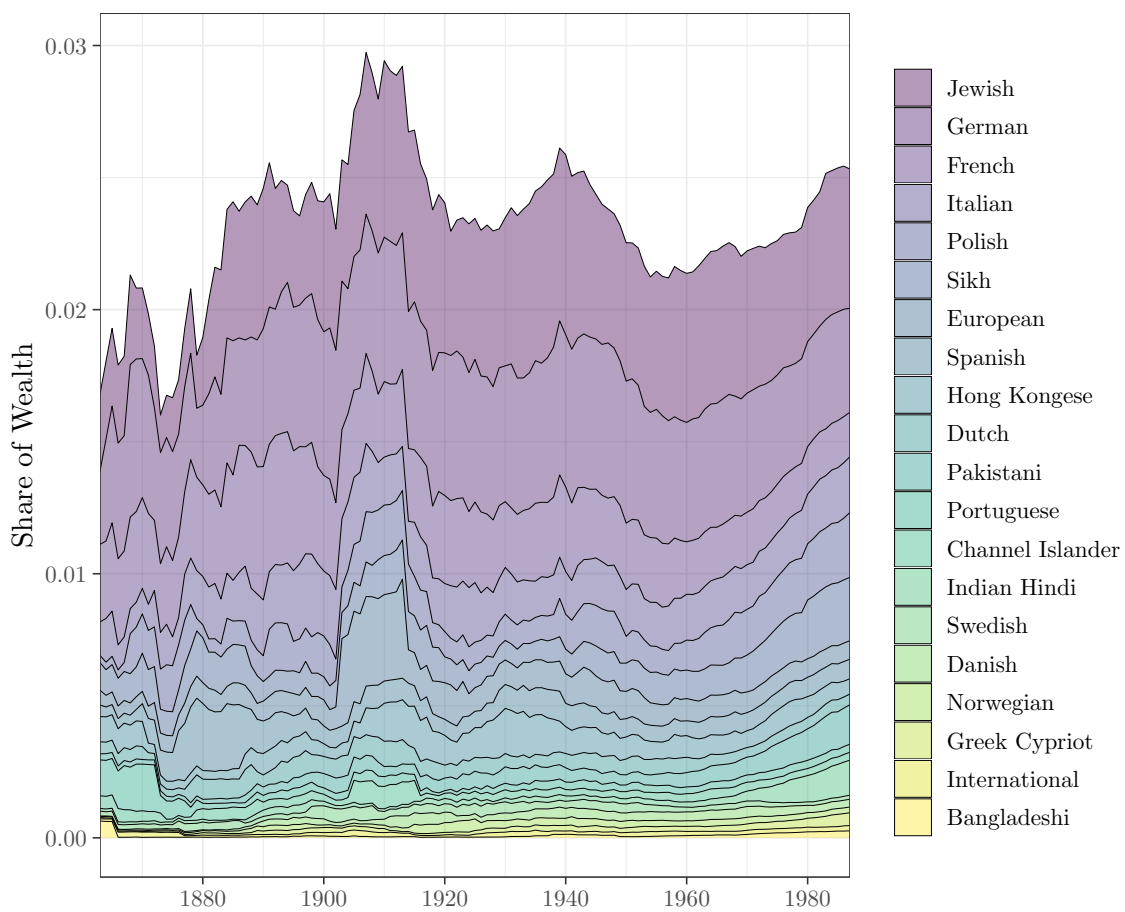


(b) Others

Figure 2: The Ethnic Composition of England and Wales, Death Records 1838-2007  
 Source: Death Register Index of All Deaths in England and Wales, 1866-2007.  $N = 75,353,417$ .  
 Subfigure B is ordered by share of deaths over the period (largest to smallest).



(a) British, Irish and Other



(b) Others

Figure 3: The Ethnic Composition of English Wealth, 1892-1992

Source: PPR Calendar and Vital Registers, 1858-1992. 5-year moving average applied to data. Subfigure B presents the 20 most numerous "other" ethnicities (by number of deaths 1858-92), and is ordered by share of wealth over the period (largest to smallest).

### 4.3 Ethnic Wealth Inequality

This section presents the ethnic breakdown of wealth along the three moments discussed in section 3. Namely the probate rate (section 4.3.1), average wealth (4.3.2) and representation amongst the top 1% of wealth holders (4.3.3).

#### 4.3.1 The Probate Rate, 1858-2018

Figure 4 reports the probate rate, the proportion of adult deaths reporting wealth above the probate threshold, by ethnicity, 1858 to 1992. As documented in earlier work (Cummins (2019b)), the majority of those dying in English do not have wealth sufficient to merit an act of probate, all the way from 1858 to 1992 (and as will be shown here, to 2018).

There is no substantial differences between the Scottish, Welsh and English. (Although the Scottish do register slightly higher probate rates from 1858 to the Second World War.) However, the Irish have substantially lower probate rates than the British every year 1858-1992.

As before I plot the 20 most numerous (by sum of deaths 1858-1992) non-British or Irish ethnicities in figure 4b, ordered by average probate rate over the period. The figure is difficult to read in specific detail but is meant only as illustrative of the broad trend and variation in the underlying data. Despite this it is clear that the war years are associated with sharp rises in the probate rate of these groups. This perhaps reflects richer-than-average refugees fleeing persecution, such as the Jewish community during World War II.

To assess more clearly these correlation, I estimate, by period of death, for individual  $j$  the probability of being probated ( $D^{Prob}$ , coded as 0/1) as a function of a set of  $i$  ethnicity group dummies as

$$D_j^{Prob} = \alpha + \sum_{i=1}^n \beta_i D_i^{Eth} + \varepsilon \quad (4)$$

where  $\alpha$  is a constant and  $\varepsilon$  is a random error. This simple linear regression ignores everything other than the ethnic variables.

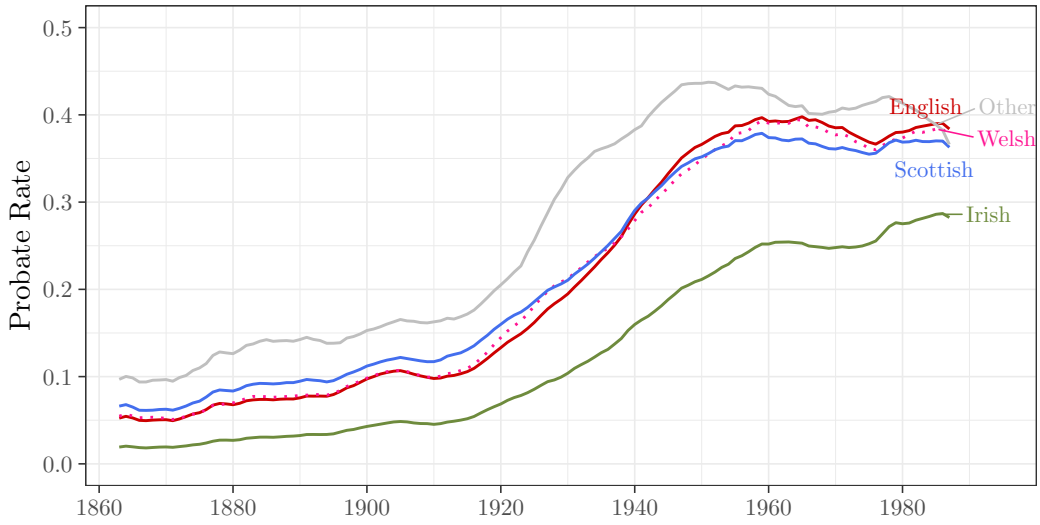
Table A14 reported in the appendix, details the results of this estimation for top 25 ethnicities (by number of deaths, 1858-1992), separately for the periods, 1858-79, 1880-1913, 1914-39, 1940-59, 1960-69, 1970-1979 and 1980-1992. Figure 5 reports the coefficient estimates and the 95% confidence interval for the first and last period, 1858-79, and 1980-92. The omitted category are the English, so the effect plotted in figure 5 is expressed relative to that group. The ethnic gradient of wealth in 1980-92 is very different from that of 1858-79. Then, most ethnic groups report wealth *above* that of the English, as measured by the proportion with probatable wealth. But in 1980-92, there is considerable variation relative to the English; many ethnicities are richer, such as Western European ethnicities and the Jewish communities. But many are significantly poorer, such as the Italian, Pakistani, Black Caribbean, Portuguese, Irish and Swedish ethnic groups. The richest group in England, by this measure 1980-1992, are the Dutch, and the poorest are the Bangladeshi.<sup>33</sup>

The emergence of this sharp ethnic gradient in wealth is difficult to precisely date but the coefficient estimates from table A14, and the general trend suggested from figure 4b suggest a post-WWII origin. This is surprising as we know that the post-war period is associated with a dramatic decline in the wealth share of the top 1%. This aggregate distributional equalizing force failed to equalize outcomes between ethnicities.

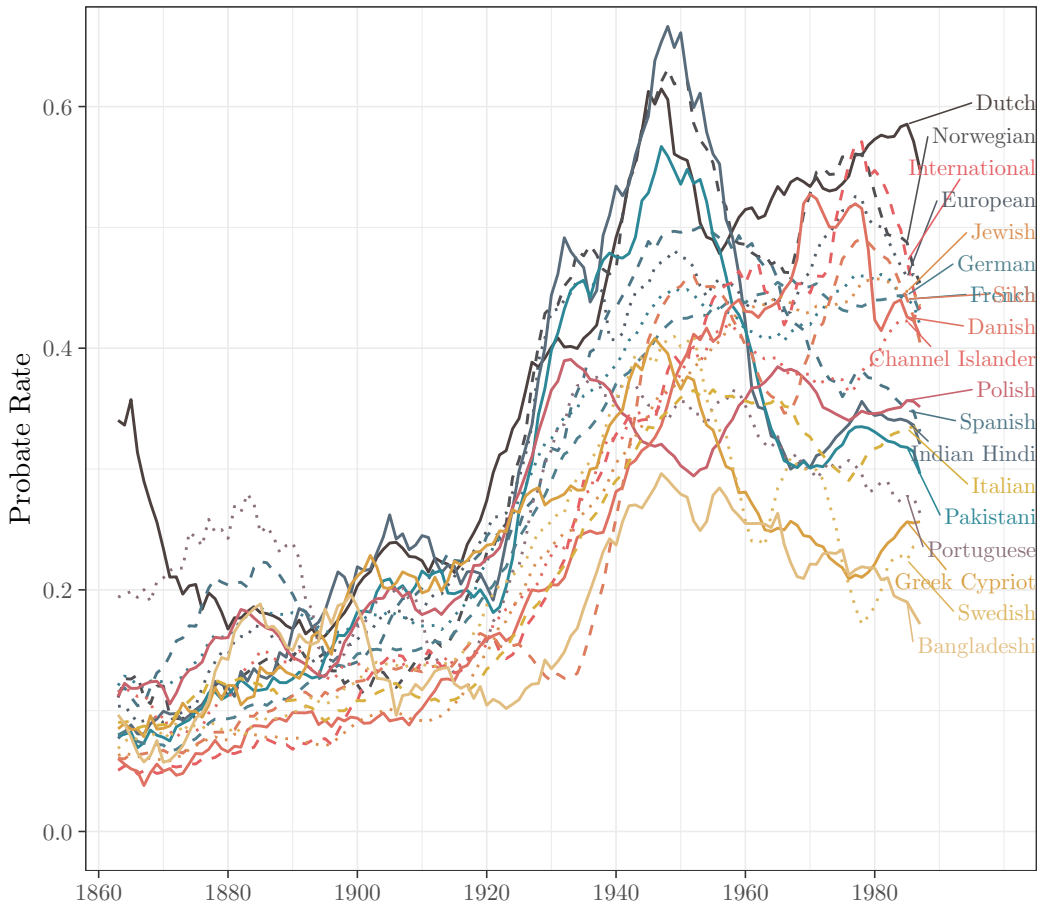
As detailed in section 2, I have complemented the PPR data 1858-1992, with counts of surnames and probates 1996-2018. By incorporating the counts of adult deaths by ethnicity to 2018, and extrapolating these counts by the average growth rate of the number of adult deaths for all those dying in England, I am able to estimate the probate rate 1996-2018. Figures 6 and 7 reports these results, interpretable as measuring the proportion of a group with significant wealth. (Please see the discussion in section 2 where I discuss the difficulty of interpreting

<sup>33</sup>A concern with Dutch names is that they could be disproportionately associated with the English aristocracy.





(a) British, Irish and Other



(b) Others

Figure 4: The Probate Rate, by Ethnicity, 1892-1992

*Note:* The probate rate is the number of probates divided by the number of adult deaths. Source: 100% sample of Probates, 1892-1992. Subfigure B presents the 20 most numerous “other” ethnicities (by number of deaths 1858-92), and is ordered by the level of the probate rate in 1992. The figure is difficult to read in specific detail but is meant only as illustrative of the broad trend and variation in the underlying data.

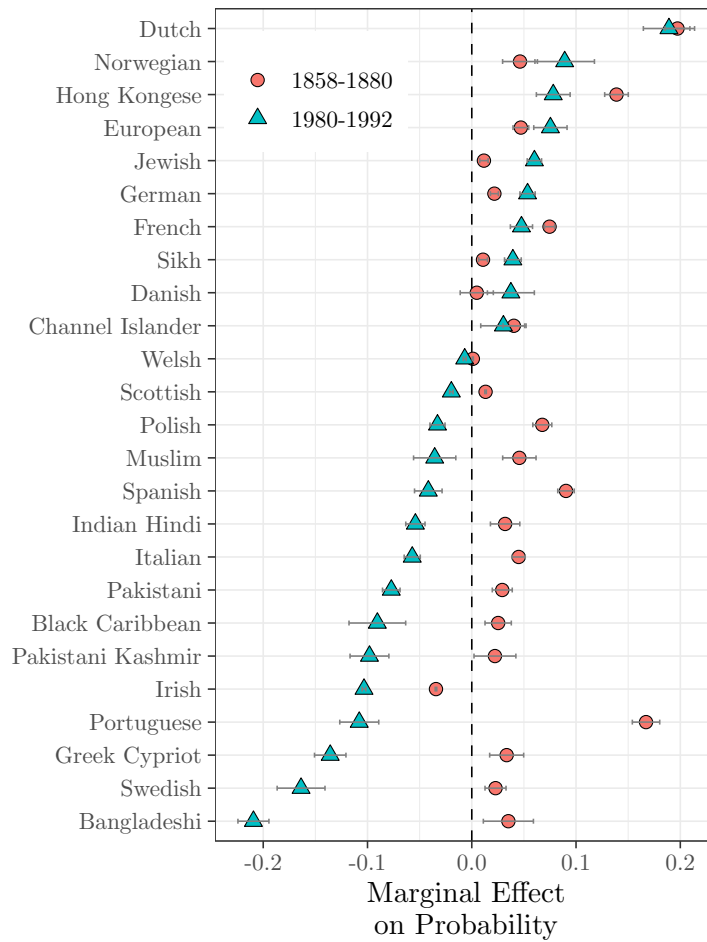
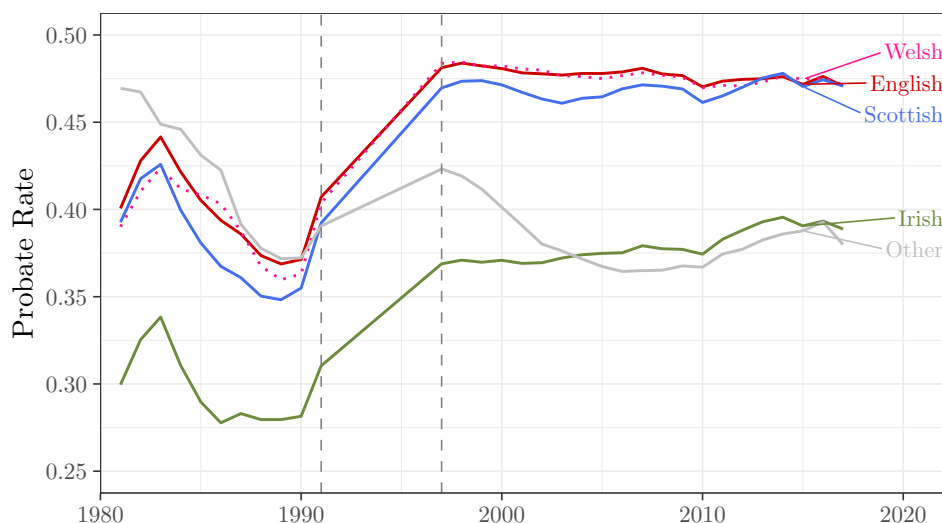


Figure 5: The Marginal Effect of Ethnic Group on Probability Probated, 1858-79 Compared with 1980-1992

Notes: "Significant" wealth is wealth sufficient to merit an act of probate upon death (e.g. £10 in 1900, £5,000 in 1990). The error bars are 95% confidence intervals. "English" are the committed category.



(a) British, Irish and Other

Figure 6: The Probate Rate, by Ethnicity, 1996-2019

Notes: “Significant” wealth is wealth sufficient to merit an act of probate upon death (e.g. £10 in 1900, £,5000 in 1990). Note that 1992 to 1996 are interpolated values. The probate rate is calculated as  $\frac{N_{Probs}^i}{N_{Deaths}^i}$ , where  $N_{Probs}^i$  is the number of probates for ethnicity  $i$ , and  $N_{Deaths}^i$  is the number of adult deaths to surname  $i$ . Note that adult deaths after 2006 are extrapolated from 2007 to 2018 by ethnicity, by simply assigning the 2006 value. The observed 1996 values are omitted from this figure as it is clear that the data are about 10% incomplete for that year (the levels for *all* the ethnic groups were about 10% lower than 1997).

probate in this period.) Note that 1992 to 1996 are interpolated values, and the values 1980 to 1992 are plotted for reference from figure 4.

Figure 6 reveals that that the Welsh, Scottish and English are indistinguishable from each other, 1996-2018. The Irish still report significantly lower probate rates however. Further the figure reveals the decline in the status of the ‘other’ ethnic groups from around the mid 1990s. From being a richer ‘group’ 1858 to 1985, they are now, like the Irish, dying with significantly lower probabilities of needing an act of probate, 1996 to 2018.

Figure 7 reports the probate rate for the top 50 ethnicities (by number of deaths), 1996 to 2018. The Channel Islander, Belgian Flemish and Jewish ethnicities have the highest probate rate in this period, and the Sri Lankan, Vietnamese and Brazilian ethnicities have the lowest.

Now most minority ethnic groups report lower probate rates than the English. This modern ethnic wealth gradient is a new feature of the wealth distribution and is a reversal of the typically high status of those with minority ethnic origin surnames before 1990.<sup>34</sup>

### 4.3.2 Average Wealth, 1858-1992

Figure 8 reports average real wealth (in £2015 pounds) by ethnicity and year, 1858-1992. The effects of the World Wars on wealth are striking with sharp drops observable for World War I in particular, but also for World War II. The surprising flat growth of wealth at death after 1950

<sup>34</sup>As discussed in section 2, cultural norms that could affect joint ownership of assets, family formation, gender dynamics, and inter-vivos bequests, could drive some of the observed ethnic gradient in the probate rate. A limitation of this analysis is that we only observe probatable wealth at death. If we assume a ‘stickiness’ to culture, relative to wealth; the secular decline in the relative probate rate for many ethnicities, as shown in figure 4 suggests a real change in relative wealth, and not just cultural difference. However, due to potential shifts in the character, and the economic and social selectivity of migration to England, over time, we cannot rule out a large role for culture in determining the level of probatable wealth.

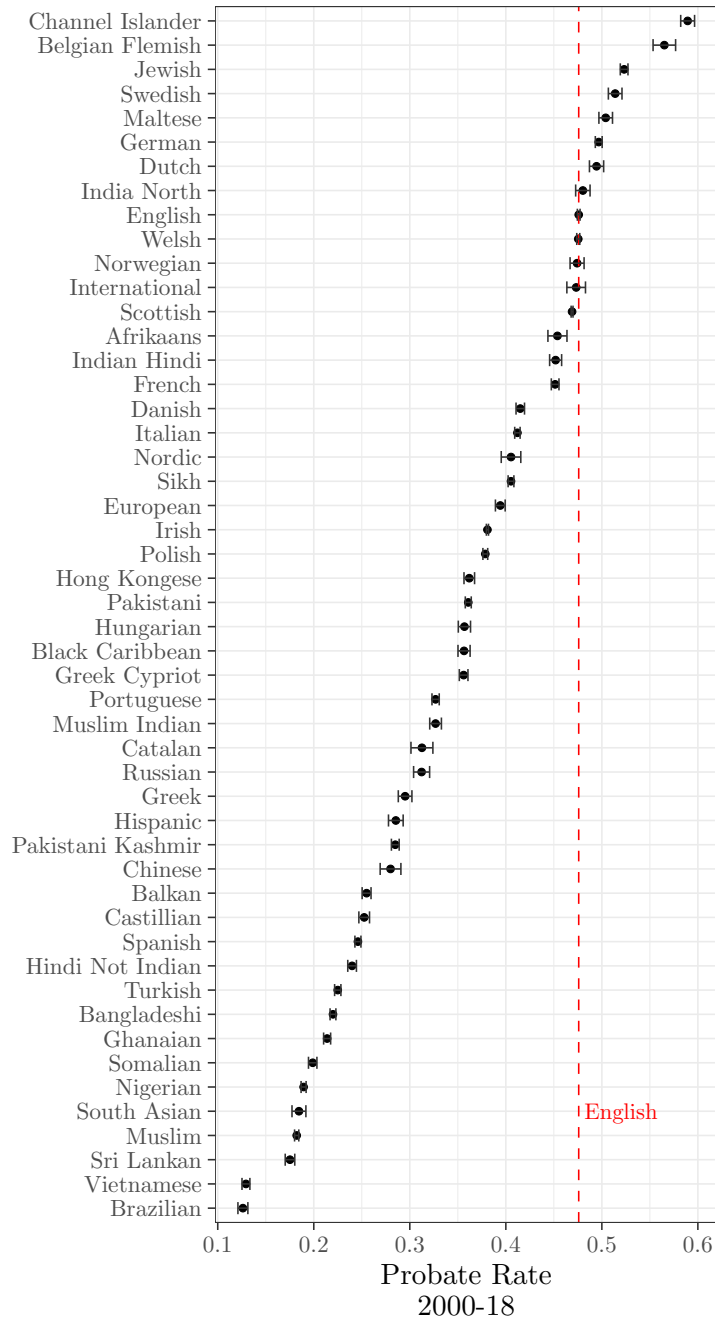


Figure 7: The Proportion of Deaths with Significant Wealth, 2000-18, by ethnicity

Notes: “Significant” wealth is wealth sufficient to merit an act of probate upon death. The probate rate is calculated as  $\frac{N_{Probs}^i}{N_{ADeaths}^i}$ , where  $N_{Probs}$  is the number of probates for ethnicity  $i$ , and  $N_{ADeaths}$  is the number of adult deaths to surname  $i$ . The error bars represent 95% confidence intervals.

is focused upon in another companion paper to this one, Cummins (2019a), which argues that a large proportion of wealth is hidden, in trusts and offshore accounts, during this period.

The ethnic group patterns are intriguing. Again the Irish are consistently and significantly always poorer than the English. The Welsh are poorer but much closer in average level to the English. The Scottish are always richer, at least until the 1980s, and the other minority ethnic groups are always significantly richer, throughout. These broad patterns mask significant variation between the ethnic groups, as revealed in figure 8b. Again, the figure cannot communicate precise details but does convey the wide spectrum of wealth status amongst the most numerous (by number of deaths) English ethnic minorities.

As with the proportion probated, I estimate for individual  $j$  the effect of ethnic group (coded as a set of dummy variables for the  $i$  ethnicities,  $D^{Eth}$  on real wealth ( $W$ ), from the regression:

$$\ln(W_j) = \alpha + \sum_{i=1}^n \beta_i D_i^{Eth} + \varepsilon \quad (5)$$

where, as before,  $\alpha$  is a constant and  $\varepsilon$  is a random error. Table A15 reported in the appendix, details the results of this estimation for top 25 ethnicities (by number of deaths, 1858-1992), separately for the periods, 1858-79, 1880-1913, 1914-39, 1940-59, 1960-69, 1970-1979 and 1980-1992. Figure 9 reports the coefficient estimates and the 95% confidence interval for the first and last period, 1858-79, and 1980-92. As with figure 9, the omitted category are the English.

In general the pattern corresponds to that revealed by the proportion probated measure; the Dutch are the richest ethnic group, the Bangladeshi are the poorest. Again the general large positive effects of ethnicity on wealth 1858-80, reverse by 1980-92, for many groups. For example the Polish, Pakistani and Bangladeshi groups. Other groups experience the opposite trajectory, their 1980-92 wealth semi-elasticity is significantly higher than 1858-79; namely the Jewish and German groups. As with figure 5, the Irish are always poorer than the English.<sup>35</sup> As with the evidence from the proportion probated, these average wealth-estimates reveal that the modern ethnic wealth gradient is of recent emergence.

### 4.3.3 Representation in the Top 1%

How well do ethnic groups assimilate into the very top of the English wealth distribution? Figure 10 reports the ethnic composition of the top 1% richest English wealth holders, by year of death, 1858-1992. The overall pattern of ethnic representation in the top 1% of wealth has been one of stability, with a marked rise in the representation of other ethnicities 1970 to 1980, before a subsequent decline 1980-92. Examining figure 10b, we can see that some ethnicities have increased their representation in the top 1% significantly, such as the German and Jewish communities.

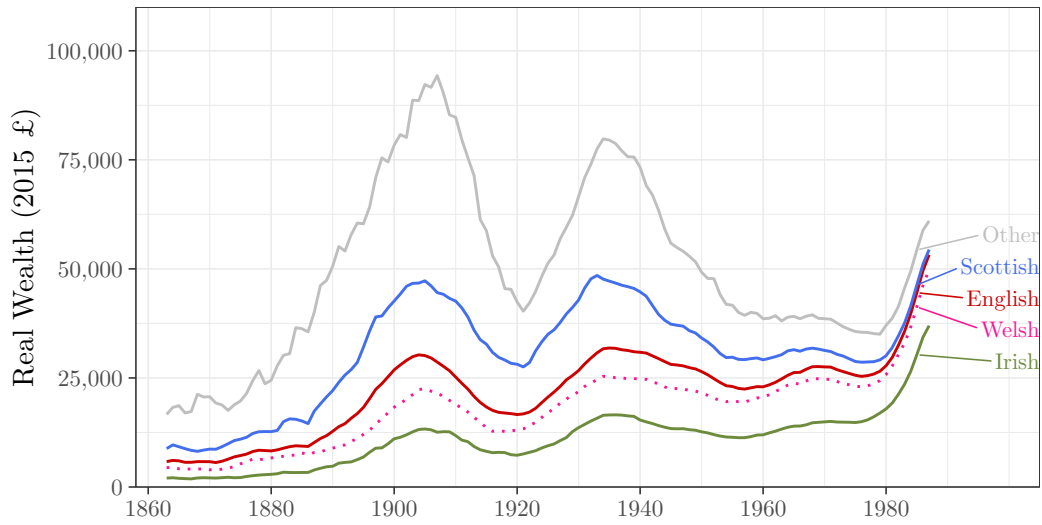
Once more, to understand ethnic group representation within the top 1%, I estimate for individual  $j$  the probability of being in the top 1% ( $D^{Top1\%}$ , coded as 0/1) as function of a set of  $i$  ethnicity group dummies as

$$D_j^{Top1\%} = \alpha + \sum_{i=1}^n \beta_i D_i^{Eth} + \varepsilon \quad (6)$$

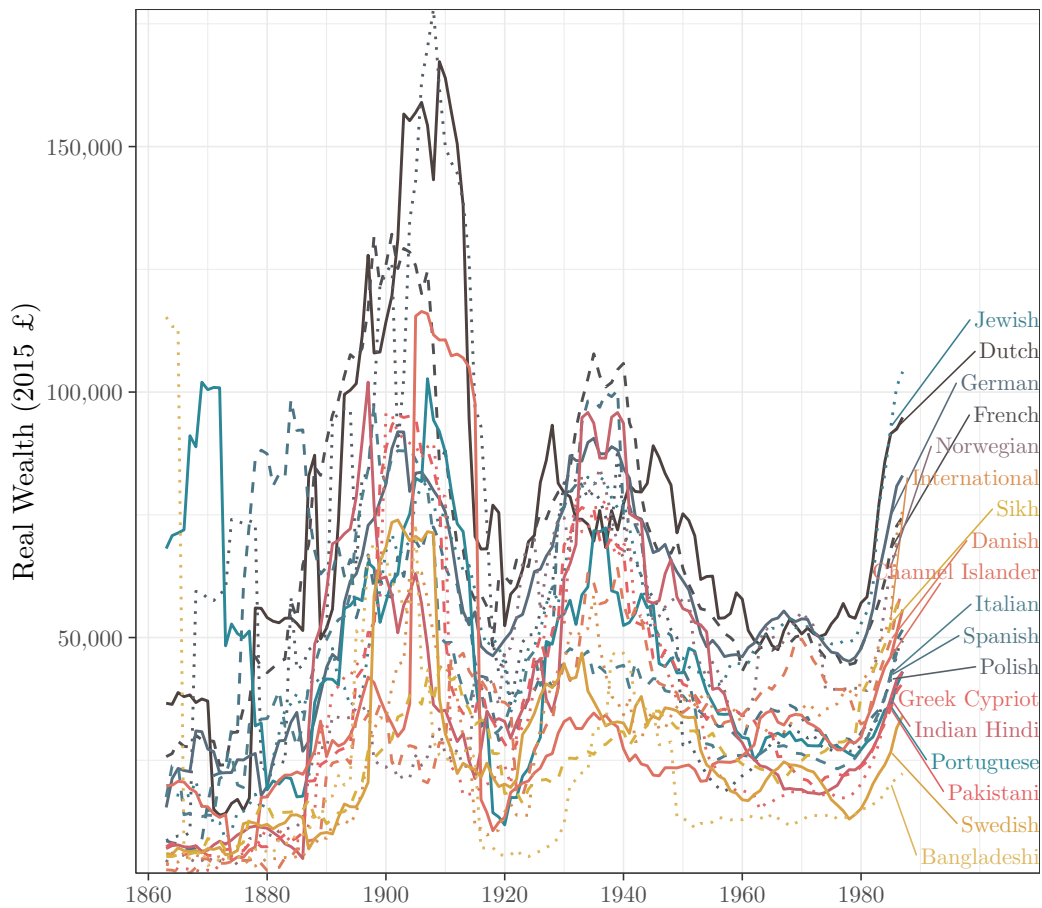
where  $\alpha$  is a constant and  $\varepsilon$  is a random error. Table A16 reported in the appendix, details the results of this estimation for top 25 ethnicities (by number of deaths, 1858-1992), separately for the periods, 1858-79, 1880-1913, 1914-39, 1940-59, 1960-69, 1970-1979 and 1980-1992.

The constant, representing those with English names is estimated at .009 to .011, depending on period. Figure 11 reports the coefficient estimates and the 95% confidence interval for the first and last period, 1858-79, and 1980-92. In 1858-80, the majority of major ethnic groups are much more likely to appear in the top 1% than the English. For some groups, this remains true in the most recent period, 1980-92, but for most the positive ethnic effect on being in the top

<sup>35</sup>Notice that the confidence intervals are very tight for this group as the observations are large; the confidence interval is smaller than the point size in figure 9.



(a) British, Irish and Other



(b) Others

Figure 8: Average Wealth by Ethnicity, 1858-1992

*Note:* Average wealth includes non-wealth holders, who are assigned an inferred wealth. Subfigure B presents the 20 most numerous “other” ethnicities (by number of deaths 1858-92), and is ordered by the level of average wealth in 1992. The figure is difficult to read in specific detail but is meant as illustrative of the broad trend and variation in the underlying data.

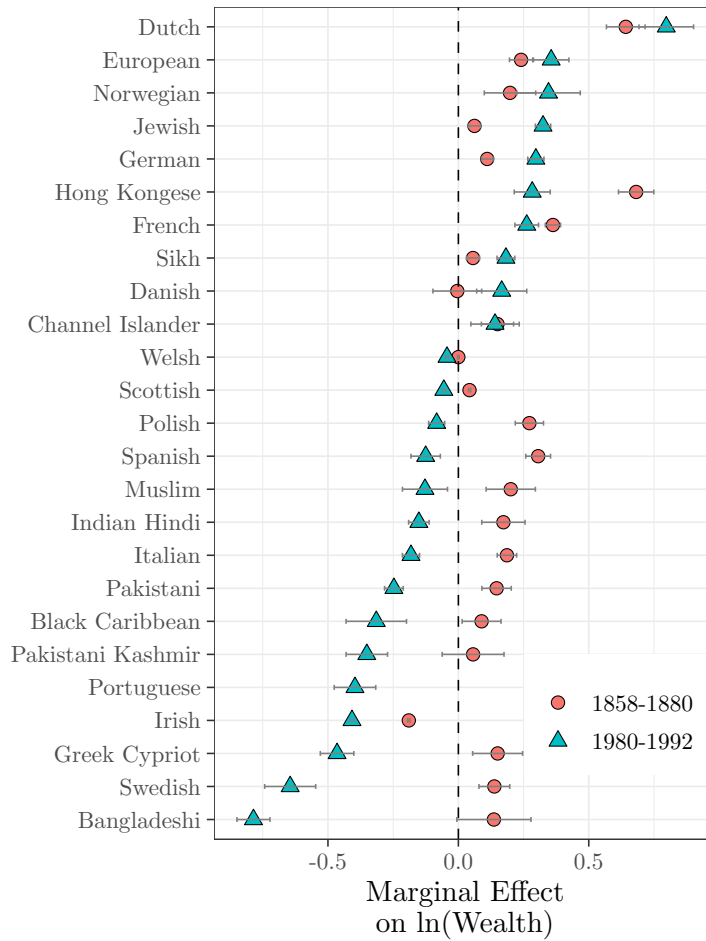
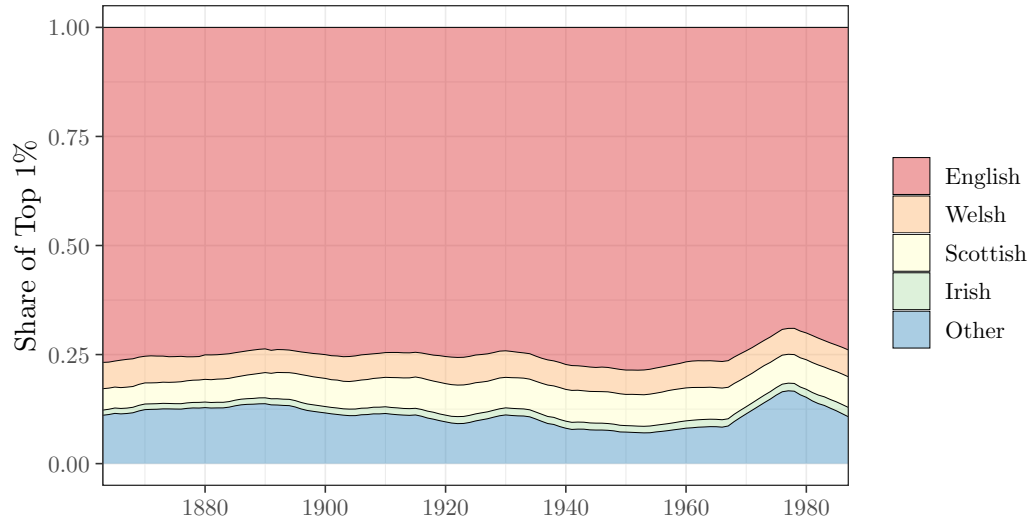
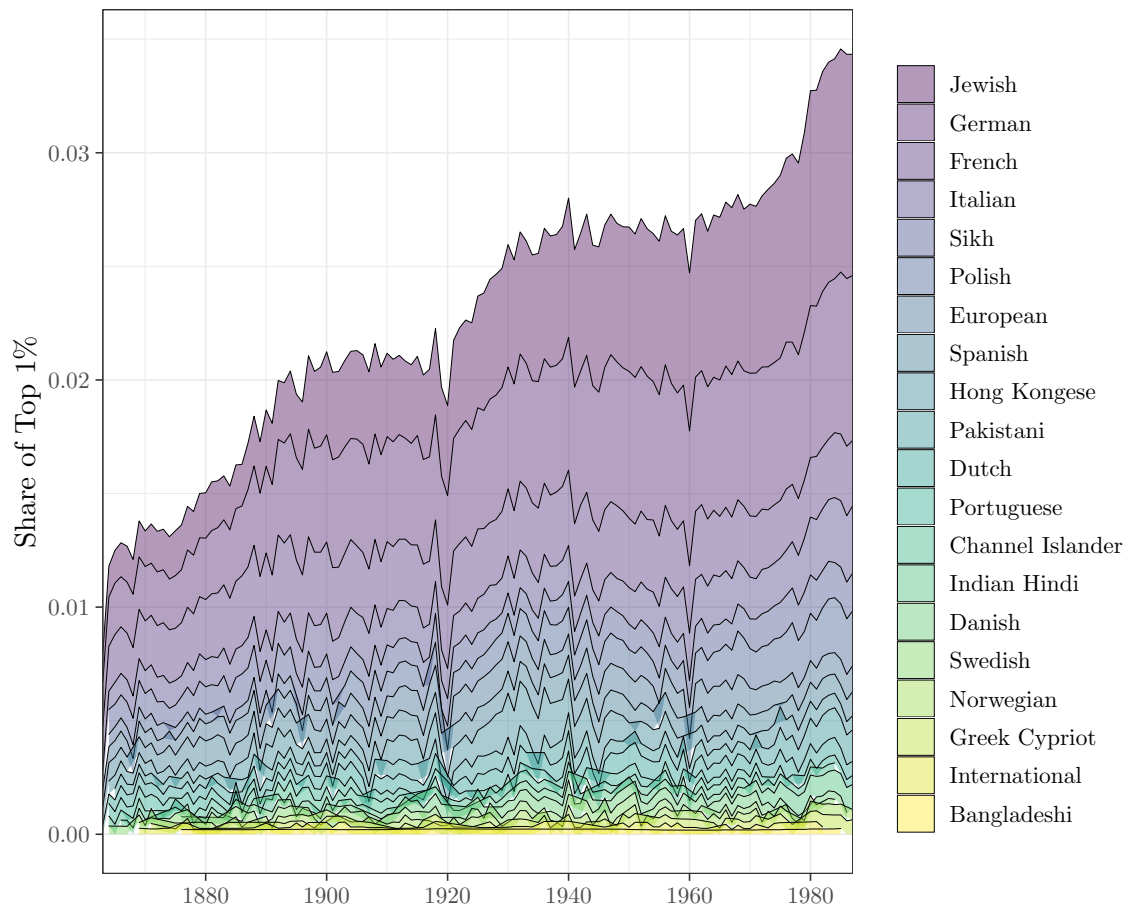


Figure 9: The Correlation of Ethnicity and Real Wealth, 1858-1880 compared with 1980-1992

Notes: Ethnic group coefficients from OLS regression  $\ln(W_j) = \alpha + \sum_{i=1}^n \beta_i D_i^{Eth} + \varepsilon$  where  $W$  is the real wealth of individual  $j$  and ethnic group categorical dummy variables are represented by  $D^{Eth}$  for  $i = 25$  most numerous by total number fo deaths 1858-1992 ethnicity,  $\alpha$  is a constant and  $\varepsilon$  is a random error.



(a) British, Irish and Other



(b) Other Ethnicities

Figure 10: The Ethnic Composition of the Top 1%

*Notes:* The figure reports the ethnic composition of the top 1% of the English wealth-at-death distribution, 1858-1992. I have smoothed the estimates to decadal moving averages in subfigure B figure for readability. Subfigure B presents the 20 most numerous “other” ethnicities (by number of deaths 1858-92), and is ordered by the share of that ethnicity in the top1% of wealth-holders over the sample period. The figure is difficult to read in specific detail but is meant as illustrative of the broad trend and variation in the underlying data.



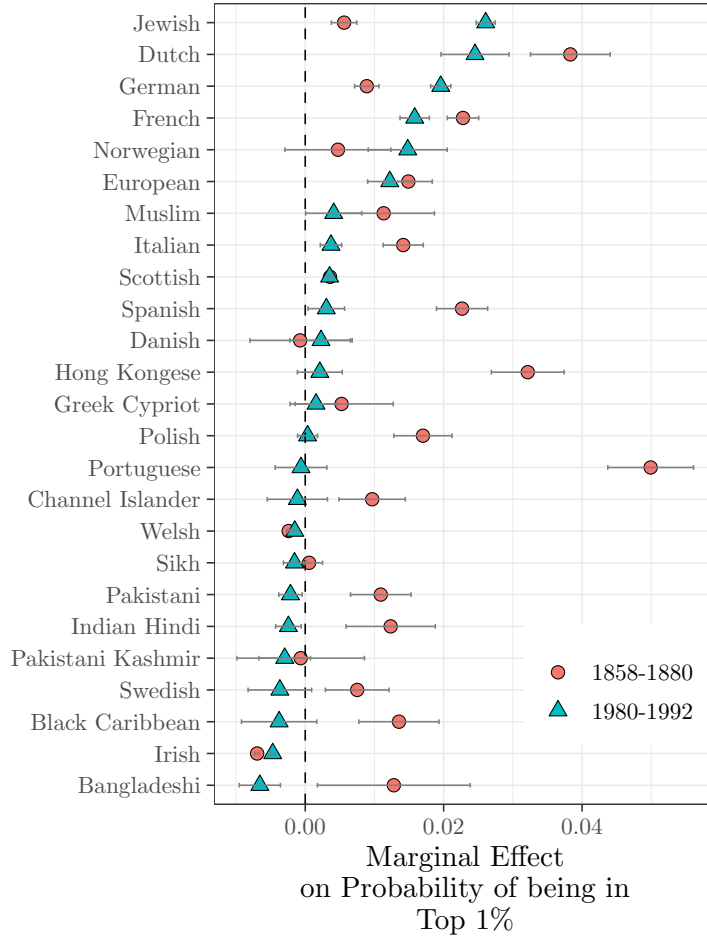


Figure 11: The Marginal Probability of being in the Top 1% of Wealth Holders, by Ethnicity, 1858-1880 compared with 1980-1992

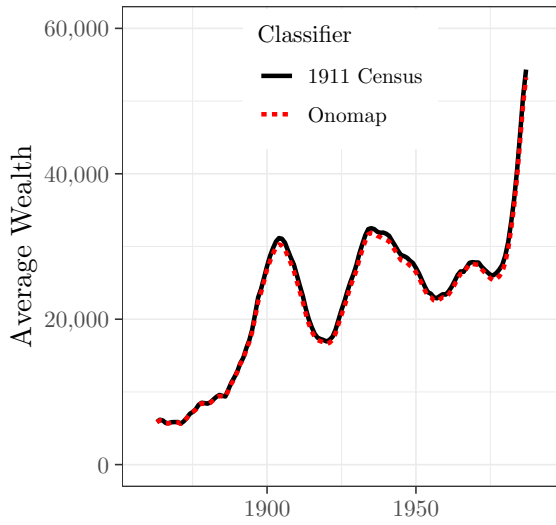
1% has disappeared. Only the Irish and Bangladeshi communities are underrepresented in the top 1%.

It is somewhat of a surprise to find that the measure of wealth inequality that displays the *least* inequality between ethnic groups is representation amongst the very richest of England, the top 1% of wealth holders.

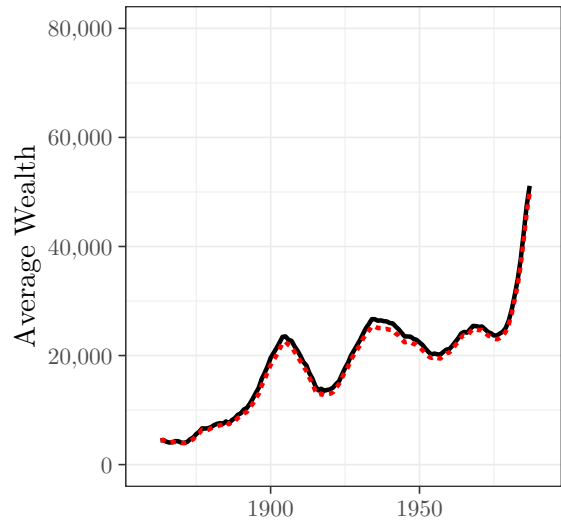
## 5 Results using an Alternative Ethnic Classification

How robust are these patterns to a different method of ethnic classification? Onomap will identify *contemporary* clusters of surnames. As mentioned in section 3, this method may classify fully assimilated ethnic groups as “English” where the forenames used by these ethnic groups are indistinguishable from the native English. Therefore the resulting wealth patterns from this classification may simply show the distinctive patterns of a *subset* of ethnicities that do not assimilate.

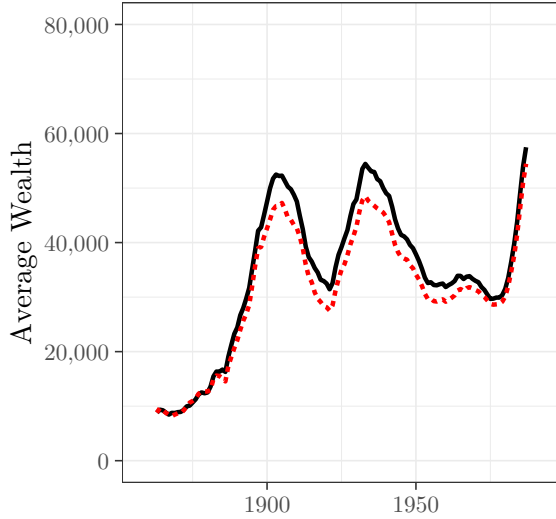
To examine this I use a classification of ethnicity based upon the distribution of place of birth for holders of a surname in the 1911 census of England and Wales. As this classifies ethnicities in 1911, it is not subject to this specific type of potential error as Onomap. The methodology to do this is presented in a companion paper (Cummins and Gráda (2022)), and is summarized in appendix section E.



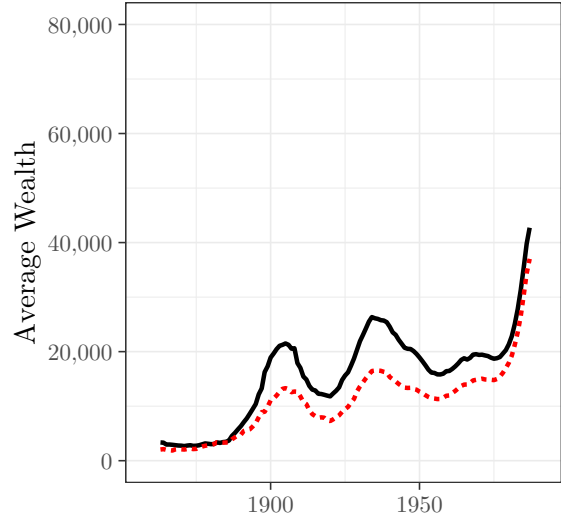
(a) English



(b) Welsh

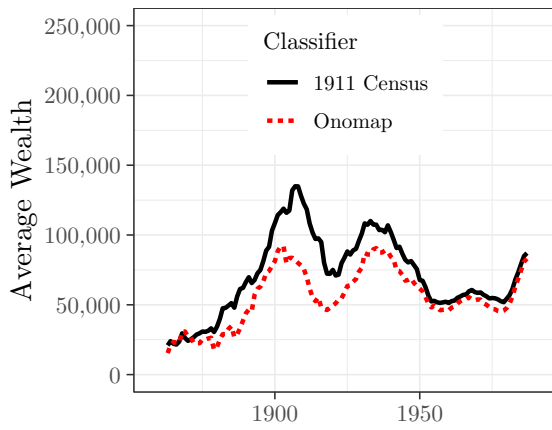


(c) Scottish

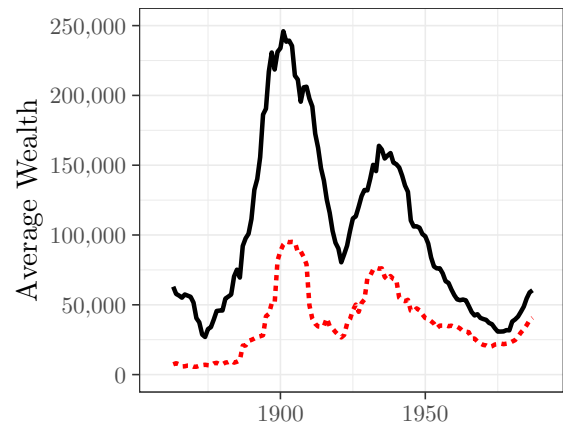


(d) Irish

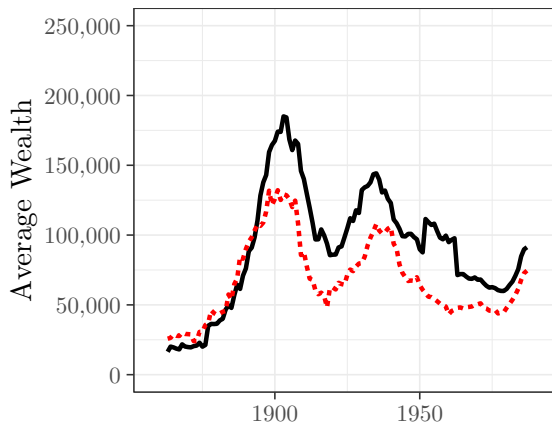
Figure 12: Comparing Average Wealth, 1858-1992, using Alternative Ethnic Classifier, I of II  
*Notes:* Average wealth includes the non-wealth holder population (2015 pounds).



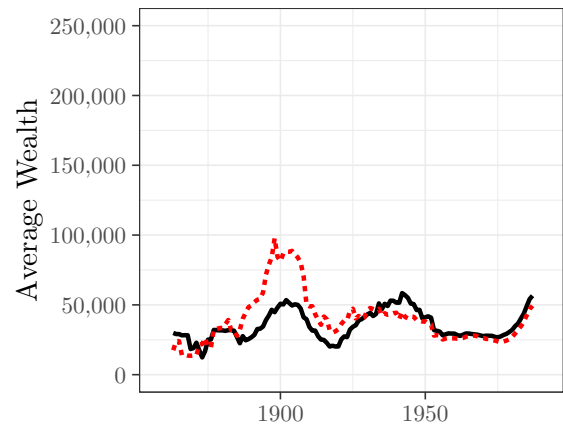
(a) German



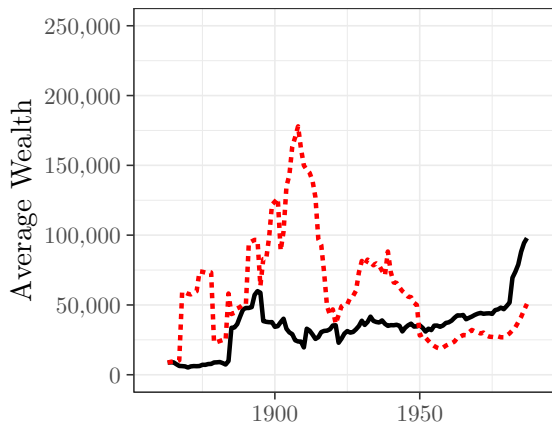
(b) Indian



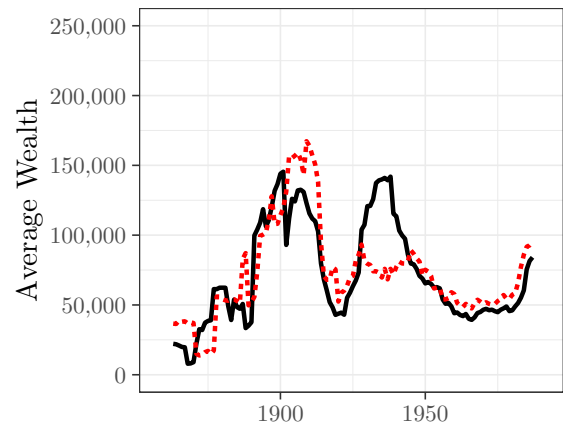
(c) French



(d) Italian

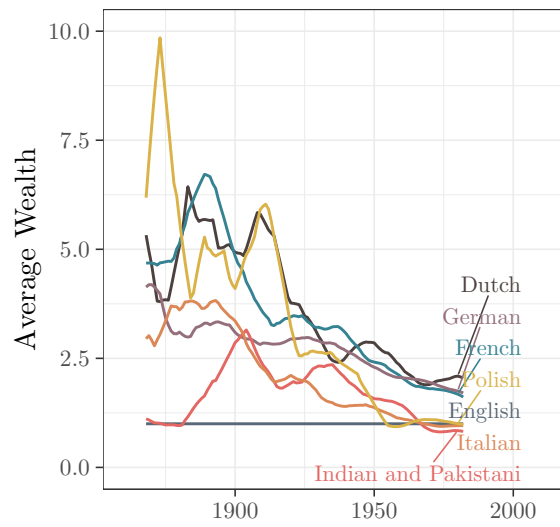
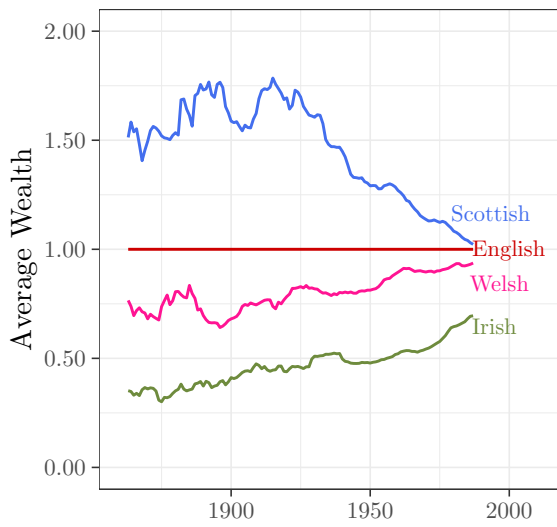


(e) Polish

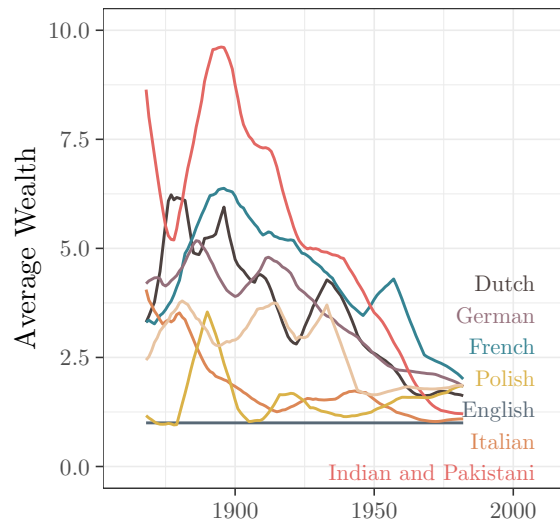
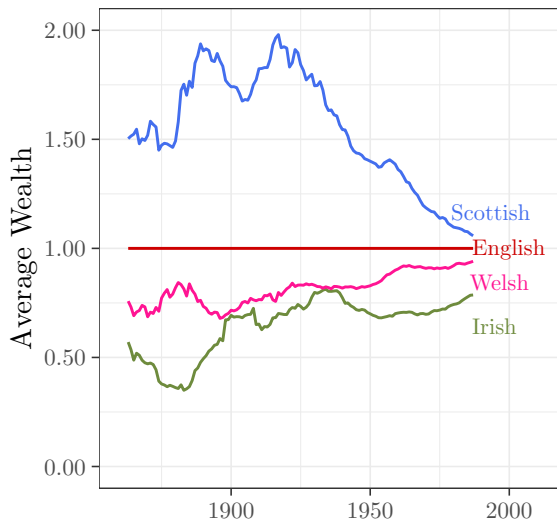


(f) Dutch

Figure 13: Comparing Average Wealth, 1858-1992, using Alternative Ethnic Classifier, II of II  
*Notes:* Average wealth includes the non-wealth holder population (2015 pounds).



(a) Onomap



(b) 1911 Census

Figure 14: Relative Wealth of Ethnicities in England, 1858-1992, using Alternative Classifiers  
*Notes:* All ethnicity wealth estimates are divided by that of the English.

Figure 12 and 13 present for a subset of ten ethnicities a comparison of average wealth, by ethnicity, by year, for both the Onomap classification, and the 1911 classification. The two classifications result in identical wealth estimates for the English, Scottish and Welsh. However for the Irish there is clear indication that Onomap significantly *underestimates* wealth. Here a subset of successful Irish migrants may have adopted English forename customs, and thus become indistinguishable from the English. This effect may also explain the patterns for the German, Indian, French and Dutch ethnicities. Or it could be that earlier migrants from these countries were positively selected (for example on labour productivity, ambition, or human capital). This historical assignment, may itself simply be picking out an ascendant, upwardly mobile, subset of the sending countries populations. Thus the historical classifier would indicate that these groups are now richer than the contemporary classifier which is based on a broader sample of contemporary individuals. Where migration was historically negatively selected, the pattern may be the reverse, and this may explain the patterns for the Italian, and Polish ethnicities.

To what extent do these inconsistencies between the classifiers, matter for the principle findings of this paper? Figure 14 illustrates the relative patterns of ethnic average wealth, by year, 1858 to 1992, for both classifiers. In all subfigures, the English in a given year are set at one. Whilst there are substantial differences in level between the classifiers, the relative position of these ethnicities are unchanged. Thus the overall picture is broadly similar, although with important exceptions (such as the level of the Irish, and the relative position of the Indian and Pakistani group). More specific interrogation of this, by ethnicity, can unearth the historical migration dynamics that give rise to these wealth patterns.

## 6 How Important is Ethnic Ancestry for Wealth?

Having established significant differences in wealth by ethnicity I analyze here how important an individual's ethnicity is for their wealth. Is the majority of inequality explained by the ethnic gradient in wealth? Or is it dwarfed by other factors? And how has this changed over time?

The Theil measure of inequality is the difference between the maximum possible entropy of a wealth distribution and that observed (Theil (1967); Cowell (2009)). Amongst inequality measures it has the advantage of additive decomposability. Here, I use the Theil index to examine the degree to which wealth inequality in England 1858-1992, is *within* ethnic groups, and how much is *between* ethnic groups.

The Theil measure  $T$  of entropy is calculated as

$$T = \frac{1}{N} \sum_{j=1}^N \frac{w_j}{\bar{w}} \ln\left(\frac{w_j}{\bar{w}}\right) \quad (7)$$

where  $N$  is the number of individuals (denoted by  $j$ ) dying in a year and  $w$  is wealth at death (Cowell (2009, p.54)). Figure 15 compares the Theil measure for all English wealth, 1858-1992, with the more widely used Gini coefficient. A value of 0 represents complete equality and the Theil index is bounded at  $\ln(N)$  where  $N$  is the sample size. The entropy based measure does appear to be much more sensitive year to year, and registers a sharper decline, than the Gini coefficient.<sup>36</sup>

The big advantage of the Theil entropy measure is that it can be decomposed into a *between* and *within* group components. The Theil measure for an individual of ethnicity  $i$  is

$$T = \left[ \sum_{i=1}^{N^E} \frac{N_i \bar{w}_i}{\bar{w}} \right] T_i + \left[ \frac{1}{N} \sum_{i=1}^{N^E} N_i \frac{\bar{w}_i}{\bar{w}} \ln\left(\frac{\bar{w}_i}{\bar{w}}\right) \right] \quad (8)$$

where  $N^E$  is the number of ethnicities,  $N_i$  is the number of people sharing that ethnicity (Theil (1967, p.95, eq. 1.9)). The first component of the right hand side of equation 8 corre-

<sup>36</sup>A separate paper (Cummins (2021)) calibrates the Gini coefficient from the PPR data with the Gini coefficient from other studies. It finds that the PPR wealth Gini is constant with exiting estimates.

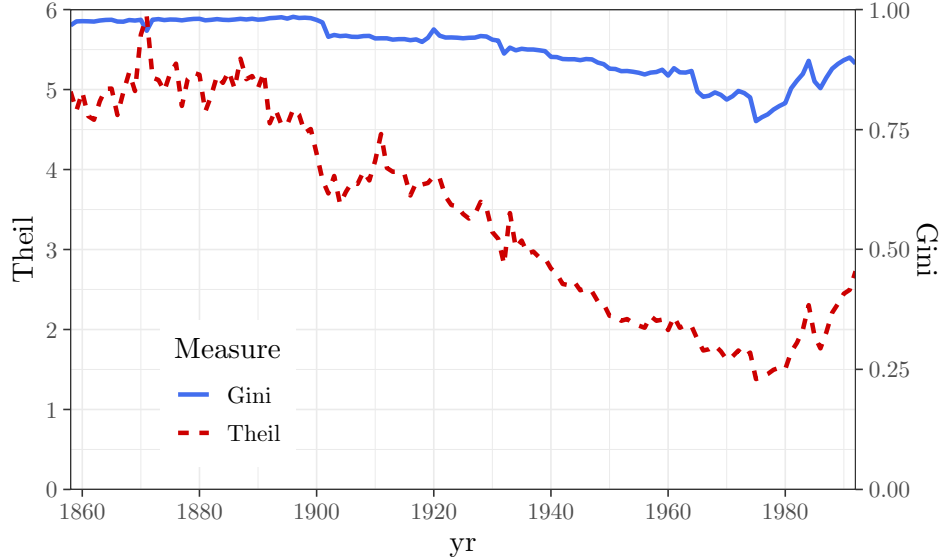


Figure 15: The Theil Index of Inequality compared with the Gini Coefficient, English Wealth, 1858-1992

sponds to the weighted sum of *within* group inequality  $T_i$ , and the second term corresponds to *between* group inequality (Shorrocks (1980, p.613)).

Figure 16 reports the Theil decomposition of wealth inequality into a *between* ethnicity component and a *within* ethnicity component, as equation 8,<sup>37</sup> and figure 16b reports *between* inequality as a percentage of overall Inequality, England 1858-1892.

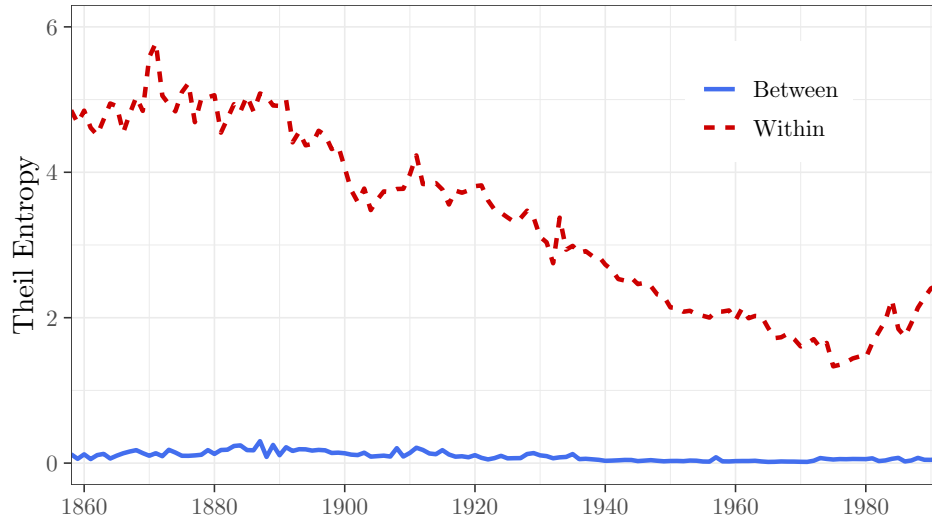
The Theil decomposition indicates that 97.5% of inequality in England from 1858 to 1992 is individual (*within* ethnicity) and about 2.5% is attributable to ethnic group differences (the *between* part). This is may come as a surprise given the large and seemingly systematic differences in the proportion probated and average wealth, by ethnicity, reported in this paper. Also surprisingly, the Theil decomposition indicates that between ethnic group inequality has not changed much 1858 to 1992.

It is easier to understand why the entropy measure attributes 97.5% of inequality to individuals. Figure 17 plots the distribution of real wealth for the major ethnic groups of England, across the sample period. All the distributions overlap substantially. This gives us an intuitive understanding for the apparent contradictory coexistence of substantial and systematically different group averages. For example the Irish in every measure presented in this paper are consistently poorer 1858 to 2018. But in any year knowing someones ethnic group does not provide much information on where they lie within the overall wealth distribution.<sup>38</sup>

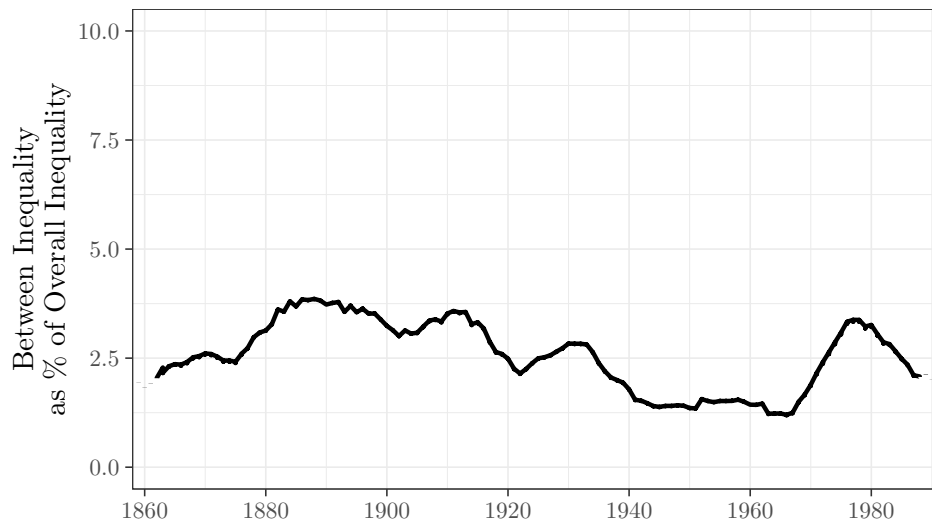
Another way to understand this is to think of a standard OLS regression. It is typically the case in empirical studies using micro data to find substantial and highly statistically significant regression coefficients yet simultaneously have vanishingly small  $R^2$  values. For example, table A15, reported in the appendix, which demonstrates this precise pattern for the PPR wealth data. Here, a set of ethnicity dummies are regressed upon the natural log of real wealth, by period. The coefficients on many of the ethnic groups are large in absolute terms, and highly statistically significant with p values below  $2e^{-16}$ , yet the proportion of variance explained ( $R^2$ ) is never above .005. This is further emphasized by repeating the regression 135 times for every year from 1858 to 1992 separately, retrieving the  $R^2$  and comparing it values over time. This is done in figure 18.

<sup>37</sup>The decomposition was calculated using the “decompGEI” package in R.

<sup>38</sup>Appendix section I tests the sensitivity of this result to different methods for attributing non-probated wealth. In summary, this attribution has no major effect upon the Theil decomposition reported here.



(a) Entropy Measures



(b) Between Proportion

Figure 16: The Theil Index of Inequality, decomposed for between and within ethnic groupings, English Wealth, 1858-1992

Notes: Calculated separately for each of the 135 years 1858-1992 using *decompGEI* in R.

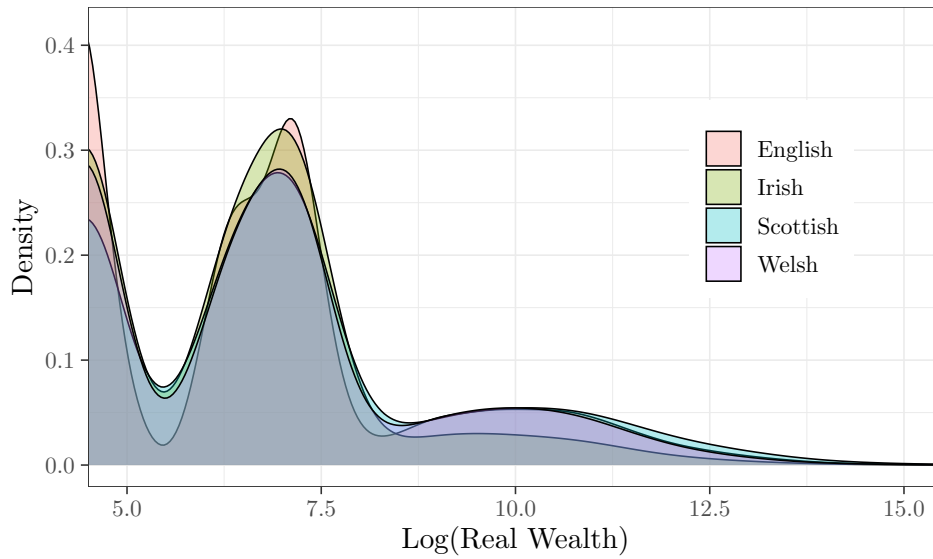


Figure 17: The Distribution of Real Wealth, by Major Ethnic Groups, England 1858-1992

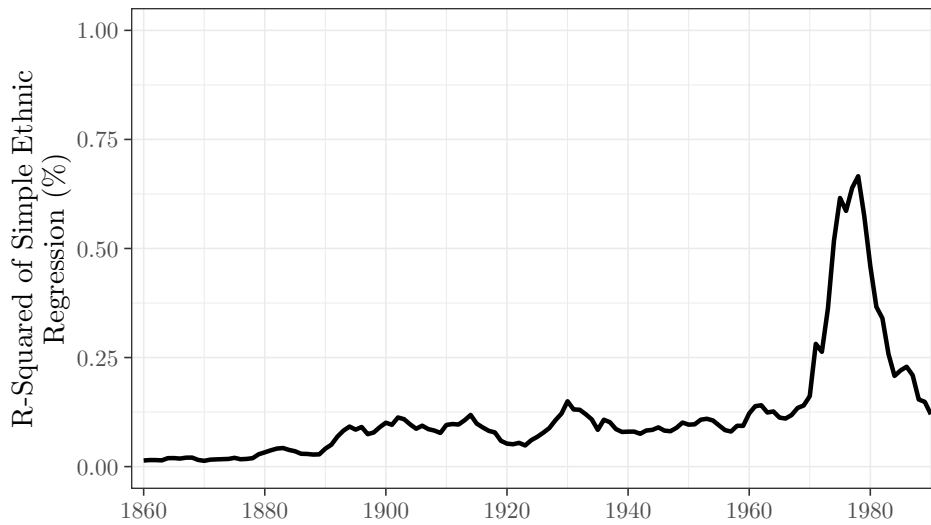


Figure 18: The Proportion of Variance explained by a simple Linear regression of Wealth on a set of Ethnic Dummies, by year, 1858-1992

Source: The  $R^2$  from 135 (one for every year 1858-1992) OLS regressions of  $\ln(W_j) = \alpha + \sum_{i=1}^n \beta_i D_i^{Eth} + \varepsilon$  where  $W$  is the real wealth of individual  $j$  and ethnic group categorical dummy variables are represented by  $D_i^{Eth}$  for  $i = 25$  most numerous by total number of deaths 1858-1992 ethnicity,  $\alpha$  is a constant and  $\varepsilon$  is a random error.



## 7 Discussion

This paper has documented significant ethnic inequalities in wealth, and their change over time. There is a distinctive wealth trajectory for several of England’s ethnic minorities. The modern wealth gradient, as revealed by contemporary surveys, has been revealed to have been of recent emergence. Many of the figures reported in this paper were visually dense and hence it was difficult to track specific ethnic minorities over time. For a sample of ethnicities, figure 19 presents a clearer visualization of average wealth (as reported earlier in section 4.3.2).

British and Irish ethnicity are illustrated in panels a (Welsh), b (Scottish), and c (Irish). The Welsh are on average poorer throughout than the English but this has almost equalized over time. The Scottish are significantly richer on average 1860-1940, and have converged since then to the English average. This is consistent with the popular perception of a ‘brain drain’ from Scotland to England. For Ireland the pattern is the reverse. The Irish are consistently poorer than the English 1858 to 2018. A companion paper explores this in greater detail (Cummins and Gráda (2022)), where it is shown that the Irish also experience significantly higher infant mortality rates. These findings are robust to place, and age at death controls. Thus the finding that the Irish are poorer is not an artifact of return migration.<sup>39</sup>

Panels d, e and f represent the ‘rich’ European migration experience to England of the Jewish, German and French ethnicities. These ethnicities are always richer than the English. However, this advantage has declined consistently for the Germans and the French over the 20th century, and has remained more or less constant for the Jewish ethnicity.

The experience of the Indian sub-continent is displayed in panels g (Sikh), h (Pakistani), and i (Indian Hindi) ethnicities. From 1880 to 1950, the Pakistani and Indian Hindi groups are significantly richer than the English. But this has reversed since about 1970. Both groups are now poorer than the English. The Sikh ethnicity are sometimes richer, but are usually close to the English average.

Panels j (Swedish), k (Bangladeshi) and l (Portuguese) illustrate the trajectory of what are by 1992, the poorest ethnicities in England. Whilst the Bangladeshi experience varies over the 20th century, the Swedish and Portuguese experience is one of secular decline since about 1940.<sup>40</sup>

What explains these patterns? The present study, designed to present an overview of ethnic inequality for all ethnicities over the 1858-2018 period cannot offer any definitive answer to this question. However it is the hope that this quantitative summary provides the context for deeper dives into specific ethnicity experience. The prospect of specific discrimination, in the labour market, and more broadly, within English society, cannot be ruled out as a driver of these patterns. A related paper examines this issue for the Irish (Cummins and Gráda (2022)). That paper also argued that remittances could only explain a small proportion of the Irish-in-England wealth gap relative to the English. Future work exploring the differential rate of remittances across ethnic groups is central to understanding the patterns documented in this paper.

A detailed economic history of a specific ethnic groups experience using Big Data, linked micro data (across generations, and perhaps across countries), and nuanced qualitative sources, would allow us to tease apart the relative roles of migrant selection, education, industrial decline, geography, ‘cultural distance’, language, religion and labour market discrimination in the determination of the socio-economic assimilation of ethnic minorities into host populations. Further, the impact of policies related to housing, immigration, employment, and education on specific ethnicities wealth accumulation should be explored. The present study demonstrates that previously invisible inequities can be uncovered using new methodologies on such novel big data sources.

---

<sup>39</sup>For the flow of Scottish doctors to England see Abel-Smith et al. (1964). Samuel Johnson stated “the noblest prospect which a Scotchman ever sees, is the high road that leads him to England!” Boswell (1791). It was also noted in 1966 in the House of Lords: “Have your Lordships ever thought of the brain drain from Scotland to this country? It went on for two hundred years. It still goes on. The noble Lord, Lord Todd, is a rather fine example in point. Quite out of proportion to the relative populations, the Scots sent us scientists, engineers, and of course all kinds of other eminent professional people.” (Lord Snow) link.

<sup>40</sup>As this analysis does not control for age it could be that those Swedes, or Portuguese, that die in England, are simply those who died young before returning home. Thus the surprising relative poverty of these groups could be a simple artifact of the character of temporary economic movements during the life-course.

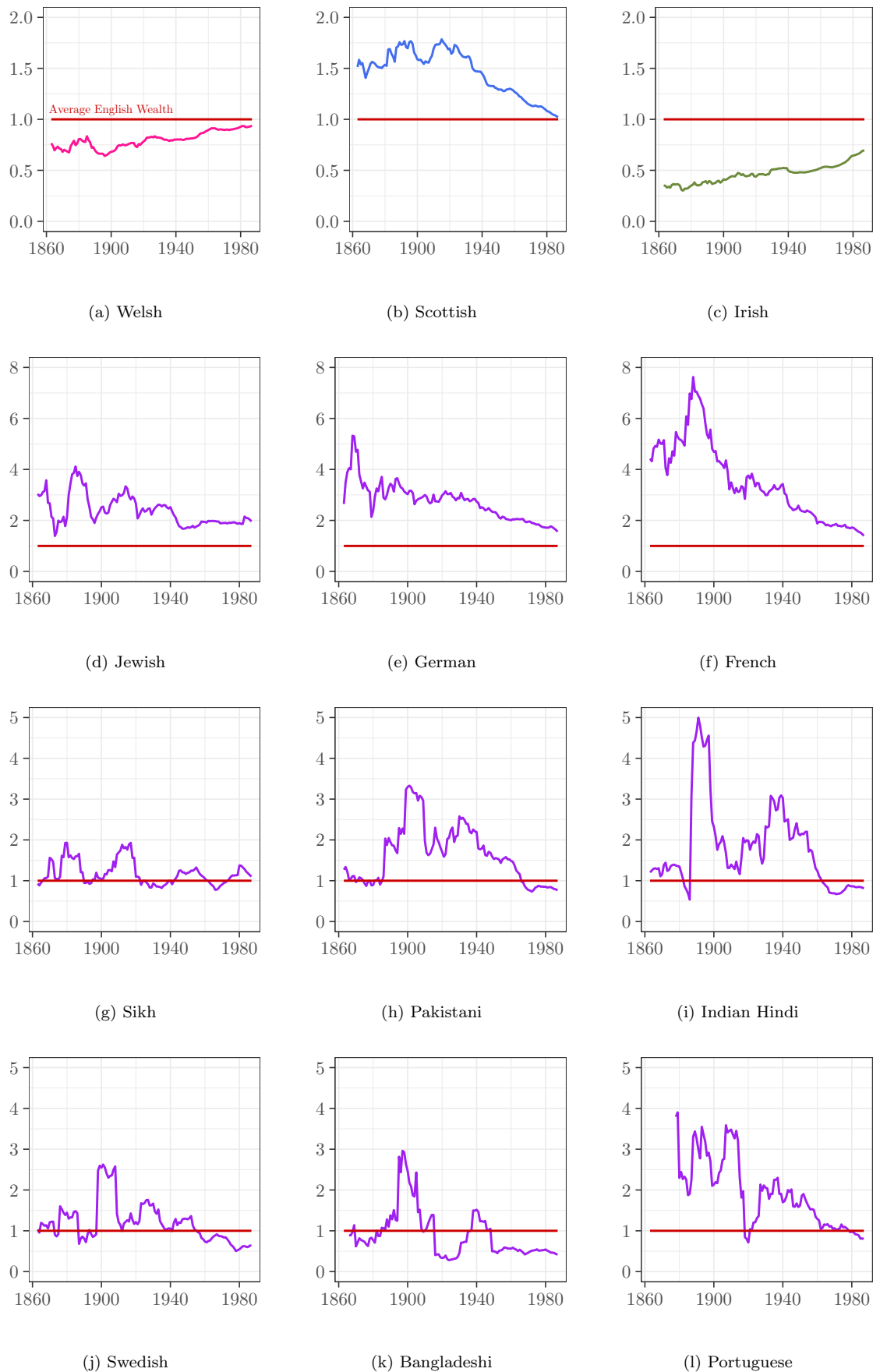


Figure 19: Relative Wealth Inequality, 1858-1992

Notes: In all subfigures, the wealth of those with “English” surnames is set to one. Ethnicities sampled with reference to the number of deaths over the sample period, and representativeness.

## 8 Conclusion

I have presented here the first study of the history of ethnic wealth inequality in England and Wales. The ‘ethnic resolution’ is significantly higher than even the best contemporary data (Office for National Statistics (2019b, 2020)). Through out the past 150 years, non-British Ethnicities have been, in general, an economic elite. After WWII, large variations emerge, and by the 1990s, many ethnicities, such as the Bangladeshi community, are significantly poorer than the English. The Irish are always poorer, and the Scottish richer, until recently. The Irish wealth-gap is consistent, persistent and large. Despite these significant inequities, most inequality is between individuals and not between ethnicities.

What is driving these inequities? The potential roles of migrant selection from the sending countries, human capital and culture, in the assimilation of ethnicities into English society, and the role of systematic discrimination are all mechanism that could be at play. Whilst the current analysis can say nothing definitive on these mechanisms, the presented patterns give us an empirical basis to discuss, and explore further, the causes of ethnic based inequities in life opportunities and outcomes.

## References

- Abel-Smith, B., K. Gales, and G. Macfarlane**, *British Doctors at Home and Abroad*. By B. Abel-Smith and K. Gales, with the Assistance of Gillian Macfarlane [Occasional Papers on Social Administration. no. 8.], Social Administration Research Trust, 1964.
- Alfani, Guido**, *As Gods Among Men: A History of the Rich in the West*, Princeton, NJ: Princeton University Press, 2023.
- Alvaredo, Facundo, Anthony B Atkinson, and Salvatore Morelli**, “Top wealth shares in the UK over more than a century,” 2017.
- , **Anthony B. Atkinson, and Salvatore Morelli**, “Top wealth shares in the UK over more than a century,” *Journal of Public Economics*, 2018, 162, 26 – 47. In Honor of Sir Tony Atkinson (1944-2017).
- Atkinson, Anthony B.**, “Wealth and Inheritance in Britain from 1896 to the Present,” 2013.
- **and Allan J. Harrison**, “Wealth Distribution And Investment Income In Britain,” *Review of Income and Wealth*, 1974, 20 (2), 125–142.
- , **James P. F. Gordon, and Alan Harrison**, “Trends in the shares of top wealth-holders in Britain, 1923-1981,” *Oxford Bulletin of Economics and Statistics*, 1989, 51 (3), 315–332.
- , **Peter G. Backus, and John Micklewright**, “Charitable Bequests and Wealth at Death,” *The Economic Journal*, 10 2017, 127 (605), F1–F23.
- Atkinson, Anthony Barnes and Allan James Harrison**, *Distribution of personal wealth in Britain*, Cambridge Univ Pr, 1978.
- Battu, Harminder and Yves Zenou**, “Oppositional Identities and Employment for Ethnic Minorities: Evidence from England,” *The Economic Journal*, 2010, 120 (542), F52–F71.
- Blake, David and J. Michael Orszag**, “Annual estimates of personal wealthholdings in the United Kingdom since 1948,” *Applied Financial Economics*, 1999, 9 (4), 397–421.
- Boswell, James**, *Boswell’s Life of Johnson* 1791. <https://www.gutenberg.org/files/1564/1564-h/1564-h.htm>.
- Casey, Teresa and Christian Dustmann**, “Immigrants Identity, Economic Outcomes and the Transmission of Identity across Generations,” *The Economic Journal*, 2010, 120 (542), F31–F51.

- Chetty, Raj, Nathaniel Hendren, Maggie R Jones, and Sonya R Porter**, “Race and Economic Opportunity in the United States: an Intergenerational Perspective,” *The Quarterly Journal of Economics*, 12 2019, *135* (2), 711–783.
- Clark, Gregory and Neil Cummins**, “Surnames and Social Mobility in England, 1170–2012,” *Human Nature*, Dec 2014, *25* (4), 517–537.
- **and** – , “Intergenerational Wealth Mobility in England, 1858–2012: Surnames and Social Mobility,” *The Economic Journal*, 2015, *125* (582), 61–85.
- **and** – , “Malthus to modernity: wealth, status, and fertility in England, 1500–1879,” *Journal of Population Economics*, 2015, *28* (1), 3–29.
- Cowell, Frank**, *Measuring Inequality*, London School of Economics Perspectives in Economic Analysis, 2009.
- Cummins, Neil**, “Hidden Wealth,” Technical Report, CEPR Discussion Paper 2019.
- , “Where is the Middle Class? Inequality, Gender and the Shape of the Upper Tail from 60 million English Death and Probate Records, 1892–2016,” *Centre for Economic Policy Research Discussion Paper 13436*, 2019.
- , “Where Is the Middle Class? Evidence from 60 Million English Death and Probate Records, 1892–1992,” *The Journal of Economic History*, 2021, *81* (2), 359–404.
- , “The hidden wealth of English dynasties, 1892–2016,” *The Economic History Review*, 2022, *75*, 667–702.
- **and Cormac Ó Gráda**, “The Irish in England,” *Centre for Economic Policy Research Working Paper DP17439*, 2022.
- Dustmann, Christian and Ian Preston**, “Attitudes to Ethnic Minorities, Ethnic Context and Location Decisions,” *The Economic Journal*, 2001, *111* (470), 353–373.
- **and** – , “Is immigration good or bad for the economy? Analysis of attitudinal responses,” *Research in Labor Economics*, 2006, *24* (1), 3–34.
- **and Tommaso Frattini**, “The Fiscal Effects of Immigration to the UK,” *The Economic Journal*, 11 2014, *124* (580), F593–F643.
- Farkas, Lilla**, “Data collection in the field of ethnicity,” 2017. [https://ec.europa.eu/newsroom/just/document.cfm?action=display&doc\\_id=45791](https://ec.europa.eu/newsroom/just/document.cfm?action=display&doc_id=45791).
- GOV.UK**, “Find a will,” 2018. <https://probatesearch.service.gov.uk/#calendar>.
- , “Dealing with the estate of someone who’s died,” 2020. <https://www.gov.uk/probate-estate>.
- Gymrek, Melissa, Amy L. McGuire, David Golan, Eran Halperin, and Yaniv Erlich**, “Identifying Personal Genomes by Surname Inference,” *Science*, 2013, *339* (6117), 321–324.
- Harbury, C. D.**, “Inheritance and the Distribution of Personal Wealth in Britain,” *The Economic Journal*, 1962, *72* (288), 845–868.
- House of Commons Library**, “Probate Fees,” 2019. <http://researchbriefings.files.parliament.uk/documents/CBP-7929/CBP-7929.pdf>.
- Jobling, Mark A.**, “In the name of the father: surnames and genetics,” *Trends in Genetics*, 2001, *17*, 353–7.
- Kandt, Jens and Paul A. Longley**, “Ethnicity estimation using family naming practices,” *PLOS ONE*, 08 2018, *13* (8), 1–24.

- Karagiannaki, Eleni**, “Recent Trends in the Size and the Distribution of Inherited Wealth in the UK,” *Fiscal Studies*, 2015, *36* (2), 181–213.
- King, Turi E. and Mark A. Jobling**, “Founders, Drift, and Infidelity: The Relationship between Y Chromosome Diversity and Patrilineal Surnames,” *Molecular Biology and Evolution*, 2009, *26* (5), 1093–1102.
- , **Stéphane J. Ballereau, Kevin E. Schürer, and Mark A. Jobling**, “Genetic Signatures of Coancestry within Surnames,” *Current Biology*, 2006, *16* (4), 384 – 388.
- Kuhn, Moritz, Moritz Schularick, and Ulrike I. Steins**, “Income and Wealth Inequality in America, 1949–2016,” *Journal of Political Economy*, 2020, *128* (9), 3469–3519.
- Lakha, F., D.R. Gorman, and P. Mateos**, “Name analysis to classify populations by ethnicity in public health: Validation of Onomap in Scotland,” *Public Health*, 2011, *125* (10), 688 – 696.
- Lasker, Gabriel Ward**, *Surnames and Genetic Structure*, Cambridge University Press, 1985.
- Lydall, H. F. and D. G. Tipping**, “The Distribution of Personal Wealth in Britain,” *Bulletin of the Oxford University Institute of Economics & Statistics*, 1961, *23* (1), 83–104.
- Manacorda, Marco, Alan Manning, and Jonathan Wadsworth**, “The Impact Of Immigration On The Structure Of Wages: Theory And Evidence From Britain,” *Journal of the European Economic Association*, 2012, *10* (1), 120–151.
- Mateos, Pablo**, “A review of name-based ethnicity classification methods and their potential in population studies,” *Population, Space and Place*, 2007, *13* (4), 243–263.
- , **Paul A. Longley, and David O’Sullivan**, “Ethnicity and Population Structure in Personal Naming Networks,” *PLOS ONE*, 09 2011, *6* (9), 1–12.
- Milanovic, Branko**, *Global Inequality: A New Approach for the Age of Globalization*, Cambridge, MA: Harvard University Press, 2016.
- Morris, Stephen, Matthew Sutton, and Hugh Gravelle**, “Inequity and inequality in the use of health care in England: an empirical investigation,” *Social Science & Medicine*, 2005, *60* (6), 1251 – 1266.
- Nicholas, Tom**, “Clogs to Clogs in Three Generations? Explaining Entrepreneurial Performance in Britain Since 1850,” *The Journal of Economic History*, 1999, *59* (3), 688–713.
- Office for National Statistics**, “Estimates of the population for the UK, England and Wales, Scotland and Northern Ireland,” 2018. [www.ons.gov.uk](http://www.ons.gov.uk): link.
- , “Deaths by single year of age tables - UK,” 2019.
- , “Ethnicity pay gaps in Great Britain: 2018,” 2019.
- , “Household wealth by ethnicity, Great Britain: April 2016 to March 2018,” 2020.
- , “Vital statistics in the UK: births, deaths and marriages,” 2021.
- Perkin, Harold**, “The Recruitment of Elites in British Society since 1800,” *Journal of Social History*, 1978, *12* (2), 222–234.
- Piazza, A., S. Rendine, G. Zei, A. Moroni, and L. L. Cavalli-Sforza**, “Migration rates of human populations from surname distributions,” *Nature volume*, 1987, *329*, 714–716.
- Piketty, Thomas**, *Capital in the Twenty-First Century*, Harvard University Press, 2014.

- Rothery, Mark**, “The Wealth of the English Landed Gentry, 1870-1935,” *The Agricultural History Review*, 2007, 55 (2), 251–268.
- Rubinstein, W. D.**, “Wealth, Elites and the Class Structure of Modern Britain,” *Past & Present*, 1977, (76), 99–126.
- Rubinstein, William D.**, “The Victorian Middle Classes: Wealth, Occupation, and Geography,” *The Economic History Review*, 1977, 30 (4), 602–623.
- , *Men of property: The very wealthy in Britain since the industrial revolution*, Croom Helm London, 1981.
- Sá, Filipa**, “Immigration and House Prices in the UK,” *The Economic Journal*, 09 2014, 125 (587), 1393–1424.
- Scheidel, Walter**, *The Great Leveler: Violence and the History of Inequality from the Stone Age to the Twenty-First Century*, Princeton University Press, 2018.
- Schurer, Kevin and Edward Higgs**, “Integrated Census Microdata (I-CeM) Names and Addresses, 1851-1911: Special Licence Access. [data collection].” 2021. 2nd Edition. UK Data Service. SN: 7856.
- and **Matthew Woollard**, “1881 Census for England and Wales, the Channel Islands and the Isle of Man (Enhanced Version) [computer file],” 2000.
- Shorrocks, A. F.**, “The Class of Additively Decomposable Inequality Measures,” *Econometrica*, 1980, 48 (3), 613–625.
- Smith, Lesley, Paul Norman, Melpo Kapetanstrataki, Sarah Fleming, Lorna K Fraser, Roger C Parslow, and Richard G Feltbower**, “Comparison of ethnic group classification using naming analysis and routinely collected data: application to cancer incidence trends in children and young people,” *BMJ Open*, 2017, 7 (9).
- Theil, Henri**, *Economics and information theory.*, Amsterdam: North-Holland, 1967.
- Turner, John D.**, “Wealth concentration in the European periphery: Ireland, 1858-2001,” *Oxford Economic Papers*, 2010, 62 (4), 625–646.
- Vallejo, Jody Agius and Lisa A. Keister**, “Immigrants and wealth attainment: migration, inequality, and integration,” *Journal of Ethnic and Migration Studies*, 2019, 0 (0), 1–17.
- Wedgwood, J.**, “The Influence of Inheritance on the Distribution of Wealth,” *The Economic Journal*, 1928, 38 (149), 38–55.

# Online Appendix to Ethnic Wealth Inequality in England and Wales, 1858-2018

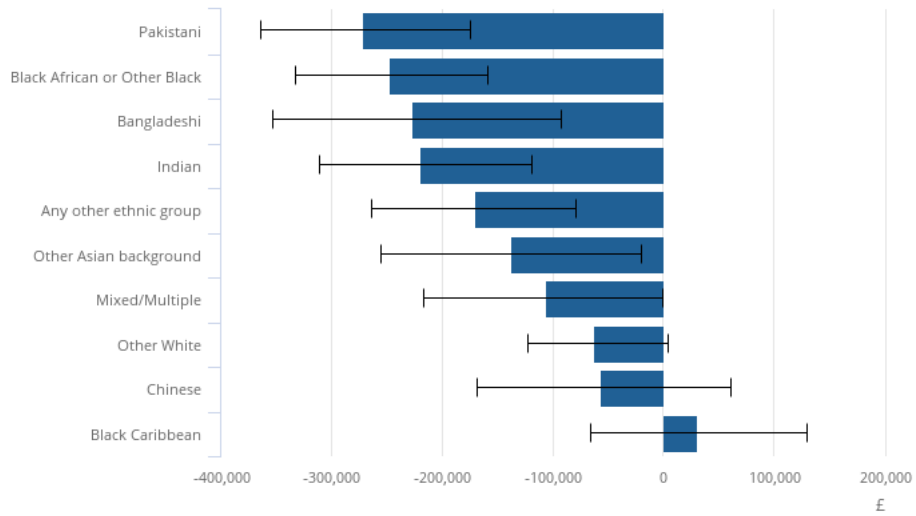
Neil Cummins

This appendix reports a set of extra background material, empirical investigations and results that supplement the main paper.

Figure A20 reports a snapshot of contemporary ethnic wealth inequality in Great Britain from the Office for National Statistics (Office for National Statistics (2020)). Figure A21 reports a set of comparisons of the individual Probate wealth data used in the main paper with existing wealth estimates from the literature. Figure A22 reports the number of individual vital records collected with official sources. Figure A23 reproduces figure 4 (p.19) from Kandt and Longley (2018), which reports the prediction accuracy of the Onomap ethnic classifier for surnames.

Figure 1: Total household wealth in the White British group is estimated to be higher than many other ethnic groups, after adjusting for several demographic and socio-economic characteristics

Estimated difference in total household wealth compared with White British, by ethnicity of the household head when controlling for other factors; Great Britain, April 2016 to March 2018

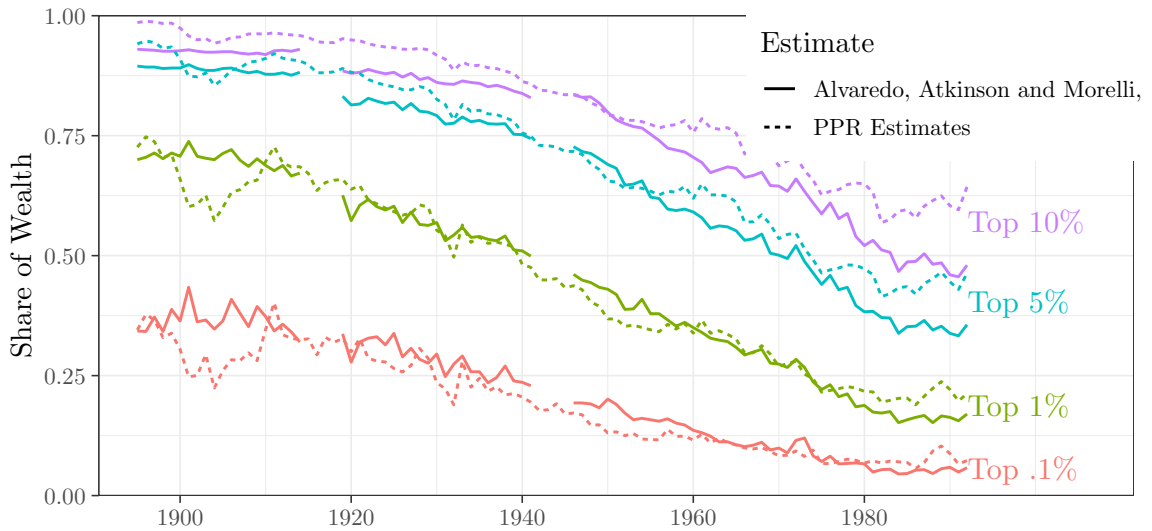


Source: Office for National Statistics – Wealth and Assets Survey

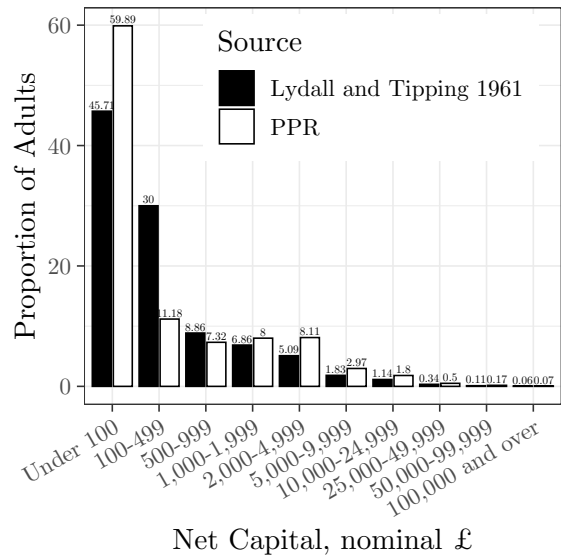
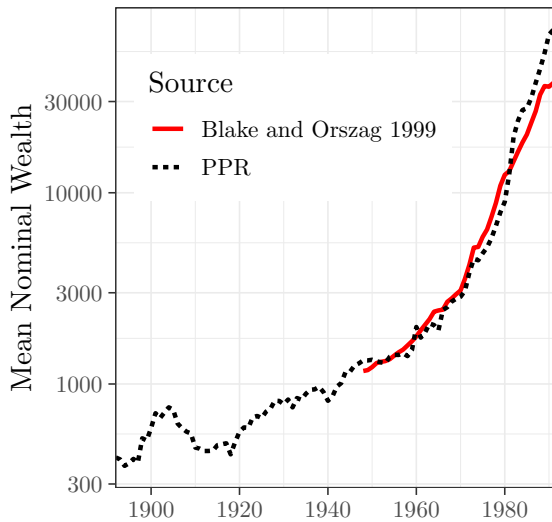
Figure A20: Reproduction of ONS 2020a, Figure 1 “Estimated difference in total household wealth compared with White British, by ethnicity of the household head when controlling for other factors; Great Britain, April 2016 to March 2018”

Source: Office for National Statistics (2020, Figure 1).





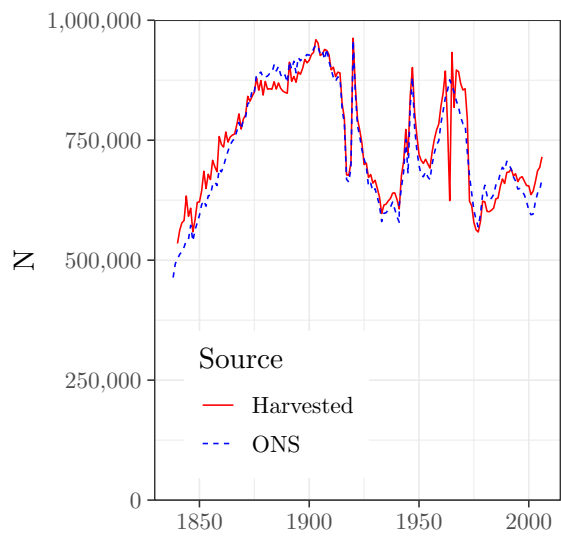
(a) Comparing Different Estimates of Top Wealth Shares, England 1892-1992



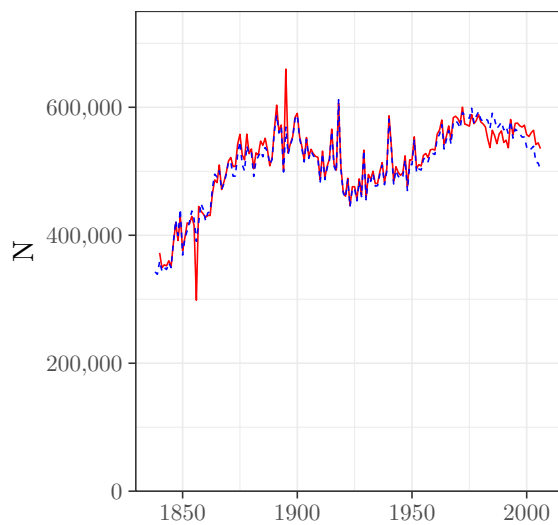
(b) Comparing Average Wealth with Blake and Orszag (1999) (c) Comparison of Net Capital with Lydall and Tipping (1961), by Wealth Band, 1950s

### Figure A21: The PPR Calendar Wealth Data, Compared with Existing Estimates

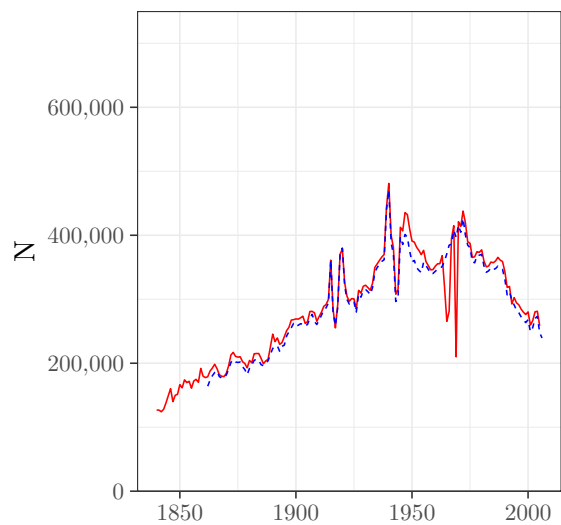
Notes: See Cummins (2021) for a detailed account of the source, construction and validation of the PPR data. Sources: PPR wealth data, Alvaredo et al. (2017) table D1, Blake and Orszag (1999, Table 12) (sum of columns 'net financial wealth', 'housing wealth' and 'consumer durable assets'). These aggregate sums were converted to a per adult measure using population data from Office for National Statistics (2018). Source for figure c: Lydall and Tipping (1961, p.89). Note that the PPR covers England, the Lydall and Tipping (1961) estimates cover Great Britain. Both estimates exclude pension wealth. These figures are also reported in Cummins, 2021.



(a) Births



(b) Deaths



(c) Marriages

Figure A22: Data Collection Verification, 'Harvested' versus Official Count Comparison  
*Notes:* The source for the Official Counts is Office for National Statistics (2021).

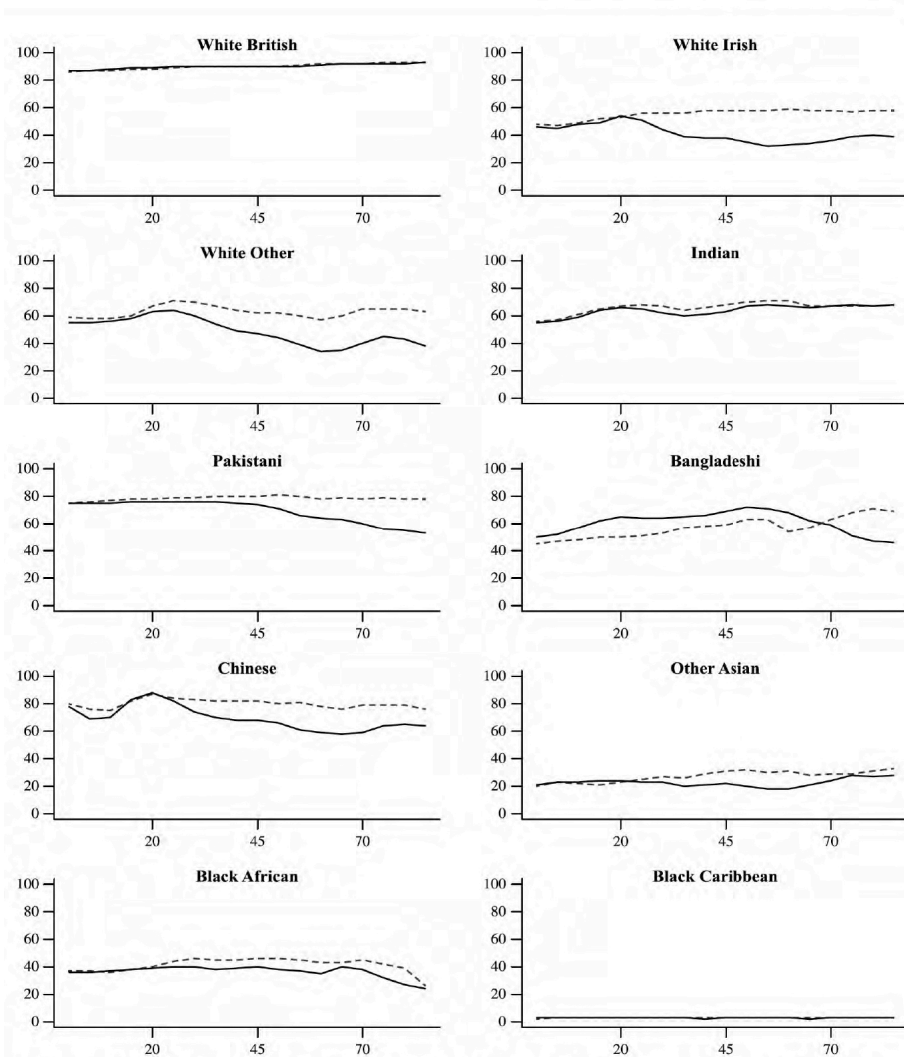


Fig 4. Prediction accuracy for women (solid line) and men (dashed line) by age in each self-ascribed ethnic category (the small Arab/Mixed/Other category is not shown).

<https://doi.org/10.1371/journal.pone.0201774.g004>

Figure A23: The Accuracy of Onomap Ethnicity Classifier, from the 2011 Census  
 Source: Reproduction of Kandt and Longley (2018, Figure 4, p.9).

Table A3: Top 30 Onomap Name Types, Great Britain, 2004

Onomap Type	Population	Names	Percentage
English	31,118,965	122,616	67.3
Scottish	4,749,864	12,980	10.3
Irish	3,172,876	8,342	6.9
Welsh	3,065,041	6,596	6.6
Pakistani	508,699	5,538	1.1
Indian Hindi	319,677	2,680	0.7
Sikh	283,657	3,792	0.6
Italian	229,931	11,377	0.5
Northern Irish	223,988	340	0.5
Norwegian	186,375	895	0.4
Bangladeshi	179,401	1,976	0.4
Polish	155,743	9,102	0.3
German	129,190	7,053	0.3
French	125,754	4,983	0.3
Hong Kongese	119,566	1,683	0.3
Unknown	110,023	3,235,452	0.2
Cornish	107,068	497	0.2
Muslim	103,514	9,509	0.2
Pakistani Kashmir	91,472	1,796	0.2
Nigerian	88,243	4,761	0.2
Portuguese	86,930	4,807	0.2
Jewish	80,522	1,508	0.2
Spanish	80,180	13,094	0.2
Greek Cypriot	79,304	4,248	0.2
India North	75,282	1,863	0.2
Sri Lankan	53,919	2,126	0.1
Turkish	50,706	2,375	0.1
Ghanaian	46,095	3,029	0.1
Celtic	45,653	2,925	0.1
Somalian	33,260	1,081	0.1

*Source:* ONOMAP

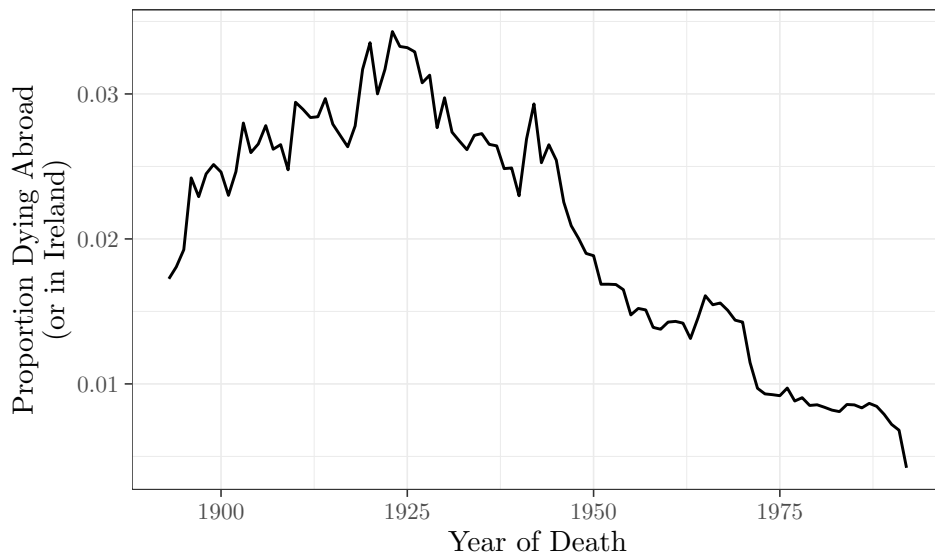


Figure A24: The Proportion of the PPR Calendar Entries Dying Abroad, 1892-1992  
*Notes:* Calculated by cross referencing the last two strings of the address entry in the individual PPR calendars with a list of country names, historical country names, continents, alternative country names, and Irish counties and towns.

## A The Proportion with ‘Significant Wealth’, 1996-2018

The PPR Calendar data was supplemented by a database of the number of deaths and the number of probates, by surname, 1996-2018. Every probate over this period is listed, by name, on <https://probatesearch.service.gov.uk/#calendar>. It was necessary to enter an exact surname on the webpage to return the count of that surname for a given year. From a 100% sample of the 1881 census ((Schurer and Woollard, 2000)) and the 100% samples of births, marriages and deaths, 1838-2007, and the probate Calendar 1892-1992, a master-list of 3,535,375 surnames was created. Of these surnames many were mistakes so a second list was created filtering the master list by the criteria that the name appeared at least 5 or more times in the death registers, 1983-2007. This resulted in 92,812 surnames which were searched individually for every year 1996 to 2020, a total of 2,320,300 searches for each of the 25 years. (As the probate process can take a few months to a year and those years are this incomplete, I do not report the post 2018 rate here.) Each surname from this master-list was entered into <https://probatesearch.service.gov.uk/#calendar> and the count recorded (GOV.UK, 2018).

As reported in table A4 the threshold estate value above which probate was legally required has been £5,000 from 1984 to today, 2020. In recent years however, the *de facto* reality is that financial institutions have exercised discretion in releasing monies to relatives and beneficiaries from the bank accounts of the recently deceased. In 2020, banks apply their own discretion upon which accounts need probate and which don’t. The value they apply as their probate limit could range from £5,000 to £50,000.<sup>41</sup>

It is not clear from existing academic literature or the archives of official Govt. websites advising on probate (<https://webarchive.nationalarchives.gov.uk>) when exactly the nominal probate went from being a flat £5,000 across all institutions, to a discretionary amount that varies in the range £5-£50 thousand, and is institution specific. In 2007-8 (see Atkinson et al. (2017, F8) and as late as 2010 (See Karagiannaki (2015, p.187)), there is evidence that the £5,000 probate threshold was generally applied.<sup>42</sup>

Before 1994, at least, and probably until at least 2010, the assumption that the non-probated estates were worth precisely less than £5,000 appears to be well justified. However, for post-

<sup>41</sup>The current official Government advice on probate states “Contact each asset holder (for example a bank or mortgage company) to find out if you’ll need probate to get access to their assets. Every organization has its own rules.” GOV.UK (2020). A list of the institution specific probate limits are reported here: <https://www.co-oplegalservices.co.uk/media-centre/articles-may-aug-2018/bank-limits-for-probate/>. A news article from 1994 states “Although the Act does not specifically apply to banks and to building societies, they usually apply their discretion in a similar way, and will normally only pay out above the pounds 5,000 limit with a grant of probate.” <https://www.independent.co.uk/news/business/why-the-bereaved-must-wait-rules-governing-the-release-of-money-when-a-person-dies-can-cause-1420519.html>. A 2017 blog post by a probate professional <https://www.todayswillsandprobate.co.uk/guest-writers/obtaining-up-to-50k-without-grant-probate/> discusses the change.

<sup>42</sup>Atkinson et al. (2017) state “We have been told by Her Majesty’s Revenue & Customs (HMRC) that the ‘small estate’ category probably accounts for the large majority of estates that do not go through probate ” (p.F8).

Years	Nominal Probate Threshold	Source
1858-1900	£10	Turner 2010 p.628
1901-1931	£50	Turner (2010) p.628
1932-1964	£100	Atkinson and Harrison (1978) p.36
1965-1974	£500	Atkinson and Harrison (1978) p.36
1975-1984	£1,500	Atkinson and Harrison (1978) p.36
1984→	£5,000	Turner (2010) p.628, Alvaredo et al. (2018) p.29 Atkinson et al. (2017) p.F8, Karagiannaki (2015) p.187

Table A4: The Minimum Probate Threshold, 1858-2017

2010, and in particular more recent years, this assumption is not reasonable. Therefore, we can only interpret the probate rate 1996-2018 as being an indicator of wealth that was significant enough for the asset holders (e.g the banks or building societies) to demand an act of probate before transferring the monies. As this could be anywhere between £5,000-50,000, the probate rate after 2010 can only be interpreted as a measure of significant wealth, and not wealth above the legal probate threshold. So I report this measure separately here and do not include it in the main analysis.

Table A5 reports the count of probates, the sum of adult deaths and the proportion probated from 1996 to 2020. The proportion of adult deaths (deaths to those aged 20 and above) requiring an act of probate to deal with their financial assets at death is consistently around 50%. (Note that the 45-47% recorded in 2016-8 may be underestimated due to the lag in recording probates.) This is consistent with the calculations of Karagiannaki (2015) in her analysis of inherited wealth, who estimates a proportion probated of about 50% for the period 2002-2007 (p.187). A figure of 50% is also reported for 2016 in House of Commons Library (2019, p.7).

Table A5: Proportion Probated, 1996-2018

	$N_{Probates}$	$N_{AdultDeaths}$	Prop. Probated
1996	266,236	556,003	0.48
1997	270,153	551,125	0.49
1998	267,581	546,765	0.49
1999	268,320	546,980	0.49
2000	260,342	531,734	0.49
2001	257,968	526,436	0.49
2002	258,379	529,468	0.49
2003	261,600	533,201	0.49
2004	250,165	508,443	0.49
2005	251,295	507,230	0.50
2006	246,889	496,696	0.50
2007	247,885	498,258	0.50
2008	250,171	503,390	0.50
2009	242,546	485,806	0.50
2010	246,748	488,040	0.51
2011	240,566	479,335	0.50
2012	248,151	494,422	0.50
2013	249,000	502,187	0.50
2014	242,478	496,853	0.49
2015	250,743	525,073	0.48
2016	242,379	520,610	0.47
2017	248,864	528,838	0.47
2018	241,124	537,228	0.45

Source: Office for National Statistics (2019a) and [probatesearch.service.gov.uk](http://probatesearch.service.gov.uk)

## B Ethnic Composition of Deaths in 2006, Top 50

Table A6 reports the top 50 ethnicities, by total number of deaths, of the ethnic for all deaths in England and Wales in 2006.

Table A6: Ethnic Composition of Deaths, England and Wales, 2006

1	English	403,211	74.94
2	Welsh	40,699	7.56
3	Scottish	31,418	5.84
4	Irish	28,199	5.24
5	Unknown	10,070	1.87
6	Pakistani	2,655	0.49
7	Polish	2,165	0.40
8	Sikh	2,120	0.39
9	Italian	1,920	0.36
10	Indian Hindi	1,792	0.33
11	German	1,519	0.28
12	Jewish	1,447	0.27
13	Bangladeshi	1,115	0.21
14	Spanish	858	0.16
15	French	769	0.14
16	Greek Cypriot	587	0.11
17	Hong Kongese	585	0.11
18	Pakistani Kashmir	570	0.11
19	Portuguese	530	0.10
20	Muslim	421	0.08
21	European	387	0.07
22	Nigerian	314	0.06
23	India North	290	0.05
24	Turkish	232	0.04
25	Ghanaian	231	0.04
26	Balkan	221	0.04
27	Black Caribbean	204	0.04
28	Danish	199	0.04
29	Muslim Indian	178	0.03
30	Sri Lankan	178	0.03
31	Castillian	162	0.03
32	Channel Islander	155	0.03
33	Dutch	155	0.03
34	Swedish	140	0.03
35	Hungarian	139	0.03
36	Vietnamese	113	0.02
37	Somalian	111	0.02
38	Hindi Not Indian	107	0.02
39	Norwegian	98	0.02
40	Maltese	91	0.02
41	International	90	0.02
42	Greek	80	0.01
43	Nordic	74	0.01
44	Chinese	67	0.01
45	Hispanic	65	0.01
46	Afrikaans	64	0.01
47	Brazilian	62	0.01
48	Ukranian	56	0.01
49	Catalan	52	0.01
50	Czech	50	0.01

Source: ONOMAP and 100% Death Records



## C The Most Numerous Names Dying in England, 1858-2007 by Ethnicity

Here I inspect the most common surnames, as measured by the number of deaths to each surname, 1858-1992, in each major ethnicity category. Table A7 reports the top 10 most frequent English names dying in England 1858 to 1992, table A8 the 10 most frequent Irish names, table A9 the most frequent Scottish names, table A10 the most frequent Welsh names and table A11 the 50 most frequent names from other ethnicities.

For each table I present the number of deaths and probates, the probate rate and the median wealth of those probated. The English names report an average probate rate of about 21% and a median printed wealth of £27,000, although there is notable variance between names. For example the “Brown” surname has a probate rate of 24% (£26,410 median real wealth), but the “Johnson” name has a probate rate of 18% (median real wealth £25,487). Irish names report both significantly lower probate rates and lower median real wealth. For example, the name “Murphy” has a probate rate of 12% and a median real wealth for those who are probated of £20,647. Scottish names display the opposite pattern; they report significantly higher probate rates and wealth, for example the name “Stewart” with a probate rate of 27% and a median probated wealth of £33,943. Welsh names display a large variance in probate rates (from 12% for “Davies” to 24% for “Jones” but the median probated wealth is around £24,000 with lower deviation.

Finally, surnames classified as belonging to ethnic groups outside of the four major nations report considerable heterogeneity in their probate rate and median wealth. Compare the “Levy” surname, which is Jewish in origin, which has a probate rate of 29% and a median real probated wealth of £39,033, considerably higher than that of English names, with that of “Ali”, a Bangladeshi surname, which has a 4% probate rate and a median probated wealth of £8,908.

Table A7: Top 10 English Names dying in England, 1858-2007

Surname	$N_d$	$N_p$	$\frac{N_p}{N_d}$	$r\tilde{w}$
Smith	1,005,904	210,367	0.21	26,488
Taylor	470,279	100,346	0.21	26,406
Brown	423,064	99,999	0.24	26,410
Johnson	276,134	49,274	0.18	25,487
Robinson	254,626	57,323	0.23	26,838
Wright	239,341	42,103	0.18	26,105
Thompson	233,729	51,352	0.22	26,458
Wood	229,983	43,129	0.19	27,661
White	229,413	48,619	0.21	27,887
Walker	228,069	47,580	0.21	28,420

*Notes:*  $N_d$  is number of deaths,  $N_p$  is the number of probates and  $PR$  is the proportion probated  
 $r\tilde{w}$  is median probated real wealth (£2015)

Table A8: Top 10 Irish Names dying in England, 1858-1992

Surname	$N_d$	$N_p$	$\frac{N_p}{N_d}$	$r\tilde{w}$
Kelly	78,288	9,493	0.12	22,634
Murphy	66,196	8,123	0.12	20,647
Sullivan	37,960	5,512	0.15	18,426
Ryan	31,533	5,249	0.17	21,135
O'Brien	31,069	3,180	0.10	18,665
McCarthy	30,395	3,458	0.11	18,790
Burke	28,714	3,753	0.13	20,544
Gough	24,280	4,962	0.20	26,251
Rice	23,513	5,658	0.24	27,281
Connor	22,880	4,071	0.18	18,769

*Notes:*  $N_d$  is number of deaths,  $N_p$  is the number of probates and  $PR$  is the proportion probated  
 $r\tilde{w}$  is median probated real wealth (£2015)

Table A9: Top 10 Scottish Names dying in England, 1858-1992

Surname	$N_d$	$N_p$	$\frac{N_p}{N_d}$	$r\tilde{w}$
Wilson	279,751	69,046	0.25	27,799
Scott	135,173	28,022	0.21	30,054
Simpson	108,566	23,125	0.21	27,773
Gibson	72,339	11,573	0.16	25,773
Graham	63,268	15,182	0.24	26,650
Murray	53,913	10,782	0.20	28,785
Campbell	46,731	10,646	0.23	34,141
Stewart	41,869	11,499	0.27	33,943
Grant	41,222	5,671	0.14	27,493
Henderson	39,546	9,465	0.24	29,383

*Notes:*  $N_d$  is number of deaths,  $N_p$  is the number of probates and  $PR$  is the proportion probated  
 $r\tilde{w}$  is median probated real wealth (£2015)

Table A10: Top 10 Welsh Names dying in England

Surname	$N_d$	$N_p$	$\frac{N_p}{N_d}$	$r\tilde{w}$
Jones	878,849	207,477	0.24	23,830
Williams	564,973	137,296	0.24	24,317
Davies	433,678	50,584	0.12	22,641
Evans	349,911	71,927	0.21	24,046
Thomas	317,632	64,799	0.20	24,574
Roberts	293,002	58,620	0.20	24,204
Edwards	221,825	40,433	0.18	25,039
Hughes	220,180	38,415	0.17	23,719
Lewis	211,949	48,939	0.23	24,812
Morris	177,541	42,374	0.24	25,592

*Notes:*  $N_d$  is number of deaths,  $N_p$  is the number of probates and  $PR$  is the proportion probated  
 $r\tilde{w}$  is median probated real wealth (£2015)

Table A11: Top 50 Names dying in England, not British or Irish

Surname	Ethnicity	$N_d$	$N_p$	$\frac{N_p}{N_d}$	$r\tilde{w}$
Cohen	Jewish	18,477	3,630	0.20	41,396
Levy	Jewish	11,144	3,274	0.29	39,033
Lund	Swedish	7,982	1,573	0.20	28,726
Isaacs	Jewish	7,648	1,172	0.15	32,997
Abrahams	Jewish	7,101	1,555	0.22	31,657
Patel	Indian Hindi	5,360	1,955	0.36	36,001
Singh	Sikh	4,979	2,580	0.52	27,473
Sales	Spanish	4,815	1,029	0.21	22,989
Silver	Jewish	4,422	1,274	0.29	33,603
Paton	Spanish	4,134	1,283	0.31	41,854
Holley	Channel Islander	3,954	972	0.25	22,709
Michael	Greek Cypriot	3,752	1,251	0.33	25,846
Gold	Jewish	3,691	645	0.17	30,018
Bloom	Jewish	3,570	986	0.28	36,962
Tong	Hong Kongese	3,429	374	0.11	28,113
Goldstein	Jewish	3,388	569	0.17	28,846
Woolf	Jewish	3,153	938	0.30	45,028
Ali	Bangladeshi	3,082	113	0.04	8,908
Mackey	Celtic	3,032	366	0.12	24,751
Osman	Muslim Indian	2,968	325	0.11	20,502
Goldberg	Jewish	2,905	439	0.15	25,290
Cornelius	Dutch	2,731	744	0.27	29,274
Khan	Pakistani	2,708	601	0.22	22,796
Hannan	International	2,666	419	0.16	22,179
Rosenberg	Jewish	2,544	744	0.29	25,737
Kaur	Sikh	2,482	383	0.15	9,759
Muller	German	2,460	933	0.38	44,181
Roy	International	2,460	800	0.33	32,584
Freedman	Jewish	2,428	603	0.25	41,901
Lazarus	Jewish	2,380	701	0.29	37,465
Solomons	Jewish	2,373	480	0.20	26,405
Olsen	Nordic	2,181	315	0.14	19,425
Hansen	Danish	2,151	619	0.29	28,588
Lally	Unclassified	2,122	169	0.08	17,212
Wagner	German	1,927	576	0.30	42,603
Bernstein	Jewish	1,879	531	0.28	28,526
Hyams	Jewish	1,869	381	0.20	37,503
Hyman	Jewish	1,830	611	0.33	38,247
Gross	Jewish	1,809	724	0.40	40,013
Levi	Jewish	1,806	456	0.25	39,475
Channing	Channel Islander	1,702	320	0.19	23,140
Hoffman	Jewish	1,695	423	0.25	28,040
Goldman	Jewish	1,670	224	0.13	31,929
Schneider	Jewish	1,568	425	0.27	32,094
Schmidt	German	1,557	339	0.22	39,669
Klein	Jewish	1,483	338	0.23	30,690
Hussain	Pakistani	1,417	391	0.28	21,642
Becker	German	1,409	349	0.25	43,056
Reddy	Indian Hindi	1,408	193	0.14	23,649
Rosenthal	Jewish	1,385	477	0.34	40,866

*Notes:*  $N_d$  is number of deaths,  $N_p$  is the number of probates and  $PR$  is the proportion probated.  $r\tilde{w}$  is median probated real wealth (£2015)

## D Irish Names

The PPR Calendar data was processed via an OCR (Optical Character Recognition) engine. The process in general worked very well and the resulting data set passed multiple data-quality tests (Cummins (2019b)). Amongst the Irish, names beginning with “O” are commonplace, and non-existent within other populations. As the OCR process and the algorithms used to extract surnames may have missed this “”, I inspected all possible candidate “O” stemmed names in the PPR calendar data. This check turned up numerous oddities. For example, there are 37,613 deaths 1838-2007 for people with the surname “O’Brien” yet only 5 probates recorded, 1858-1992, for this surname. Yet, there are 3,175 probates recorded for the name “Brien” but only 2,304 deaths. I cross-checked all Irish names and assigned any possible stemmed names to the most common occurrence, as measured by the count of all deaths to that name, 1838-2007. Mechanically this was done by summing all deaths in the death data, all probates in the PPR Calendar data and inspecting all 5,805 Irish names for anomalies. This meant that all “Briens” were updated to “O’Brien”, “Neill” to “O’Neill”, but all “O’Sullivans” were updated to “Sullivan”, “O’Daly” to “Daly”. Surnames were only updated where both the stemmed and non-stemmed version were of Irish ancestry. (thereby grouping over Irish doesn’t make any difference to the results.

## E The 1911 Census Ethnicity Classifier

I use the 36 million de-anonymized individual records from the special access version of the 1911 census, to examine the distribution of place of birth for the over 500,000 surnames (Schurer and Higgs (2021)). Table A12 reports the top 25 most numerous countries of birth listed in 1911. Nearly 90% of those enumerated were born in England, 6.5% were born in Wales, 1.5% in Scotland, 1.2% in Ireland. All other countries each represent far less than 1%.

Based on table A12 we pick 11 countries of birth to attribute an ethnicity to surnames. The countries are England, Wales, Scotland, Ireland, India (which includes contemporary Pakistan as it’s before 1947), Germany, France, Italy, Poland and the Netherlands. How to know whether a given surname corresponds to a origin country? Table A13 presents the matrix of the proportions born in each of the 11 countries for a set of well known surnames.

As here we do not observe the global distribution of surnames in 1911 but the distribution *within* England, we cannot simply assign the most frequent country-of-birth to a surname. This would classify many names incorrectly. For example, Stewart (Scottish), Cohen (Russian and Jewish), Murphy (Irish), Durand (French), Van Gelder (Dutch), Becker (German), Posner (Russian and Polish) would all incorrectly be classified as English.<sup>43</sup>

To more correctly attribute Surnames to ethnic origins I therefore cross reference the complete matrix of surnames by country of birth (as represented by the example Surnames in table A13), with the average proportions born in England from table A12. We first attribute to each surname an ethnic origin based upon the most frequent country of birth. Where there is a country other than England or Wales which accounts for 5% or over of the births of that surname, we update the ethnic origin to that country. This procedure works to attribute correctly all of the example surnames in table A13).<sup>44</sup>

---

<sup>43</sup>An example of how this happens is to imagine a migrant couple, from Ireland, with a unique surname, moving to London in 1900, and having 5 children. By 1911, 5/7, or 71%, of the holders of the name, in England, would be born in England, even if this is arguably an “Irish” family.

<sup>44</sup>Upon inspection, it was apparent that this method incorrectly assigned many Welsh surnames as “English” (e.g. Jones, Edwards and Hughes). This is because of the very unequal population sizes of the two neighboring countries. 44% of Jones, 37% of Hughes and 25% of Edwards, are born in Wales. We therefore updated a surname to “Welsh” if more than 20% of the holders of a surname were born in Wales. As 6.5% of the population of England and Wales was born in Wales (table A12), the 20% cutoff here implies that the holders of a “Welsh” surname are at least 300% more likely to have been born in Wales than the average English.

Table A12: Distribution of 1911 Adult Census Population by Country of Birth

Country	N	%
England	28,052,691	89.19
Wales	2,052,922	6.53
Scotland	458,153	1.46
Ireland	372,708	1.18
Russia	72,533	0.23
India and Pakistan	58,598	0.19
Germany	55,237	0.18
France	40,242	0.13
United States	36,025	0.11
Isle of Man	35,111	0.11
Australia	21,410	0.07
Italy	18,412	0.06
Poland	18,253	0.06
Canada	17,493	0.06
South Africa	16,650	0.05
Austria	12,196	0.04
Switzerland	9,877	0.03
Netherlands	7,859	0.02
At Sea	6,082	0.02
Malta	5,863	0.02
Sweden	5,724	0.02
Norway	5,493	0.02
Belgium	5,397	0.02
New Zealand	5,282	0.02
Gibraltar	4,758	0.02

Source: 1911 Census

Table A13: Example Surnames for Attributing Ethnicity from the 1911 Census

Surname	N	Country of Birth										
		England	Wales	Scotland	Ireland	Russia	Germany	India	France	Italy	Poland	Netherlands
Churchill	4,957	<b>.789</b>	.049	.003	.009	.000	.000	.002	.001	.000	.000	.000
Davies	215,938	.348	<b>.559</b>	.004	.002	.000	.000	.001	.000	.000	.000	.000
MacDonald	8,027	.638	.021	<b>.144</b>	.027	.000	.001	.008	.000	.000	.000	.000
Murphy	24,697	.638	.055	.015	<b>.144</b>	.000	.000	.003	.000	.000	.000	.000
Cohen	14,816	.568	.011	.003	.003	<b>.194</b>	.010	.000	.001	.000	.050	.003
Becker	861	.584	.014	.006	.010	.030	<b>.138</b>	.002	.006	.000	.003	.001
Singh	186	.032	.000	.000	.000	.000	.000	<b>.790</b>	.000	.000	.000	.000
Durand	258	.581	.004	.004	.004	.000	.000	.035	<b>.198</b>	.000	.000	.000
Ferrari	252	.425	.020	.000	.000	.000	.000	.000	.000	<b>.226</b>	.000	.000
Posner	328	.579	.003	.000	.003	.155	.015	.003	.000	.000	<b>.113</b>	.000
Van Gelder	98	.653	.000	.000	.000	.010	.000	.000	.000	.000	.000	<b>.122</b>

Note: Calculated from the 1911 census. **Bolded** text indicates assigned ethnicity.

## F Banded Wealth

The majority of PPR Calendar wealth estimates after 1980 are reported within a “band”. These are £25,000, £40,000, £70,000, £100,000, £115,000 and £125,000, with each entry listed as “Not Exceeding” the quoted amount. Figure A25 reports the distribution of wealth within each of these bands for the last observed year where exact values were used for all wealth estimates. As is evident the distributions within the bands are neither normal, nor uniform. A simple average would not be optimal given these distributional characteristics.

Therefore I extract the characteristics of the 1980 wealth distribution, from zero to £125,000 using a Kernel Density Estimate, and used those characteristics to infer a specific wealth for the banded observations after 1980. This was done for 1,521,608 observations out of a total of 2,506,371, 1981-1992 inclusive.

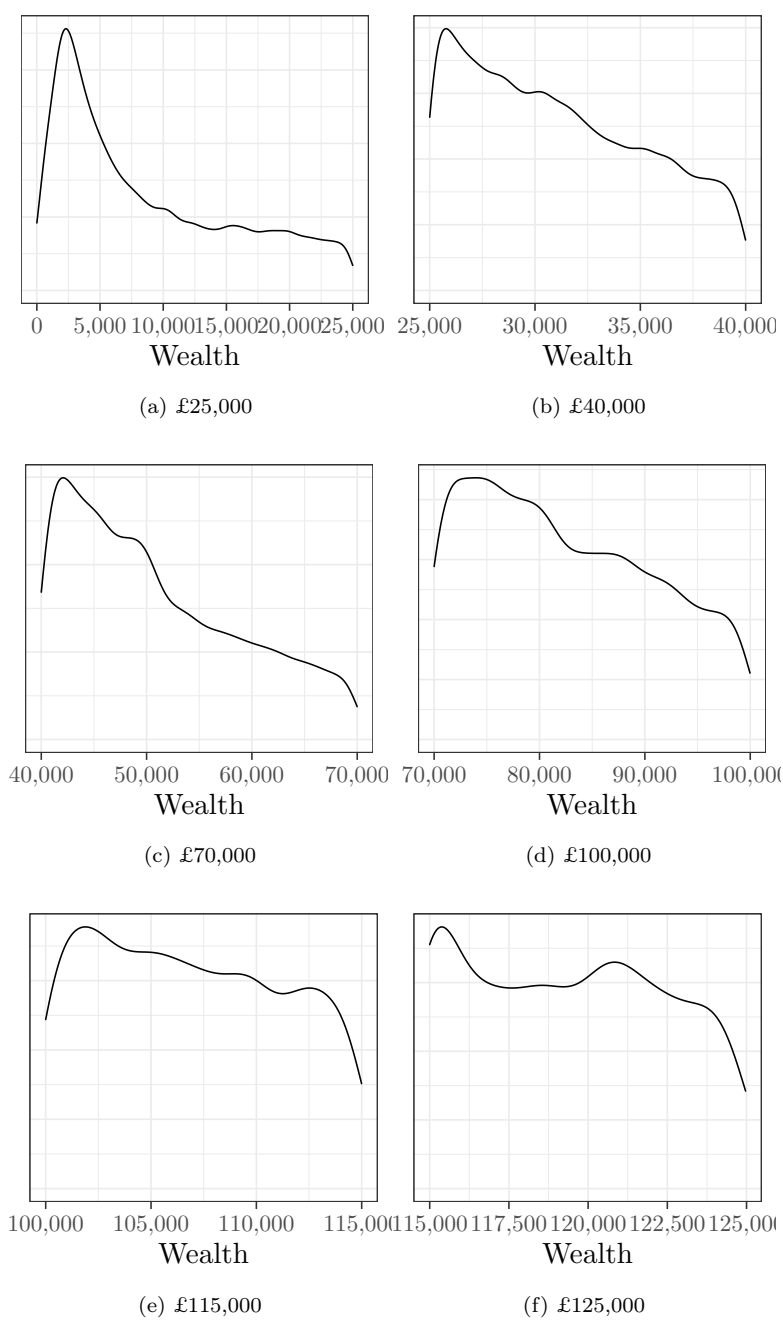


Figure A25: The Distribution of Wealth within the post 1980 Defined Wealth Bands, for 1980

## G Extra Results

Table A14 reports a linear probability model of the proportion probated as a simple function of a set of ethnic dummies, and table A15 does the same for  $\ln(Wealth)$ . Table A16 gives the full regression details of a simple linear regression of the probability of being in the top 1% of wealth holders in your year of death against the top 25 ethnic dummies. Note that in all tables ethnic groups are ranked in descending order of the sum total of deaths, 1866-2007.



Table A14: Probability Probated for the Top 25 Ethnic Groups

	Probability Probated						
	1858-1879	1880-1913	1914-39	1940-59	1960-69	1970-79	1980-92
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Welsh	.001*** (.0003)	.002*** (.0002)	.012*** (.0004)	-.013*** (.001)	-.003*** (.001)	-.007*** (.001)	-.007*** (.001)
Scottish	.013*** (.0004)	.017*** (.0003)	.020*** (.001)	-.011*** (.001)	-.024*** (.001)	-.014*** (.001)	-.020*** (.001)
Irish	-.034*** (.0004)	-.048*** (.0003)	-.083*** (.001)	-.148*** (.001)	-.144*** (.001)	-.118*** (.001)	-.103*** (.001)
Jewish	.012*** (.002)	-.002 (.001)	.027*** (.002)	.011*** (.003)	.046*** (.004)	.067*** (.004)	.060*** (.004)
German	.022*** (.002)	.044*** (.002)	.079*** (.002)	.056*** (.003)	.056*** (.004)	.062*** (.004)	.054*** (.004)
Sikh	.011*** (.002)	.029*** (.002)	-.033*** (.003)	.068*** (.005)	-.055*** (.005)	.103*** (.005)	.039*** (.004)
Italian	.045*** (.003)	.024*** (.002)	.004 (.003)	-.010* (.004)	-.041*** (.005)	-.075*** (.005)	-.057*** (.004)
Polish	.068*** (.005)	.081*** (.004)	.120*** (.006)	-.044*** (.005)	-.013* (.006)	-.031*** (.005)	-.033*** (.004)
French	.075*** (.003)	.088*** (.002)	.102*** (.003)	.074*** (.005)	.013* (.006)	.082*** (.006)	.048*** (.005)
Pakistani	.029*** (.005)	.072*** (.005)	.133*** (.006)	.152*** (.009)	-.092*** (.008)	-.046*** (.006)	-.077*** (.004)
Black Caribbean	.025*** (.006)	.065*** (.007)	.062*** (.011)	.049** (.015)	-.110*** (.017)	-.086*** (.016)	-.091*** (.014)
Indian Hindi	.032*** (.007)	.098*** (.007)	.150*** (.012)	.190*** (.015)	-.093*** (.012)	-.038*** (.007)	-.054*** (.005)
Spanish	.090*** (.004)	.097*** (.004)	.111*** (.006)	.130*** (.007)	.070*** (.009)	-.006 (.008)	-.042*** (.007)
European	.047*** (.004)	.077*** (.004)	.165*** (.006)	.096*** (.007)	.036*** (.010)	.146*** (.009)	.075*** (.008)
Hong Kongese	.139*** (.006)	.406*** (.006)	.646*** (.006)	.595*** (.009)	.399*** (.013)	.453*** (.010)	.078*** (.008)
Bangladeshi	.035** (.012)	.046*** (.010)	-.047*** (.011)	-.090*** (.015)	-.165*** (.014)	-.137*** (.011)	-.209*** (.008)
Channel Islander	.040*** (.005)	.050*** (.005)	.036*** (.006)	.016 (.009)	-.018 (.012)	.010 (.012)	.030** (.011)
Swedish	.023*** (.005)	.026*** (.004)	.056*** (.007)	.017 (.009)	-.122*** (.013)	-.140*** (.012)	-.164*** (.012)
Greek Cypriot	.033*** (.008)	.086*** (.008)	.084*** (.012)	-.001 (.012)	-.154*** (.011)	-.156*** (.010)	-.136*** (.008)
Portuguese	.167*** (.007)	.102*** (.006)	.078*** (.009)	-.011 (.011)	-.076*** (.013)	-.071*** (.011)	-.108*** (.010)
Danish	.005 (.008)	.004 (.005)	.016* (.008)	.010 (.010)	.041** (.013)	.140*** (.013)	.038** (.011)
Dutch	.197*** (.006)	.107*** (.005)	.156*** (.008)	.178*** (.010)	.136*** (.014)	.151*** (.014)	.189*** (.013)
Pakistani Kashmir	.022* (.010)	.238*** (.010)	.140*** (.010)	.239*** (.016)	.026 (.017)	-.064*** (.013)	-.098*** (.010)
Muslim	.046*** (.008)	.137*** (.008)	.214*** (.012)	.154*** (.015)	.020 (.017)	.272*** (.013)	-.036*** (.010)
Norwegian	.046*** (.009)	.047*** (.006)	.139*** (.009)	.200*** (.012)	.068*** (.017)	.182*** (.016)	.089*** (.015)
Constant	.055*** (.0001)	.086*** (.0001)	.175*** (.0001)	.355*** (.0002)	.398*** (.0002)	.369*** (.0002)	.388*** (.0002)
Observations	10,522,167	17,903,960	12,661,899	10,188,348	5,523,080	5,645,397	6,624,590
R <sup>2</sup>	.001	.002	.003	.004	.004	.003	.002

Note: \*p<0.05; \*\*p<0.01; \*\*\*p<0.001

OLS, English is the omitted Group

Table A15: Ethnic Group Correlations with Wealth, by Sub-period, 1858-1992

	ln(Real Wealth)						
	1858-1879	1880-1913	1914-39	1940-59	1960-69	1970-79	1980-92
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Welsh	-.0002 (.001)	-.003 (.002)	.014*** (.002)	-.067*** (.002)	-.019*** (.003)	-.038*** (.003)	-.044*** (.003)
Scottish	.043*** (.002)	.114*** (.002)	.111*** (.002)	.018*** (.003)	-.017*** (.004)	-.016*** (.003)	-.056*** (.004)
Irish	-.190*** (.002)	-.232*** (.002)	-.312*** (.002)	-.556*** (.003)	-.521*** (.004)	-.406*** (.003)	-.408*** (.004)
Jewish	.062*** (.012)	.195*** (.010)	.144*** (.009)	.131*** (.010)	.317*** (.015)	.351*** (.013)	.325*** (.015)
German	.110*** (.012)	.355*** (.010)	.360*** (.010)	.336*** (.012)	.350*** (.017)	.321*** (.014)	.297*** (.016)
Sikh	.056*** (.013)	.140*** (.014)	-.100*** (.014)	.249*** (.019)	-.107*** (.021)	.253*** (.017)	.182*** (.017)
Italian	.186*** (.019)	.364*** (.016)	.051*** (.013)	.004 (.016)	-.049* (.019)	-.193*** (.015)	-.182*** (.017)
Polish	.272*** (.028)	.495*** (.026)	.509*** (.022)	-.234*** (.019)	-.041 (.022)	-.097*** (.015)	-.083*** (.016)
French	.363*** (.015)	.553*** (.015)	.470*** (.014)	.431*** (.017)	.218*** (.024)	.348*** (.020)	.262*** (.023)
Pakistani	.147*** (.029)	.399*** (.031)	.501*** (.026)	.528*** (.035)	-.141*** (.032)	-.178*** (.020)	-.247*** (.018)
Black Caribbean	.089* (.038)	.318*** (.047)	.319*** (.046)	.276*** (.059)	-.299*** (.066)	-.251*** (.054)	-.315*** (.059)
Indian Hindi	.173*** (.042)	.401*** (.048)	.514*** (.047)	.693*** (.057)	-.101* (.047)	-.153*** (.022)	-.151*** (.020)
Spanish	.306*** (.024)	.522*** (.025)	.443*** (.023)	.495*** (.028)	.304*** (.036)	-.015 (.026)	-.125*** (.029)
European	.240*** (.023)	.453*** (.025)	.679*** (.023)	.449*** (.028)	.292*** (.038)	.548*** (.031)	.356*** (.035)
Hong Kongese	.682*** (.035)	1.990*** (.039)	2.303*** (.025)	2.108*** (.035)	1.419*** (.051)	1.362*** (.034)	.283*** (.035)
Bangladeshi	.136 (.073)	.513*** (.068)	-.252*** (.045)	-.416*** (.057)	-.455*** (.055)	-.447*** (.036)	-.786*** (.032)
Channel Islander	.150*** (.031)	.339*** (.031)	.133*** (.026)	.084* (.033)	.037 (.047)	.069 (.041)	.141** (.047)
Swedish	.138*** (.030)	.188*** (.029)	.200*** (.026)	.107** (.035)	-.407*** (.051)	-.401*** (.041)	-.645*** (.050)
Greek Cypriot	.151** (.049)	.390*** (.053)	.383*** (.048)	-.026 (.047)	-.445*** (.045)	-.412*** (.033)	-.465*** (.033)
Portuguese	.749*** (.041)	.524*** (.041)	.266*** (.036)	.022 (.044)	-.121* (.052)	-.182*** (.039)	-.396*** (.041)
Danish	-.004 (.048)	.176*** (.035)	.091** (.031)	.099** (.037)	.231*** (.053)	.533*** (.043)	.166*** (.049)
Dutch	.642*** (.038)	.529*** (.035)	.641*** (.031)	.780*** (.038)	.625*** (.053)	.566*** (.046)	.797*** (.053)
Pakistani Kashmir	.057 (.060)	1.324*** (.067)	.468*** (.042)	.831*** (.062)	.209** (.065)	-.187*** (.042)	-.351*** (.041)
Muslim	.201*** (.048)	.646*** (.057)	.763*** (.049)	.719*** (.060)	.168* (.067)	.852*** (.043)	-.128** (.044)
Norwegian	.198*** (.050)	.417*** (.042)	.461*** (.038)	.776*** (.048)	.304*** (.068)	.585*** (.055)	.346*** (.062)
Constant	4.649*** (.0004)	5.867*** (.0005)	7.324*** (.0005)	7.778*** (.001)	8.078*** (.001)	8.475*** (.001)	8.773*** (.001)
Observations	10,468,704	17,864,595	12,661,899	10,188,348	5,523,080	5,645,397	6,624,589
R <sup>2</sup>	.001	.001	.003	.004	.003	.003	.002

Note: \*p<0.05; \*\*p<0.01; \*\*\*p<0.001

OLS, English is the omitted Group

Table A16: Probability of being in the Top 1% of Wealth, for the Top 25 Ethnic Groups

	Wealth in Top 1% (1/0)						
	1858-1880	1880-1914	1914-40	1940-60	1960-70	1970-80	1980-92
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Welsh	-.002*** (.0001)	-.003*** (.0001)	-.002*** (.0001)	-.003*** (.0001)	-.002*** (.0002)	-.002*** (.0001)	-.002*** (.0001)
Scottish	.004*** (.0002)	.005*** (.0001)	.006*** (.0001)	.005*** (.0001)	.005*** (.0002)	.004*** (.0002)	.004*** (.0002)
Irish	-.007*** (.0002)	-.005*** (.0001)	-.005*** (.0001)	-.005*** (.0002)	-.005*** (.0002)	-.005*** (.0002)	-.005*** (.0002)
Jewish	.006*** (.001)	.006*** (.001)	.009*** (.001)	.011*** (.001)	.014*** (.001)	.019*** (.001)	.026*** (.001)
German	.009*** (.001)	.014*** (.001)	.017*** (.001)	.016*** (.001)	.016*** (.001)	.017*** (.001)	.020*** (.001)
Sikh	.001 (.001)	.004*** (.001)	.002* (.001)	.003** (.001)	-.001 (.001)	.0002 (.001)	-.002 (.001)
Italian	.014*** (.001)	.010*** (.001)	.005*** (.001)	.002** (.001)	.003** (.001)	.001 (.001)	.004*** (.001)
Polish	.017*** (.002)	.015*** (.001)	.020*** (.001)	.003** (.001)	.002 (.001)	.001 (.001)	.0003 (.001)
French	.023*** (.001)	.028*** (.001)	.022*** (.001)	.019*** (.001)	.015*** (.001)	.015*** (.001)	.016*** (.001)
Pakistani	.011*** (.002)	.013*** (.002)	.018*** (.002)	.009*** (.002)	.001 (.002)	-.002 (.001)	-.002* (.001)
Black Caribbean	.014*** (.003)	.010*** (.002)	.012*** (.003)	.021*** (.003)	.007* (.003)	.004 (.003)	-.004 (.003)
Indian Hindi	.012*** (.003)	.012*** (.002)	.018*** (.003)	.011*** (.003)	.00002 (.002)	-.002 (.001)	-.002** (.001)
Spanish	.023*** (.002)	.021*** (.001)	.016*** (.001)	.011*** (.001)	.007*** (.002)	.002 (.002)	.003* (.001)
European	.015*** (.002)	.019*** (.001)	.022*** (.001)	.018*** (.001)	.019*** (.002)	.014*** (.002)	.012*** (.002)
Hong Kongese	.032*** (.003)	.048*** (.002)	.041*** (.002)	.029*** (.002)	.011*** (.003)	.015*** (.002)	.002 (.002)
Bangladeshi	.013* (.006)	.007* (.003)	-.003 (.003)	-.001 (.003)	-.007* (.003)	-.005* (.002)	-.007*** (.002)
Channel Islander	.010*** (.002)	.013*** (.002)	.002 (.002)	-.004* (.002)	.005* (.002)	.003 (.002)	-.001 (.002)
Swedish	.008** (.002)	.003* (.001)	.003* (.002)	.005** (.002)	-.002 (.003)	-.003 (.002)	-.004 (.002)
Greek Cypriot	.005 (.004)	.007** (.003)	.010*** (.003)	.008*** (.002)	-.002 (.002)	-.005* (.002)	.002 (.002)
Portuguese	.050*** (.003)	.025*** (.002)	.004 (.002)	.013*** (.002)	.004 (.003)	.002 (.002)	-.001 (.002)
Danish	-.001 (.004)	.001 (.002)	.013*** (.002)	.011*** (.002)	.010*** (.003)	.006* (.003)	.002 (.002)
Dutch	.038*** (.003)	.026*** (.002)	.022*** (.002)	.030*** (.002)	.018*** (.003)	.021*** (.003)	.025*** (.003)
Pakistani Kashmir	-.001 (.005)	.033*** (.003)	.008** (.003)	.019*** (.003)	.002 (.003)	.002 (.002)	-.003 (.002)
Muslim	.011** (.004)	.023*** (.003)	.017*** (.003)	.014*** (.003)	.006 (.003)	.009*** (.003)	.004* (.002)
Norwegian	.005 (.004)	.007*** (.002)	.013*** (.002)	.013*** (.003)	.017*** (.003)	.007* (.003)	.015*** (.003)
Constant	.011*** (.00003)	.009*** (.00002)	.009*** (.00003)	.010*** (.00003)	.010*** (.00005)	.009*** (.00004)	.009*** (.00004)
Observations	10,468,704	17,864,595	12,661,899	10,188,348	5,523,080	5,645,397	6,624,589
R <sup>2</sup>	.0004	.001	.001	.001	.0005	.0005	.001

Note: \*p<0.05; \*\*p<0.01; \*\*\*p<0.001

OLS, English is the omitted Group

## H On Changes in Ethnic Classification

Is the error rate for the *Onomap* ethnic classification of a surname correlated with the average wealth of that name? If so, this could indicate a mechanism where the economic success of a subset of an ethnicity result in their integration into the English socio-economic landscape via the naming patterns of their children. In other words, their naming patterns become indistinct from the native English thus clustering them with the English in the *Onomap* methodology.

Here I examine those surnames classified by the 1911 census method whose assigned ethnicity is different from that estimated by *Onomap*. Three categorical variables are constructed. The first is where the surname classification is different between the 1911 method, and the *Onomap* method (1). The second is where a surname is Non-English in the 1911 method, and “English” according to *Onomap* (2). Thirdly, I look at surnames which are coded as ‘English’ in 1911, but are non-English according to *Onomap* (3). I then estimate the following equation at the *surname* level

$$D_i^{Change} = \alpha + \beta_1 \ln(\bar{W}_i^{20thC}) + \beta_2 N_i + \varepsilon \quad (9)$$

where  $D^{Change}$  is one of the three surname classification inequalities just described for surname  $i$ ,  $\alpha$  is a constant,  $\bar{W}^{20thC}$  is the average probated wealth of a surname over the years 1912-1992, and  $N$  is the number of people with a given surname in 1911.

Table A17: Ethnic Classification Differences, Correlated to Wealth at the Surname Level

	$D^{Change}$	$D^{Change\ to\ English}$	$D^{Change\ from\ English}$
	(1)	(2)	(3)
$\ln(W)$	.024*** (.003)	-.047*** (.010)	-.011 (.008)
N	-0.00000*** (0.00000)	-0.00000** (0.00000)	-0.00000*** (0.00000)
Observations	38,069	3,205	5,846
R <sup>2</sup>	.003	.009	.003

Note: \*p<0.05; \*\*p<0.01; \*\*\*p<0.001  
OLS, Wealth is averaged 1912-1992.

Table A17 reports the results of these three regressions. Whilst there are statically significant coefficients for the classification inequality dummies for (1.) any difference, and (2.) a change from a Non-English to an English classification, the point estimates and their standard error in all 3 regressions indicate that the magnitude of the correlation of classification differences and surname wealth is minor. Column 2 indicates that a 100% change in wealth results in a 5% lower probability of being coded as English (for surnames that change classification between the two methods). This is the opposite of the intuition described above. Overall wealth correlates with the probability of a classification difference between the two methods of assigning ethnicity (column 1). One interpretation is that richer families are more likely to be more international, and to potentially marry partners from outside their home country, and thus adopt more international forenames for their offspring. Regardless, in all examined classification scenarios, the effects are minor.

# I On the Sensitivity of the Decomposition of Inequality Between and Within Ethnic Groups to Different methods of Inferring Wealth for the Non-Probated Population

How sensitive is the wealth inequality decomposition from section 6 to different assumptions about the non-probated population? Recall that a uniform value, equal to half of the average value of wealth observed in the probate calendar in a given year, is applied to all those who die with wealth insufficient to merit probate. As this is a uniform attribution, perhaps it is biasing the Theil decomposition into finding a small *between* ethnicity component, relative to the individual *within* component. In this section, I compare this with three other, different methods for assigning wealth to the non-probated population.

The first variation (method 2) assigns a random value between zero pounds, and the probate threshold for a given year. This attribution is based upon the distribution of wealth for those observed above the threshold. Due to the disproportionate share of the top 1%, I estimate the distribution parameters excluding the top 1%. For this I use the R package `truncnorm`.

As a visual example of this method, figure A26 (a) compares the uniform wealth attribution for the non-probated population, with the inferred distribution using the distribution parameters of the bottom 99% of the probated wealth distribution, for 1920 (when the probate threshold was £50). Figure A26 (b) reports a comparison of the smoothed density curve for the entire wealth distribution, again for 1920.

The nice feature of the truncated distribution method is that it can be extended to infer the missing wealth information *by ethnic group*. This is the third method. Figure A27 illustrates how this works for a sample of ethnicities for 1920 (a) and 1970 (b). Finally, I apply a method to examine the maximum possible difference between the non-probated populations between the English and all other ethnicities. Here all non-probated English are assigned an inferred wealth just below the the probate threshold value for their year of death (I subtract £1 from that threshold). Whilst all non-English non-probated deaths are assigned precisely £0.

Table A18 summarizes the four different methods. Figure A28 reports estimates of the between and within Theil decomposition exercise, as section 6, for each of the four alternative Non-Probated Wealth Attribution methods. There is minor variation between the attribution methods, with the uniform method (1) resulting in the highest *within* component, and the maximum possible method resulting in the lowest *within* component. However, the Theil decompositions for methods 1-3 are virtually identical. And the smaller *within* component and higher *between* component of the maximum possible method (4), does not change the conclusion that the vast majority of variation in wealth is attributable to individual variation, and not ethnic group variation.

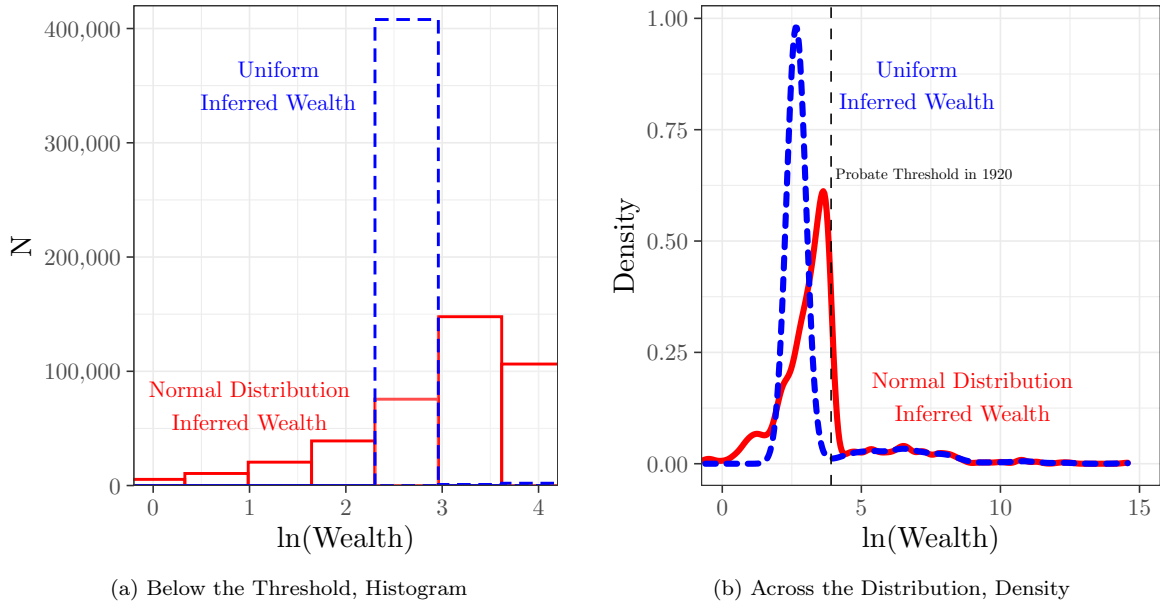
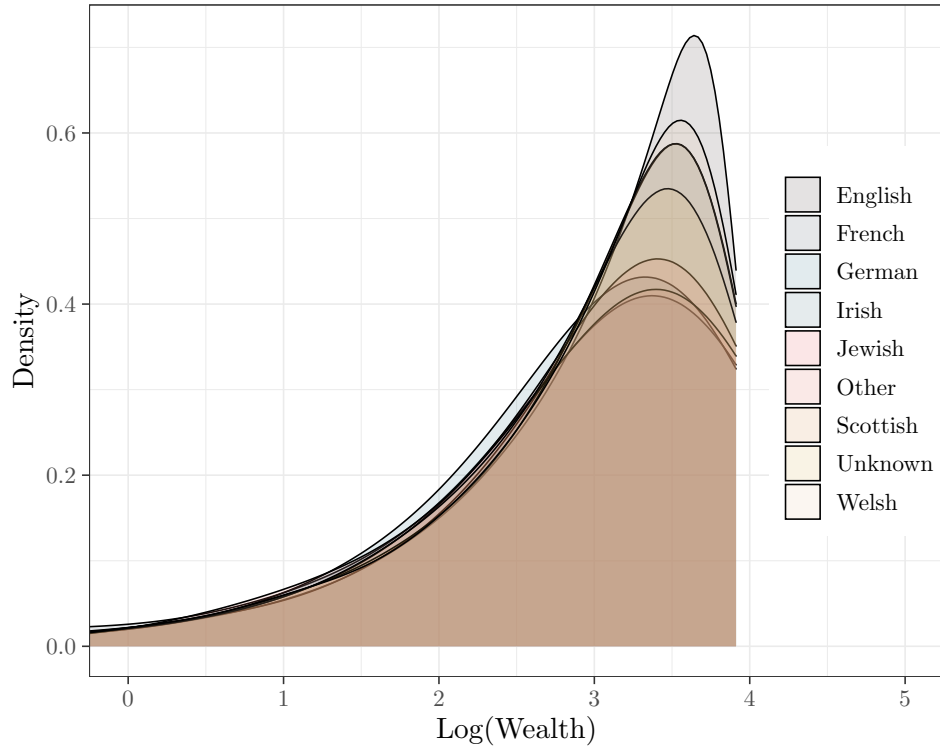


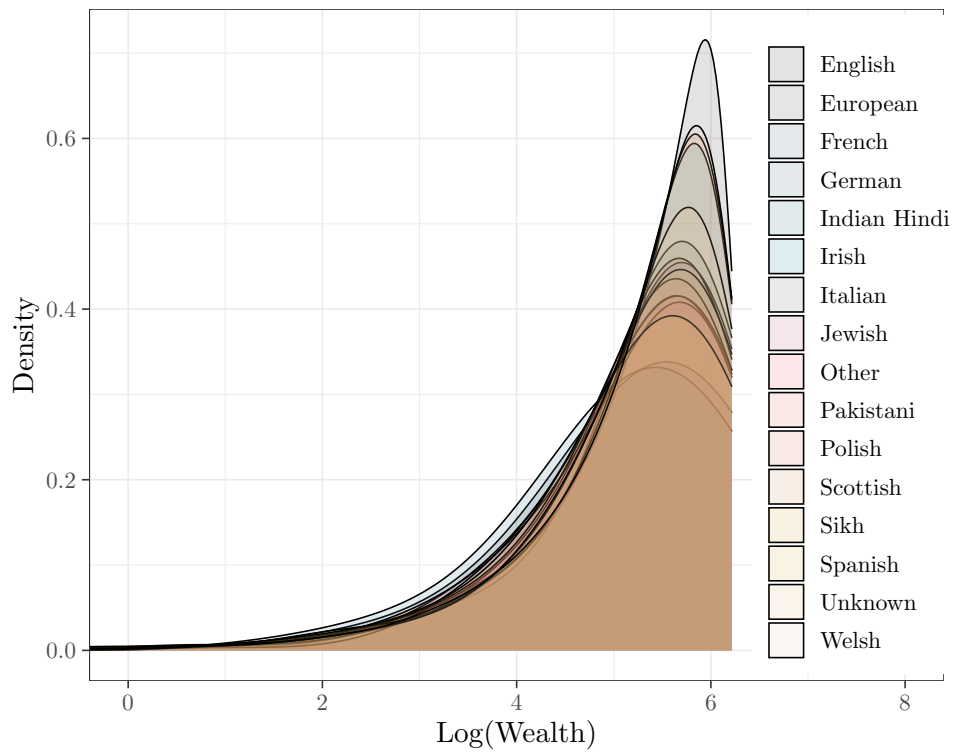
Figure A26: The Distribution of Wealth for Different Methods of Inferring Wealth for the Non-Probated Population, Visualized for 1920

Non-Probated Wealth Attribution		Description	Ethnic Variation
1	Uniform	Half the level of wealth observed in the PPR Calendars for the year of death, that was below the probate threshold	No
2	Random, Truncated Distribution	Random from 0 to threshold based upon distribution for those above threshold, by year	No
3	Random, Truncated Distribution Attribution, by Ethnicity	Random from 0 to threshold based upon distribution for those above threshold, by year and ethnicity	Yes
4	Maximum Possible Difference English to Other	Non-Probated “English” are assigned the probate threshold value -£1. All non-English are assigned £0	Between English and Others, Maximum Possible

Table A18: Scenarios for Different Wealth Attribution for the Non-Probated Population



(a) 1920



(b) 1970

Figure A27: Inferring Wealth below the Probate Threshold, by Ethnicity, Distribution Visualized for 1920 and 1970

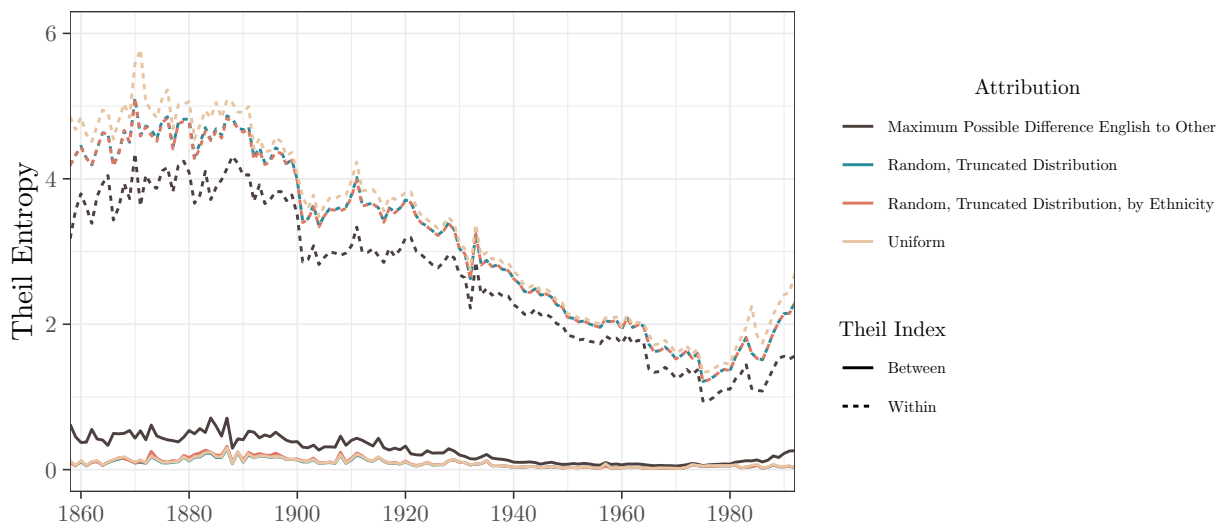


Figure A28: The Theil Index of Inequality, decomposed for between and within ethnic groupings, English Wealth, 1858-1992, for Different Methods of Inferring Wealth for the Non-Probated Population

Notes: Calculated separately for each of the 135 years 1858-1992 using *decompGEI* in R.