Inequality and Social Mobility: Long Run Evidence

Nineteenth Winter School on Inequality, Alba di Canazei

Neil Cummins

Dept. of Economic History

THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

- The *perceived* time trend of inequality over the Industrial Revolution has motivated the most severe critiques of capitalism over the past centuries
  - e.g. Marx (1867); Engels (1884); Piketty (2014)
- Yet there are major empirical gaps, where are our consistently composed data covering the long run?
  - And we know less about *social mobility*

The Plan; A Primer on the Long Run

1. **Inequality**
   - *Long Run*: Evidence from Wills
   - 19th Century on $\rightarrowtail$ Probate Data
     - 1.1 Middle Class
     - 1.2 Hidden Wealth
     - 1.3 Ethnicity
2. **Social Mobility**
   - *Long Run*: Surnames
     - Latent Factor Model
   - 19th Century on $\rightarrowtail$ Marriage Records

Outline

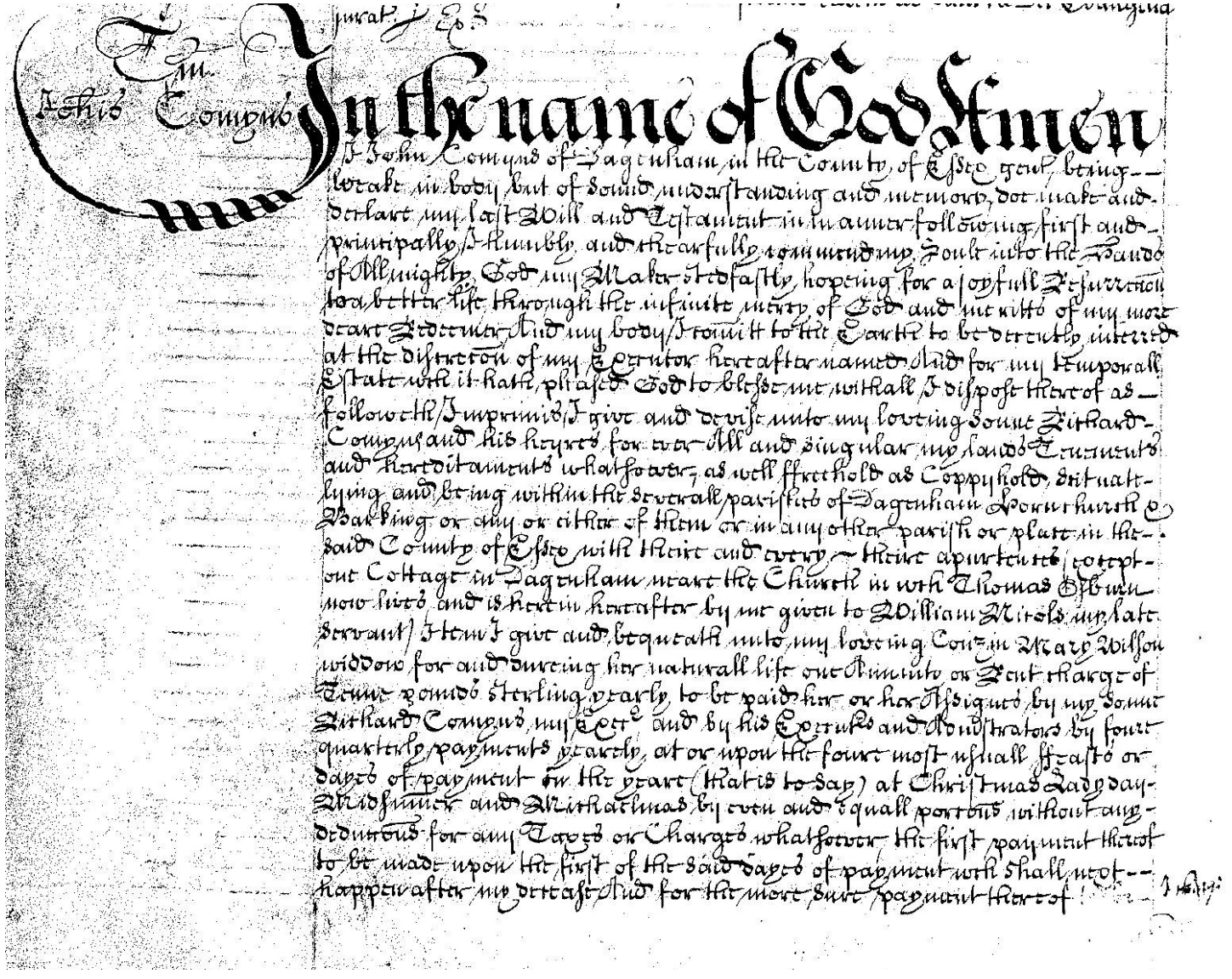

In the name of God Amen

Figure: The Will of John Comyns, Gentleman of Dagenham, Essex, 2 Nov 1688

The Last Will and Testament

- For the past 1,000 years, "last Will and Testament" guided the transmission of wealth at death in England
  - These typically listed descriptively in about 1,000 of their *own words* an individual's assets and inheritors
- These wills represent our best record of individual lives and the economic, family, social and religious influences which mattered most to them as they contemplated their death

Outline

Using Wills to Infer Wealth Inequality

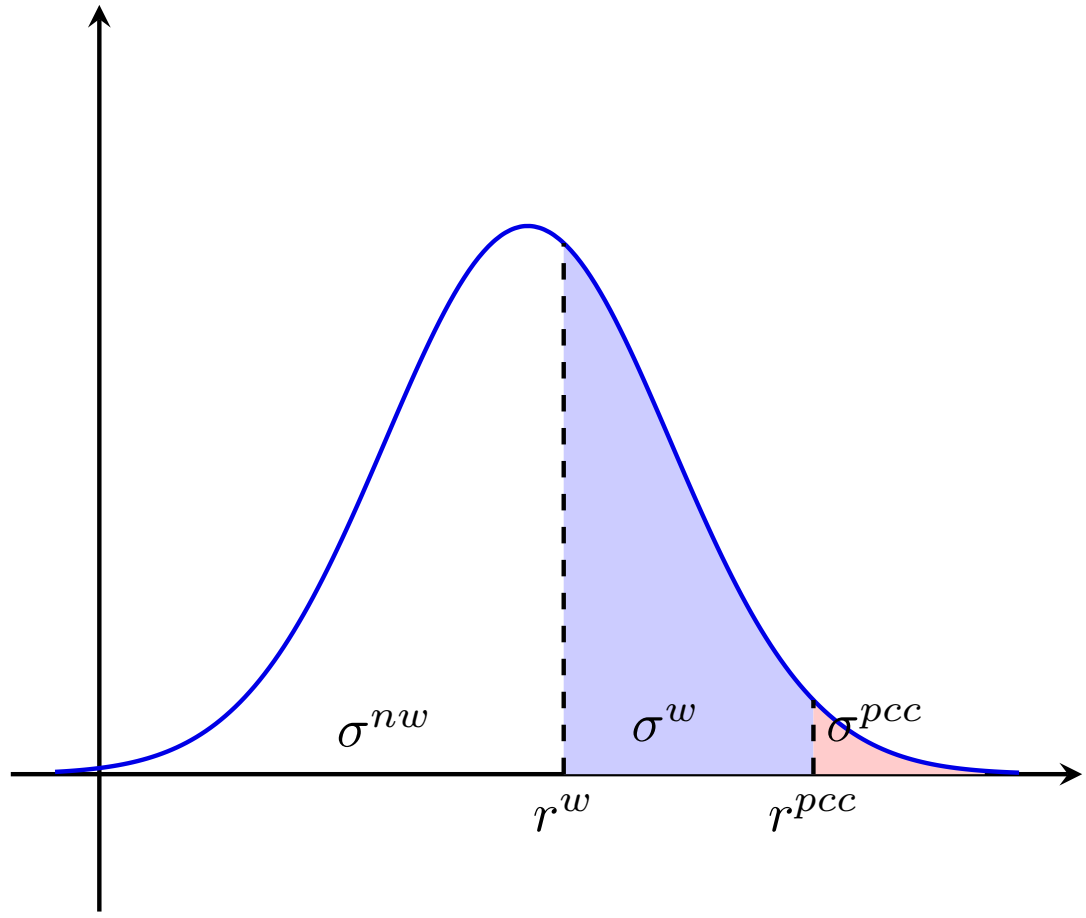
- There are Many Problems
  - Not everyone made wills
    - *And we do not know the scale of this*
  - People don't know when they are going to die
  - Inter-vivos bequests
  - Not All Property is included
    - e.g. Land
  - Assets are often only vaguely described
  - The Full Texts are hard to find at scale
  - Each will needs to be read and coded by a skilled researcher

Using Wills to Infer Wealth Inequality

- **But these selection and detail issues are common to all studies of wealth, even contemporary**
  - So there seems little reason to *not* do this
  - And the more you examine it the more these intractable problems melt away, e.g. Land
  - With or without a will, land passed by primogeniture. Under common law, land could not be devised by will.
    - **But the benefits from a piece land could be bequeathed, and this is frequently done.**

Objections

- There is hostility towards the use of wills for characterizing the past
  - **Representativeness**
- But we can model that, and how to do so is relatively clear for wealth...


σ^{nw}
σ^{w}
σ^{pcc}
r^{w}
r^{pcc}

What We (Will) Do

- Rates

1. r^w: The Population Will Making Rate 1300-1858 [Observed]
2. r^{pcc}: The Population PC Will Making Rate 1300-1858 [Observed]

- Shares

1. σ^{nw}: The share of wealth held by non-will makers [Inferred]
2. σ^w: The share of wealth held by local will makers [Observed]
3. σ^{pcc}: *The share of wealth held by PC will makers* [Observable]

Preview of Results

1. r^w: The Population Will Making Rate rises 1300-1600, then **declines significantly 1600-1858**
    - Shifts in inequality over the early Industrial Revolution era immiserated the majority of the population such that assets at death, for 90% or more, were not significant enough to merit a written will.
2. r^{pcc}: The Population PCC Will Making Rates are either **rising or stable** 1600-1800
    - If the decline of population will making is a cultural turn unreflective of wealth holding, then why does the higher court rate also not decline and can be seen to actually rise
3. The Will Making Rate Spikes Suddenly over the Horizon of the Black Death

- The evidence from Wills suggests that over the period of the Industrial Revolution, Inequality Explodes

Outline

TABLE 6 Wealth inequality in England, 1327–32 and 1524–5 (Gini indexes and relevant percentiles)

Year	Gini	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	Top 5%	Top 1%
1327–32 (observed)	0.423	2.8	3.7	4.3	5.5	6.1	7.9	9.0	11.5	15.4	34.1	22.8	8.3
1524–5 (observed)	0.657	1.3	1.9	2.6	2.6	3.2	4.2	5.6	7.8	12.8	58.0	49.7	29.8
1327–32 (reconstructed)*	0.725 (0.723–0.726)	0.2	0.6	1.1	1.7	2.5	3.7	5.6	8.7	15.6	60.3	45.8	21.5
1327–32 (reconstructed, alternative estimate)*	0.753 (0.752–0.755)	0.2	0.5	0.9	1.4	2.1	3.2	5.0	8.0	14.8	63.9	49.6	24.4
1524–5 (reconstructed)*	0.756 (0.754–0.758)	0.2	0.5	0.9	1.4	2.1	3.2	4.9	8.0	14.8	64.1	49.8	24.5

Notes: *95% confidence intervals between parentheses; see notes to table 3 for additional details.

Figure: 1280-1525

Source: (Alfani and García Montero, 2022, p.1324), medieval lay and Tudor subsidies.

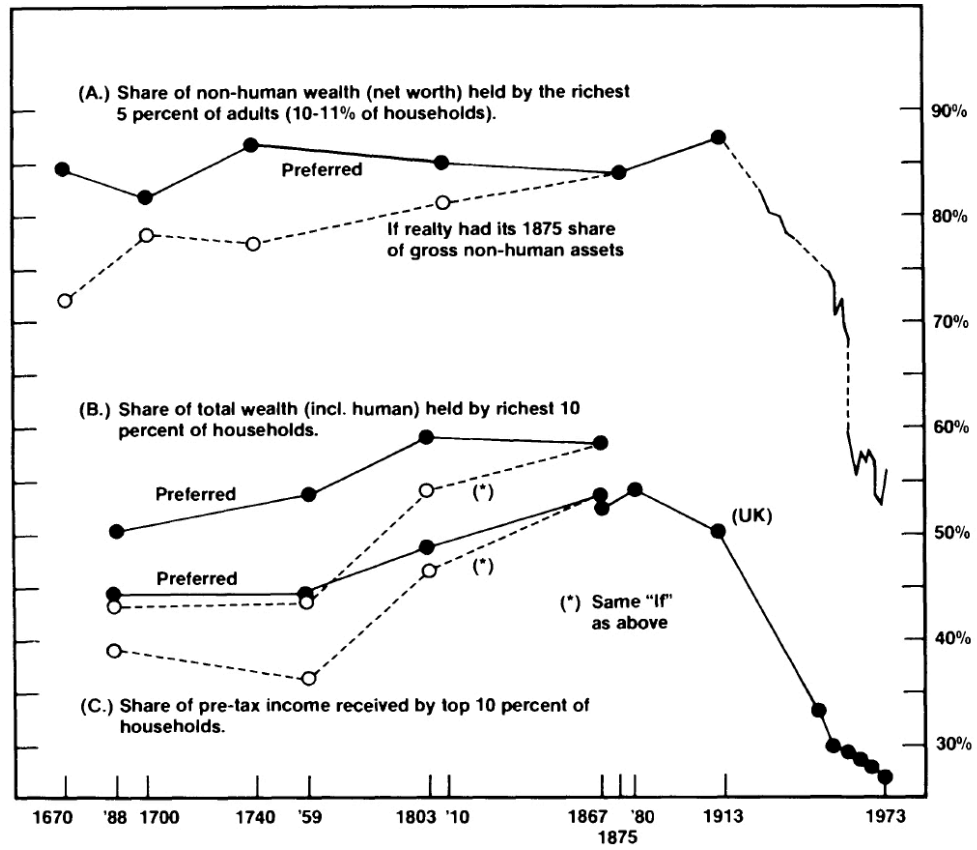


FIG. 2.—Shares of nonhuman wealth, total wealth, and income received by the top 10 percent of households, both sexes, England and Wales, 1670–1973. See notes to table 5.

Figure: 1670-1973

Source: (Lindert, 1986, p.1146), probated wealth and land ownership.

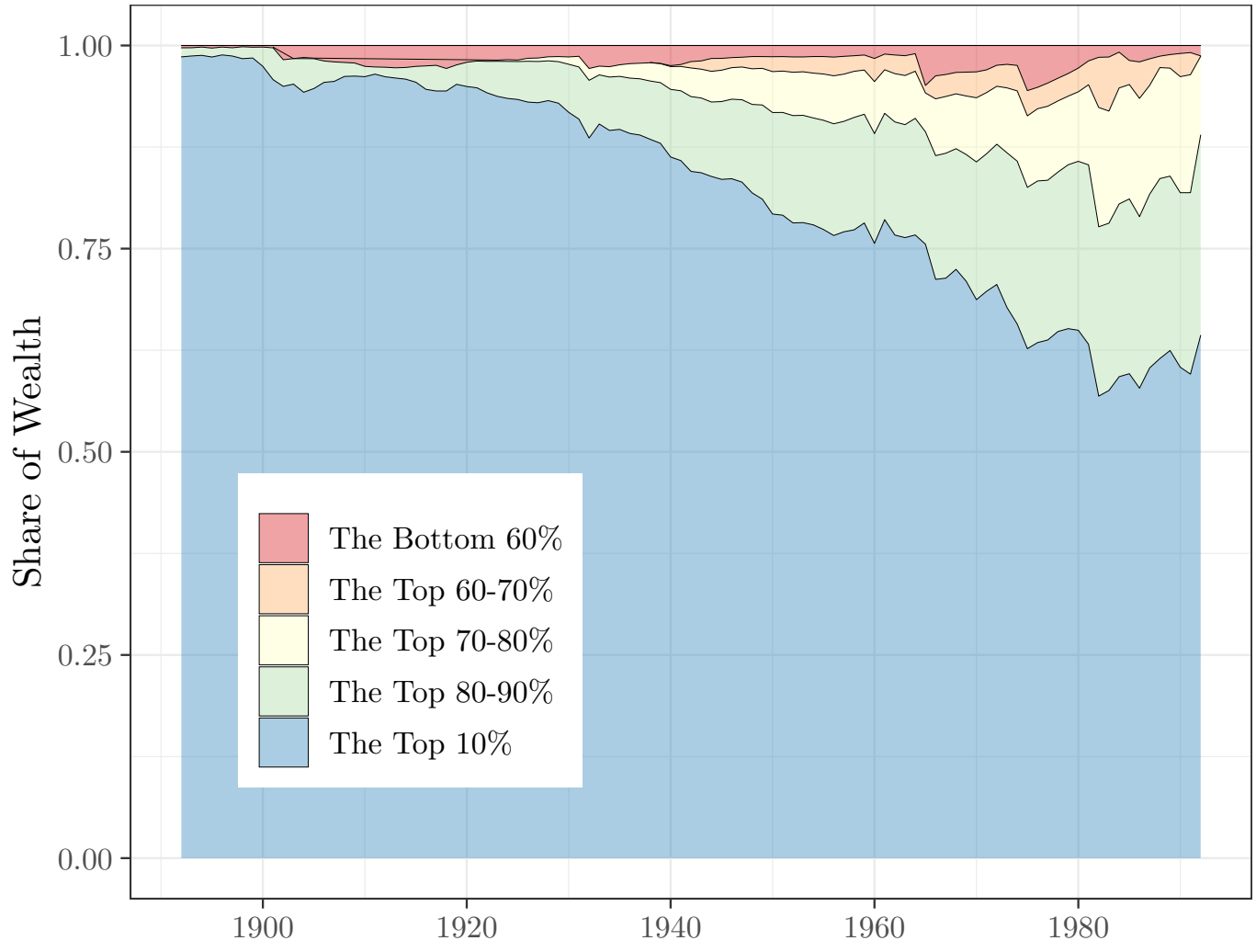


Figure: 1892-1992

Source: Cummins (2021), probated wealth. By 1892, Wealth in England displays spectacular levels of inequality.

Outline

The Wealth of those who Made Wills

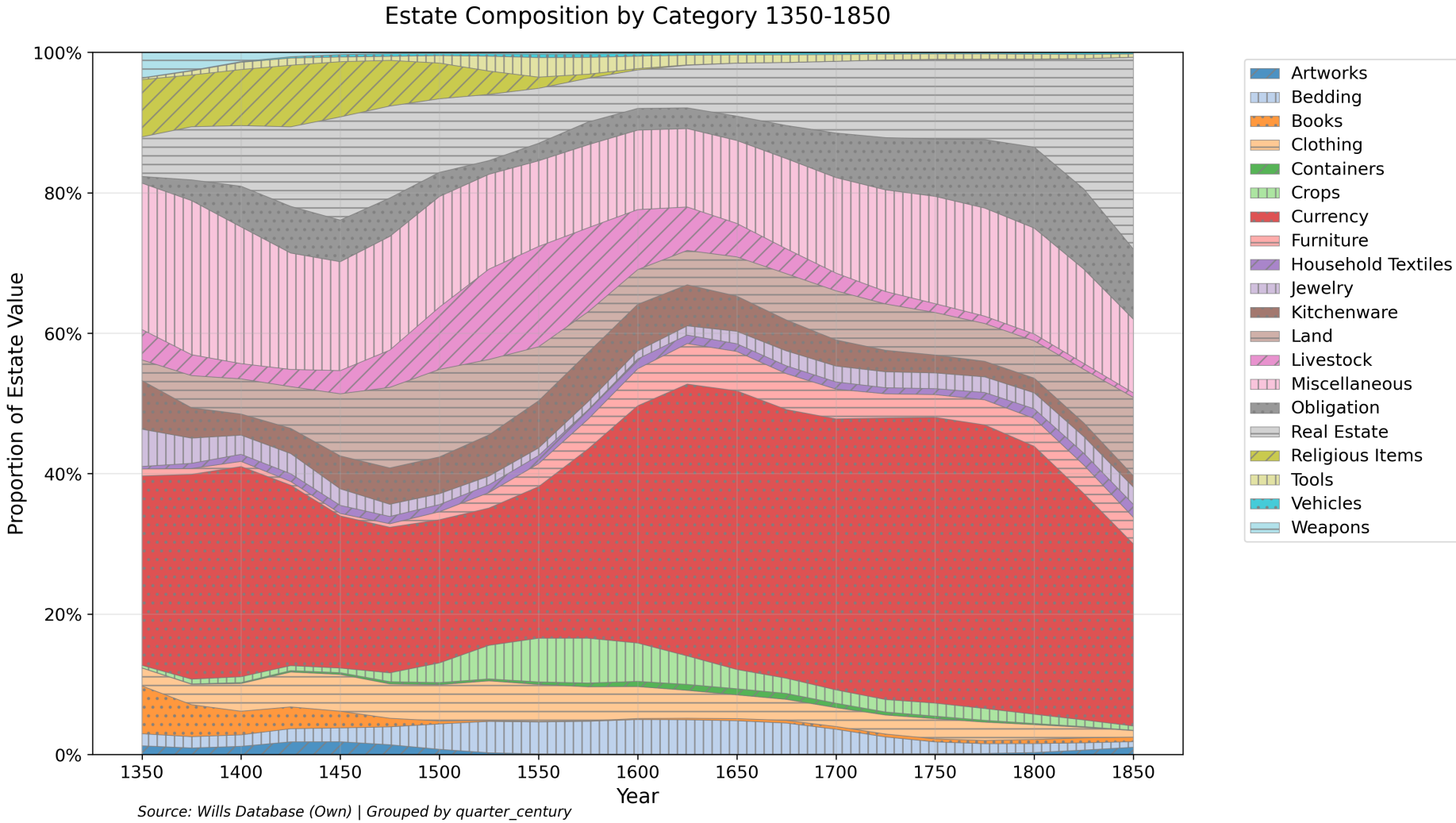
- We estimate wealth for 23,910 wills
- Each will is read, itemized, and coded, using Gemini Flash-2.0 large language model
  1. Biographical Variables
     - Name, aristocratic titles, professional titles, occupation, marital status, year of will-making, and county of residence
  2. Bequest Variables
     - One of 313 item categories commonly found in wills (which are then coded into 19 super-categories); the county of the item (if mentioned); the name of the recipient; the relation of the recipient (in one of 7 categories); the **quantity** of items; the unit the quantity is given in; the **monetary value** of the item (in pounds, shillings and pence); whether the bequest involves a recurring provision; whether the item is in London; whether it is in an urban area (including London); and whether the bequest is a summarization of items included elsewhere in the will.

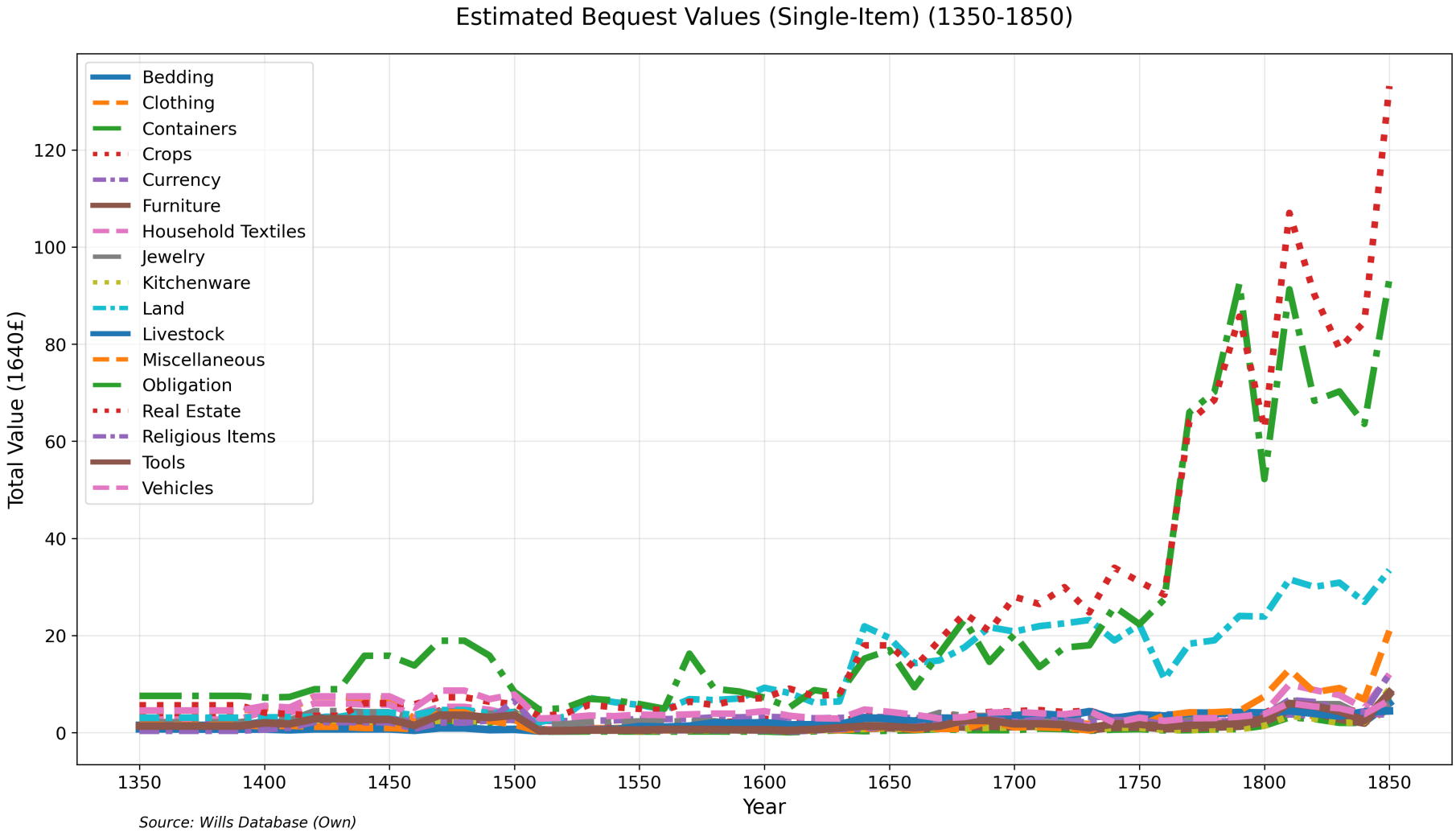
Systematic Evaluation Against a Human annotated Ground Truth Dataset

- We calculate metrics under strict matching criteria
  - The evaluation shows that automated extraction is highly accurate across all dimensions.

Prices

- Most wills (98.8%) contain information on individual bequests, in total we record 317,309 bequests for the 23,910 wills in the current sample, an average of 13.3 bequests per will. However, only 36,967 of those bequests give values for the bequeathed items (11.7%)
- We use values for these items (in combination with observations for land and real estate values from Clark (2005) to estimate the value of each bequest where a monetary value is not specified
- To estimate the values for the 95% of bequests with no stated monetary value, we use a gradient-boosted decision tree algorithm rather than simple averages due to the substantial heterogeneity in our data. Bequest values follow highly skewed distributions and vary dramatically across categories, locations, and time periods
- **All Estate Values are Expressed in £1640s**


Estate Composition by Category 1350-1850
Proportion of Estate Value
100%
80%
60%
40%
20%
0%
1350
1400
1450
1500
1550
1600
1650
1700
1750
1800
1850
Year
Source: Wills Database (Own) | Grouped by quarter_century
Artworks
Bedding
Books
Clothing
Containers
Crops
Currency
Furniture
Household Textiles
Jewelry
Kitchenware
Land
Livestock
Miscellaneous
Obligation
Real Estate
Religious Items
Tools
Vehicles
Weapons


Estimated Bequest Values (Single-Item) (1350-1850)
Bedding
Clothing
Containers
Crops
Currency
Furniture
Household Textiles
Jewelry
Kitchenware
Land
Livestock
Miscellaneous
Obligation
Real Estate
Religious Items
Tools
Vehicles
Total Value (1640£)
0
20
40
60
80
100
120
1350
1400
1450
1500
1550
1600
1650
1700
1750
1800
1850
Year
Source: Wills Database (Own)

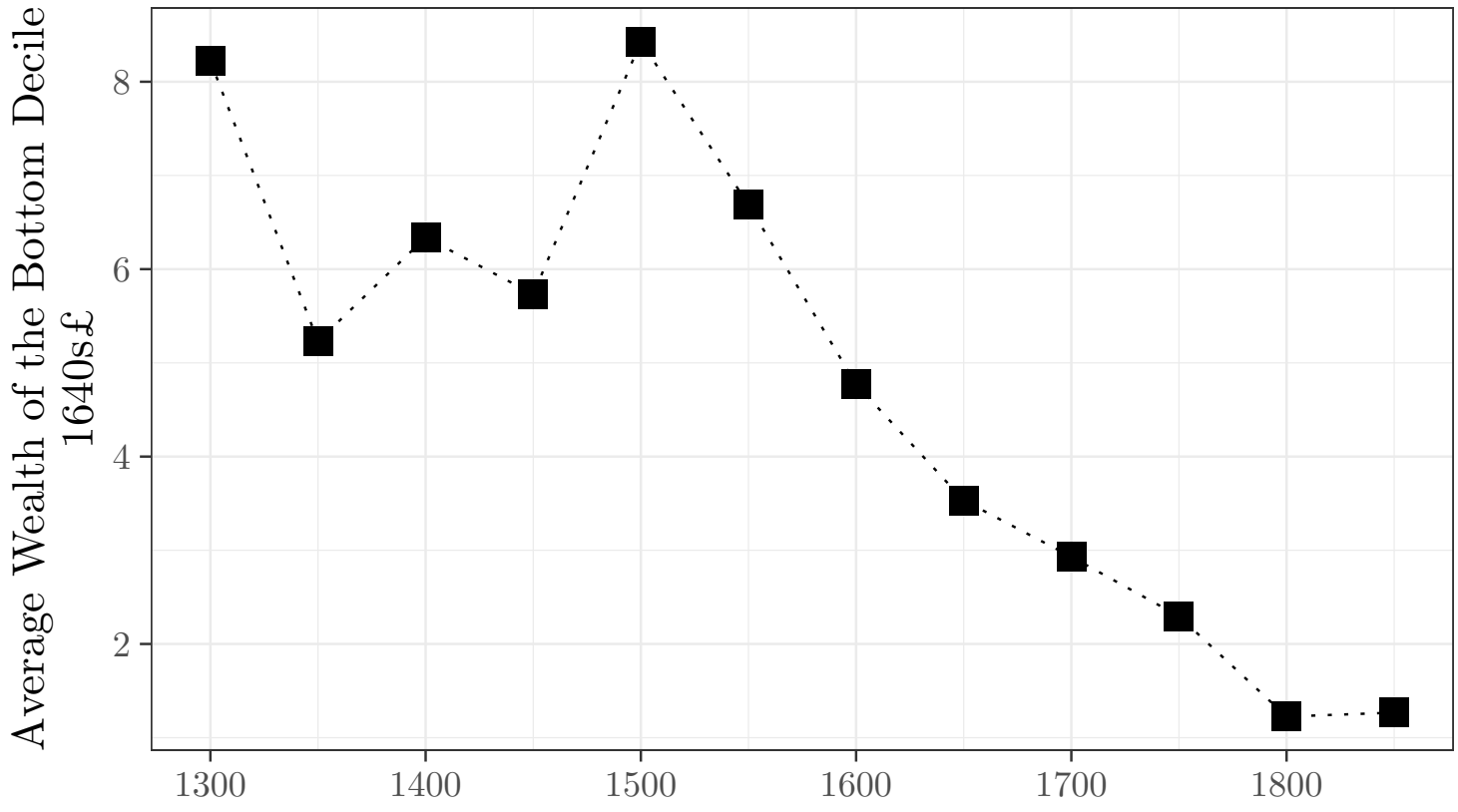


Figure: The Average Wealth of the Bottom Decile of Will Makers, Making Rate 1300–1850

Note: We use this wealth trend to correct the will making series before ...

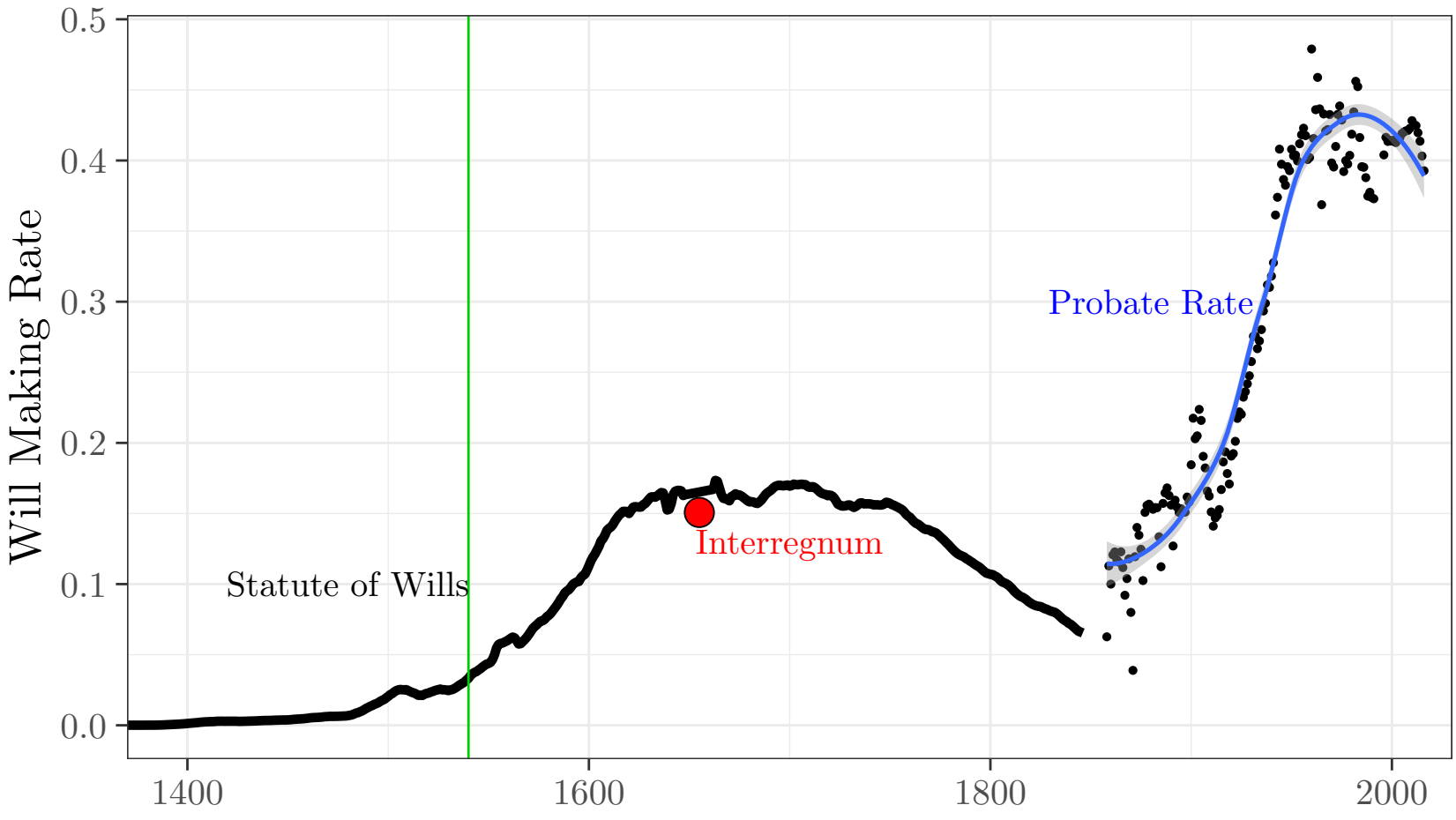
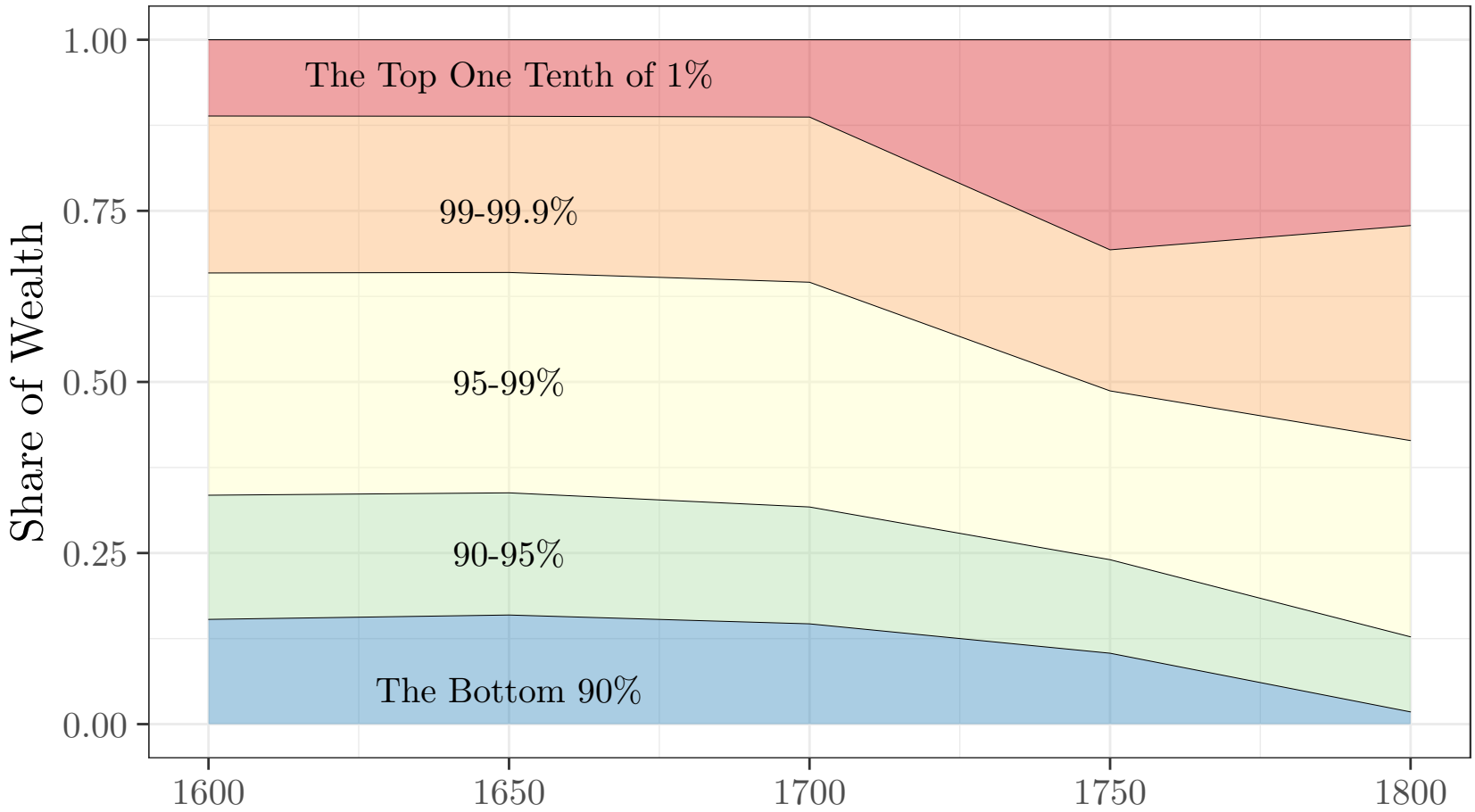


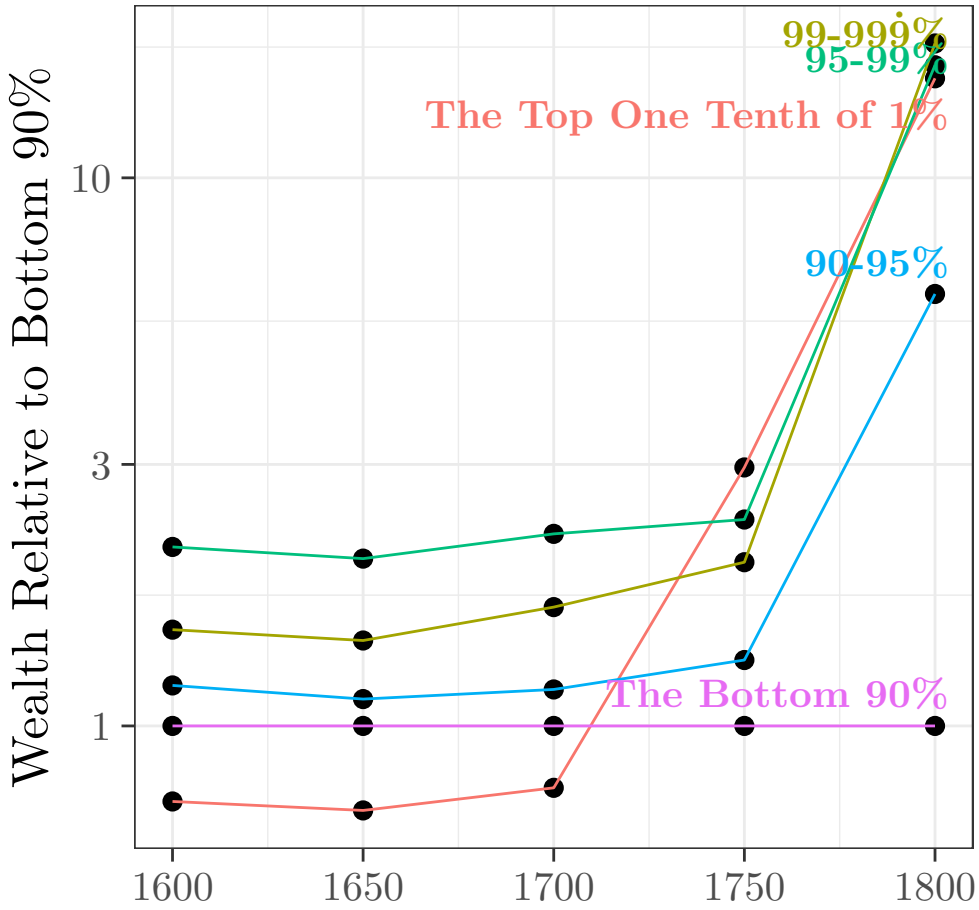
Figure: Predicted Will Making Rates, National

MtoM Data



(a) Stacked Area

Figure: Population Wealth Percentile Shares, 1600-1800



(a) Ratio to Bottom 90%

Figure: Population Wealth Percentile Shares, 1600-1800

Notes: Estimated for the male population only. The extent of the rise in inequality is likely underestimated due to the sampling of

Outline

Table: Wealth Share Comparison with Lindert 1986

			Lindert 1986 Estimate		
Period	Percentile	Wealth Share	lower	preferred	higher
1650	The Top 10%	84.1	61.3	82.7	97.7
1650	The Top 5%	66.2	45.8	73.4	89.5
1650	The Top 1%	34.0	22.1	48.9	57.4
1700	The Top 10%	85.3	66.4	80.8	97.7
1700	The Top 5%	68.3	50.5	71.4	90.1
1700	The Top 1%	35.4	24.5	39.3	71.2
1750	The Top 10%	89.6	70.3	86.0	99.0
1750	The Top 5%	76.0	54.2	73.6	83.1
1750	The Top 1%	51.3	29.3	43.6	71.8
1800	The Top 10%	98.2	61.4	83.4	97.0
1800	The Top 5%	87.2	50.7	74.3	86.2
1800	The Top 1%	58.6	26.1	54.9	68.4

Lindert estimates are for household level.

Outline

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.
- The probate index records all those who died with wealth above a minimum threshold

The Probate Threshold

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: The Minimum Probate Threshold, 1858-2017

Outline

Buillding 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
- I downloaded, and digitized, all of them, 1.5 million image files

509 **CUMING** William Herbert of "Nani 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 17*s.* 8*d.* Resworn £19856 17*s.* 8*d.*

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 3*s.* 4*d.*

CUMMING John of 113 Club Garden-road Sheffield died 23 June 1905 Probate **Wakefield** 11 August to Thomas Henry Cumming elementary-teacher Effects £235 4*s.* 2*d.*

CUMMING John Morley of Pathali Ram Assam India died 28 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 Cummi.g colonel in H.M. army Effects £1242 0*s.* 9*d.*

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" Cheltenham 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 1*s.* 4*d.* Resworn £108,291-9-0.

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** 9 September to Alexander Mark Turnbull solicitor and Thomas Gray solicitor's-clerk Effects £456 12*s.* 6*d.*
Resworn £ 716 " 7 " 6

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 7*s.* 3*d.*

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 £252 11*s.* 6*d.*

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 13*s.*

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 16*s.* 7*d.*

CUMMINS Isabella of 8 Tehidy-terrace Falmouth Cornwall (wife of Henry Royse Cummins) died 27 November 1905 Administration **Bodmin** 28 December to the said Henry Royse Cummings physician and surgeon Effects £400. Resworn £975-7-7

CUMMINS Robert of 40 Newgate-street Bishop Auckland county Durham died 9 November 1900 Probate **Durham** 13 April to Mary Ann Cummins widow Effects £56 12*s.* Resworn £ 163.12.-

938 **CUMNER** Thomas of 15 Prince-street Deptford Kent died 3 March 1905 Probate **London** 27 July to John Stagg warehouseman Effects £15 4*s.*

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 18*s.* 6*d.*

CUNDILL George of Scagglethorpe Yorkshire farmer died 4 November 1905 Probate **York** 16 December to Robert Cundill and George Arthur Cundill farmers Effects £610 16*s.*

CUNDY Francis of the "Gables" Watering 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 16*s.*

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 5*s.* 11*d.* 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 14*s.* 10*d.*

1905.

509 **CUMING** William Herbert of "Nani 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 17*s.* 8*d.* Resworn £19856 17*s.* 8*d.*

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 3*s.* 4*d.*

CUMMING John of 113 Club Garden-road Sheffield died 23 June 1905 Probate **Wakefield** 11 August to Thomas Henry Cumming elementary-teacher Effects £235 4*s.* 2*d.*

CUMMING John Morley of Pathali Ram Assam India died 28 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 0*s.* 9*d.*

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" Cheltenham 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 1*s.* 4*d.* Resworn £108,291-9-0.

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** 9 September to Alexander Mark Turnbull solicitor and Thomas Gray solicitor's-clerk Effects £456 12*s.* 6*d.*
Resworn £ 716 " 7 " 0

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 7*s.* 3*d.*

1511 **CUMMINGS** Richard of the "Half Moon" inn Billingham county Durham died 12 December 1905 Probate **London** 27 December to

No rise of a broad based 'middle' class.

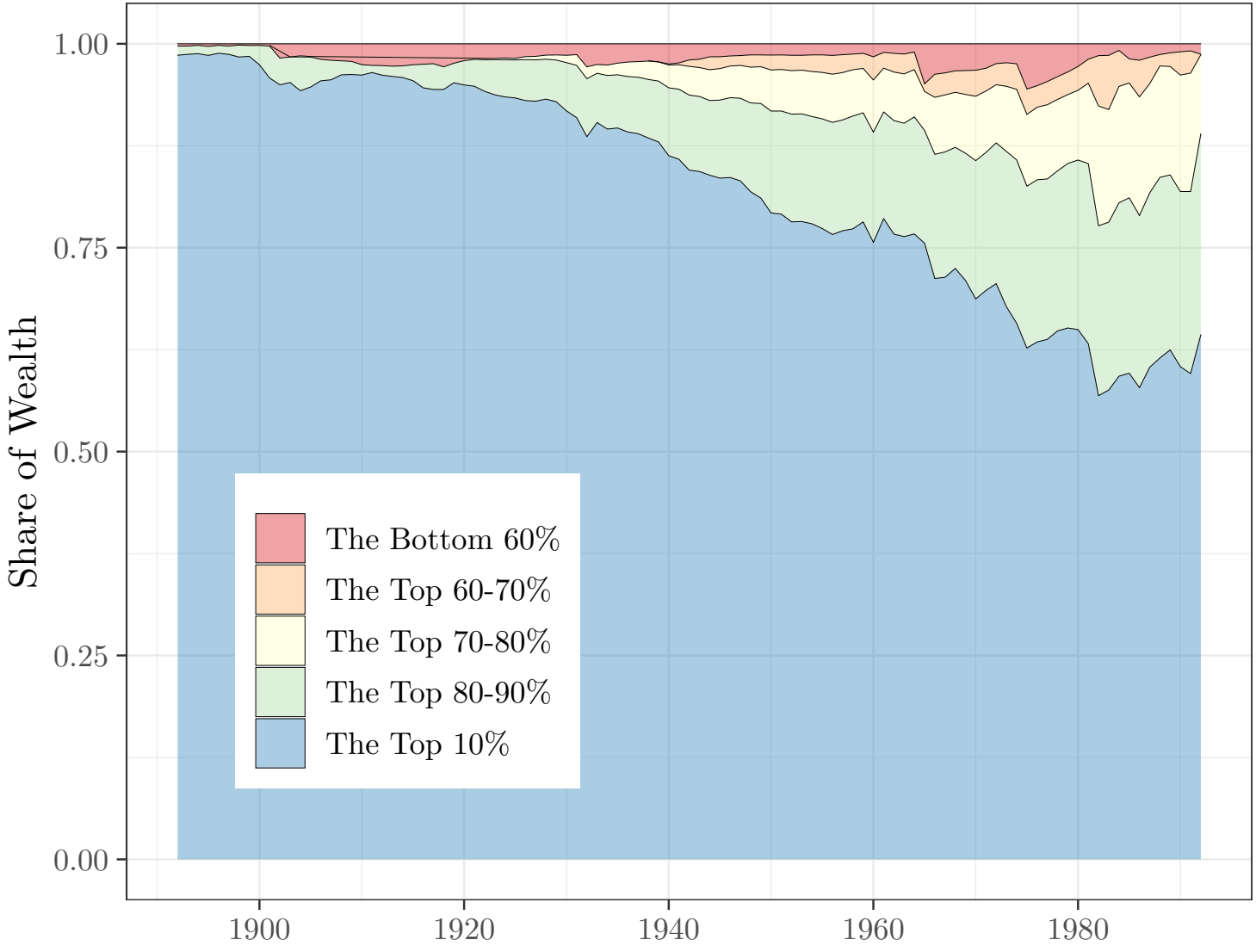


Figure: English Wealth Holding by Decile, 1892-1992

Outline

Outline

The Idea

- I observe all wealth at death in England, 1892-1992
  - Using this, I estimate 'hidden' inherited wealth, in *any* form
  - $>$ 20% is hidden
- This 'hidden' inherited wealth predicts appearance in:
  - Offshore Banking Leaks of 2013-6,
  - House values in 1999,
  - Oxbridge attendance, 1990-2016.
- Eliminates 1/3 of the decline of top 10% wealth-share over the 20th century
- I find 8,549 dynasties that are hiding £7.7 Billion

Wealth at Death

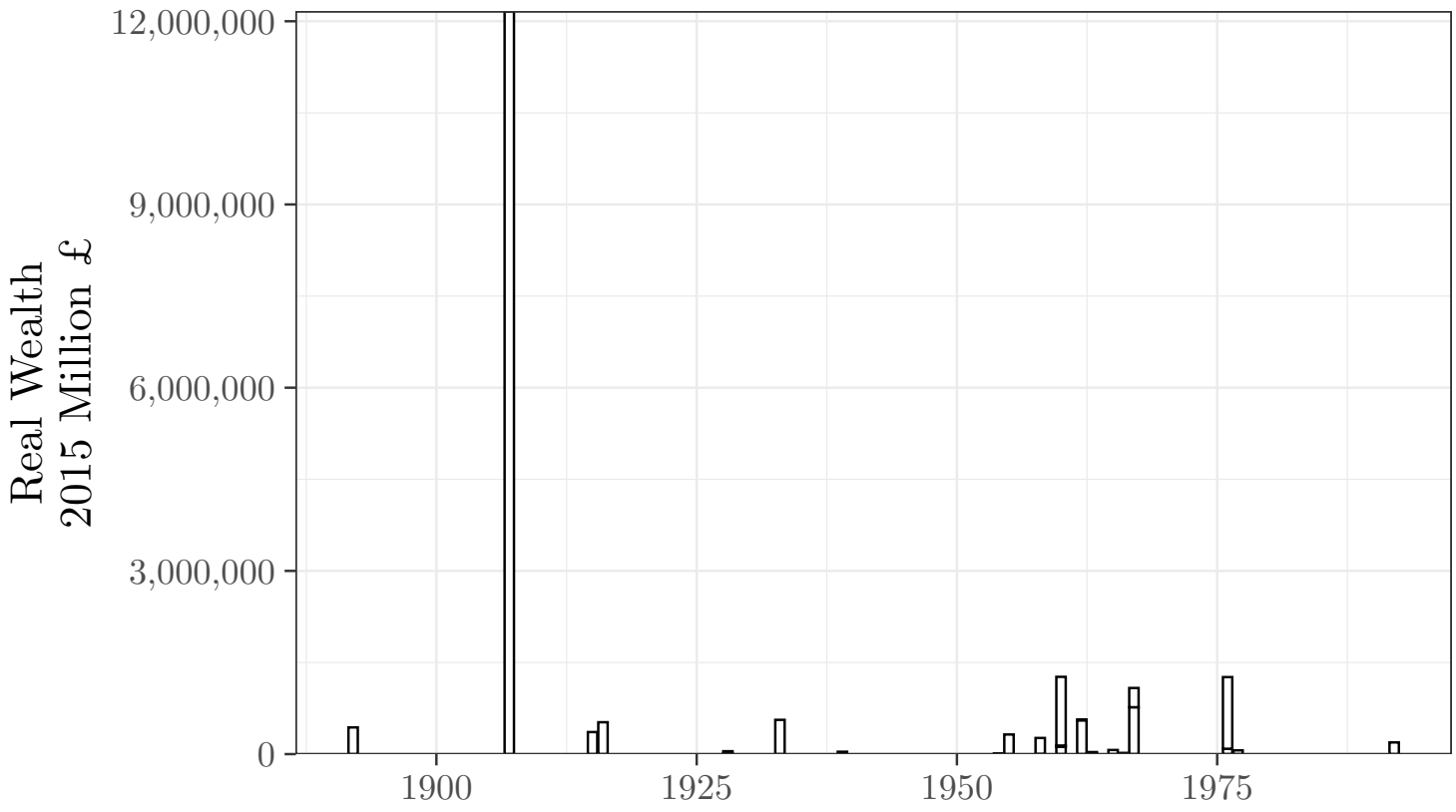


Figure: The Wealth at Death of Members of the Axxxxxxx Dynasty

Source: 100% PPR Calendar Sample.

The Concept

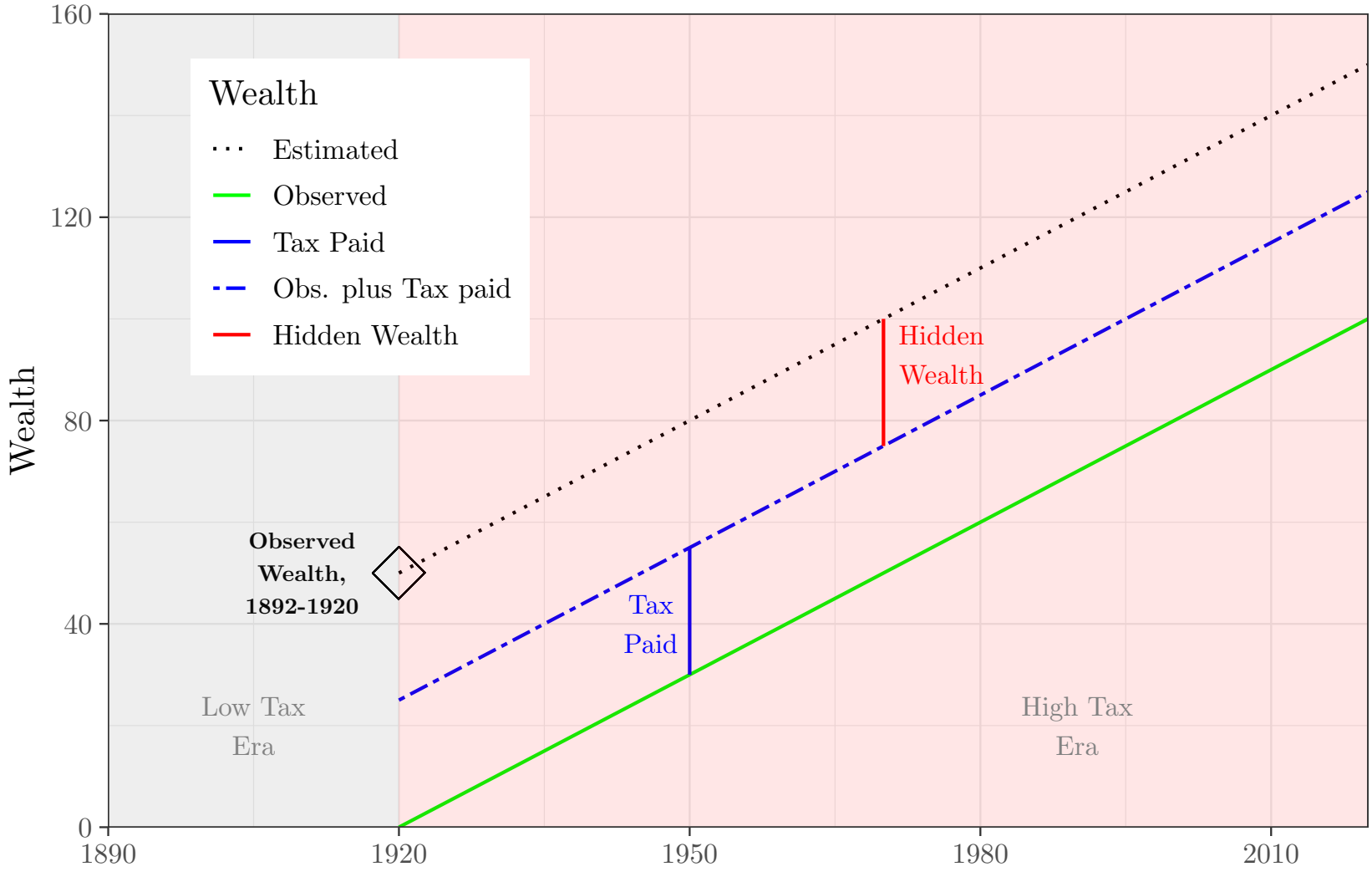


Figure: The Concept

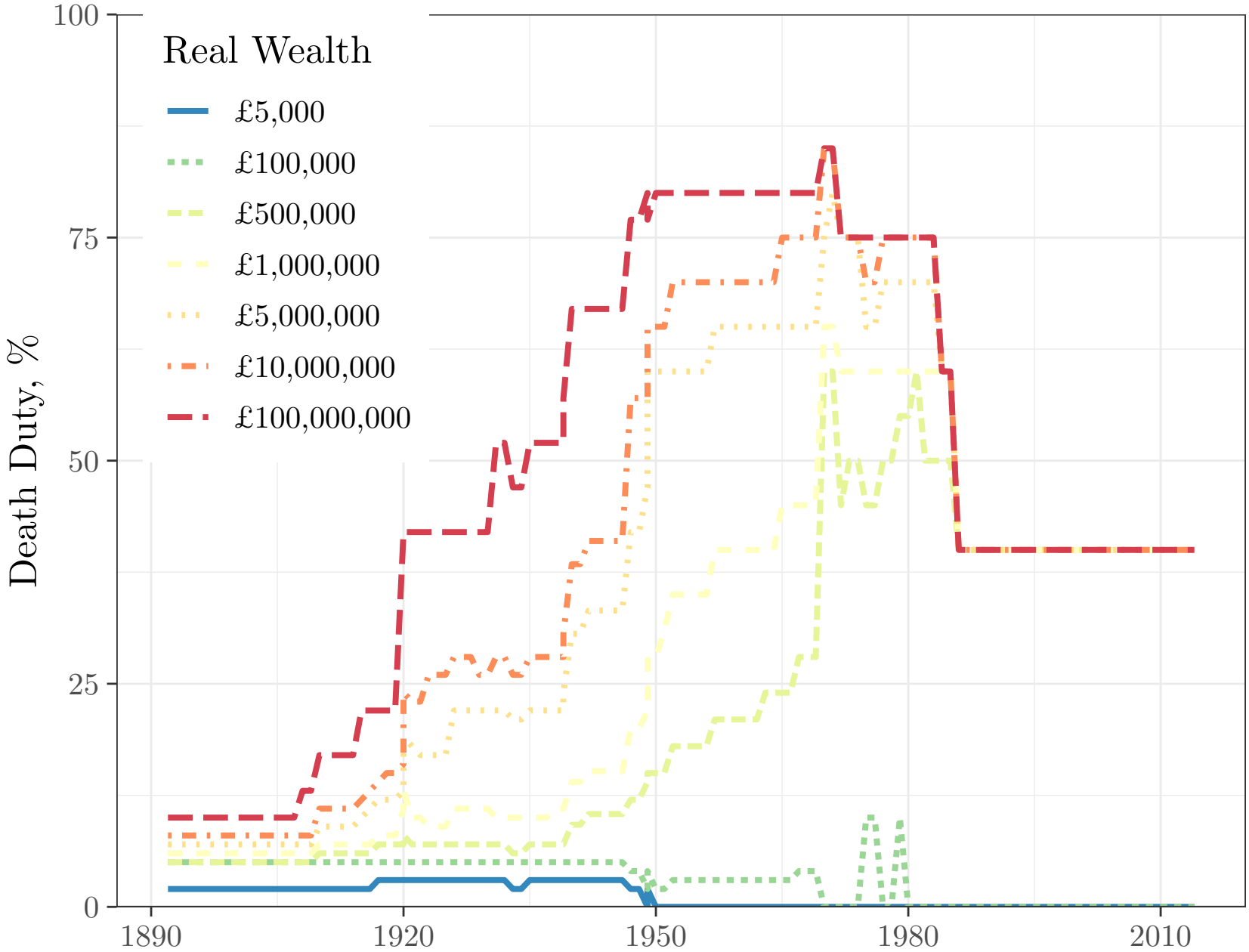


Figure: Death Duties, 1892-2015

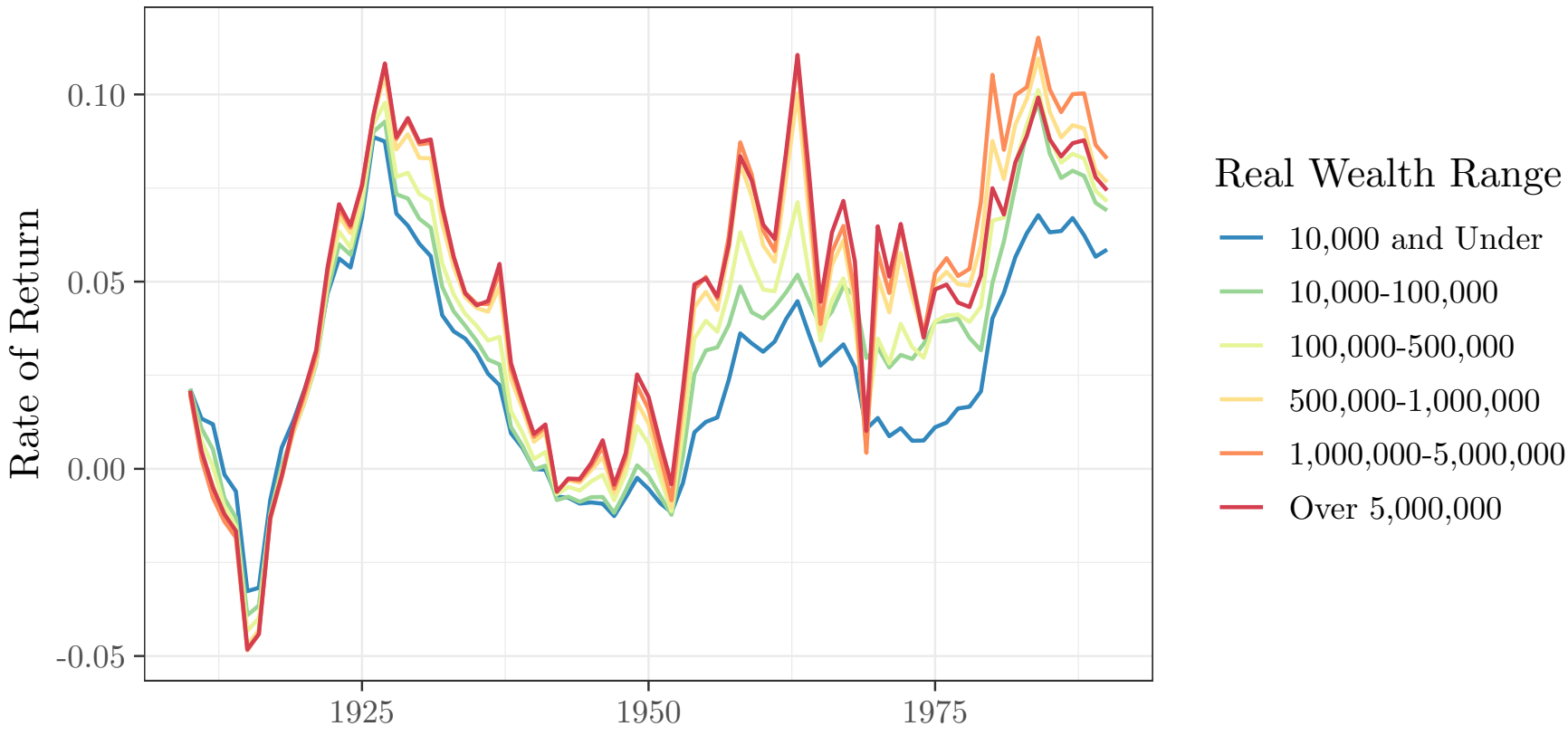


Figure: Rates of Return by Estate Value Range, 1910-1990

Outline

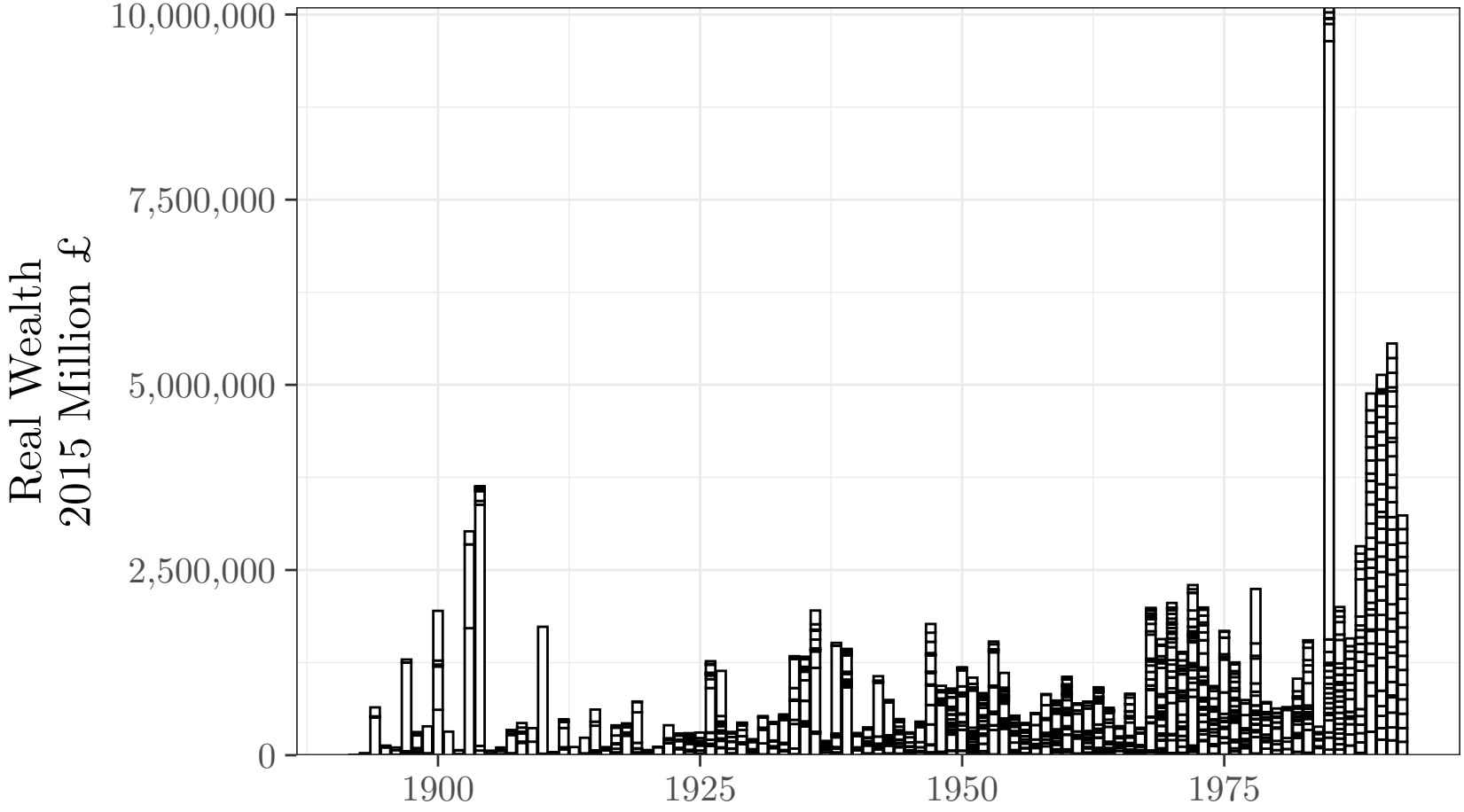


Figure: Cummins's

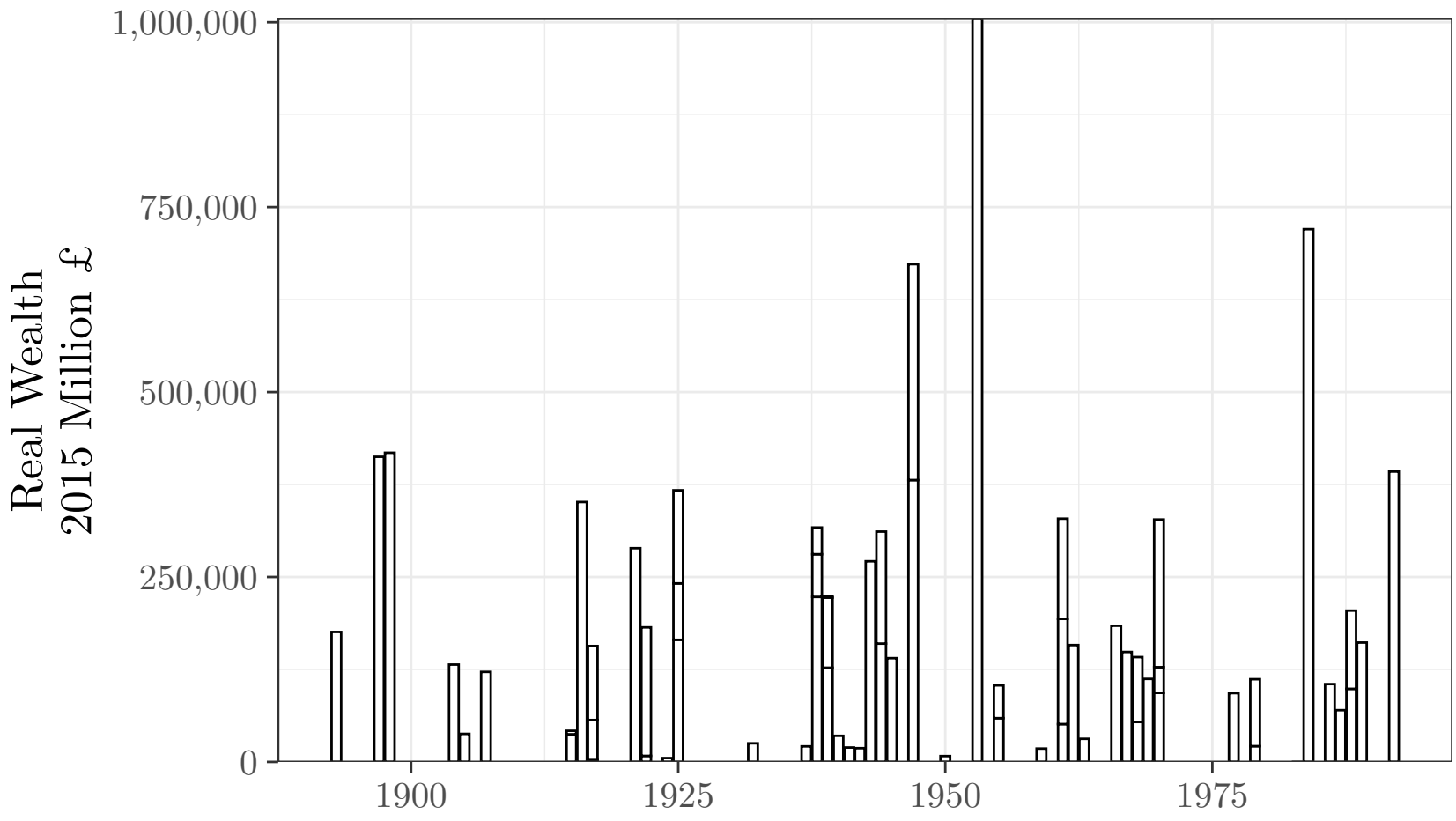


Figure: Comyns's

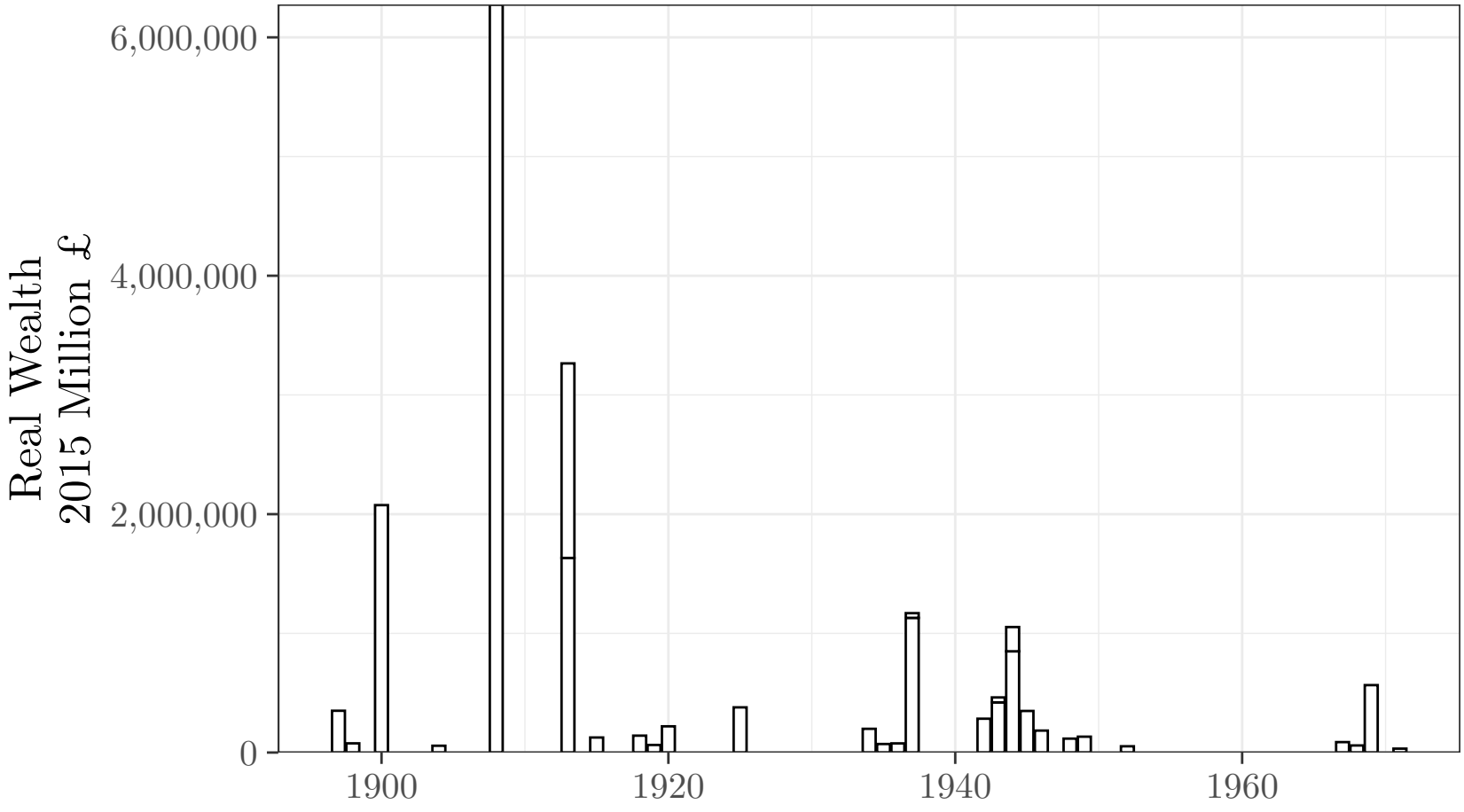


Figure: XXXXX's

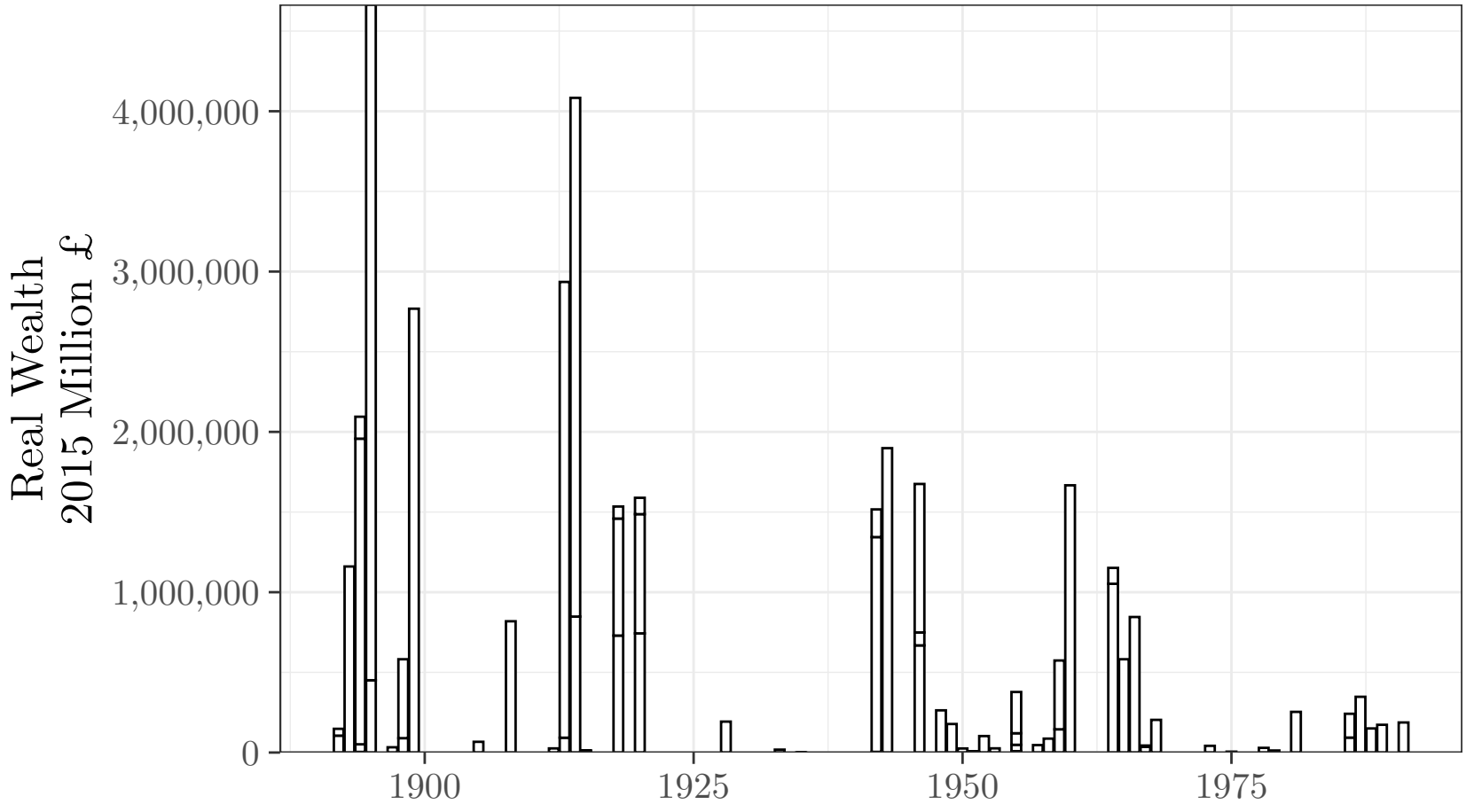


Figure: XXXXX's

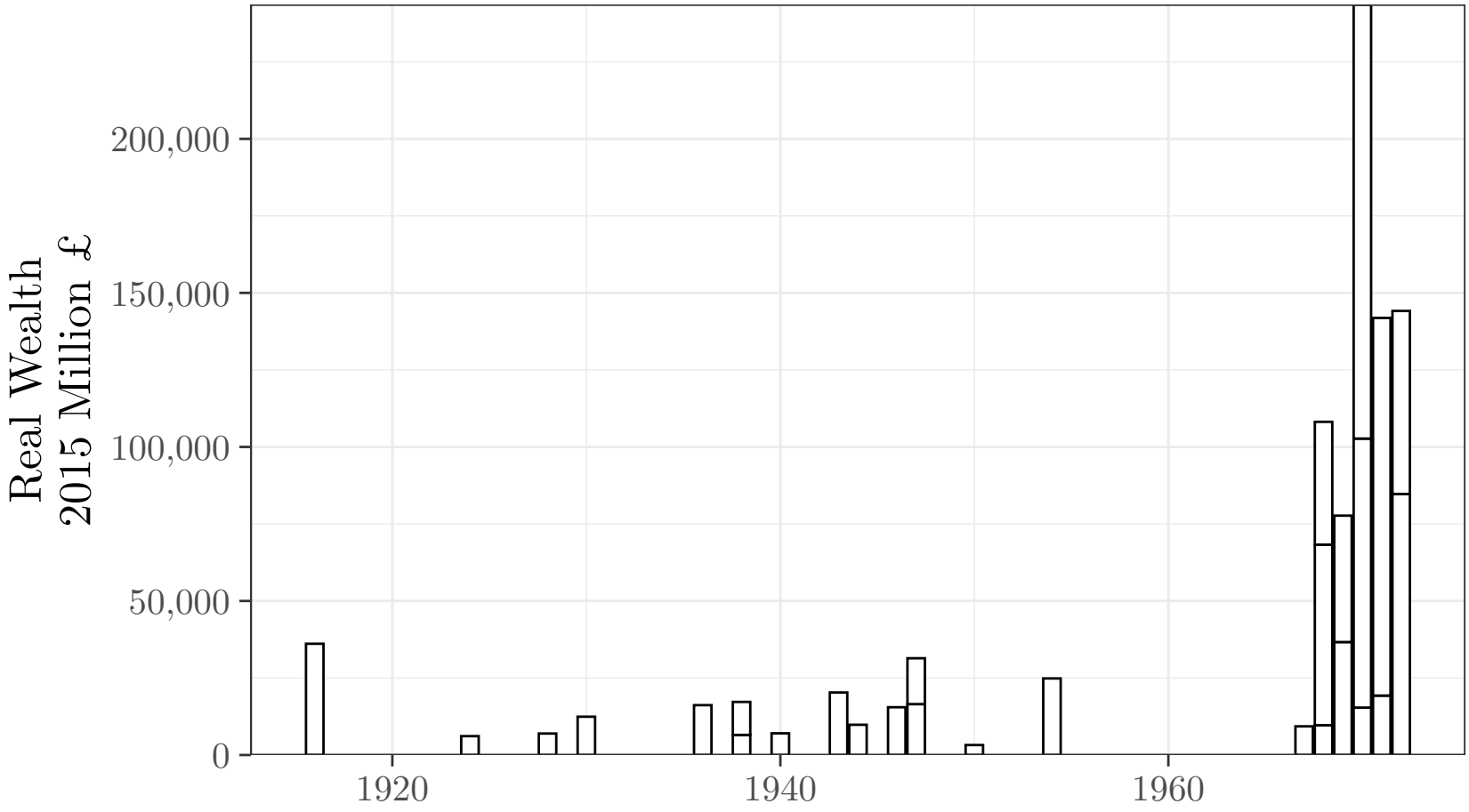


Figure: Dumpleton's

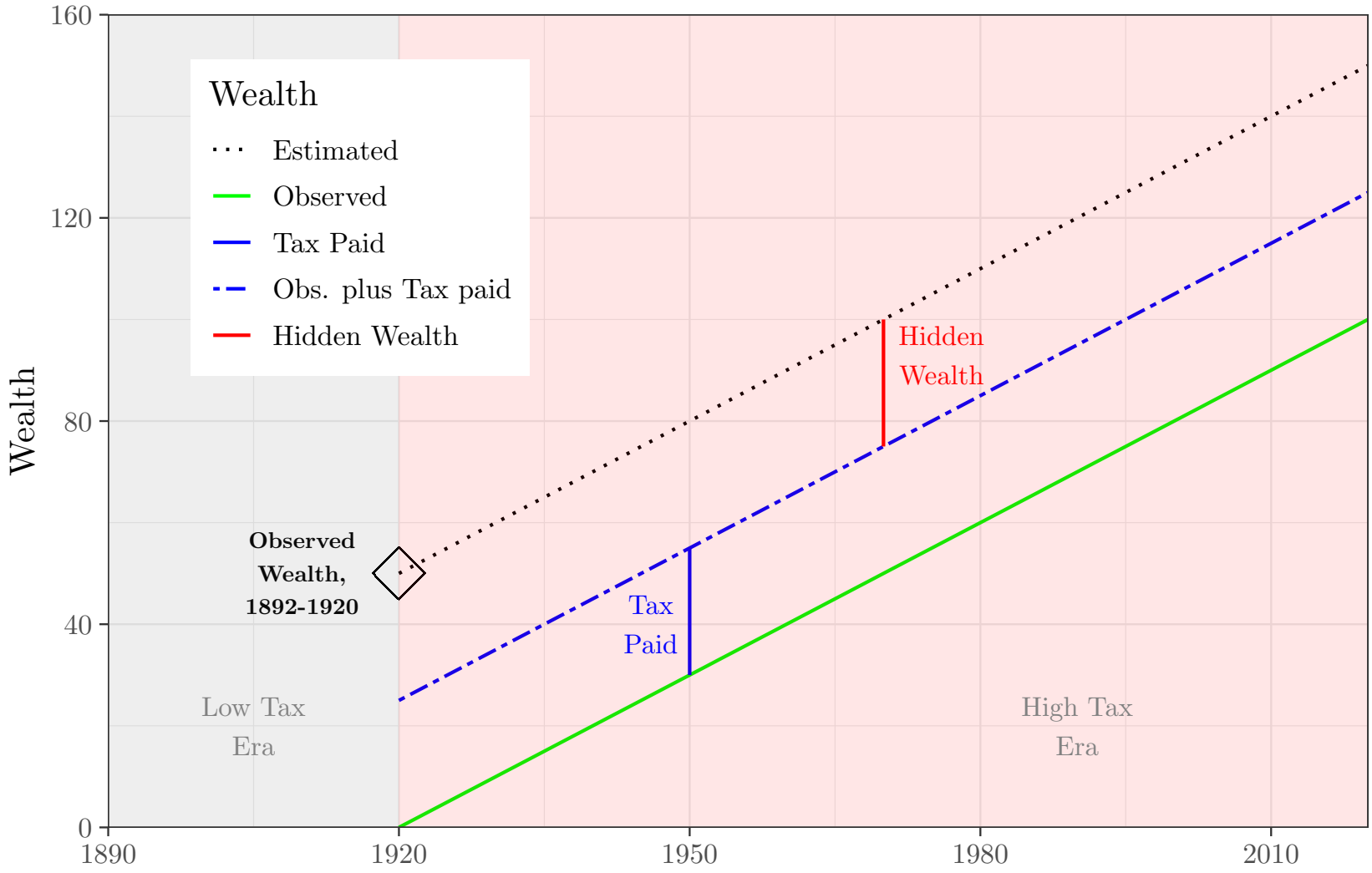


Figure: The Concept

Note: To estimate hidden inherited wealth, I project observed dynastic capital during the low-tax era forward using the net-of-taxes rate-of-return on capital. I subtract wartime destruction and inheritance tax paid. Hidden wealth is the difference between this expected wealth and that actually observed. I assume that wealth observed during

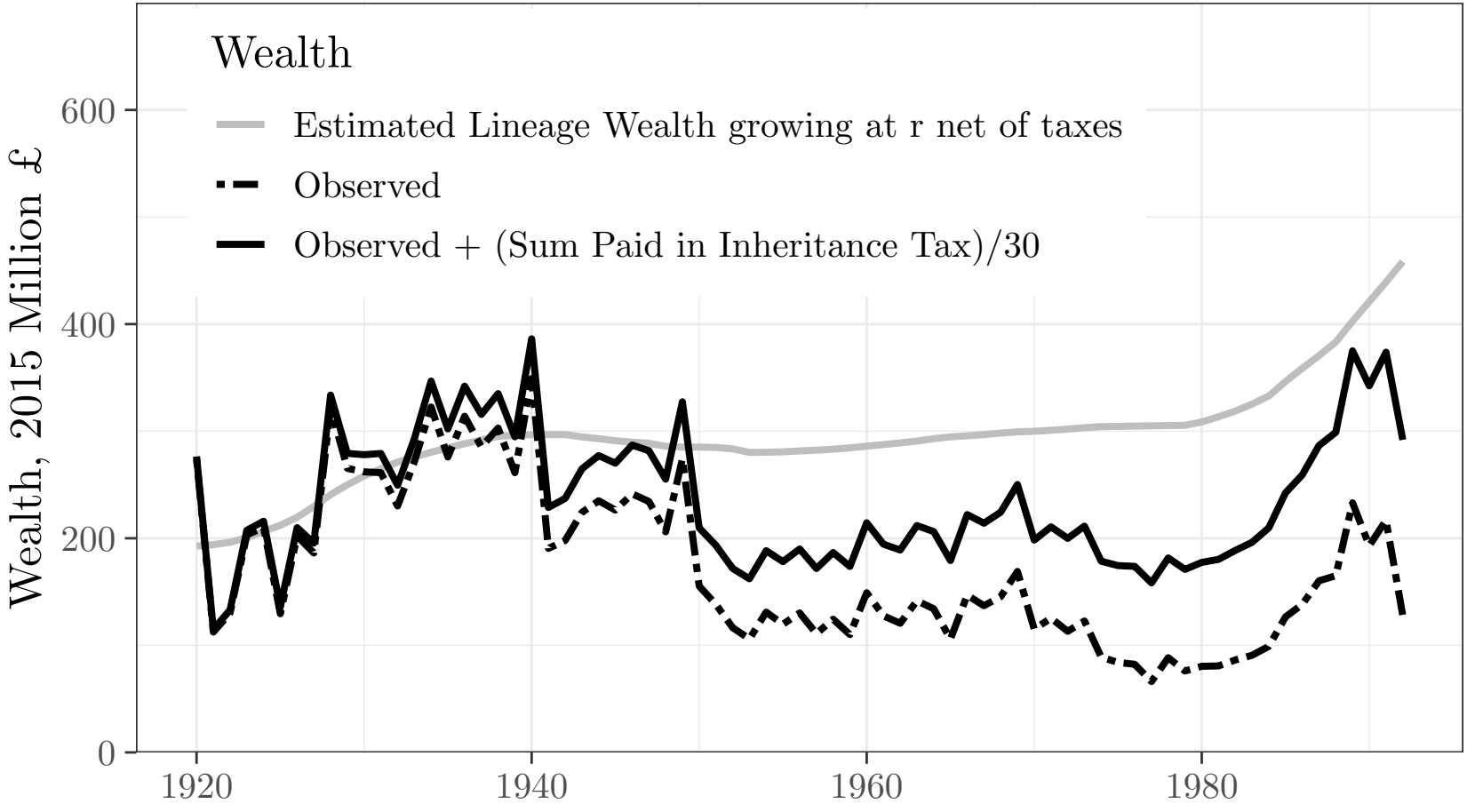


Figure: Victorian Elite Lineages

Wealth, Hidden Wealth and the Proportion of Hidden Wealth as Predictors of Presence in the Paradise Papers

Table: Wealth, Hidden Wealth and the Proportion of Hidden Wealth as Predictors of Presence in the Paradise Papers

	Paradise Dummy				
	(1)	(2)	(3)	(4)	(5)
Wealth Percentile: 70-80	1.110	1.081	1.104	1.114	1.115
	[1.328]	[.997]	[1.260]	[1.377]	[1.391]
80-90	1.111	1.075	1.107	1.113	1.116
	[1.343]	[.928]	[1.288]	[1.352]	[1.388]
The Top 10%	1.407	1.349	1.398	1.401	1.411
	[4.442]***	[3.942]***	[4.348]***	[4.389]***	[4.470]***
D_{Hider}	1.267				
	[3.633]***				
Hidden Wealth		1.034			
		[2.300]*			
HW: 0-.2m			1.134		
			[1.212]		
HW: .2m-1m			1.301		
			[2.798]**		
HW:>1m			1.418		
			[2.951]**		
Prop. Hidden, α				1.399	
				[3.320]***	
α: 0-.5					1.182
					[1.774]
α .5-.75					1.333
					[2.506]*
α. 75-.9					1.440
					[2.700]**
α>.90					1.211
					[1.198]
N_{2002}	1.004	1.004	1.004	1.004	1.004
	[13.106]***	[12.944]***	[13.121]***	[13.079]***	[13.119]***

Outline

Wealth, Hidden Wealth and the Proportion of Hidden Wealth as Predictors of House Price in 1999

Table: Wealth, Hidden Wealth and the Proportion of Hidden Wealth as Predictors of House Price in 1999

	House Price				
	(1)	(2)	(3)	(4)	(5)
Wealth Percentile: 70-80	14.9	12.2	13.6	15.9*	15.6*
	(7.6)	(7.6)	(7.6)	(7.7)	(7.7)
80-90	48.0***	44.6***	46.6***	48.7***	48.5***
	(7.8)	(7.8)	(7.8)	(7.9)	(7.9)
The Top 10%	93.8***	88.4***	91.7***	93.8***	94.1***
	(8.2)	(8.2)	(8.2)	(8.2)	(8.2)
D_{Hider}	37.7***				
	(6.3)				
Hidden Wealth		10.5***			
		(1.9)			
HW: 0-.2m			1.8		
			(9.8)		
HW: .2m-1m			43.1***		
			(9.4)		
HW:>1m			86.6***		
			(12.5)		
Prop. Hidden, α				59.0***	
				(9.9)	
α: 0-.5					24.1**
					(9.3)
α .5-.75					56.8***
					(11.5)
α. 75-.9					50.1***
					(14.1)
α>.90					27.9
					(15.3)
N_{2002}	−.6***	−.6***	−.6***	−.6***	−.6***
	(.03)	(.03)	(.03)	(.03)	(.03)

Wealth, Hidden Wealth and the Proportion of Hidden Wealth as Predictors of Oxbridge Attendance Rate 1990-2016

Table: Wealth, Hidden Wealth and the Proportion of Hidden Wealth as Predictors of Oxbridge Attendance Rate 1990-2016

	Oxbridge Attendance Rate (Z)				
	(1)	(2)	(3)	(4)	(5)
Wealth Percentile: 70-80	.100***	.095***	.097***	.103***	.103***
	(.022)	(.022)	(.022)	(.022)	(.022)
80-90	.116***	.111***	.113***	.118***	.118***
	(.023)	(.023)	(.023)	(.023)	(.023)
The Top 10%	.263***	.253***	.259***	.264***	.264***
	(.024)	(.024)	(.024)	(.024)	(.024)
D_{Hider}	.068***				
	(.018)				
Hidden Wealth		.028***			
		(.005)			
HW: 0-.2m			−.022		
			(.028)		
HW: .2m-1m			.092***		
			(.027)		
HW:>1m			.175***		
			(.036)		
Prop. Hidden, α				.115***	
				(.028)	
α: 0-.5					.030
					(.027)
α .5-.75					.120***
					(.033)
α. 75-.9					.079*
					(.040)
α>.90					.072
					(.044)
N_{2002}	.002***	.002***	.002***	.002***	.002***
	(.0001)	(.0001)	(.0001)	(.0001)	(.0001)
Observations	18,921	18,921	18,921	18,921	18,921

Outline

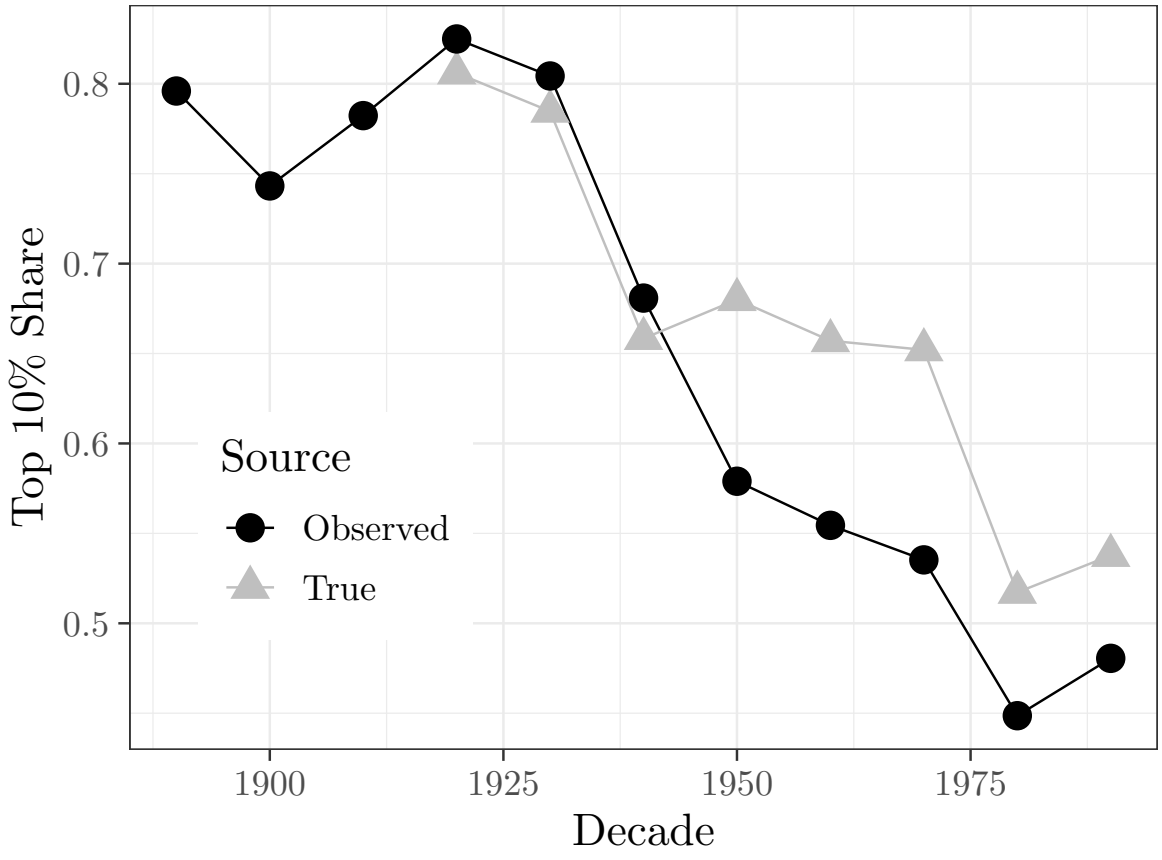


Figure: Top 10% Shares, Observed and 'True'

Notes: Calculated at the Rare Surname Level. *Source*: 100% PPR Calendar Sample.

Outline

Ethnic Wealth Inequality

- We know little about ethnic based socioeconomic inequality
- And nothing about the history of this
  - Because we have very little data
- What to Do?

The Key Idea

Surnames Can tell us *something* about ethnicity

- **Very Imperfect, a huge and varying error (lots of counter examples)**
  - But **properly** handled; there is information here
- Has the potential to reveal previously **invisible** inequalities

The Irish in England

- The successful assimilation of ethnic minorities into Western economies is one of the biggest challenges facing the Modern World.
- The Irish, in England, provide an historical example of this process.
  - However, this topic has received surprisingly little scholarly attention.

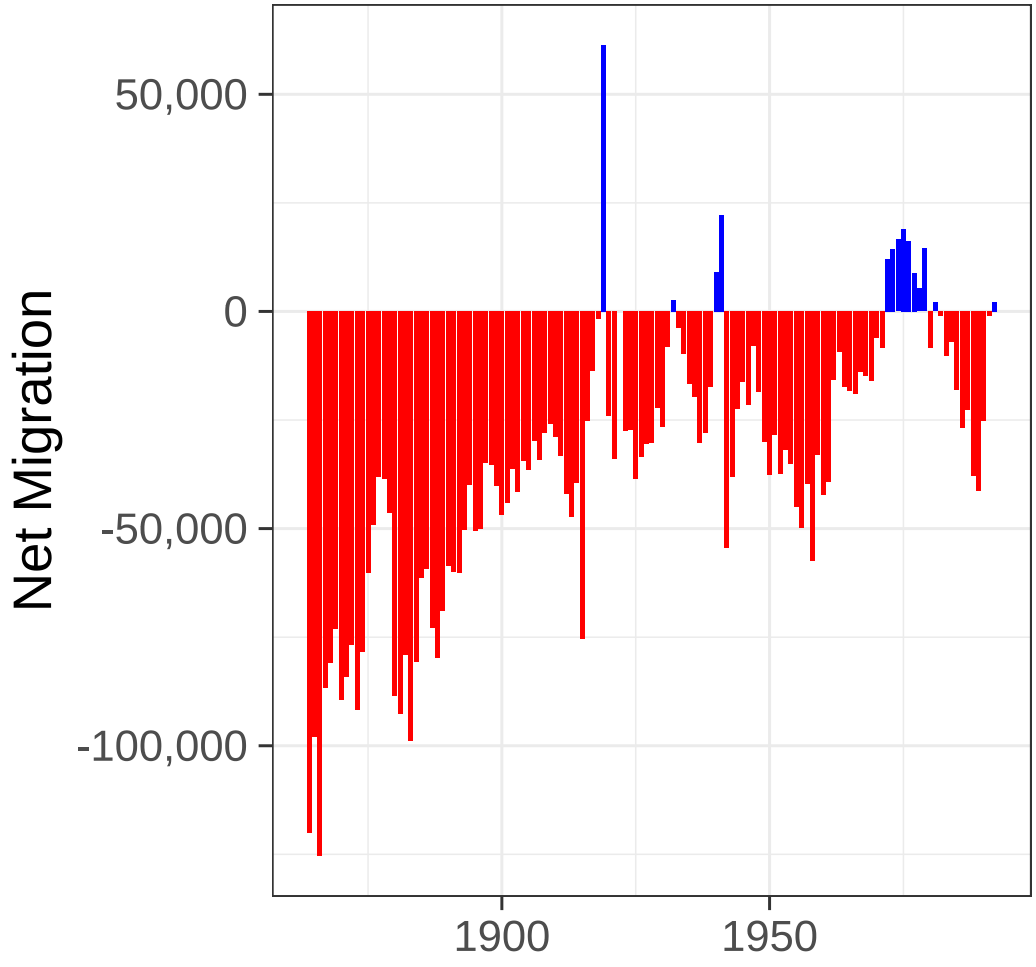


Figure: Net Migration, Ireland, 1850-1992

Sources: Ireland, 1600-1850 Ó Gráda (1979), 1850-1951 Rothenbacher (2005), 1951-2019: cso.ie. England and Wales, 1541-1851: Wrigley and Schofield (1981), 1851-1871Rothenbacher (2005), 1971-2019: ONS.

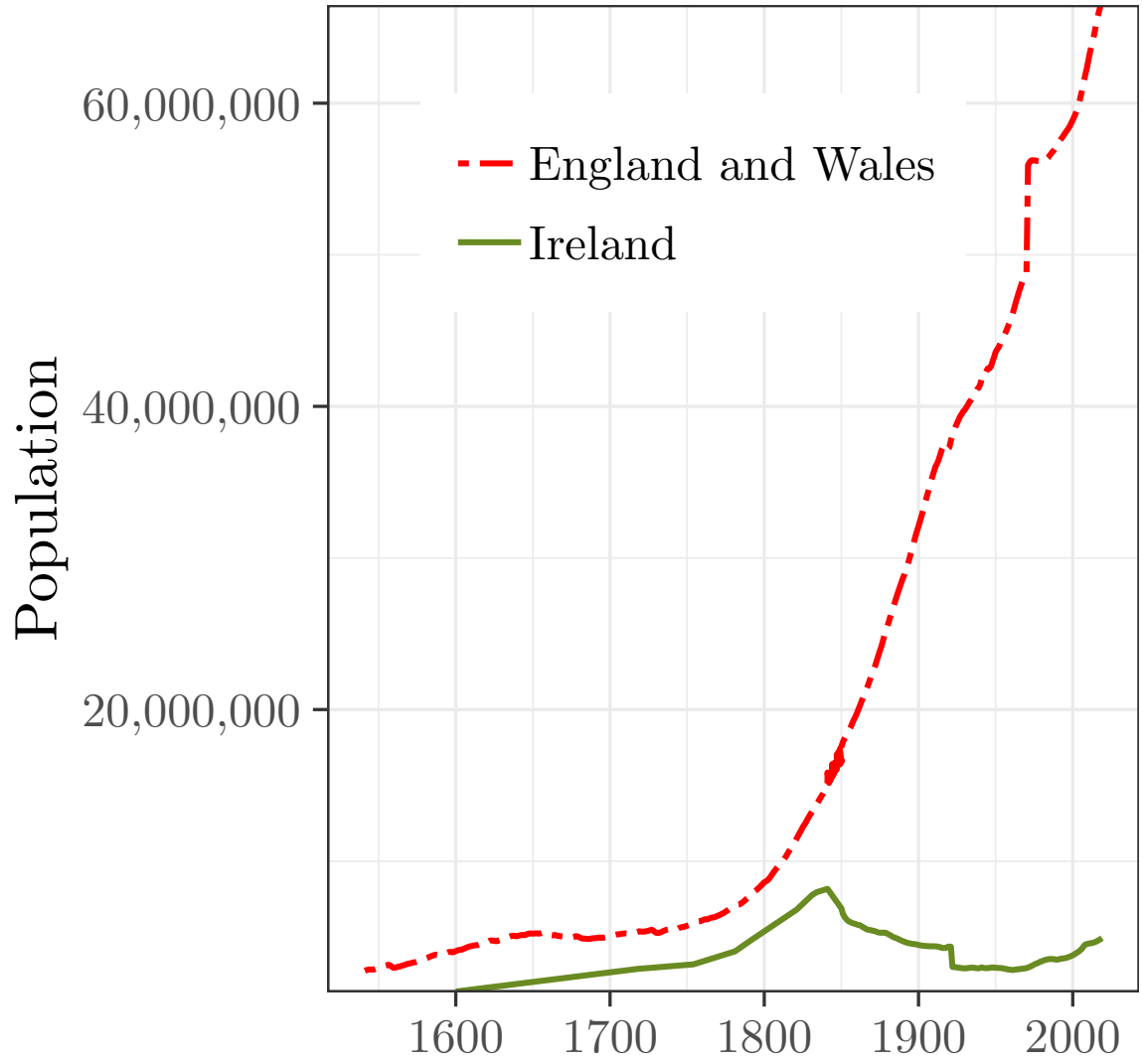


Figure: Population Growth, Ireland compared with England & Wales, 1600-2019

Sources: Ireland, 1600-1850 Ó Gráda (1979), 1850-1951 Rothenbacher (2005),

An underclass?

> *[...] in all the big industrial centres in England there is profound antagonism between the Irish proletariat and the English proletariat. The average English worker hates the Irish worker as a competitor who lowers wages and the standard of life. He feels national and religious antipathy for him. He regards him somewhat like the poor whites of the Southern States of North America regard their black slaves.*

[Karl Marx *28 March 1870* (confidential communication) as quoted by (Deleixhe, 2019, p.223).]

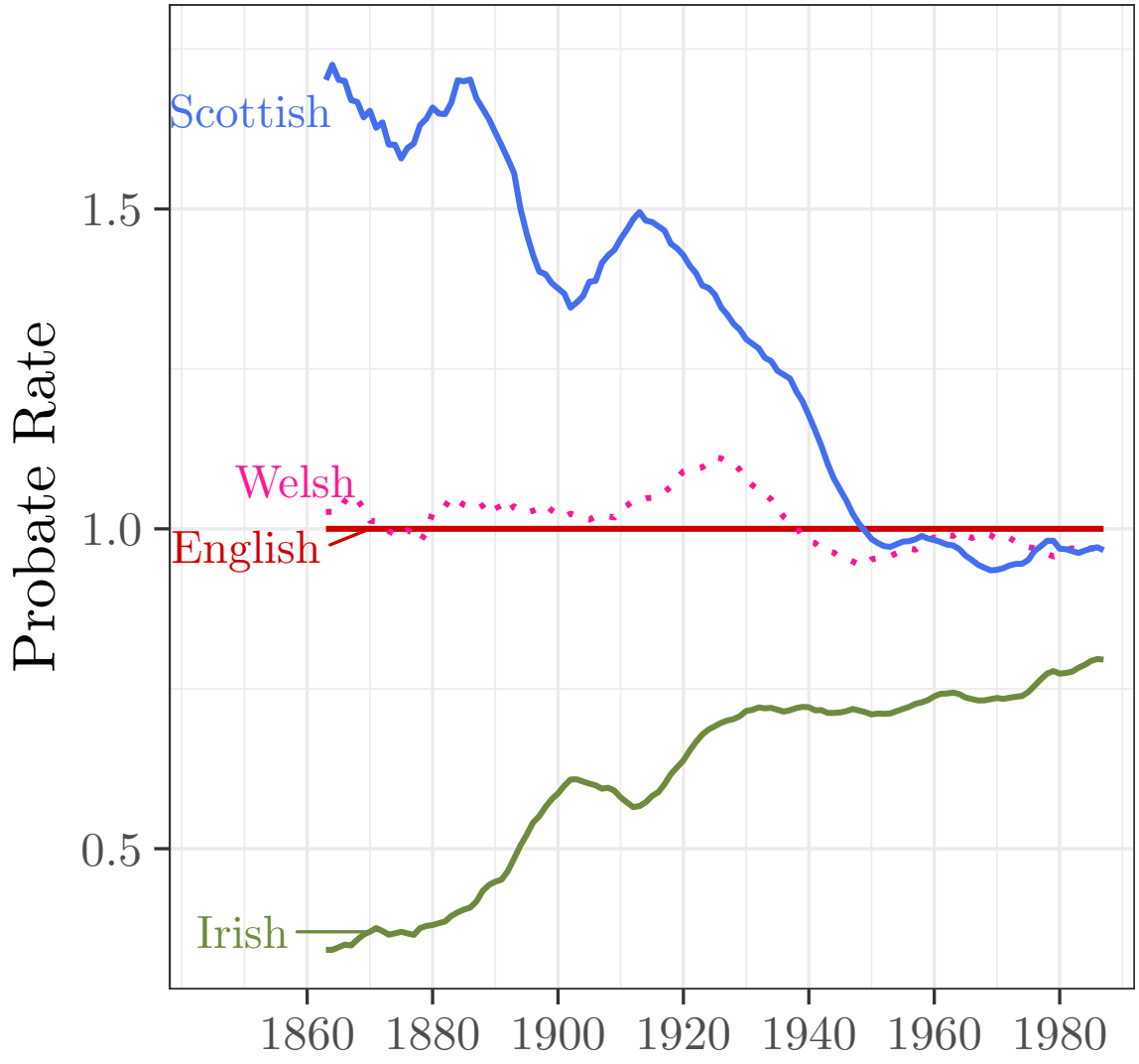


Figure: Probate Rate, 1858-1992

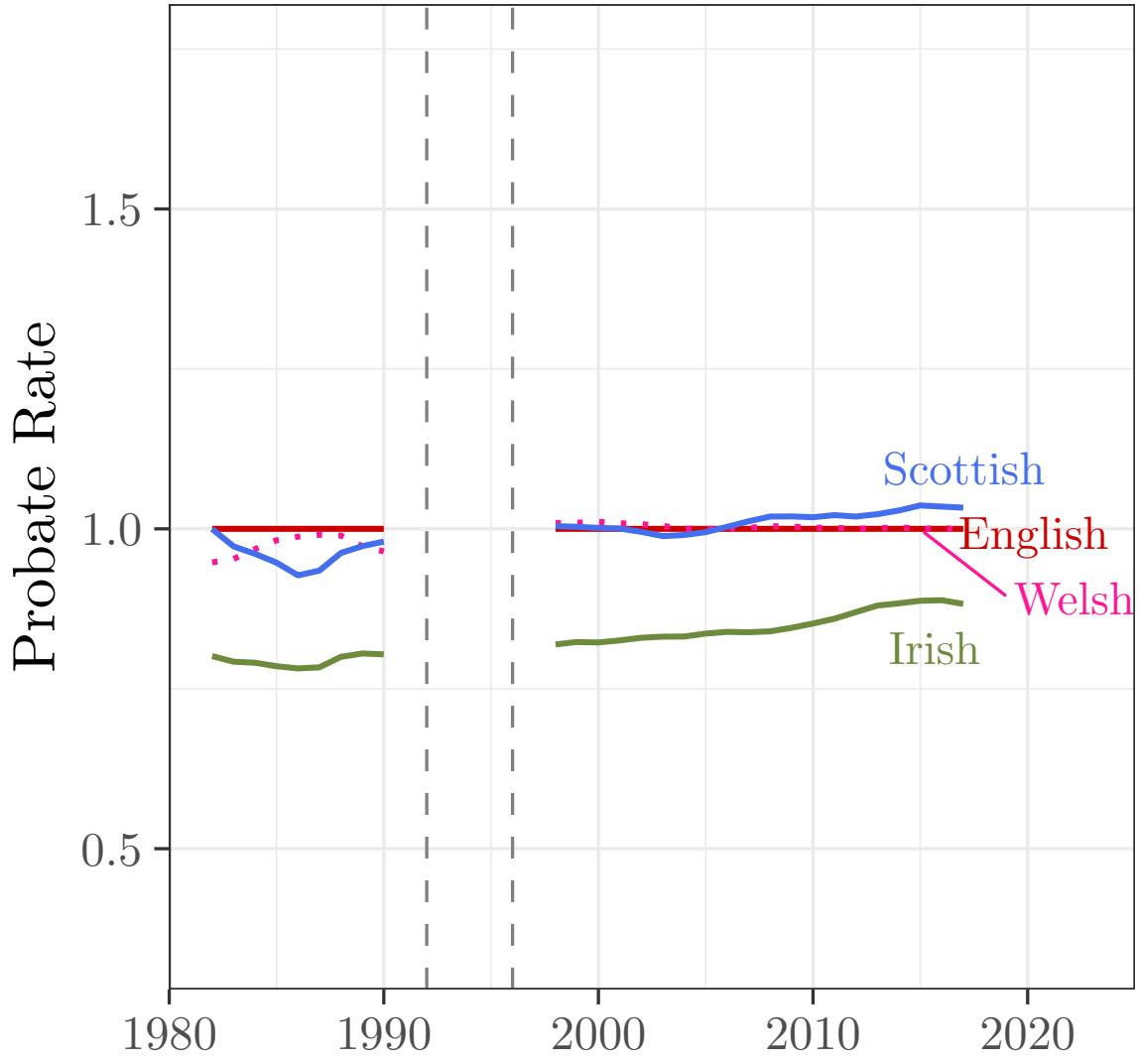


Figure: Probate Rate, 1996-2019

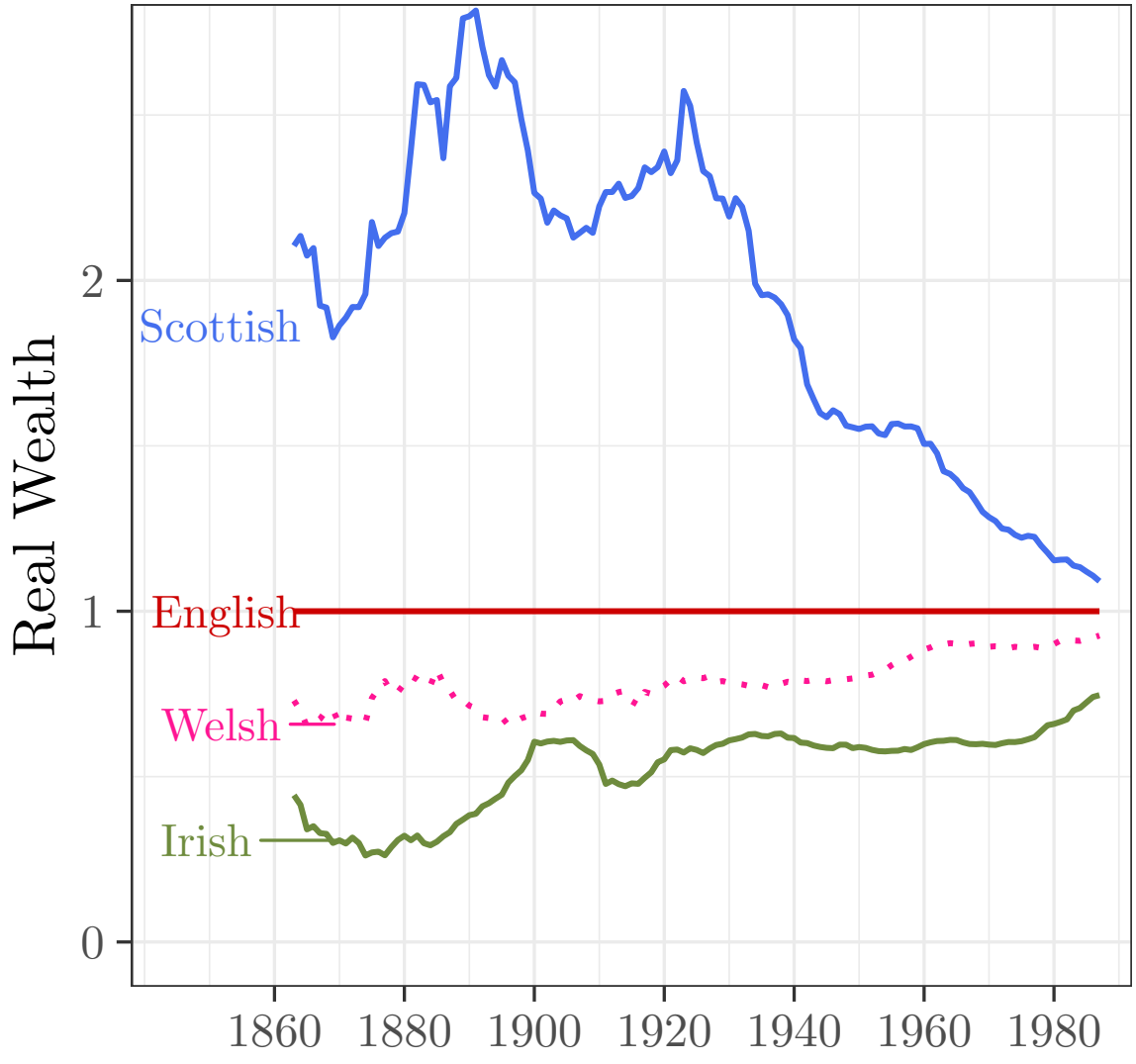


Figure: Average Wealth

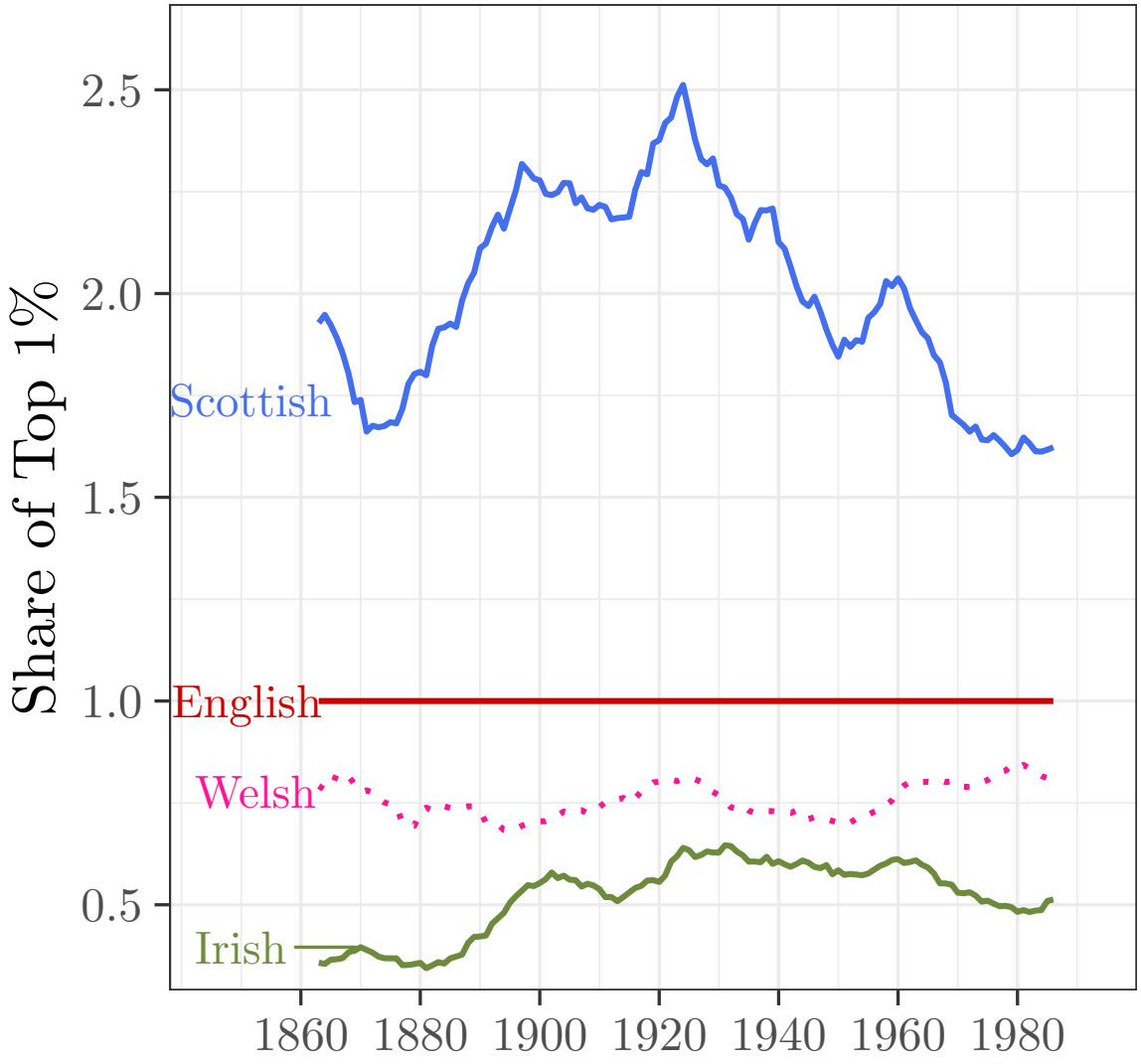


Figure: Share of the Top 1%

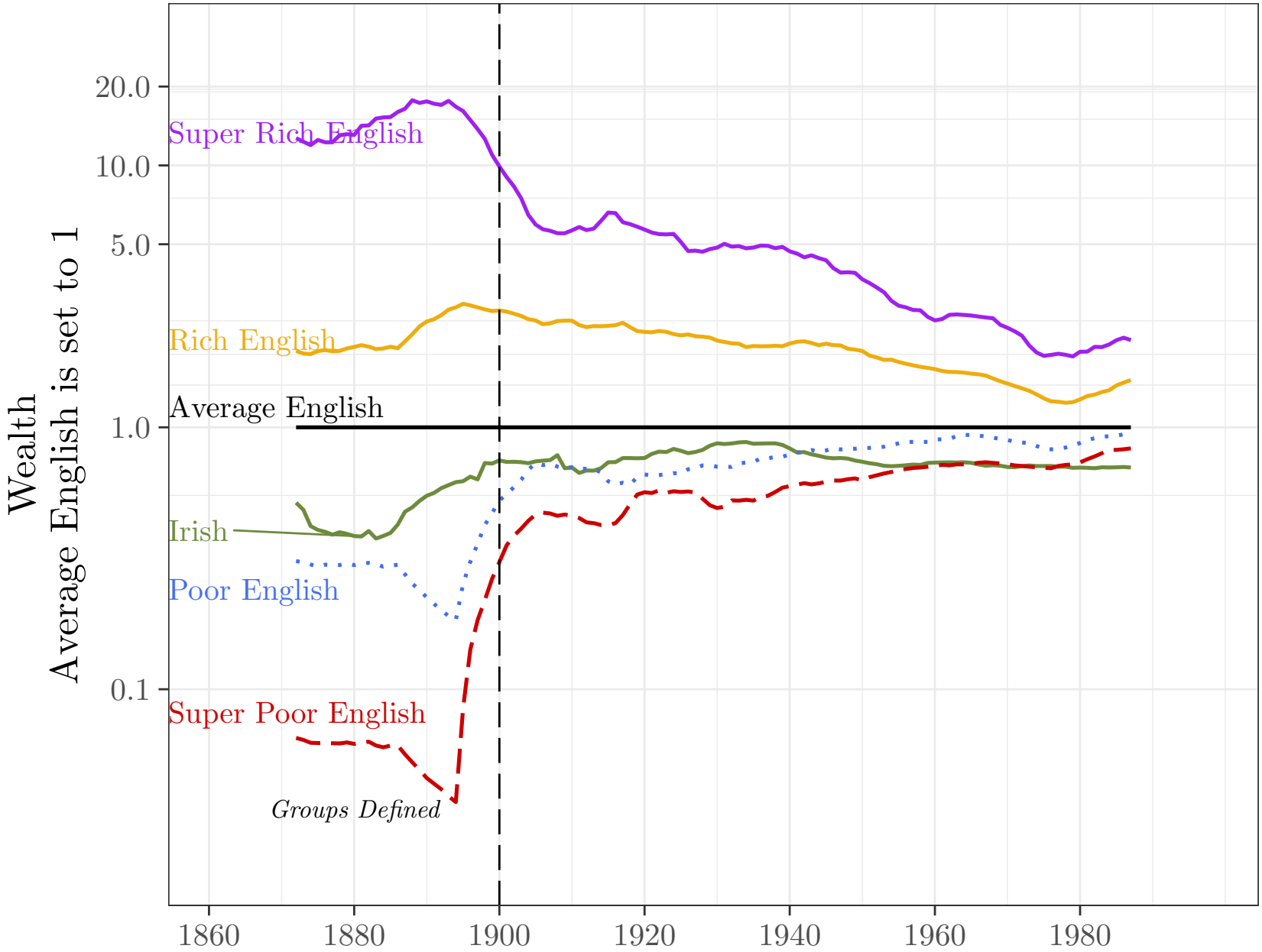


Figure: A Distinctive Irish Wealth Pattern

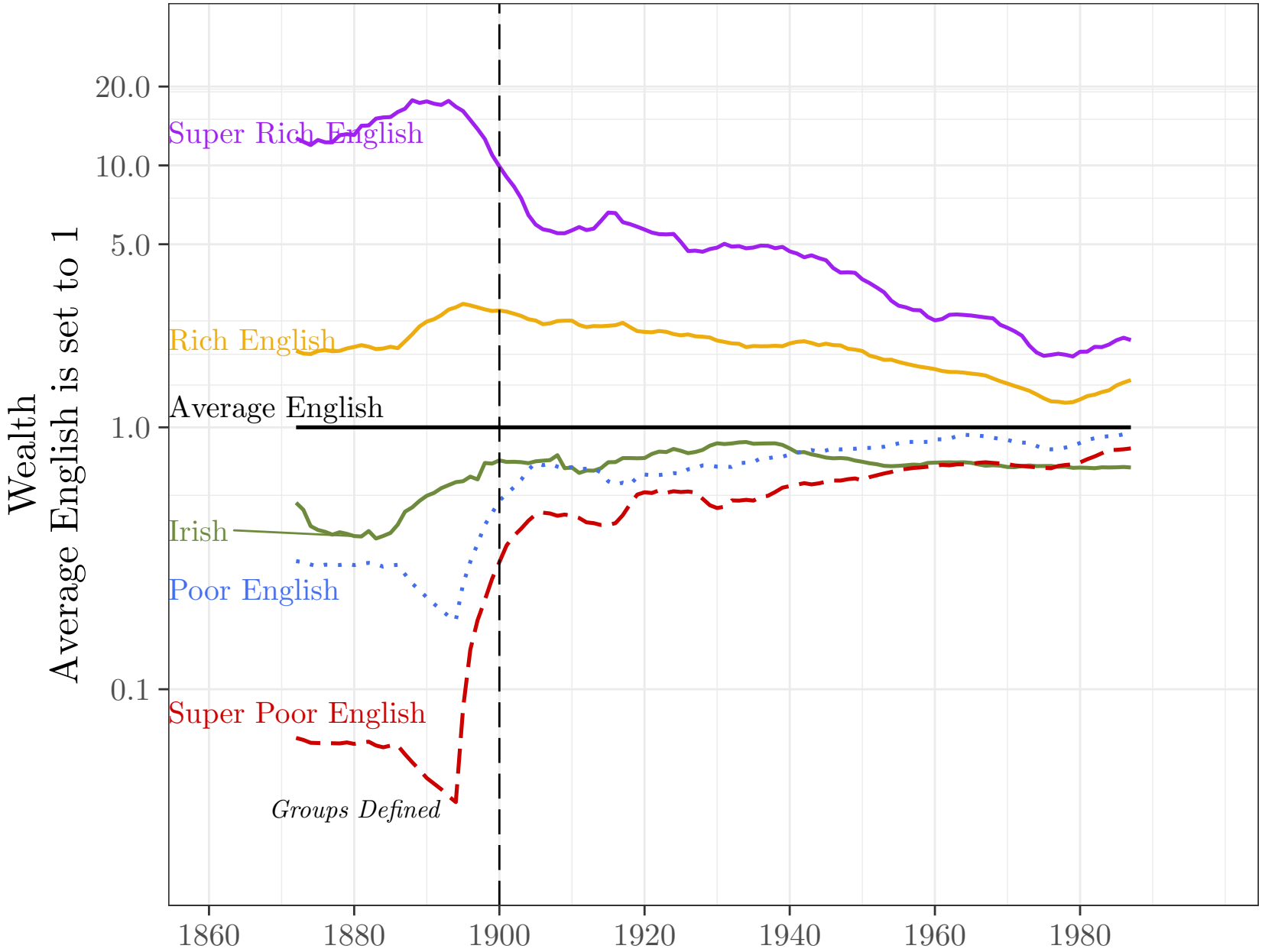


Figure: A Distinctive Irish Wealth Pattern

Framework for Measuring Intergenerational Correlations

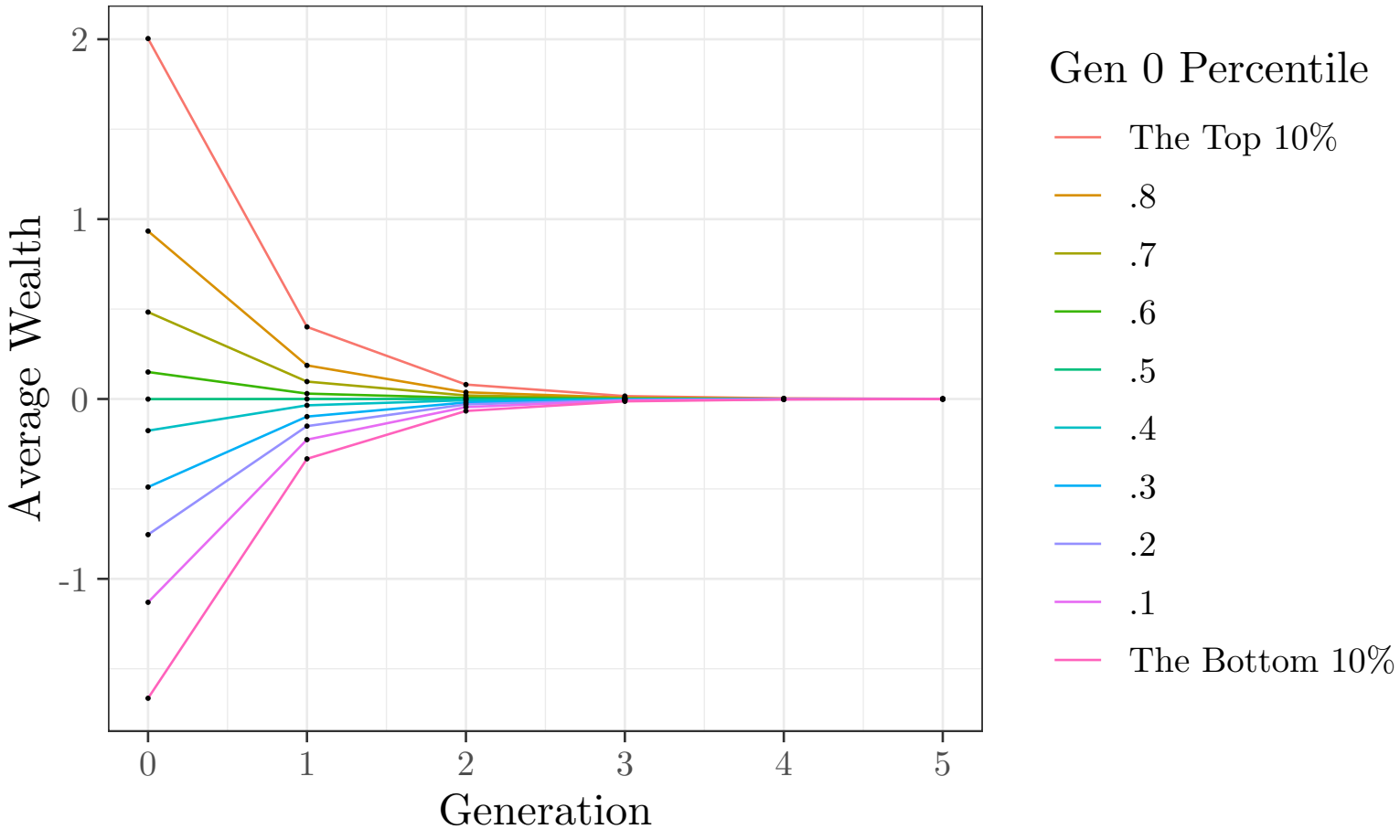
$$y_{t+1} = \beta y_t + u_t$$

β=intergenerational correlation
t=generation

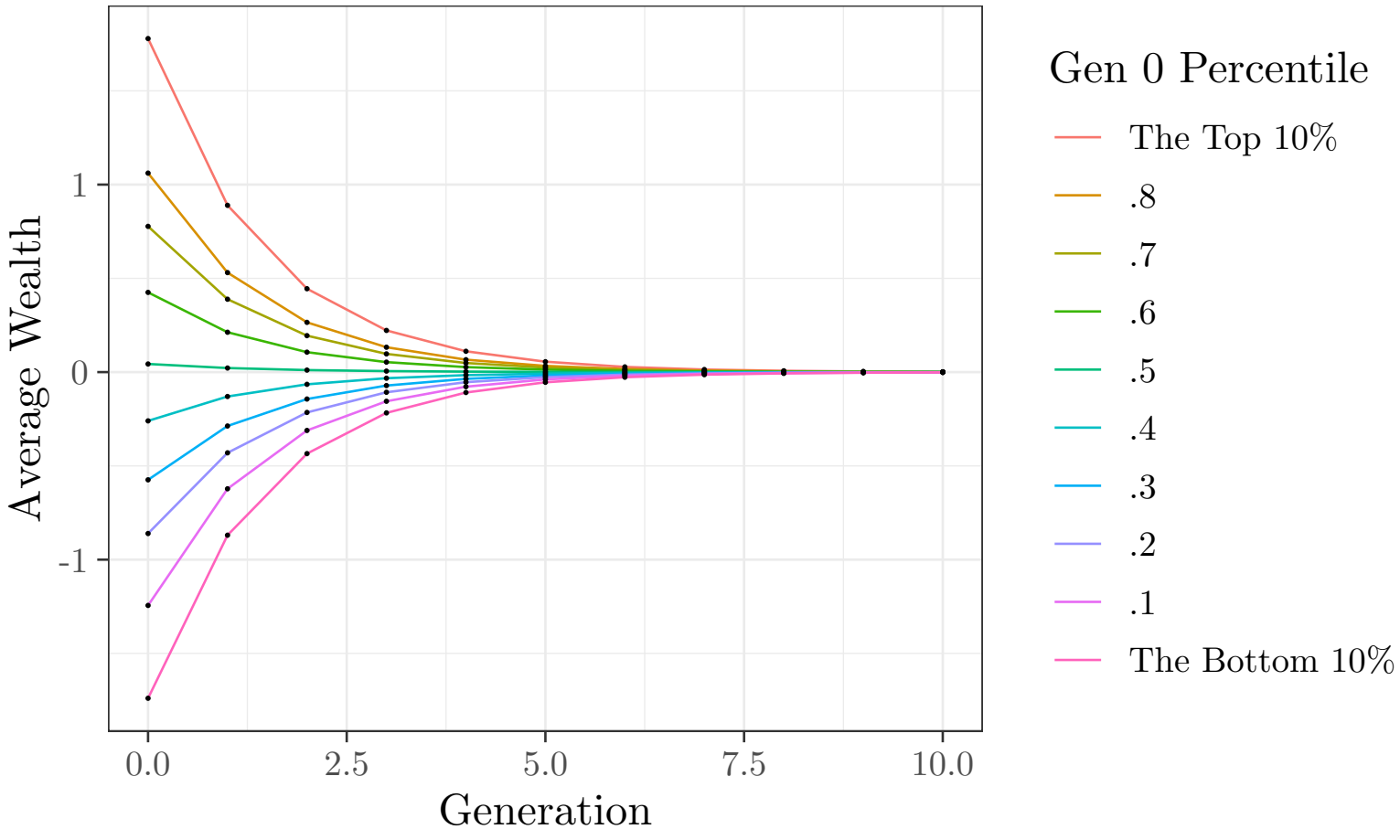
What do different values of b imply?

- b will lie somewhere between 0 and 1.
  - $b = 1$ No convergence, regression towards the mean: The rich stay rich forever, as do the poor.
  - $b = 0$ Complete convergence in 1 generation, Complete and instant regression towards the mean: We live in a classless society (over time, not necessarily in cross section).
- What about other values of b?

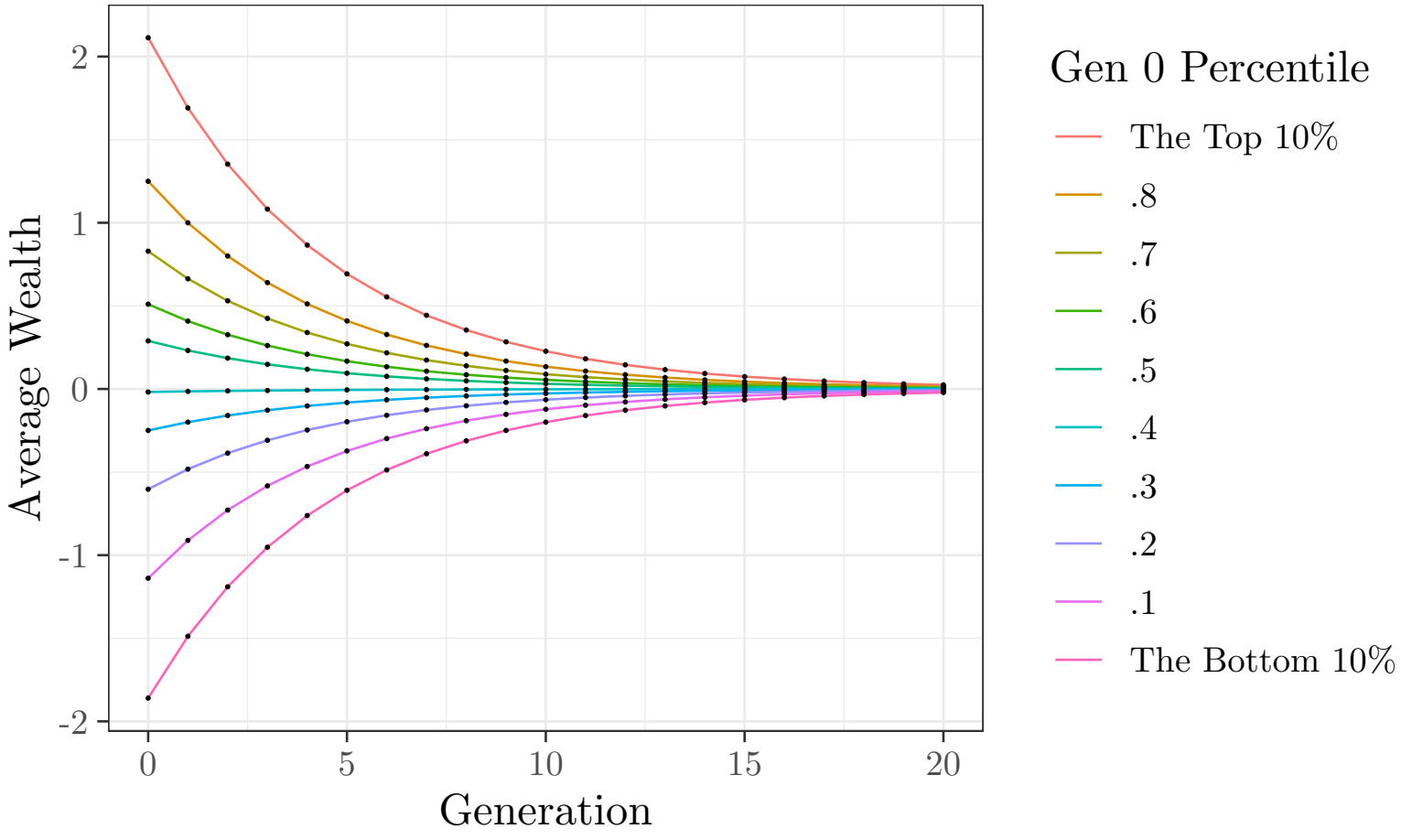
$b = .2$: Convergence in 3/4 gens



$b = .5$: Convergence in 6/7 gens.



$b = .8$: Convergence in >20 generations



Time for Complete Convergence of Initial groups

- (Within 10% of Avg. Wealth)
- b=.2 2/3 gens
- b=.5 6/7 gens
- b=.8 >20 gens
  - Small differences in b have major consequences for social mobility

Friedman (1962, p. 171)

A major problem in interpreting evidence on the distribution of income is the need to distinguish two basically different kinds of inequality; temporary, short-run differences in income, and differences in long-run income status. Consider two societies that have the same annual distribution of income. In one there is great mobility and change so that the position of particular families in the income hierarchy varies widely from year to year. In the other there is great rigidity so that each family stays in the same position year after year. The one kind of inequality is a sign of dynamic change, social mobility, equality of opportunity; the other, of a status society

The Great Gatsby Curve

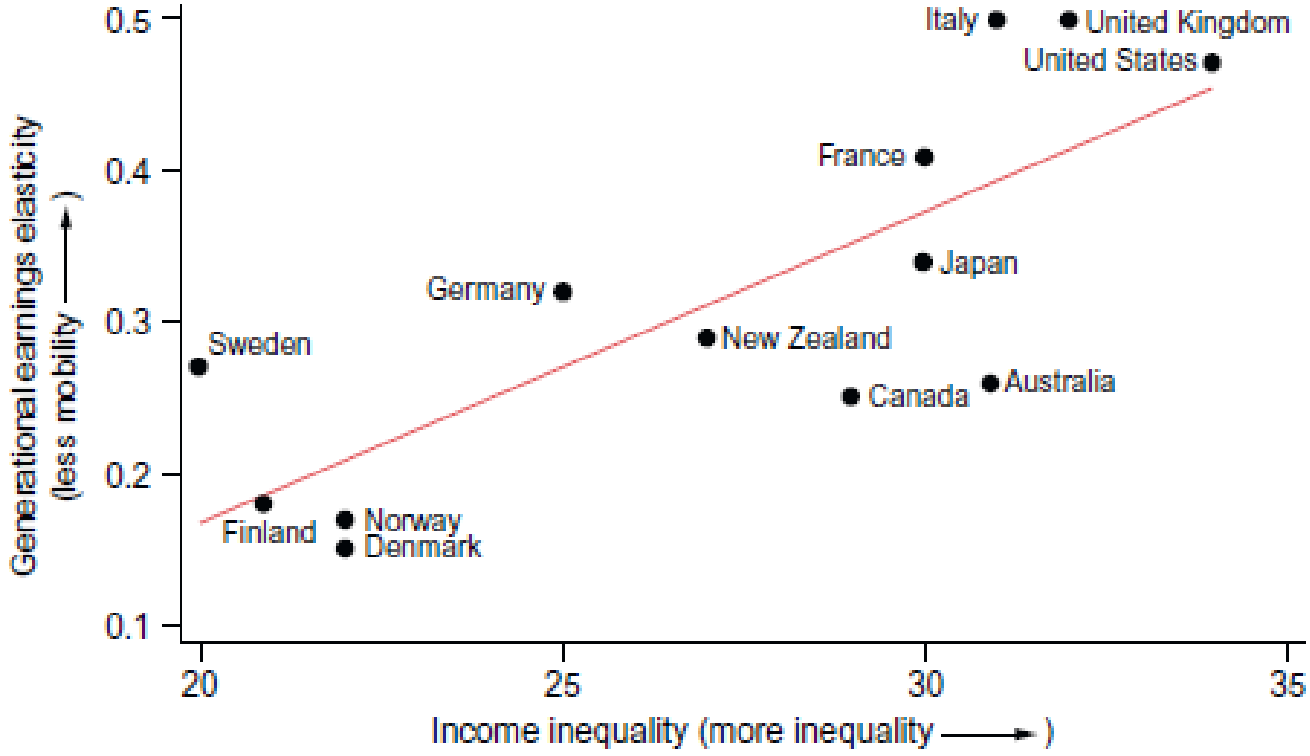
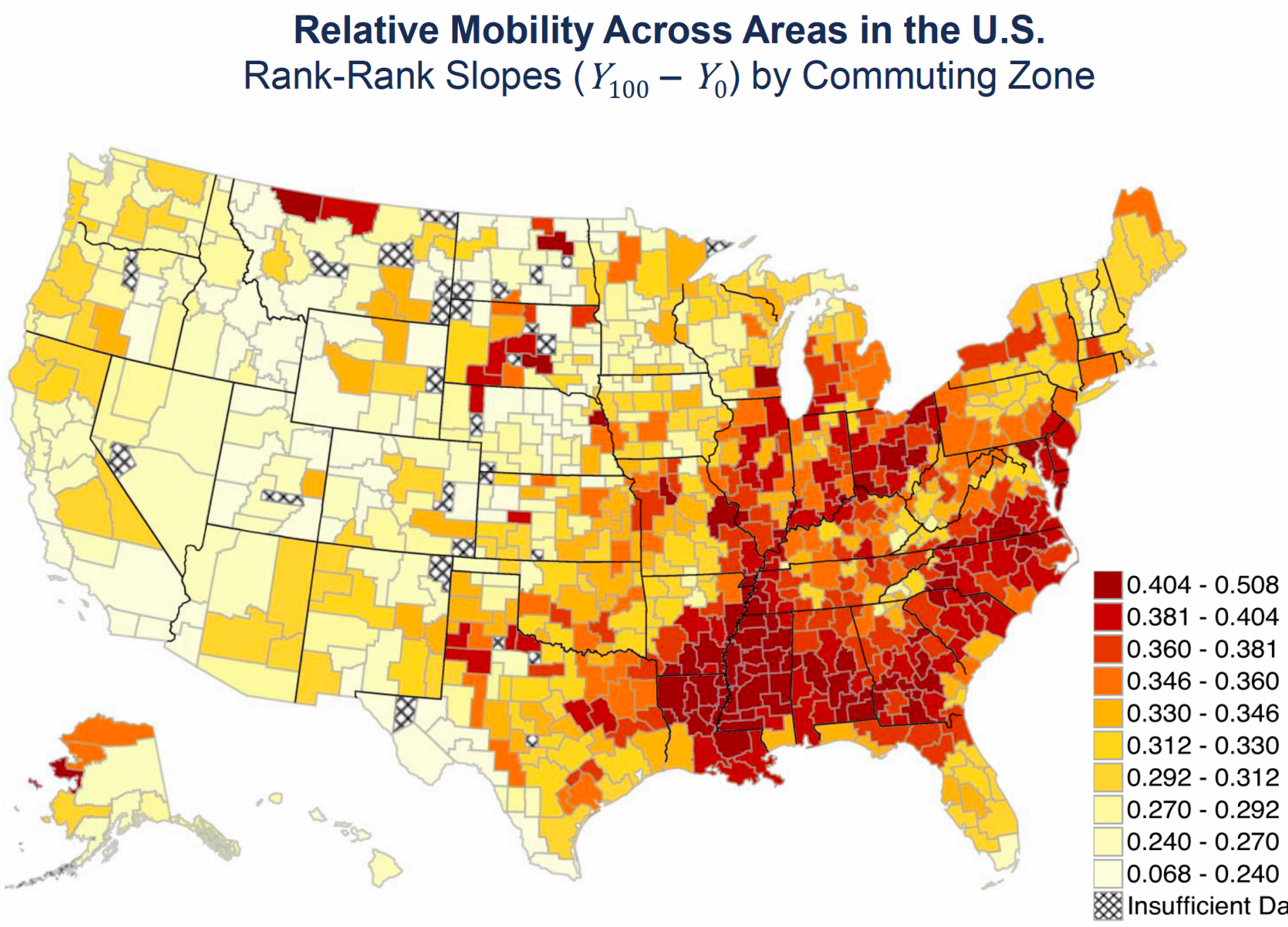


Figure 10.13 The Great Gatsby curve: the relationship between intergenerational earnings persistence and cross-sectional income inequality. *Note: Income inequality is measured by the Gini coefficient of disposable household income in 1985 taken from the OECD. Persistence is measured as the Beta of parental and son earnings. Sons are born in early 1960s, and outcomes for them are measured in late 1990s. See Corak (2013a,b) for further detail. Source: Corak (2013a, Figure 1).*

b within the US



Intergenerational Correlations *look* like they are rising (e.g. US):

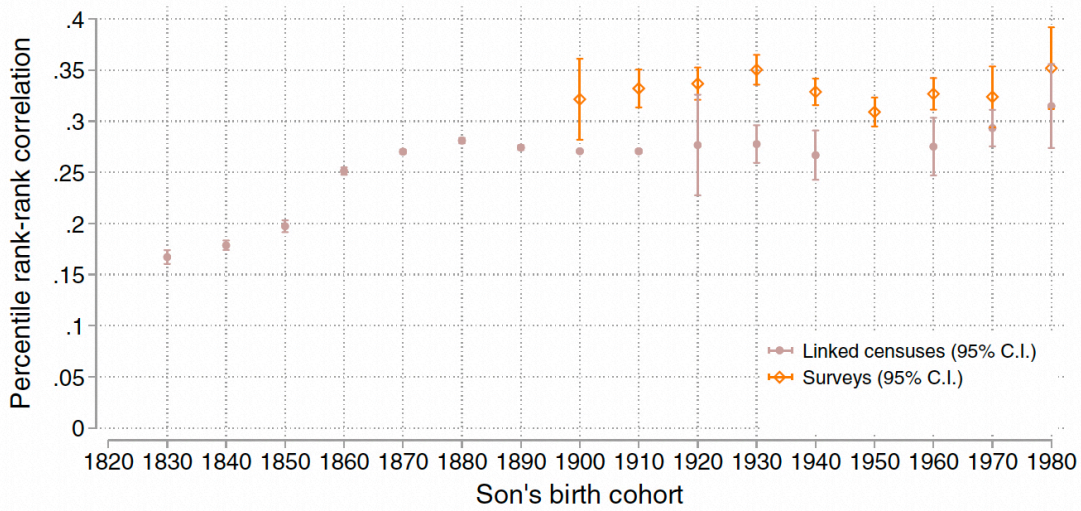


Fig. 2. Trends in intergenerational rank–rank correlation over son's birth cohort. Data sources: Linked censuses include linked historical census data 1850–1880, 1880–1910, 1910–1940; linked contemporary data using census 1940 to CPS (1973, 1979, 1980–1990) and CPS to census 2000/ACS 2001–2015. Survey data include General Social Surveys 1972–2016; Health and Retirement Study 1992–2010; National Longitudinal Survey–Older Men 1966–1990; National Longitudinal Survey–Young Men 1966–1981; National Longitudinal Survey of Youth 79, 1979–2012; National Survey of Families and Households, 1987, 1993, 2002; Occupational Changes in a Generation I and II; Panel Study of Income Dynamics (Survey Research Center [SRC] sample), 1968–2015; Survey of Income and Program Participation, 1986, 1987, 1988; Wisconsin Longitudinal Study, 1957–2011. Notes: The figure plots the estimated correlation in occupational percentile ranks between fathers and sons over 16 birth cohorts of sons. The series of solid circles plots the point estimates of the rank–rank correlations between fathers and sons for the 1830–1980 birth cohorts of sons. The 1950 birth cohort (sons born between 1946 and 1955) is missing because these individuals were not observed during their childhoods in either the linked historical censuses or the contemporary linked CPS-Census/ACS data, which is necessary to make the father–son link. The series in diamond symbols plots the rank–rank correlations estimated from weighted OLS regressions for the 1900–1980 birth cohorts from pooled contemporary social surveys. The weights are constructed from the original sampling probability weight variable in each survey as well as an additional weight variable that takes into account variation in sample size by birth cohort across surveys. The capped spikes refer to 95% confidence intervals of the correlation estimates. Because of the smaller sample size of the survey data relative to the linked census data, the survey estimates show a higher level of uncertainty reflected in wider confidence intervals. See *SI Appendix*, Table S10 for the exact numbers of the estimates.

Source: Long-term decline in intergenerational mobility in the United States since the 1850s

The Problem

- Almost all existing measures of social mobility are based on observational correlations
- They miss the deeper dynamics
  - They mis-estimate social mobility
  - They are fundamentally flawed
- We can show this using surname based estimators for England over the past millennium

Outline

Surnames

- **Surnames** are hereditary cultural labels that typically denote an individual's membership of a genetically related family.
  - In England, surnames originated amongst the aristocracy
- The more unusual the surname, the more likely holders are close genetic relatives: 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))
  - Modern forensics has famously solved decades old 'cold-cases' exploiting the new art of genetic genealogy (Gymrek et al. (2013))

Surnames and Social Mobility

- It is hard to get individual level data but surnames provide a shortcut
  - Surnames link us to previous generations though the patriline – in England we can link some people alive now to their ancestors in 1086
- With the high rates of social mobility typically found [e.g, .2-.4], common surnames should rapidly lose status information
  - We can define rare surnames by their frequency in elite surname lists

Oxbridge as a measure of Elite Status

Our First Example: 'Oxbridge' Type names

Table 2 Examples of rare Oxbridge versus non-Oxbridge surnames, 1800–1829

Oxbridge		Non-Oxbridge	
Agassiz	Brickdale	Agnerv	Bodgett
Anquetil	Brooshooft	Allbert	Boolman
Atthill	Bunduck	Arfman	Bradsey
Baitson	Buttanshaw	Bainchley	Breckill
Barnardiston	Cantis	Bante	Callaly
Bazalgette	Casamajor	Barthorn	Capildi
Belfour	Chabot	Bavey	Carville
Beridge	Charretie	Bedborne	Cavet
Bleeck	Cheslyn	Bemond	Chanterfield
Boinville	Clarina	Berrton	Chesslow
Boscawen	Coham	Bideford	Chubham
Bramston	Conyngham	Bisace	Clemishaw

Oxbridge Elite Surnames, 1830-2012

Table 3 Calculating intergenerational correlation for the rare surnames (500 or fewer in the 1881 census)

Period	Share of Oxbridge attendees (English surnames)	Share of population	Relative representation	Oxbridge elite (%)	Implied mean status	Implied *b*
1830–1859	11.86	1.18	10.04	0.62	1.05	–
1860–1889	8.18	1.15	7.11	0.53	0.76	0.72
1890–1919	5.23	1.11	4.72	0.48	0.58	0.76
1920–1949	3.24	1.06	3.06	0.70	0.43	0.75
1950–1979	1.96	1.01	1.94	1.16	0.26	0.60
1980–2009	1.38	0.86	1.60	1.27	0.19	0.72
2010–2012	1.42	0.86	1.65	1.19	0.20	1.12

The Slow Decline of Oxbridge Elites, 1830-2012

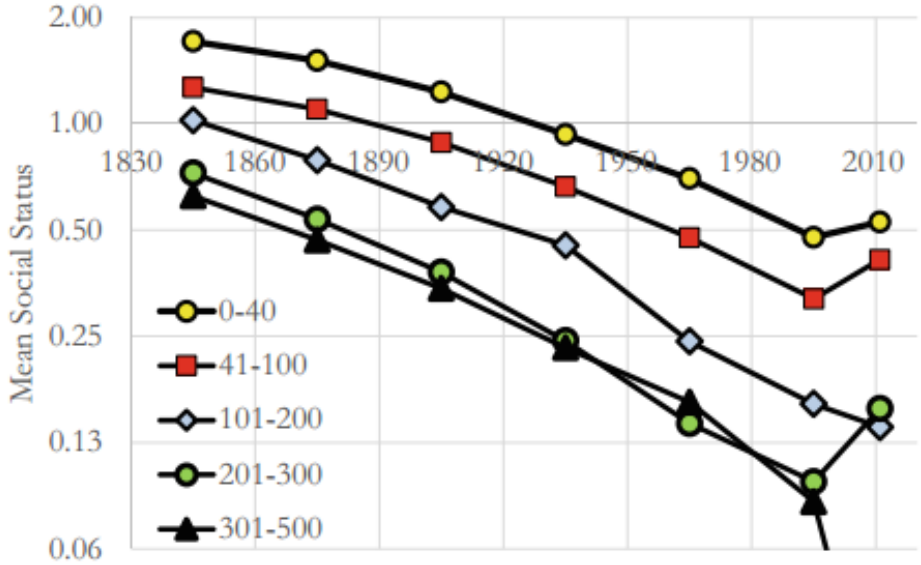


Fig. 4 Estimated mean status of surname groupings, Oxbridge, 1830–2012. Status is measured in terms of where (how many standard deviations above the mean) each group lies in terms of social status. The slope of each line measures social mobility rates

b estimates, 1730-2009

Table 4 *b* estimates, 1830–2009

Group*	Surname holders, 1881	1830–2009 *b*	1830–2009 *b* 5%CI	Relative population share, 2010 vs. 1880
High status				
0–40	12,948	0.77	0.74–0.81	0.61
41–100	7,838	0.76	0.72–0.79	0.60
101–200	8,050	0.69	0.65–0.73	0.76
201–300	11,703	0.66	0.64–0.69	0.72
301–500	136,925	0.68	0.64–0.72	0.81
0–500	177,464	0.72	0.69–0.75	0.78
Low status				
2001–5000	501,773	0.62	0.57–0.66	0.82

* Based on the frequency of surnames in the 1881 census

Regression to the Mean of Low Status Names

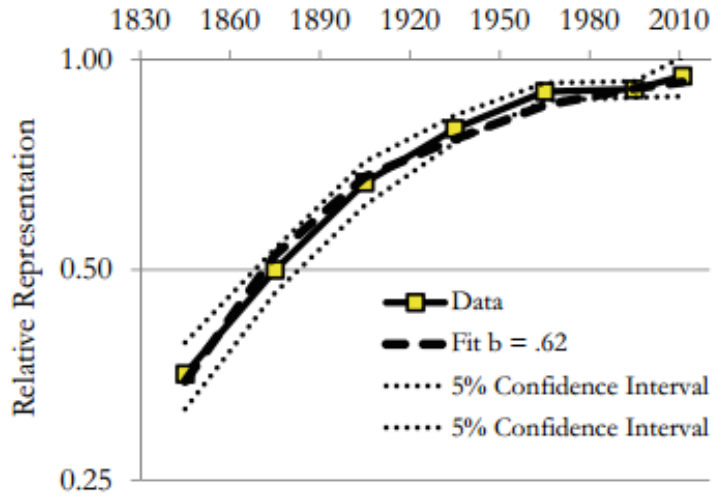


Fig. 5 Regression to the mean of low-status surnames, 1830–2012. Groups of surnames that appear at low rates at Oxbridge rise in status over time in a process that is close to a mirror image of the high-status groups. Both high- and low-status groups regress to the mean at slow rates

English Surname Types

- In the 12th century, surnames were a recent invention
- Examples of Surname Types:
  - *Patroynimic*: Johnson, ..son etc.
  - *Local*: Hill, Wood etc. (local topography)
  - *Occupational*: Smith(e), Taylor, Butcher...
  - *Nickname*: Long, Short, Beardsly, Stout...

Types of Elite Surnames looked for at Oxbridge

- *Norman*: Domesday book (1086) records the names of the Norman conquerors of England.
  - E.g. Baskerville, Darcy, Mandeville, Montgomery, Neville, Percy, Punchard, and Talbot
- *Inquisitions Post Mortem* of 1236-1299. Listed high status landholders. E.g. : Berkeley, Pakenham
- *Locative* Surnames: E.g. Atherton, Puttenham, Beveridge

How we measure b: Relative Representation

$$RelativeRepresentation = \frac{shareElite}{sharePop}$$

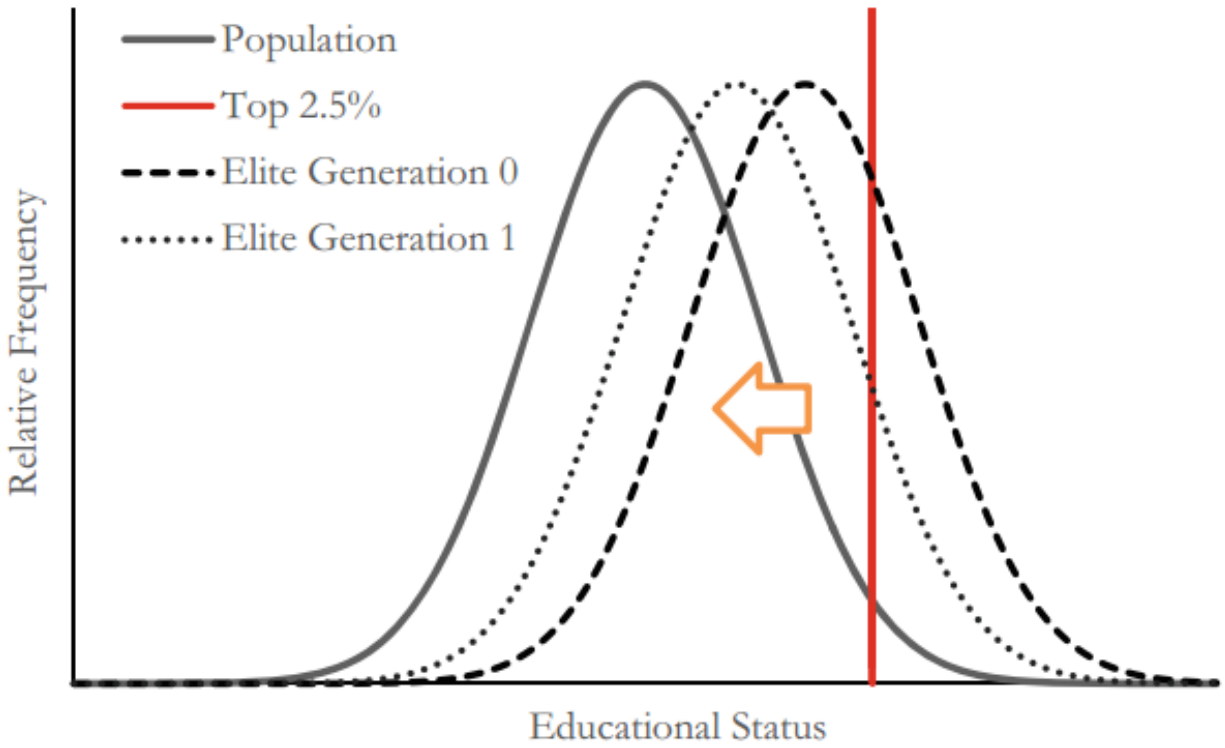


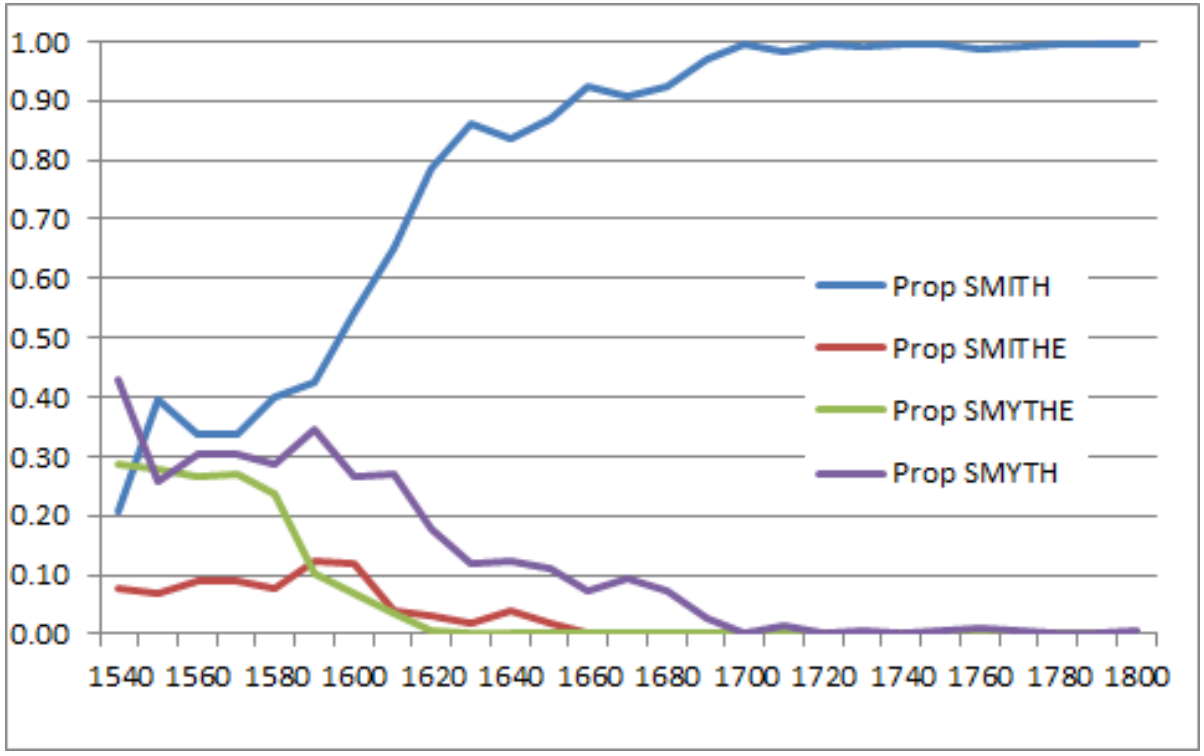
Fig. 1 Regression to the mean of surname status. The strength of the intergenerational correlation in status, *h*
can be measured by the speed of decline of the overrepresentation of initial elite surnames among social elite

N over time

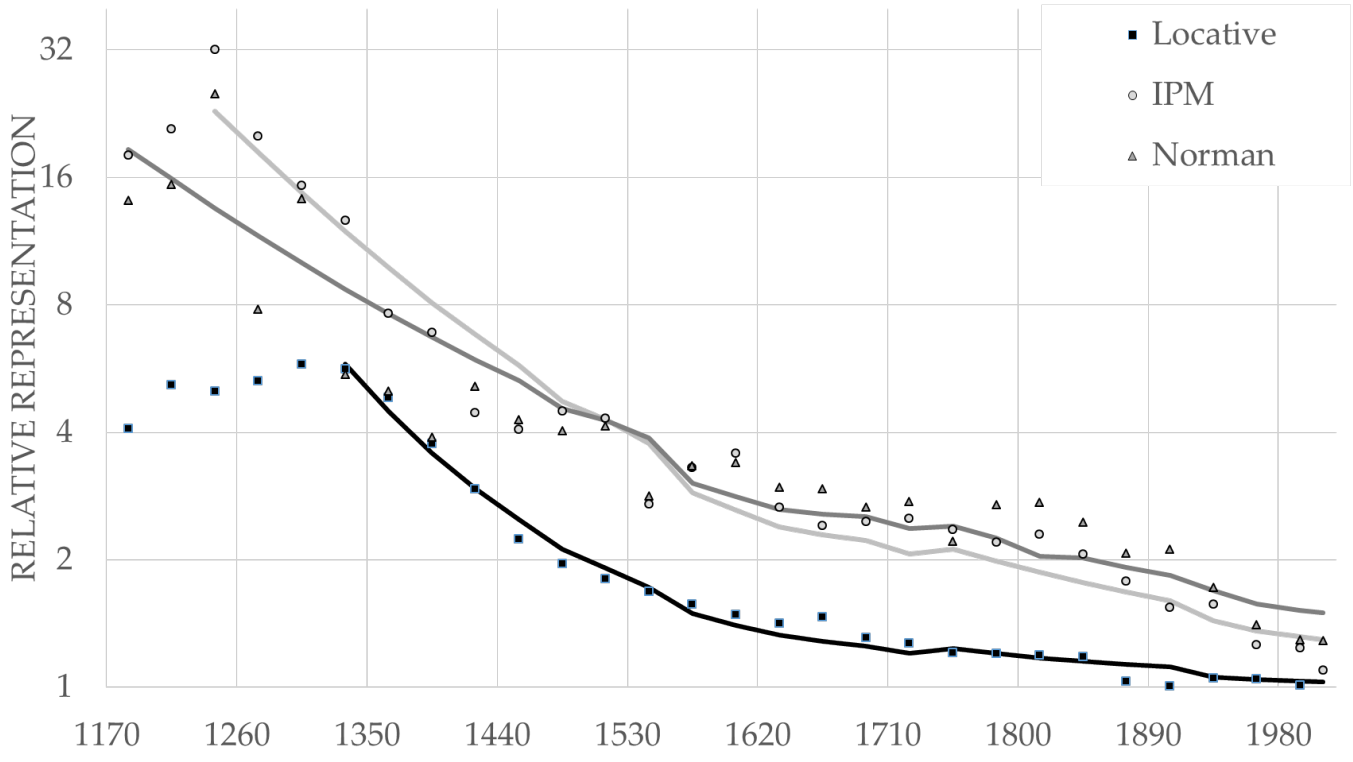
Table 1 Surnames at Oxbridge, 1170–2012

Generation	Oxbridge students observed	Estimated total Oxbridge students	Assumed domestic share	Population students drawn from	Oxbridge cohort share (%)
1170–1199	107	–	1.00	–	0.80
1200–1229	260	7,510	1.00	853,400	0.80
1230–1259	386	8,742	1.00	993,407	0.80
1260–1289	787	9,514	1.00	1,081,095	0.80
1290–1319	1,317	11,934	1.00	1,356,162	0.80
1320–1349	2,284	12,590	1.00	1,430,674	0.80
1350–1379	1,746	9,991	1.00	1,135,318	0.80
1380–1409	3,332	7,241	1.00	822,842	0.80
1410–1439	2,115	6,333	1.00	719,703	0.80
1440–1469	5,454	5,744	1.00	652,724	0.80
1470–1499	6,146	6,146	1.00	628,280	0.89
1500–1529	5,684	5,684	1.00	654,964	0.79
1530–1559	6,477	6,477	1.00	789,152	0.71
1560–1589	19,349	19,349	1.00	849,960	2.01
1590–1619	22,327	22,327	1.00	1,009,277	2.06
1620–1649	24,232	24,232	1.00	1,273,656	1.85
1650–1679	23,908	23,908	1.00	1,462,187	1.75
1680–1709	17,042	17,042	1.00	1,479,698	1.13
1710–1739	16,021	16,021	1.00	1,492,885	1.00
1740–1769	10,519	10,519	1.00	1,583,707	0.61
1770–1799	11,994	11,994	0.99	1,793,974	0.55
1800–1829	18,649	18,649	0.99	2,246,609	0.64
1830–1859	24,415	24,415	0.99	3,245,746	0.62
1860–1889	38,678	38,678	0.96	7,085,936	0.53
1890–1919	30,962	47,526	0.93	9,265,992	0.48
1920–1949	67,927	92,854	0.88	11,589,095	0.70
1950–1979	156,645	192,254	0.86	14,209,853	1.16

Names Standardize over Time



Relative Representation at Oxbridge 1170-2012



What does this mean?

- There is regression to the mean: $b < 1$
- The rate of regression to the mean seems constant
- Social mobility is slow
- Social mobility rates appear similar for modern and feudal England
- The IR, democracy and 20th century education policies have done nothing to speed the process up

Outline

Rare Surnames Indicate Status

Table 2

The Rare Surname Groups, 1858–87

Rich	Prosperous	Poor
Ahmuty	Agace	Adson
Angerstein	Agar-Ellis	Aller
Appold	Aglen	Almand
Auriol	Allecock	Angler
Bailward	Aloof	Anglim
Basevi	Alsager	Annings
Bazalgette	Bagnold	Austell
Beague	Beridge	Backlake
Benthall	Berthon	Bagwill
Berens	Brettingham	Balsden
Berners	Brideoake	Banbrook
Bigge	Broadmead	Bantham
Blegborough	Broderip	Bawson
Blicke	Brouncker	Beetchenow
Boger	Brune	Bemmer

THE ECONOMIC JOURNAL

The Economic Journal, **125** (*February*), 61–85. Doi: 10.1111/ecoj.12165 Published by John Wiley & Sons, 9600 Garsington Road, Oxford OX4 2DQ, UK and 350 Main Street, Malden, MA 02148, USA.

INTERGENERATIONAL WEALTH MOBILITY IN ENGLAND, 1858–2012: SURNAMES AND SOCIAL MOBILITY*

Gregory Clark and Neil Cummins

This article uses a panel of 18,869 people with rare surnames whose wealth is observed at death in England and Wales 1858–2012 to measure the intergenerational elasticity of wealth over five generations. We show, using rare surnames to track families, that wealth is much more persistent than standard one generation estimates would suggest. There is still a significant correlation between the wealth of families five generations apart. We show that this finding can be reconciled with standard estimates of wealth mobility by positing an underlying first order Markov process of wealth inheritance with an intergenerational elasticity of 0.70–0.75 throughout the years 1858–2012.

The Process for that paper

1. RQ: What is the correlation of status across generations in Britain
2. New Idea: How about averaging over rare surnames
3. Use 1881 census to count names -rare -> N=5-20
4. Get all deaths for these names from `freebmd.com`
5. Look for these people in RCJ PPR calendars (pen + paper, then entered into excel)
6. Fill in their wealth
7. Analyze in Stata

Individual Surname Fortunes and Surname Averages

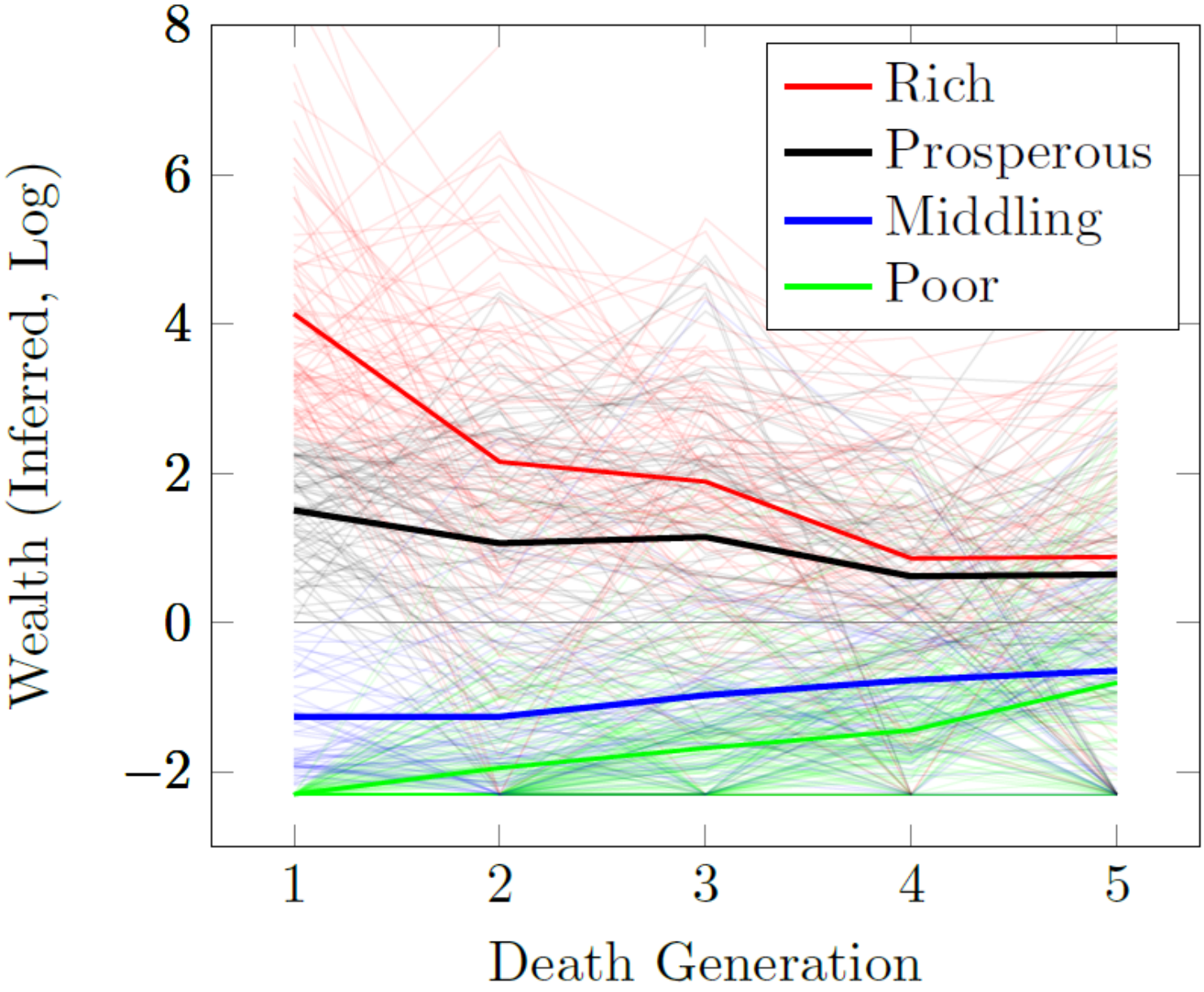


Figure: Regression to the Mean in Wealth across 5 gens

Regression Towards the Mean

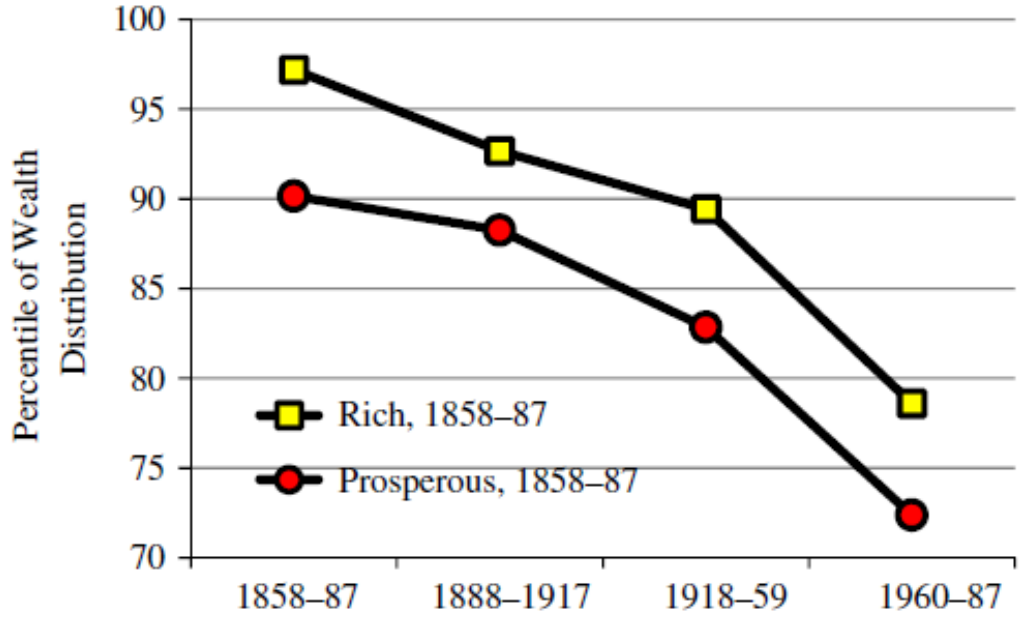


Fig. 2. *Location by Wealth Percentiles, Surname Types, by Generation*
Note. See Appendix for details of this calculation.

Summary Stats

Table 3

Summary of the Surname Samples

Period	Surnames	Probates	Deaths 21+
Rich and prosperous			
1858–87	182	1,137	1,661*
1888–1917	180	1,070	1,785
1918–59	182	1,962	2,874
1960–93	176	1,115	1,769
1994–2012	170	526	890
Poor			
1858–87	226	19	1,054*
1888–1917	199	151	1,327
1918–59	201	571	2,364
1960–93	201	875	2,635
1994–2012	164	548	1,393

Notes. All surnames were held by 40 or fewer people in the 1881 census. Deaths are from the General Registry Office (See References Section). *Where age was unknown 1858–65 (97 and 363 cases), the fraction above 21 was estimated from the 1866–87 ratio of deaths 21+ to all deaths.

Table 4

Proportion Probated by Surname Group

Period	Average birth year	Rich	Prosperous	Poor	All deaths
1858–87	1814	0.83	0.56	0.02	0.15
1888–1917	1843	0.66	0.54	0.11	0.22
1918–59	1872	0.73	0.63	0.24	0.40
1960–93	1900	0.66	0.61	0.33	0.46
1994–2012	1926	0.62	0.57	0.39	0.43

Regression Towards the Mean: Probate Rates

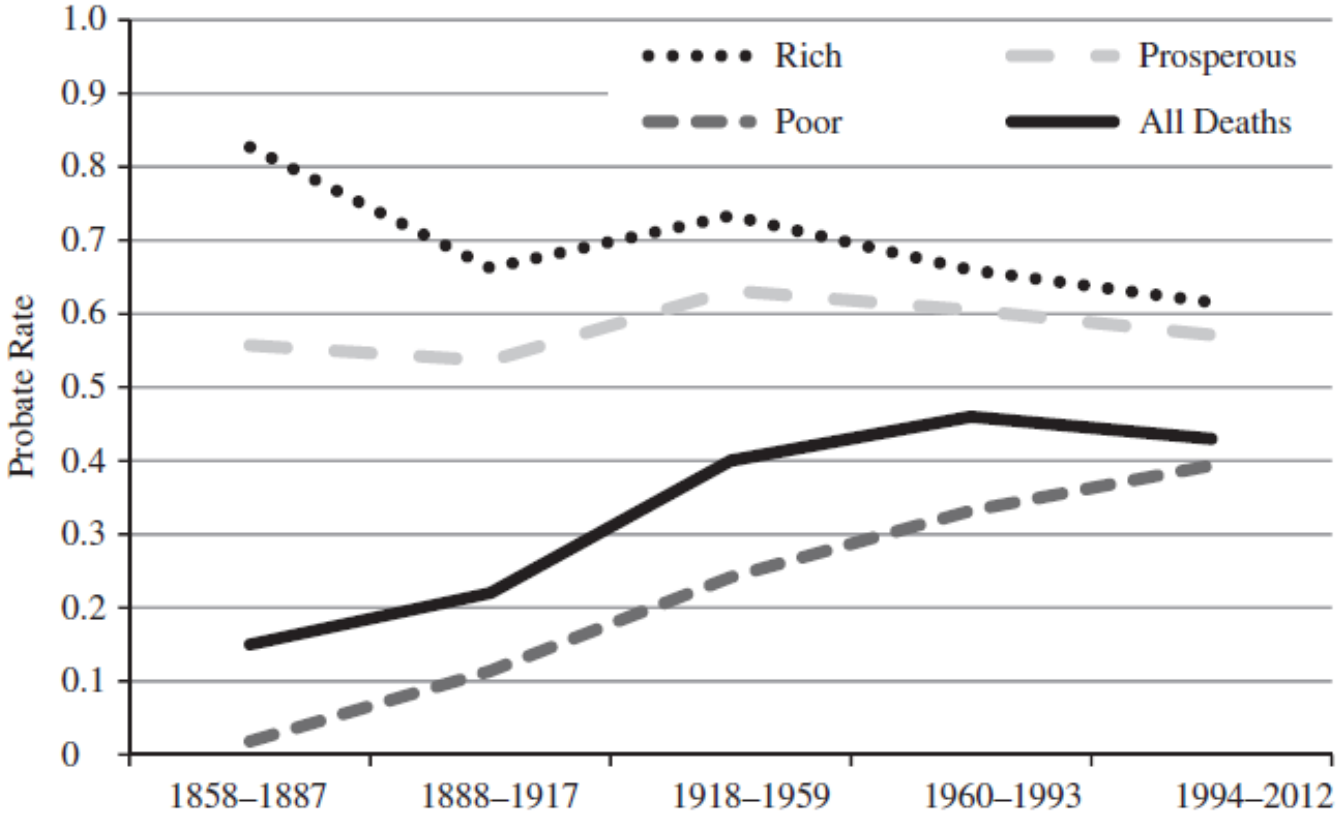


Fig. 3. *Probate Rates of Surname Types, by Generation*

Notes. The probate rate in a given generation is the number of people recorded in the probate registry divided by the number of people dying.

Source. Principal Probate Registry and GRO.

Regression Towards the Mean: Average Wealth

Table 5

Average ln Wealth, All Adult Deaths, by Death Generations

Generation	Rich	Prosperous	Poorer
1858–87	5.20	3.05	−0.64
1888–1917	3.32	2.49	−0.43
1918–52	2.29	1.68	−0.43
1953–87	1.69	1.19	−0.10
1999–2012	1.36	1.03	−0.11

Notes. Wealth is measured relative to estimated average wealth. Those not probated are assigned an imputed wealth as described in the text. The years 1988–98 are omitted for the reasons described in the text.

Regression Towards the Mean: Total Wealth

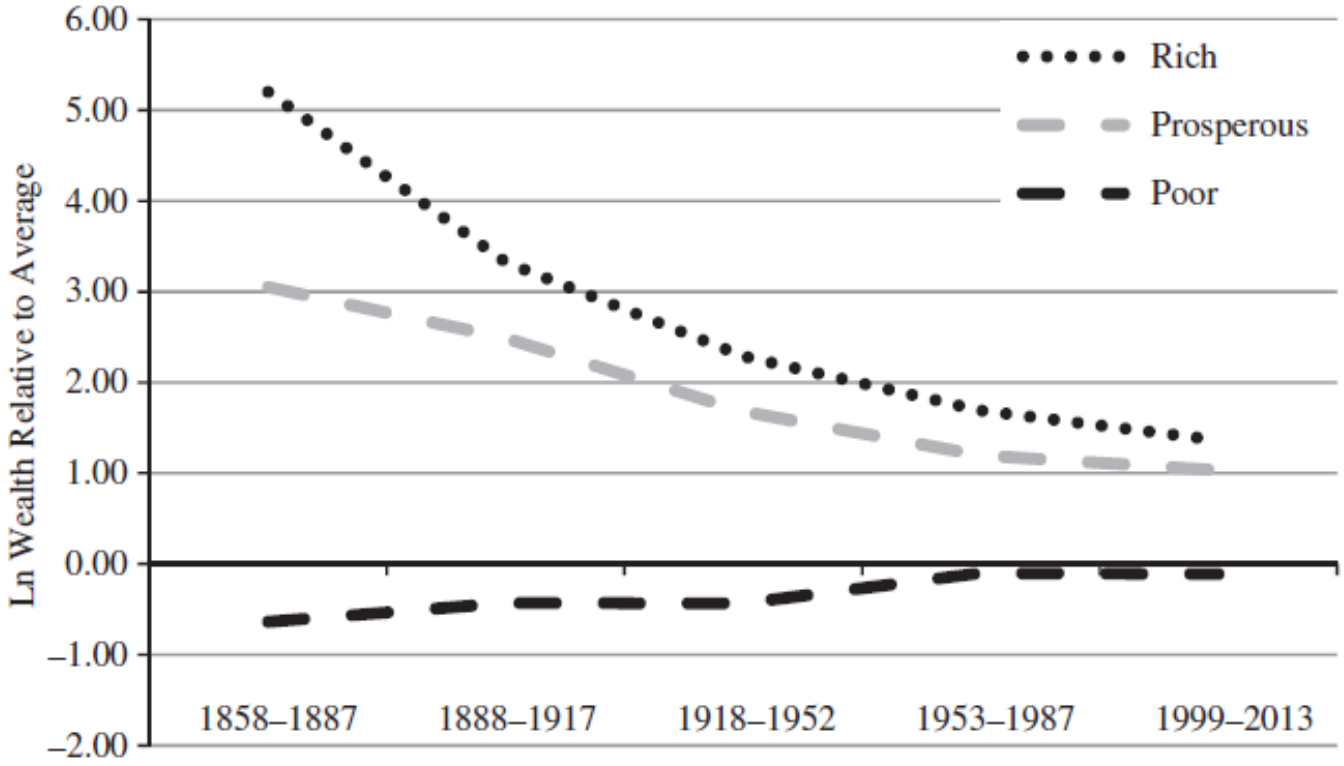


Fig. 5. *Average Log Probate Value, Including Those Not Probated, by Generation*
Source. Table 5.

Estimating b via Individual Links

Table 6

Conventional Estimates of Intergenerational Wealth Elasticities, 1858–2012, Individual Families

Period of child death	All N	All $\hat{\beta}$	Father probated, N	Father probated $\hat{\beta}$
1858–87	237	0.376	160	0.483
		(0.063)		(0.134)
1888–1917	902	0.494	581	0.672
		(0.028)		(0.057)
1918–59	2,109	0.389	1,230	0.473
		(0.017)		(0.036)
1960–87	1,126	0.383	567	0.370
		(0.023)		(0.039)
1999–2012	449	0.419	207	0.539
		(0.055)		(0.086)

Note. Robust standard errors in parentheses.

Estimaing b via Surname Averages

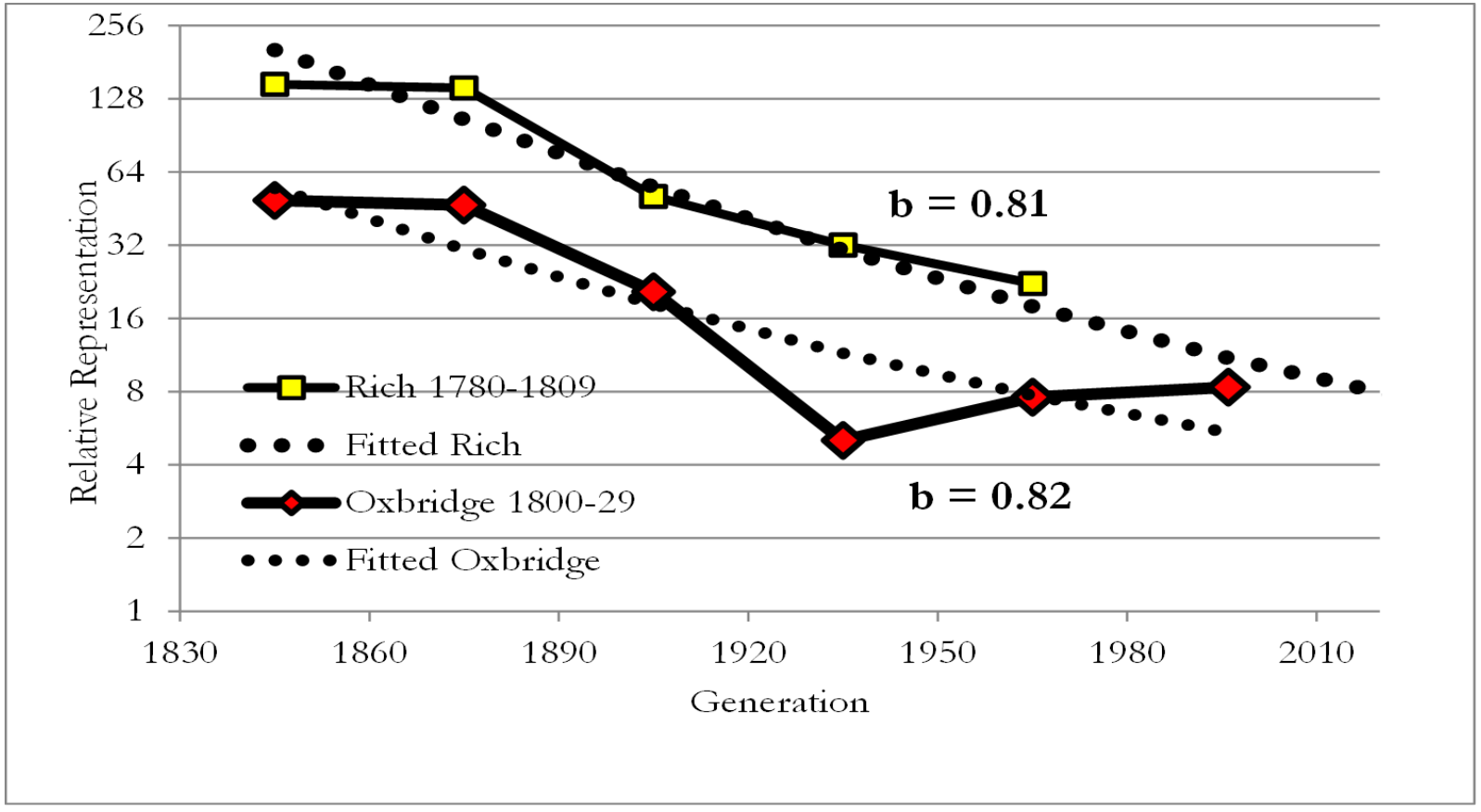
Table 7

Wealth Elasticities Between Death Generations, Surname Groups

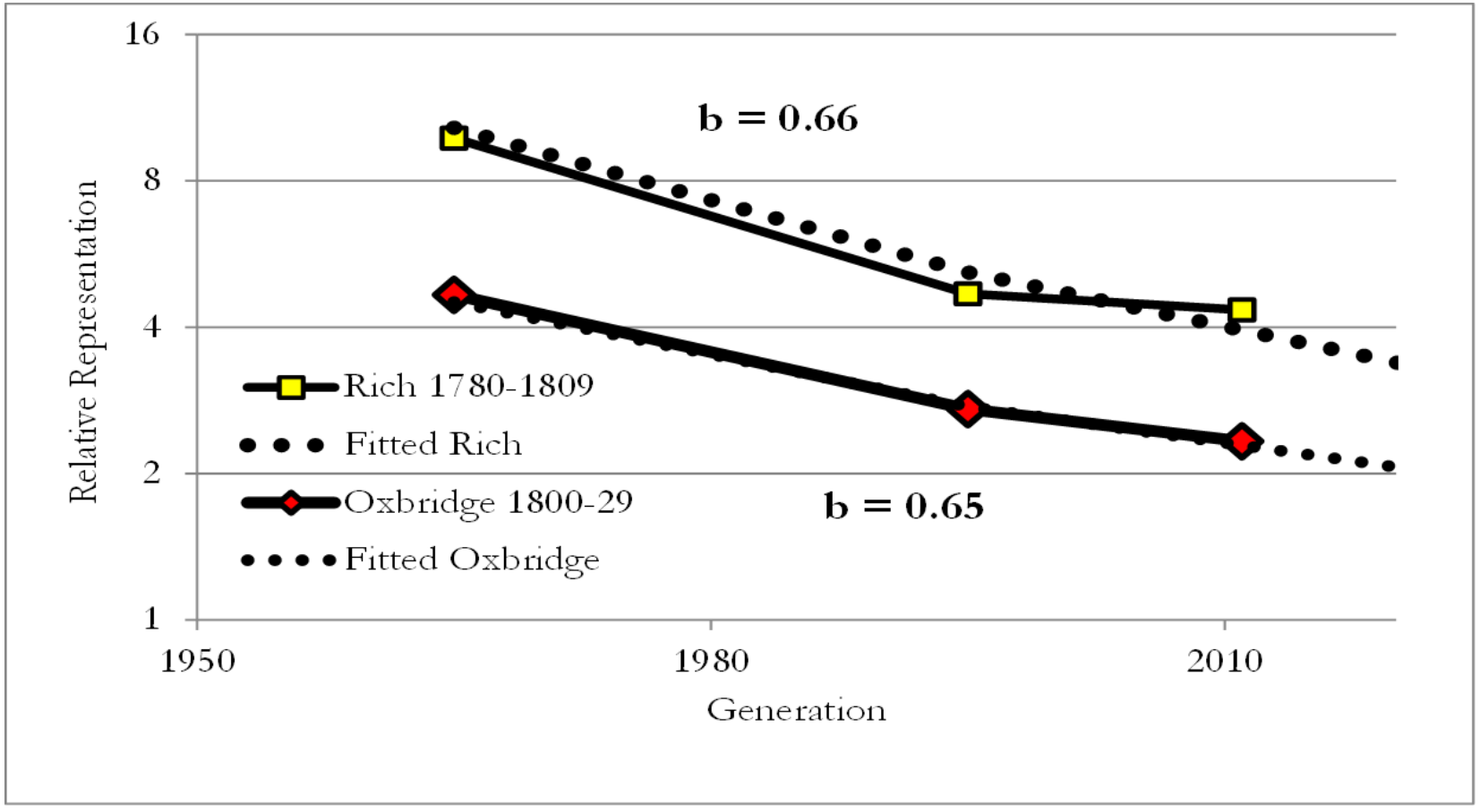
Year of death	Rich	Prosperous	Rich/prosperous	Poor
1888–1917	0.64	0.81	0.71	0.67
	(0.027)	(0.051)	(0.026)	(0.060)
1918–59	0.69	0.68	0.68	1.01
	(0.043)	(0.042)	(0.028)	(0.104)
1960–87	0.74	0.71	0.72	0.23
	(0.054)	(0.052)	(0.032)	(0.070)
1999–2012	0.81	0.87	0.83	1.28
	(0.095)	(0.126)	(0.077)	(3.698)
Average	0.71	0.76	0.73	0.60
	(0.019)	(0.027)	(0.016)	(0.162)

Note. Bootstrapped standard errors in parentheses.

Relative Representation in the House of Commons, 1830-2012



Relative Representation among Solicitors and Barristers, 1950-2012



In Sum

- Social mobility rates appear constant over time in England
- Conventional estimates overestimate mobility
- The World is substantially more socially rigid than we expected

And we can look at other places, globally

Country	Measure	Period	B
USA	Attorneys	1950-2011	0.67-0.77
USA	Doctors	1950-2011	0.73-0.74
England	Attorneys, Doctors	1950-2012	0.69-1.00
England	Wealth	1950-2012	0.70
England	Education	1950-2012	0.77
England	Education	1300-1500	0.75
Chile	Occupations	1940-2010	0.74
China	Education	1905-2011	0.71
Japan	Education, Occupations	1940-2012	0.76-0.84
India	Doctors	1860-2009	**0.89-1.00**

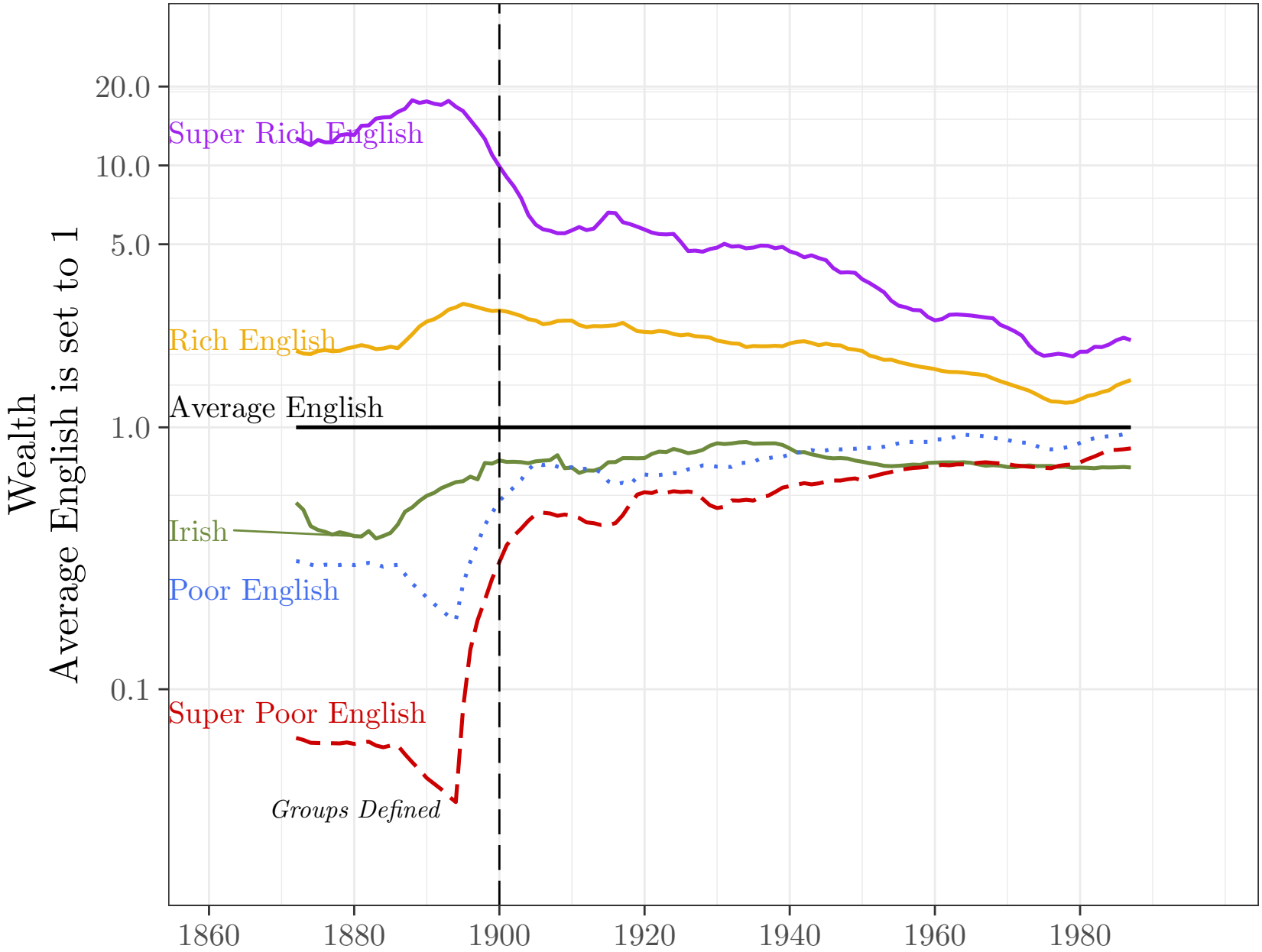


Figure: A Distinctive Irish Wealth Pattern

Outline

Outline

Why are Surname estimates of Mobility so different from Conventional estimates?

Clark-Cummins: $b = .8$
Everyone else: $b = .3 - .5$

- **These are wildly different and suggest different Worlds**

1. Life-course effects? - Comparing father and son's at different points
2. Income v wealth?
3. Unrepresentative Sample? *Becker's Dying words!*
4. Can you think of others?

A clue: Individual vs Group

Table 1: Individual and Group Level Correlations of Wealth across Generations, England, 1858-2012.

Period of child death	N	Father-Son Wealth Correlation	Correlation - Surname Groupings
1888-1917	902	0.49	0.71
		(0.028)	(0.026)
1918-59	2,109	0.39	0.68
		(0.017)	(0.028)
1960-87	1,126	0.38	0.72
		(0.023)	(0.032)
1999-2012	449	0.42	0.83
		(0.055)	(.077)

Note: Robust standard errors in parentheses.
Source: Clark and Cummins, 2015, table 6.

Figure: Individual vs. Group

People Trade off aspects of their latent status for lifestyle

Status is a Vector of Different but often Correlated Attributes

$$Status = f(Income, Wealth, Education, Occupation, Address, ...)$$

- One aspect of status will correlate more weakly than an average
  - Over say surnames/ surname groups
- People will perhaps trade off one aspect of status for another

The Latent Factor Model of Social Mobility

- Conventional

$$y_{t+1} = \beta y_t + u_t$$

  y is some aspect of status – income, wealth, years of schooling

- But

$$y_t = \tau X_t + e_t$$

- where X_t is an **underlying** status that the various y_t measure imperfectly

New Work

- Other scholars are now working with new methods to estimate underlying social mobility rates
  - For example Zhu for England 1851-1911
  - Sthuler for Germany
  - and Zach Ward for the US
- They all find an underlying b much greater than the observational correlation

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

Ziming Zhu

London School of Economics

z.zhu11@lse.ac.uk

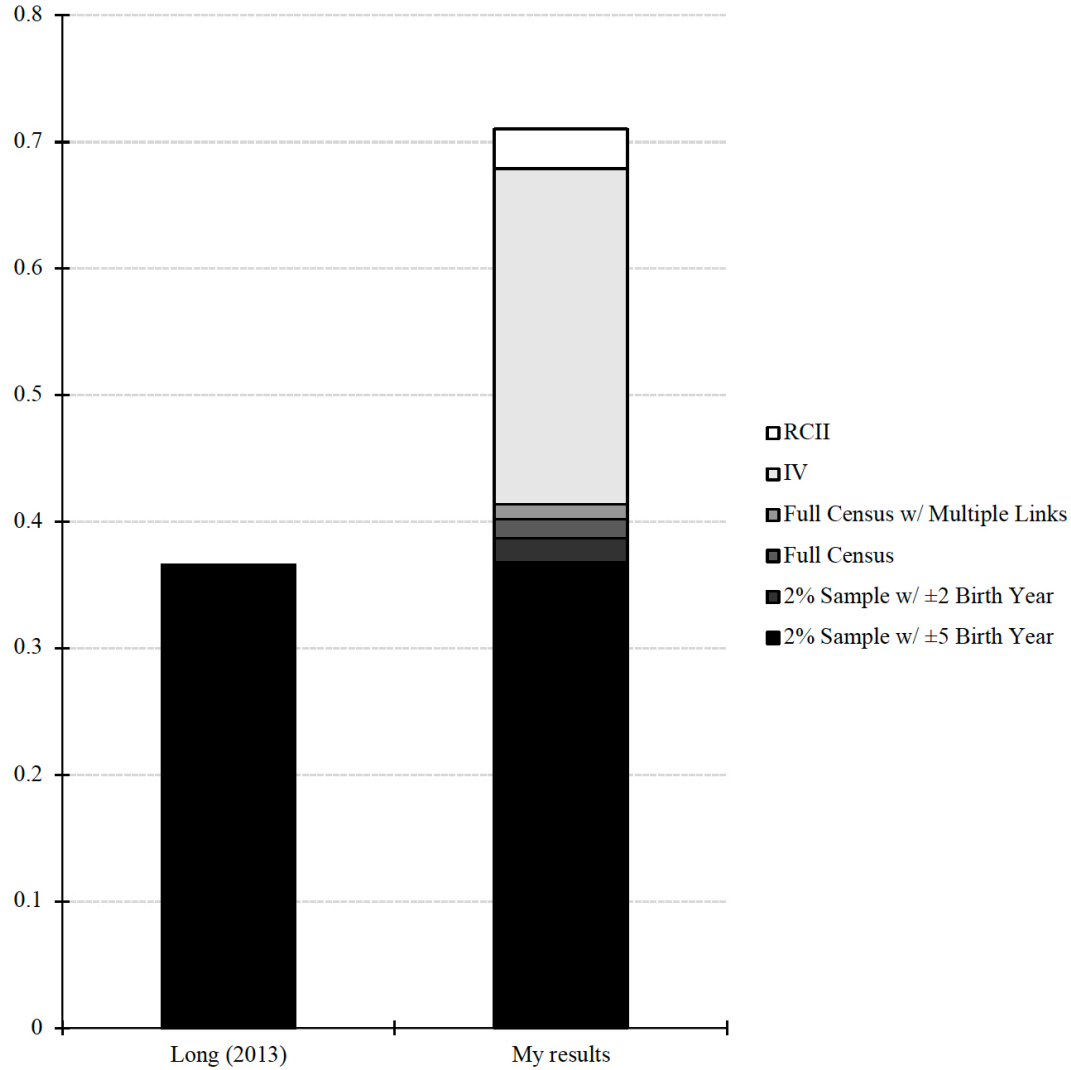
Abstract

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

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

JEL Classification: J62, N33

Figure 2: Comparison of my results for 1851-1881 with Long (2013)



Notes: '2% Sample w/ ± 5 Birth Year' refers to using a 2% sample of the 1851 census and allowing for birth year to differ by at most plus or minus 5 years – this is the approach taken with census linkage in Long (2013) which I have also replicated in my work; I use the same 2% sample that I have created through randomisation but with the further restriction of only allowing the birth year to differ by 2 years to produce the '2% Sample w/ ± 2 Birth Year' estimate; 'Full Census' estimate is taken from Table 2, Column 1; 'Full Census w/ Multiple Links' is taken from Table 2, Column 2; 'IV' estimate is taken from Table 2, Column 3; 'RCII' is the estimate obtained using both the IV strategy and the RCII scores instead of HISCAM. Sources: Long (2013) and author's analysis of I-CeM (UKDA, SN 7481) and I-CeM Names and Addresses (UKDA, SN7856).

Intergenerational Mobility in American History: Accounting for Race and Measurement Error*

Zachary Ward†

Baylor University

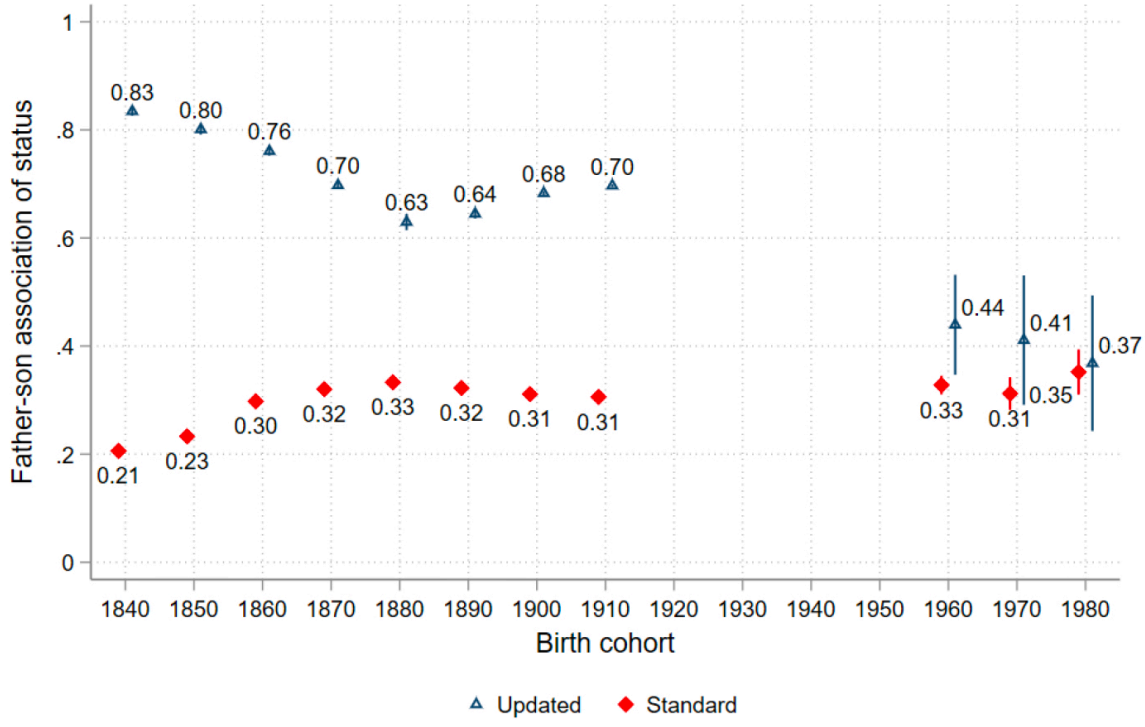
January 2023

Abstract: A large body of evidence finds that relative mobility in the US has declined over the past 150 years. However, long-run mobility estimates are usually based on white samples and therefore do not account for the limited opportunities available for non-white families. Moreover, historical data measure the father's status with error, which biases estimates toward greater mobility. Using linked census data from 1850-1940, I show that accounting for race and measurement error can double estimates of intergenerational persistence. Updated estimates imply that there is greater equality of opportunity today than in the past, mostly because opportunity was never that equal.

Keywords: intergenerational mobility, measurement error, persistence

JEL Codes: J62, N31, N32

Figure 1. Updated estimates of intergenerational mobility



Notes: The plotted estimates are the slope coefficient of the son's status (on a 0-100 scale) on the father's status. Estimates are presented by the son's birth cohort, rounded to the nearest decade. The "standard" estimates are based on data from Song et al. (2020) and reflect standard historical mobility estimates in the literature. For example, only white males are in the data, one occupation observation is assumed to capture permanent status, and within-occupation differences is status by race or region are ignored. The "updated" estimates make multiple changes to the standard estimates, which are described in Appendix D. The most important changes are: (1) Black families are pooled with white families, (2) measurement error is accounted for via instrumental variables, and (3) the status measure allows for within-occupational differences by race and region. Other differences in estimates include weighting and the linking method, but these differences are not as important for the long-run trend between 1840 and 1980 birth cohorts.

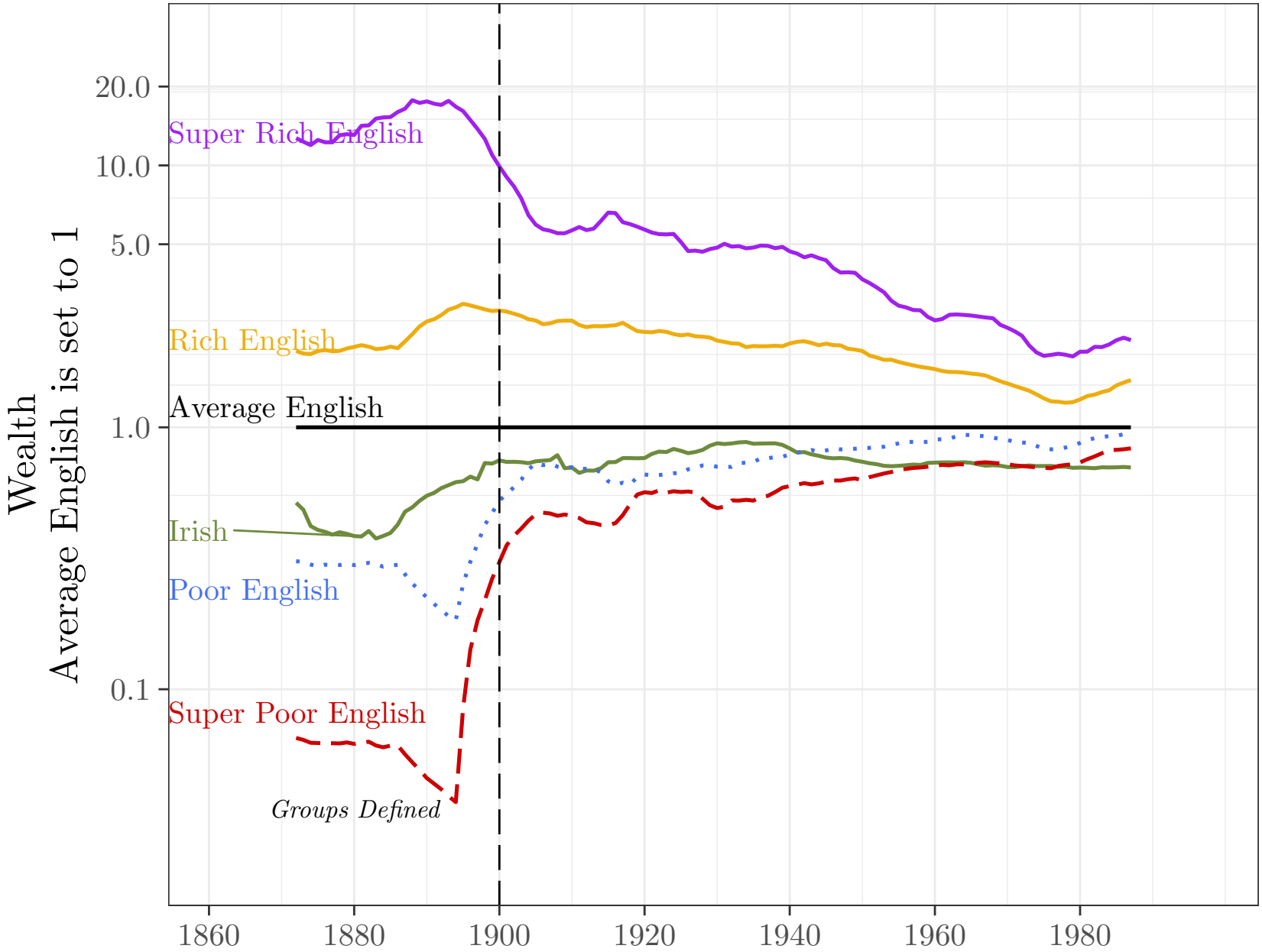
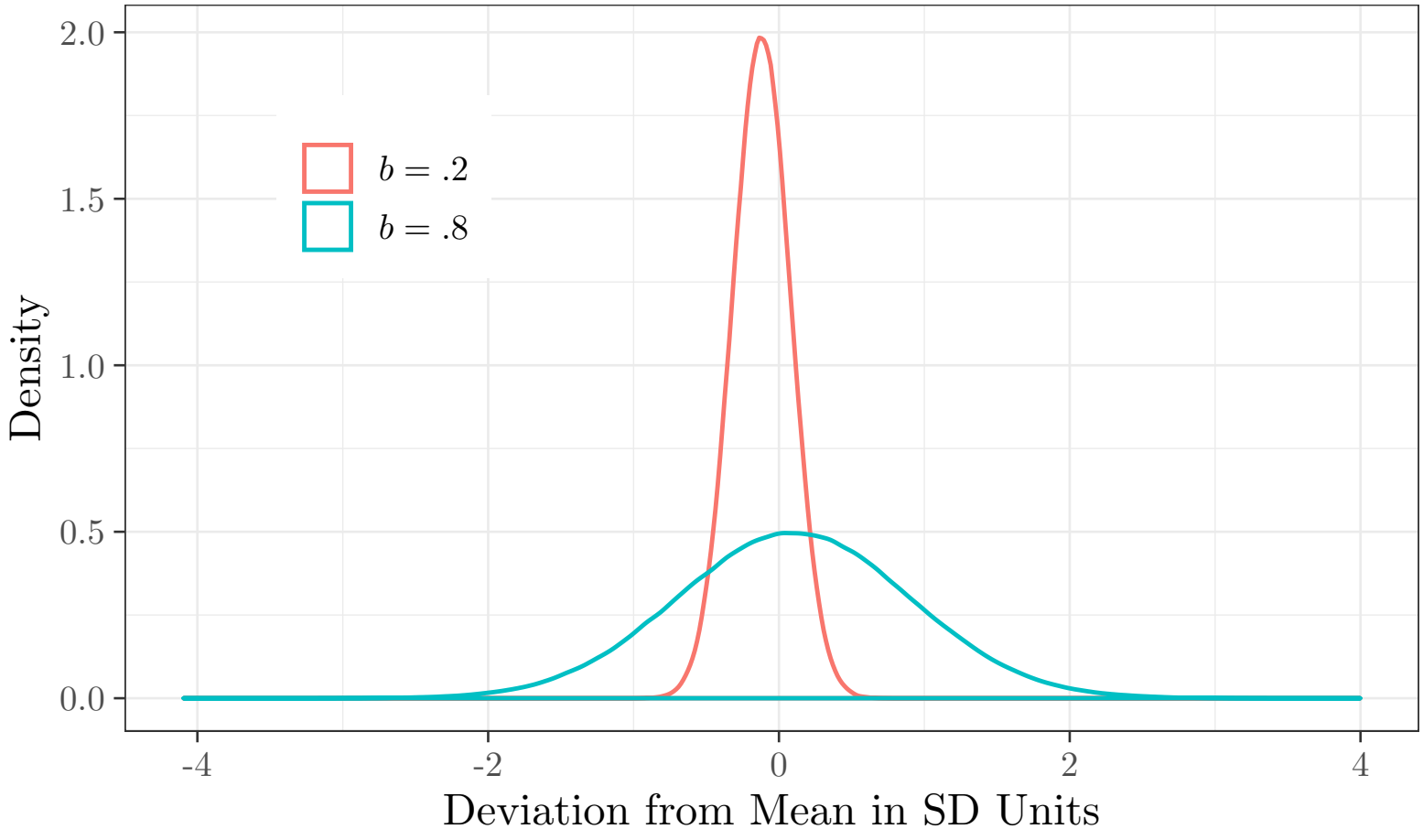


Figure: A Distinctive Irish Wealth Pattern

Inequality and Social Mobility: Simple Model Predicts ->

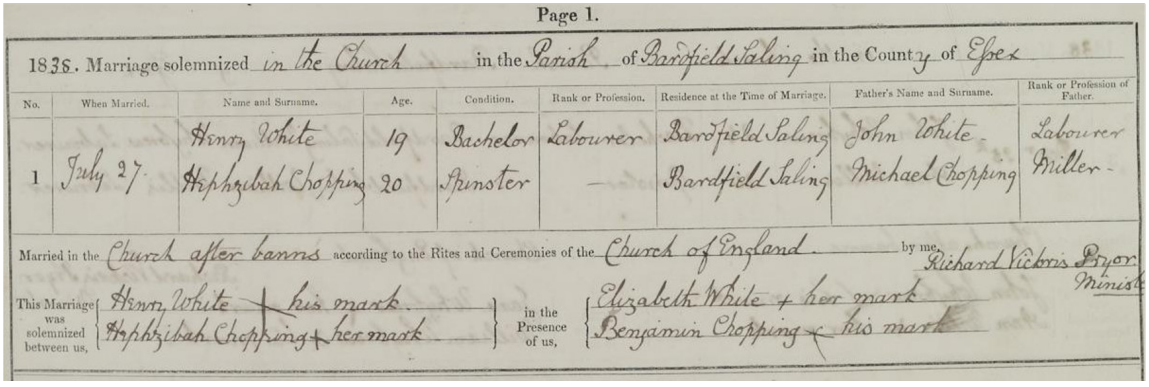


Outline

Hypergamy

- It is widely believed that women value social status in marital partners more than men, leading to female marital hypergamy, and more female intergenerational social mobility
  - For example Almas et al. 2023 for Norway

Church Marriage Registers, marriages 1837-2020



Page 1.

1838. Marriage solemnized in the Church in the Parish of Bardfield Saling in the County of Essex

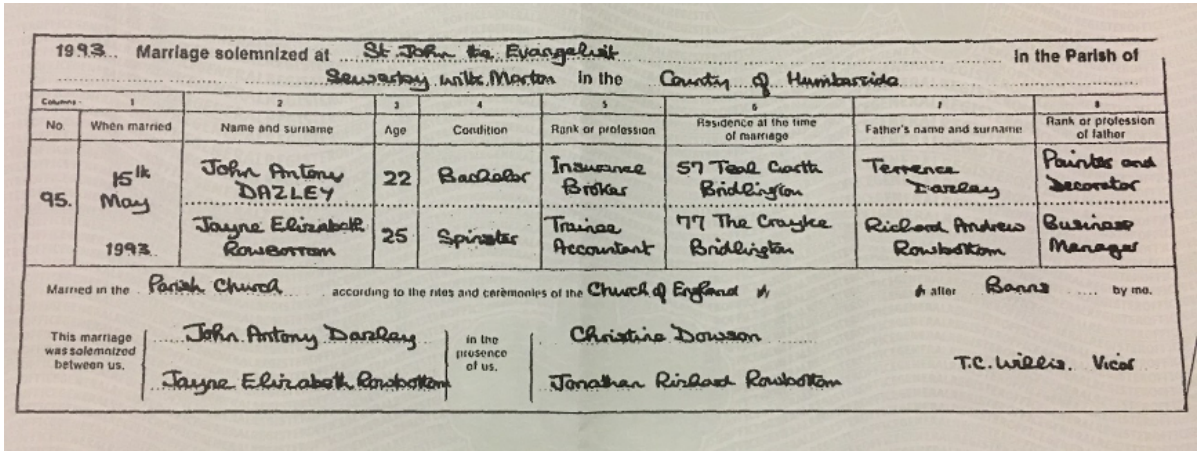
No.	When Married.	Name and Surname.	Age.	Condition.	Rank or Profession.	Residence at the Time of Marriage.	Father's Name and Surname.	Rank or Profession of Father.
1	July 27	Henry White	19	Bachelor	Labourer	Bardfield Saling	John White	Labourer
		Hephzibah Chopping	20	Spinster	—	Bardfield Saling	Michael Chopping	Miller

Married in the Church after banns according to the Rites and Ceremonies of the Church of England by me, Richard Victoris Pryor Minist

This Marriage was solemnized between us, Henry White his mark / Hephzibah Chopping her mark

in the Presence of us, Elizabeth White her mark / Benjamin Chopping his mark

Figure: Example of Marriage Certificate, 1838



1993. Marriage solemnized at St John the Evangelist Sewerby with Marton in the Parish of in the County of Humberside

Columns:	1	2	3	4	5	6	7	8
No.	When married	Name and surname	Age	Condition	Rank or profession	Residence at the time of marriage	Father's name and surname	Rank or profession of father
95.	15th May 1993	John Antony DAZLEY	22	Bachelor	Insurance Broker	57 Teal Croft Bridlington	Terrence Dazley	Painter and Decorator
		Jayne Elizabeth Rowbottom	25	Spinster	Trainee Accountant	77 The Crayke Bridlington	Richard Andrew Rowbottom	Business Manager

Married in the Parish Church according to the rites and ceremonies of the Church of England after Banns by me.

This marriage was solemnized between us, John Antony Dazley / Jayne Elizabeth Rowbottom

in the presence of us, Christine Dawson / Jonathan Richard Rowbottom

T.C. Willis. Vicar

Figure: Example of Marriage Certificate, 1993

And Big Data: Registers of Births, Marriages and Deaths, 1837-2007

47,082,406 marriages from 1837 to 2007:

- freebmd.com (1837-1980)
- famliysearch.org (1980-2007).
  - Here we attribute status to surnames using **wealth [Cummins 2020,2021]**

Example Occupations

Rank	FOE-Occupation	RCII	PCA
1	Titled	100.00	96.08
2	Esquire	97.67	69.71
3	Member Of Parliament	96.79	98.42
4	Bishop-Church Of England	93.47	97.46
5	General Army	90.58	82.18
6	Colonel Army	90.39	79.81
7	Deacon-Church Of England	89.37	100.00
8	Admiral Rn	89.29	78.43
9	Judge	88.25	95.17
10	Lieutenant-Colonel Army	87.55	79.13
432	Nail Forger	15.88	10.27
433	Mine Laborer	15.02	31.03
434	Spade Maker	14.37	35.96
435	Puddler	11.27	19.98
437	Framework Knitter	9.03	21.13
438	Chainmaker	1.50	24.55
439	Coal Miner	0.43	22.39
440	Ore Dresser	0.34	13.72
441	Nailer	0.00	15.00

Table 4.3: Top and Bottom Ten Ranked Occupations under FOE-RCII index

Status Differences in Marriage in England

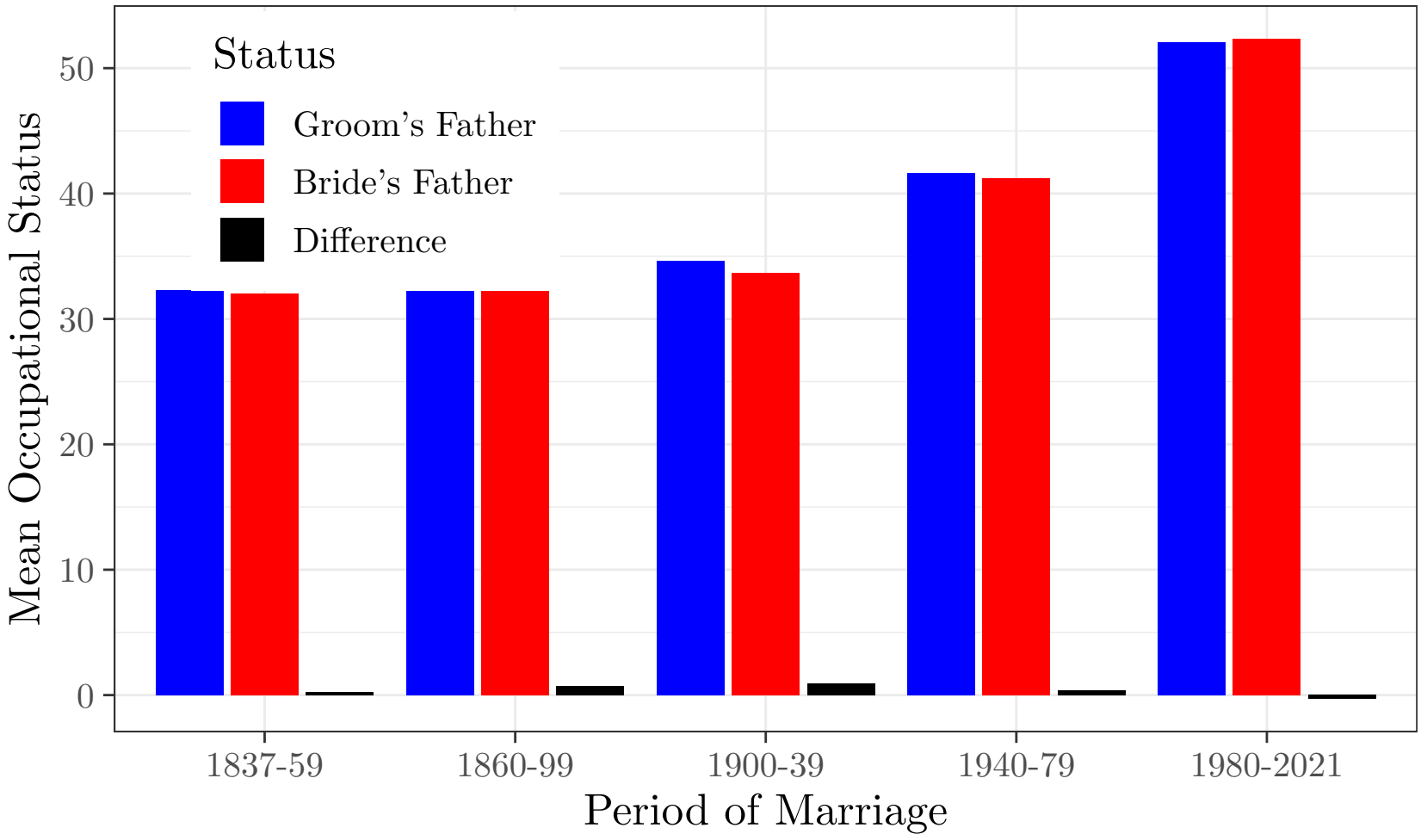


Figure: Parish Marriage Records, 1837-2020

Status Differences in Marriage in Engtand

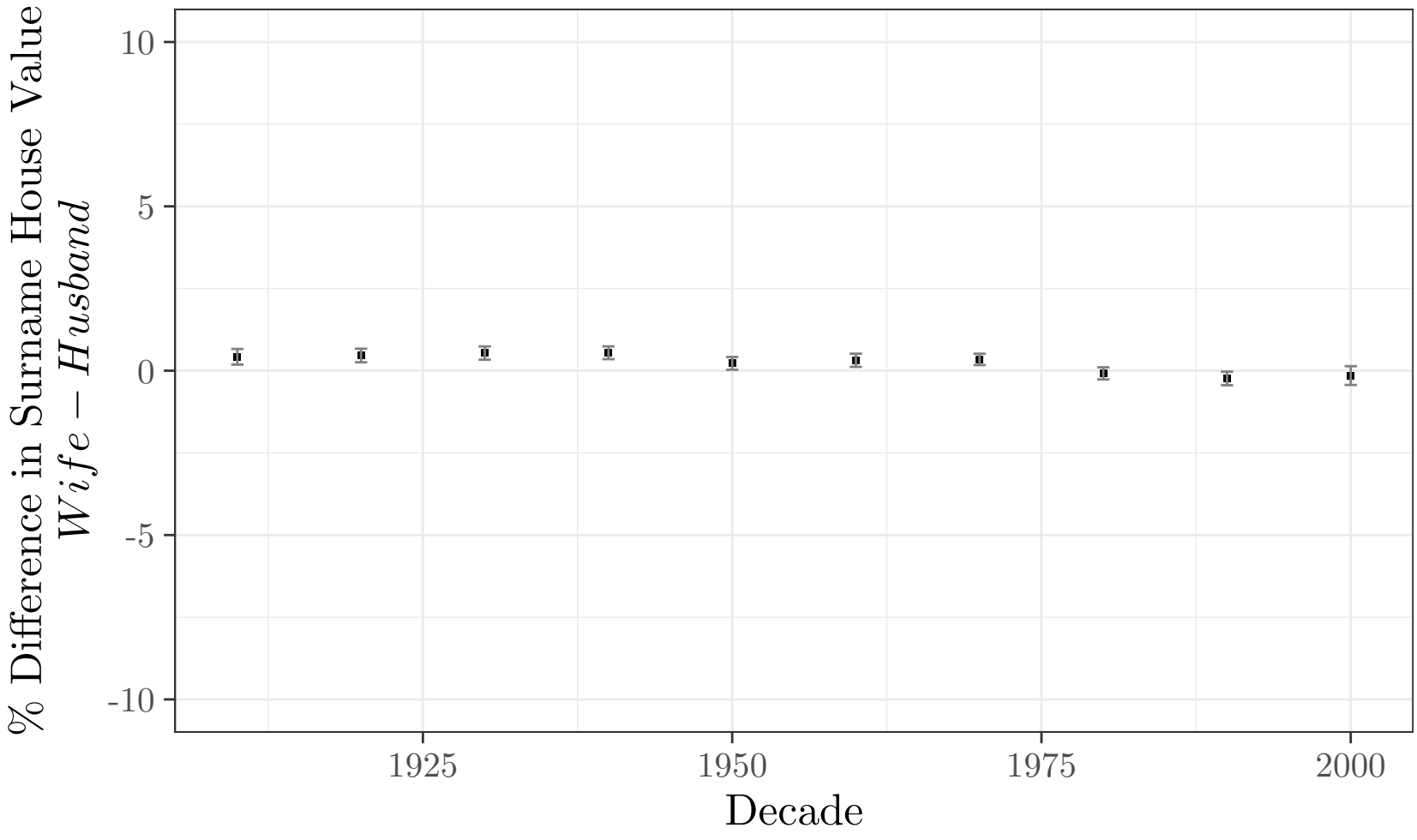
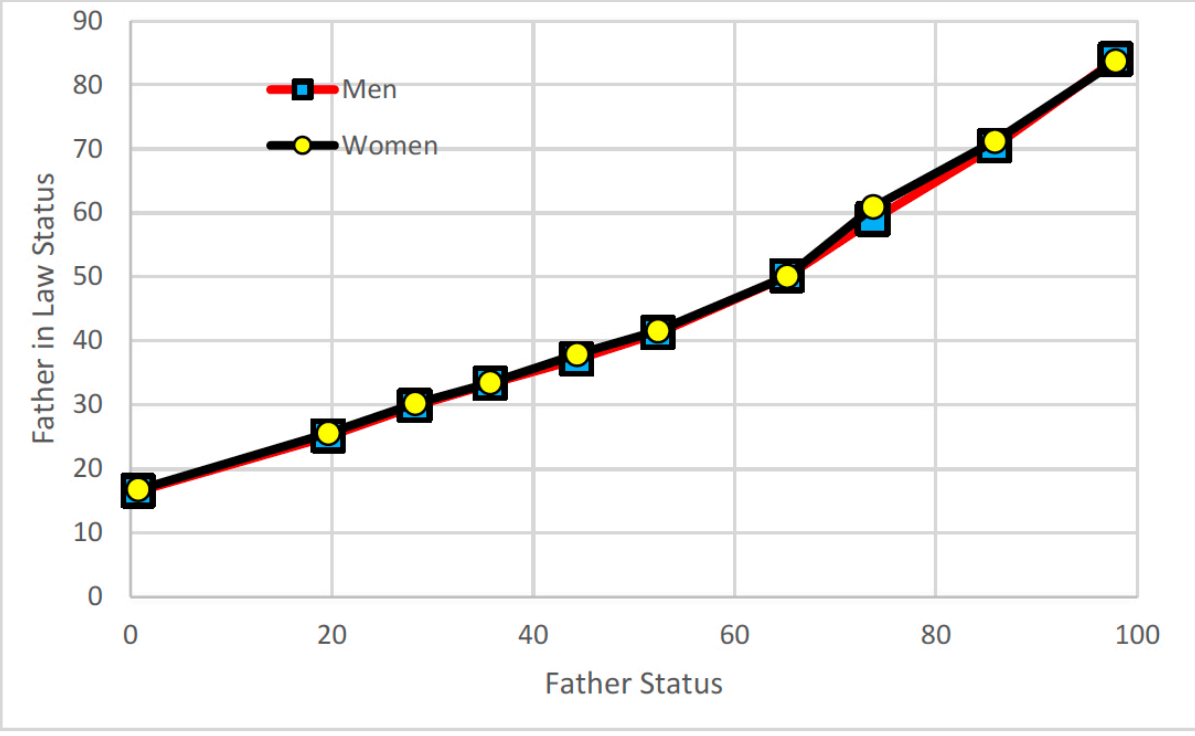


Figure: Universe of Marriage Registers, 1912-2007

Hypergamy?

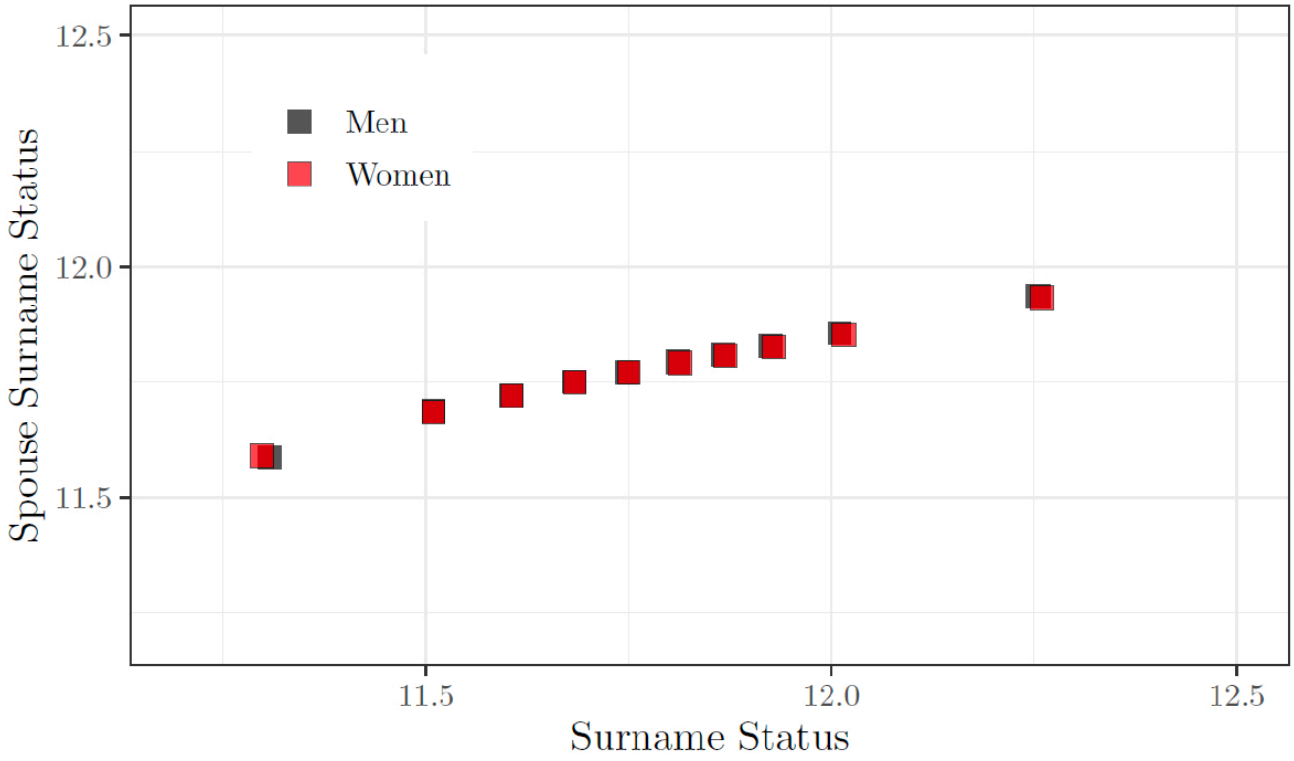
Figure 4: Occupational Status, Father-in-Laws, men and women, by decile, 1837-59



Note: The horizontal axis shows average father status for men and women by status decile. The vertical axis shows the average father-in-law status.

Hypergamy?

Figure 5: Surname Status, men and women, by decile, marriages 1980-2007



Note: Measured for marriages where bride and groom surname has a frequency 10-500 in the 1999 electoral register.

Hypergamy?

- There is no evidence for hypergamy in England 1838 to today
  - Men and Women Marry 'Up' and Marry 'Down' Equally

Hypergamy?

1. There is no significant hypergamy by women in English marriage throughout 1837-2021, as measured by bride and groom fathers' occupational status, or by bride and groom surname house value.
2. Women show no more social mobility in their marital pairings than do men. Across the parent status distribution, women match to men in just the same way as men match to women.
3. Across the family status distribution male and female marriage rates are the same. There is no differential tendency to marry across family status for women compared to men.
4. There is ancillary evidence that in England 1837-2021 the physical appearance of women was a modest determinant of matching in marriage. The underlying matching on social abilities was high and constant at 0.8 1837-2021. Such a high correlation would not be possible if men valued physical appearance in women strongly, and this was uncorrelated with social abilities.

Outline

Who Marries Who: Marital Assortment

- Degree of Marital Assortment matters for:

1. Household level inequality
2. Intergenerational mobility
3. Supply of "upper-tail human capital" (Mokyr et al.) which may matter for IR

- There has surprisingly little study of the degree of marital assortment over the long run
  - Post 1940
  - Years of Education ~.4-.5

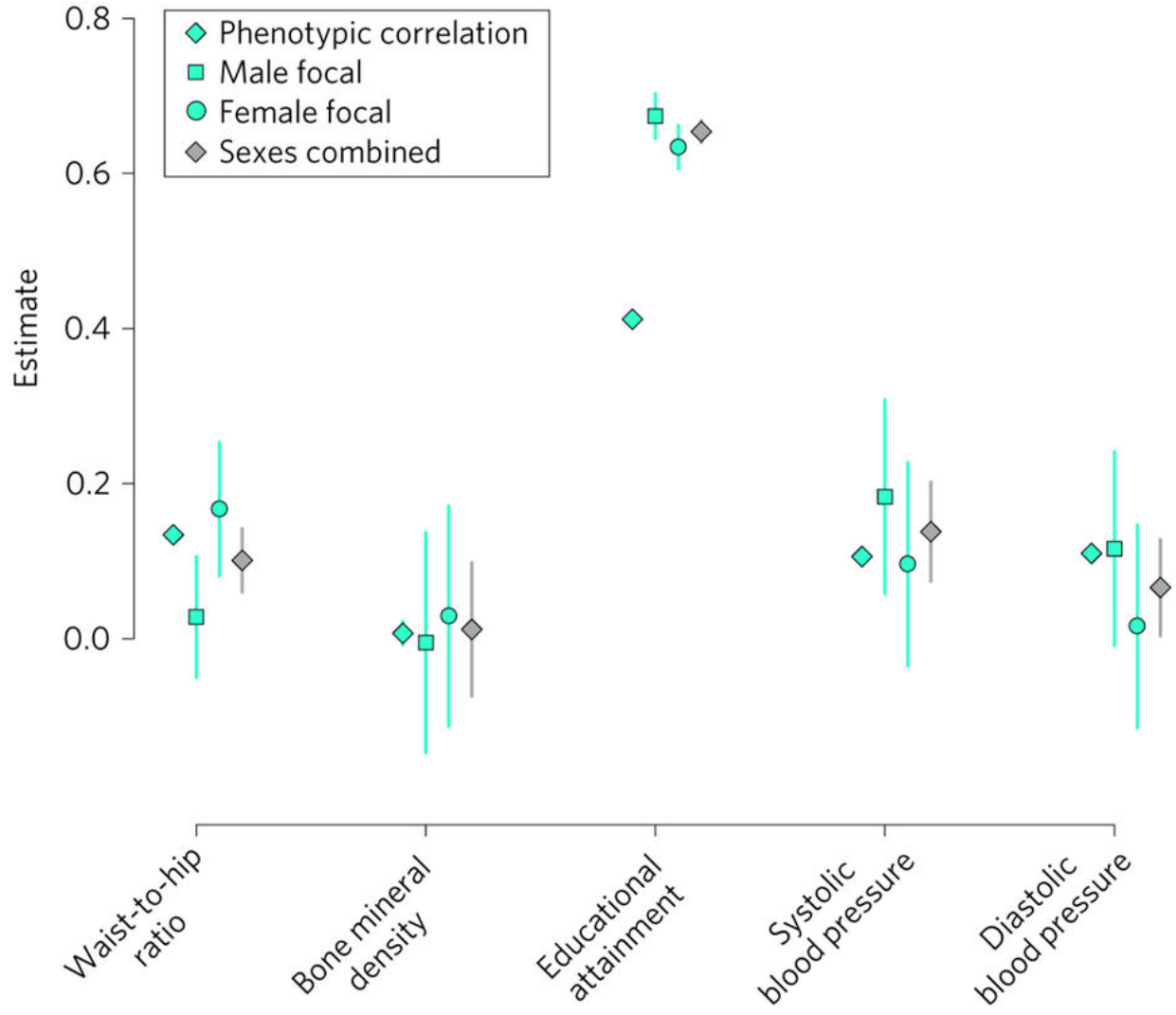
A Latent Variable Model of Marriage Choice

- People have different characteristics
  - Looks, intelligence, social connections, honesty, charm, wisdom; y
- All play a part in mate choice
- But each individual characteristic is but one expression of some underlying overall quality, X
  - Perhaps represented by a vector of these characteristics; $X \approx y_1, y_2, ..., y_n$
- People match on some this **latent factor (X)** that individual observable characteristics ($y_1, y_2, ..., y_n$) can give an imperfect picture of...

Anna Karenina principle

Happy families are all alike; every unhappy family is unhappy in its own way

Evidence of Marital Assortment, 2017



Source

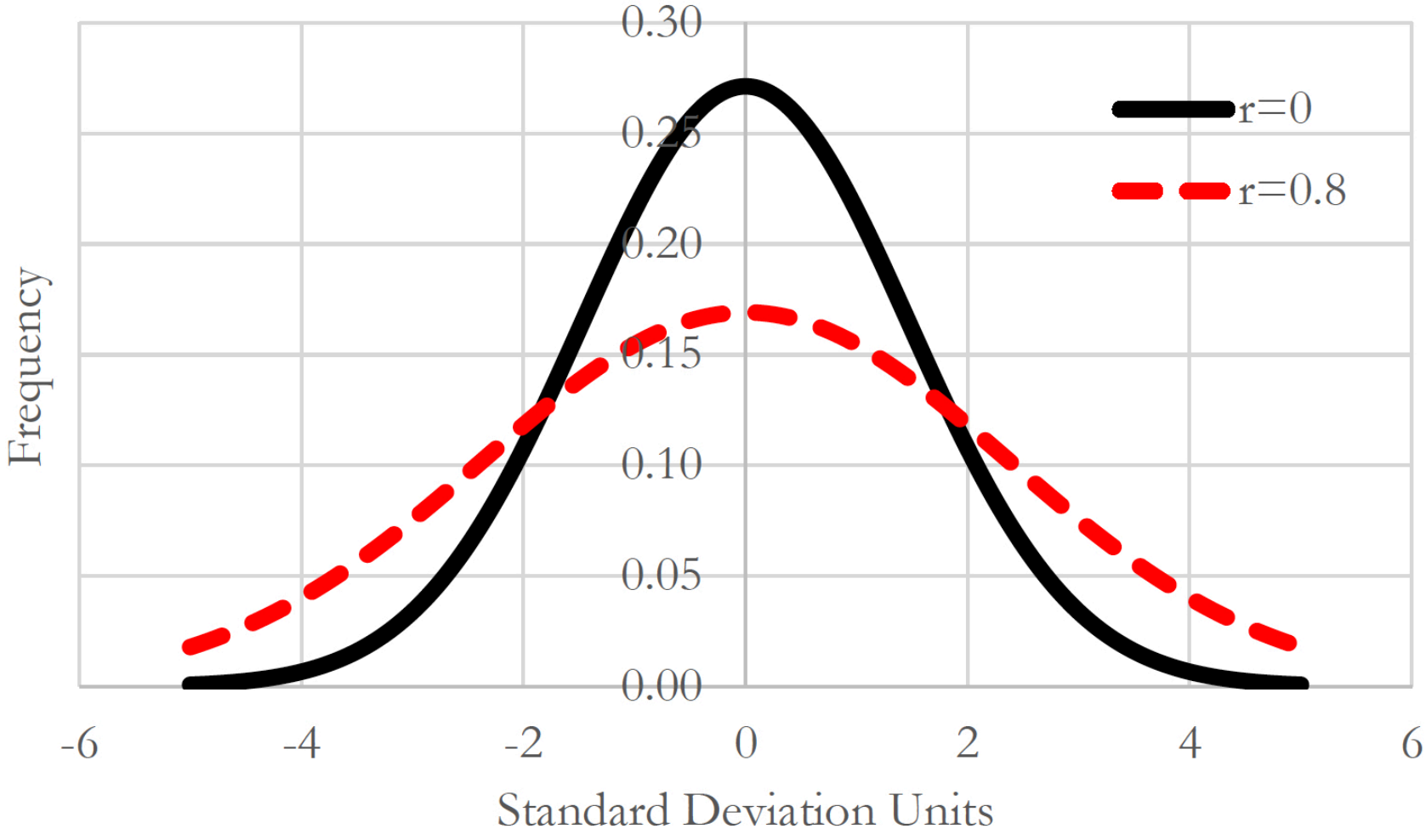


Figure: Marital Assortment and the Social Distribution of Abilities

Note: The standard deviation units are for the situation of zero assortment in marriage.

Effects of Changed Assortment on Social Distribution take time

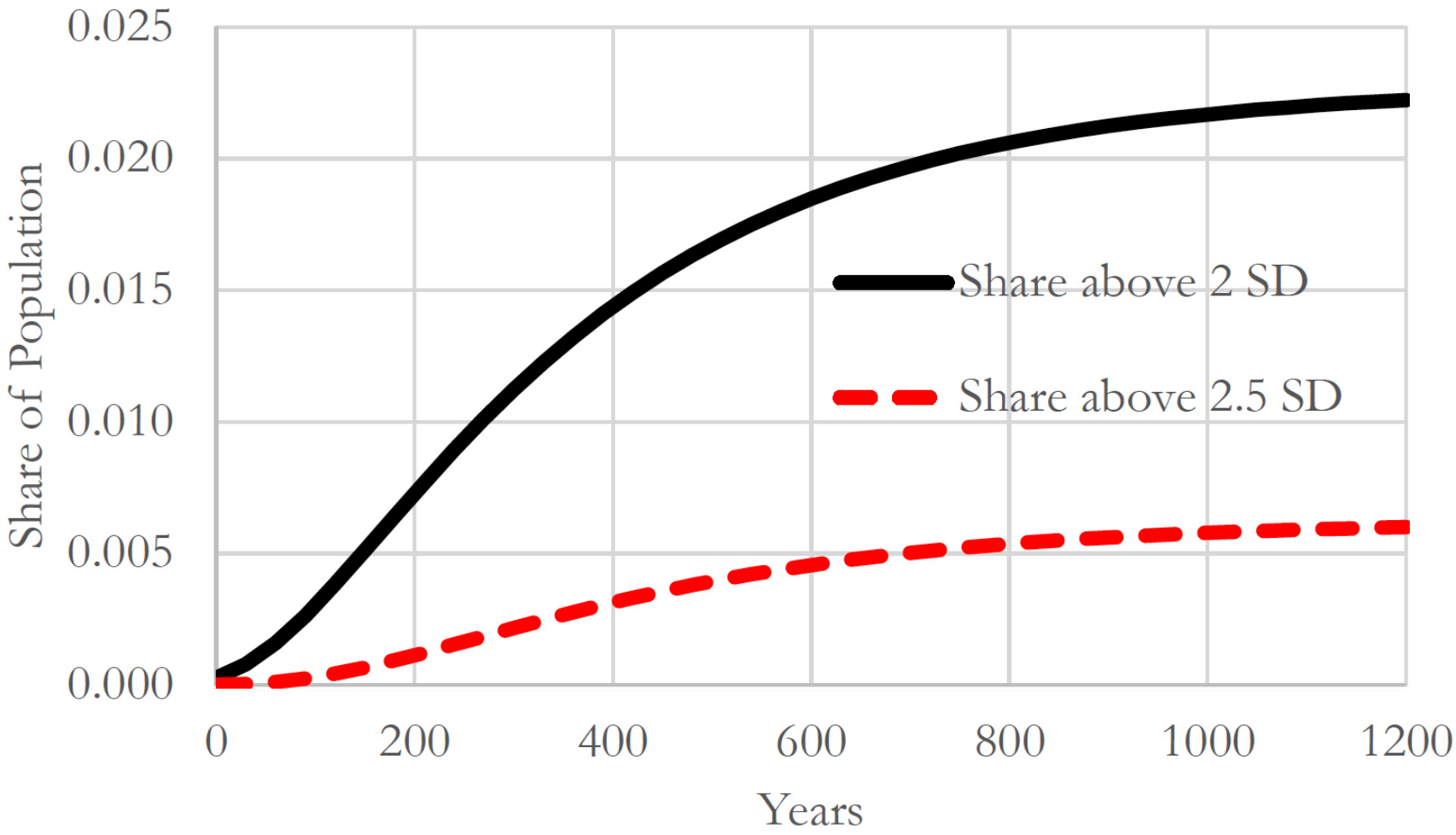


Figure: Dynamics of an Increase in Marital Assortment

Note: The figure shows the share of the population with abilities above 2 and 2.5

Interesting Questions

- Did the adoption of highly assortative marriage in NW Europe in the late middle ages create pre-conditions for the later Industrial Revolution?
- How assortative was marriage in most pre-industrial societies?

Pattern of Correlations around marriage

$r =$ marital correlation

$b =$ intergenerational correlation

The Latent Factor Model of Social Mobility

- Conventional

$$y_{t+1} = \beta y_t + u_t$$

  y is some aspect of status – income, wealth, years of schooling

- But

$$y_t = X_t + e_t$$

- where X_t is an **underlying** status that the various y_t measure imperfectly

Measurement Error Plagues Social Mobility Studies

- Why are surname based estimates so very different from indvidual level correlations?
- And of course!
  - What of change over **time**? And over **space**?

Occupational Labels Change in Status Over Time

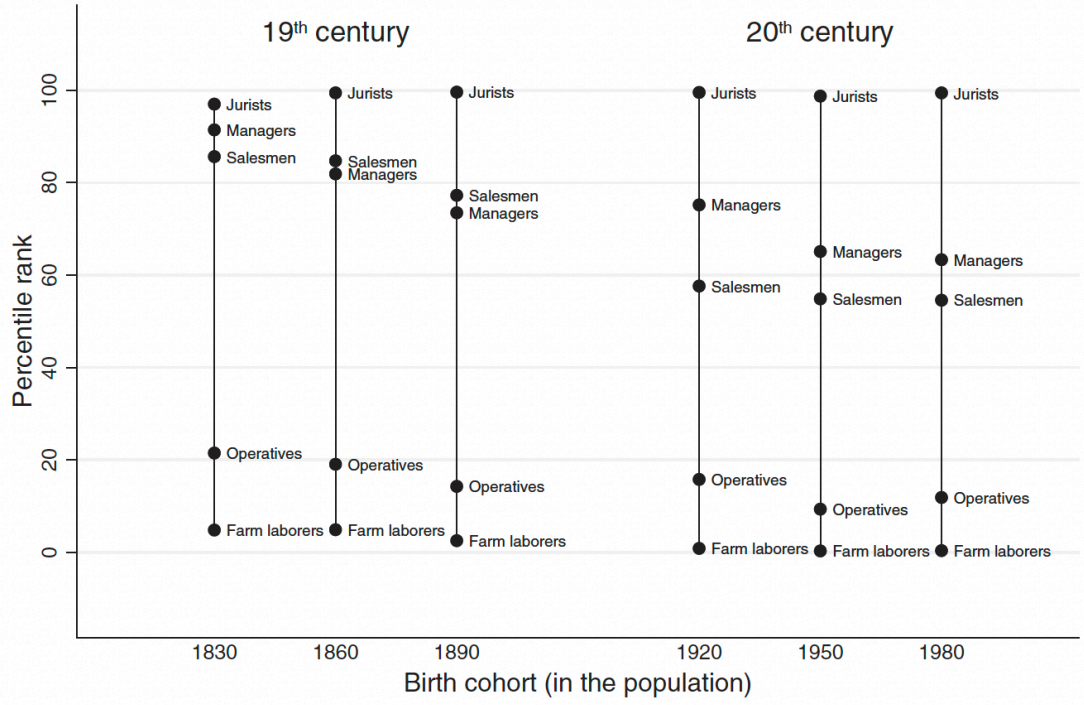


Fig. 1. Trends in occupational percentile ranks among select occupations. Data sources: IPUMS US Population Censuses full count 1850, 1880, 1900, 1910, 1920, 1930, 1940; 1% samples 1860, 1870, 1950; 5% samples, 1960, 1980, 1990, 2000; 6% sample, 1970; ACS 2001–2015. Notes: The figure shows changes in percentile ranks for 5 out of the 70 microclass occupations defined in *SI Appendix*, Table S1. The percentile ranks are estimated from occupation-specific educational distributions by birth cohort using all men and women aged 25 to 64 in population censuses. The methodology is described in *SI Appendix*, section S4. Percentile rank changes for all of the birth cohorts and microclass occupations are shown in *SI Appendix*, Fig. S1. The graph suggests that distances between top-ranked occupations, such as between jurists and managers, have increased over time because of the growth of professional, managerial, and upper nonmanual occupations at the top end of the distribution.

Source: Long-term decline in intergenerational mobility in the United States since the 1850s

And Intergenerational Correlations look like they are rising (e.g. US):

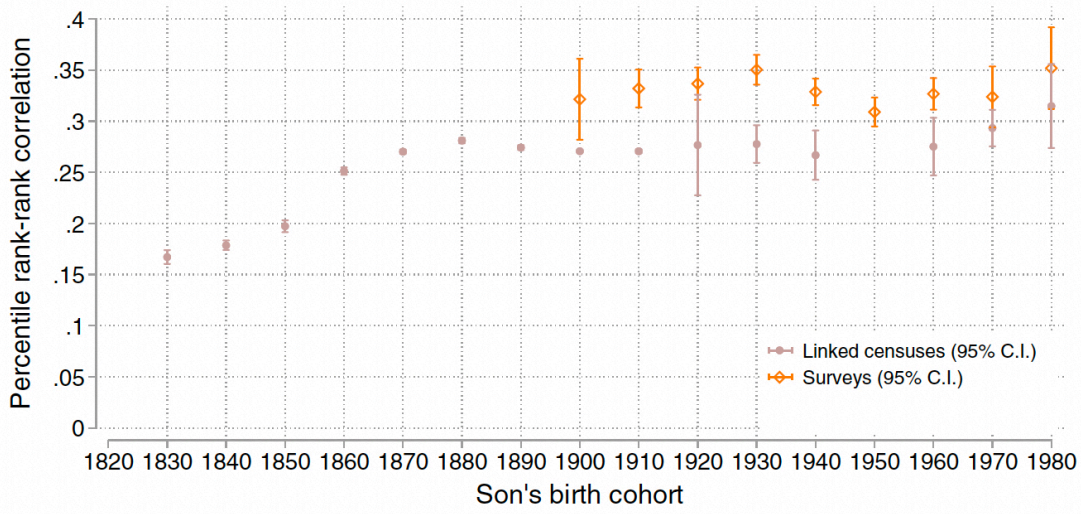
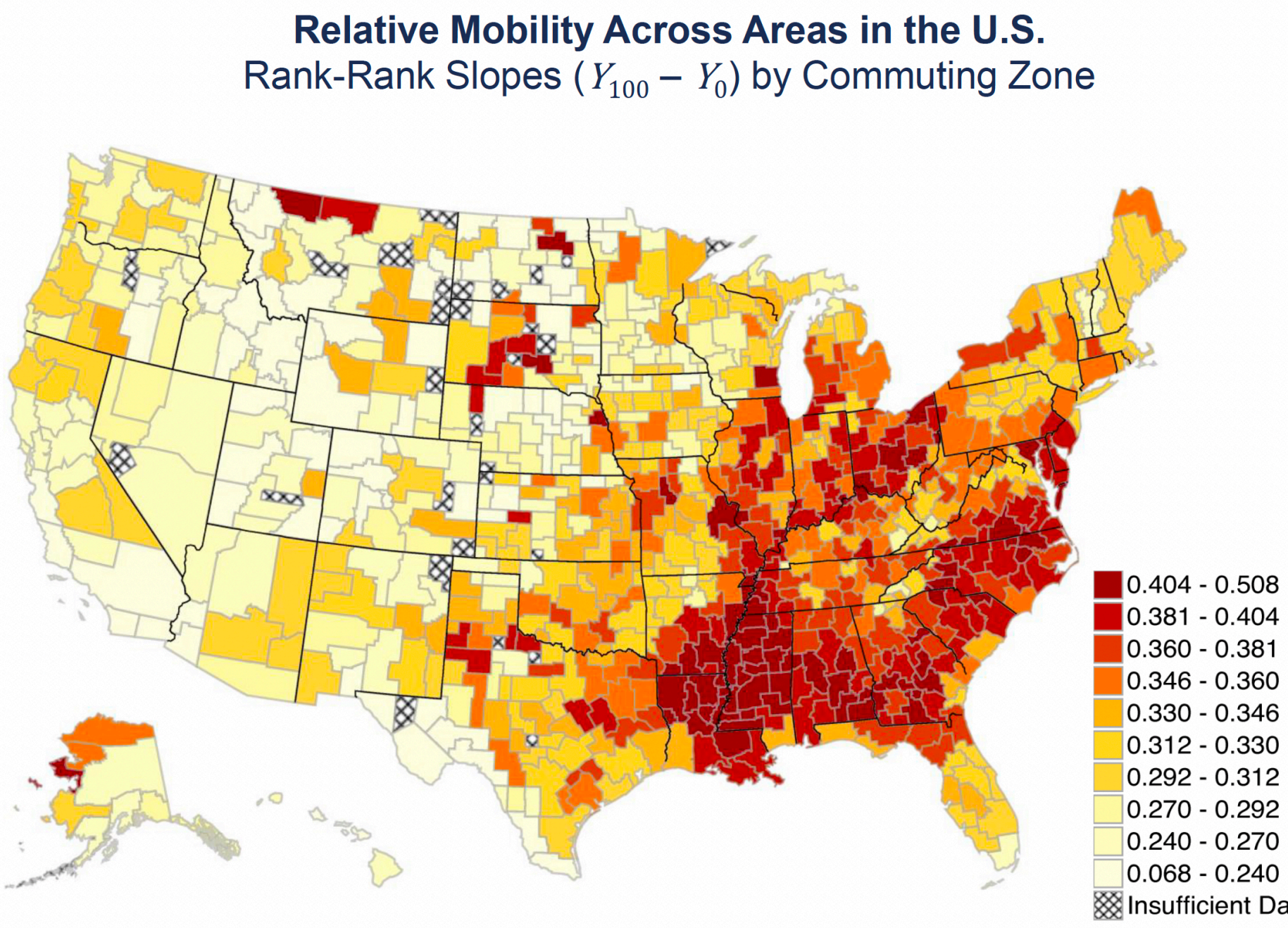


Fig. 2. Trends in intergenerational rank–rank correlation over son's birth cohort. Data sources: Linked censuses include linked historical census data 1850–1880, 1880–1910, 1910–1940; linked contemporary data using census 1940 to CPS (1973, 1979, 1980–1990) and CPS to census 2000/ACS 2001–2015. Survey data include General Social Surveys 1972–2016; Health and Retirement Study 1992–2010; National Longitudinal Survey–Older Men 1966–1990; National Longitudinal Survey–Young Men 1966–1981; National Longitudinal Survey of Youth 79, 1979–2012; National Survey of Families and Households, 1987, 1993, 2002; Occupational Changes in a Generation I and II; Panel Study of Income Dynamics (Survey Research Center [SRC] sample), 1968–2015; Survey of Income and Program Participation, 1986, 1987, 1988; Wisconsin Longitudinal Study, 1957–2011. Notes: The figure plots the estimated correlation in occupational percentile ranks between fathers and sons over 16 birth cohorts of sons. The series of solid circles plots the point estimates of the rank–rank correlations between fathers and sons for the 1830–1980 birth cohorts of sons. The 1950 birth cohort (sons born between 1946 and 1955) is missing because these individuals were not observed during their childhoods in either the linked historical censuses or the contemporary linked CPS-Census/ACS data, which is necessary to make the father–son link. The series in diamond symbols plots the rank–rank correlations estimated from weighted OLS regressions for the 1900–1980 birth cohorts from pooled contemporary social surveys. The weights are constructed from the original sampling probability weight variable in each survey as well as an additional weight variable that takes into account variation in sample size by birth cohort across surveys. The capped spikes refer to 95% confidence intervals of the correlation estimates. Because of the smaller sample size of the survey data relative to the linked census data, the survey estimates show a higher level of uncertainty reflected in wider confidence intervals. See *SI Appendix*, Table S10 for the exact numbers of the estimates.

Source: Long-term decline in intergenerational mobility in the United States since the 1850s

And vary Across Space, like in the US:



But

- *All of these US estimates suffer from measurement error*
  - The translation of underlying status to achieved status
- The observational correlations matter
  - But the true underlying correlation matter much much more
- Where will the descendants of the rich be in 100/500 years time?
  - The observational correlations are useless
  - The underlying correlations are what matter

Two Problems

1. Most women did not have formal educational attainments, or formal occupations, until the mid 20th century
2. It is hard to get consistent measures of educational and occupational status over time

- We don't need direct measures of bride's attributes
  - How?

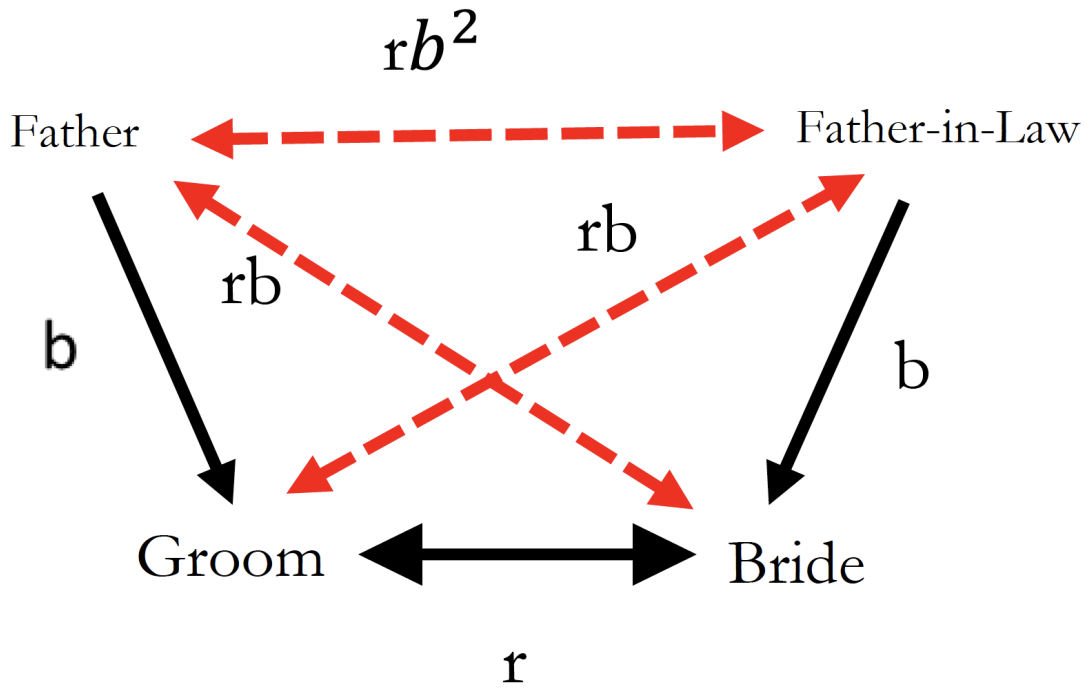


Figure: Underlying Correlations

Note: Causal correlations are indicated by black lines, and non-causal ones by red.

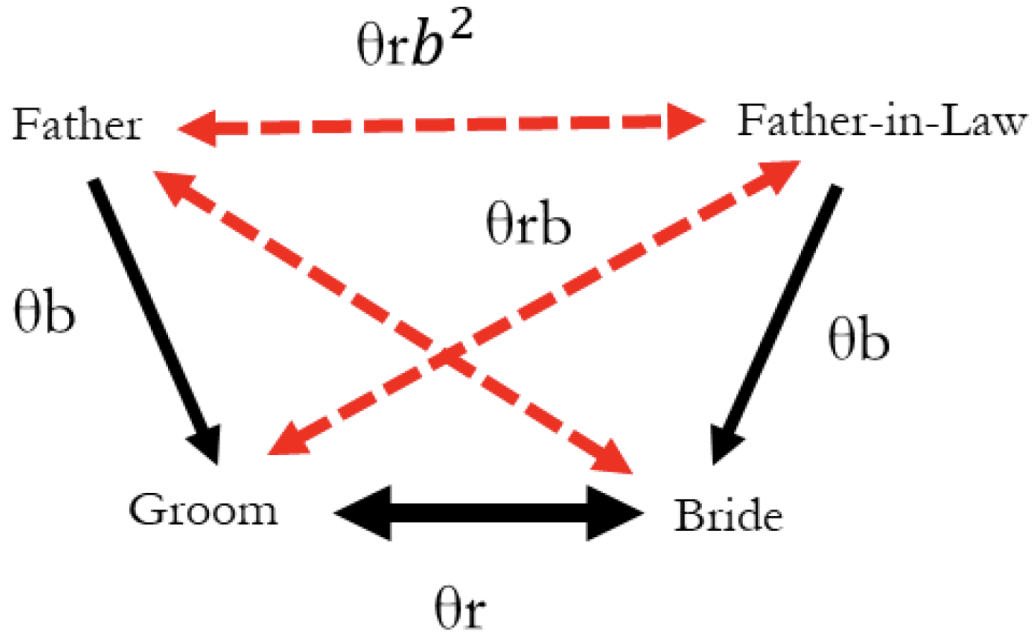


Figure: Observed Correlations with Same Measurement Error for all

Note: Causal correlations are indicated by black lines, and non-causal ones by red.

The Method

- The correlation in status of son and father in law will be θrb,
  - compared to the son-father observed correlation of θb
- Thus the underlying matching correlation in marriage can be estimated from

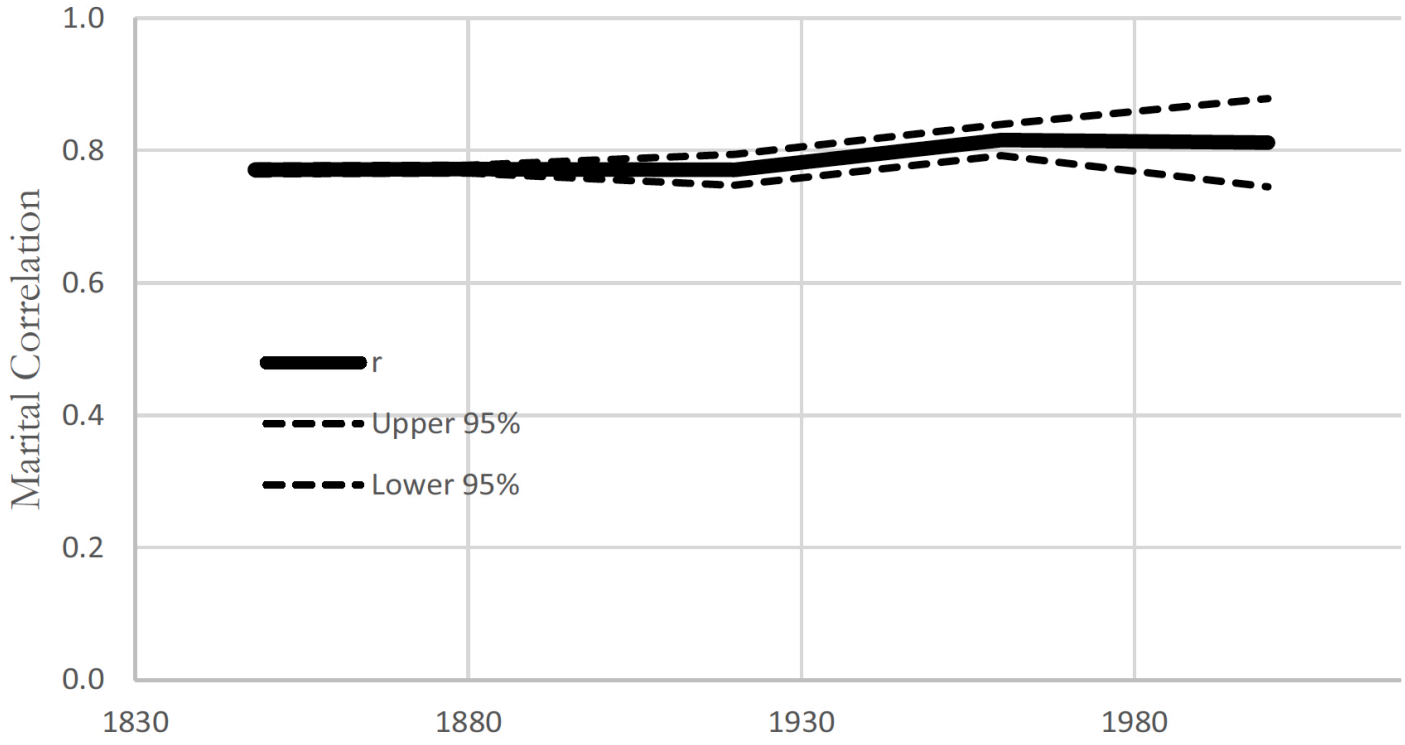
$$r = \frac{\rho_{sl}}{\rho_{sf}} = \frac{\theta rb}{\theta b} \tag{1}$$

- The underlying intergenerational correlation of status in this case as

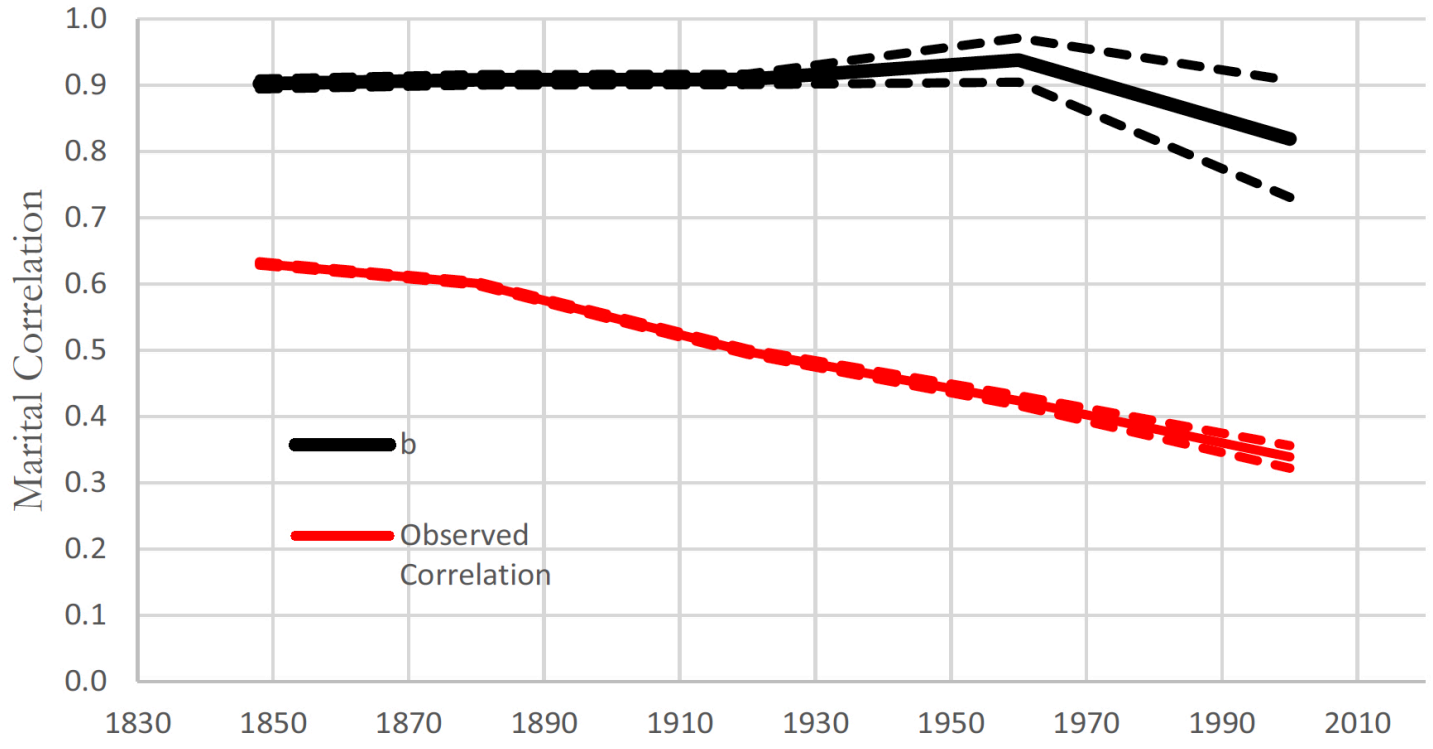
$$b = \frac{\rho_{fl}}{\rho_{sl}} = \frac{\theta rb^2}{\theta rb} \tag{2}$$

- The measurement attenuations cancel out

Marital Occupational Status Correlations



Intergenerational Occupational Status Correlations



A History of Marriage

European Marriage Pattern

- England average age of marriage for women 25, men 27. Consent of parents not required for marriages 21+
- 1837-79 both parties aged 21 and above in 77% of marriages Many men, women living independent of parents aged 16 and above
- Large fraction of men, women never marry
- Matching is bride-groom

Other Pre-Industrial Marriage Systems

- Age of marriage of parties often younger before they can reveal individual social competencies
- Cousin Marriage common – Islamic World, Amerindian Groups
- Arranged marriages between families very common – with again often very young brides

A New Theory for the Origin of Growth?

- England 1750-2021 had a marriage pattern where grooms matched to brides very strongly in social abilities.
- This implied both low rates of social mobility, and high inequality
  - a wide variance in social abilities
- Most pre-industrial marriage systems would have much less close correlations in bride and groom social abilities
- Emergence of European Marriage Pattern circa 1400 would cause substantial rise in share with upper level abilities by 1800
  - And thus could be the fundamental causal force via "Upper tail" Human capital for the origin of economic growth

Fin

Alfani, Guido and Hector García Montero, "Wealth inequality in pre-industrial England: A long-term view (late thirteenth to sixteenth centuries)," *The Economic History Review*, 2022, *75* (4), 1314–1348.

Clark, Gregory, "The Condition of the Working Class in England, 1209-2004," *The Journal of Political Economy*, 2005, *113* (6), pp. 1307–1340.

Cummins, Neil, "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.

Deleixhe, Martin, "Marx, the Irish Immigrant-Workers, and the English Labour Movement," *Historical Materialism*, 2019, *27* (2), 222 – 247.

Engels, Friedrich, *The Origin of the Family, Private Property and the State* 1884. 1901 version at hreflinkwikisource.org.

Gymrek, Melissa, Amy L. McGuire, David Golan, Eran Halperin, and Yaniv Erlich, "Identifying Personal Genomes by Surname Inference," *Science*, 2013, *339* (6117), 321–324.

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.

Lindert, Peter H., "Unequal English Wealth since 1670," *The Journal of Political Economy*, 1986, *94* (6), 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.

Ó Gráda, Cormac, "The population of Ireland 1700-1900: a survey," in "Annales de démographie historique" JSTOR 1979, pp. 281–299.

Piketty, Thomas, *Capital in the Twenty-First Century*, Harvard University Press, 2014.

Rothenbacher, Franz, *The European population since 1945*, Palgrave Macmillan, 2005.

Wrigley, Edward. A. and Roger S. Schofield, *The Population History of England 1541-1871*, Cambridge Universtiy Press, 1981.