

# **DISCUSSION PAPER SERIES**

DP13436

**WHERE IS THE MIDDLE CLASS?  
INEQUALITY, GENDER AND THE SHAPE  
OF THE UPPER TAIL FROM 60 MILLION  
ENGLISH DEATH AND PROBATE  
RECORDS, 1892-2016**

Neil Cummins

**ECONOMIC HISTORY**

# WHERE IS THE MIDDLE CLASS? INEQUALITY, GENDER AND THE SHAPE OF THE UPPER TAIL FROM 60 MILLION ENGLISH DEATH AND PROBATE RECORDS, 1892-2016

*Neil Cummins*

Discussion Paper DP13436  
Published 09 January 2019  
Submitted 08 January 2019

Centre for Economic Policy Research  
33 Great Sutton Street, London EC1V 0DX, UK  
Tel: +44 (0)20 7183 8801  
[www.cepr.org](http://www.cepr.org)

This Discussion Paper is issued under the auspices of the Centre's research programme in **ECONOMIC HISTORY**. Any opinions expressed here are those of the author(s) and not those of the Centre for Economic Policy Research. Research disseminated by CEPR may include views on policy, but the Centre itself takes no institutional policy positions.

The Centre for Economic Policy Research was established in 1983 as an educational charity, to promote independent analysis and public discussion of open economies and the relations among them. It is pluralist and non-partisan, bringing economic research to bear on the analysis of medium- and long-run policy questions.

These Discussion Papers often represent preliminary or incomplete work, circulated to encourage discussion and comment. Citation and use of such a paper should take account of its provisional character.

Copyright: Neil Cummins

# WHERE IS THE MIDDLE CLASS? INEQUALITY, GENDER AND THE SHAPE OF THE UPPER TAIL FROM 60 MILLION ENGLISH DEATH AND PROBATE RECORDS, 1892-2016

## Abstract

This paper analyses a newly constructed individual level dataset of every English death and probate from 1892-2016. The estimated top wealth shares match closely existing estimates. However, this analysis clearly shows that the 20th century's 'Great Equalization' of wealth stalled in mid-century. The probate rate, which captures the proportion of English with any significant wealth at death rose from 10% in the 1890s to 40% by 1950 and has stagnated to 2016. Despite the large declines in the wealth share of the top 1%, from 73% to 20%, the median English person died with almost nothing throughout. All changes in inequality after 1950 involve a reshuffling of wealth within the top 30%. Further, I find that a log-linear distribution fits the empirical data better than a Pareto power law. Finally, I show that the top wealth shares are increasingly and systematically male as one ascends in wealth, 1892-1992, but this has equalized over the 20th century.

JEL Classification: N00, N33, N34, D31

Keywords: inequality, economic history, Big Data

Neil Cummins - [n.j.cummins@lse.ac.uk](mailto:n.j.cummins@lse.ac.uk)  
*London School of Economics and CEPR*

# Where is the Middle Class? Inequality, Gender and the Shape of the Upper Tail from 60 million English Death and Probate Records, 1892-2016

Neil Cummins\*

January 8, 2019

## Abstract

This paper analyses a newly constructed individual level dataset of every English death and probate from 1892-2016. The estimated top wealth shares match closely existing estimates. However, this analysis clearly shows that the 20th century’s ‘Great Equalization’ of wealth stalled in mid-century. The probate rate, which captures the proportion of English with any significant wealth at death rose from 10% in the 1890s to 40% by 1950 and has stagnated to 2016. Despite the large declines in the wealth share of the top 1%, from 73% to 20%, the median English person died with almost nothing throughout. All changes in inequality after 1950 involve a reshuffling of wealth within the top 30%. Further, I find that a log-linear distribution fits the empirical data better than a Pareto power law. Finally, I show that the top wealth shares are increasingly and systematically male as one ascends in wealth, 1892-1992, but this has equalized over the 20th century.

JEL: N00, N33, N34, D31

Keywords: inequality; economic history; big data

## 1 Introduction

The distribution of the gains from capitalism, globalization and technological progress preoccupies academic and popular economics (Marx (1867); Piketty (2014)). Within countries, the driving force behind the 20th century’s drop in inequality were the declines in the wealth-shares of the top 1% (Alvaredo et al. (2018), Saez and Zucman (2016) and Piketty (2014)).<sup>1</sup> From this, a ‘patrimonial (or propertied) middle class’ arose (Piketty (2014) p.260).

---

\*Economic History, LSE. Version 1.5. Thanks to Facundo Alvaredo and Cormac Ó Gráda for detailed feedback, John Turner, Greg Clark, John Devereux and Morgan Kelly for helpful comments and seminar participants at the University of Luxembourg and the University of Hong Kong. I thank the International Inequalities Institute at LSE for a grant to finish the database construction.

<sup>1</sup>“The twentieth century’s dramatic drop in economic inequality was entirely due to falls in inequality in income from capital” (Piketty (2014) p.366)”.

This paper shows that for Britain, it was not the rise of a broad ‘middle’ class which characterized the changes in the 20th century wealth distribution but a reshuffling of wealth away from the top 1% to the rest of the top 20-30%.<sup>2</sup> The vast majority die with nothing.

I introduce and analyze a new individual level dataset of every English adult death and probate (60m and 18m respectively) from 1892-2016, a period which captures the decline of wealth inequality in Britain.<sup>3</sup> The top 1% share declines from 73% to 20%. Despite this ‘Great Equalization’, the *relative* gains from the decline of the elite are limited to the top 30% after 1950. The median English person dies with no significant wealth, throughout the entire period. Inferring via Pareto power law extrapolations that the decline of the 1% led to a rise in median wealth is mistaken. If anything, I show that the wealth distribution is log-normal, not Pareto. I also describe the equalization of gender-based wealth inequality, 1892-1992, although I note a systematic and persistent *increasing* underrepresentation of women among the top percentiles. The higher the wealth cut-off, the top 10, top 1% and so on, the less women are observed.

This paper follows existing empirical work on the historical English wealth distribution by Lindert (1986), Atkinson and Harrison (1978); Atkinson et al. (1989); Atkinson (2013). It complements recent work, using a different but related source, by Alvaredo et al. (2018). The individual level data developed here allows for broader claims about the wealth distribution outside the top 10%, its distributional form and gender.<sup>4</sup>

There are 5 principal contributions of this paper. Firstly, I detail the construction of the probate and wealth database. The process involved the transformation of over 1.5 million scanned images into a set of text files. Next the text was algorithmically parsed and formed into a database suitable for economic analysis. This largely automated data-building methodology has the potential to be applied to many other scanned historical sources. Any historical document with a standard structure can be converted in this way.

Secondly I contribute new estimates of the top wealth shares. They match closely recent estimates from separate sources, validating the constructed data. For the first time wealth shares beyond the top 20% are estimated<sup>5</sup>.

Thirdly, I combine the probate data with all English deaths, 1892-2016, to calculate the probate rate. By analyzing the probate rate and the changes in the wealth shares of the top 30%, I am able to characterize the relative ‘winners’ from the 20th century’s ‘great equalization’ of wealth. The ‘winners’ are not a broad ‘middle class’. The distributional gains are exclusively confined to the top 30% of the wealth distribution.

The individual level data allows the precise testing of theoretical distributions to English wealth. I find that the log-normal is a better approximation of the wealth distribution than the Pareto power law. The distributional tests executed here have importance for those using incomplete data that needs to make distributional assumptions.<sup>6</sup> They also matters for theory.

---

<sup>2</sup>There is no single, widely accepted definition of the ‘middle class’. Piketty defines his ‘patrimonial middle class’ as the 50th - 90th percentile (2014 p.347-8). There are a universe of alternative definitions.

<sup>3</sup>I observe the universe of individual estate valuations from 1892-1992. From 1992 to 2016 I observe simply whether an individual made a probate (a legal requirement for having assets greater than £5,000 at death in 2018).

<sup>4</sup>Green et al. (2009) use a sample from the death duty registers to examine gender and wealth, 1870-1902.

<sup>5</sup>Estimates for the top 20% wealth share are presented by Atkinson et al. (1989), table 1, p.318.

<sup>6</sup>To quote Atkinson (2005), p.332, on Pareto interpolation; “The standard current practice ... is to assume a Pareto distribution... This follows a venerable tradition: for example a Pareto interpolation was used in the report of the House of Commons Committee on Income Tax (House of Commons (1906), Appendix, pages 222 and 245–6). Studies that use Pareto interpolations include Kuznets (1953), Atkinson and Harrison (1978), Feenberg and Poterba (2000), Piketty (2001, 2003) and Piketty and Saez (2003). A recent paper developing *generalized* Pareto curves to analyze tabulated data is Blanchet et al. (2017). Saez and Zucman (2016) estimate ‘capitalized’ wealth, from income

A log-normal wealth distribution implies different generation mechanisms than a Pareto form (Gabaix (1999)).<sup>7</sup>(As one example, Stiglitz (1969) shows how the operation of primogeniture can lead to a stable wealth distribution with a Pareto upper tail.)

Finally, I examine the gender breakdown of English wealth. I show that the top percentiles are systematically increasingly male. as one ascends the wealth distribution. However, this has been changing, towards greater equality between the sexes, over the century between 1892 to 1992.

This paper consists of 7 sections. Section 2 discusses the source for the probate data; The Principal Probate Registry Calendars, 1892-2016. Section 3 describes the many stages behind the data construction. The resulting data is cross checked with existing estimates and wealth shares are described in section 4. Section 5 tests whether a log-normal or Pareto distribution fit the data better, for both the complete probated population and the extreme upper tail, the top point one percent. Gender aspects of wealth are reported in section 6. Section 7 concludes.

## 2 The Source

### 2.1 The Principal Probate Registry Calendars

The data for analysis is a complete digitization of the Principal Probate Registry (PPR) Calendar for England and Wales from 1892-1992. In order for a will to be executed and assets transferred an act of probate must be granted. The probate index records all those who died with wealth above a minimum threshold (see table 2.2).

Together with the name, address and date of death of the deceased were details of the executioner of the estate and an estimate of the estate value. Not everyone who dies has assets. Therefore the probate data is supplemented with complete death registers 1838-2006. The centralized, national, printed calendars begin in 1858 and extend (in microfiche form after 1979) until 1996 - The index is now in a digital form. The data was only extracted between 1892 and 1992 as the format of entries is consistent during this interval.

Table 2.1 summaries the type of assets included in the probate valuations. The values in the index are ‘gross’ - where the net value accounts for debts and funeral expenses. The biggest consistent omission is ‘unsettled personalty’ - for example trust funds (Rubinstein (1974)). p.70). Also, there is no information on inter-vivos gifts.<sup>8</sup> It is also worth noting that transfers to spouses or charity were never subject to inheritance tax reducing the incentive to mis-report estates (Alvaredo et al. (2018) p.39). As noted by Rubinstein:

Although imperfect in several respects the probate valuations offer comprehensive and objective information on the personal wealth of the entire British population in the modern period. They are, moreover, probably unique among advanced industrial nations in presenting probate valuations for the whole population... *It is a mystery why so little use has been made of them* (Rubinstein (1977a) p.100-1, my emphasis).

---

data, in the US, 1913-2013, which requires distributional assumptions (see their discussion p.540).

<sup>7</sup>On the relationship of Pareto wealth and income distributions to theory see Jones (2015) and Benhabib et al. (2011).

<sup>8</sup>Wealth in trusts has been estimated at <1% of all wealth (Alvaredo et al. (2018) p.40). See also the same paper for a discussion of the importance of inter-vivos gifts (p. 39). Karagiannaki (2011), using the British Household Panel Survey, estimates inter-vivos transfers at 2.3% of total wealth for 2004.

Years	Valuation
1858-1898	Unsettled Personalty + Let Freeholds
1898-1925	” + Unsettled realty
1925→	” + Settled Land

Table 2.1: The Probate Valuations

*Notes:* Based on information from Rubinstein (1974, 1977a); Turner (2010). ‘Unsettled refers to cash from the sale of an asset where as ‘settled’ refers to assets that are unsold but held in trust for successive beneficiaries (see <https://www.gov.uk/guidance/inheritance-tax-manual/section-8-settled-property> for more details on the legal definitions).

It is important to note here that existing work on the distribution of wealth in England and Wales, such as Atkinson and Harrison (1978), Atkinson et al. (1989), Atkinson (2013) and Alvaredo et al. (2018) use a different data source; aggregated summary data from the Inland Revenue. There, estates are aggregated into sizes and types, and published in tabulations in the Annual Reports of the Inland Revenue.<sup>9</sup> The key difference between the PPR calendar valuations and the Inland Revenue data is that the latter are anonymous and grouped (and will be closed to the public for the next 150 years) while the former are not (Rubinstein (1974), Rubinstein (1981), Harbury (1962), Harbury and Hitchens (1979)). This allows the direct *individual*, family and surname analysis of wealth in England. The Inland Revenue estate valuations are also different from the probate valuations here: they include the property that is excluded from the Calendar valuations (table 2.1) (Rubinstein (1974) p.70).<sup>10</sup>

Previous work directly using the individual probate valuations includes Harbury (1962), Rubinstein (1977a,b, 1981) Nicholas (1999), Rothery (2007), Turner (2010), Clark and Cummins (2015a) and (2015b).

Only estates at death above a specified minimum value required an act of probate to transfer the assets. Estates below the threshold were known as ‘small estates’. Table 2.2 reports the changing definition of a ‘small estate’ from 1858-2017.<sup>11</sup> The treatment of Non-Probated wealth is described in the next section.

## 2.2 Estimated Probated Wealth at Death and its relationship to Actual Wealth during Life

The PPR Calendar valuations record a portion of wealth at death (see table 2.1). There are a number of conceptual problems extending the patterns and trends of the probated wealth of the dead to the total wealth of the living.

<sup>9</sup>Alvaredo et al. (2018) list their exact data sources by year in their online appendix pages 3-4. They use the Annual Report of the Inland Revenue, 1895-1968, the Inland Revenue Statistics Division 1968-97 and the National Archive and HMRC 1997-2013.

<sup>10</sup>See also Alvaredo et al. (2018) p.43: “Probate values tend to understate the HMRC/IR figures as they are only intended to cover all those assets which an executor must dispose of in accordance with the testator’s will (or the intestacy rules)”. They compare the Probate valuations of 74 deceased members of the *Sunday Times Rich list* with their *Rich List* estimated wealth. They find “.. for the largest estates in probates (above £200 million at 2015 prices) the List considerably *underestimates* wealth.”

<sup>11</sup>Atkinson and Harrison state that “prior to 1965 the limit was £100” (Atkinson and Harrison (1978) p.36) and Turner states that the limit “ranged from £50 to £500” (Turner (2010, p.628)) However it was possible to deduce the minimum value from the probate records themselves - there was significant clumping at the minimum value threshold and by tabulating the 5 most frequently recorded probate values by year of probate and manually checking the resulting output it was possible to work out what the limit was between 1932 and 1964.

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

Table 2.2: The Minimum Probate Threshold, 1858-2017

First the dead are not randomly sampled. Older people die in greater proportions than younger people. Therefore any claim about wealth inequality in, for example, 1980 based on probated wealth, really corresponds to those dying in 1980, born on average in 1910 and experiencing their young life during both World Wars, having families in the 1940s and 50s, working and saving from the Great Depression to the Thatcher era. It does not tell us about the average experience of someone living in the 1980s. And of course death cohorts are mixtures of different birth cohorts. There is also a potentially large life-course pattern to wealth accumulation. Failing to account for these potential effects may lead to a substantial difference between a person probated wealth and their actual wealth during their life.<sup>12</sup>

The traditional solution to these *age-composition* issues is to re-weight the observed wealth-at-death to match the age distribution of the living. Mortality multipliers (the inverse of the death rate by age) can be applied if age is available, as done in Atkinson and Harrison (1978).

Unfortunately, age at death is not reported in the PPR calendars. But as Alvaredo et al. (2018) show emphatically in their figure 6 (p.18), there is no substantive difference in the level or trend of the wealth shares by the application of mortality multipliers to the Inland Revenue estate data. No attempt to re-weight the PPR wealth data is attempted here.

The PPR data used here end in 1992. However, there has been no change in the relative shares from 1980-1985 to 2015 (see Alvaredo et al. (2018) figure 2, p. 29). Whatever process has driven the huge shift in wealth shares over the past 100 years was complete in 1980.

### 3 Building the Data

The original printed volumes of the Principal Probate Registry, calendar from 1858 to 1996, have been digitized as scanned images and are made available at <https://probatesearch.service.gov.uk> [last accessed 25 April 2016]. The data are free for academics to exploit (see the Open Government License for public sector information). Examples of the webpage interface and resulting images of the original index are shown in figure 3.1. In order to create a database of the PPR calendar suitable for academic use I created a script to mass download all the image files (e.g. figure 3.1). The search engine had the feature that all results for a two letter string code starting sequence were returned: E.g. searching for “SM” in 1850 returned all those with “SM” as the first two letters

<sup>12</sup>Another issue are status based selection effects. In England, as everywhere, the rich die older than the poor. Further, this socio-economic gradient in lifespan has changed over time (See for example ONS, 2015).



1905.

- 509 **CUMING** William Herbert of "Kani Tal" College-place Southampton retired major-general in H.M. army died 20 December 1904 Probate **London** 11 April to Robert Octavius Cumming retired colonel in H.M. army and Robert Stevenson Dalton Cumming captain in the Royal Navy Effects £19416 17s. 8d. *Resworn* £19856 17s. 8d.
- STEWART or GORDON**  
**CUMMING** Elizabeth Newton. See "STEWART."
- 28 **CUMMING** Hugh Macpherson of Belmont Shanklin Isle-of-Wight died 26 November 1904 Probate **London** 18 January to John Charles Cumming esquire and John Durham solicitor Effects £849 3s. 4d.
- CUMMING** John of 118 Club Garden-road Sheffield died 23 June 1905 Probate **Wakefield** 11 August to Thomas Henry Cumming elementary-teacher Effects £285 4s. 2d.
- CUMMING** John Morley of Pathah Ram Assam India died 29 February 1905 Administration **London** 19 April to John Cumming journalist Effects £254.
- 1419 **CUMMING** Mary of 3 St. Stephen's-mansions Monmouth-road Bayswater Middlesex widow died 2 October 1905 Probate **London** 6 November to William Gordon Cumming colonel in H.M. army Effects £1242 0s. 3d.
- CUMMING** Robert of 14 Murrayfield-road Murrayfield Midlothian died 28 July 1905 at Edinburgh Confirmation of Janet Campbell Thomson Clark or Cumming widow William Oliphant Cumming solicitor and Mary Berwick Cumming and Jessie Grace Cumming spinsters and George Watt Sealed **London** 23 November.
- CUMMING** Robert Octavius of "Coulter" Gbeltenham retired lieutenant-colonel in H.M. army died 7 October 1905 Probate **Gloucester** 17 November to Robert Stevenson Dalton Cumming captain in the Royal Navy Walter Charles Cumming cleric and Annie Elizabeth Cumming spinster Effects £109270 1s. 4d. *Resworn* £108291-9-0.
- CUMMINGS** Ann of 3 Dewsbury-court York-street Swansea widow died 20 May 1905 at Terrace-road Swansea Administration **London** 27 November to Margaret Ann Ellery (wife of James John Ellery) Effects £30.
- CUMMINGS** Catherine Emily of 119 Hamilton-street Newcastle-upon-Tyne spinster died 2 August 1905 Probate **Newcastle-upon-Tyne** 3 September to Alexander Mark Turnbull solicitor and Thomas Gray solicitor's-clerk Effects £456 12s. 6d. *Seaworn* £ 716 " 7 " 8
- CUMMINGS** Emily Sophia of Lurganboy county Leitrim widow died 30 March 1905 Probate Ballina to Edwin Tyrrell Cummings M.D. Sealed **London** 14 November Effects £208 in England.
- CUMMINGS** George John of 32 Park-road Newcastle-upon-Tyne gentleman died 16 November 1904 Probate **Newcastle-upon-Tyne** 10 January to John Ditchburn grocer and Alexander Mark Turnbull solicitor Effects £2679 7s. 3d.
- 1541 **CUMMINGS** Richard of the "Half Moon" inn Billingham county Durham died 12 December 1905 Probate **London** 27 December to Janet Hannah Cummings widow Effects £232 11s. 6d.
- CUMMINGS** Sarah Jane of 9 Froddington-road Southsea Portsmouth (wife of Richard Cummings) died 17 January 1905 Probate **Winchester** 10 February to the said Richard Cummings pensioned-leading-stoker from the Royal Navy Effects £162 13s.
- 805 **CUMMINGS** Thomas of 19 Albert-embankment Lambeth Surrey died 29 March 1905 Probate **London** 29 June to Louisa Spicer (wife of Robert Spicer) Effects £87 16s. 7d.
- CUMMINS** Isabella of 8 Tchidy-terrace Falmouth Cornwall (wife of Henry Royle Cummins) died 27 November 1905 Administration **Bodmin** 28 December to the said Henry Royle Cummings physician and surgeon Effects £400. *Resworn* £ 475-7-7
- CUMMINS** Robert of 40 Newgate-street Bishop Auckland county Durham died 9 November 1900 Probate **Durham** 13 April to Mary Ann Cummins widow Effects £66 12s. *Resworn* £ 1638-12-
- 938 **CUMNER** Thomas of 15 Prince-street Deptford Kent died 3 March 1905 Probate **London** 27 July to John Stagg warehouseman Effects £15 4s.
- CUNDELL** Charles John of Haughton-le-Skerne Darlington died 19 January 1905 Administration **Durham** 28 February to Jane Cundell widow Effects £217.
- CUNDELL** Leonard of Hungerford Berkshire gentleman died 14 October 1905 Probate **Oxford** 23 December to Henry Stratton Cundell bank-manager and Matthew Henry Cundell agent Effects £1909 18s. 6d.
- CUNDILL** George of Scagglethorpe Yorkshire farmer died 4 November 1905 Probate **York** 16 December to Robert Cundill and George Arthur Cundill farmers Effects £610 16s.
- CUNDY** Francis of the "Gables" Watling St. Austell Cornwall retired farmer died 13 July 1904 Probate **Bodmin** 8 July to Annette Buckthorp singlewoman and Ernest Horatio Richards general-merchant Effects £1274 18s.
- 936 **CUNDY** the reverend Henry George of "St. Margaret's" Shortlands Kent clerk D.D. died 6 June 1905 Probate **London** 11 July to Robert Moxon esquire and Richard Howlett barrister-at-law Effects £5039 5s. 11d. *Resworn* £ 5230-9-4
- CUNDY** Robert of the Joint-counties-lunatic-asylum Carmarthen died 30 August 1905 Administration **London** 28 September to Emma Jane Jones (wife of Thomas Jones) Effects £344 14s. 10d.

Figure 3.1: The Scanned Images

**WOODHAMS** Sarah. 16 January. **The Will of** Sarah Woodhams late of 61 Rose-Hill-terrace Brighton in the County of Sussex Personal Estate £1,210. Spinster who **died** 25 December 1883 at 61 Rose-Hill terrace was proved at L6W6S by Edward Saxby Woodhams of 12 East-street Brighton Broker the Nephew and Emma Scrase of 10 ..Brunswick-place North Brighton Widow the Sister the Executors.

**LOCKETT** John of 1 Bath-street Burslem Staffordshire leather cutter **died** 13 November 1893 **Probate** Lichfield 21 December to Mary Lockett widow Effects £148 6s. 7d.

**ASTON** Alfred Charles of 3 St. Annes Road Willenhall Staffordshire **died** 20 August 1906 **Probate** Birmingham 19 October to James Henry Edwards floor moulder and Joseph Price production controller. £1772.

**LICCINS** Edward of 150 Whitmore-park Holbrooks-lane Coventry **died** 19 April 1929 **Probate** Birmingham 8 June to Maria Liggins widow. **Effects** £48 8s.

**KNEEBONE** Julia of 89 Meneage-street Helston Cornwall widow **died** 12 May 1934 **Administration** Bodmin 18 June to Harold Dennis Kneebone hotel proprietor. **Effects** £86 Is. 7d.

**ALLIX** Elizabeth Barbara of 4 Hanover Court Hanover-street London W.1 spinster **died** 19 April 1945 **Probate** Llandudno 28 August' to Martins Bank Limited **Effects** £53380 Is. 3d.

**LACEY** John Ethelbert of 35 Wynfield-road Leicester **died** 30 November 1952 **Probate** Leicester 27 January to Midland Bank Executor and Trustee Company Limited. **Effects** £6883 0s. 2d

**KEAM** John Arnold of 42 Berkeley Avenue Barklngsldde Ilford Essex **died** 19 August 1961 at Dagenham Hospital Essex **Administration** London 5 February to Agnes Alice Keam widow. **Effects** £3542 12s. 6d.

**BECKHAM**, Ronald Arthur of 117 Halfway Rd Sheenness Kent **died** 1 May 1975 **Administration** Brighton 18 June £3179 750519140E

**BECKHAM**, Robinson of 58 Woodside Barnard Castle County Durham **died** 10 July 1980 **Probate** Newcastle-upon-Tyne 17 October £24275 802403103M

Figure 3.2: Examples of Entries, 1880s-1990s

of their last name probated in 1850. I searched <https://probatesearch.service.gov.uk> for every two letter combination from “AA” to “ZZ” for every year from 1858 to 1996 inclusive, and recorded the number of result pages. This information allowed me to construct the base urls that would lead me to each of the scanned images e.g. <https://probatesearch.service.gov.uk/Calendar?surname=Cummins&yearOfDeath=1905&page=1#calendar> directs to the image in figure 3.1. This led to the creation of 1,013,056 urls that were used to download the index images. This process was automated.

Optical Character Recognition (OCR) was performed for each of the index pages using *ABBYY* fine-reader 12 software. This software performed by far the best in terms of fidelity and consistency (as compared to free software such as Google’s *tesseract*). Over one million images were collected into 504 PDF files of 2,000 pages each. This process was also automated.

The OCR process resulted in 504 ‘dirty’ (unformatted, full of duplicates) text files. These files were merged into 10 larger text files. Next the text patterns underlying the PPR Calendar entry structure were deduced by inspection (as indicated by the bolded text in figure 3.2). The patterns are reported in table 3.1. These features were found, marked and parsed using regular expression in the *Perl* computing language (executed as command line .bat files). Following this a set of touch-ups was conducted using macros in the text manager *Ultra-edit*. The resulting ‘semi-clean’ text files were imported to R and marked patterns in the text were found using *SQL* language (via the R package *squidf*). The code converted the text to a database of rows and columns. Unique entries were identified and duplicates discarded. After this, individual fields were constructed from the relative position patterns. Years and wealth were attributed. Much code was then devoted to cleaning the wealth measure as many entries had £ and shilling values conflated and post 1970 many

<i>Description</i>	Example
A sequence of capital letters, delimits a new record/line	KAYE
delimits address	Willie " of "
delimits death date	Church View Kirkburton " died " 12 December 1946
Type or Record (always "Probate" or "Administration")	Administration London
delimits executor	3 May " to "
delimits taxable estate amount	Benjamin Kaye boot and shoemaker Effects £1082 9s. Id.

*Notes:* Many variations of the above were employed.

Table 3.1: General Entry Patterns

records had the effects value conflated with a 10-digit record number (resulting in many decedents with astronomical levels of wealth, see the last 2 entries in figure 3.2 for example). This cleaning process was a combination of manual, by-eye, checking to discover patterns of problems and coding via SQL in R to clean the main data.

All nominal millionaires (over 4,000) were checked one by one via [www.ancestry.com](http://www.ancestry.com) and the probate service website. Some examples of problematic and unusually rich entries are reported in figure 3.3. This process, summarized in figure 3.4, resulted in a database of  $N = 15,152,822$ , all with full name, street address, date of death and wealth at death.

In addition I formed a database of the number of probates granted by surname from 1997 to 2017. From a 100% of the 1881 census ((Schurer and Woollard, 2000)) and the 100% samples of births, marriages and deaths and the probate calendar 1892-1992 detailed previously, a master list of 2,590,144 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 in one of the birth, marriage or death registers, 1838-1973 at least once. This resulted in a final, 'non-zero register' surname master-list of 389,249. Each surname from this master-list was entered into <https://probatesearch.service.gov.uk/#calendar> and the count recorded(GOV.UK, 2018).

**SICH** Francis of Corney House Chiswick **Middlesex** died 13 February 1912 Probate **London** 20 March to Alexander Sich Alfred Sich and George Sich brewers. Effects £118107 1s. 3d.

viscount

**PORTMAN** the right honourable Gerald Berkeley seventh of The Manor Healing **Lincolnshire** died 3 September 1948 Probate (limited to settled land) **London** 23 April to the honourable Ian Leslie Melville banker and the most noble Edward William Spencer K.G. duke of Devonshire. Effects £4249000. Former Grant P.R. 21 February 1949.

**VANDERVELL** Guy Anthony of Brockhurst Park Stoke Poges **Buckinghamshire** and of 24 Sloane Avenue **London S.W.3** died 10 March 1987 at 18 Bryanston Square London W.1 Probate **London** 14 April to Gerard Wilfred White and Rudolph Edgar Francis de Trafford bankers and Joseph Leonard Reed solicitor. £10950148. Memorandum of Order dated 21 June 1988 under Section 26 of the Matrimonial Causes Act 1965 Annexed to Will.

**BERRY, Benjamin** of 3 Old Bank Ripponden nr Halifax West **Yorks** died 24 March 1953 Probate Leeds 22 April £821305196c

**BAMBRIDGE, Elsie** of Wimpole Hall Arrington Royston Herts died 23 May 1976 Probate **London** 27 October £1600896 760119907N

**ELLERMAN** C.H. sir John Reeves baronet of 1 South Audley-street **Middlesex** died 16 July 1933 at Hotel Royal Dieppe France Probate **London** 12 August to Frederick George Burt secretary and sir John Reeves Ellerman baronet. Effects £17224425 3s. 8d. Resworn £25817786 11s. 8d.

**VERNON** Thomas Thorneycroft of Shotwick Park near **Chester** died 24 January 1919 Probate **London** 5 April to John Herbert Vernon flour miller Thomas Cooper bank manager and Elizabeth Cooper Vernon spinster. Effects £324388 12s. 11d.

**CAYZER** sir Charles William of St. Lawrence Hall **Isle of Wight** of Gartmore **North Britain** and of 109 Hope-street **Glasgow** baronet died 28 September 1916 at Gartmore Probate **London** 24 January to August Bernard Tellefsen Cayzer and Herbert Robin Cayzer shipowners Admiral sir John Rushworth Jellicoe G.C.B. O.M. G.C.V.O. first sea lord of the Admiralty and Robert Alexander Hill solicitor. Effects £2204148 9s. 3d. Further Grant 3 October 1917.

**CAYZER** sir Charles William of St. Lawrence Hall **Isle of Wight** of Gartmore **North Britain** and of 109 Hope-street **Glasgow** baronet died 28 September 1916 at Gartmore Probate **London** 3 October to Harold Stanley Cayzer shipowner. Effects £1946676 11s. 7d. Former Grant January 1917.

**SICH** Francis of Corney House Chiswick **Middlesex** died 13 February 1912 Probate **London** 20 March to Alexander Sich Alfred Sich and George Sich brewers. Effects £118107 1s. 3d.

Figure 3.3: Examples of Problematic and unusually rich Entries

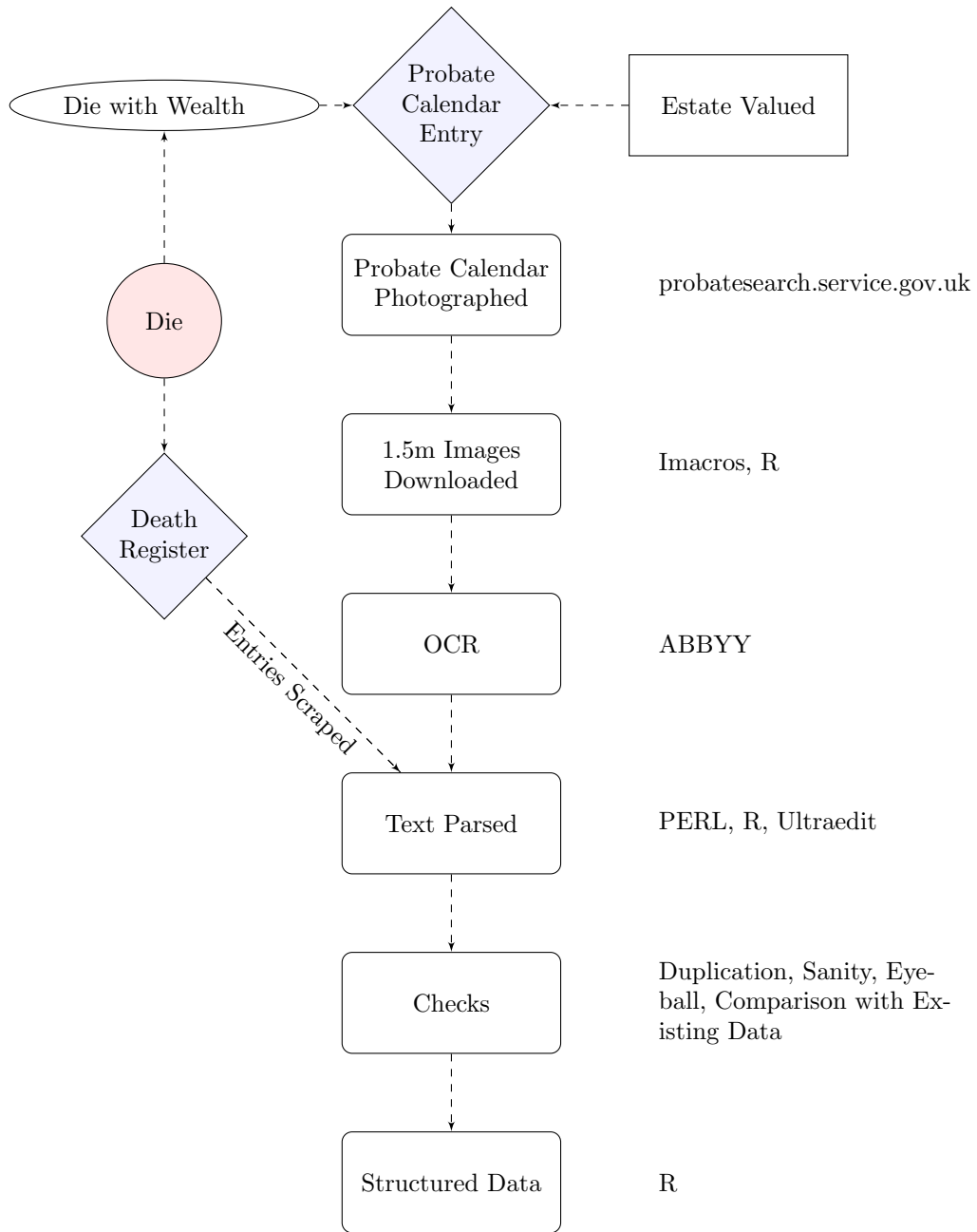


Figure 3.4: The Data-Building Process

## 4 Calculating the Components of the Wealth Distribution

### 4.1 Combining Probates and Deaths, Inferring Wealth

To understand the evolution of wealth inequality at death in England and Wales it was necessary to combine the probate data with the complete death registers, also collected for this project. This process is described in the appendix sections D and E. Table 4.1 reports the number of probates and deaths over 20 by decade from 1892 to 1992. The proportion of adults receiving probate after death rises from 15% in the 1890s to 40% by 1992. For the analysis of overall wealth inequality, the wealth of these omitted decedents has to be inferred. Following the standard method of the HMRC, I assigned each non-probated adult a wealth equal to half the level of wealth observed in the probate calendars for the year of death that was below the threshold<sup>13</sup> (Turner (2010, p.628-9)). Figure 4.1 reports the total value of real wealth, 1892-1992 for the probated population and the inferred excluded population. Despite the fact that the probated population is a minority, only 15-40% of the total death population in any year, the application of an inferred minimum wealth to the remainder implies that on average 98% of all wealth is being captured by those probated. (Changes in wealth inequality over time can thus be captured by accounting for the changing probate rate and the shifting shares of the top wealth holders within the probated class.) 30,696,529 observations representing the excluded population were added to the probate data and assigned wealth as described above. Wealth shares were calculated by finding the percentile wealth at various cutoffs, assigning a dummy to each wealth observation indicating which percentile range it fell into then summing wealth across these ranges, by year.

### 4.2 Descriptive Results

#### 4.2.1 Comparison with Existing Estimates

Figure 4.2 compares my estimates of the top .1, 1, 5 and 10% shares of the wealth distribution with recent estimates from Alvaredo et al. (2018). In general, my estimates for the share of the top 10 and 5% wealth shares appear to be over-estimated relative to Alvaredo et al. (2018) but there is a striking correspondence for our estimates of the top .1 and 1% shares.<sup>14</sup> Further the trends in all series are the same, until the 1980s. The estimates agree that inequality stopped rising but my estimates show a higher share of wealth for the upper percentiles that is even, possibly, increasing. However, post 1980, the PPR calendars increasingly use ‘banded’ wealth estimates that are clearly loosely applied (this is discussed further in the appendix).<sup>15</sup> The picture is clear and well known. The ‘Great Equalization’ of the 20th century is driven by declines in the concentration of wealth and income from capital. But how were the spoils distributed between the wealth-holding class, those who had wealth sufficient to warrant probate at death, and those who had nothing?

---

<sup>13</sup>Many estates were valued below the threshold level in the Probate Calendar.

<sup>14</sup>Due to the exclusion of settled property before 1925, as reported in table 2.1, the shares of the top on percent are likely underestimated (see also Alvaredo et al. (2018) p.33).

<sup>15</sup>As discussed, Alvaredo et al. (2018) use the ‘estate multiplier method to adjust decedents estate tax data to match that of the living population. My series is constructed without this adjustment. This adjustment does not make a serious difference to their estimates.

Period	N Probates	N Deaths > 20	Probate Rate
1892-1900	376,292	2,498,315	.15
1900s	624,951	3,558,901	.18
1910s	588,213	3,591,196	.16
1920s	811,868	3,696,695	.22
1930s	1,196,711	4,222,475	.28
1940s	1,715,975	4,627,277	.37
1950s	2,034,470	4,980,193	.41
1960s	2,292,231	5,376,468	.43
1970s	2,324,704	5,676,328	.41
1980-1992	2,929,142	7,330,740	.40
All	14,894,557	45,558,588	.33

Table 4.1: Counts of Probates and Adult Deaths, 1892-1992

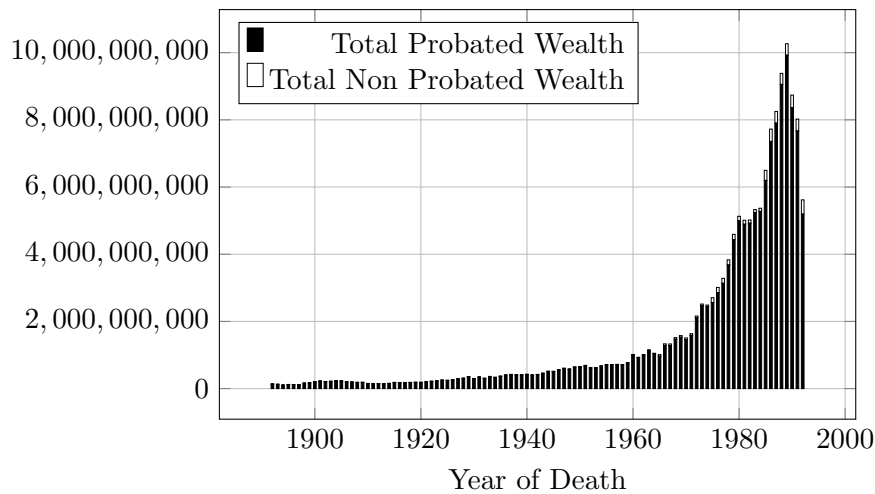


Figure 4.1: Nominal Probated and Non Probated Wealth in England and Wales, 1892-1992

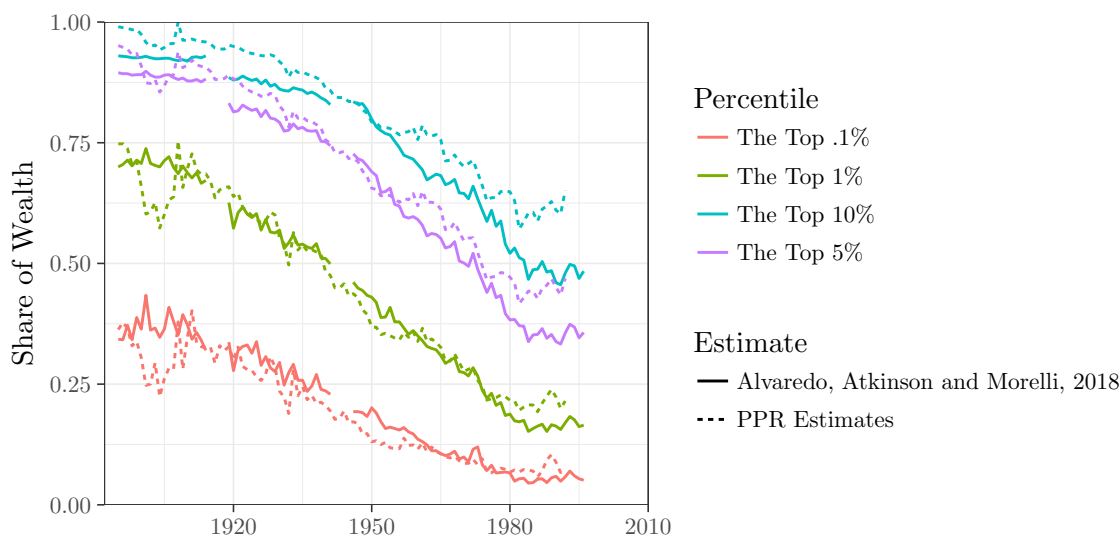


Figure 4.2: Comparing Different Estimates of Top Wealth Shares, England 1892-1992

#### 4.2.2 Overall Inequality Over Time

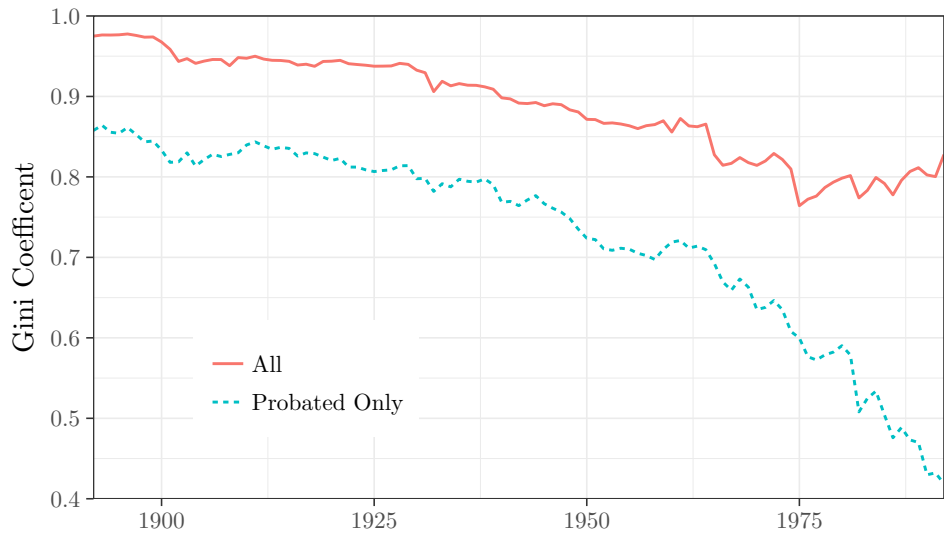
Figure 4.3 (a) reports the Gini coefficient of probated and total wealth by year of death and figure 4.3 (b) reports the proportion of English dead probated over the same period. There are three immediate facts that these two figures indicate about the evolution of the English wealth distribution. Firstly, total wealth inequality declined significantly from a Gini of over .9 to .8, between 1892 and 1980. Second, after the mid 1970s, inequality in probated wealth fell but total inequality plateaued.<sup>16</sup>

Third, and most importantly for the average English: the proportion of decedents that have wealth significant enough to merit probate rate has been flat since the end of the second World War to 2016. This simple finding is quite stark. Despite the great equalization of wealth over the 20th century, most English have no significant wealth at death. This is even more surprising considering the fact that the nominal threshold for probate (now £5,000) was only upwardly revised sporadically (see table 2.2) and was the same from 1984 to 2018 (this is the threshold today). A modest rise in the wealth of the ‘middle class’, the average English decedent, coupled with inflation, should have resulted in a rocketing probate rate and a far greater increase in the wealth share. Yet this is not evident.

As figure 4.2 and 4.3 indicate this is a story of a reshuffling of the share of the top .1 to 10 % to the rest the probated population. The bottom 60% of English have seen no increase whatsoever in their wealth share in the ‘great equalization’. Of course, inter-vivos bequests could obscure the true pattern of wealth holding. But if this is the case we would expect the results to find a *greater* rise of the middle class. Given the ‘progressive’ (/confiscatory) nature of the top marginal rate of inheritance taxes after 1950, figure A.1 in the appendix, we would expect the rich and the very rich

<sup>16</sup>Caution should be exercised with the post 1980 inequality results due to the decline in the quality of the valuations.





(a) Gini Coefficient in Wealth



(b) Proportion Probated

Figure 4.3: Overall Wealth Inequality, England 1892-2016

Source: 100% Probate Calendar sample and 100% death index, 1892-1007. ONS death series 1838-2017 and probate counts 1997-2018. Probate rates are interpolated from 1992-1996.

to have a greater proportional incentive to dispose of as much wealth as possible before death.

What of the reshuffling of wealth within the top 30%? Uniquely, the PPR data allow us to estimate wealth share below the top 10%.

### 4.2.3 Percentile Wealth Shares Over Time, Detailed Breakdown

Figure 4.4 reports the shares of the top percentiles of the wealth distribution annually from 1892 to 1992. In contrast to figure 4.2, these estimates are for non-overlapping percentiles (hence the top .1-1% do not include the top .1% and so on). The results for the top 10% mirror earlier work by Atkinson and Harrison (1978), Atkinson et al. (1989) and Atkinson (2013). The top .1% and top .1-1%, graphed in figure 4.4 (a) account for a consistently decreasing share of all wealth until about 1975. The top .1% hold 36% of all wealth in 1892 and the top .1-1% hold 38% - meaning that the top 1% hold 74% of all wealth in England and Wales. This declines to 22% by 1975 (7 and 15% of all wealth is held by the top .1% and the top .1-1% respectively). Thereafter their shares are roughly constant to 1992.<sup>17</sup>

Two aspects of the decline of these very top shares are surprising. Firstly, the decline is apparent well before 1940. Secondly, the plateaux in the decline of the share of the super rich coincide with the oil shocks of the 1970s and the end of the European ‘Golden Age’ of post-war economic growth.

The share of the top 5-1%, graphed in figure 4.4 (b) has held roughly constant over the observed century, as also noted by Atkinson et al. (1989) p.319, (but does rise and decline in the series reported here), whilst the share of the top 10-5% has consistently risen, from 4% of all wealth in 1892 to 17% in 1991.

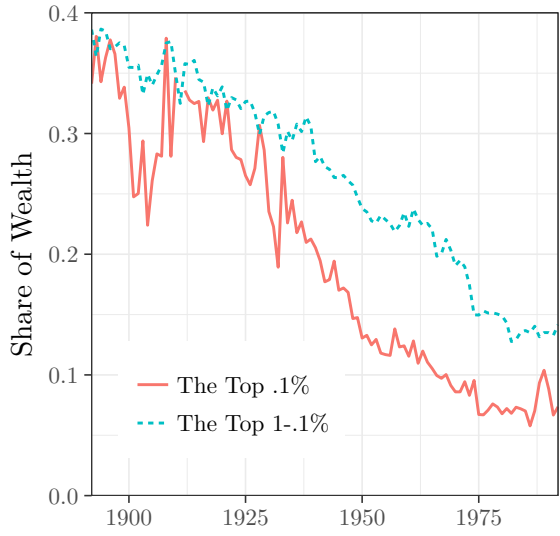
Figure 4.4 (c) reports the dynamics of even 10% bins of the top half of the wealth distribution. What emerges here - and is emphasized in figure 4.4 (d) - is that the decline of the share of the top 1% of wealth is entirely absorbed by the top 10-5%, the top 80-90% and the top 70-80%. Despite the choppiness of the estimates for the lower percentiles, it is clear that for the top 50-70% there astonishingly little growth in the wealth share. Further, the rate of increase of the share of wealth held by all percentiles below the top 10% is negatively related to the percentile. For example, the 80-90th percentile increase their share from 1.5% in 1892 to 25% in 1992, the 70-80th percentile go from <1% to 12% in 1992.

Figure 4.5 summarizes the trends in the complete distribution of English wealth at death as revealed by the PPR calendars. Over the century 1892-1992, the top 10% reduce their wealth share from over 99% to 65%.<sup>18</sup>The relative winners are the 20% immediately below them. Where is the growth of the ‘middle class’? Below the top 30%, the bottom 70% go from around .02% of all English wealth to around 2% on 1992, a ten-fold increase, but tiny in absolute terms. These trends, while unprecedented and transformational, do not translate into a rise of a broad based ‘middle’ class.

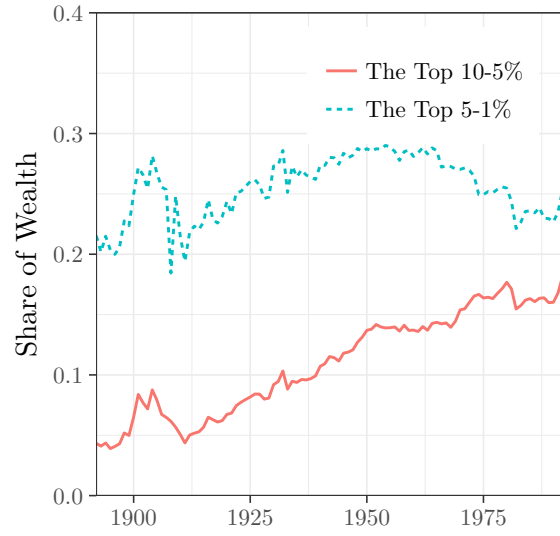
---

<sup>17</sup>The top .1% share moves around a bit before 1900 - it is by definition a small group - there are 345 members of the top .1% in 1900.

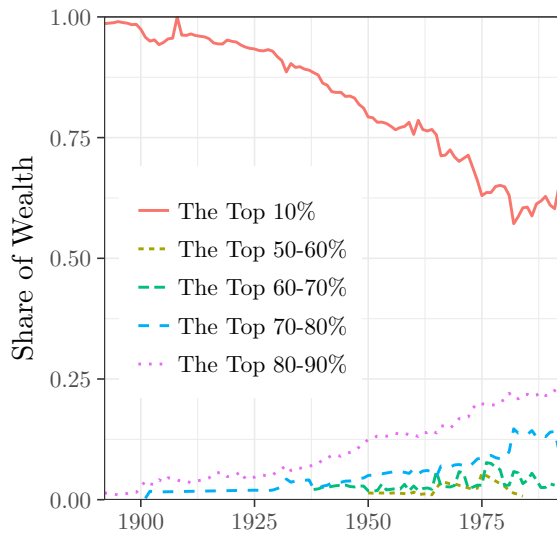
<sup>18</sup>The bottom 90% estimate of the wealth share at 35% in 1980-1992 matches estimates from the US by Saez and Zucman (2016).



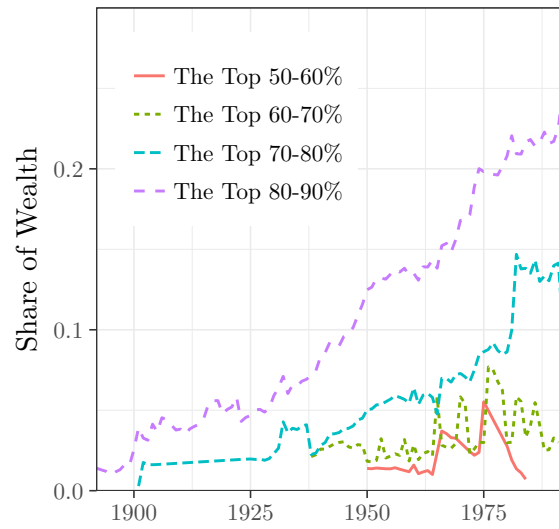
(a) The top .1% and .1-1%



(b) The 90-95th and 95-99th Percentile



(c) The Shares of the Top 50%, Equal 10% Intervals



(d) The Shares of the Top 50-90%, Bigger Scale

Figure 4.4: Top Wealth Shares, England 1892-1992

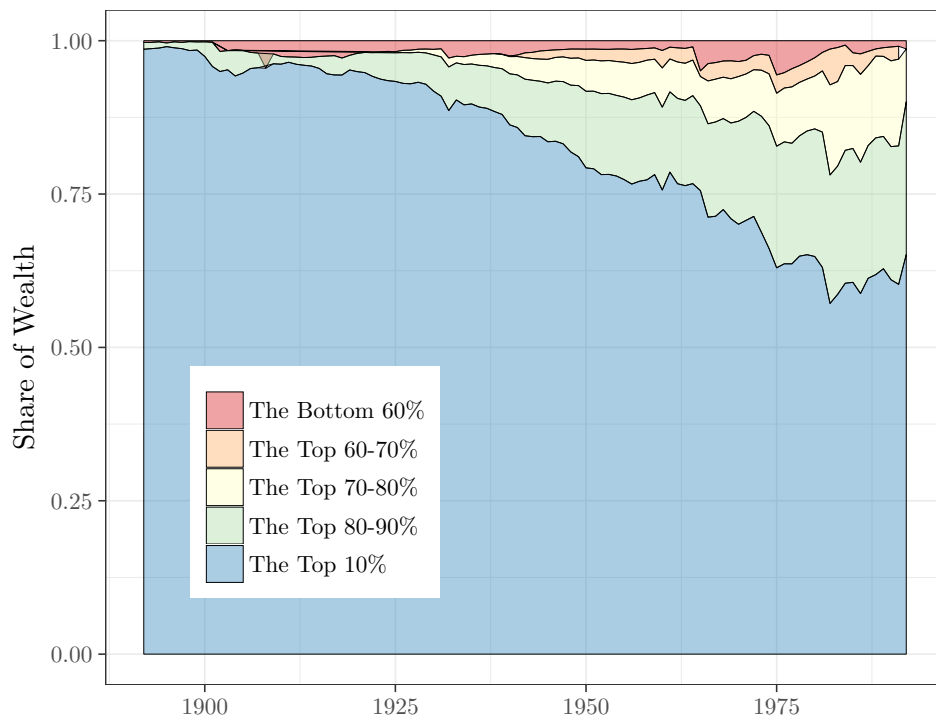


Figure 4.5: English Wealth Holding by Decile, 1892-1992

## 5 The Shape of Wealth

This section characterizes the shape of English wealth, 1892-1992. Figure 4.4 (a) suggests 8 natural periods, based on historical events and structural breaks in the share of the top 1%, to examine the shape of the wealth distribution: Pre War (1892-1914), WWI (1914-8) The inter-war years (1918-39), WWII (1939-45), the post war years and the 1950s (1945-60), 1960s, 1970s and 1980-1992.

### 5.1 Lorenz Curves; All Wealth and Wealth-Holders Only

Figure 5.1 reports Lorenz Curves, by period for English wealth for all decedents over 20 (a) and for wealth-holders only (b). Consistently the Lorenz curves shift over time towards the line of perfect equality. As with the earlier share calculations, the vast majority of the distributive gains from the dilution of the extreme concentration of elite wealth in the 1890s and 1900s are the wealth-holding class above the median English. The Lorenz curves reveal 3 phases in the Great Equalization of the English Wealth distribution: A slow but persistent movement towards greater equality 1892 to 1945 followed by two great leaps, from the 1950/60s to the 70/80s.<sup>19</sup>

### 5.2 The Shape of the Upper Tail

In 1896, Vilfredo Pareto noted that 80% of Italian land was held by 20% of Italians. This fact, together with other income and wealth data he collected, from other countries and times, led him to develop his famous power law model of distribution. In the Pareto distribution the probability of observing wealth above some cut-off,  $w$ , is equal to  $w$  raised to some power:

$$Pr[Wealth \geq w] = w^{-1/\alpha}$$

where  $\alpha$  is the critical ‘shape’ parameter. In Pareto’s initial observation of the inequity of Italian landownership,  $\alpha = \log_4(5)$ . If a distribution is Pareto distributed we can easily calculate the share going to the top  $p$  % as  $(\frac{100}{p})^{\alpha-1}$  (Jones (2014)).

Alternatively, wealth will have a lognormal distribution if  $\ln(Wealth)$  is normally distributed. If this is the case in reality; calculation shortcuts based on Pareto distributional assumptions will give incorrect inferences regarding the shares of the unobserved percentiles.

With the complete English wealth distribution at the individual level at my disposal I can define whether Pareto or log normal describe the data better - both for the probated population (depending on period, this is between the top 20 to 40%) and separately the extreme upper tail, the top point-one-percent.<sup>20</sup>

Here I assume that any ‘missing’ or hidden wealth (say in trust finds or in offshore accounts) is directly proportional to observed wealth. A separate paper, Cummins (2018), directly estimates the amount of wealth hidden by English dynasties.

---

<sup>19</sup>The results for the 1980s are not reported for all because the banding problem (discussed in the appendix) prevent accurate estimates of the wealth distribution. However by discarding the banded estimates and assuming that the remaining are a unbiased sample the Lorenz curve in figure 5.1 (b) can be estimated. Of course, caution should be exercised before over-interpreting these estimates.

<sup>20</sup>Alvaredo et al. (2018) estimated Pareto coefficients from “shares within shares” (p.35). They warn that assuming a Pareto distribution may “miss a potentially important element of the change...the assumption of Pareto-distributed wealth might not be a compelling one”. Further, they find that their estimated Pareto coefficients fail to fit the extreme upper tail as revealed by the Sunday Times Rich List (p.42).

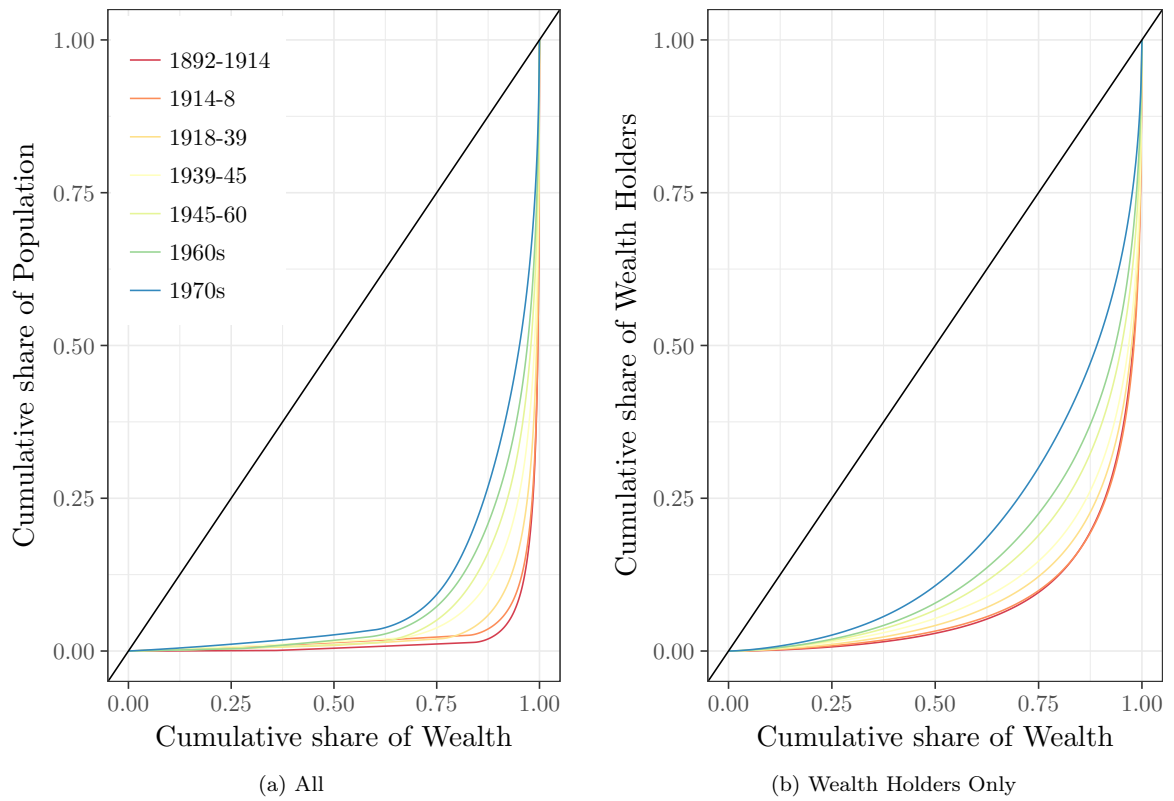


Figure 5.1: Lorenz Curves for Real Wealth, England, 1892-1992

In 1892, in England, the top 20% held 99.9% of English wealth. The top 1% held 74%. Does the English wealth distribution follow a Pareto type power law? Or is it log-normal? To answer this I compared the empirical distribution of probated estates with those predicted by 2 distributional forms: log normal and the classical power law distribution of Pareto. <sup>21</sup>

Figures 5.2 and 5.3 report quantile-quantile plots of the empirical versus theoretical distribution of decedent estate values for the log-normal and Pareto distributions, by period. Both axes are in log scale and the straight line indicates a perfect fit. The closer the theoretical distributions are to this line - the better the fit. Of course, neither distributions describe the distribution perfectly. However it is clear that the log normal distribution is a better approximation of the data than Pareto. This is true for every period to 1945 - After that both theoretical distributions fit poorly.

To put more precision on this judgement, the fitted distributions were then assessed on their absolute difference from the empirical distribution. As this is wealth and we have a special interest in predicting the tails, Anderson-Darling statistics were calculated as:

$$n \int_{-\infty}^{\infty} \frac{(F_n(x) - F(x))^2}{F(x)(1 - F(x))} dx$$

Anderson-Darling statistics equally emphasize the tails as well as the main body of a distribution and are often used in risk assessment (Delignette-Muller et al. (2014) p.7). They are reported for each of the fitted distributions in table 5.1. The lower the number, the better the fit.

Table 5.1: Anderson-Darling Statistic for Fitted Distributions, All Wealth Holders, by Decade

Period	Log mal	Nor- Pareto
1892-1914	3,851.42	2,955.15
1914-8	1,461.46	1,398.35
1918-39	8,961.96	11,302.69
1939-45	4,936.47	8,038.87
1945-60	5,123.81	15,188.11
1960s	1,194.14	7,230.32
1970s	4,018.89	7,120.75
1980-92	23,134.10	6,660.65

The result is clear - Despite the gigantic level of English wealth inequality, it's distribution is at least as well - if not better - described by a log-normal distribution as opposed to a Pareto power law<sup>22</sup>. Surprisingly, the graphical analysis reveals that the log-normal does a better job of approximating the upper tail in every period apart from 1945-60 and 1960-70.

However, it has been observed that many process can be described as log-normal over a lower range and power law distributed above a threshold (Bee et al. (2011)). The PPR data are unique in that the individual estate values of the top point-one-percent are directly observed. Does the distribution of that extremely rich group follow a power law?

<sup>21</sup>These distributions were fitted using the 'fitdistrpluis' package in R (Delignette-Muller et al. (2014)).

<sup>22</sup>Of course, I could test the fit for many more complex distributions such as Burr and other Generalized Pareto distributions. Further, the distributions could be checked for the top 1, 10, 20 and so on, percentiles. The virtue of the (log) normal and Pareto distributions are their simplicity and for this reason I restrict my analysis to them. Future research, using the individual data that I will make available upon publication, could proceed in the direction of a more accurate (but at the expense of complexity) fit.

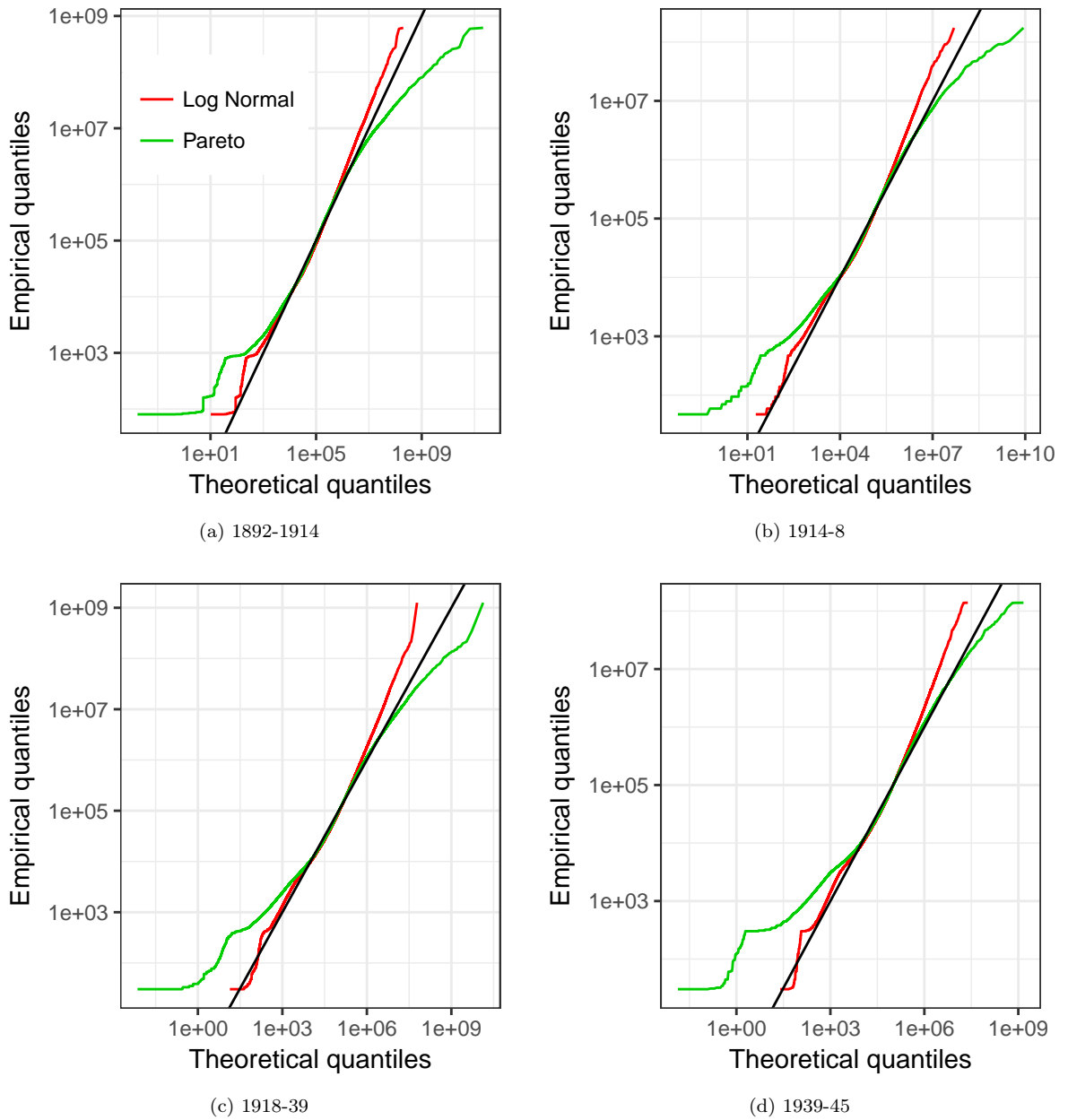
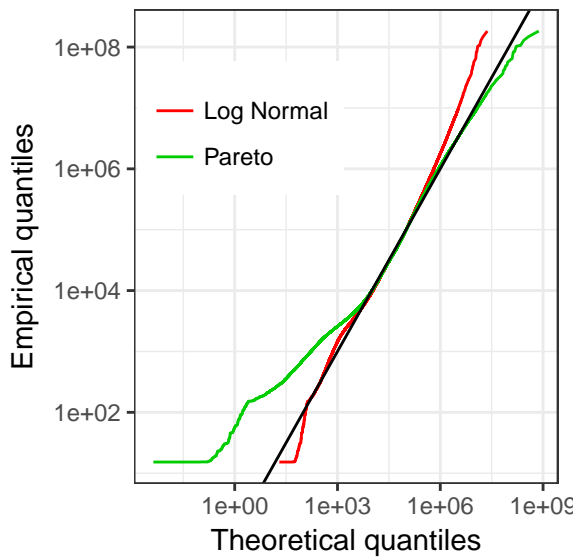
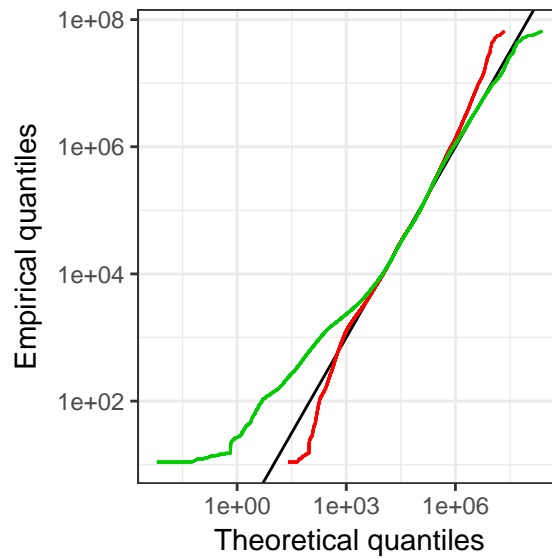


Figure 5.2: Empirical v Fitted Wealth Distributions, by Period, England, 1 of 2

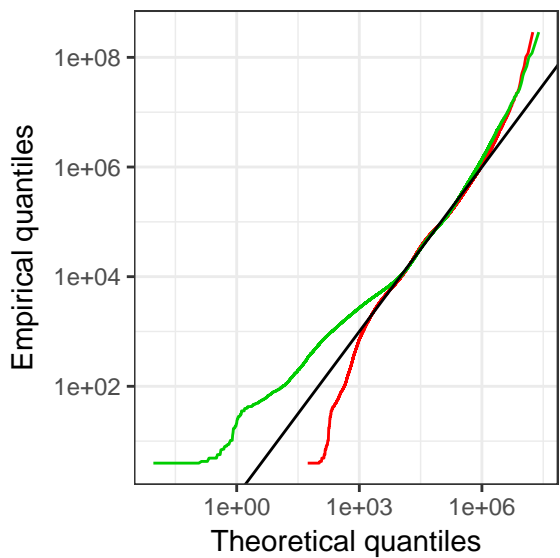




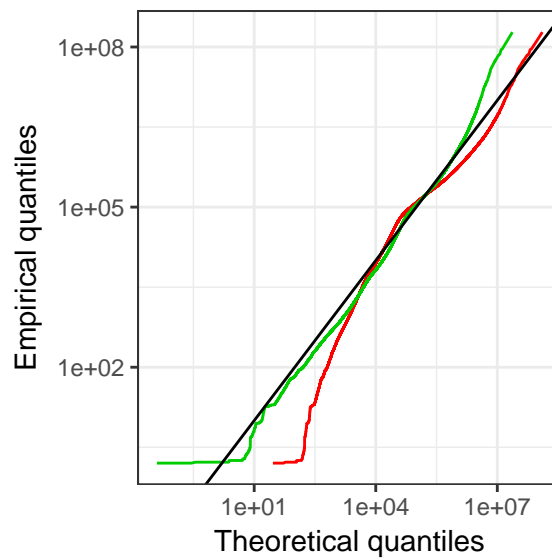
(a) 1945-60



(b) 1960s



(c) 1970s



(d) 1980s

Figure 5.3: Empirical v Fitted Wealth Distributions, by Period, England, 2 of 2

Rank	Name	Death Year	Place	Real Wealth
1	Sir John Reeves Ellerman, Baronet	1933	London	1,257,371,575
2	Henry Overton Wills	1911	Somerset	433,864,083
3	James Buchanan, 1st Baron Woolavington	1935	Cornwall	345,796,993
4	Charles Cross	1974	Hampshire	288,513,883
5	Wentworth Beaumont, Baron	1907	London	279,678,377
6	John Gretton	1899	London	265,401,845
7	Sir Andrew Barclay Walker	1893	Liverpool	260,701,343
8	Panaghi Athanarius Vagliano	1902	London	256,213,849
9	William Orme Foster	1899	Shropshire	238,162,639
10	William Louis Winans	1897	Brighton	233,821,183

Notes: 2015 prices.

Table 5.2: The 10 Richest English, 1892-1992

### 5.3 The Shape of the Extreme Upper Tail

The 10 richest English who died in the sample period are reported in table 5.2. As discussed in section 2, all extreme wealth values were also checked by eye. Further all 40,074 of the top point 1 percent, 1892-1992 were checked for duplicates by eye.<sup>23</sup> Of the top 10 listed in table 5.2 all are known to be wealthy (apart from #4, Charles Cross).<sup>24</sup> Figure 5.4 reports quantile-quantile plots of the empirical versus theoretical distribution of decedent estate values of the top .1% for the log-normal and Pareto distributions, for a selection of periods. Table 5.3 reports the same Anderson-Darling goodness-of-fit statistics as before. Again, and even more surprisingly, the log-normal distribution does a much better job of approximating the true distribution of the wealth of the top point one percent. This result is driven by the Pareto distribution badly underestimating the share of the left tail of the top point one percent.

A simple Pareto approximation of extreme English wealth provides a poor fit to the tails of both the complete wealth distribution - and that of the top point one percent. Log-normal is a significantly better approximation than Pareto. However, for estates above £100,000,000 (2015 pounds), all hell breaks lose and the log-normal badly underestimates the tail (as indicated by the predicted fits lying to the left of the 45 degree line in figure). For this group, Pareto is closer to the empirical distribution but the difference is not large 5.4.

This analysis of the individual English wealth data suggest a log-normal approximation should always be preferred to a Pareto power law. This has important implications, not only other studies which use Pareto shortcuts to calculate unobserved percentile wealth shares but also, more deeply,

<sup>23</sup>Of the 41,696 top point one percent 1892-1992, 1,621 entries were duplicates. Duplication had two potential sources: 1. A rich person could be probated more than once (and have two or more entries in the PPR calendars). 2. The process failed to deduplicate correctly. For the rest of the wealth distribution, problem 1 and 2 will result in some duplication, although the probability of multiple probate is likely highly correlated with the size and complexity of the estate.

<sup>24</sup>Number 1, John Reeves Ellerman was so reclusive that even the newspapers he owned could not find a photograph for his obituary ([www.telegraph.co.uk/news/uknews/1519047/Was-this-the-richest-and-most-secretive-British-tycoon-ever.html](http://www.telegraph.co.uk/news/uknews/1519047/Was-this-the-richest-and-most-secretive-British-tycoon-ever.html)). Note that his original entry is one of the original entries reproduced in figure 3.3.

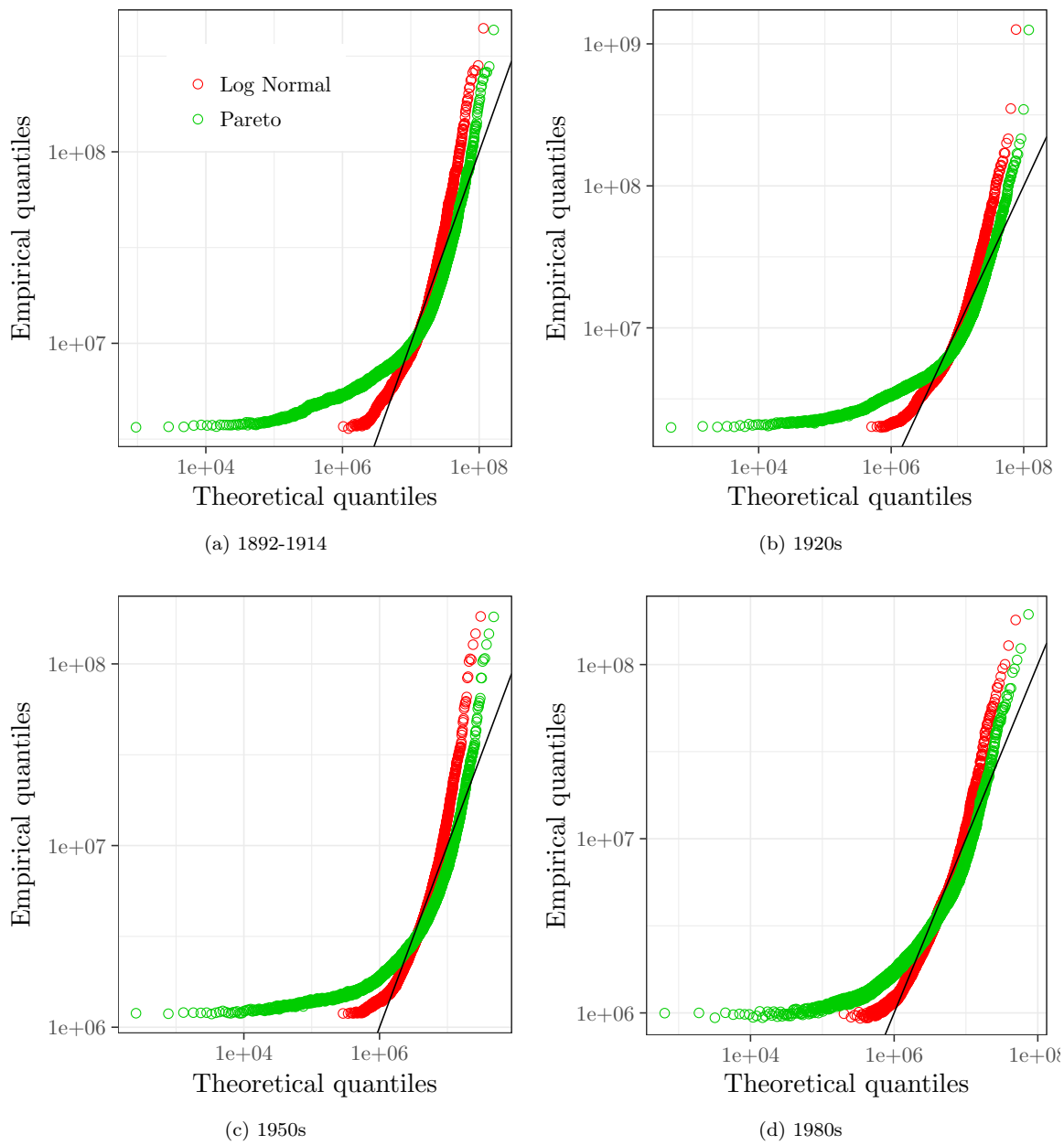


Figure 5.4: Testing Pareto for the Top .1%

Table 5.3: Anderson-Darling Statistic for Fitted Distributions, top .1%, by Decade

Period	Log Normal	Pareto
1892-1914	155.42	680.78
1914-8	31.62	117.03
1918-39	183.75	737.78
1939-45	77.63	280.19
1945-60	129.96	690.35
1960s	200.72	717.96
1970s	112.90	582.41
1980-92	37.99	188.04

for understanding the inequality generating process behind the wealth distribution (see for example the simple theoretical discussion in Jones (2015)). Understanding these dynamics - whether they be  $r$  versus  $g$  based (Piketty (2014)) or driven by some other set of mechanisms is fundamental to the academic and social mission of economics.

## 6 Gender and Wealth

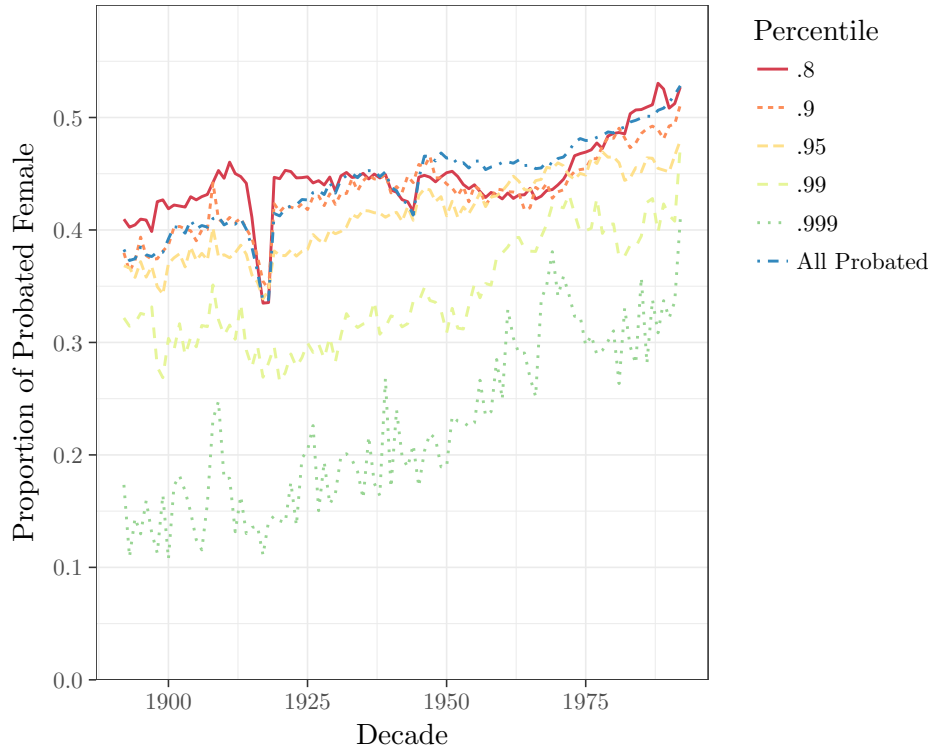
Gender has been a neglected aspect of the recent focus on top income and wealth. W.D Rubinstein omitted women entirely from his study of the very wealthy in Britain from the same source used in this paper (Rubinstein (1981), as noted also by Laurence et al. (2009) p.1). After the Married Women’s Property Acts of 1870 and 1882, men and both single and married women were fully equal before the law with respect to their rights over their wealth.

The PPR calendar contains the full name of all probated decedents. I assigned a gender, based on first name.<sup>25</sup> Figure 6.1 (a) reports the proportion female amongst the probated English and the top wealth percentiles, by year, 1892-1992. The proportion of probates that are female rises from under 40% in 1892 to over 50% by 1992, when females have a higher probate rate than men. After 1970, this series rise accelerates. (Notice also that it drops during the wars, reflecting higher male mortality.) Strikingly, there is a time-consistent, sharply decreasing female proportion, by percentile, through the century 1892-1992. The trend in all top percentile female ratios, over time, is towards equality but even in 1992 women are underrepresented amongst the top 5%. The convergence undoubtedly reflects the rising status of women in the labour force and as wealth holders within English society. The under-representation amongst the very top wealth holding class may reflect specific discrimination against female entrepreneurs, discrimination against women receiving large bequests from their fathers, gender preferences or differences between the sexes at the extreme right tail of innate talent in accumulating wealth.<sup>26</sup>

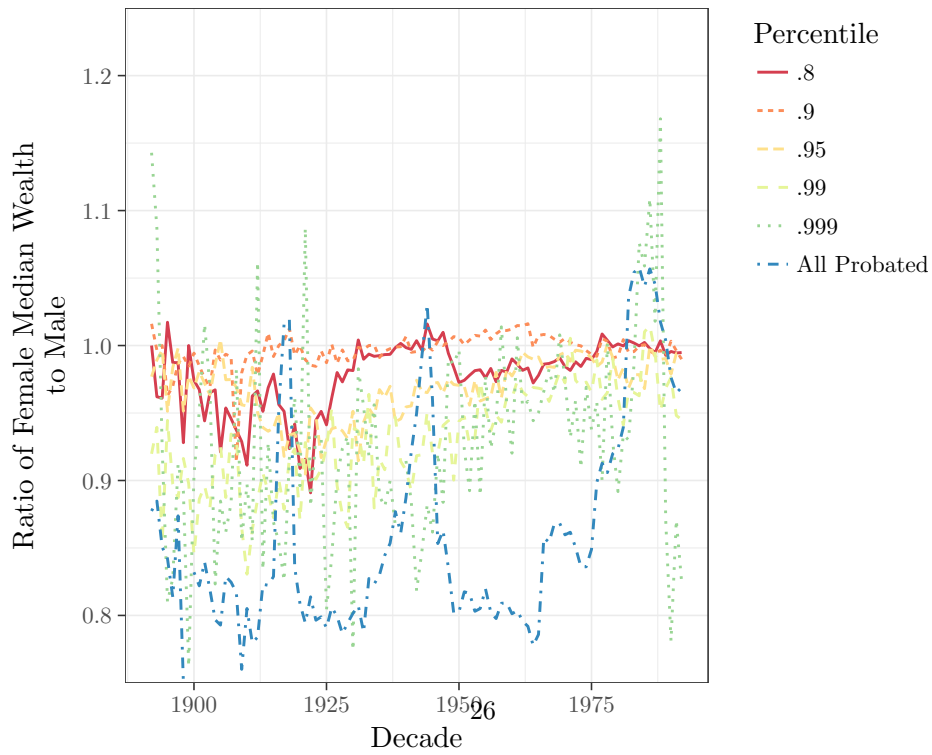
Figure 6.1 (b) reports the ratio of female median wealth to male median wealth, by year, 1892-1992. Here the picture is messier. Amongst the probated population, women are consistently considerably less wealthy than men (the equalization post 1980 may reflect the ‘banding issue,

<sup>25</sup>This was done using the ‘gender’ package (<https://CRAN.R-project.org/package=gender>) in R

<sup>26</sup>This observation is consistent with Green et al. (2009) analysis of the death duty registers, 18701-902. For their sample they report that women represent about 40%, with a mean wealth around half that of men (table 1, p.320). Sine the ratio of mean wealth is about 90%, this implies that the large wealth-holders are increasingly male.



(a) Proportion Female, Probated Population and Top Percentiles



(b) Ratio of Female Wealth to Male

Figure 6.1: Gender and Wealth, England, 1892-1992

discussed in greater detail in the appendix). However amongst the 80-90 percentile there is a greater equality between these sexes than for the top 5%, just as revealed by the proportions probated. The annual variation is bigger but the trend towards greater equality is clear and even stronger than figure 6.1 (a). It would appear that for those women who do make great wealth - their wealth is close to equality with that of men.

## 7 Conclusion

Using novel population-scale data, this paper has freshly characterized the English wealth distribution between 1892 and 2016. A period that captures the ‘Great Equalization’ of wealth. I find that  $r-g$  dynamics failed to create a broadly based ‘middle’ class England. Even in the post war years, the period of the massive drop in the wealth-share of the top 1% from 73% to 20%, two thirds of English decedents lived, worked, then died with nothing to their name.<sup>27</sup>

This result is surprising given the well-known increase in home ownership rates over the 20th century (ONS, 2013). Perhaps mortgage and other debt can explain this divergence. Or maybe the missing middle class from the probate records have just not died yet. A weakness of the wealth data used here is that it can only speak about the wealth of the dead. However, the picture painted by the PPR data will be familiar to many. In terms of wealth accumulation, lifetime spending and investment in human capital; the median English person has little surplus wealth at death.

This paper is an opening salvo for a research agenda examining the determinants of the English wealth distribution over the past century and more. Future research that can fully exploit the rich individual level detail of the PPR entries has great promise. The data contain compelling surname information for example as well as exact street addresses of decedents. A population analysis of the controversial social mobility claims of Clark and Cummins (2015a) could be attempted for example.

Why did the share of the top point one percent decline so dramatically? Every member of the top point-one-percent is listed in the PPR data. Their family history, life choices, demography and luck can now be detailed and tracked precisely. Theoretically, linking the wealth distribution to theories of social mobility that are consistent with the empirical facts, given by the PPR data, is another direction.

The methodology applied here to constructing a new dataset can be applied to any set of images of historical records that contain consistent formatting. There are many millions of these images lying on website servers all over the world. As Optical Character Recognition software continues to become more accurate, there is now remarkable potential for new, big data analysis in economic history.

## References

- Alvaredo, Facundo, 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, A. B., “Top Incomes in the UK over the 20th Century,” *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 2005, 168 (2), 325–343.

---

<sup>27</sup>Separate evidence supporting this comes from Karagiannaki (2015), who estimates from contemporary survey data that the median English inherit £0 during their lifetime (table 4, p.198). (The 43% of inheritors receive a median sum of £9,400.)

- Atkinson, Anthony B**, “Wealth and Inheritance in Britain from 1896 to the Present,” 2013.
- Atkinson, Anthony B., 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.
- Atkinson, Anthony Barnes and Allan James Harrison**, *Distribution of personal wealth in Britain*, Cambridge Univ Pr, 1978.
- Bee, Marco, Massimo Riccaboni, and Stefano Schiavo**, “Pareto Versus Lognormal: a Maximum Entropy Test,” 08 2011, 84, 026104.
- Benhabib, Jess, Alberto Bisin, and Shenghao Zhu**, “The Distribution of Wealth and Fiscal Policy in Economies With Finitely Lived Agents,” *Econometrica*, 2011, 79 (1), 123–157.
- Blanchet, Thomas, Juliette Fournier, and Thomas Piketty**, “Generalized Pareto Curves : Theory and Applications,” Working Papers 201703, World Inequality Lab March 2017.
- Clark, Gregory and Neil Cummins**, “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.
- Cummins, Neil**, “The Decline of the top 1% and the Hidden Wealth of the Elite: English Dynasties, 1892-2018,” 2018.
- Delignette-Muller, M., R. Pouillot, J. Denis, and C. Dutang**, “fitdistrplus: Help to Fit of a Parametric Distribution to Non-Censored or Censored Data,” 2014.
- Feenberg, Daniel R. and James M. Poterba**, “The Income and Tax Share of Very High-Income Households, 1960-1995,” *The American Economic Review*, 2000, 90 (2), 264–270.
- Gabaix, Xavier**, “Zipf’s Law for Cities: An Explanation,” *The Quarterly Journal of Economics*, 1999, 114 (3), 739–767.
- GOV.UK**, “Find a will,” 2018.
- Green, David R., Alastair Owens, Josephine Maltby, and Janette Rutterford**, “Lives in the balance? Gender, age and assets in late-nineteenth-century England and Wales,” *Continuity and Change*, 2009, 24 (2), 307–335.
- Harbury, C. D.**, “Inheritance and the Distribution of Personal Wealth in Britain,” *The Economic Journal*, 1962, 72 (288), 845–868.
- Harbury, CD and DM Hitchens**, *Inheritance and wealth inequality in Britain*, Allen & Unwin (London and Boston), 1979.
- House of Commons**, “Report of the Committee on Income Tax,” 1906.
- Jones, Charles I.**, “Pareto and Piketty: the macroeconomics of top income and wealth inequality,” 2014.

- , “Pareto and Piketty: The Macroeconomics of Top Income and Wealth Inequality,” *The Journal of Economic Perspectives*, 2015, 29 (1), 29–46.
- Karagiannaki, Eleni**, “The magnitude and correlates of inter-vivos transfers in the UK,” 2011, (151).
- , “Recent Trends in the Size and the Distribution of Inherited Wealth in the UK,” *Fiscal Studies*, 2015, 36 (2), 181–213.
- Kuznets, Simon**, *Shares of Upper Income Groups in Income and Savings* 1953.
- Laurence, Anne, Josephine Maltby, and Janette Rutterford**, “Introduction,” in “Women and their Money 1700-1950,” Routledge, 2009, pp. 1–30.
- Lindert, Peter H.**, “Unequal English Wealth since 1670,” *The Journal of Political Economy*, 1986, 94 (6), pp. 1127–1162.
- Marx, Karl**, “Das Kapital: kritik der politischen ökonomie,” *Germany: Verlag von Otto Meisner*, 1867, 1885, 1894.
- 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.
- Piketty, Thomas**, *Les hauts revenus en France au XXème siècle*, Grasset, 2001.
- , “Income Inequality in France, 1901-1998,” *The Journal of Political Economy*, 2003, 111 (5), pp. 1004–1042.
- , *Capital in the Twenty-First Century*, Harvard University Press, 2014.
- **and Emmanuel Saez**, “Income Inequality in the United States, 1913-1998,” *The Quarterly Journal of Economics*, 2003, 118 (1), 1–39.
- Rothenbacher, Franz**, *The European population since 1945*, Palgrave Macmillan, 2005.
- 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.**, “Probate Valuations: A Tool for The Historian,” *The Local Historian*, 1974, 11 (2), 68–72.
- , “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.
- Saez, Emmanuel and Gabriel Zucman**, “Wealth Inequality in the United States since 1913: Evidence from Capitalized Income Tax Data \*,” *The Quarterly Journal of Economics*, 2016, 131 (2), 519–578.



**Schurer, Kevin and Matthew Woollard**, “1881 Census for England and Wales, the Channel Islands and the Isle of Man (Enhanced Version) [computer file],” 2000.

**Stiglitz, J. E.**, “Distribution of Income and Wealth Among Individuals,” *Econometrica*, 1969, 37 (3), 382–397.

**Turner, John D.**, “Wealth concentration in the European periphery: Ireland, 1858-2001,” *Oxford Economic Papers*, 2010, 62 (4), 625–646.

## List of Tables

2.1	The Probate Valuations . . . . .	4
2.2	The Minimum Probate Threshold, 1858-2017 . . . . .	5
3.1	General Entry Patterns . . . . .	8
4.1	Counts of Probates and Adult Deaths, 1892-1992 . . . . .	12
5.1	Anderson-Darling Statistic for Fitted Distributions, All Wealth Holders, by Decade . . . . .	20
5.2	The 10 Richest English, 1892-1992 . . . . .	23
5.3	Anderson-Darling Statistic for Fitted Distributions, top .1%, by Decade . . . . .	25
A.1	Proportion of All Decedents and those Probated with Banded Probate Valuations . . . . .	32
B.1	Wealth Share Estimates, 1892-1992 . . . . .	34

## List of Figures

3.1	The Scanned Images . . . . .	6
3.2	Examples of Entries, 1880s-1990s . . . . .	7
3.3	Examples of Problematic and unusually rich Entries . . . . .	9
3.4	The Data-Building Process . . . . .	10
4.1	Nominal Probated and Non Probated Wealth in England and Wales, 1892-1992 . . . . .	12
4.2	Comparing Different Estimates of Top Wealth Shares, England 1892-1992 . . . . .	13
4.3	Overall Wealth Inequality, England 1892-2016 . . . . .	14
4.4	Top Wealth Shares, England 1892-1992 . . . . .	16
4.5	English Wealth Holding by Decile, 1892-1992 . . . . .	17
5.1	Lorenz Curves for Real Wealth, England, 1892-1992 . . . . .	19
5.2	Empirical v Fitted Wealth Distributions, by Period, England, 1 of 2 . . . . .	21
5.3	Empirical v Fitted Wealth Distributions, by Period, England, 2 of 2 . . . . .	22
5.4	Testing Pareto for the Top .1% . . . . .	24
6.1	Gender and Wealth, England, 1892-1992 . . . . .	26
A.1	Inheritance Taxes in England . . . . .	31
C.1	The Top .1% in International Context . . . . .	35
D.1	Data Collection Verification, Official Count Comparison . . . . .	36
E.1	Data Collection Verification, Official Count Comparison, Deaths 1973-2007 . . . . .	38

## A Extra

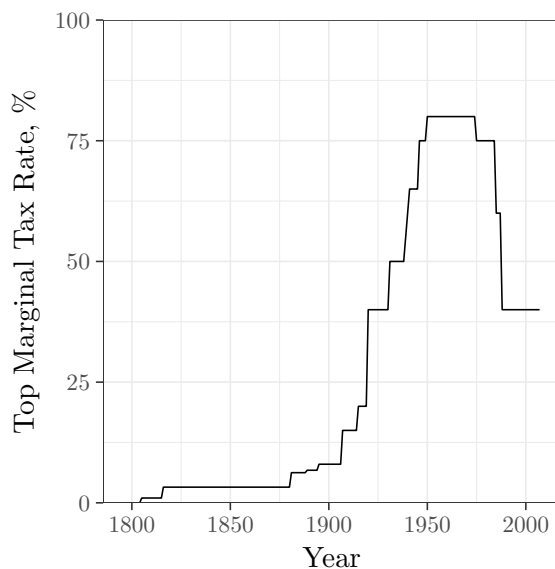


Figure A.1: Inheritance Taxes in England

Source: Cummins (2018) (Working Paper forthcoming)

### A.1 Banded Probate Valuations After 1980

After 1980 there was a change in the system for valuing probates in the calendars. As opposed to an exact valuation, which was the practice 1892-1979, a proportion of valuations appear as bands. Namely, these are £25,000, £40,000, £70,000, £100,000, £115,000 and £125,000 with each entry listed as “Not Exceeding” the named amount. Table A.1 reports the overall incidence of the banded values, 1892-1992. It is clearly evident that these bands are loosely applied. The proportion of probated values that were entered as bands was about 60% in the 1980s and inspection of the data revealed that using these numbers for wealth distribution analysis was pointless. For example, calculated mini coefficients using the band values suggest a sudden and dramatic large drop in economic inequality in 1980. Outside of war, revolution and natural disagree we would not expect such huge drops in inequality on a year to year basis. Further, the values of wealth necessary to enter the various top percentiles change dramatically after 1980. Occam’s Razor suggest that such a sudden change in the years immediately following a new valuation system is probably a direct result of that valuation system. Attributing a value based upon the average wealth of observed specific values also led to implausible drops in measured wealth inequality during the ‘banded’ years. It was also clear by examining the entries by eye did not transmit much information about the true wealth of the decedents. Therefore all banded values were dropped entirely and the post 1980 analysis release on the distribution of specific probated values only. (The ‘excluded’ population post 1980 had to adjusted by the proportion of those probated who have a banded value).

The assumption here is that the specific values are an unbiased measure of the wealth distribution. This is probably incorrect. It appears that using specific values instead of the bands is likely to oversample the rich -who can pay to have their decades family members estate professionally valued. Caution must be exercised before interpreting the the post 1980 trends. The direction

Years	All	Probated
1892-1979	0	0
1981	.13	.30
1982	.25	.55
1983	.27	.59
1984	.28	.63
1985	.23	.59
1986	.22	.56
1987	.23	.6
1988	.24	.63
1989	.25	.67
1990	.29	.77
1991	.30	.80
1992	.25	.84

Table A.1: Proportion of All Decedents and those Probated with Banded Probate Valuations

of the bias however can be gauged by considering who is been selectively purged from the data by dropping the wealth banded valuations. The rich are more represented as noted before - the poor are still represented but we are taking people out of the upper middle of the distribution - which other things being equal is likely to bias the inequality estimates upward. As the finding is of flat/declining inequality post 1980, I claim that the expected direction of the bias does not contradict this analysis' post 1980 results.

## B Wealth Share Estimates, 1892-1992, Data Table

yr	.70	.80	.90	.95	.99	.999
1892		0.014	0.043	0.215	0.386	0.341
1893		0.013	0.041	0.202	0.364	0.380
1894		0.012	0.044	0.215	0.387	0.343
1895		0.010	0.039	0.204	0.385	0.363
1896		0.012	0.041	0.200	0.370	0.378
1897		0.013	0.043	0.206	0.372	0.366
1898		0.016	0.052	0.228	0.375	0.329
1899		0.015	0.050	0.224	0.373	0.338
1900		0.026	0.065	0.250	0.355	0.305
1901	0.003	0.039	0.084	0.272	0.355	0.248
1902	0.018	0.033	0.077	0.266	0.357	0.250
1903	0.016	0.031	0.072	0.254	0.332	0.294
1904	0.016	0.041	0.088	0.282	0.349	0.224
1905	0.016	0.037	0.079	0.267	0.340	0.261
1906		0.045	0.067	0.256	0.349	0.283
1907		0.044	0.065	0.253	0.356	0.281

Continued on next page

yr	.70	.80	.90	.95	.99	.999
1908			0.062	0.184	0.375	0.379
1909		0.038	0.057	0.248	0.376	0.281
1910		0.038	0.051	0.215	0.350	0.347
1911		0.035	0.044	0.194	0.325	0.402
1912		0.038	0.050	0.218	0.358	0.336
1913		0.040	0.052	0.223	0.357	0.328
1914		0.041	0.053	0.220	0.361	0.325
1915		0.045	0.057	0.227	0.345	0.327
1916		0.054	0.065	0.245	0.343	0.293
1917		0.056	0.063	0.229	0.323	0.330
1918		0.056	0.061	0.226	0.338	0.319
1919		0.048	0.062	0.231	0.331	0.328
1920		0.050	0.067	0.244	0.339	0.300
1921		0.052	0.068	0.234	0.319	0.327
1922		0.058	0.075	0.251	0.330	0.287
1923	0.019	0.043	0.077	0.252	0.328	0.280
1924	0.020	0.046	0.080	0.256	0.321	0.278
1925	0.020	0.047	0.082	0.260	0.327	0.265
1926	0.019	0.050	0.084	0.262	0.327	0.258
1927	0.020	0.051	0.084	0.257	0.317	0.271
1928	0.019	0.049	0.080	0.246	0.299	0.307
1929	0.020	0.051	0.081	0.247	0.315	0.286
1930	0.023	0.059	0.092	0.273	0.317	0.235
1931	0.026	0.064	0.094	0.274	0.318	0.223
1932	0.043	0.071	0.103	0.286	0.308	0.189
1933	0.036	0.060	0.088	0.251	0.284	0.280
1934	0.039	0.066	0.095	0.273	0.302	0.226
1935	0.038	0.065	0.094	0.264	0.294	0.245
1936	0.040	0.068	0.096	0.270	0.308	0.218
1937	0.041	0.069	0.096	0.266	0.302	0.227
1938	0.023	0.072	0.097	0.264	0.314	0.210
1939	0.023	0.075	0.099	0.262	0.306	0.213
1940	0.028	0.083	0.107	0.274	0.277	0.205
1941	0.030	0.086	0.109	0.274	0.281	0.195
1942	0.035	0.092	0.115	0.280	0.273	0.177
1943	0.035	0.092	0.114	0.280	0.270	0.179
1944	0.036	0.090	0.111	0.274	0.264	0.194
1945	0.039	0.096	0.118	0.284	0.263	0.170
1946	0.039	0.098	0.119	0.280	0.265	0.172
1947	0.040	0.101	0.121	0.282	0.261	0.168
1948	0.044	0.109	0.127	0.288	0.257	0.147
1949	0.045	0.116	0.131	0.285	0.247	0.148
1950	0.050	0.125	0.137	0.287	0.238	0.131
1951	0.051	0.126	0.138	0.285	0.235	0.133

Continued on next page

yr	.70	.80	.90	.95	.99	.999
1952	0.053	0.132	0.142	0.287	0.228	0.125
1953	0.054	0.132	0.140	0.287	0.226	0.129
1954	0.055	0.131	0.139	0.290	0.232	0.118
1955	0.057	0.134	0.139	0.289	0.229	0.117
1956	0.059	0.137	0.140	0.285	0.225	0.116
1957	0.058	0.136	0.136	0.278	0.219	0.138
1958	0.057	0.138	0.141	0.285	0.224	0.123
1959	0.054	0.134	0.137	0.286	0.234	0.124
1960	0.064	0.135	0.137	0.281	0.223	0.115
1961	0.054	0.131	0.136	0.284	0.238	0.128
1962	0.059	0.139	0.140	0.288	0.229	0.110
1963	0.060	0.139	0.137	0.283	0.224	0.120
1964	0.058	0.143	0.143	0.288	0.226	0.111
1965	0.048	0.138	0.144	0.286	0.221	0.105
1966	0.070	0.152	0.142	0.272	0.198	0.100
1967	0.070	0.154	0.143	0.273	0.200	0.097
1968	0.067	0.148	0.139	0.273	0.212	0.100
1969	0.072	0.156	0.145	0.271	0.203	0.091
1970	0.073	0.168	0.154	0.270	0.191	0.086
1971	0.071	0.168	0.155	0.271	0.195	0.086
1972	0.068	0.171	0.160	0.269	0.190	0.094
1973	0.075	0.189	0.165	0.265	0.175	0.083
1974	0.084	0.200	0.167	0.249	0.150	0.095
1975	0.086	0.198	0.164	0.249	0.150	0.067
1976	0.088	0.199	0.164	0.252	0.153	0.067
1977	0.092	0.197	0.163	0.251	0.151	0.071
1978	0.087	0.196	0.168	0.254	0.151	0.076
1979	0.085	0.202	0.171	0.256	0.150	0.073
1980	0.086	0.208	0.177	0.255	0.149	0.068
1981	0.100	0.221	0.171	0.244	0.143	0.072
1982	0.147	0.209	0.155	0.221	0.128	0.068
1983	0.138	0.209	0.157	0.226	0.129	0.073
1984	0.138	0.217	0.162	0.235	0.135	0.072
1985	0.135	0.218	0.163	0.236	0.137	0.070
1986	0.143	0.214	0.161	0.234	0.135	0.058
1987	0.130	0.216	0.163	0.239	0.140	0.070
1988	0.133	0.223	0.164	0.230	0.132	0.093
1989	0.130	0.216	0.160	0.229	0.135	0.104
1990	0.140	0.217	0.160	0.227	0.135	0.088
1991	0.141	0.226	0.168	0.235	0.133	0.067
1992	0.099	0.249	0.182	0.254	0.142	0.074

Table B.1: Wealth Share Estimates, 1892-1992

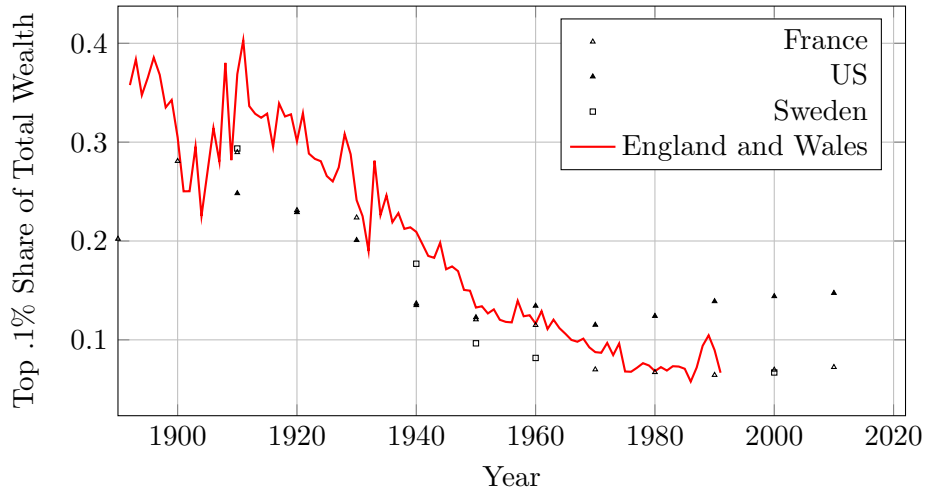


Figure C.1: The Top .1% in International Context

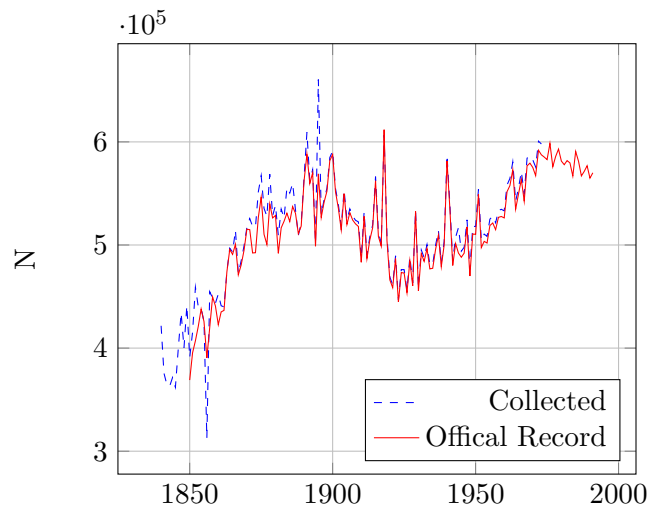
## C The Top .1% in International Comparison

Figure C.1 compares the top .1% wealth share in England and Wales (from figure 4.4) with estimates for France, Sweden and the US. Pre World War II, the top 1% in England and Wales held a greater share of wealth than anywhere we have empirical evidence for. Since 1920, this share has been in consistent decline. After 1960, England and Wales exhibits the same trend and levels of top .1% shares as France and Sweden which are less than half those of the US share.

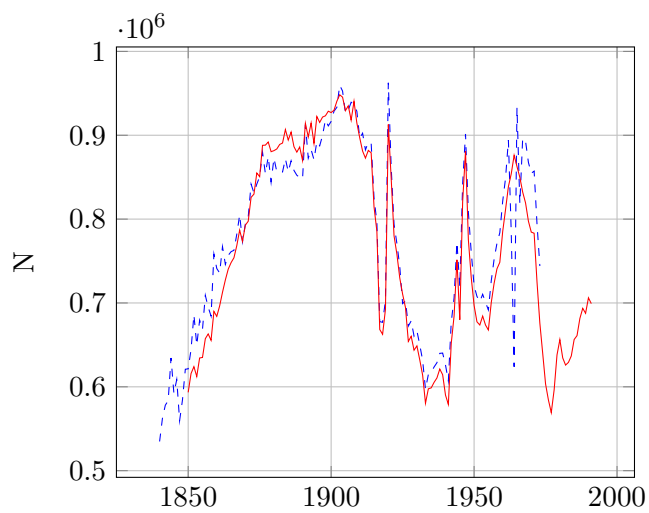
## D Complete Birth, Marriage and Death Registers for England and Wales, 1858-1973

The probate registry records those who die above a certain threshold level of wealth. Those who die below the threshold are invisible. Therefore I downloaded the individual records of all deaths in England and Wales from freebmd.com and familysearch.

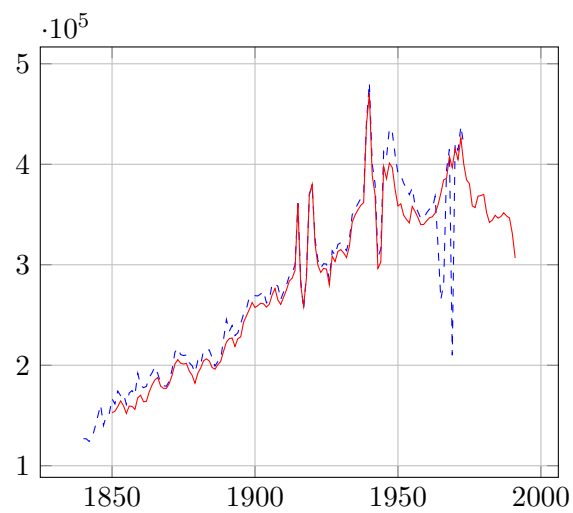
A web-scraper was set up to automate the download of all vital records from freebmd.com which had a complete index from 1838 to 1973 (reference). Reported in maximum batches of 3,000 results the scraper looked up names by first string of first name and surname, quarter and year of event. All vital records reported name and surname, quarter, year and registration district. Birth records also reported the surname of the mother (after 1912), marriage records reported the surname of the spouse (after 1912) and death records recorded the age at death of the deceased (after 1866). Figure D.1 reports by year separately for each vital series a comparison of the numbers collected by this procedure versus that recorded by the official records (the official counts were taken from (I calculated the counts by multiplying the crude death rates by the population)). The collected marriage counts were divided by 2 because there were 2 index entries for every 1 marriage.



(a) Deaths



(b) Births



(c) Marriages

Figure D.1: Data Collection Verification, Official Count Comparison

Notes: The source for the Official Counts is Rothenbacher (2005)

## **E 100% Death Registers, 1973-2007**

The death registers 1973-2007 were scraped automatically from famliysearch.org "England and Wales Death Registration Index 1837-2007." Database. FamilySearch. <http://FamilySearch.org>: 14 June 2016. From "England & Wales Births, 1837-2006." Database. findmypast. <http://www.findmypast.com>: 2012. Citing General Register Office, Southport, England. Figure E.1 reports a comparison of the deaths series form 1838-2007 with the official count from the ONS.



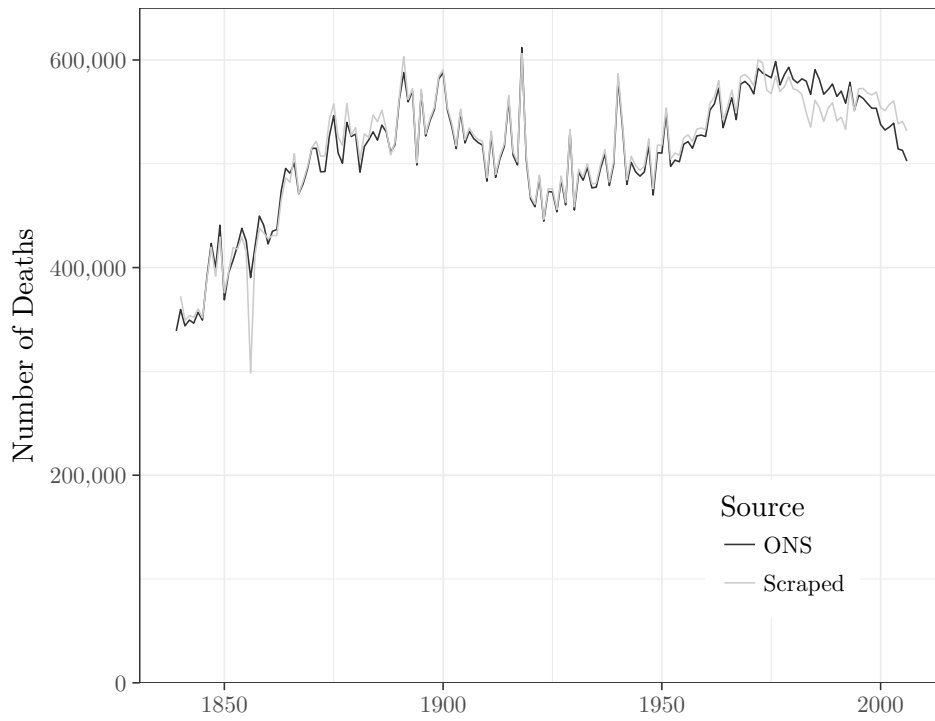


Figure E.1: Data Collection Verification, Official Count Comparison, Deaths 1973-2007

Notes: The source for the Official Counts is <https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths>