Stayin' Elite: Wealth, Multigenerational Persistence, and Political Power Among the US Hyperwealthy, 1875-1950 *

Julián Costas-Fernandez† Kirill Shakhnov‡

March 24, 2026

Abstract

This paper studies the intergenerational transmission of wealth among the top wealthiest US individuals between 1875-1950. We combine historical rich-lists with crowd-sourced genealogies to link top wealthy individuals with both their vertical and extended kinship across cross-sections of rich-lists. For 18% of individual in these lists we can find a vertical descendant in a rich list twenty-five years later. This declines to 14% when considering seventy-five year periods. Persistence is much higher when we consider extended families, we find that over 40% of wealthy individuals are linked to a wealthy individual twenty-five years later when we consider extended kinship. We provide evidence supporting that these extended kinship connections are strategic and improve the likelihood of wealth persistence at the very top. We document that, among the hyperwealthy, the offspring of wealthier individuals have a social connectivity advantage, meaning that they serve as bridges between wealth dynasties through extended family relations. This connectivity is valuable, IV estimates exploiting within-dynasty higher-order connectivity shocks show that one standard deviation increase in children connectivity increases the likelihood of wealth persistence by 22% and total observed wealth in the next cohort by 31%. Finally, children connectivity increases the likelihood of wealthy elites branching out into another elite: US Congress dynasties.

Keywords: Wealth Persistence, Economic Elites, Social Networks

JEL codes: D31, N32, N31, Z13

*Hung (Harry) Vu provided excellent research assistance. Authors declare no conflicts of interest.

†University of Surrey

‡University of Surrey

1 Introduction

Recent literature has shown that intergenerational persistence in economic outcomes is stronger than previously thought, with assortative mating, where advantaged families form links with other advantaged families, playing a significant role (Collado et al., 2023; Barone and Mocetti, 2021; Clark, 2014; Goñi, 2022). Moreover, evidence suggests the underlying processes of intergenerational mobility differ for the top 1% compared to the rest of the population (Corak, 2013). Based on Scandinavian data, Björklund et al. (2012), for example, show that persistence is particularly high at the very top of the income distribution, with dynasties occupying the top 1 %, where intergenerational wealth transmission appears to be the dominant mechanism. However, there remains limited evidence on how persistent wealth is among the wealthiest families, especially in the US over the long run, and the extent to which strategic family link formation contributes to maintaining that wealth across generations.

We study the long-run persistence of wealth among the richest families in the United States by combining historical "rich lists" with genealogical data. For this, we combine rich lists with information for top wealthy individuals during the late 19th to mid-20th century with crowd-source genealogical datasets to reconstruct family trees and trace both direct descendants and extended kinship connections across nearly a century. Our analysis shows that wealth persistence is substantial, not only through direct inheritance but even more strongly through extended kinship ties. Families at the top repeatedly appear across generations, forming a tightly interconnected web of dynasties that links economic elites with one another, and at times, with political elites as well. This evidence points to the importance of family networks in sustaining elite positions well beyond individual lifetimes.

We provide evidence consistent with individuals engaging in strategic family network formation seeking to maximize the connectivity of their offspring. We show that, among the hyperwealthy, wealthier individuals generate offspring that serve as connectors between wealth dynasties. OLS estimates suggest that these connections might be valuable in terms of improving wealth persistence: increasing children connectivity by one standard deviation increases the likelihood of wealth persistence by 27% and total

observed wealth in the next cohort by 63%. These estimates, however, may reflect unobserved heterogeneity among better-connected dynasties. Indeed, we provide evidence that wealthy individuals with better connected children are more likely to have a parent appearing in a previous cohort of the wealth data. We address this concern with an IV strategy that exploits within-dynasty shocks in higher-order connectivity. Here we instrument children's connectivity with the average connectivity of their extended kinship controlling for the expected connectivity of the latter. Thus, we follow Borusyak and Hull (2023) by recentering our instrument around its observation-specific expectation by treating connectivity among the extended kinship of an individual's nephews, children in law, and their own children as draws from the same counterfactual distribution of connectivity. Importantly, we show that the variation we isolate is orthogonal to initial wealth and previous persistence. With this strategy, we find that a one standard deviation increase in children's connectivity increases the likelihood of wealth persistence by 22% and total observed wealth in the next cohort by 31%.

In addition, by linking our data to Dal Bó et al. (2009), we show that better connected children increase the likelihood of wealthy dynasties diversifying into the political elite by forming extended kinship links with established political families. Our findings further suggest that these links to Congress are strategic: Congress members with extended kinship connections in the wealth elite are significantly more likely to have had family members in Congress before them and to have family members serving after them. All in all, our evidence supports the idea that economic elites engage in strategic marriage, and that these strategies are successful in increasing elite persistence within the economic elite and diversifying into other dimensions of eliteness.

We contribute to several strands of literature. Despite interest on intergenerational persistence of wealth, especially at the top, evidence has been typically focused on Scandinavian countries due to data availability.[1] For the US, Kalsi and Ward (2025) use Census data for the 19th and 20th century along with various proxies of wealth to study wealth

[1] Collado et al. (2023) examines intergenerational and assortative processes using extended family data from Swedish administrative records. By comparing relatives of varying degrees of kinship within the same generation, the authors adopt a horizontal approach to quantify the persistence of economic status. Adermon et al. (2021) estimates long-run intergenerational persistence in human capital by analyzing outcomes across four generations in Sweden. Using comprehensive population-level data, the study captures

persistence at the top. They find that about 7% of grandchildren with a grandfather in the top 1% stay in the top 1%. Furthermore, Kalsi and Ward (2025) also document inequality at the very top with grandchildren of the top 0.1% having a likelihood of being in the top 1% of about 14%. Other research on wealth persistence at the top in the US draws from Forbes 400. For example, Kaplan and Rauh (2013) research based on the Forbes 400 list reveals that while some billionaires are self-made, a substantial portion inherit their wealth, indicating limited wealth mobility at the very top. Longitudinal studies show that intergenerational persistence remains strong despite periods of economic upheaval. Moretti and Wilson (2023) show that estate taxes and tax policies significantly affect the geographical distribution of ultra-wealthy individuals. In states with high estate taxes, billionaires are more likely to relocate to minimize tax burdens, which impacts local wealth accumulation. This paper focuses on the US and brings a multigenerational perspective and extended kinship to understand the multigenerational persistence in the US at the very top of the wealth distribution.

What is less known is the mechanism behind the higher persistence at the very top. Beckert (2022) argues that the mechanism of wealth preservation at the very top is different and is not simply a matter of individual achievement but is maintained by a complex system of legal, financial, and social mechanisms that reinforce economic inequality across generations. We add to this debate by providing evidence consistent with wealthy American families historically engaging in strategic family formation, and that family connectivity gave these families advantages in terms of wealth persistence and elite diversification. In this direction, Goñi (2022) shows that elites rely on matching technologies to improve sorting in marriage. Our evidence further supports this point by showing that elite families have incentives, in the form of improved elite persistence, from marrying their children strategically. In a similar vein, our findings relate to Michelman et al. (2022) who show that belonging to elite clubs improves the likelihood of holding elite jobs and entering other elite groups later. Our research shows that even within the elite there is

the dynastic transmission of human capital through both direct parental influence and extended family dynamics.

heterogeneity: those individuals better connected within the elite enjoy higher elite persistence than those worst connected.

2 Data

Measuring wealth at the very top over long periods is challenging. As described by Kopczuk (2015), there are four methods of measuring wealth at the very top of the distribution. First, one can carry out a survey that oversamples high-net-worth taxpayers. To the best of our knowledge, the Survey of Consumer Finances (SCF) is the only source of that kind in the United States. Second, while the United States does not have an annual wealth tax, it does have an estate tax. Estate tax records provide a snapshot of the distribution of wealth at the time of death. Third, while wealth itself is not reported to tax authorities, much of the capital income that wealth generates is taxable and observable, which provides an opportunity to estimate the underlying wealth distribution based on the annual flows of capital income. Finally, lists of named top wealth-holders exist, with Forbes being the best-known such list since 1982. All these methods, however, have limitations. SCF undercounts the ultra-wealthy, estate records only capture wealth at death, income-based estimates rely on assumptions, and rich lists lack transparency in their sampling design.

Alternatively, mid to late XIX century US Census capture measures of wealth. In 1850, enumerators collected information on real estate wealth only. In 1860 and 1870, questions on personal property were added to the census. However, US Census does not contain direct measures of wealth for later periods. On the other hand, SCF and Forbes are only available for the modern period: Forbes 400 started systematically counting ultra-wealthy from 1982 and the first SCF survey was conducted in 1983. These two sources have been used complementarily to overcome the undersampling of the very wealthy (see Saez and Zucman, 2016; Smith et al., 2021). Estate tax and capitalization methods allow for constructing estimates going back to the beginning of the 20th century: the US income tax

was introduced in 1913, and the estate tax was introduced in 1916 (Kopczuk, 2015)[2]. To overcome this limitation, we rely on a unique, manually collected list of wealthy individuals, which provides a rare window into the distribution and dynamics of wealth at the very top.

2.1 Historical Rich Lists

Our rich lists come from A Classification of American Wealth (CAW), an online compilation of top wealth-holders in the US compiled by Drew Caradine Shouter (a pseudonym). The now-defunct website contained lists of wealthy Americans along with their biographies and family trees, drawing on various historical sources. The data were previously available at `http://www.raken.com/american_wealth/index.asp` as a subscription-based product and have been used in prior economic research (see Edlund and Kopczuk, 2009; Kopczuk, 2015; Turchin, 2016). Each edition typically includes between 600 and 1,500 individuals; the earliest data we use are from 1875 and the latest 1950. Thus, our sample spans close to a century covering several interesting periods: the first wave of globalisation from 1870 to the First World War; the Gilded Age and roaring 1920s, the great depression, the Second World War, and the post-war economic expansion. During this period the inequality has increased in the initial period, to then flatten and finally decreased after the Second World War, see Figure 2.

The best way to conceptualize the list is to think about them as Forbes 400, as Forbes has published the best-known such list since 1982, prior to Forbes 400 lists. Table 1 highlights several important trends. First, the share of individuals whose wealth originates from inheritance rises substantially across cohorts, indicating stronger dynastic persistence at the top of the wealth distribution. The proportion of women also increases,

[2]The confidentiality protections of tax returns do not automatically expire, even after 75 years. The IRS still cannot release individual-level returns unless Congress changes the law or grants specific exceptions. In addition, wealthy individuals utilized estate planning strategies to minimize potential estate taxes. For example, Henry Ford transferred a significant portion of his wealth to the Ford Foundation, a charitable organization, during his lifetime and in his will, which significantly reduced the taxable value of his estate. This allowed him to pass on control of the Ford Motor Company to his family while minimizing the impact of estate taxes.

TABLE 1. Descriptive Statistics by Cohort

	(1) Mean	(2) Standard Deviation	(3) Minimum	(4) Maximum
Panel A. 1875 Cohort				
Age	54.71	14.44	16	88
Male	0.93	0.26	0	1
New York Resident	0.36	0.48	0	1
Wealth (2010 USD millions)	67.04	118.09	17	1,730
Inherited Wealth	0.16	0.36	0	1
Observations		585		
Panel B. 1900 Cohort				
Age	52.90	14.77	16	95
Male	0.77	0.42	0	1
New York Resident	0.33	0.47	0	1
Wealth (2010 USD millions)	146.52	275.95	22	5,468
Inherited Wealth	0.41	0.49	0	1
Observations		1,402		
Panel C. 1925 Cohort				
Age	54.72	14.99	16	95
Male	0.68	0.47	0	1
New York Resident	0.34	0.47	0	1
Wealth (2010 USD millions)	151.03	313.32	21	6,370
Inherited Wealth	0.53	0.50	0	1
Observations		1,567		
Panel D. 1950 Cohort				
Age	53.90	15.86	16	92
Male	0.62	0.49	0	1
New York Resident	0.24	0.43	0	1
Wealth (2010 USD millions)	175.54	364.78	23	6,816
Inherited Wealth	0.66	0.47	0	1
Observations		1,064		

Notes: This table reports descriptive statistics for wealthy individuals in the CAW dataset, disaggregated by cohort of entry into rich lists (1875, 1900, 1925, and 1950). For each cohort we present the mean, standard deviation, minimum, and maximum of key demographic and economic characteristics, including wealth, age, and family attributes. Data quality varies: residence information is unavailable for 5% of individuals. The statistics provide a first look at the changing composition of the wealthy over time.

reflecting both greater visibility of female wealth holders and shifts in legal and social norms.[3] Second, although longevity in the general population improves markedly over this period,[4] the average age of wealthy individuals remains surprisingly stable. This suggests that entry into the top tail of wealth is not strongly associated with life expectancy gains. Third, while nominal wealth values grow, real wealth holdings do not increase significantly after 1900, consistent with the idea that relative positions within the elite are more stable than absolute levels of fortune.

We investigate the validity of the data by comparing it with a comprehensive set of alternative sources on wealth or income, available for the same period, including comprehensive lists, such the US 1870 census, which provide information on wealth; the comprehensive list of millionaires: 1892 Tribune list of 4056 millionaires and World Almanac list of 3734 millionaires; the United States Individual Income Tax Payments, 1923-1924 for New York Chicago and Washington. We also match our data to most well-known specialized lists, such as Forbes 1918, Morgenthau 1941, and Fortune 1957. This way for each list of CAW we have at least one additional source of information. Table C.2 summarizes all other alternative list which we use.

In Table 2, we compare the CAW data with smaller wealthy lists that we have manually linked at the individual level. Column 2 shows that, for alternative datasets also reflecting wealth, the share of individuals who are contained in the alternative list and CAW is always above 90%. Moreover, in column (3) of the same table, we report a strong correlation between measures of wealth in the CAW data and the alternative rich list. For all datasets we find strong correlations, particularly for alternative rich list that reflect wealth, i.e. Forbes 1918 and Fortune 1975. In addition, we show that the CAW lists provide particularly good coverage at the very top of the wealth distribution, especially among deca- and centimillionaires. Table C.4 compares the share of CAW millionaires, decamillionaires, and centimillionaires with the general population. As can be seen, the

[3] Edlund and Kopczuk (2009) argue that the increase in the share of women among the very wealthy in the United Stats that peaked in 1960's reflects an increase in the importance of dynastic wealth at the very top.

[4] Life expectancy at birth in the United States rose from 47 years in 1900 to 78 years in 2000 (Cutler et al., 2006).

CAW lists capture at least 70% of these more affluent groups, indicating a high degree of representativeness at the upper end of the wealth scale.

Moreover, Table C.5 in the Appendix provides differences in terms of observable characteristics contained in the 1875 CAW and the 1870 Census. First, in the 1875 CAW we observe 585 individuals after we select those that are known to be between 16 and 95 years of age. This is about 316 less individuals than individuals in the 1870 Census with total wealth above about one million.[5] We also finde that, when compared to the 1870 Census, the 1875 CAW data under counts wealthy females, contains somehow older individuals and there are sizeable differences in terms of industry. The latter should be interpreted with caution: we have done a manual reclassification from IPUM 1950 industry codes into CAW industry labels, this reclassification might be far from perfect. For example, this might explain the large share of Census observations that we classify in the *Other* industry group. Figure D.4 compares the geographical distribution of individuals in CAW and the 1870 Census. We observe that individuals from New York and Pennsylvania are over-represented in the CAW data. Finally, Figure D.1 provides wealth distributions for the 1875 cohort of the CAW data where we provide a raw density and a density where we re-sample the CAW data to have the same distribution of observables as the 1870 Census. This figure also suggests that CAW under-samples the less wealthy among the very wealthy.[6]

[5] In the Census data we have access to wealth is censored at $999,997.

[6] Table C.6 in the Appendix compares characteristics observable in in the 1900 CAW and the 1892 Tribute List. CAW has a larger share of females, reports larger shares of old money and significant differences in terms of industry. In Figure D.5 in the Appendix, we compare the geographical coverage of these two sources. Compared with the 1892 Tribune List, CAW oversamples individuals living in Pennsylvania and undersamples individuals in California and Ohio. Differently from Census, which samples the whole population, the sampling framework of the 1892 Tribune List is unknown. It is therefore unclear, whether CAW or the 1892 Tribune list provides a more accurate description of the population of interest.

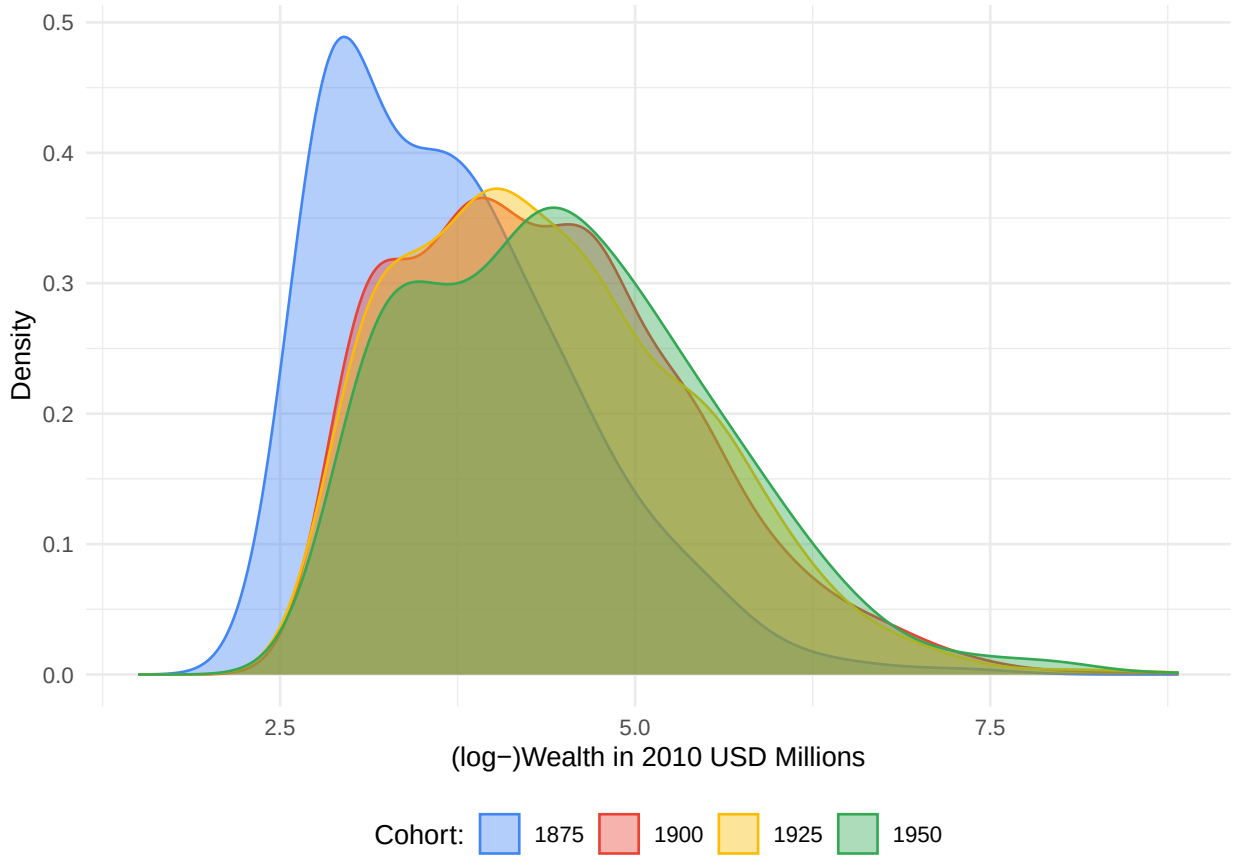


FIGURE 1. Wealth Densities

Notes: Figure provides the density of (log-)wealth for individuals in the CAW data by cohort.

Figure 1 plots the kernel densities of log-wealth (in 2010 USD millions) for individuals in the CAW dataset, separately by cohort of entry (1875, 1900, 1925, and 1950). The distribution for the 1875 cohort (blue) is tightly concentrated at lower levels of wealth, with a sharp peak around 3–4 million. By contrast, later cohorts (1900, 1925, 1950) exhibit broader distributions with fatter right tails, reflecting the emergence of extremely wealthy individuals. Although the bulk of the mass remains relatively stable across cohorts, the 1950 distribution shows both greater dispersion and higher maximum observed wealth, consistent with widening inequality at the very top. Taken together, the figure

suggests that while median fortunes did not rise substantially after 1900, the upper tail stretched significantly, suggesting persistence of wealth combined with growing heterogeneity among the elite.

TABLE 2. Comparison with Other "Rich Lists"

(1) Dataset Size	(2) Share Contained in CAW	(3) Wealth Correlation	(4) Rank Correlation
Detroit Evening 1892			
42	0.929		
Forbes 1918			
30	1.000	0.685	0.444
Morgenthau† 1940-1941			
98	0.582	0.517	0.294
Fortune 1957			
76	1.000	0.675	0.251

Notes: Table provides descriptive statistics comparing the Classification of American Wealth (CAW) data with other "Rich Lists" for the same period. Column (1) provides the number of observations in the alternative rich lists, column (2) provides the share of these observations that are also contained in CAW. When available, column (3) and (4) provide wealth and wealth-rank correlations between the given alternative rich list and CAW where we use the closest available CAW cohort. †Morgenthau 1940-1941 does not provide wealth but net income.

Figure 2 compares estimates of wealth concentration in the United States, focusing on the top 0.1% and 0.01% wealth share, across multiple sources. The CAW samples, shown in black, represent around 1,000 wealthy individuals, or about 0.0015% of the U.S. population, which is two orders of magnitude smaller than the typical top 0.1% group examined in wealth inequality studies. The figure provides a comparison of wealth share from CAW samples to other estimates in the literature: based on tax returns (Saez and Zucman, 2016; Kopczuk and Saez, 2004) or Census from 1850 to 1940 (Kalsi and Ward, 2025). Notice that those 3 different sources are not directly comparable for several reasons. First, It is well known that estimates of top income and wealth shares in US household survey data are generally lower than the estimates from US administrative income tax



FIGURE 2. CAW Samples vs Top Wealth Shares

Notes: The figure provides a comparison of wealth share from CAW samples to other estimates in the literature: based on tax returns (Saez and Zucman, 2016; Kopczuk and Saez, 2004) or Census from 1850 to 1940 (Kalsi and Ward, 2025).

data. There is an issue of undersampling of the rich, see Figure D.1, which is particularly severe for Census data. In addition, census data provide very limited information about individual wealth, with perhaps the exception of 1860 and 1870 Censuses. Using Census one has to deal with the twin problem of differential non-response and underreporting. Second, wealth can be estimated using income tax data. Those estimates are measured indirectly with the help of a model that "capitalizes" taxable income into wealth, and calibrates to external wealth aggregates. CAW samples instead are lists of named top wealth-holders. They rely on either the value of the company stock or the estimates of the estate upon the individual's death, which are mentioned in their biographies or other written sources. CAW samples provide the largest estimates for wealth inequality. Some might find it surprising, however, we get similar levels of inequality if we compare with the Forbes lists, e.g., in 2021, the collective wealth of the 400 richest Americans on the Forbes 400 list was $4.5 trillion, which corresponds to 3.75% of total household wealth in the US compared to the highest top 0.001% (or 3300 individuals) of the population wealth share 4.5%

3 Methodology

In an ideal case, when we observe the entire population of the US across many decades (generations), P_k^*, the estimation of intergenerational persistence becomes a trivial exercise. We are interested on the likelihood of an individual belonging in the population of hyper-wealthy individuals (U_k^*) being connected through family relations to a future generation of hyper-wealthy $U_{k+\tau}^*$

$$\begin{pmatrix} U_{k+\tau}^* \\ P_{k+\tau}^* \setminus U_{k+\tau}^* \end{pmatrix} = \begin{pmatrix} \rho_\tau & (1-\delta_\tau) \\ (1-\rho_\tau) & \delta_\tau \end{pmatrix} \begin{pmatrix} U_k^* \\ P_k^* \setminus U_k^* \end{pmatrix}$$

There are two issues with the data. First, unfortunately, we only observe a sample of the population of ultra-wealthy individuals $U_k \subseteq U_k^*$ in different periods. Moreover, we do not know the sampling design behind the CAW data with respect to the population of interest. Thus, as we discuss in detail below, CAW data may both under or over-estimate

true persistence. Furthermore, we can only produce estimates for ρ_τ, i.e., we estimate persistence for those that are ultra wealthy U_k. The second issue is more subtle: we can only estimate ρ_τ conditionally on being able to link individuals to the genealogical data.

To start with lets abstract from sampling considerations, i.e. assume that if one is hyperwealthy $Y^*_{it} \equiv \mathbb{1}[i \in U^*_t] = 1$, then one does appear in the CAW data $Y_{it} = Y^*_{it}$. Here we face a sample selection problem driven by non-random linking to the genealogical data. Without further assumptions, the best we can identify is the joint likelihood of being linked to the genealogical data and having a hyperwealthy descendent conditional on oneself being hyperwealthy. Assuming that CAW is a sample of the hyperwealthy, it is trivial to see that the object of interest,[7] i.e. having a hyperwealthy descendent, is lower bounded by the joint likelihood of persistence and being linked to the genealogical data. We report this lower bound in Panel A, *Raw Estimates*, of Table C.11 in the Appendix.

We can recover point identification if we have conditional independence. Formally, we can recover point identification if there is a set of observable characteristics, X_i, such that the likelihood of having a hyperwealthy descendent, $Y^*_{it+\tau} = 1$, conditional on one being hyperwealthy, $Y^*_{it} = 1$, is independent of linkage to the genealogical data, $L_{it} = 1$, conditional on set of observables, X_{it}. In such case, the object of interest can be recovered by inverse probability sampling weighting as per equation 1. Figure D.3 represents balance of observed characteristics between those linked to the genealogical data and those that have not being linked before and after re-weighting.

$$\begin{aligned}
\mathbb{E}[Y^*_{i,t+\tau}|Y^*_{it} = 1] =& \mathbb{E}\left[\frac{Y^*_{it+\tau}L^*_{it}}{\mathbb{E}[L^*_{it}|Y^*_{it} = 1, X_{it}]}\middle| Y^*_{it} = 1\right] \\
=& \mathbb{E}\left[\frac{Y^*_{it+\tau}L_{it}}{\mathbb{E}[L_{it}|Y^*_{it} = 1, X_{it}]}\middle| Y_{it} = 1\right] \mathbb{E}\left[\frac{L_{it}}{\mathbb{E}[L_{it}|Y^*_{it} = 1, X_{it}]}\middle| Y^*_{it} = 1\right]^{-1}
\end{aligned} \tag{1}$$

[7] There is some non-zero likelihood that an individual we failed to link to the genealogical data has a wealthy descendent. In this discussion, we assume that the likelihood of wrongfully linking a CAW individual to a non-CAW individual who has a hyperwealthy descendent contained in CAW is negligible.

A second source of bias arises due to cross-sectional sampling; i.e., the CAW does not sample every hyperwealthy individual who is a descendant of the previous sample of hyperwealthy individuals. Random sampling from cross-sections induces a downward bias in persistence. This can be seen clearly in the example in Figure 3, where we move from a true persistence of 1 to 0.5 because of random sampling from the cross-section.

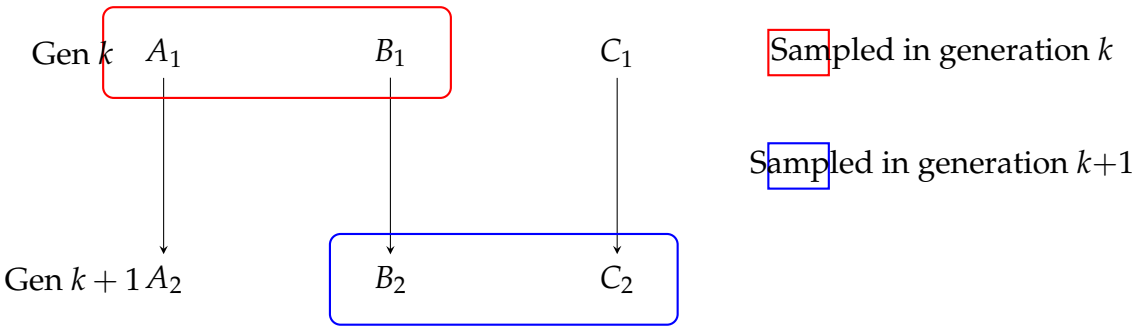


FIGURE 3. Bias from Cross-Sectional Sampling

Even with complete linkage to the genealogical data and with representative cross-sectional samples, we will underestimate the extent of true persistence by looking at the likelihood of individuals being linked to future cross-sectional samples of wealthy individuals. In our empirical exercise we produce two exercises that should mitigate this point. First, we produce estimates of the likelihood of being linked to an US Congress member, for whom we have near population level data. Second, we produce estimates of elite persistence by measuring expected wealth from Wikipedia bios, in this exercise no inclusion in Wikipedia is a signal of non-eliteness.

The last source of bias we consider is that of non-random sampling with respect to outcomes of interest, i.e. that the CAW over/under-samples persistent dynasties. To be able to address this source of bias, we impose an additional sufficiency condition, namely that there is a set of observable characteristics, X_{it}, such that persistence among wealthy individuals in CAW, i.e. all i such that $Y_{it} = 1$, conditional on X_{it} is the same as for all hyperwealthy individuals $Y^*_{it} = 1$. For persistence estimates, Table C.11 Panel D and

Table C.13 Panel D, we also provide estimates where we re-weight the data to have the same distribution of observables as the 1870 Census.

4 Definition of Kinship and Linking CAW to Genealogical Data

We infer persistence by measuring whether a wealthy individual in CAW can be linked through vertical, parent-child, relationships into a future CAW cross-section. Moreover, we also study the role that extended kinship plays in maintaining socio-economic advantage. Following Collado et al. (2023), we define the extended kinship as those that can be reached through sibling-sibling and spouse-spouse relations within at most five steps from the root individual or from an individual in the vertical kinship of the root. These definitions are exemplified in Figure 4 where we represent the genealogy of John Davidson Rockefeller Sr (1839-1937) obtained from WikiTree. In Figure 4, the descendents of John Rockefeller, the black node, are represented as blue nodes. John Rockefeller's extended kinship members are represented in green, with the exception of spouses of both John Rockefeller and his descendents who are represented in yellow. This figure shows that in the genealogical tree of John Rockefeller there is substantial wealth to which he is connected through extended kinship relations.

We obtain genealogical data from three sources: Wikidata, WikiTree and Find A Grave. For this, we need to first identify entries in these datasets that correspond to entries in the CAW data. We do this sequentially to take advantage that for some entries Wikidata reports identifiers for WikiTree and Find A Grave. This allow us to use the Wikidata data to train and test a link discriminator for WikiTree and Find A Grave. The first step we take in this direction is, therefore, to link CAW to Wikidata. We have carried this linkage by manually searching Wikidata for every name contained in CAW, checking that birth/death dates in CAW and Wikidata are consistent and using information about residence and source of wealth to discriminate between linkages. At the end of this process we are able to link 2,932, (62.96%) of entries in CAW to Wikidata.

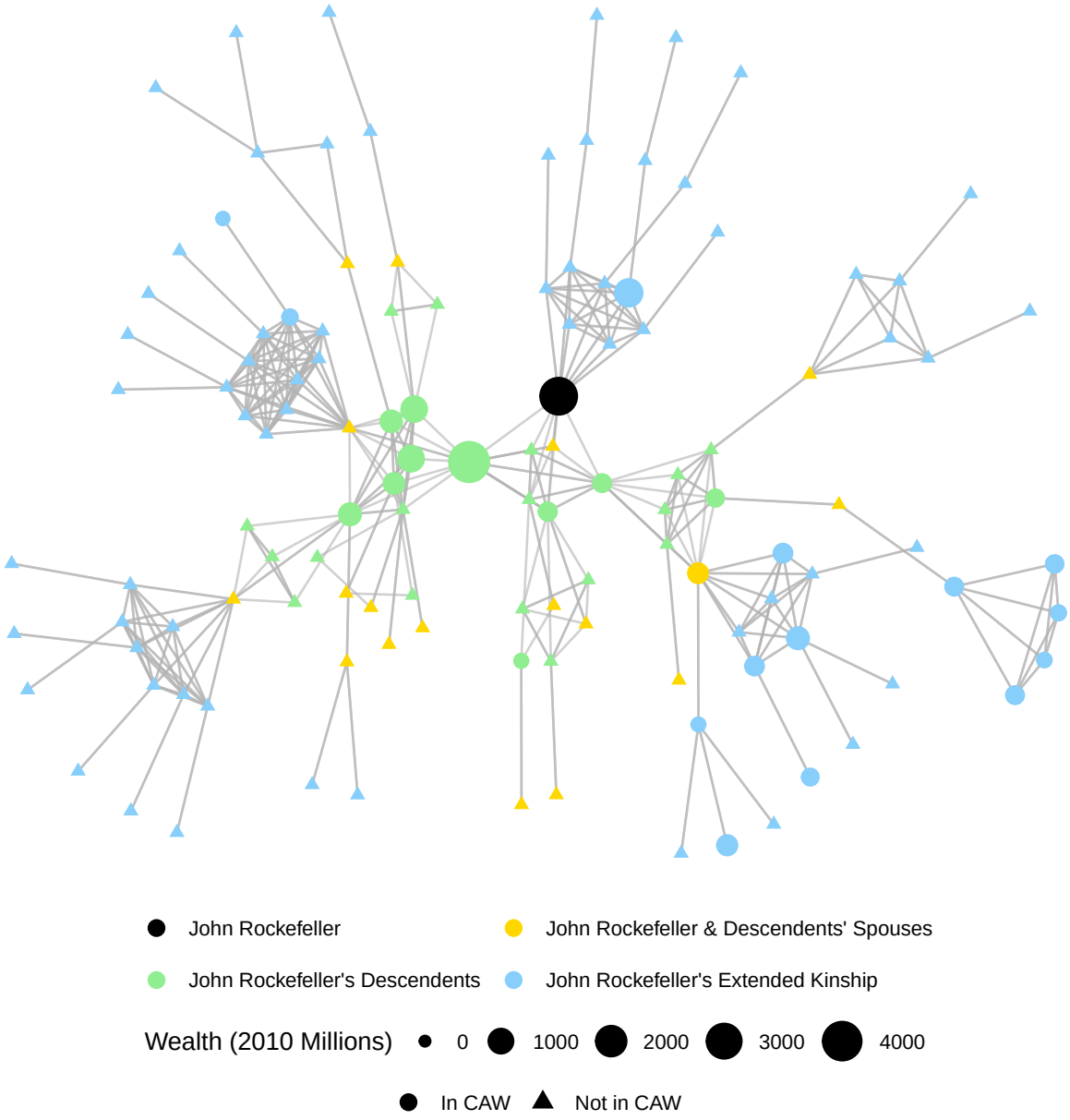


FIGURE 4. John Davison Rockefeller Sr's WikiTree Genealogy

Notes: Figure provides the genealogy of John Davison Rockefeller Sr (1839 - 1937), black node, that we extracted from WikiTree. For compactness we limit the extended kinship to those at most three steps from either John Davison Rockefeller or his descendents, instead of the five steps used through the paper. John Davison Rockefeller's descendents are represented with green nodes. Spouses of both John Rockefeller and his descendents are represented with yellow nodes. Members of John Rockefeller's extended kinship are represented in blue. Nodes belonging to individuals contained in the CAW data are represented with circles, while those not in the CAW data are represented with triangles. Larger nodes indicate larger individual wealth.

Figure D.2 reports differences in terms of observables between CAW observation linked to Wikidata and those that have not being linked. Given that Wikidata tends to reflect notable, in some dimension, individuals it is intuitive to find that those with higher wealth are more likely to be linked. Males are also more likely to be linked to Wikidata than females, reflecting a bias in Wikidata towards notable males as compared to notable females that has been documented in the literature (Wagner et al., 2016; Zhang and Terveen, 2021). Somehow counter-intuitively, we find that those that are reported in CAW as to have inherited their wealth are less likely to be linked to Wikidata.

As mentioned earlier, from Wikidata we do not only obtain genealogical data but also links to WikiTree and Find A Grave. Of those linked to Wikidata, 1,858 (63.37%) of them have a link to Find A Grave and 2,067 (70.50%) have a link to WikiTree. We use these links to train a discriminator to find unique matches between CAW and WikiTree/Find A Grave.

Before we do so, we need to obtain candidate links from both datasets. We do so by sending request to the websites hosting these datasets. We automatically break full-name strings in the CAW data into names and first names, where we use the position in the string along with relative frequencies in the 1990 US Census to decide whether a word, token for now on, within a full name string is a name or a surname. Furthermore, we have collected a set of common nicknames from various sources that we attach to every name and surname. Once we have broken full names into name and surname tokens, we send a search request for every possible triad composed by a name token, surname token and year of birth. We limit searches to those that are within 5 years of the date of birth reported in CAW and allow non-exact matching on name and surname strings. If this search returns no candidate then we substitute the name and surname, with known nicknames or aliases and send a new query. At the end of this process we have for every individual in CAW a set of non-unique candidate links in both WikiTree and Find A Grave, among which we want to find the true link.

We decide what is the best link using a Random Forest discriminator that we train on the links reported by Wikidata. For individuals linked to Wikidata for which Wikidata reports a link to one of the other two datasets, we have a set of candidate links and

know which one of them is the true one. With this information we predict the likelihood of a candidate link being the true link, where predictions are made based on weighted string distances[8] and distances between CAW birth/death year and the birth/death year reported in the target genealogical dataset. To this end, we choose hyper-parameters through cross-validation. Tables C.8 and C.9 report linking errors from our discriminator in the testing dataset. We have manually chosen thresholds on the estimated likelihood of being a true match for acceptance so that we minimise the likelihood of choosing a false match when a true match does not exist while preserving a high likelihood of choosing the true match when a true match exists. For this, we select a candidate link if the largest estimated probability of being a true link is larger than 80% and this probability is larger than the second best candidate link by at least 10 percentage points. Figures D.6 and D.7 provide the density for the likelihood of the first and second best candidate link being the true match in both the WikiTree and Find A Grave data, respectively.

Table C.3 provides overall linking rates to genealogy data by cohort. Compared with other papers linking historical cross-sections, we do quite well; mostly considering that the target genealogical datasets are most likely non-comprehensive. The high linking rate we obtain is most likely due to the elite nature of the individuals under consideration.

Figure D.3 reports differences in terms of observables between CAW observation linked to any genealogical source and those that have not linked. These show patterns similar to those observed for the likelihood of being manually linked to Wikidata.

[8]For two strings s_1 and s_2 that may contain multiple tokens, we compute the Jaro-Winkler distance between every token in s_1 and every token in s_2. We then rank all these dyads by descending distance, and we select the token dyad with the smallest distance as a match, say (t_i^1, t_j^2). After this we extract from both t_i^1 from s^1 and t_j^2 from s^2 and repeat the process. We do this until no token is left in s^1, if s^1 contains more tokens than s^2 then those additional tokens are given a maximum distance of one. Let $d(t_i^1|s_1, s_2)$ be the distance for token t_i^1, then we compute the distance between string s_1 and s_2 as $\frac{\sum_{t \in s_1} d(t|s_1, s_2)\omega(t)}{\sum_{t \in s_1} \omega(t)}$ where $\omega(t)$ is the inverse of the frequency of token t among the names of all candidates. This over-weights unusual name tokens in the construction of distances when a name contains multiple tokens. For example, Carol Averell Harriman married Richard Penn Smith on 11th July 1917, upon marriage she takes the very common Smith surname. Still in the CAW data she appears as Carol Averell (Harriman) Smith Stewart. The Harriman token is more informative than the Smith token. Say that there are two candidate links one Carol Averell Smith Stewart and a Carol Averell Harriman Stewart, the distance we use will prescribe the second as being closer to Carol Averell (Harriman) Smith Stewart.

Once we have linked an individual from CAW into one of the genealogies we do snowball sampling to retrieve their genealogies. Thus, for a given individual we sample their parents, spouses, siblings and children. For the siblings and children we found in the first step we also sample their parents, spouses, siblings and children. We iterate this process until no further individual is found. After this, we select all spouses in the resulting dataset and search for their parents, siblings, spouses and children. This sampling, however, it is only done once, for example we do not search for the spouse of the sibling of a spouse. We impose this limitation because sampling individuals from the genealogical data is time consuming and sampling relations horizontally without constraints may lead to very big and time consuming genealogical trees. Finally, we use the tree obtained from these two steps to grow the genealogical tree horizontally. In particular, first, we select all the wealthy roots and their children. Second, we select all siblings, spouses and siblings of spouses that have already being sampled and are within five degrees of separation of the wealthy individuals and their children. Third, for everyone selected in the first and second step we search for their spouses and children. We recur this third step for as long as we find an individual that is within five degrees of separation of a rich individual or their children.

At the end of this process we have obtained genealogical trees for 3,720 (81%) entries in the CAW data. Before we do analysis on them we prune the trees so that there are no inconsistencies in terms of dates, for example children being born before their parents or much after their parents death, and that they have desirable transitivity properties, such as a child not being able to rich their parents through parent-child walks. How we exactly prune the genealogical trees is described in detail in Appendix A. Table 3 provides descriptive statistics for the vertical genealogies of linked CAW individuals by cohort. Table C.7 in the appendix provides analogous statistics for the extended kinship.

TABLE 3. Vertical Kinship Connections Among CAW Individuals

	Wikidata				WikiTree				Find A Grave			
	Avg.	S.D.	Min.	Max.	Avg.	S.D.	Min.	Max.	Avg.	S.D.	Min.	Max.
1875 Cohort												
Any Link	0.65	0.48	0	1	0.84	0.36	0	1	0.88	0.33	0	1
# 1° Neigh.	1.38	1.85	0	12	2.43	2.85	0	22	3.58	3.07	0	19
# 2° Neigh.	1.72	2.89	0	20	3.30	4.90	0	34	5.50	6.13	0	43
# All Neigh.	11.27	31.42	0	318	10.71	17.66	0	173	19.68	27.77	0	289
Avg. Distance	2.26	1.07	1	6	2.02	0.64	1	4	2.18	0.59	1	4
Max. Distance	3.44	1.84	1	8	3.22	1.34	1	6	3.60	1.16	1	6
1900 Cohort												
Any Link	0.55	0.50	0	1	0.74	0.44	0	1	0.82	0.39	0	1
# 1° Neigh.	0.93	1.47	0	11	1.60	2.23	0	14	2.67	2.62	0	22
# 2° Neigh.	0.97	2.14	0	21	1.68	3.29	0	36	3.64	4.78	0	31
# All Neigh.	4.98	16.63	0	309	4.42	7.75	0	85	9.47	12.89	0	116
Avg. Distance	1.87	0.88	1	5	1.62	0.51	1	3	1.79	0.49	1	3
Max. Distance	2.71	1.55	1	7	2.39	1.12	1	6	2.79	1.05	1	6
1925 Cohort												
Any Link	0.49	0.50	0	1	0.69	0.46	0	1	0.77	0.42	0	1
# 1° Neigh.	0.75	1.27	0	10	1.13	1.71	0	12	1.99	2.01	0	11
# 2° Neigh.	0.78	2.25	0	31	0.71	1.84	0	21	1.91	3.10	0	25
# All Neigh.	3.09	9.60	0	111	2.06	3.81	0	36	4.63	6.73	0	84
Avg. Distance	1.64	0.72	1	4	1.34	0.39	1	3	1.48	0.39	1	3
Max. Distance	2.25	1.27	1	6	1.77	0.86	1	5	2.12	0.86	1	5
1950 Cohort												
Any Link	0.47	0.50	0	1	0.54	0.50	0	1	0.65	0.48	0	1
# 1° Neigh.	0.80	1.43	0	11	0.73	1.37	0	10	1.30	1.68	0	11
# 2° Neigh.	0.89	2.74	0	31	0.29	1.15	0	19	0.68	1.60	0	11
# All Neigh.	2.60	7.74	0	88	1.11	2.64	0	35	2.07	3.16	0	21
Avg. Distance	1.46	0.56	1	3	1.20	0.31	1	2	1.24	0.30	1	2
Max. Distance	1.93	1.05	1	5	1.51	0.72	1	4	1.59	0.71	1	4

Notes: Table provides descriptive statistics for the vertical kinship connections of individuals in the CAW data. Column *Avg.* provides averages, *S.D.* standard deviations, *Min.* minimums and *Max.* maximums. *Avg. Distance* and *Max.Distance* provide individuals' average and maximum family distances descriptive statistics for those with at least one link.

5 Results

We start by documenting the likelihood of persistence in the CAW, i.e. the probability that one has a vertical descendent in a future cohort of the CAW data given that one appears in the CAW data.

The persistence estimates in Table 4 reveal several notable patterns. First, persistence probabilities are substantial, particularly for the early cohort, we find one generation ahead persistence estimates that range from 32% for 1875 cohort to 16% for 1925 cohort. Second, persistence rates decline considerably across generations. Comparing one generation ahead persistence with two generations we observe on average a decline of 12% across cohorts, where we select only cohorts for which both 25 and 50 year ahead persistence is observed. This decline is driven by the 1900 cohort. For the 1875 cohort, we observe persistence increasing when we compare 25 years ahead persistence with 50 years ahead persistence. In appendix, Table C.11 provides our baseline estimates along with raw estimates and re-weighted estimates to have the same distribution of observed characteristics as the 1875 CAW and 1870 Census. With exception of the latter, our estimates are robust to reweighing. Undersampling of the less wealthy within the very wealthy is a reason behind the 1870 Census weights producing lower persistence, see Figure D.1.

Persistence estimates in Table 4 may be difficult to interpret in the sense that the size and share of national wealth contained in the CAW data, see Figure 2, varies across cohorts. To address this point, we provide two exercises. In Table C.10, we compute persistence for individuals at the top 400, thus emulating the modern sample size of Forbes. We then estimate the likelihood of one of these individuals having a vertical descendent in the top 400 at 25, 50 and 75 years later. These estimates show an average persistence at 25 years of 17%. Furthermore, differently from Table 4 we do not observe much of a decline across cohorts. For the 1875 cohort, we still observe an increase between 25 and 50 years ahead persistence, but this is now not statistically significant with 95% confidence intervals almost fully overlapping. In a similar vein, Table 5 refines persistence estimates of Table 4 by splitting both roots and persistence in terms of narrow wealth groups. In particular, we categorise every individual in a given cohort by whether they belong in the

top 100 wealthies of a given cohort, the 101-200, 201-300, 301-400, to those below the top 400 or whether they do not appear in the CAW data.

TABLE 4. Vertical Kinship Persistence

	Root Cohort		
	1875	**1900**	**1925**
1900	0.323		
	[0.284,0.363]		
1925	0.338	0.214	
	[0.297,0.379]	[0.191,0.237]	
1950	0.152	0.153	0.161
	[0.121,0.183]	[0.133,0.173]	[0.141,0.180]
Observations	585	1,402	1,567
Share with Genealogy	0.867	0.815	0.805

Notes: Table reports the likelihood of a root having a vertical descendant in a future CAW cohort estimated from those roots that we have been able to link to genealogical data and re-weighting to account for selection induced by said linkage.

Estimates in Table 5 show that there is substantial heterogeneity at the very top. Of those root individuals at the top 100 about 49% of them have a vertical descendent in the next cohort of the CAW data. This is in contrast to 14% for roots below the 400 wealthiest in CAW. There is also inequality in terms of the destination of wealth group for vertical descendents: 14% of roots in the top 100 have a vertical descendent in the top 100 in the next CAW cohort, this is in contrast to 3% for roots in the top 200. When we look at 50 years ahead persistence, Panel B of Table C.12, we observe broadly comparable patterns.

Estimates in Tables 4 and 5 broadly align with those reported in Kalsi and Ward (2025) for the US and Ozkan et al. (2023) for Norway. Based on Census data, Kalsi and Ward (2025) provide evidence of strong non-linearities between initial wealth and persistence at the top 1%: for grandparents at the of the wealth distribution the probability of having a grandchild at the top 1% is about 14%, see Figure 11 Panel B in Kalsi and Ward (2025). This is remarkably similar to the persistence likelihood that we find when looking at 50

year gaps and transitions from top-100 to top-100, Panel B of Table C.12. Furthermore, in line with Kalsi and Ward (2025) we also find that at the very top persistence is still strongly monotonically increasing on initial wealth. It is reassuring that the CAW data produces persistence patterns that are broadly comparable to those obtained with the Census.

TABLE 5. Vertical Kinship Persistence by Wealth Group

Target Wealth Group	**Root Wealth Group**				
	1-100	**101-200**	**201-300**	**301-400**	**400+**
Panel A: 25 Years Ahead					
1-100	**0.143**	0.033	**0.017**	0.011	0.008
	[0.103,0.183]	[0.012,0.054]	[0.002,0.032]	[-0.001,0.024]	[0.004,0.011]
101-200	0.091	**0.047**	0.014	**0.014**	**0.010**
	[0.058,0.124]	[0.022,0.072]	[0.000,0.028]	[0.000,0.028]	[0.006,0.015]
201-300	0.075	0.039	0.026	0.012	0.011
	[0.045,0.105]	[0.017,0.062]	[0.007,0.044]	[-0.002,0.025]	[0.006,0.016]
301-400	0.055	0.053	0.022	0.016	**0.010**
	[0.029,0.081]	[0.027,0.079]	[0.005,0.039]	[0.000,0.032]	[0.005,0.014]
400+	0.130	0.199	0.223	0.169	0.098
	[0.092,0.169]	[0.152,0.246]	[0.174,0.272]	[0.124,0.215]	[0.084,0.111]
Broadly Wealthy	0.494	0.372	0.302	0.223	0.136
	[0.436,0.551]	[0.315,0.428]	[0.247,0.356]	[0.172,0.273]	[0.120,0.151]
Observations	300	300	300	300	2,354
Share with Genealogy	0.973	0.927	0.913	0.863	0.768

Notes: Table provides wealth group transition probabilities at 25 year gaps conditional on the starting wealth group. For every column and panel we highlight the largest persistence within detailed wealth groups in bold. Below the dashed line we report persistence into a residual wealthy group, below the solid horizontal line we report overall persistence. Data has been re-weighted to account for selection induced by the linkage of the CAW data to the genealogical data. 95% confidence between square brackets account for pre-estimation of these weights, some cells are empty and for these we report a zero with zero length confidence intervals. *Share with Genealogy* reports the share of observations, by baseline wealth group, that we have been able to link to at least one genealogical source.

6 The Role of Extended Kinship

Our main departure from the existing literature, and in particular from Kalsi and Ward (2025), is that we focus on extended kinship mechanisms leading to elite persistence. Extended kinship has been shown to be important to maintain elite status (Padgett and Ansell, 1993). Here, we explore the role that extended kinship plays in shaping intergenerational persistence among the top wealthy. In Table 6, we start by replicating the vertical persistence estimates reported in Table 4 but using extended kinship relations instead of vertical relations. An individual is within the extended kinship of another individual, if they are within five horizontal degrees of separation, i.e. sibling-sibling or spouse-spouse, from a vertical descendent of the focal individual or the focal individual themself.

Estimates in Table 6 show that the likelihood of being connected to a wealthy individual in a future cohort of the CAW data through extended kinship is significantly higher than through vertical kinship. When we look at 25 years ahead, we find that about 40% of individuals can be linked to a future wealthy individual through extended kinship, see Panel A of Table C.13. This estimated persistence increases to 48% when we exclude individuals who we have not been able to link to the genealogical data and re-weight the data to have the same distribution of observables between the linked and non-linked subsamples, see Table 6.

Similar to the patterns observed with vertical persistence we also observe that persistence through extended kinship has declined over time. As exemplified by Table C.13 in Appendix, this is true even when we re-weight the data so that all cohorts have the same distribution of covariate as the 1875 cohort or the 1870 Census. We also observe extended kinship persistence declining with the length of the persistence window when we limit our analysis to the top 400, see Table C.14 in the Appendix.

All results in this section suggests that wealthy individuals net an interconnected web of social relations among them through extended kinship. To provide further evidence of who connects with whom and whether these connections give any advantage in terms of wealth persistence we now focus on the extended kinship of the children of a wealthy individual. In particular, we focus on children connectivity measured as the betweenness

TABLE 6. Extended Kinship Persistence

	Root Cohort		
	1875	**1900**	**1925**
1900	0.538		
	[0.494,0.582]		
1925	0.487	0.511	
	[0.443,0.532]	[0.482,0.540]	
1950	0.385	0.329	0.438
	[0.343,0.428]	[0.302,0.356]	[0.410,0.465]
Observations	585	1,402	1,567
Share with Genealogy	0.867	0.815	0.805

Notes: The table reports the likelihood of a root having an extended kinship descendant in a future CAW cohort estimated from those roots that we have been able to link to genealogical data and re-weighting to account for selection induced by said linkage

of roots' children when considering shortest paths between individuals belonging in the CAW cohort of the root.

To be more specific on what connectivity means. For a given CAW cohort we select all wealthy roots. We then select all their children and all extended kinship of their children that are within five degrees of separation. For each children and children's kinship node we compute their betweenness restricting shortest paths to those going from and to wealthy roots in the given cohort. We use undirected shortest paths so that paths can traverse parent-child relations in both directions. We measure connectivity for every cohort and for every genealogical source separately. Furthermore, within source and cohort we standardize betweenness by mapping it into the unit interval. Finally, we aggregate betweenness to the cohort-node level by taking averages for each individual and cohort across genealogical sources.

In Figure 5, we illustrate the kind of connectivity we measure using John D. Rockefeller's (1839-1937) children extended kinship and the other 1875 CAW individuals that he is linked to. In that figure, we have three wealthy roots that appear in the 1875 CAW: McLean with $299M wealth, McCornick with $898M wealth and Rochefeller with $598M wealth, all in 2010 USD. Rockefeller has five children, four women and a man. One of his daughters, Edith (Rockefeller) McCormick, married to Harold Fowler McCormick, the second to youngest children of Cyrus Hall McCormick. This marriage places both Edith and Harold in the shortest path between two 1875 roots: their respective parents John Davison Rockefeller and Cyrus Hall McCormick. This connection, however, could be in-

ferred from the vertical kinship of these two roots, as Edith and Harold had five children together.

Figure 5 exemplifies the importance of extended kinship in connecting elite families through the connections between McCormick and McLean. These two individuals, both in the 1875 CAW, are not directly linked through their vertical kinship, but rather through the extended kinship of their children. McCormick has seven children, one of them, Anita (McCormick) Blaine marries Emmons Blaine. Emmons' sister, Harriet Stanwood (Blaine) Beale, is married to Truxtun Beale whose sister, Emily Truxton (Beale) McLean, is married to John Roll McLean, the second children of Washington McLean. Thus, through a sequence of marriages both, McLean and McCormick are connected through their extended kinship.[9]

A final point that Figure 5 exemplifies is that there is substantial heterogeneity in connectivity even within narrowly defined families. Take the children of McCormick. Out of his seven children, only two of them are well-connected. For Rockefeller and McLean only one of their respective children is well-connected. This is despite some of their other, worst connected, children having large extended kinships. This example suggest that connectivity may be scarce, and, if it is an important feature for elite persistence, there may be a market for it on which the wealthiest among the wealthiest may have an advantage (see Goñi, 2022). This is a point we investigate systematically in the next section.

6.1 Strategic Family Formation

Table 7 reports correlations between root's children connectivity and root's characteristics. First, even among the very wealthy, wealthier individuals generate better connected offspring. This positive association between children's mediating role among wealthy individuals in the previous CAW cohort and their parents' wealth is stable across specifications, including one where we control for the whether their grand-father was in the CAW data two cohorts prior. Moreover, we see that female roots have an advantage in generat-

[9] Table C.15 in the Appendix provides the names of the top ten best connected individuals in each CAW cohort.

ing well-connected children and that advantage is also stable across specifications. Similarly, female advantage in generating social connections is also carried by their children: wealthy roots' who have a larger share of female children have better connected children. We also observe that rich roots with more children, conditional on having any children in the genealogical data, generate better connected children. Finally, in column (4) of Table 7 we report correlations between roots' children connectivity and a dummy for whether roots' parents were in the previous CAW cohort. We find a positive, albeit marginally significant, correlation between these two variables, suggesting that wealthy dynasties are more successful at connecting their children, even conditional on roots' wealth and children structure.

Estimates in Table 7 provide suggestive evidence of strategic marriage, with wealthier individuals more successful in generating better connected offspring. We are interested in what are the returns of such connections in terms of elite persistence. In this direction, estimates in Table 8 provide partial correlations between various measures of elite persistence roots' wealth and roots' children connectivity. All estimates point in the same direction: dynasties originating from wealthier individuals with better connected children are more likely to stay wealthy 25 years ahead. We find higher persistence if we measure both whether a vertical descendent is included in future CAW cohorts, Panel A, and the total wealth of vertical descendents in future CAW cohorts, Panel B, where roots' without any vertical descendents in future CAW cohorts are coded as zeros. Estimates are remarkably stable to the inclusion of controls for children characteristics and the size of children's extended kinship. We find similar estimates when we look at persistence outcomes with a 50 year gap, see Table C.16 in the Appendix.

Panel E of Table 8 introduces a new margin of eliteness, whether a roots' children has any extended kinship connection seating in Congress between the CAW cohort on which the wealth of the root is measured and the next 25 years. To code this outcome, we have linked the Congress genealogies of Dal Bó et al. (2009) to Wikidata, WikiTree and FindAGrave in an analogous manner in which we linked the CAW data to these genealogical sources. This margin of elite persistence has the advantage that the data on Congress members is, almost, universal, and it should contain everyone who has held a

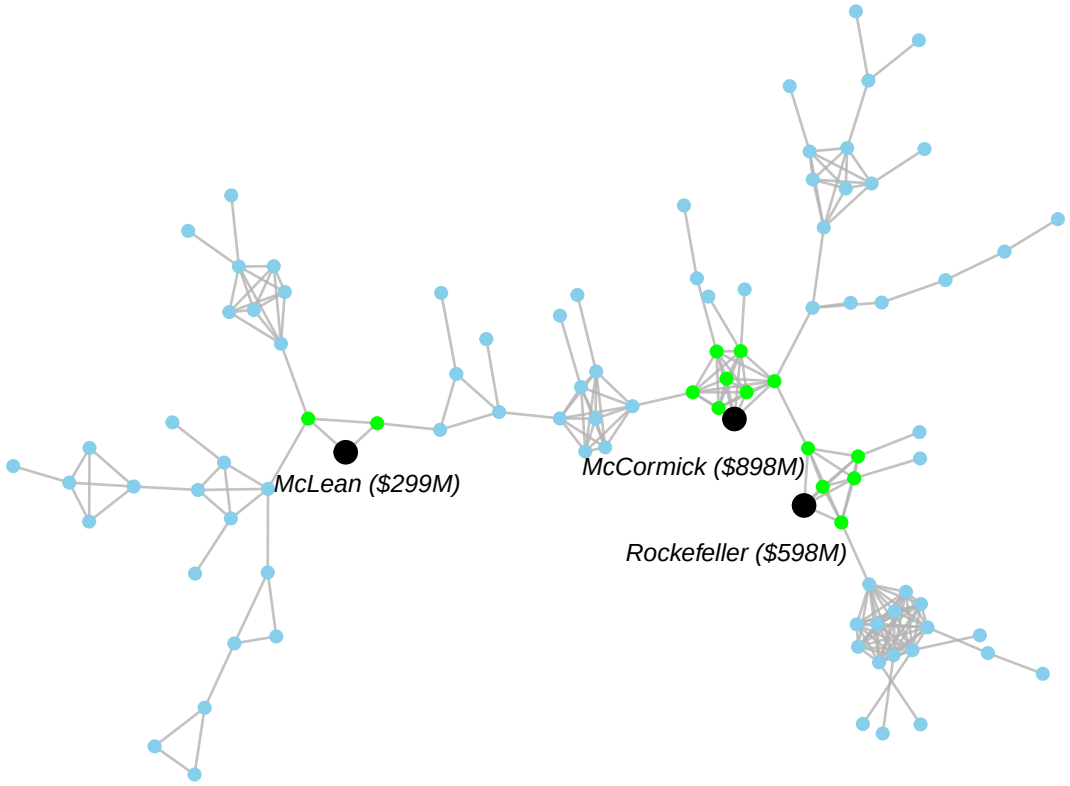
TABLE 7. Wealth Connectivity and Root's Wealth and Realised Wealth Persistence

	(1)	(2)	(3)	(4)
(log-)Wealth	0.133***	0.128***	0.107***	0.097***
	(0.020)	(0.020)	(0.020)	(0.023)
Inheritance	0.227***	0.237***	0.240***	0.178***
	(0.050)	(0.050)	(0.048)	(0.053)
Parent in Previous CAW				0.342***
				(0.044)
F-stat	29.469	28.970	23.807	36.152
F-stat (p-value)	0.000	0.000	0.000	0.000
Observations	3,554	3,554	3,554	2,969
Included Controls:				
Any Children in Gen.Data	✓	✓	✓	✓
Number of Children		✓	✓	✓
Children Sex Composition		✓	✓	✓
Kinship Size			✓	✓

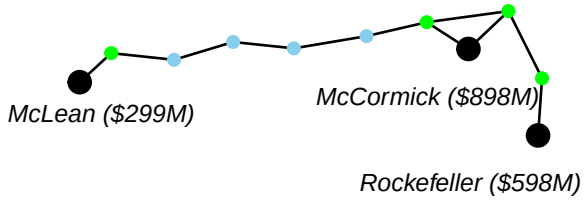
Notes: Table reports partial correlations between root's children connectivity and root's characteristics from OLS regressions where we re-weight the data to account for sample selections when linking to the genealogical data. *Parent in Previous CAW* is a dummy taking value one if at least one of the root's parents appears in the previous CAW cohort, as such is not available for roots only appearing in the first CAW cohort. F-stat reports the F-statistic for joint significance of reported coefficients. Wealth connectivity has been standardised to have mean zero and variance one. All specifications include controls for year, age (quadratic), regions of residence, industry, sex, whether they inherited wealth and a dummy for whether any children is present.

seat over the 1875-1994 period. This avoids biases in estimated persistence when we have only random samples from the cross-section of elite members, as discussed in Section 3. Estimates in Panel E show that both roots' wealth and roots' children connections are also positively correlated with the likelihood of having extended kinship connections in Congress. This further supports the CAW based evidence in that both wealth and family connections are important in determining elite persistence, with the added margin that here we observe a jump into a new sphere of eliteness, the political elite.

Dal Bó et al. (2009) coded a number of Congress members' characteristics, including whether congress members had a family member in congress before or after them, their political alignment, whether they stay in congress for more than one term and whether they have won a re-election. In Panel B of Table 9, we compare these characteristics for Congress members who belong in the extended kinship of a CAW individual and those who do not. Estimates in that table show that Congress members linked through extended kinship to a CAW individual are disproportionally likely to come from a dynasty of Congress members, in the sense of having a family member who has seated in Congress before them. They are also disproportionally likely to have a family member that will seat in Congress after them. Thus, evidence in Tables 7 and **??** suggest that well connected children helps the wealthy elite to create family connections to the dynasties in the political elite.



(A) Weakly Connected Component



(B) Shortest Paths Among Roots

FIGURE 5. John Davison Rockefeller Sr's Connections through Children's Kinship in 1875 CAW, source WikiTree

Notes: Panel A of figure represents the weakly connected component in the network defined by 1875 CAW roots, their children and their children's extended kinship to which John Davison Rockefeller Sr (1839 - 1937) belongs to. Black nodes are wealthy roots included in the 1875 CAW cohort, green nodes are their children and blue nodes are their children's extended kinship. Panel B of figure displays shortest paths among 1875 CAW roots together with nodes along these paths. Nodes along these paths, excluding root nodes, have strictly positive connectivity, while those outside have zero connectivity. Wealth in 2010 USD between parenthesis.

TABLE 8. Children Connections: Wealth and Elite Persistence

	(1)	(2)	(3)
Panel A. Descendent in CAW 25 years ahead			
(log-)Wealth	0.374***	0.374***	0.369***
	(0.037)	(0.037)	(0.037)
Wealth Connectivity	0.298***	0.298***	0.270***
	(0.038)	(0.038)	(0.040)
Average		0.209	
Observations		3,554	
Panel A. Descendent in top 400 CAW 25 years ahead			
(log-)Wealth	0.887***	0.887***	0.883***
	(0.080)	(0.080)	(0.080)
Wealth Connectivity	0.441***	0.441***	0.418***
	(0.074)	(0.074)	(0.075)
Average		0.083	
Observations		3,554	
Panel B. Descendent wealth 25 years ahead			
(log-)Wealth	1.276***	1.276***	1.272***
	(0.226)	(0.226)	(0.227)
Wealth Connectivity	0.784***	0.784***	0.761***
	(0.220)	(0.220)	(0.229)
Average		7,887,275.967	
Observations		3,554	
Included Controls:			
Any Children in Gen.Data	✓	✓	✓
Number of Children		✓	✓
Children Sex Composition		✓	✓
Kinship Size			✓

Notes: Table reports partial correlations between various measures of elite persistence and roots' wealth and connectivity of their children. Estimates are produced by OLS re-weighting the data to account for sample selection in genealogical linking. Panel A measures elite persistence by whether the root as a vertical descendent in the next, 25 year ahead, CAW cohort. Panel B has analogous time horizon but we substitute a dummy with the total wealth held by a root vertical descendents. Panel C measures elite persistence by coding whether any of a root's children extended kinship has hold a seat in Congress within the next 25 years. All outcomes are divided by their respective averages reported under *Average*. Robust standard errors reported between parenthesis account for the pre-estimation of weights to account for selection in genealogical linking. *Children* includes controls for the number of children, including a dummy for whether any children is included in the genealogy data, and their sex composition. *Kinship size* includes control for children's extended kinship size, including a dummy for whether any extended kinship is present in the genealogical data.

TABLE 9. Wealth, Connectivity and Political Connections

	(1)	(2)	(3)
Panel A. Links to US Congress			
(log-)Wealth	0.209***	0.209***	0.202***
	(0.043)	(0.043)	(0.043)
Wealth Connectivity	0.746***	0.746***	0.707***
	(0.053)	(0.053)	(0.057)
Average		0.166	
Observations		3,554	
Included Controls:			
Any Children in Gen.Data	✓	✓	✓
Number of Children		✓	✓
Children Sex Composition		✓	✓
Kinship Size			✓

Panel B. Characteristics of Wealth Connected Congress Members	**Wealth Connected**		**Difference**
	Yes	**No**	
Democrat	0.380	0.535	-0.156***
			(0.019)
Republican	0.620	0.457	0.164***
			(0.019)
Longterm	0.893	0.917	-0.024**
			(0.012)
Postrelative	0.281	0.075	0.206***
			(0.017)
Prerelative	0.388	0.076	0.312***
			(0.019)
Win	0.859	0.883	-0.024
			(0.021)
Members	163	6,078	
Seats	693	25,130	

Notes: Table provides correlations between wealth and wealth connectivity using OLS, columns (1) and (2), and reduced form estimates from an IV strategy where we exploit connectivity shocks to distance extended kinship, columns (3) and (4). All specifications include controls for year, age (quadratic), regions of residence, industry, sex and whether they inherited wealth. Estimates in Panel A come from inverse probability weighted specifications to account for selection induced by linkage to the genealogical data. Standard errors provided between parenthesis account for the first stage estimation of said weights.

6.2 Exploiting Shocks to Higher Order Connections

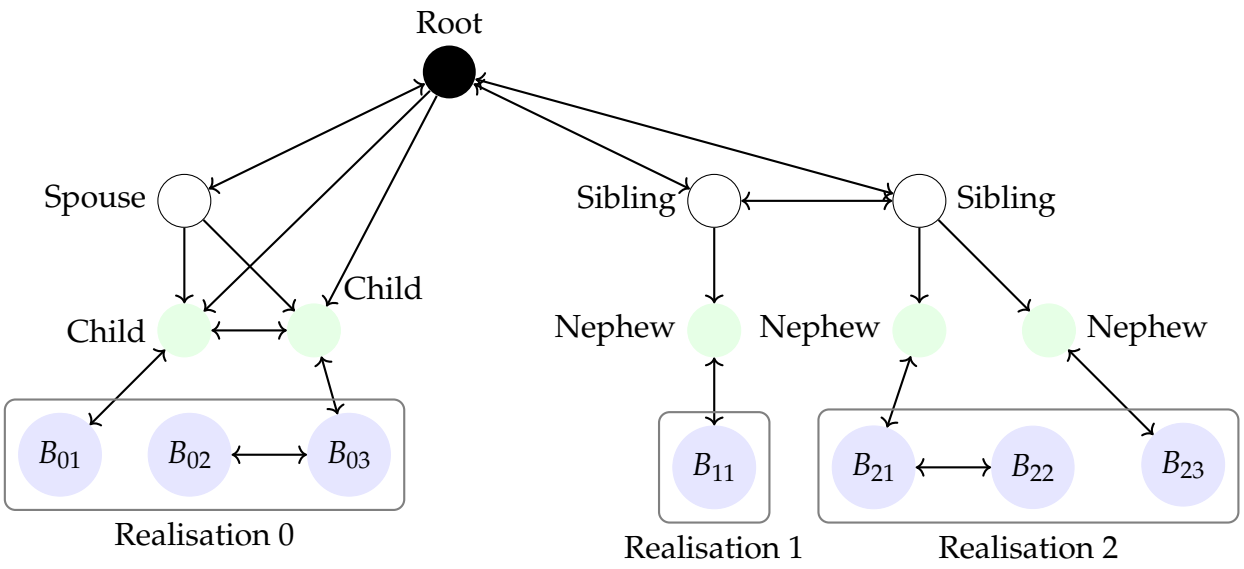
Children connectivity may be endogenous to unobserved characteristics of a dynasty. In this sense, Table 7 shows that wealthier, longer dynasties produce better connected children. This makes estimates in Table 8 showing a positive association between children connectivity and elite persistence difficult to interpret. In particular, wealthy roots with better connected children may have a persistence advantage that reflects ex-ante unobserved heterogeneity influencing both children connectivity and persistence.

In this section, we exploit the richness of the genealogical data to find moments that are likely more exogenous. For this, we produce estimates of the effect of children connectivity exploiting variation in connectivity from members children's extended kinship. Going back to our example for Rockefeller in Figure 5, we are interested in the connectivity of the green nodes, i.e. children, on the wealth persistence of the black nodes, i.e. the CAW wealthy roots. Children connectivity, however, may reflect unobserved characteristics such as other dimensions of eliteness that are not captured by wealth alone. Instead, we will exploit variation in the extended kinship of the the children, i.e. the blue nodes. Still, connectivity of children's extended kinship may reflect persistent positions of said dynasties in the social network of the wealth elite. For example, persistent dynasties with a long history of wealth eliteness may have position themselves overtime in a well-connected neighbourhood of the elite social network to guarantee that their descendents stay in the elite. In this sense, roots with well connected children will have extended kinship that are well connected due to their historical position and this position gives them a persistence edge.

To account for individual specific unobserved heterogeneity, we use the re-centring strategy of Borusyak and Hull (2023). Thus, we exploit deviations in children's extended kinship connectivity with respect to their observation-specific expectation. We generate these conditional expectations by measuring the average connectivity, not only among roots' children extended kinship, but also among the extended kinship of the roots' nephews and the children of the root's spouses from other marriages. This is exemplified in Figure 6. In this Figure, we represent a root who has two siblings and a spouse, all of whom may or may not be contained in CAW. Together with their spouse they have two children. Fur-

thermore, one of the root's siblings has a single child while the other has two children. We measure root's children extended kinship connectivity as the average kinship across all nodes B_{0j}, i.e. the average of *Realisation 0*. This measure of connectivity may, still, reflect ex-ante differences in the position of the root in the social network. To take this into account, we treat each of the root's siblings children extended kinship as realizations from the same data generating process. We then take the average across all these realisations, including the root's children extended kinship themselves, and include this average as an additional control in our regressions. We call this additional control *Expected Kinship Connectivity* in our tables.

Table **??** provides partial correlations between a root's children extended kinship and pre-determined characteristics of the root. Column 1 and 2 of that table show that without controlling for *Expected Kinship Connectivity*, children extended kinship connectivity is positively correlated with the roots wealth, although marginally significant, and, more strongly, with whether any of the root's parents were included in a previous cohort of the CAW data. In columns 3 and 4, we report the same correlations but we also add as an additional control the expected connectivity of kinship. We observe that this additional control reduces the correlation between children extended kinship connectivity and wealth, and the parents of root individuals being in a previous cohort of the CAW data. For the later correlation, we observe a large decline with the partial correlation going from positive and significant to negative and only marginally significant. We still observe, however, a positive and significant correlation between children's extended kinship connectivity and number of children, conditional on any child being present in the genealogical data. This correlation dissipates when we control for the size of the children's extended kinship, including both a linear term and a dummy for whether any extended kinship is observed in the genealogical data. Once we control for kinship size and expected kinship connectivity, columns 5 and 6, we see that all reported covariates are jointly statistically insignificant. We take this as suggestive evidence that extended kinship connectivity, conditional on expected extended kinship connectivity and extended kinship size, is uncorrelated with predetermined characteristics of root individuals.



$$CB_i \equiv \frac{\sum_j B_{ij}}{\sum_j 1}$$

$$CB_0 = \underbrace{\beta \sum_{i=0}^{2} \frac{CB_i}{3}}_{\text{Expected Kinship Connectivity}} + \underbrace{\epsilon}_{\text{Kinship Connectivity Shock}}$$

FIGURE 6. Identifying Variation in Kinship connectivity

Notes: Figure represents an example of a tree where a root has a spouse, two siblings, a child, a stepchild and two nephews. The connectivity of the root's children extended kinship is given by the average across B_{0j}. We take into account that a root may have children with ex-ante differently connected extended kinship by controlling for expected extended kinship connectivity of the child. For this, we treat the realised kinship of each of the root's siblings and the stepchildren of the spouse as a random draw from the same process. We then take averages across all these realisations, including the children kinship itself, and add this average as an additional control in our specifications.

TABLE 10. Wealth Connectivity IV: Balancing and First Stage

	(1)	(2)	(3)	(4)
Panel A. Connectivity of Distant Kinship Conditional on Counterfactual				
(log-)Wealth	0.019	0.018	0.010	0.006
	(0.013)	(0.013)	(0.012)	(0.014)
Inheritance	0.014	0.018	0.033	0.043
	(0.036)	(0.036)	(0.034)	(0.038)
Parent in Previous CAW				-0.027
				(0.032)
Kin Counterfactual	0.700***	0.697***	0.650***	0.643***
	(0.022)	(0.023)	(0.025)	(0.027)
F-stat	1.105	0.999	0.694	0.611
F-stat (p-value)	0.331	0.368	0.500	0.608
Panel B. First Stage (all controls in respective column of panel B included)				
Kin Wealth Connectivity	0.528***	0.527***	0.538***	0.535***
	(0.030)	(0.030)	(0.030)	(0.031)
F-stat	305.257	305.737	314.315	303.995
Observations	3,554	3,554	3,554	2,969
Included Controls:				
Any Children in Gen.Data	✓	✓	✓	✓
Number of Children		✓	✓	✓
Children Sex Composition		✓	✓	✓
Kinship Size			✓	✓

Notes: Panel A provides correlations between distant extended kinship wealth connectivity and a root's wealth and realised wealth persistence including as a control the counterfactual distant extended kinship wealth connectivity. Panel B provides first stage estimates where the left-hand-side is a root's wealth connectivity through their children and the instrument is distance extended kinship wealth connectivity conditional on all controls, in the given column, in Panel A. For Panel A *F-Stat* is the F-statistic for the joint significance of root wealth and realised wealth persistence. For Panel B *F-Stat* is for the significance of the excluded instrument. Estimates are from inverse probability weighting to take into account selection induced by linkage to the genealogical data, standard errors provided between parenthesis take into account the first stage estimation of these weights. All specifications include controls for year, age (quadratic), regions of residence, industry, sex, whether they inherited wealth and a dummy for whether any children is present.

6.2.1 Estimates Exploiting Higher Order Connectivity Shocks

Table 11 provides first stage estimates for children connectivity instrumented with children's extended kinship connectivity without controlling for expected extended kinship connectivity, column 1, controlling for this, column 2, and also controlling for extended kinship size, column 3. First stage estimates are positive, significant and stable across these three specifications. Moreover, for all three specifications, we find that the F-statistic for excluded instruments is well above the typical rule of thumb of ten.

Seeing as extended kinship connectivity provides enough variation in terms of children connectivity, in panels B to C we report reduced form estimates for various elite persistence measures. In panel B, we report the effect of extended kinship connectivity on persistence measured as whether the root has any vertical descendent in the next CAW data. Panel C, provides the effect of extended kinship connectivity on how much wealth the root's vertical descendents have in the next CAW data. Finally, Panel C reports the reduced form effect of children connectivity on whether the extended kinship includes Congress members seating in Congress between the year of the CAW cohort and the next 25 years. Across measures of eliteness and specifications we observe the same effect: roots with better connected extended kinship are more likely to generate descendents who are themselves part of the economic elite and are also more likely to generate extended kinship connections into Congress. We observe similar estimates when we consider 50 year gaps, see Table C.17 in the Appendix. In terms of magnitudes, for the most restrictive specification in column 3, we observe that a one standard deviation increase in terms of extended kinship connectivity increases the likelihood of having a vertical descendent in the CAW data by 13% with respect to the average. A change of the same magnitude in extended kinship connectivity results in a 17% increase with respect to the mean in terms of total wealth controlled by vertical descendents in the next CAW data. Finally, the same magnitude change in the variable of interest leads to an 23% increase with respect to the average on the likelihood that the root's extended kinship contains a member of Congress

sitting in Congress between the CAW cohort and the next twenty-five years, see Panel D of Table 11.[10]

6.2.2 Robustness

This section provides robustness estimates for our main result of children connectivity improving wealth elite persistence.

Panel 7a of Figure 7 provides IV[11] and reduced form estimates for alternative measures of wealth persistence. In particular, we provide estimates for the effect of children connectivity on the likelihood of whether at both 25 and 50 years ahead the root has a vertical descendent in the top 100 and 400 of the CAW data. Estimates come from our most restrictive specification, i.e. column 3 of Table 11. All estimates are positive and significant with the exception of the likelihood of having a vertical descendent in the top 100 within the next 25 years, for which we find a positive albeit insignificant effect. In the longer run, this is 50 years after the root CAW cohort, we do find a positive, large and statistically significant persistence effect at the top 100 from children connectivity.

Panel 7a of Figure 7 also includes an additional measure of wealth persistence, where we predict total wealth of a root's descendents based on their inclusion in Wikipedia and their Wikipedia descriptions. For this, we have collected the Wikipedia descriptions for every CAW member for whom we were able to find an English Wikipedia page. After retrieving Wikipedia descriptions, we use SentenceTransformers[12] to obtain embeddings of the Wikipedia description in a vector space. We extract features from these embeddings that are indicative of wealth by using LASSO to select the embedding dimensions most predictive of wealth. Specifically, we select 113 dimensions out of the 768 dimensions returned in the Sentence Transformers' embeddings obtained from Wikipedia descriptions. We find that these embeddings predict about 34% of the variation in log-wealth for CAW members with a Wikipedia description. Figure D.8 provides a scatter plot of

[10] Table C.18 in the Appendix provides IV estimates where we do not correct for linking to the genealogical data using inverse probability weighting.

[11] IV estimates do not adjust for selection induced by linking to the genealogical data using inverse probability weighting.

[12] With pre-trained model all-mpnet-base-v2.

TABLE 11. The Effect of Children Connectivity: Reduced Form Estimates

	(1)	(2)	(3)
Panel A. Descendent in CAW 25 years ahead			
(log-)Wealth	0.381***	0.381***	0.376***
	(0.037)	(0.037)	(0.037)
Kin Wealth Connectivity	0.188***	0.188***	0.133***
	(0.046)	(0.046)	(0.050)
Average		0.209	
Observations		3,554	
Panel B. Descendent in top 400 CAW 25 years ahead			
(log-)Wealth	0.905***	0.905***	0.902***
	(0.080)	(0.080)	(0.080)
Kin Wealth Connectivity	0.241***	0.241***	0.199**
	(0.075)	(0.075)	(0.082)
Average		0.083	
Observations		3,554	
Panel C. Descendent wealth 25 years ahead			
(log-)Wealth	1.317***	1.317***	1.312***
	(0.231)	(0.231)	(0.232)
Kin Wealth Connectivity	0.466***	0.466***	0.423**
	(0.157)	(0.157)	(0.173)
Average		7,887,275.967	
Observations		3,554	
Panel C. Links to US Congress			
(log-)Wealth	0.219***	0.219***	0.214***
	(0.044)	(0.044)	(0.044)
Kin Wealth Connectivity	0.376***	0.376***	0.319***
	(0.075)	(0.075)	(0.080)
Average		0.166	
Observations		3,554	
Included Controls:			
Kin Counterfactual	✓	✓	✓
Any Children in Gen.Data	✓	✓	✓
Number of Children		✓	✓
Children Sex Composition		✓	✓
Kinship Size			✓

Notes: Table reports reduce form estimates for the effect of children connectivity on various measures of elite persistence. Estimates are produced by OLS re-weighting the data to account for sample selection in genealogical linking. Panel A measures elite persistence by whether the root as a vertical descendent in the next, 25 year ahead, CAW cohort. Panel B measures a similar outcome constraining to have a descendent in the top 400 of CAW. Panel C has analogous time horizon but we substitute a dummy with the total wealth held by a root vertical descendents. Panel D measures elite persistence by coding whether any of a root's children extended kinship has hold a seat in Congress within the next 25 years. All outcomes are divided by their respective averages reported under *Average*. Robust standard errors reported between parenthesis account for the pre-estimation of weights to account for selection in genealogical linking. *Kinship size* includes control for children's extended kinship size, including a dummy for whether any extended kinship is present in the genealogical data. All specifications include controls for year, age (quadratic), regions of residence, industry, sex, whether they inherited wealth and a dummy for whether any children is present. Significance codes: ***: 0.01, **: 0.05, *: 0.1.

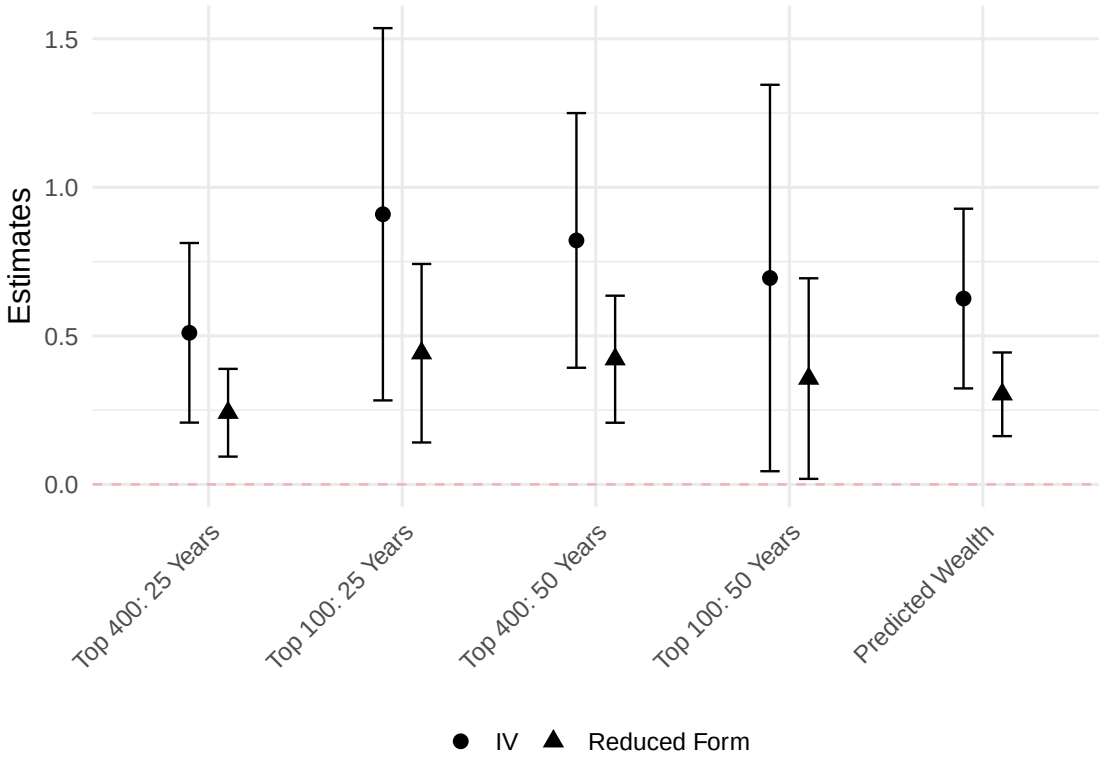
actual observed wealth against predicted wealth from LASSO estimates using Wikipedia embeddings. We use these estimates to predict the wealth of every vertical descendent of a CAW individual with a known Wikipedia entry. Then for every root we aggregate the total predicted wealth among their vertical descendents and report reduced form and IV estimates for the effect of children connectivity on this outcome in the two right-most points of Panel 7a Figure 7. This measure shows also a positive effect of children connectivity on persistence, with magnitudes with respect to average predicted wealth similar to those found for persistence at the top 400.

Panel 7b of Figure 7 reports baseline reduced form estimates in column 3 of Table 11 together with estimates where we re-weight genealogical linked observations to have the same distribution of observable characteristics as millionaires in the 1870 Census. In line with baseline estimates, we find that roots with better connected children are more likely to have a vertical descendent in future cohorts of the CAW data, and that their vertical descendents have larger total wealth observed in future cohorts of the CAW data. Estimates with 1870 Census weights are considerably noisier, and we can only reject the null of no effect for persistence at 25 years ahead and total wealth at 50 years ahead. Nonetheless, point estimates are, with the exception of persistence 50 years ahead that is marginally smaller, always larger than baseline ones. These estimates should be interpreted with caution, as they assume that the distribution of observables among the economic elite has stayed stable over the 1870-1925 period.

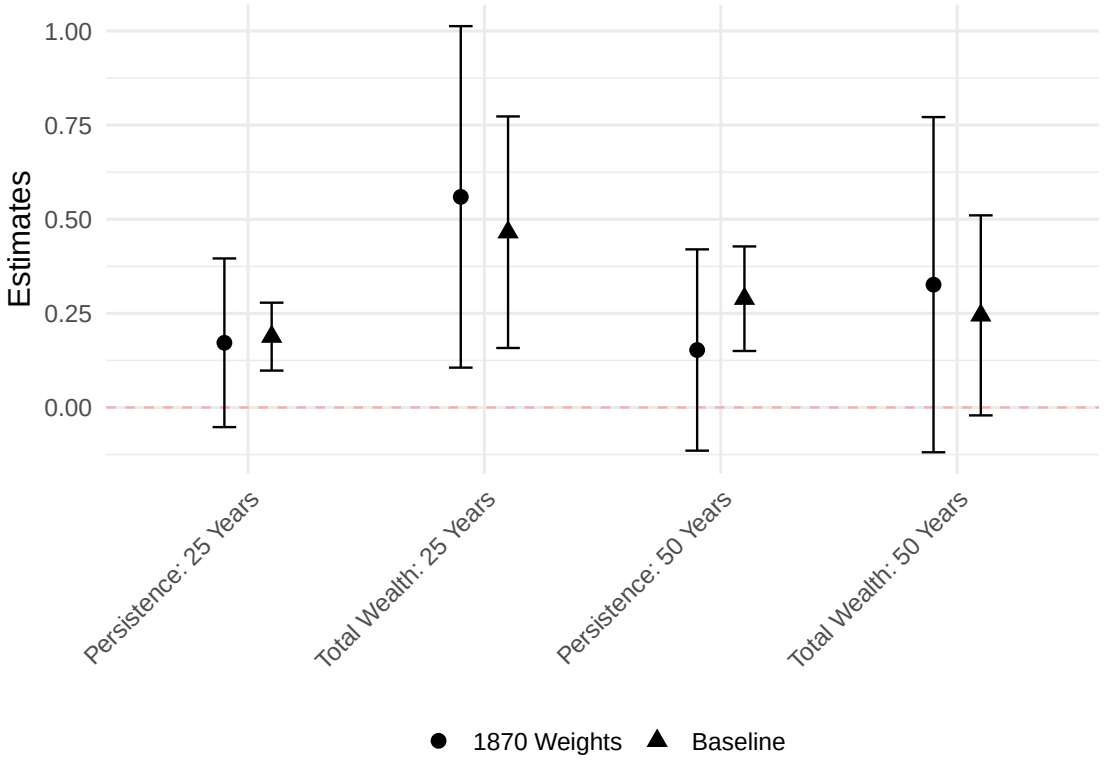
Finally, Table C.19 in the appendix provides reduced form estimates analogous to those in Table 11 but restricting the sample to the top 400 in a given CAW cohort and coding wealth persistence estimates to the top 400 of CAW 25 years ahead. For the dummy of persistence in the top 400 we find an reduced form effect of children extended kinship connectivity that is broadly comparable to that in Panel B of Table 11. For wealth persistence, we find point estimates that are positive although smaller in magnitude, with respect to the average,[13] than those in Panel C of Table 11, furthermore for our most restrictive specification the estimate for wealth persistence in the top 400 in Table C.19 is not

[13] Average wealth of descendents 25 years ahead is \$7,887,275.967 for the whole sample used in Table 11, while for the top 400 in Table C.19 average wealth of descendents in the top 400 of CAW 25 years ahead is \$12,934,976.913.

statistically significant, despite being economically meaningful in terms of magnitude. Finally, we find a positive and significant, although marginally for the most restrictive specification, for the reduced form effect of children extended kinship connectivity on the likelihood of Congress extended kinship connections.



(A) Alternative Measures of Wealth Persistence



(B) 1870 Census Re-Weighted (reduced form estimates)

FIGURE 7. Robustness Estimates

Notes: Figure provides robustness estimates where we provide estimates of the effect of children connectivity on various alternative measures of persistence, Panel 7a, and where using the baseline measures of persistence we re-weight the data to have the distribution of observables of millionaires in the 1870 Census, Panel 7b. 95% confidence intervals as vertical lines.

7 Conclusions

Linking historical "rich lists" with crowd-sourced genealogical data we document patterns of persistence among the wealth elite. We document wealth persistence estimates similar to those in Kalsi and Ward (2025) and also a strong gradient between initial wealth and persistence, even among the very economic elite. New to the literature, we provide evidence of the historical US economic elite netting an tightly connected web of extended family connections. We show that over 40% of wealthy individuals are linked to another wealthy individual twenty-five years later through extended family connections. We observe that wealthier individuals generate better connected children who serve as hubs with other wealthy individuals. OLS and IV estimates exploiting within family connectivity shocks both point in the same direction: individuals with better connected children are more likely to generate descendents who will also be at the top of the wealth distribution. Furthermore, we provide evidence that better connected children also help building bridges with other elites. In particular, we provide evidence that wealthy individuals with better connected children are more likely to have extended family members who contemporaneously hold a seat in Congress. These connections to the political elite are neither build at random. We provide evidence of congress members linked to the economic elite being more likely to have had a family member who served in Congress before them, and more likely to have a family member that will serve in Congress after them.

References

Adermon, A., Lindahl, M., and Palme, M. (2021). Dynastic human capital, inequality, and intergenerational mobility. *American Economic Review*, 111(5):1523–1548.

Barone, G. and Mocetti, S. (2021). Intergenerational mobility in the very long run: Florence 1427–2011. *The Review of Economic Studies*, 88(4):1863–1891.

Beckert, J. (2022). Durable wealth: Institutions, mechanisms, and practices of wealth perpetuation. *Annual Review of Sociology*, 48:561–580.

Björklund, A., Roine, J., and Waldenström, D. (2012). Intergenerational top income mobility in sweden: Capitalist dynasties in the land of equal opportunity? *Journal of Public Economics*, 96(5-6):474–484.

Borusyak, K. and Hull, P. (2023). Nonrandom exposure to exogenous shocks. *Econometrica*, 91(6):2155–2185.

Clark, G. (2014). *The Son Also Rises: Surnames and the History of Social Mobility*. Princeton University Press.

Collado, M. D., Ortuño-Ortín, I., and Stuhler, J. (2023). Estimating intergenerational and assortative processes in extended family data. *The Review of Economic Studies*, 90(3):1195–1230.

Corak, M. (2013). Income inequality, equality of opportunity, and intergenerational mobility. *Journal of Economic Perspectives*, 27(3):79–102.

Cutler, D., Deaton, A., and Lleras-Muney, A. (2006). The determinants of mortality. *Journal of economic perspectives*, 20(3):97–120.

Dal Bó, E., Dal Bó, P., and Snyder, J. (2009). Political dynasties. *The Review of Economic Studies*, 76(1):115–142.

Edlund, L. and Kopczuk, W. (2009). Women, wealth, and mobility. *American Economic Review*, 99(1):146–178.

Goñi, M. (2022). Assortative matching at the top of the distribution: Evidence from the world's most exclusive marriage market. *American Economic Journal: Applied Economics*, 14(3):445–487.

Kalsi, P. and Ward, Z. (2025). The gilded age and beyond: The persistence of elite wealth in american history. Technical report, National Bureau of Economic Research.

Kaplan, S. N. and Rauh, J. D. (2013). Family, education, and sources of wealth among the richest americans, 1982-2012. *American Economic Review: Papers and Proceedings*.

Kopczuk, W. (2015). What do we know about the evolution of top wealth shares in the united states? *Journal of Economic Perspectives*, 29(1):47–66.

Kopczuk, W. and Saez, E. (2004). Top wealth shares in the united states, 1916–2000: Evidence from estate tax returns. *National Tax Journal*, 57(2):445–487.

Michelman, V., Price, J., and Zimmerman, S. D. (2022). Old boys' clubs and upward mobility among the educational elite. *The Quarterly Journal of Economics*, 137(2):845–909.

Moretti, E. and Wilson, D. J. (2023). Taxing billionaires: Estate taxes and the geographical location of the ultra-wealthy. *American Economic Journal: Economic Policy*.

Ozkan, S., Hubmer, J., Salgado, S., and Halvorsen, E. (2023). Why are the wealthiest so wealthy? a longitudinal empirical investigation.

Padgett, J. F. and Ansell, C. K. (1993). Robust Action and the Rise of the Medici, 1400-1434. 98(6):1259–1319.

Saez, E. and Zucman, G. (2016). Wealth inequality in the united states since 1913: Evidence from capitalized income tax data *. *The Quarterly Journal of Economics*, 131(2):519–578.

Smith, M., Zidar, O. M., and Zwick, E. (2021). Top wealth in america: New estimates and implications for taxing the rich. Technical report, National Bureau of Economic Research.

Turchin, P. (2016). *Ages of Discord: A Structural-Demographic Analysis of American History*. Beresta Books, Chaplin, CT.

Wagner, C., Graells-Garrido, E., Garcia, D., and Menczer, F. (2016). Women through the glass ceiling: gender asymmetries in wikipedia. *EPJ data science*, 5(1):5.

Zhang, C. C. and Terveen, L. (2021). Quantifying the gap: a case study of wikidata gender disparities. In *Proceedings of the 17th International Symposium on Open Collaboration*, pages 1–12.

A Pruning Genealogical Trees

Here we describe how we clean the genealogical data obtained from Wikidata, WikiTree, and Find A Grave.

The following date restrictions are imposed on **spouse pairs**:

1. If both partners' birth and death dates are known, they must have been alive at the same time, allowing a tolerance of ± 1 year.
2. If one partner is missing either a birth or death date, the available dates must not imply that the two individuals could not have been alive at the same time. Overlap is checked only on the side where dates are present.
3. If birth or death dates are known for both partners, their birth dates must not differ by more than 90 years, and likewise for their death dates.
4. If all overlap conditions cannot be evaluated because relevant dates are missing, or if all date information is missing on one side, the pair is retained.
5. A final restriction enforces that the two individuals are distinct.

The following date restrictions are imposed on **parent–child pairs** (identical for child–parent but with ego inverted):

1. The ego is born between 11 and 66 years after the parent's birth and is born before the parent's death (allowing a $+1$ year tolerance); or the ego is born between 11 and 66 years after the parent's birth, and the parent's death date is unknown.

2. The parent's birth date is missing, but the child is born before the parent's death (again allowing a $+1$ year tolerance).

3. The child's birth date is missing, but the child's death occurs more than 11 years after the parent's birth.

4. If the above conditions cannot be evaluated because the required dates are missing, or if all date entries are missing on one side, the case is retained.

5. The ego cannot be their own parent.

The following date restrictions are imposed on **sibling–sibling pairs**:

1. The ego and the alter must not be born more than 66 years apart.

2. The ego cannot be their own sibling.

After checking the consistency of dates, we perform further cleaning. We start by eliminating pairs in which the relationship is not unique. For example, a pair may be duplicated, indicating that two individuals are both spouses and siblings. In such cases, we eliminate all pairs that do not have a unique label: sibling-sibling, spouse-spouse, or parent-child. We then eliminate possible superfluous nodes, i.e. subnetworks that are not connected to at least one of the roots and that have been sampled by mistake, e.g. they were only reachable through duplicated edges. After this, we check that trees defined by parent-child relations are not strongly connected; that is, one cannot reach a parent from their children through directed parent-child walks. If we find cycles, we break them by eliminating the link that is furthest from a wealthy individual. This implicitly assumes that links further from a CAW individual are more likely to be incorrect. At this stage, we have a clean network in the sense that dates for family relations are consistent, all nodes connect to at least one CAW individual, and no cycles appear in vertical relations. We use this network to infer espousal relations that may not have been contained in the data. For this, for every node we look at their inward nodes through parent-child relations. We check that there are at most two such nodes and impute a spouse-spouse edge between

these two nodes. After this, we perform a final cleaning in which we retain all nodes that are vertical descendants of CAW individuals, including the CAW individuals themselves, as well as every other node that can be reached within at most five spouse-spouse or sibling-sibling steps from a node on the vertical tree of the CAW individuals.

B US Census

High-level overview of the linkage code

At a high level, the R code constructs consolidated, de-duplicated linkage tables that connect individuals across multiple historical census years, and stores these tables in a SQLite database.

The workflow can be summarised as follows:

1. **Input: pairwise linkage files between census years.** For every ordered pair of years (y_0, y_1) with $y_1 > y_0$, the code reads a compressed file of record links (e.g. `./CensusTree/1870_1900.zip`). Each file contains individual-level identifiers for year y_0 and year y_1, representing matched persons across the two censuses.

2. **Standardisation of identifiers.** The person identifiers from each file are renamed into a common schema:

$$\texttt{histid}y_0 \to \texttt{unperson_id}, \qquad \texttt{histid}y_1 \to \texttt{histid_t}.$$

   This yields a uniform two-column structure for every mapping: one column for the origin-year person identifier (`unperson_id`) and one for the target-year identifier (`histid_t`).

3. **One consolidated table per target year.** For each target year y_1, the code builds and maintains a database table

$$\texttt{XMAP}y_1,$$

which aggregates mappings from all earlier years $y_0 < y_1$ into that same target year. The earliest origin year initialises `XMAP`y_1 directly from its pairwise linkage file; subsequent origin years contribute additional rows.

4. **De-duplication by target-year identifier.** When incorporating new mappings for a given target year y_1 from another origin year y_0, the code only retains those rows whose target-year identifier `histid_t` is not already present in `XMAP`y_1. In effect, for each target-year person, the table avoids creating multiple entries that differ only by their origin year. The result is a consolidated crosswalk in which each `histid_t` in year y_1 appears at most once.

5. **Storage and maintenance in SQLite.** All intermediate and final linkage tables are stored in a SQLite database, accessed via an R database connection. After processing all year pairs and finalising all `XMAP`y_1 tables, the script issues a `VACUUM;` command to compact the database file and reclaim unused space created by temporary tables.

In summary, the code integrates many separate pairwise linkage files into a set of unified, non-duplicated cross-year linkage tables, one for each target census year, and stores them in a single SQLite database for efficient querying and further analysis.

C Additional Tables

TABLE C.1. Surname Persistence Among the Wealthiest

	Root Cohort						
	1875	**1892**	**1900**	**1902**	**1924**	**1925**	**1950**
Panel A. Surname Persistence							
1892	0.922						
1900	0.847	0.411					
1902	0.823	0.719	0.779				
1924†	0.634	0.552	0.759	0.549			
1925	0.589	0.357	0.711	0.347	0.337		
1950	0.327	0.250	0.379	0.239	0.254	0.530	
Observations	587	4,043	1,407	3,624	3,268	1,593	1,070
Panel B. Share with Surname in US Congress							
	0.678	0.649	0.625	0.602	0.548	0.577	0.584
Observations	587	4,043	1,407	3,624	3,268	1,593	1,070
Panel C. Assortative Marriage							
			0.469			0.544	0.512
Observations			213			395	336

Notes: Table provides in Panel A the likelihood of observing an wealthy individual at a future point in time that carries the same surname. Panel B provides the share of individuals in each dataset with a surname that is contained in the US Congress data. We do these two exercises based on the last surname within the name string. Panel C uses the CAW data exclusively to select women with a maiden name that appeared in a previous CAW cohort, for these it provides the share of them for whom their married surname was also contained in a previous CAW cohort.

TABLE C.2. Alternative list historical lists of wealthy and high-income individuals in the U.S.

Name	Target list	Source	Observations	Details
Census1870	1875	wealth	913†	Census contains the list of individuals reputed to be worth a million or more.
Tribune1892	1875/1900	wealth	4056	The Tribune's list of all persons in the United States reputed to be worth a million or more.
Detroit1900	1900	wealth	42	Detroit Evening published the list of all Detroit millionaires with wealth.
Almanac1902	1900	wealth	3734	World Almanac: all persons in the United States reputed to be worth a million or more.
Forbes1918	1900/1925	wealth	30	The first Forbes list contains all millionaires with wealth above 50 million.
Fortune1957	1950	wealth	76	People with wealth above 75 million.
Fortune1968	1950	wealth	66	Fortune 1968 published a list of America's centimillionaires; only the 66 above $150 million are listed.
Nye1915	1900/1925	income	176	All U.S. taxpayers with more than $1 million in taxable income during the years 1915 to 1920.
TaxReturns1924	1925	income	–	All tax returns for U.S. residents in 1923 and 1924.
Morgenthau1941	1925/1950	income	98	The largest taxpayers in 1941.

Notes: Publicly available data provides personal and real estate value truncated at $ 999997, the reported figure is comprised of individuals for whom the sum of the two is at least $ 999997.

TABLE C.3. Linking Rates

1875	1900	1925	1950
Overall			
0.869	0.818	0.806	0.764
Wikidata Genealogies			
0.750	0.632	0.616	0.580
WikiTree Genealogies			
0.673	0.576	0.527	0.461
Find a Grave Genealogies			
0.744	0.696	0.679	0.615

Notes: Table provides linking rates for roots in the historical rich lists. Top panel provides overall linking rates. Lower panels provide linking rates by genealogical source.

TABLE C.4. Relative proportions of CAW millionaire and multi-millionaires and general population

Year	Millionaires	Decamillionaires	Centimillionaires	Population Source
1875	63%	98%	100%	1870 Census
1900	38%	141%	67%	1982 Tribune List
1925	9%	72%	69%	1929 Tax Returns
1950	4%	84%	77%	1953/1957 Fortune list

TABLE C.5. Differences in Observed Characteristics between 1875 CAW and 1870 Census List

	CAW	1870 Census	Difference
Female	0.070	0.101	-0.031**
			(0.015)
Age	49.708	46.863	2.844***
			(0.768)
Finance, Insurance and Law	0.164	0.053	0.111***
			(0.017)
Manufacturing	0.108	0.065	0.042***
			(0.015)
Mining	0.115	0.004	0.110***
			(0.013)
Oil	0.010	0.003	0.007
			(0.005)
Railroad	0.161	0.014	0.146***
			(0.016)
Real Estate	0.113	0.014	0.098***
			(0.014)
Unknown Industry	0.048	0.283	-0.235***
			(0.017)
Other	0.282	0.562	-0.280***
			(0.025)
Observations	585	901	

Notes: Table provides means for every category both in CAW and the 1870 Census from which we select every entry with property value equal or above a million. Industry information in 1870 Census has been manually recoded from IPUMS 1950 basis industry codes and should be taken as approximate.

TABLE C.6. Differences in Observed Characteristics between 1900 CAW and 1892 Tribune List

	CAW	1892 Tribune List	Difference
Female	0.233	0.099	0.134***
			(0.012)
Inheritance	0.413	0.214	0.199***
			(0.015)
Finance, Insurance and Law	0.130	0.058	0.072***
			(0.009)
Manufacturing	0.145	0.212	-0.067***
			(0.011)
Mining	0.132	0.032	0.100***
			(0.009)
Oil	0.042	0.016	0.025***
			(0.006)
Railroad	0.170	0.072	0.099***
			(0.011)
Real Estate	0.083	0.188	-0.105***
			(0.009)
Other	0.298	0.423	-0.124***
			(0.014)
Observations	1,402	4,043	

Notes: Table provides means for every category both in CAW and the 1892 Tribune rich list. Industry information in 1892 Tribune appear as free text that we categorize into the CAW industry classification using a zero shot classifier, thus it should be interpreted with caution.

TABLE C.7. Extended Kinship Connections Among CAW Individuals

	Wikidata				WikiTree				Find A Grave			
	Avg.	S.D.	Min.	Max.	Avg.	S.D.	Min.	Max.	Avg.	S.D.	Min.	Max.
1875 Cohort												
Any Link	0.69	0.46	0	1	0.99	0.10	0	1	0.97	0.16	0	1
# 1° Neigh.	0.73	1.01	0	8	3.13	3.15	0	16	4.20	3.32	0	15
# 2° Neigh.	1.22	1.97	0	11	5.03	5.07	0	26	8.02	6.08	0	30
# All Neigh.	21.33	60.03	0	638	72.92	127.53	0	1,376	204.04	303.35	0	3,554
Avg. Distance	3.34	1.72	1	7	3.80	1.17	1	6	4.47	1.11	1	7
Max. Distance	4.80	2.62	1	10	6.06	1.80	1	10	6.92	1.70	1	10
1900 Cohort												
Any Link	0.66	0.47	0	1	0.97	0.17	0	1	0.98	0.12	0	1
# 1° Neigh.	0.89	1.42	0	9	2.90	3.18	0	14	4.44	3.47	0	16
# 2° Neigh.	1.24	2.49	0	24	3.94	4.46	0	29	7.71	6.03	0	57
# All Neigh.	12.15	32.51	0	301	37.96	63.03	0	527	127.20	167.36	0	2,279
Avg. Distance	2.75	1.42	1	7	3.25	1.10	1	6	4.06	0.88	1	6
Max. Distance	4.02	2.24	1	9	5.17	1.69	1	9	6.17	1.35	1	9
1925 Cohort												
Any Link	0.67	0.47	0	1	0.97	0.16	0	1	0.99	0.11	0	1
# 1° Neigh.	1.05	1.59	0	9	2.72	3.30	0	23	4.28	3.28	0	20
# 2° Neigh.	1.54	3.06	0	26	3.62	4.69	0	38	7.30	5.85	0	41
# All Neigh.	11.22	30.73	0	596	28.28	46.87	0	448	88.08	106.47	0	1,506
Avg. Distance	2.49	1.25	1	7	2.99	1.04	1	5	3.78	0.78	1	6
Max. Distance	3.65	2.01	1	9	4.63	1.58	1	8	5.62	1.11	1	9
1950 Cohort												
Any Link	0.68	0.47	0	1	0.96	0.20	0	1	0.98	0.15	0	1
# 1° Neigh.	1.20	1.94	0	23	2.21	2.75	0	15	3.66	2.97	0	17
# 2° Neigh.	1.93	4.37	0	45	2.81	3.96	0	29	6.06	5.57	0	32
# All Neigh.	13.28	42.75	0	596	18.77	34.43	0	303	57.57	66.11	0	593
Avg. Distance	2.33	1.17	1	6	2.64	1.06	1	5	3.48	0.79	1	5
Max. Distance	3.37	1.87	1	8	4.04	1.64	1	7	5.07	1.11	1	7

Notes: Table provides descriptive statistics for the extended kinship connections of individuals in CAW in each cohort for every of our genealogical sources. Column *Avg.* provides averages, *S.D.* standard deviations, *Min.* minimums and *Max.* maximums. *Avg. Distance* and *Max.Distance* provide individuals' average and maximum family distances descriptive statistics for those with at least one link.

TABLE C.8. Genealogy Linking Errors: WikiTree

Panel A. Link Actual and Assigned Status

Linking Decision	Actual Link	
	No	**Yes**
Link Disregarded	1.000	0.006
Link Chosen	0.000	0.994
	[89,457]	[1,990]

Panel B. Distribution of Assigned Matches

False Match	No Match	True Match
0.000	0.220	0.780
[1]	[452]	[1602]

Panel C. Assigned Matches When All True Matches Are Removed

False Match	No Match
0.002	0.998
[3]	[1672]

Notes: Table provides rates for various miss-classification errors from our genealogical link discriminator. Whether a link is a true one or not is taken from data provided by Wikidata. We train the link discriminator through cross-validation on data manually matched to Wikidata. We train a random forest to discriminate true and false links based on birth and death year distances and name string distances. We select a candidate link if the largest estimated probability of being a true link is larger than 80% and this probability is larger than the second best candidate link by at least 10 percentage points. †In *Panels B* and *C* we eliminate all true matches from the testing data and compute the share of root individuals that have been linked to an incorrect entry in the genealogical data, thus *Observations* reports the number of unique root individuals to be linked rather than the total number of candidate links. Discriminator has been trained and tested with both individuals from the CAW and Dal Bó et al. (2009) data.

TABLE C.9. Genealogy Linking Errors: Find A Grave

Panel A. Link Actual and Assigned Status

Linking Decision	Actual Link	
	No	**Yes**
Link Disregarded	1.000	0.043
Link Chosen	0.000	0.957
	[147,366]	[1,746]

Panel B. Distribution of Assigned Matches

False Match	No Match	True Match
0.008	0.284	0.708
[15]	[520]	[1296]

Panel C. Assigned Matches When All True Matches Are Removed

False Match	No Match
0.049	0.951
[70]	[1366]

Notes: Table provides rates for various miss-classification errors from our genealogical link discriminator. Whether a link is a true one or not is taken from data provided by Wikidata. We train the link discriminator using cross-validation on data manually matched to Wikidata. We train a random forest to discriminate true and false links based on birth and death year distances and name string distances. We select a candidate link if the largest estimated probability of being a true link is larger than 80% and this probability is larger than the second best candidate link by at least 10 percentage points. †In *Panels B* and *C* we eliminate all true matches from the testing data and compute the share of root individuals that have been linked to an incorrect entry in the genealogical data, thus *Observations* reports the number of unique root individuals to be linked rather than the total number of candidate links. Discriminator has been trained and tested with both individuals from the CAW and Dal Bó et al. (2009) data.

TABLE C.10. Vertical Kinship Persistence: Top 400 to Top 400

	Root Cohort		
	1875	**1900**	**1925**
1900	0.156		
	[0.120,0.192]		
1925	0.163	0.190	
	[0.125,0.201]	[0.151,0.230]	
1950	0.114	0.159	0.156
	[0.081,0.146]	[0.122,0.196]	[0.120,0.193]
Observations	400	400	400
Share with Genealogy	0.907	0.920	0.930

Notes: Table reports the likelihood of a CAW individual in the top 400 having a vertical descendant in the top 400 of a future CAW cohort estimated from those roots that we have been able to link to genealogical data and re-weighting to account for selection induced by said linkage.

TABLE C.11. Vertical Kinship Persistence

	Root Cohort		
	1875	**1900**	**1925**
Panel A. Raw Estimates			
1900	0.297		
	[0.260,0.334]		
1925	0.306	0.185	
	[0.269,0.343]	[0.165,0.206]	
1950	0.137	0.132	0.135
	[0.109,0.164]	[0.114,0.150]	[0.118,0.152]
Panel B. Linking Corrected			
1900	0.323		
	[0.284,0.363]		
1925	0.338	0.214	
	[0.297,0.379]	[0.191,0.237]	
1950	0.152	0.153	0.161
	[0.121,0.183]	[0.133,0.173]	[0.141,0.180]
Panel C. 1875 Covariates Distribution			
1900	0.325		
	[0.285,0.366]		
1925	0.340	0.181	
	[0.299,0.381]	[0.148,0.215]	
1950	0.152	0.129	0.146
	[0.121,0.183]	[0.102,0.157]	[0.055,0.237]
Panel D. 1870 Census Covariates Distribution			
1900	0.210		
	[0.136,0.284]		
1925	0.278	0.172	
	[0.196,0.359]	[0.124,0.219]	
1950	0.111	0.144	0.091
	[0.064,0.158]	[0.078,0.210]	[0.061,0.121]
Observations	585	1,402	1,567
Share with Genealogy	0.867	0.815	0.805

Notes: Table reports the likelihood of having a descendent in a future cohort of the CAW data given that one is the CAW data. We do this for every root cohort, in columns, and for various time spans, in rows. Panel A, the joint likelihood of a root being genealogically linked and having a vertical descendent in a future CAW cohort. Panel B reports the likelihood of a root having a vertical descendent in a future CAW cohort estimated from those roots that we have been able to link to genealogical data and re-weighting to account for selection induced by said linkage. Panel C (D) re-weights the data to have the same distribution of observed characteristics as the 1875 CAW (1870 Census). Between parenthesis we report 95% confidence intervals, where re-weighting is employed we produce these from standard errors accounting for pre-estimation of

TABLE C.12. Vertical Kinship Persistence by Wealth Group

Target Wealth Group	Root Wealth Group				
	1-100	**101-200**	**201-300**	**301-400**	**400+**
Panel A: 25 Years Ahead					
1-100	**0.143**	0.033	**0.017**	0.011	0.008
	[0.103,0.183]	[0.012,0.054]	[0.002,0.032]	[-0.001,0.024]	[0.004,0.011]
101-200	0.091	**0.047**	0.014	**0.014**	**0.010**
	[0.058,0.124]	[0.022,0.072]	[0.000,0.028]	[0.000,0.028]	[0.006,0.015]
201-300	0.075	0.039	0.026	0.012	0.011
	[0.045,0.105]	[0.017,0.062]	[0.007,0.044]	[-0.002,0.025]	[0.006,0.016]
301-400	0.055	0.053	0.022	0.016	**0.010**
	[0.029,0.081]	[0.027,0.079]	[0.005,0.039]	[0.000,0.032]	[0.005,0.014]
400+	0.130	0.199	0.223	0.169	0.098
	[0.092,0.169]	[0.152,0.246]	[0.174,0.272]	[0.124,0.215]	[0.084,0.111]
Broadly Wealthy	0.494	0.372	0.302	0.223	0.136
	[0.436,0.551]	[0.315,0.428]	[0.247,0.356]	[0.172,0.273]	[0.120,0.151]
Observations	300	300	300	300	2,354
Share with Genealogy	0.973	0.927	0.913	0.863	0.768
Panel B: 50 Years Ahead					
1-100	**0.148**	**0.047**	0.026	**0.042**	**0.023**
	[0.098,0.197]	[0.017,0.078]	[0.003,0.049]	[0.006,0.078]	[0.013,0.032]
101-200	0.082	0.042	**0.028**	0.011	0.016
	[0.044,0.121]	[0.014,0.071]	[0.004,0.053]	[-0.004,0.027]	[0.008,0.024]
201-300	0.046	0.037	0.040	0.011	0.013
	[0.017,0.075]	[0.010,0.064]	[0.011,0.068]	[-0.004,0.026]	[0.006,0.020]
301-400	0.042	0.016	0.027	0.000	0.011
	[0.013,0.070]	[-0.002,0.033]	[0.004,0.050]	[0.000,0.000]	[0.005,0.018]
400+	0.117	0.179	0.208	0.126	0.067
	[0.072,0.162]	[0.124,0.234]	[0.148,0.268]	[0.076,0.175]	[0.051,0.083]
Broadly Wealthy	0.435	0.322	0.329	0.190	0.131
	[0.365,0.504]	[0.255,0.388]	[0.260,0.397]	[0.129,0.251]	[0.109,0.152]
Observations	200	200	200	200	1,187
Share with Genealogy	0.970	0.940	0.900	0.845	0.773

Notes: Table provides wealth group transition probabilities at 25, Panel A, and 50, Panel B, year gaps conditional on the starting wealth group. For every column and panel we highlight the largest persistence within detailed wealth groups in bold. Below the dashed line we report persistence into a residual wealthy group, below the solid horizontal line we report overall persistence. Data has been re-weighted to account for selection induced by the linkage of the CAW data to the genealogical data. 95% confidence between square brackets account for pre-estimation of these weights, some cells are empty and for these we report a zero with zero length confidence intervals. *Share with Genealogy* reports the share of observations, by baseline wealth group, that we have been able to link to at least one genealogical source.

TABLE C.13. Extended Kinship Persistence

	Root Cohort		
	1875	**1900**	**1925**
Panel A. Raw Estimates			
1900	0.460		
	[0.419,0.500]		
1925	0.427	0.417	
	[0.387,0.467]	[0.391,0.443]	
1950	0.347	0.276	0.352
	[0.310,0.384]	[0.253,0.299]	[0.329,0.376]
Panel B. Linking Corrected			
1900	0.538		
	[0.494,0.582]		
1925	0.487	0.511	
	[0.443,0.532]	[0.482,0.540]	
1950	0.385	0.329	0.438
	[0.343,0.428]	[0.302,0.356]	[0.410,0.465]
Panel C. 1875 Covariates Distribution			
1900	0.534		
	[0.490,0.577]		
1925	0.486	0.476	
	[0.442,0.530]	[0.431,0.520]	
1950	0.390	0.316	0.356
	[0.348,0.432]	[0.273,0.359]	[0.262,0.450]
Panel C. 1870 Census Covariates Distribution			
1900	0.437		
	[0.332,0.542]		
1925	0.378	0.452	
	[0.275,0.482]	[0.348,0.555]	
1950	0.339	0.242	0.305
	[0.238,0.439]	[0.164,0.320]	[0.243,0.367]
Observations	585	1,402	1,567
Share with Genealogy	0.867	0.815	0.805

Notes: Below the diagonal, table provides estimates of likelihood of an individual being linked to the top 600 wealthies in future generations. *Panel A.* provides estimates for vertical kinship, i.e. from parents to their children. *Panel B.* provides estimates for extended kinship links, i.e. from parents to their children but also between siblings and spouses up to 5 degrees of separation from the vertical kinship of the root individual, see Section XX for detailed description.

TABLE C.14. Extended Kinship Persistence: Top 400 to Top 400

	Root Cohort		
	1875	**1900**	**1925**
1900	0.318		
	[0.271,0.366]		
1925	0.297	0.460	
	[0.250,0.344]	[0.409,0.510]	
1950	0.311	0.344	0.404
	[0.264,0.359]	[0.296,0.392]	[0.355,0.453]
Observations	400	400	400
Share with Genealogy	0.907	0.920	0.930

Notes: Table reports the likelihood of a CAW individual in the top 400 having an extended kinship relation in the top 400 of a future CAW cohort estimated from those roots that we have been able to link to genealogical data and re-weighting to account for selection induced by said linkage.

TABLE C.15. Top Ten Connected CAW Individuals

1875 CAW	**1900 CAW**
Laura (Astor) Delano	Joanna Hone (Howland) Grinnell
Mary Alida (Astor) Carey	Dr William Seward Webb
William (Backhouse) Astor (jr)	Harry Payne Whitney
Henry Astor	William Douglas Sloane
John Jacob Astor III	Hamilton McKown Twombly
William Henry Vanderbilt	Alice Claypoole (Gwynne) Vanderbilt
William Backhouse Astor	John Gerard Heckscher
William Edgar Howland	Oliver Harriman
Hamilton McKown Twombly	Oliver Harriman jr
Richard Thornton Wilson	Richard Thornton Wilson

1925 CAW
John Davison Rockefeller
Harold Fowler McCormick
Edith (Rockefeller) McCormick
John Davison Rockefeller jr
Alta (Rockefeller) Prentice
Margaret (Strong) de Cuevas de Larrain
Cornelius Vanderbilt Whitney
Alice Claypoole (Gwynne) Vanderbilt
Alexander Blair Thaw
Josiah Copley Thaw

Notes: Table provides the names for the top ten best connected individuals in each of the CAW cohorts. For 1875 CAW, William Earl Dodge and Melissa (Phelps) Dodge, who are spouses, have the same children connectivity thus we have chosen to report eleven individuals instead of ten.

TABLE C.16. Children Connections: Wealth and Elite Persistence

	(1)	(2)	(3)
Panel A. Descendent in CAW 50 years ahead			
(log-)Wealth	0.329***	0.329***	0.327***
	(0.055)	(0.055)	(0.055)
Wealth Connectivity	0.356***	0.356***	0.318***
	(0.058)	(0.058)	(0.061)
Average		0.208	
Observations		1,987	
Panel B. Descendent in top 400 CAW 50 years ahead			
(log-)Wealth	0.625***	0.625***	0.624***
	(0.100)	(0.100)	(0.099)
Wealth Connectivity	0.575***	0.575***	0.544***
	(0.101)	(0.101)	(0.102)
Average		0.104	
Observations		1,987	
Panel C. Descendent wealth 50 years ahead			
(log-)Wealth	0.589***	0.589***	0.588***
	(0.172)	(0.172)	(0.171)
Wealth Connectivity	0.445***	0.445***	0.429***
	(0.115)	(0.115)	(0.108)
Average		14,072,470.739	
Observations		1,987	
Included Controls:			
Any Children in Gen.Data	✓	✓	✓
Number of Children		✓	✓
Children Sex Composition		✓	✓
Kinship Size			✓

Notes: Table reports partial correlations between various measures of elite persistence and roots' wealth and connectivity of their children. Estimates are produced by OLS re-weighting the data to account for sample selection in genealogical linking. Panel A measures elite persistence by whether the root as a vertical descendent in the next, 50 year ahead, CAW cohort. Panel C has analogous time horizon but we substitute a dummy with the total wealth held by a root vertical descendents. All outcomes are divided by their respective averages reported under *Average*. Robust standard errors reported between parenthesis account for the pre-estimation of weights to account for selection in genealogical linking. *Kinship size* includes control for children's extended kinship size, including a dummy for whether any extended kinship is present in the genealogical data.

TABLE C.17. The Effect of Children Connectivity: Reduced Form Estimates

	(1)	(2)	(3)
Panel A. Descendent in CAW 50 years ahead			
(log-)Wealth	0.328***	0.328***	0.328***
	(0.055)	(0.055)	(0.055)
Kin Wealth Connectivity	0.289***	0.289***	0.230***
	(0.071)	(0.071)	(0.075)
Average		0.208	
Observations		1,987	
Panel A. Descendent in top 400 CAW 50 years ahead			
(log-)Wealth	0.625***	0.625***	0.625***
	(0.099)	(0.099)	(0.099)
Kin Wealth Connectivity	0.421***	0.421***	0.372***
	(0.109)	(0.109)	(0.115)
Average		0.104	
Observations		1,987	
Panel B. Descendent wealth 50 years ahead			
(log-)Wealth	0.599***	0.599***	0.599***
	(0.168)	(0.168)	(0.168)
Kin Wealth Connectivity	0.245*	0.245*	0.211
	(0.136)	(0.136)	(0.132)
Average		14,072,470.739	
Observations		1,987	
Included Controls:			
Kin Counterfactual	✓	✓	✓
Any Children in Gen.Data	✓	✓	✓
Number of Children		✓	✓
Children Sex Composition		✓	✓
Kinship Size			✓

Notes: Table reports reduce form estimates for the effect of children connectivity on various measures of elite persistence. Estimates are produced by OLS re-weighting the data to account for sample selection in genealogical linking. Panel A measures elite persistence by whether the root as a vertical descendent in the next, 50 year ahead, CAW cohort. Panel B measures a similar outcome constraining to have a descendent in the top 400 of CAW. Panel C has analogous time horizon but we substitute a dummy with the total wealth held by a root vertical descendents. All outcomes are divided by their respective averages reported under *Average*. Robust standard errors reported between parenthesis account for the pre-estimation of weights to account for selection in genealogical linking. *Kinship size* includes control for children's extended kinship size, including a dummy for whether any extended kinship is present in the genealogical data. All specifications include controls for year, age (quadratic), regions of residence, industry, sex, whether they inherited wealth and a dummy for whether any children is present. Significance codes: ***: 0.01, **: 0.05, *: 0.1.

TABLE C.18. Children Connections, Wealth and Elite Persistence: IV Estimates

	(1)	(2)	(3)
Panel A. Descendent in CAW 25 years ahead			
(log-)Wealth	0.390***	0.390***	0.386***
	(0.038)	(0.038)	(0.038)
Wealth Connectivity	0.388***	0.388***	0.278***
	(0.090)	(0.090)	(0.096)
Average		0.209	
Observations		2,910	
Panel A. Descendent in top 400 CAW 25 years ahead			
(log-)Wealth	0.962***	0.962***	0.959***
	(0.083)	(0.083)	(0.083)
Wealth Connectivity	0.510***	0.510***	0.419**
	(0.154)	(0.154)	(0.163)
Average		0.083	
Observations		2,910	
Panel B. Descendent wealth 25 years ahead			
(log-)Wealth	1.425***	1.425***	1.422***
	(0.255)	(0.255)	(0.256)
Wealth Connectivity	1.024***	1.024***	0.933**
	(0.351)	(0.351)	(0.379)
Average		7,887,275.967	
Observations		2,910	
Included Controls:			
Any Children in Gen.Data	✓	✓	✓
Number of Children		✓	✓
Children Sex Composition		✓	✓
Kinship Size			✓

Notes: Table reports partial correlations between various measures of elite persistence and roots' wealth and connectivity of their children. Estimates are produced by IV exploiting shocks within the distant extended kinship of a root's children. Panel A measures elite persistence by whether the root as a vertical descendent in the next, 25 year ahead, CAW cohort. Panel B has analogous time horizon but we substitute a dummy with the total wealth held by a root vertical descendents. Panel C measures elite persistence by coding whether any of a root's children extended kinship has hold a seat in Congress within the next 25 years. All outcomes are divided by their respective averages reported under *Average*. Robust standard errors reported between parenthesis. *Kinship size* includes control for children's extended kinship size, including a dummy for whether any extended kinship is present in the genealogical data.

TABLE C.19. The Effect of Children Connectivity: Restricted to top 400

	(1)	(2)	(3)
Panel A. Reduced Form: Descendents in CAW Top 400 25 Yeas Ahead			
(log-)Wealth	1.191***	1.176***	1.176***
	(0.109)	(0.109)	(0.107)
Kin Wealth Connectivity	0.305***	0.301***	0.246**
	(0.095)	(0.093)	(0.101)
Average		0.168	
Observations		1,200	
Panel B. Reduced Form: Wealth in Top 400 25 Years Ahead			
(log-)Wealth	2.572***	2.566***	2.558***
	(0.535)	(0.537)	(0.540)
Kin Wealth Connectivity	0.753**	0.740**	0.665**
	(0.294)	(0.292)	(0.335)
Average		12,934,976.913	
Observations		1,200	
Panel C. Reduced Form: Extended Kinship Connections in Congress			
(log-)Wealth	0.244***	0.247***	0.238***
	(0.090)	(0.089)	(0.088)
Kin Wealth Connectivity	0.391***	0.382***	0.279**
	(0.108)	(0.108)	(0.115)
Average		0.215	
Observations		1,200	
Included Controls:			
Kin Counterfactual	✓	✓	✓
Any Children in Gen.Data	✓	✓	✓
Number of Children		✓	✓
Children Sex Composition		✓	✓
Kinship Size			✓

Notes: Table reports reduced form estimates for the effect of children connectivity on persistence restricting the sample to individuals in the top 400. Panel A, regresses a dummy for whether an individual has a vertical descendent in the top 400 of CAW 25 ahead on children's extended kinship connectivity. Panel B regresses total wealth of a root's vertical descendents observed within the top 400 of the CAW data 25 year ahead on root's children extended kinship connectivity. Panel D, regresses a dummy for whether a root's children extended kinship contains a Congress member sitting in Congress between the baseline period and 25 years ahead. All left-hand-side variables have been divided by their respective average to be able to interpret estimates as changes with respect to averages. Estimates are re-weighted to account for selection in genealogical linking. Robust standard errors accounting for pre-estimation of these weights are provided between parenthesis. All specifications include controls for year, age (quadratic), regions of residence, industry, sex, whether they inherited wealth and a dummy for whether any children is present. Significance codes: : ***: 0.01, **: 0.05, *: 0.1.

D Additional Figures

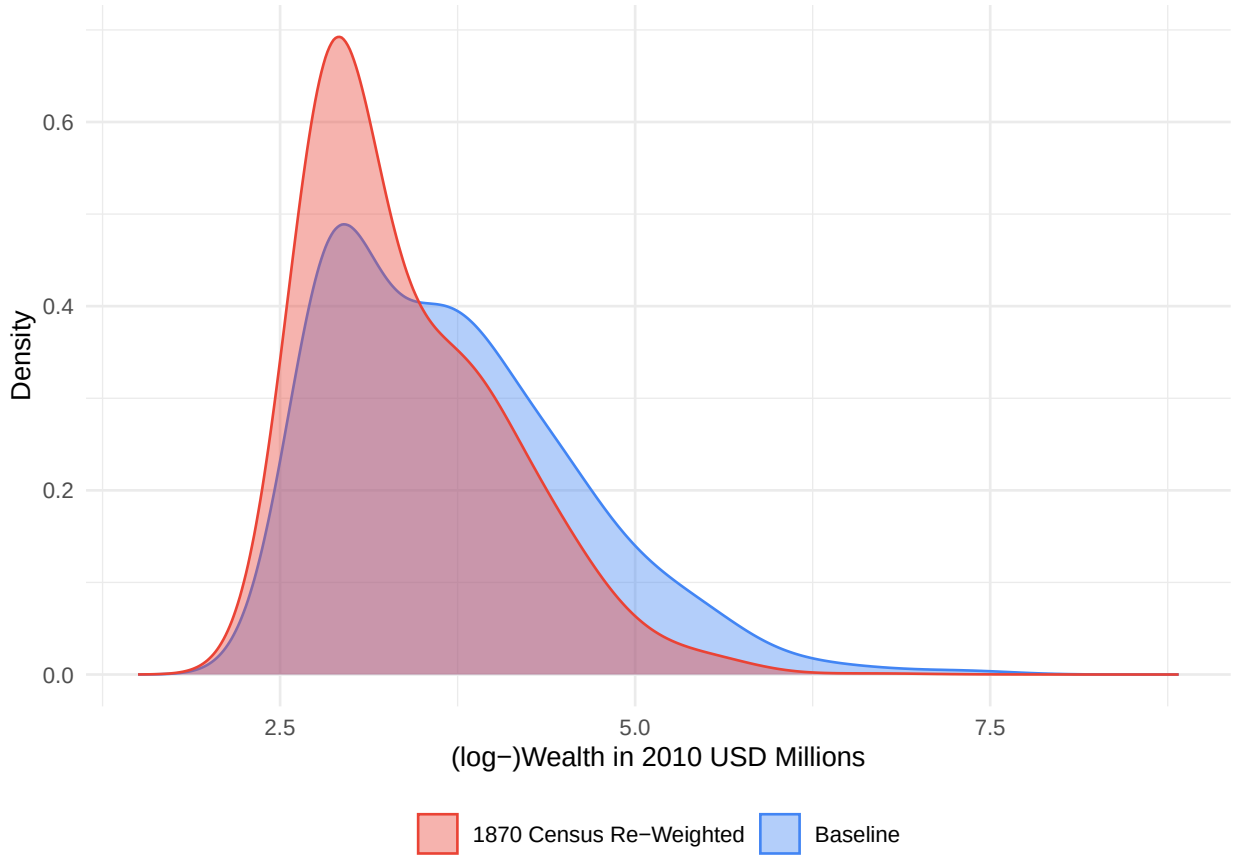


FIGURE D.1. 1875 CAW Wealth Distribution: Raw vs 1870 Census Re-weighted

Notes: Figure provides the density of log wealth in the 1875 cohort of the CAW data with, red figure, and without, blue figure, re-weighting so that CAW observations have the same distribution of observables, other than wealth, as the 1870 census.

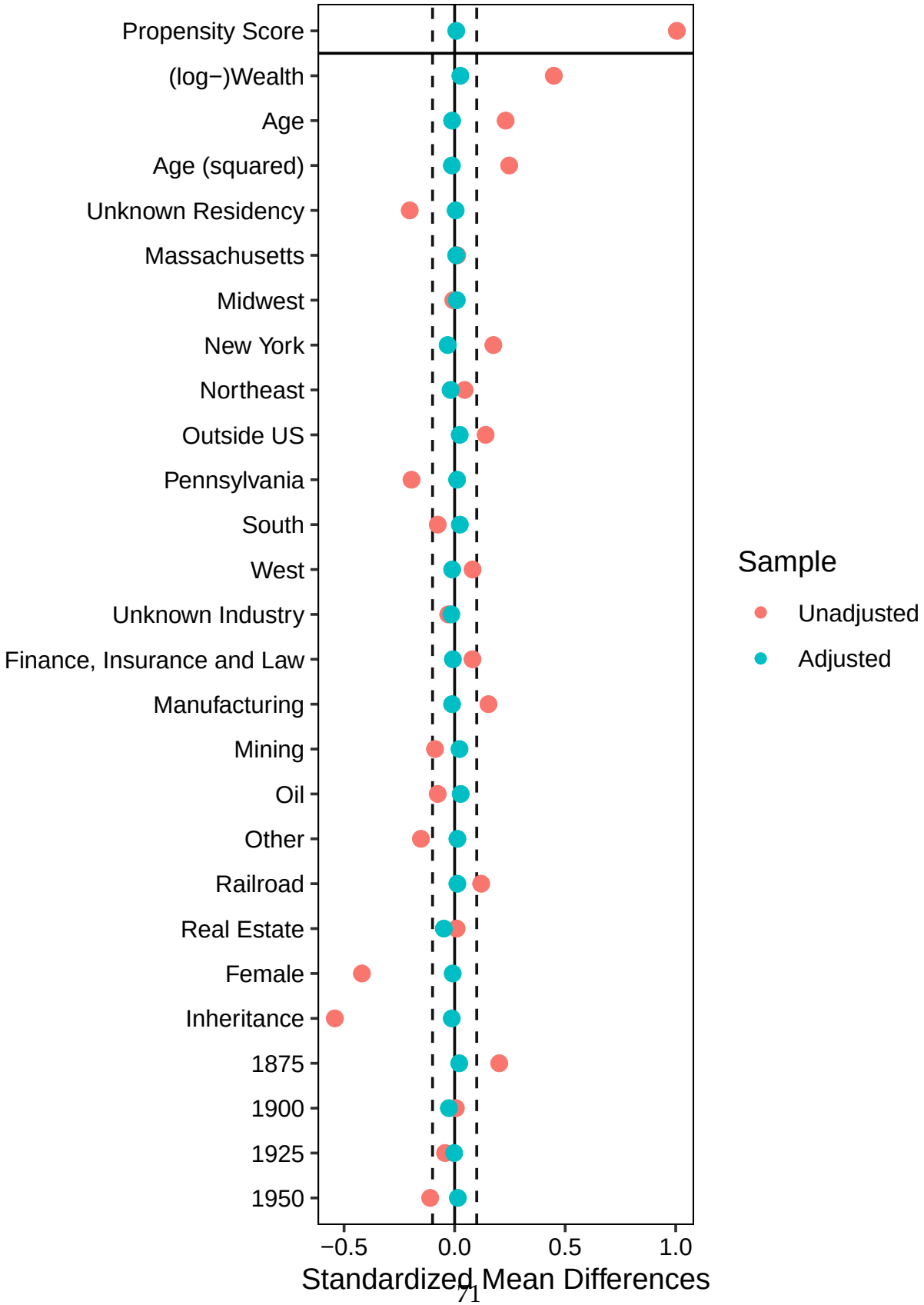


FIGURE D.2. Covariance Balance Between Linked and Unlinked to WikiData Data

Notes: Figure provides balancing of observed characteristics between observations linked to the WikiData data and those that have been not. Figure provides unadjusted differences (red) and adjusted differences (blue) using propensity score.

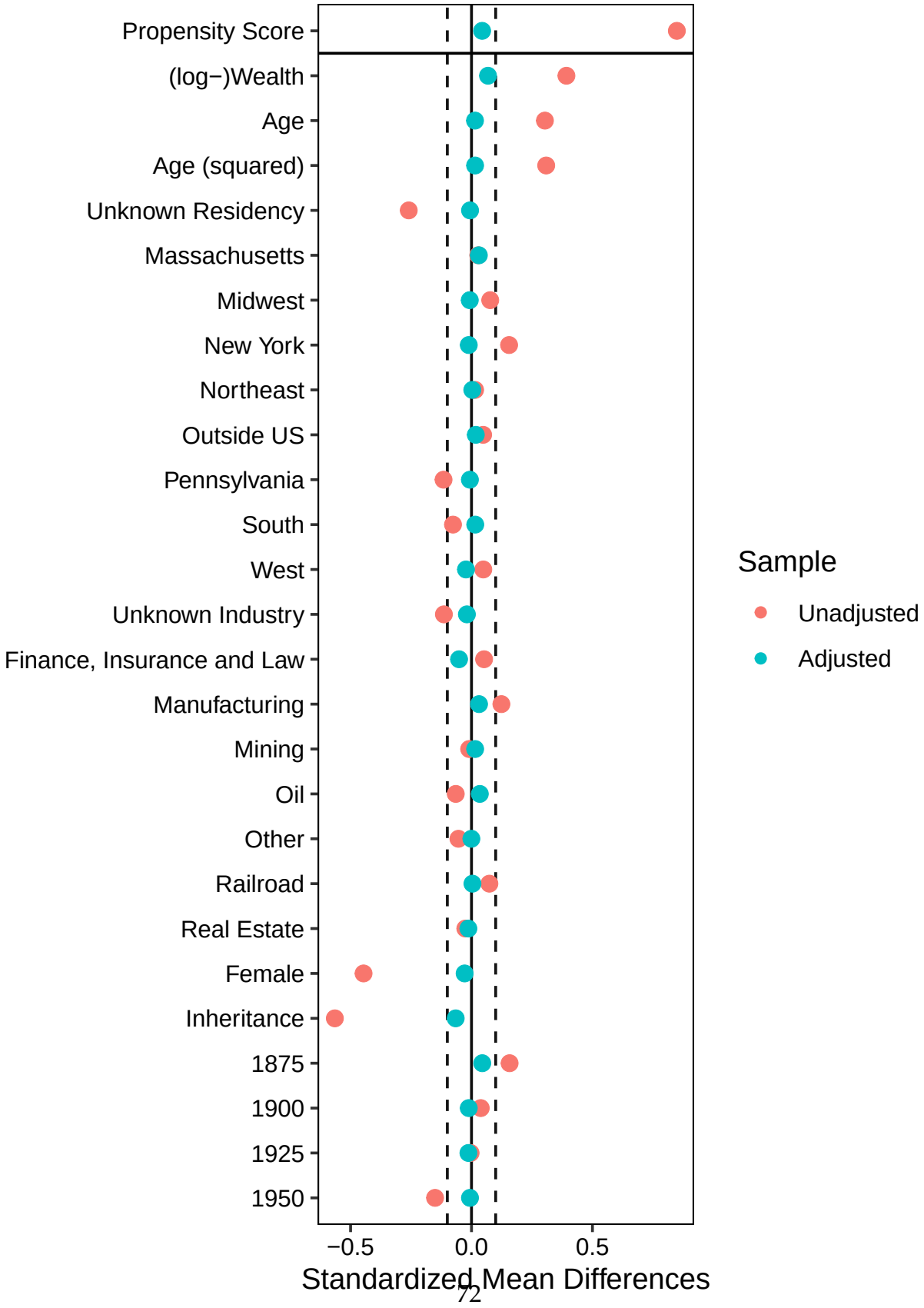


FIGURE D.3. Covariance Balance Between Linked and Unlinked to Genealogical Data

Notes: Figure provides balancing of observed characteristics between observations linked to the genealogical data and those that have been not. Figure provides unadjusted differences (red) and adjusted differences (blue) using propensity score.

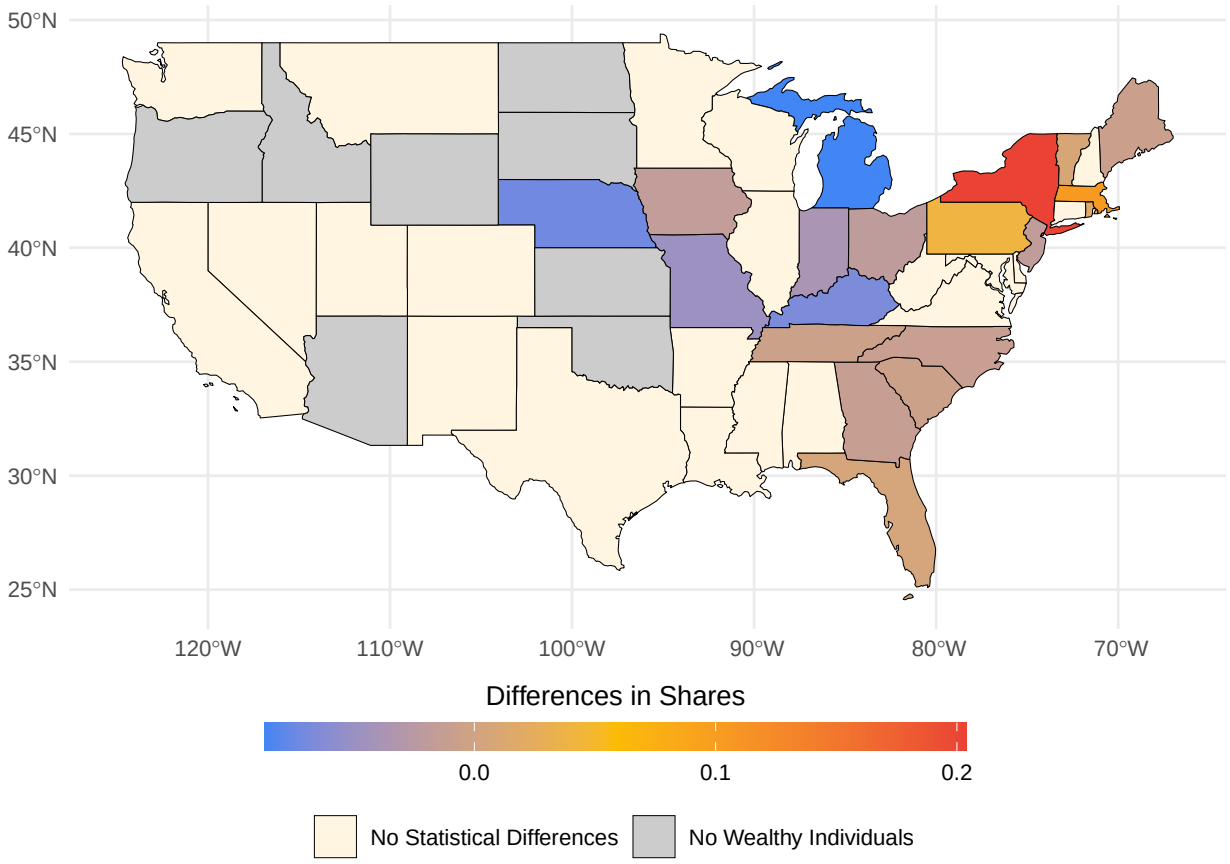


FIGURE D.4. Differences Between CAW and 1870 Census Geographic Coverage

Notes: Figure provides differences in geographic coverage between the CAW 1875 data and the 1870 Census for individuals with property value equal or above one million by testing for difference in terms of the share of sample that each state contains. For differences that are statistically significant at 90% significance level we plot the actual difference, states with no significant differences in shares between the two samples are represented in broken white, states with no observations in either dataset are represented in grey.

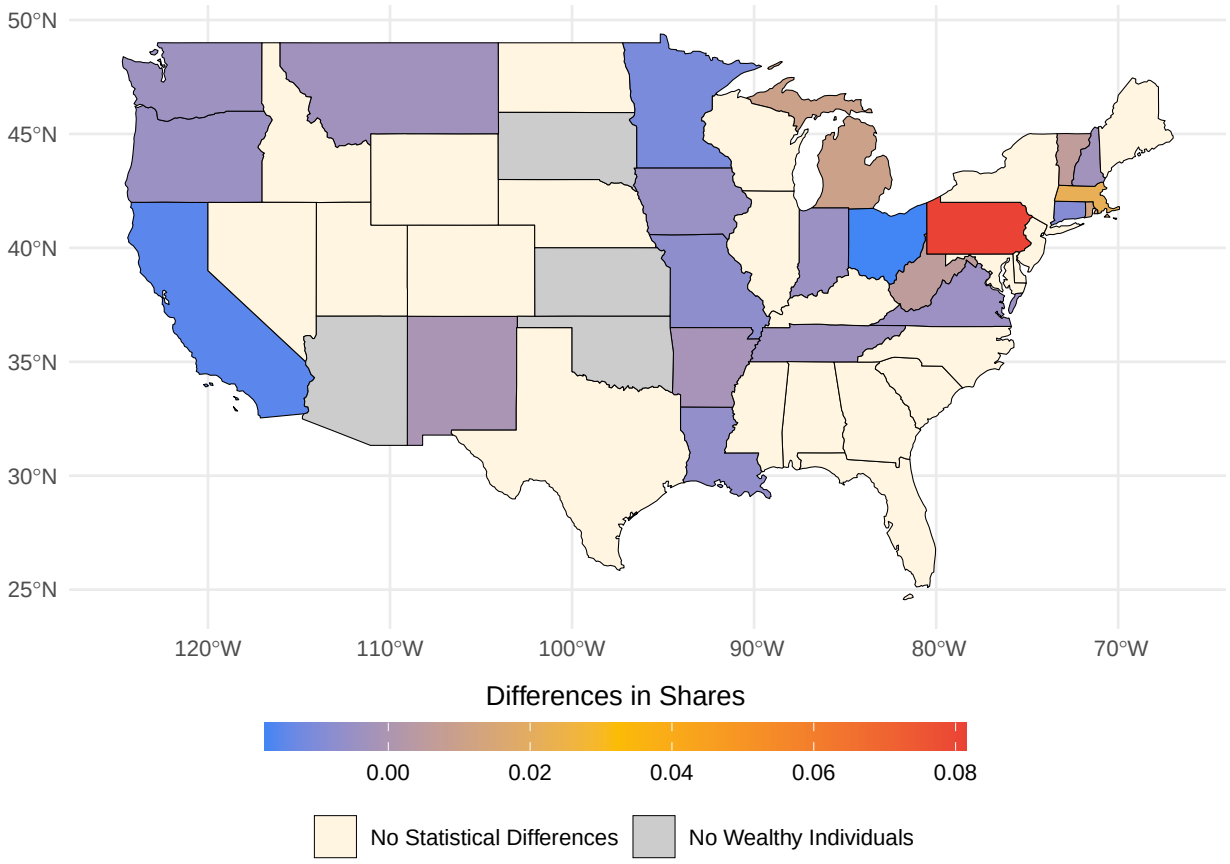


FIGURE D.5. Differences Between CAW and 1892 Tribune List Geographic Coverage

Notes: Figure provides differences in geographic coverage between the CAW 1900 data and the 1892 Tribune list by testing for difference in terms of the share of sample that each state contains. For differences that are statistically significant at 90% significance level we plot the actual difference, states with no significant differences in shares between the two samples are represented in broken white, states with no observations in either dataset are represented in grey.

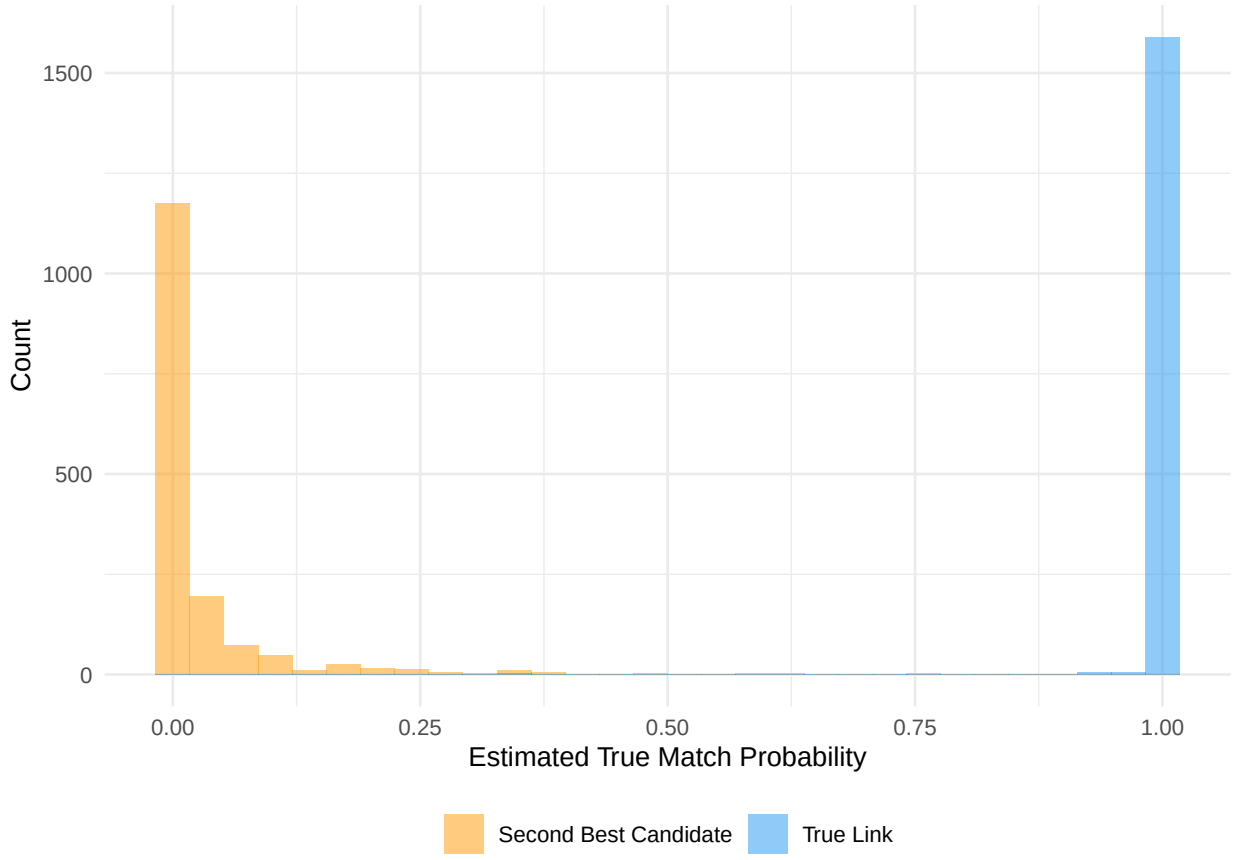


FIGURE D.6. WikiTree: Estimated Likelihood of Link Being a True Match for True Matches and Second Best Candidates

Notes: Figure provides Random Forest estimates for the likelihood of a candidate link being the true link for both actual true links and their second best candidates. For this, we use a sample of 1,610 wealthy individuals for which the true link is known from Wikidata and for which we have found a second best candidate when searching WikiTree. By construction, the number of *True Links* and *Second Best Candidates* is balanced.

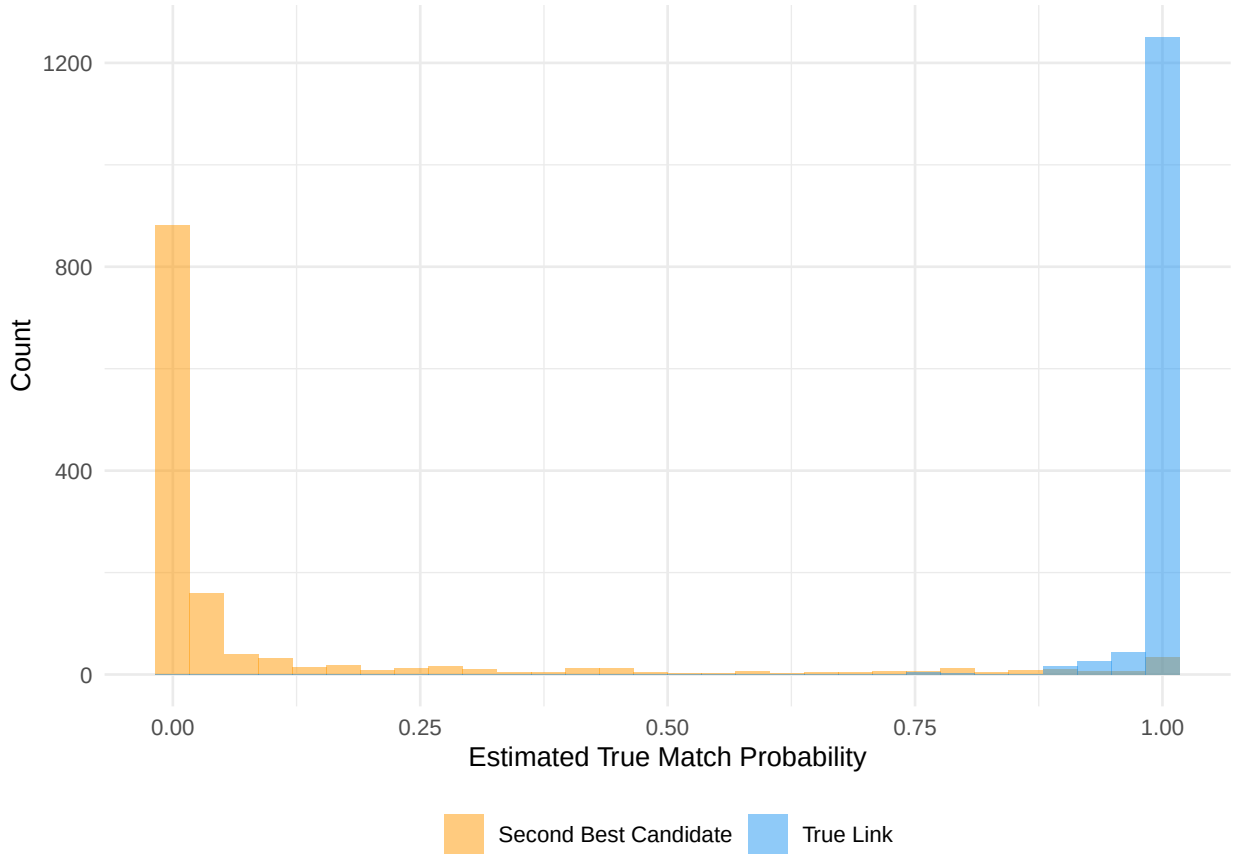


FIGURE D.7. Find a Grave: Estimated Likelihood of Link Being a True Match for True Matches and Second Best Candidates

Notes: Figure provides Random Forest estimates for the likelihood of a candidate link being the true link for both actual true links and their second best candidates. For this, we use a sample of 1,351 wealthy individuals for which the true link is known from Wikidata and for which we have found a second-best candidate when searching Find A Grave. By construction, the number of *True Links* and *Second Best Candidates* is balanced.

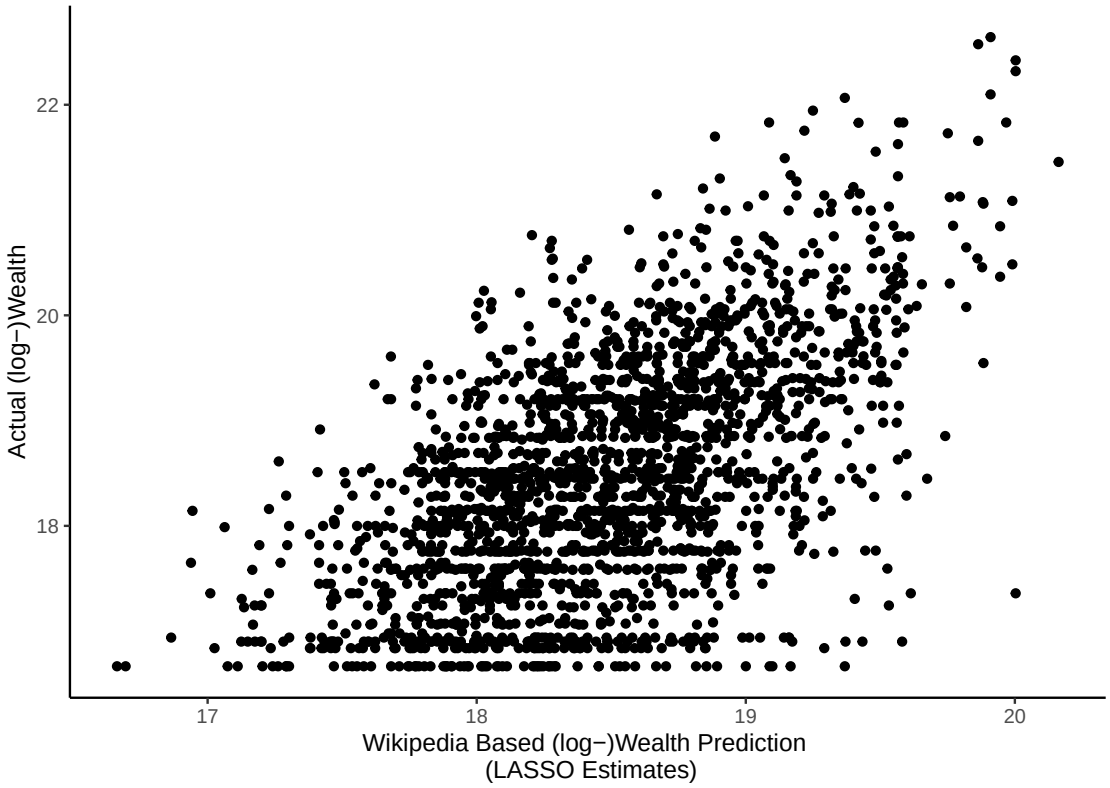


FIGURE D.8. Wikipedia Wealth Scores and Actual Wealth