Surnames and Distinction*

Surname-based measures of the Persistence of the French Elite 1791-1902, Wealth, Honours, and Office

Noah Sutter (LSE)†

November 21, 2026

Preliminary and incomplete; please do not circulate or cite. The latest version is available here.

Abstract

Bi-generational estimates of intergenerational wealth mobility imply that economic advantages dissipate within only a few generations. Surname-based methods, by contrast, uncover much higher rates of intergenerational persistence. I extend the surname-based method to 19$^{\text{th}}$ century France, using a newly digitised dataset of 4.6 million death records (1.9 million with a positive amount of wealth at death). I find high rates of persistence comparable to the ones for England during the same period — between 0.74 and 0.87. I use data on 22 elite registers linked at the surname level, yielding more than 240,000 observations, to independently test these results and confirm the persistence levels. I do not find evidence for the economic decline of the *ancien régime* nobility. While sorting on wealth gives a persistence rate of 0.739 (0.156) for the highest ventile, sorting on hereditary noble titles results in no detectable regression to the mean. I propose the hypothesis that the latent factor can be interpreted as a durable endowment with social and cultural capital.

JEL: C38, D63, J62, N33, Z13

Keywords: Social Mobility, Wealth Mobility, Social Capital, Cultural Capital, French Revolution, Persistence of Status

*I want to thank Neil Cummins, Mohamed Saleh, Thomas Piketty, Jean-Laurent Rosenthal, Gilles Postel-Vinay, Guido Alfani, Chris Minns, Eric Schneider, Gabriel Zucman, Ellora Derenoncourt, Arnaud Lefranc, Aurelius Noble, Jan Stuhler, Andrea Del Pizzo, Taylan Alpkaya, and Hillary Vipond for supervision and comments at different stages of this project, as well as the participants of the LSE Economic History Seminar, the SPEECH Seminar at LSE, the International Inequalities Institute Graduate Seminar, the IT19 Winter School on Inequality and Social Welfare Theory, and the Seminar on *"New Directions in the Historical and Comparative Sociology of Elites"* at LSE.

†E-mail: n.w.sutter@lse.ac.uk

"[Capital] is what makes the game of society — not least, the economic game — something other than simple games of chance offering at every moment the possibility of a miracle." (P. Bourdieu, 2011, p. 78)

1 Introduction

The question of the degree of persistence of economic advantage across generations has long occupied social scientists as well as economists. Initially, studies focused on correlations between two successive generations (see Black & Devereux, 2011; Solon, 1999, for reviews of the literature). These studies have been interpreted as implying that economic advantages disappear in a few generations. In recent years a new literature has emerged that has taken a more long-term perspective — either by incorporating several linked generations or by adopting a rare surnames framework. These studies find significantly more persistence of economic advantage across generations (Adermon, Lindahl, & Palme, 2021; Adermon, Lindahl, & Waldenström, 2018; Barone & Mocetti, 2021; Braun & Stuhler, 2018; Clark, 2015; Clark & Cummins, 2014, 2015; Clark, Cummins, Hao, & Vidal, 2015; Eeckhout, Markevich, & Schoors, 2023b; Güell, Mora, & Telmer, 2015; Güell, Rodríguez Mora, & Solon, 2018; Güell, Pellizzari, & Rodríguez Mora, 2018; Lersch, Longmuir, & Schnitzlein, 2024; Lindahl, Palme, Massih, & Sjögren, 2015; Long & Ferrie, 2018; Stuhler, 2024). This means that inequality of wealth and income is potentially much more persistent than assumed.

In this paper, I extend this approach to a new historical context: 19$^{\text{th}}$ century France (1791-1902). Three features make this setting an interesting object of study. First, France in this period was characterised by more radical institutional change and critical junctures than the corresponding period in England investigated in the canonical study by Clark and Cummins (2015), one could therefore expect higher levels of intergenerational mobility. Second, France during the period under investigation has featured an exceptionally unequal wealth distribution, which has been a focus of the recent revival of studies of economic inequality (Piketty, 2001, 2011, 2014; Piketty, Postel-Vinay, & Rosenthal, 2006, 2014). Third, whether the French Revolution was associated with the economic ascendancy of a new capitalist wealth elite at the expense of an old, declining feudal elite or not, was a classic debate in French and international historiography (Comninel, 1991, p. 2). A classical view emphasizing the former (Jaurès, 1924; Lefebvre, 1939; Mathiez, 1922; Soboul, 1953, 1988) was challenged by an emerging revisionist camp arguing for the latter in the 1960s (Beck, 1981; Cobban, 1964; Forster, 1967). I find evidence that despite the more radical institutional change, social mobility is not higher in France than in England. This implies that the high levels of wealth inequality found by Piketty

(2001, 2011, 2014) and Piketty et al. (2006) were not just persistent but one was also very likely to find oneself in a similar position to one's parents. Finally, I find evidence for high levels of persistence of the *ancien régime* nobility at the top of the wealth distribution, thereby lending support to the *revisionist* camp.

I contribute a newly digitised dataset of 4.6 million death records (1.9 million with a positive amount of wealth at death) for France during the period 1791-1870. I use this data to apply the approach pioneered by Clark and Cummins (2015) to this new dataset and find surname-group-based wealth persistence rates of 0.74–0.87 per generation. This is comparable with the rates of 0.70-0.75 found by Clark and Cummins (2015) for England. The persistence rates for France in the period 1791-1870 appear to be even slightly higher than the ones for England, despite its more turbulent political history. A likely explanation for this is that none of the French Revolutions of the long 19^{th} century involved substantial economic redistribution. Even the property seized during the Revolution starting in 1789 was eventually largely returned through the indemnity billion. By contrast, J. Bourdieu, Kesztenbaum, Postel-Vinay, and Suwa-Eisenmann (2019) find bi-generational estimates of between 0.324 and 0.365 for the same period using the *Enquête TRA* data.

In the interpretation proposed by the latent factor model (Clark & Cummins, 2014, 2015; Collado, Ortuño-Ortín, & Stuhler, 2023; Stuhler, 2024), this discrepancy results from an attenuation of the latter number while the former uncovers the higher persistence of an underlying slow-moving latent factor of which wealth is a signal. I find independent evidence for this interpretation by employing another dataset and approach to uncover the persistence of the underlying latent factor developed by Eeckhout, Markevich, and Schoors (2023a). I use data on 22 elite outcomes including 269,734 individual observations. Assuming that these outcomes are also signals of the underlying latent factor, they can be used to estimate the intergenerational persistence of the latent factor. Replicating the approach by Eeckhout et al. (2023a), I find a persistence rate of between 0.92 and 0.94, virtually indistinguishable from the surname-group-based wealth persistence rates. This result is robust to fixed-effects and balanced-panel specifications. The persistence rate is much higher than the rates estimated by Eeckhout et al. (2023b) for the Russian Empire, the Soviet Union, and the Russian Federation. In light of the more radical redistributive policies and institutional change in the Soviet Union, this is not surprising.

I propose a new interpretation of the latent factor as broad capital endowment of a family. Building on P. Bourdieu (1984, 2011), I assume that families hold social and cultural capital that decays slowly, which generates an economic return in each generation and thereby refills the stock of economic capital, which, in isolation, decays at a faster rate. This would imply that *"random"* windfall profits decay

faster than wealth which is part of a wider family endowment with capital which includes cultural and social capital. I find evidence for this interpretation. If I sort groups on wealth alone, I find a persistence parameter of 0.74 (0.16) for the top ventile of wealth holders. If I sort on an *ancien régime* hereditary title, I find a persistence parameter that is statistically indistinguishable from one, meaning no detectable mean reversion. At the same time, I find that the *ancien régime* nobility is far more overrepresented in parliament than the surnames of the top 5% of the wealth distribution. This indicates that among the merely rich some of the economic capital consists of windfall profits and is not associated with a wider family endowment with social and cultural capital, while the rich former *ancien régime* nobility manages to persist through their wider family endowment. This model is consistent with the literature, which has found that randomly allocated positive wealth shocks do not have a long-term effect (Bleakley & Ferrie, 2016) and that elites with high social and cultural capital recover from negative wealth shocks (Ager, Boustan, & Eriksson, 2021; Alesina et al., 2020; Eeckhout et al., 2023b).

This paper is a contribution to three strands of the literature. The first is the literature on the measurement of long-run social mobility. The results of the social mobility literature focussing on bi-generational correlations imply that advantages dissipate after two or three generations (Becker & Tomes, 1979; Black & Devereux, 2011; Chetty, Hendren, Kline, & Saez, 2014; Chetty, Hendren, Kline, Saez, & Turner, 2014; Solon, 1999). J. Bourdieu et al. (2019) study social mobility in France using this approach. But studies taking a multigenerational view (Braun & Stuhler, 2018; Lersch et al., 2024; Lindahl et al., 2015; Long & Ferrie, 2018; Stuhler, 2024) or using surname-based methods (Adermon et al., 2021, 2018; Barone & Mocetti, 2021; Braun & Stuhler, 2018; Clark, 2015; Clark & Cummins, 2014, 2015; Clark et al., 2015; Güell et al., 2015, 2018; Güell et al., 2018) consistently find very high persistence of 0.6-0.9 that does not dissipate significantly over generations. I estimate the first surname-based wealth-persistence estimates for post-revolutionary France, based on 4.6 million newly digitised wealth-at-death records linked at the surname-level to 5 different registers of different elite groups. I find group-level persistence rates of 0.74–0.87 per generation, lending support to the idea of a latent factor slowly regressing to the societal mean.

The second literature that the paper contributes to is the literature on elite persistence through shocks and critical junctures. Wealth and occupational elites recover from revolution, expropriation, or emancipation consistently in a variety of historical settings (Ager et al., 2021; Alesina et al., 2020; Dupraz & Simson, 2024; Eeckhout et al., 2023b; Galli, Theodoridis, & Rönnbäck, 2024). Finley, Franck, and Johnson (2021) and Franck and Michalopoulos (2017) focus on the French Revolution, specifically on land redistribution and on emigration respectively. Van Leeuwen, Maas, Rébaudo,

and Pélissier (2016) take a long-term perspective that includes the French Revolution. These studies raise hypotheses about the channels through which the elites recover their position in the wealth distribution. Alesina et al. (2020), Ager et al. (2021), and Eeckhout et al. (2023b) all hypothesize that a conversion of social or tacit cultural capital into financial capital (via credit networks or transmitted attitudes) are the reason for the elite recovery. I provide evidence that specific forms of social capital generate returns across regimes, and thereby can serve as a channel for elite recovery, while others are tied to a specific regime and cease to serve this function after the end of this regime. I find that a surname tied to *ancien régime* nobility is associated with wealth that shows no detectable reversion to the mean and with representation in parliament across seven different constitutional orders, while noble titles awarded by the Napoleonic regime are associated with only a small wealth advantage, which disappears by the 1840s.

Third, the paper contributes to the literature on the interpretation of the latent factor theorised by Clark and Cummins (2014). The latent factor has been interpreted as unobservable "social status" (Clark & Cummins, 2014) or as statistical object that is generated by multiple intergenerational transmission channels (Braun & Stuhler, 2018; Collado et al., 2023; Stuhler, 2024).[1] I propose an economic interpretation of the latent factor as a portfolio of durable family assets, different forms of capital, financial, social, and cultural capital, that can yield economic returns without depleting and transmitted across generations. The role of social capital in generating an economic return (through the marriage market) has been researched by Goñi (2022), and its role in social mobility by Chetty, Friedman, Saez, Turner, and Yagan (2020); Chetty and Hendren (2018a, 2018b); Chetty et al. (2022a), and Chetty et al. (2022b). I propose that this asset logic proposed by Goñi (2022) can be generalised and serve as an interpretation of the latent factor.

The rest of the paper is organised as follows. Section 2 describes the data that is employed in the paper and presents descriptive statistics. Section 3 introduces the empirical approach applied to the data. Section 4 presents the empirical results on wealth and status persistence. Section 5 discusses results and proposes a new theoretical interpretation of the latent factor. Section 6 concludes.

[1] I discuss how my results and interpretation fit into the framework proposed by Braun and Stuhler (2018); Collado et al. (2023); Stuhler (2024), and Del Pizzo, Nybom, and Stuhler (2026) in appendix D.

2 Data

2.1 Wealth Data

The wealth data is mostly based on newly digitised data from the *Tables des Successions et Absences*, the French inheritance register, for the period 1791-1870. These registers list wealth at death in the *département* of the last legal residence for every individual dying in France.[2] This includes women, who are registered separately from their husbands, and children. There is no lower wealth threshold to be included and I therefore also capture the poor. The dataset covers 20 *départements*[3] and the digitisation process is described in Noble and Sutter (tba). The entire dataset encompasses approximately 4.6 million individual-level death records between 1791 and 1870, 1.9 million of which record a positive amount of wealth at death. The digitisation of the wealth data has been collaborative work with Aurelius Noble (Noble & Sutter, tba).

2.2 Elite Status Data

This wealth dataset is then linked at the surname-level to various datasets indicating endowments with elite cultural and social capital. I combine data on noble status, the *Légion d'Honneur*, membership in parliament, book authors, colonial administrators, as well as pre-revolutionary elites, i.e. the *Ferme Générale*, and the *avocats* at the Paris *Parlement*.

For Bourdieu, a noble title is *the* example of social capital in its institutionalised form (P. Bourdieu, 2011, p. 79). In France, a *nobility* had existed since the middle ages that began to be verified by the crown, who also awarded noble titles, in the 17$^{\text{th}}$ century. Until the end of the *Ancien Régime*, no definite list of the *French Nobility* existed. I construct a dataset of French *Ancien Régime* noble families by combining several sources. I have digitised all 19 editions of the Dictionnaire de la Noblesse by La Chesnaye des Bois and De Badier (1783). Additionally, I have digitised 16 *Annuaires de la Noblesse*[4]. For both of these digitisations, I used the `Google Gemini Flash 2.4` API. This is then supplemented with data from the *Liste des familles subsistantes de la noblesse française* on Wikipedia. This list contains all noble families (both *ancien régime* and ennobled later) with date of ennoblement,

[2]Since only wealth in the département of last legal residence is registered, real estate outside this *département* is registered at the location of said real estate and I therefore do not capture it. *Mobile wealth*, however, is fully captured for each individual.

[3]Bouches-du-Rhône, Cantal, Cher, Creuse, Deux-Sèvres, Essonne, Eure-et-Loir, Haute-Vienne, Landes, Loire, Manche, Nièvre, Nord, Paris, Pas-de-Calais, Saône-et-Loire, Seine-Saint-Denis, Territoire de Belfort, Val-de-Marne, and Yvelines

[4]I digitised the *Annuaire* for the years 1843-1848, 1851, 1853, 1855, 1857, 1859, 1861, 1863, 1867, 1869, and 1872.

Table 1: **Summary of the newly digitised wealth data**

Period	Deaths	Non-zero Wealth Obs.	In %	Avg. normalised log wealth	SD log wealth
1791-1799	85,270	14,303	16.8%	6.773	2.101
1800-1809	218,954	53,498	24.4%	6.258	1.792
1810-1819	392,273	113,755	29.0%	6.098	1.658
1820-1829	485,047	187,154	38.6%	6.051	1.776
1830-1839	631,036	307,481	48.7%	6.005	1.868
1840-1849	793,285	364,909	46.0%	6.040	1.857
1850-1859	849,601	390,131	45.9%	6.028	1.861
1860-1870	1,105,960	500,232	45.2%	6.124	1.883
All periods	4,561,426	1,931,463	42.3%	6.070	1.849

which allows for filtering for *ancien régime nobles*, of France which still feature at least one living descendant — either in France or abroad[5] and is based on different national and regional dictionaries and armorials.[6]

After the nobility was abolished by the Revolution a new *Imperial* nobility was introduced by Napoleon. To get a complete picture of the 19th century French nobility, Higgs (1987) mainly relies on the compilation of all grants of titles and ennoblements during the First Empire and the Restoration by Albert de Révérend (1895; 1901), which he calls *"the most valuable single reference work for nineteenth-century titles"* (Higgs, 1987, p. 237). I have completely digitised Campardon (1889), which also contains a full list of Napoleonic ennoblements, and use it as the basis for the list of surnames associated with the Imperial nobility.

[5] But not families with noble titles awarded by foreign sovereigns living in France.

[6] Authier, Galbrun, and Dell'Acqua (1973), Chaix d'Est Ange (1903), Jougla de Morenas and de Warren (1934), du Puy de Clinchamps (2005), de Séréville and de Saint-Simon (1975, 1977), Valette (1959, 1989, 2002, 2007), de Vaulchier, de Saulieu, and de Bodinat (2004), de Woelmont de Brumagne (1928), Frotier de La Messelière (1912), de Saint-Houardon (2007), de Rivoire de La Bâtie (1867), de Barrau (1853), and de Foras and de Mareschal de Luciane (1863).

The political elite encompasses all members of the successive national parliaments, starting with the *États Généraux* and ending with the *Assemblée Nationale*, as well as all cabinet ministers and members of the upper house[7]. The members of parliament were scraped from the *Base de données des députés français depuis 1789*. The data on cabinet members comes from Wikidata.[8] The resulting dataset contains 18,517 individuals with 13,451 unique surnames. The members of the upper house were scraped from the *Sénat*'s website.

The *Légion d'Honneur* is France's highest order of merit. Initially, it was introduced by Napoléon with the intention of creating an institution similar to the Praetorian Guard or the Janissaries (Higgs, 1987, p. 6). It was, however, not exclusively military, but also awarded to civilians. Later it was made inheritable after three generations. All available information on the members of the legion are available in the so-called *Base Léonore*, which has been scraped completely for the present paper. This yields a dataset of 381,104 awarded individuals with 109,175 unique surnames.

The *Magistrature* was the professional judiciary that was installed after the Revolution. Members of the *Magistrature* have been compiled in the *Annuaire rétrospective de la magistrature française*, which has been compiled by Farcy (2010). It contains 59,717 individuals, 29,883 of which fall within the period under investigation in this paper.

Another elite group that is included in the elite database are the authors of books — holders of significant amounts of cultural and social capital. Especially in the *Ancien Régime*, when *hommes de lettres* could not support themselves financially through writing, patronage networks were crucial. They could support themselves as authors at semi-official periodicals like the *Mercure* or the *Journal des Sçavans*, as official historiographers, or by obtaining a royal pension to *Men of Letters* (Darnton, 1982). I obtain information on authors of published books from Quérard's (1827–1839) *La France Littéraire*. This publication was compiled by the bibliographer Joseph-Marie Quérard, who pursued the plan of assembling a *"complete national bibliography"*. Quérard (1827–1839) claims to have assembled all works published in the French language, inside or outside France, between 1700 and 1830. The project was semi-official, as it was supported financially by a government subsidy granted by François Guizot, the dominant political figure of the July Monarchy. Quérard's *La France Littéraire* contains information on 104,549 titles written by 29,447 authors.

Both the *Collège de France* (Collège de France, 2007) and the *Académie Française* publish (Académie française, 2026) lists of the chairs (*Collège de France*) and the *immortels* (*Académie Française*) re-

[7]The upper house changed its name several times: *Conseil des Anciens*, *Sénat conservateur*, *Chambre des pairs*, then the *Sénat* of the Second Empire and Third Republic.

[8]Qualifiers: member of cabinet, P5054, in France, P27

spectively on their websites. Both of these have been digitised for this paper. The *Collège de France* is a university institution in Paris which is dedicated to public teaching, it has no students and its professors are free in the pursuit of their research. In the past it has been the pinnacle of many French academic careers. Its members are chosen by the existing members. The *Académie Française* is a learned society charged with the standardisation of the French language. Being appointed to one of its 40 chairs as a so-called immortal is considered to be indicative of very high cultural and social prestige. The membership in a variety of other academies can be found in prosopographical directory of the *Comité des travaux historiques et scientifiques*, which was established in 1834 to federate all provincial learned societies, which has been digitised as *La France Savante* and includes 31,617 individuals.

Educational elites are captured by the registers of professors at French law faculties, the *Corps des Mines*, and the graduates of the *École des Chartes*. The register of French law faculty professors comes from the `SIPROJURIS` project, which listed the professors active between 1804 and 1950. The *Corps des Mines* is a very selective French technical corps that recruits their members from the top-ranked graduates of the prestigious *École Polytechnique (X)* and has existed since 1783. The *École des Chartes* is the elite university *(Grande École)*, which trains archivists.

The French patent law of 1791 created the so-called *brevets d'invention*, which have been digitised. This means that a public database of 44,962 patent holders is available via the *Institut National de la Propriété Industrielle*.

Finally, from Wikidata, I obtain lists of members of the *Académie des sciences* (8,030), *Académie de médecine* (4,301), the *pairs de France* (peerage) (1,881), the *conseillers d'état* (1,194), the *préfets* (572), as well as military officers (8,932). These are not complete lists.

A detailed description of these different elite groups and the number of observation per dataset can be found in appendix A.5.

3 Empirical Approach

3.1 Wealth

Building on Clark and Cummins (2015), I construct a measure of population average normalised wealth at death for each person i dying at time t:

$$w_{it} = \ln(wealth_{it}) - \overline{\ln(wealth_t)} \tag{1}$$

I then define wealth bands for the different surname groups based on their wealth at death within the 1821-1845 window. This is the second 25-year generation window, as the data starts in the 1790s. The *Tables des Successions et Absences* from the first generational window have the thinnest coverage of the entire sample. This is both because older records are less likely to have survived to the present day and because of an administrative change. In 1825, the *enregistrement* switched from the so-called *successions acquittées* to the *tables des successions et absences* proper. The earlier format had a slight bias towards over-representing the more affluent. The TSA proper was conceived precisely to counteract this bias and create a synthetic document combining all information available to the *enregistrement* (J. Bourdieu, Kesztenbaum, & Postel-Vinay, 2014, p. 100). This means that the 1821-1845 generation is the earliest available high-quality data. This allows me to calculate two persistence measures: one forward, to the generation 1846-1870, and one backward to the generation 1796-1820.

Because I only have a handful of observations per surname — my sample is restricted to rare surnames — each surname mean is measured from only a handful of estates during any given decade and the measured mean carries a lot of symmetric sampling noise, meaning individuals who are rich (or poor) for random reasons of luck — or a high u_{it} term in equation 2:

$$w_{it} = x_{it} + u_{it} \tag{2}$$

Where w_{it} is wealth at death, u_{it} is measurement error, and x_{it} is underlying latent status. This means that by ranking surname groups on these noisy means, I select, in part, based on noise u_{it}. Clark and Cummins (2015) resolve this problem by looking at several generations. After a few generations, the "spurious" regression to the mean washes out and the underlying persistence b of the latent factor x can be uncovered. Because my data only features three generations and bands are defined in the middle generation due to data considerations, this is not possible in the case at hand. I therefore apply a sample split inspired by Angrist and Krueger (1995) to sort surnames into wealth bands. The idea behind this approach is to filter out luck.[9]

I assign the surnames to 4 bands: the top 5%, the top quarter excluding the top ventile, the middle 50%, and the bottom quartile. Table 2 alphabetically lists the first eighteen surname groups in each wealth band. The importance of this table is to show that most of these names do not contain any explicit signal about their wealth and social status. It is worth noting, however, that like in England (Clark et

[9]For every surname group, deaths between 1821 and 1845 are allocated at random to two halves, A and B. Half A is then used to rank the surname groups by wealth, while half B is used to compute band-specific 1821-45 means. The procedure is repeated for 50 random splits; reported estimates are medians across splits.

al., 2015), (potentially) locative surnames (*Aubusson, Angles etc.*) are more prevalent at the top of the wealth distribution than among the bottom of the wealth distribution, where originally patronymic first-name-derived surnames (*Amable, Ambrois etc.*) are more prevalent. As I am investigating the 19th century, non-French surnames do not play a significant role yet.

Table 2: The Rare Surname Groups

Top 5%	5–25%	Bottom 25%
Achte	Adenot	Abbe
Acquart	Adillon	Abe
Aguttes	Adrot	Abeille
Alin	Agache	Albaron
Amery	Agnes	Albisson
Amon	Aguillon	Allamelle
Andron	Aix	Allec
Angelard	Alamartine	Allegraud
Angiboust	Alary	Allegros
Angles	Alavoine	Alluchon
Antheaume	Alazard	Almon
Ardaillon	Albert	Alouis
Arm	Albessard	Alric
Armieux	Alexandre	Amable
Asseline	Alivon	Amadieu
Aublet	Allegre	Ambrois
Aubusson	Allonneau	Amel
Aucaigne	Almy	Amiel

Notes: The first eighteen surname groups within each wealth band in alphabetical order, as example of the 11,569 rare surname groups which have been ranked by mean of normalised log wealth at death within the definition window 1821-1845. In case of several spelling variants which have been grouped, the displayed name corresponds to the most frequent spelling. The middle band (25th-75th percentile) was omitted.

Table 3 conducts the same exercise for our different definitions of *aristocratic* surnames. The first column contains surnames from the wealth dataset with the highest confidence of representing *ancien régime aristocracy*: These names feature the *de-* particle and appear at least twice with an entry in the *ancien régime* nobility registers described in section 2.2. The second column shows, in alphabetical order, the first 15 surnames in the wealth dataset with a particle indicating nobility (*de*, *d'*, or *du*) but

that do not fulfill the strict requirements of column 1. This reveals the inherent difficulty of pinning down an *ancien régime* nobility in the absence of an official definite list or definition of the old regime aristocracy. Names like *d'Anneville*, *d'Apremont*, or *de Barville* in the second column may well belong to noble families that are missing from the lists. On the other hand side, the list includes Flemish names likes *d'Eyck* or *d'Hondt*, where the particle does not indicate nobility, as well as probably some self-styled nobles without *ancien régime* pedigree or later ennoblement.

Table 3: Aristocratic Rare Surname Groups

Old nobility (particle + ≥2 lists)	Particle, <2 lists	≥2 lists, no particle
Abon	d'Agnel	Abelly
d'Aboville	d'Anin	Abon
Saint-Abre	d'Anneville	Abot
d'Abzac	d'Anon	Agard
d'Agneaux	d'Apremont	Aignan
Agoult	d'Argaud	Aiguillon
Saint-Aignan	du Aron	Ailhaud
d'Aigremont	d'Aucourt	Aix
Aiguillon	d'Aude	Alban
Ailly	de Aute	Alberti
Aix	d'Autun	Albon
d'Albignac	de Avon	Alencon
d'Albis de Razengues	de Bans	Amaury
d'Albon	de Barnot	Ambert
Alencon	de Barville	Amedee

Notes: The first fifteen families (alphabetical by family name) under each definition, applied to each surname *as written* in the registers. *Old nobility*: the surname carries a particle (*de*, *d'*, *du*, *des*) and the name after it, or the patronym before it, is listed in at least two of the four *ancien régime* sources described in section 2.2; the name is rare (fewer than 2,000 INSEE births) and not a common first name; names are shown as printed in the lists. *Particle, <2 lists*: surnames written with a particle whose family name is in fewer than two lists (later adopters, Flemish names, unlisted families). *≥2 lists, no particle*: listed names written without a particle, which the joint requirement excludes. Columns 2 and 3 are restricted to INSEE surnames with at least three decedents.

3.2 Multiple Elite Outcomes

To analyse the persistence of social and cultural capital directly, I employ an approach developed by Eeckhout et al. (2023a) that uses data on multiple elite outcomes and measured over-representation at the surname-level to quantify intergenerational persistence of the latent factor. The method of using over-representation among elite groups as status measure was pioneered by Clark (2015) and Clark

and Cummins (2014), multiple outcomes have been used by Vosters and Nybom (2017) and Eeckhout et al. (2023b).

The goal of Eeckhout et al.'s approach is to uncover the persistence of the latent factor z_t. As this latent factor is assumed to be decaying by a constant factor, γ in their notation,the following pooled OLS regression can uncover γ:

$$log(\tilde{z}_{k,t}) = \alpha + \kappa \cdot \frac{t - t_1}{30} + \epsilon_{k,t} \tag{3}$$

Where k is one of the different elite groups described in 2.2, t is the median year of one of the generations described in the previous section 3.1, and $\tilde{z}_{k,t}$ is estimated status of elite group k at time t. With estimated status being measured by averaging relative over-representation among n different elite outcomes j. γ is defined as $\gamma = e^{\kappa}$. The detailed derivation of the regression framework and the measure of status can be found in appendix C.1.

An elite groups k is defined as a group that was member of an elite in t_1, in our case in the *ancien régime*. These are the groups whose elite persistence this approach uncovers. Elite outcomes j are outcomes that are realised in the subsequent periods. The data on elite groups and elite outcomes is described in section 2.2. A detailed overview of this data can be found in appendix A.5.

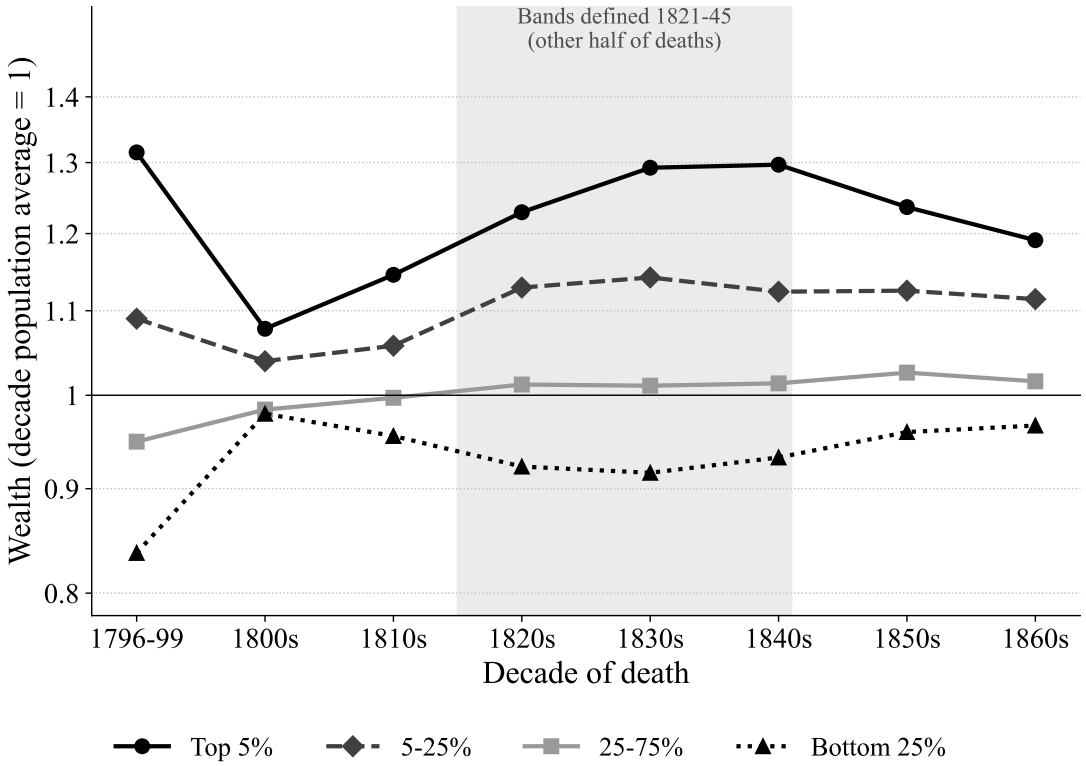
4 Empirical Results

4.1 Wealth

Figure 1 plots the wealth trajectories of these four wealth bands.

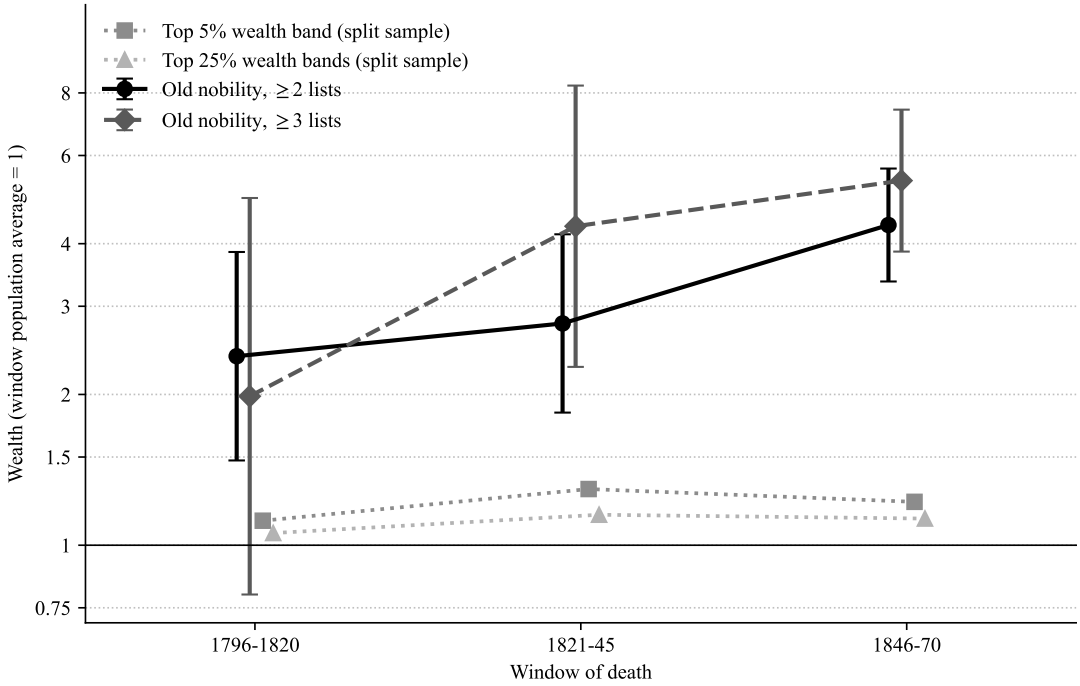
The results are similar to the ones obtained by Clark and Cummins (2014, 2015), N. Cummins (2024), and N. J. Cummins and Gráda (2025). After the generation in which the wealth bands have been defined, the top and bottom bands revert only slowly towards the mean: the top 5% fall from 1.30 times the population average in the 1840s to 1.19 times in the 1860s, while the bottom quartile rises from 0.93 to 0.97 times the average. The middle band remains close to the mean throughout. The strict definition of *ancien régime* nobility — a particle *and* being verified in two sources on the *ancien régime* aristocracy — is applied in figure 2. Because the definition is strict, the group is small: 1,313 decedents from 581 families. Here, I can observe no mean reversion at all between generation 1 and generation 2. This is consistent with the idea that aristocratic wealth represents a less noisy signal of the underlying latent factor than non-aristocratic wealth, as noise has been filtered out across generations (Clark & Cummins, 2014, 2015).

Figure 1: Wealth trajectories of surname groups defined by wealth



Notes: Person-weighted mean wealth at death by decade of death, log scale, relative to the mean of all decedents in the same decade (= 1). The wealth bands are defined in 1821-45 over rare surname groups and held fixed for all decades. Bands are ranked on one random half of each surname's 1821-45 deaths and the 1821-45 decades are measured on the other half (Angrist & Krueger, 1995); means over 50 random splits.

Figure 2: Wealth trajectories of *Ancien Régime* noble surname groups



Notes: Mean wealth at death relative to the population mean of the same 25-year window (log scale), for decedents whose surname as written carries a particle and whose family name is listed in at least two (black) or three (dark grey) of the four *ancien régime* sources; rare names only. 95% confidence intervals from 1,000 bootstrap draws over families. Grey: the top 5% and top 25% wealth bands, ranked on one random half of each surname's 1821-45 deaths and measured on the other, averaged over 50 random splits. Nobility membership does not depend on wealth.

In a next step, I calculate intergenerational persistence parameters. Clark and Cummins (2015) have derived an estimate of intergenerational wealth elasticity through measuring the rate of movement of wealth among wealth bands towards the societal wealth mean. As Clark and Cummins (2015) show, this uncovers an estimate of the persistence rate b of the underlying latent factor (See equation 2).

$$\hat{b}_A = \beta_A = \frac{\bar{w}_{ikt+1}}{\bar{w}_{ikt}} \tag{4}$$

As groups are defined in generation 1 (1821-1845), this can be done forward, as well as backward, which yields two persistence estimates each:

$$\hat{b}_A^{\text{fwd}} = \frac{\bar{w}_{i,g_2}}{\bar{w}_{i,g_1}}, \qquad \hat{b}_A^{\text{rev}} = \frac{\bar{w}_{i,g_0}}{\bar{w}_{i,g_1}} \tag{5}$$

The results of this exercise can be found in table 4. The intergenerational elasticity estimates of the calculated at the surname group level come very close to the estimates that Clark and Cummins (2015) have found for England for the period of 1858-2012.[10] The persistence estimates are very high and indicate a very sclerotic social world. Both the top ventile and the top quartile are extraordinarily stable with forward persistence rates of 0.74 and 0.87 respectively. This is even slightly higher than the rates found by Clark and Cummins (2015) for England. These coefficients are significantly higher than the ones by J. Bourdieu et al. (2019), who use bi-generational links from the *Enquête TRA* dataset and find coefficients between 0.3 and 0.42. This difference is, according to Clark and Cummins (2015), due to attenuation.

[10] Clark and Cummins (2015) find persistence parameters between 0.64 and 0.81 for the *rich*, and between 0.71 and 0.87 for the *prosperous* wealth bands, as well as between 0.23 and 1.01 for the *poor* wealth band.

Table 4: Intergenerational wealth persistence: surname groups

	Reverse (1821-45 → 1796-1820)	Forward (1821-45 → 1846-70)
Surname-group level (person-weighted mean ratios)		
Top 5%	0.448 (0.242)	0.739 (0.156)
Top 25% (rich composite)	0.337 (0.138)	0.869 (0.097)
Bottom 25%	0.503 (0.220)	0.438 (0.117)
Name-level (surname mean on lagged surname mean)		
Split-sample IV[§], by transition[†]	0.263 (0.135)	0.817 (0.164)
Split-sample IV[§], pooled	0.625 (0.114)	
Grouping estimator (individual on lagged surname-group mean)		
Split-sample IV[§], by transition[†]	0.271 (0.102)	0.760 (0.087)
Split-sample IV[§], pooled	0.563 (0.075)	

[†]Forward transitions: g0→g1 and g1→g2.

[§]Lagged surname-generation mean split into two random halves;

half A instrumented by half B, removing sampling noise in the lagged mean.

Notes: Groups defined 1821-45. Under a stationary AR(1) latent factor the group mean decays symmetrically either side of the definition window, so both ratios estimate b. 25-year windows; $b_{30} = b_{25}^{1.2}$. Middle and 5-25% bands were omitted. Split-sample: bands ranked on one random half of each surname's 1821-45 observations; the 1821-45 band mean uses the other half (Angrist & Krueger, 1995). Split-sample rows report the median over 50 random half-splits; standard errors in parentheses are the median of $\sqrt{se_k^2 + (b_k - \text{median})^2}$, adding the split-to-split variation (Chernozhukov et al. 2018); within a split, delta-method SEs for the band ratios and surname-clustered SEs for the IVs.

The backward estimates in Table 4 are considerably lower than the forward estimates. This difference is largely due to improvements in wealth measurement, however. Before 1824, most registers record a single declared total wealth number for each estate; from 1824 onwards, mobile wealth and real estate are recorded separately. Table 5 compares the two formats within the same generation. Surname mean wealth in 1821–1847 computed from the declared totals is correlated with surname mean wealth

in 1848–1870 at only 0.20, net of sampling noise, whereas surname mean wealth in the same window computed from the separately recorded mobile wealth and real estate correlates with it at 0.80. The same families therefore appear more mobile because the *Tables des Successions et Absences* prior to 1824 measured their position less precisely. In equation 2, the older format carries a larger noise component u_{it}, which attenuates the backward estimates, while the persistence of the underlying latent factor is unchanged.

Table 5: Wealth measured before and after the register reform of 1824

Surname mean wealth in 1821–1847 computed from	Correlation with 1848–1870	Correlation with
Declared totals (the pre-1824 format)	0.200 (0.105)	0.304 (0.2
	7,130 surnames	5,485 surna
Mobile wealth and real estate (recorded separately from 1824)	0.804 (0.087)	0.476 (0.1
	12,074 surnames	6,936 surna

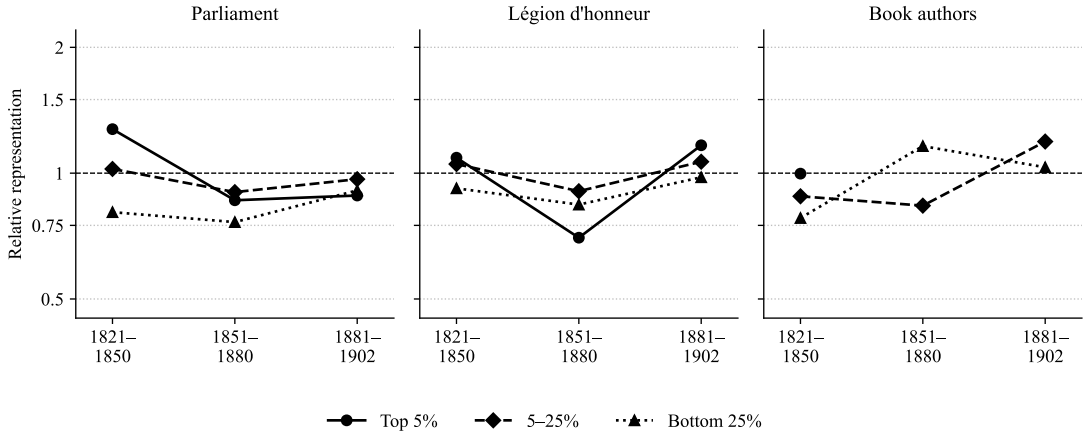
Notes: Correlation between surname mean log wealth gaps in two windows (1791–1820, 1821–1847, 1848–1870), divided by the square root of the product of the split-half reliabilities of the two window means, which removes the attenuation from sampling noise in the surname means. In 1821–1847 the surname mean is computed either from the declared totals or from the separately recorded mobile wealth and real estate; the other window uses all recorded amounts. Rare surnames with at least three decedents in each window. Standard errors in parentheses from a bootstrap over surnames.

The high intergenerational persistence estimates reveal that the underlying latent factor is very persistent — as persistent as in Britain. This is consistent with any latent factor model, not necessarily the Bourdieusian model introduced in section 5. This is where the elite status data described in section 2.2 becomes useful to investigate whether the *Bourdieusian* model accurately describes what I am observing.

Figures 3 and 4 show that different families hold very different capital portfolios. The *ancien régime* nobility accounts for between 2% and 16% of the deputies entering parliament under the different regimes (Figure 4), far more than its share of about 0.03% of all decedents — names with a *de*

particle at times represent more than 70% of the deputies entering parliament.[11] The surnames of the wealthiest families, by contrast, are only modestly overrepresented in the elite registers (Figure 3): the top 5% of the wealth distribution are 1.3 times overrepresented in parliament in 1821–1850 and slightly underrepresented thereafter, and they are proportionally represented among the book authors listed in the "national bibliography" (*La France Littéraire*) Quérard (1827–1839). This suggests that the old regime nobility indeed is not only a high-wealth group, but also relatively highly endowed with upper tail social capital.

Figure 3: Relative representation (share of the elite list divided by share of INSEE births; population = 1, dashed line) of the wealth bands, 1821–1902



Notes: "Relative representation" is defined as the group's share of the membership of a given elite register relative to its share of births recorded by INSEE between 1891 and 1900. Relative representation of 1 denotes proportional representation, as a log scale is employed. The counts come from the complete elite registers and are not contingent on the availability of wealth data. The representation of the *ancien régime* nobility, identified from the surnames as written, is shown in Figure 4.

Table 6 compares the observed wealth persistence of the identifiable old nobility with the *"rich"* group's persistence. The underlying assumption is that while the *"rich"* group was selected on wealth, the old nobility is selected on a persistent, old, and large endowment with social capital — the noble title as such *is* a form of institutionalised social capital. The observed wealth persistence of families which have been selected on (assumed) higher total capital, is higher than the one for families selected on wealth alone, although the difference is imprecisely estimated. This is the case despite the period under consideration being one of significant institutional change and instability.

[11] The share of deputies with a *de* particle fluctuates over time but between 1791 and 1870 averages 28%, with deputies dated by the start of their mandate. It is highest during the Restoration, when it peaks at 72% in 1822, and lowest in the Third Republic, when only 16% of deputies have a surname with a *de* particle.

Table 6: Wealth persistence of Nobles (Institutionalised social capital)

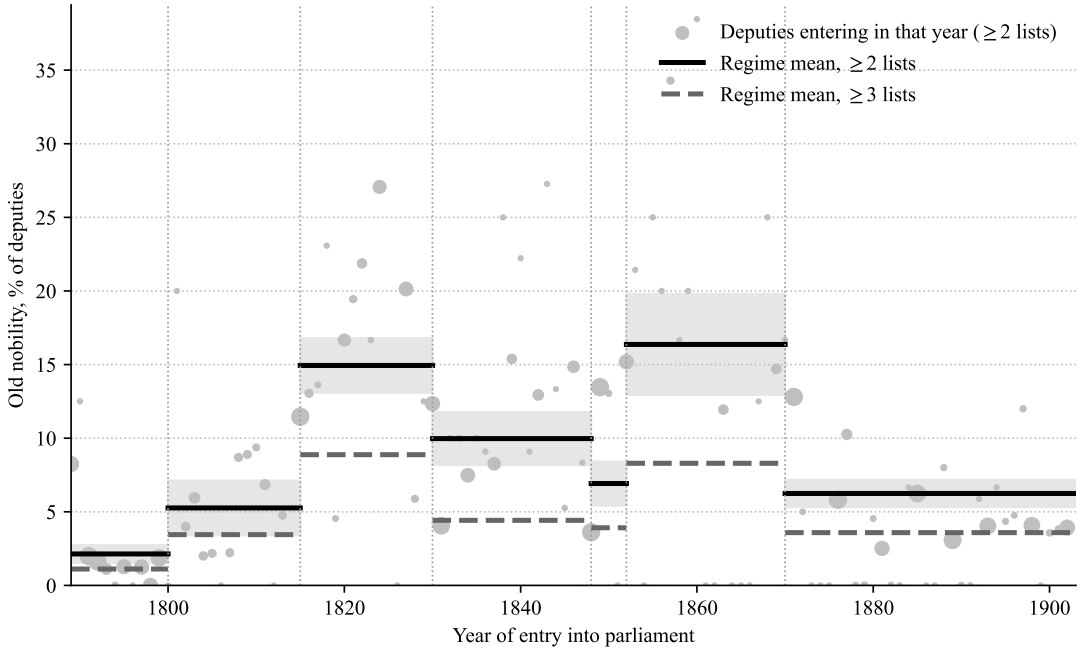
Group	$\hat{b}$ (1796-1820 → 1821-45)	$\hat{b}$ (1821-45 → 1846-70)	Pooled $\hat{b}$	Families (by window)
Old nobility, ≥2 lists	1.173 (0.529)	1.444 (0.426)	1.302 (0.235)	43/164/269
Old nobility, ≥3 lists	–	1.143 (0.335)	–	18/68/137
≥2 lists, dép. × window mean	–	1.399 (0.428)	–	43/164/269
≥3 lists, dép. × window mean	–	1.080 (0.318)	–	18/68/137
Memo: top 25% wealth bands	n.d.‡	0.869 (0.097)	0.552 (0.123)	

–: bootstrap SE > 1 (the 1796-1820 base has few families and a mean near zero); suppressed.

Notes: Ratios of mean wealth gaps (log wealth minus the population mean of the same 25-year window) between windows; pooled $\hat{b} = \sqrt{\bar{w}_{g_2}/\bar{w}_{g_0}}$. Old nobility: decedents whose surname as written carries a particle and whose family name is listed in at least two (or three) of the four *ancien régime* sources; rare names only. Standard errors from 1,000 bootstrap draws over families; ratios with a standard error above 1 are suppressed. Families: number of families with a wealth observation in 1796-1820 / 1821-45 / 1846-70. Nobility groups require no split sample, as membership is wealth-independent. Memo row: rich composite, median over 50 random half-splits (Table 4).

‡Not defined: this margin is the selection run-up forward into the definition window ($\approx 1/b$).

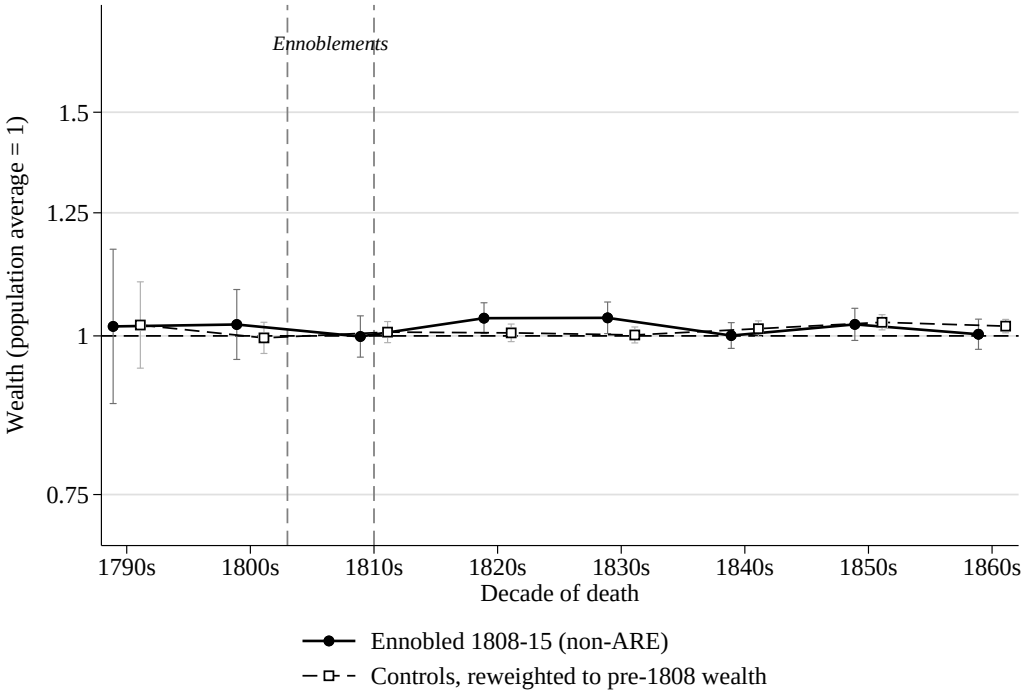
Figure 4: Noble political representation across seven regimes



Notes: Share of deputies entering parliament in a given year (grey, marker size proportional to the number of entrants) whose surname as written carries a particle and whose family name is listed in at least two of the four *ancien régime* sources described in section 2.2; rare names only. Horizontal lines: regime means with binomial 95% confidence intervals (shaded), for at least two (solid) and three (dashed) lists. Deputies are dated by the start of their mandate. The rule is conservative, so the shares are lower bounds.

Figure 5 shows, however, that social capital can be regime-dependent in its capacity to generate economic revenue. It compares the evolution of the wealth of rare surname groups associated with a Napoleonic noble title with a control group of rare "commoner" surnames selected such that the pre-1808 — i.e. pre-ennoblement — wealth distribution matches the ennobled subset. Napoleonic ennoblement is associated with wealth slightly but significantly above the population average — by about 3% — in the 1820s and 1830s, and slightly above a control group matched on pre-ennoblement wealth, although the difference from the control group is only marginally significant. The control and treatment groups are balanced before the treatment. Their slight advantage disappears in the 1840s and is no longer statistically distinguishable from zero until 1870. A Napoleonic noble title is no longer associated with above average wealth, indicating that the title had ceased to be a revenue-generating asset by the 1840s, although it was still associated with higher wealth during the Restoration and the early July Monarchy. This contrasts with the effect of *ancien régime* titles documented previously.

Figure 5: Imperial ennoblement event study



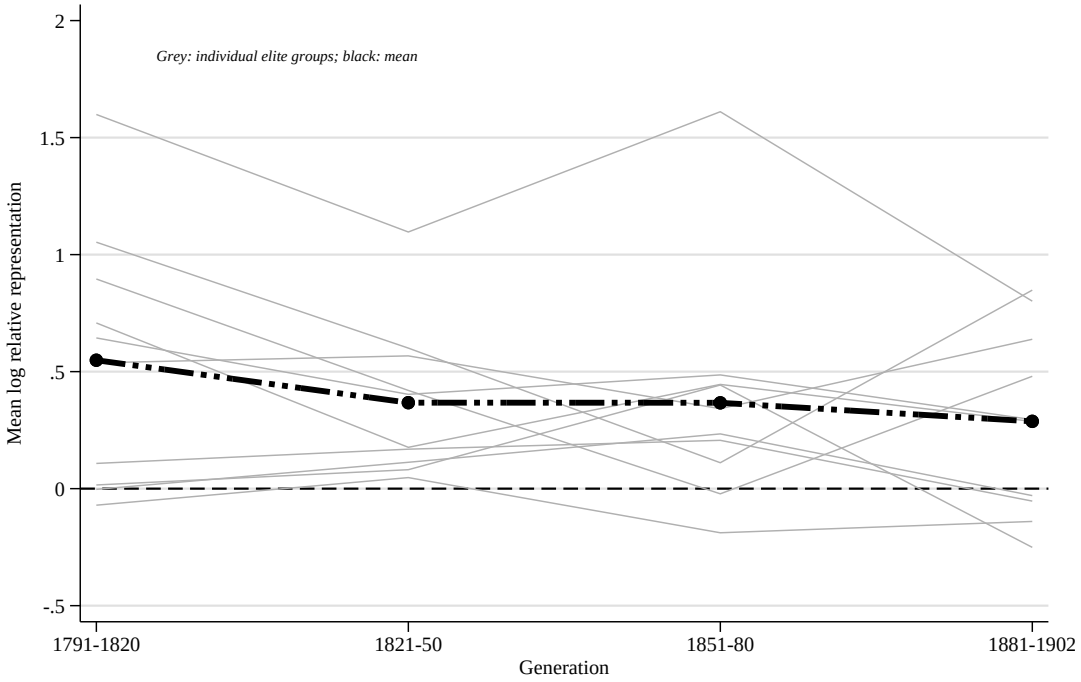
Notes: The treatment are 255 rare surnames ennobled during the Napoleonic period 1808-1815. The control group are 2,477 rare unparticled surnames without ennoblement, reweighted to match the treatment group's wealth distribution 1791-1807. 95% confidence intervals are calculated from cluster-robust standard errors on the level of the surname group.

4.2 Elite Representation

Section 3.2 described Eeckhout et al.'s (2023a) method of recovering the persistence of the latent factor from data on multiple elite outcomes. Their approach can be applied to data on elite outcomes described in section 2.2: the 11 *ancien régime* elites that can then be traced through the registers of elite outcomes (for both, see table 16 in appendix A.5) over the period 1791-1902, containing four 30-year generations.

Figure 6 shows the trajectory of the pre-revolutionary elite groups over the course of the 19$^{\text{th}}$ century by plotting their mean log relative representation among the 15 elite outcome registers, such as the parliament, the *Légion d'Honneur*, or the *Corps des Mines*. The mean elite status never converges with the societal mean at 0 until 1902 and their decay rate is very slow. The mean declines from 0.55 to 0.29 in logs, which corresponds to a relative representation of 1.7 to a relative representation of 1.3.

Figure 6: Trajectory Elite Over-Representation



Notes: Mean log relative representation by pre-revolutionary elite surname group k over the course of four "generations" of 30 years each (1791-1902). Grey lines are individual groups, the black line is the unweighted mean. The sample is restricted to sharply defined elite groups, meaning fewer than 600 surname groups. The unrestricted figure is shown in appendix C.2 figure 11.

Table 7 reports the results from the regression specified in equation 3. The headline result is the persistence coefficient $b_z = 0.942$ per 30-year generation. This is statistically indistinguishable from the persistence result derived from the wealth data in section 4.1. This is an independent confirmation of the high persistence result in the preceding section using a different method and outcome space. This result is robust to different specifications. Table 7 and figure 6 focus on sharply defined elite groups with fewer than 600 different surname groups. The results are robust to including all groups, however (0.936).[12]

[12]See figure 11 in appendix C.2.

Table 7: Decay Rate Elite Over-Representation

	$\hat{\kappa}$ (SE)	$\hat{b}_{30}$	$\hat{b}_{25}$	N
Main: sharply defined elite groups				
Trend only	-0.060 (0.013)	0.942	0.951	44
Robustness (all trend only)				
All elite groups	-0.066 (0.015)	0.936	0.946	64
Scraped registers only (no Wikidata)	-0.027 (0.011)	0.973	0.978	44
Excluding registers with incomplete coverage	-0.058 (0.010)	0.943	0.952	44
Within group × register fixed effects	-0.053 (0.018)	0.948	0.957	368
Balanced pairs only (all four generations)	-0.042 (0.016)	0.958	0.965	264
By register (trend only)				
Parliament	-0.047 (0.018)	0.954	0.962	56
Légion d'honneur	-0.006 (0.011)	0.994	0.995	64
Magistrature	-0.003 (0.012)	0.997	0.997	58
Savants (CTHS)	-0.049 (0.025)	0.952	0.960	57
Sénat	-0.087 (0.049)	0.916	0.930	43
Corps des Mines	0.032 (0.054)	1.033	1.027	29
École des chartes	-0.121 (0.092)	0.886	0.904	28
Law professors	-0.019 (0.069)	0.981	0.984	33

Notes: 16 *ancien régime* elite surname groups observed through 15 registers over four 30-year generations, 1791–1902, yielding 632 group×register×generation cells with at least three group members. The main specification restricts the analysis to 11 sharply defined groups (defined as containing less than 600 surname groups each), 368 cells; the size threshold is set ex ante, since selecting on initial RR would manufacture decay by regression to the mean. Relative representation RR is defined as the group's representation in the register over its INSEE population share. Outcome values are averaged across registers within a group-generation, so the pooled regressions have $N = 44$ group-generation observations clustered on 11 elite groups; by-register rows use the underlying cells. $\log RR_{k,t}$ is regressed on the generation index by pooled OLS, with $b_{30} = e^{\kappa}$ and $b_{25} = b_{30}^{25/30}$. All rows are trend-only and therefore directly comparable.

Political and honorific status thus persists at a rate similar to wealth. Table 8 asks whether it is the same latent factor. If wealth and status in the elite registers were signals of one and the same latent factor, a family's representation in the registers would help predict the wealth of the next generation

once its current wealth is known, because current wealth measures the latent factor with error (Braun & Stuhler, 2018). I regress surname mean wealth in 1848–1870 on surname mean wealth in 1821–1847, instrumented with the split-sample approach described in section 3.1, and on the surname's relative representation in parliament and in the *Légion d'Honneur*. The registers add nothing to the prediction: their coefficients are close to zero and precisely estimated, also when they are measured a generation earlier, in 1791–1820, and the persistence of wealth is unchanged at 0.83. Political and honorific status persist alongside wealth rather than through it.

Table 8: Status beyond wealth: elite registers and the next generation's wealth

	(1)	(2)	(3)	(4)
Surname mean wealth 1821–1847	0.833 (0.156)	0.835 (0.157)	0.833 (0.156)	0.833 (0.157)
Parliament, 1821–1850		−0.021 (0.018)		
Légion d'Honneur, 1821–1850		0.012 (0.017)		
Any deputy, 1821–1850			−0.019 (0.029)	
Any *Légion* member, 1821–1850			0.012 (0.031)	
Parliament, 1791–1820				−0.013 (0.016)
Légion d'Honneur, 1791–1820				0.020 (0.017)
First-stage F	52	52	52	52
Surnames	11,385	11,385	11,385	11,385

Notes: Dependent variable: surname mean log wealth gap in 1848–1870. Wealth in both windows is the sum of the separately recorded mobile wealth and real estate. Surname mean wealth in 1821–1847 is instrumented: the surname's deaths are split into two random halves, the mean of one half is instrumented by the mean of the other; reported estimates are medians over 50 random splits, standard errors in parentheses include the split-to-split variation (heteroskedasticity-robust within each split). Parliament and *Légion d'Honneur*: $\log(1 + RR)$, where RR is the surname group's share of the register in that generation divided by its share of INSEE births 1891–1900 (0 if absent). Any deputy / any member: indicators. Rare surnames (fewer than 50,000 INSEE births) with at least three decedents in each window. All coefficients are given to three decimals; first-stage F rounded.

5 Discussion and Theory

The results in section 4 establish five facts. First, the surname-group wealth persistence coefficient is between 0.74 and 0.87. This is roughly two or three times the level of the bi-generational estimates

around 0.3 found by J. Bourdieu et al. (2019). Second, wealth and political status are only loosely linked. The surnames of the top 5% of the wealth distribution are only modestly overrepresented in parliament (1.3 times in 1821–1850) and proportionally represented among the book authors listed in Quérard (1827–1839), while the *ancien régime* nobility accounts for between 2% and 16% of the deputies, far more than its share of decedents; and representation in the elite registers does not predict a family's future wealth once its current wealth is known. Third, families with surnames traditionally associated with hereditary *ancien régime* noble titles show no detectable mean reversion, while families sorted on wealth itself show a mean reversion in the range found by Clark and Cummins (2015) (0.74-0.87). Fourth, by contrast, surnames associated with a Napoleonic nobility title command a small wealth premium during the Restoration and July Monarchy [13], but not thereafter. Finally, rare surnames associated with the *ancien régime* nobility remain strongly overrepresented in parliament throughout the century, although their share of deputies varies with the regime, peaking under the Restoration and the Second Empire (Figure 4); their wealth shows no sign of decline across the three generations (Figure 2).

I propose a framework which can account for all these five observations, as well as the mechanisms for elite persistence proposed in Alesina et al. (2020) and Ager et al. (2021). Families hold portfolios with different forms of capital with different decay rates and different returns on capital. Besides economic capital, families hold stocks of what P. Bourdieu (1984, 2011) has called social and cultural capital. The former consists of networks, connections, and reputation, but also has institutionalised forms, such as in the form of noble titles. The latter manifests in education, as well as attitudes or culture.

These stocks are capital or an asset in the economic sense because of three of their properties. First, they take time to accumulate.[14] Second, acquiring education, tacit human capital, or social prestige requires an investment of time and resources. Third, like other assets, they generate an economic return in each period. These economic returns can come from wages from appointments to positions of prestige, *patronage*, advantageous marriages and associated dowries, access to credit networks etc.

Like financial capital, cultural and social capital can be inherited across generations. Like wealth, it cannot be inherited perfectly, as it depreciates over time and the inheritance is less direct than in the case of financial capital. The different forms of capital differ in terms of their persistence rates, their returns and their vulnerability to shocks. Financial capital and material property have relatively high variance. It can be gained in a windfall but is vulnerable to political shocks and critical junctures like revolutionary expropriations. Social and cultural capital by contrast, has less variance and is

[13] They were most commonly awarded for military achievements (Higgs, 1987).

[14] As P. Bourdieu puts it, *"[t]he accumulation period [of cultural capital] covers the whole period of socialization."*(P. Bourdieu, 2011, p. 81)

less vulnerable to shocks. Given these differing persistence rates, accumulation timelines, variance, and returns on investment, families with a dynastic motive choose a capital investment portfolio that maximizes intergenerational total capital.[15] Because of their higher persistence and security from political infringement, allocating capital to non-economic forms can be an attractive strategy to preserve capital — especially in times of institutional change and political shocks.

Each stock of non-economic capital evolves according to the following dynamic:

$$z_t = b_z z_{t-1} + \varepsilon_t^z \tag{6}$$

Where b_z is the persistence rate of capital of form z and ε_t^z is the specific error term. b_z of cultural and social capital is high and the variance $\text{Var}(\varepsilon^z)$ of the error term is smaller than for economic capital: windfall profits or sudden losses of cultural capital are not possible and unlikely for social capital.

Economic capital K_t evolves according to the following dynamic:

$$K_t = b_e K_{t-1} + r_t^z z_{t-1} - I_t + \varepsilon_t^K \tag{7}$$

with observable wealth being a noisy projection of underlying economic capital: $w_t = K_t + u_t$. Each period, the stock of economic capital is replenished by returns on all forms of capital and depleted by investments in them. The decay rate b_e is faster than b_z and the variance of the error term, $\text{Var}(\varepsilon^K)$ is high because of the possibility of windfall profits and losses. Windfall profits ε_t^K decay at the fast rate b_e. r_t^z, the economic return on capital of form z is institutionally specific and changes over time t.

As families with a broader capital endowment profit from their stock of economic capital being replenished through flows from their wider endowment with social and cultural capital. Families with a broader capital stock, such as established noble families, thus have higher apparent persistence and a lower variance of the error term. Families that are *merely rich* but do not have a broader endowment of capital exhibit faster apparent decay and a high variance of the error term. These might only be part of the *rich* group because of windfall profits which decay fast.

Observed wealth persistence is therefore not just attenuated by the attenuation factor φ, it is also different from b_e because of the reserve bonus Δ:[16]

[15] This is what P. Bourdieu (2011) terms a *reproduction strategy*.

[16] The derivation of the expression of observed wealth persistence in equation 8 can be found in appendix B.2. The model proposed in this paper is a special case of the framework developed by Del Pizzo et al. (2026). How the model relates to their framework is explored in appendix D.

$$\text{plim}\,\hat{\beta}_w = \varphi(b_e + \Delta) \tag{8}$$

Δ is larger for families with a large stock of broader capital z_t as their economic capital is replenished by returns on z_t each period. This means that families without a broader capital stock exhibit faster apparent decay — a low plim $\hat{\beta}_w$ — and high variance in the error term ε_t^K, while families with more broader capital stock exhibit higher apparent persistence and lower variance of the error term.[17]

The theoretical framework can explain the empirical facts established in section 4. Top 5% families sorted on wealth alone regress to the mean at a rate comparable to the one found by Clark and Cummins (2015) for England in the 19$^{\text{th}}$ century. Families sorted on *ancien régime* noble titles show no detectable sign of reversion to the societal wealth mean. Sorting on wealth alone partly selects on economic windfall profits ε_t^K, which depreciate at the faster rate b_e. These families potentially have no wider social and cultural capital z_t, which would depreciate more slowly. Selecting on a noble title by contrast, selects on z directly, as a noble title can be seen as an embodiment of social capital. Noble families, however, besides the title *being* high social capital empirically also hold more political capital in our sample. The *ancien régime* nobility accounts for a far larger share of deputies than of decedents (Figure 4), whereas the surnames of the wealthiest commoner families are only modestly overrepresented in parliament (Figure 3). Representation in the elite registers, however, adds little to the prediction of a family's future wealth once its current wealth is known (Table 8), which suggests that the return on z_t runs through the family's name and network rather than through these positions themselves.

The higher persistence b_z of z_t is evident in the slow decay of elite over-representation. Across the fifteen elite outcomes, the over-representation of *ancien régime* elite surnames decays at $b_z = 0.94$ per generation (Table 7). The *ancien régime* nobility remains strongly overrepresented in parliament under every regime, although the extent of its representation varies with the regime (Figure 4). Its wealth shows no detectable reversion to the mean: the ratio of its mean wealth gap in 1846–70 to that in 1821–45 is 1.44 (0.43), and 1.14 (0.34) under the stricter definition (Table 6). These estimates are imprecise, but they are consistent with an observed wealth persistence close to one for these families. In the model, this is the result of a continuous return on z_t, which could take the form of advantageous marriages, appointments, and access to business and credit networks.

Napoleonic noble titles illustrate the role of r_t^z from the opposite direction. Families ennobled between 1808 and 1815 were slightly richer than commoner families of comparable pre-1808 wealth in the 1820s

[17] This is generalising the framework by Clark and Cummins (2015). The two models are the same if I assume b_e to approach zero and conceptualise the return stream rz as the latent factor x.

and 1830s, but this small advantage had disappeared by the 1840s (Figure 5). A title conferred by a new regime, without the accumulated social and cultural capital it was meant to signal, earned only a small and short-lived return r_t^z.

Both the results and the theoretical framework are in line with the existing literature on elite persistence, as it organises the proposed channels of elite persistence into one coherent framework. Bleakley and Ferrie (2016), studying the small intergenerational effects of the Cherokee Land Lottery of 1832, provide empirical evidence for the low persistence of windfall profits ε_t^K. Alesina et al. (2020) argue that the remarkable persistence of Chinese elites across the Communist and Cultural Revolution in China can be explained through the transmission of *"informal human capital"* and *"social capital"*, finding evidence for the former in value surveys. This is evidence for the state having been able to confiscate financial capital quite effectively — a negative shock to ε_t^K — while the more persistent z_t remained stable, even during a period when its economic return r_t^z was close to zero. Ager et al. (2021) too find suggestive evidence that Southern slaveholders recovered from emancipation, a drastic reduction of their financial capital K, through access to credit through inherited social networks — or high social capital and thereby a high z_t. Eeckhout et al. (2023b) shows how elites from the Tsarist period persisted through the Soviet period as educational elites. This is an investment response predicted by the model: In periods of high uncertainty of economic capital, a large $\text{Var}(\varepsilon^K)$ and a small r_{econ}, families will decide to increase I_t and invest more in z_t — cultural and social capital — with its slower decay rate in anticipation of an increase in r_t^z, eventually happening in 1991, priming them for their economic resurgence. Goñi (2022), studying the marriage market for peers in Victorian London, empirically documents one channel through which social capital z_t is translated into economic capital — economically advantageous marriages.

The present paper and the French 19$^{\text{th}}$ century confirm these findings and adds both an old social capital asset which remains highly valuable across the period of observation, as well as a new social capital asset whose economic value was small and short-lived, both priced continuously using evidence from 4.6 million estates.

Two important predictions of the model cannot be tested with the data at hand. I cannot test whether families increase investments I_t into social and cultural capital in times of uncertainty. This would require data on education, marriages etc. which is not contained in the probate records at hand. Moreover, the episode with the biggest economic uncertainty and upheaval, the French Revolution of 1789, predates the start of the wealth data series. It would be possible to identify families which had suffered economic losses and specifically investigate whether they recover from this negative wealth shock.

6 Conclusions

Using rare surnames methodology, this article presents evidence on wealth mobility in France during the 19th century. Using the group-level methodology of Clark and Cummins (2015) and a newly digitised dataset of wealth at death with 4.6 million death records, the largest such dataset to date, I find wealth persistence parameters of 0.74–0.87 per generation. This is comparable to the results for England, despite France's more turbulent institutional history during the period, and contrasts with the bi-generational estimates of around 0.3 found by J. Bourdieu et al. (2019) using traditional measures.

I confirm these results in a different outcome space, using a newly assembled dataset of registers of 15 elite outcomes. I find that surname-level over-representation decays at $b = 0.942$[18]. This result is robust to fixed-effects and balanced-panel specifications and is indistinguishable from the surname-group-level wealth persistence estimate.

I propose a new interpretation of the latent factor as a broader family-level endowment with different forms of capital, including social and cultural capital. This is inspired by P. Bourdieu (2011) and his idea that *"[c]apital, which in its objectified or embodied forms, takes time to accumulate and which, as a potential capacity to produce profits and to reproduce itself in identical or expanded form, contains a tendency to persist in its being, is a force inscribed in the objectivity of things so that everything is not equally possible or impossible."* (P. Bourdieu, 2011, p. 78)

To investigate this hypothesis, I connect the wealth data to the elite outcome data at the surname-level. Using data on the *ancien régime* nobility from a variety of sources, I find that the group sorted on top 5% wealth alone regresses to the mean at a rate of 0.74, while the group sorted on hereditary *ancien régime* title, a prime example of institutionalised social capital, shows no detectable reversion to the mean throughout the period.[19] I also find that these families are far more overrepresented among politicians than the surnames of the wealthiest commoner families, but that representation in the elite registers does not predict a family's future wealth once its current wealth is known. In this interpretation, the latent factor is slow, but its economic returns are institutionally contingent.

[18] $b = 0.95$ for shorter generations.

[19] This lends support to the *revisionist* camp in the historiography of the French Revolution (Beck, 1981; Cobban, 1964; Forster, 1967).

References

Académie française. (2026). *Les immortels: Les quarante aujourd'hui.* Retrieved from https://www.academie-francaise.fr/les-immortels/les-quarante-aujourdhui (Accessed: 2026-03-25)

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

Adermon, A., Lindahl, M., & Waldenström, D. (2018). Intergenerational wealth mobility and the role of inheritance: Evidence from multiple generations. *The Economic Journal*, *128*(612), F482–F513.

Ager, P., Boustan, L., & Eriksson, K. (2021). The intergenerational effects of a large wealth shock: white southerners after the civil war. *American Economic Review*, *111*(11), 3767–3794.

Alesina, A., Seror, M., Yang, D. Y., You, Y., Zeng, W., et al. (2020). *Persistence through revolutions.* National Bureau of Economic Research.

Angrist, J. D., & Krueger, A. B. (1995). Split-sample instrumental variables estimates of the return to schooling. *Journal of Business & Economic Statistics*, *13*(2), 225–235.

Authier, M., Galbrun, A., & Dell'Acqua, J. (1973). *état de la noblesse française subsistante.* (43 vols., 1973–2018)

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

Beck, T. (1981). The french revolution and the nobility: a reconsideration. *Journal of Social History*, *15*(2), 219–233.

Becker, G. S., & Tomes, N. (1979). An equilibrium theory of the distribution of income and intergenerational mobility. *Journal of political Economy*, *87*(6), 1153–1189.

Black, S. E., & Devereux, P. J. (2011). Recent developments in intergenerational mobility. *Handbook of labor economics*, *4*, 1487–1541.

Blanchet, T., Fournier, J., & Piketty, T. (2022). Generalized pareto curves: theory and applications. *Review of Income and Wealth*, *68*(1), 263–288.

Bleakley, H., & Ferrie, J. (2016). Shocking behavior: Random wealth in antebellum georgia and human capital across generations. *The Quarterly Journal of Economics*, *131*(3), 1445–1495.

Boillet, M., Kermorvant, C., & Paquet, T. (2021, January). Multiple Document Datasets Pre-training Improves Text Line Detection With Deep Neural Networks. In *2020 25th international conference on pattern recognition (icpr)* (p. 2134-2141). doi: 10.1109/ICPR48806.2021.9412447

Bourdieu, J., Kesztenbaum, L., & Postel-Vinay, G. (2014). *L'enquête tra, histoire d'un outil, outil*

pour l'histoire: Tome i. 1793-1902. INED.

Bourdieu, J., Kesztenbaum, L., Postel-Vinay, G., & Suwa-Eisenmann, A. (2019). Intergenerational wealth mobility in france, 19th and 20th century. *Review of Income and Wealth*, *65*(1), 21–47.

Bourdieu, P. (1984). *Distinction*. Routledge.

Bourdieu, P. (2011). The Forms of Capital. In M. Granovetter & R. Swedberg (Eds.), *The Sociology of Economic Life* (3rd ed., pp. 78–92). Routledge.

Braun, S. T., & Stuhler, J. (2018). The transmission of inequality across multiple generations: testing recent theories with evidence from germany. *The Economic Journal*, *128*(609), 576–611.

Campardon, É. (1889). *Liste des membres de la noblesse impériale: dressée d'après les registres de lettres patentes conservés aux archives nationales*. Au siège de la Société.

Chaix d'Est Ange, G. (1903). *Dictionnaire des familles françaises anciennes ou notables à la fin du xix^{e} siècle*. Évreux: C. Hérissey. (20 vols., 1903–1929, BNF 34209118)

Chetty, R., Friedman, J. N., Saez, E., Turner, N., & Yagan, D. (2020). Income segregation and intergenerational mobility across colleges in the united states. *The Quarterly Journal of Economics*, *135*(3), 1567–1633.

Chetty, R., & Hendren, N. (2018a). The impacts of neighborhoods on intergenerational mobility i: Childhood exposure effects. *The Quarterly Journal of Economics*, *133*(3), 1107–1162.

Chetty, R., & Hendren, N. (2018b). The impacts of neighborhoods on intergenerational mobility ii: County-level estimates. *The Quarterly Journal of Economics*, *133*(3), 1163–1228.

Chetty, R., Hendren, N., Kline, P., & Saez, E. (2014). Where is the land of opportunity? the geography of intergenerational mobility in the united states. *The quarterly journal of economics*, *129*(4), 1553–1623.

Chetty, R., Hendren, N., Kline, P., Saez, E., & Turner, N. (2014). Is the united states still a land of opportunity? recent trends in intergenerational mobility. *American economic review*, *104*(5), 141–147.

Chetty, R., Jackson, M. O., Kuchler, T., Stroebel, J., Hendren, N., Fluegge, R. B., ... others (2022a). Social capital ii: determinants of economic connectedness. *Nature*, *608*(7921), 122–134.

Chetty, R., Jackson, M. O., Kuchler, T., Stroebel, J., Hendren, N., Fluegge, R. B., ... others (2022b). Social capital i: measurement and associations with economic mobility. *Nature*, *608*(7921), 108–121.

Clark, G. (2015). *The son also rises: Surnames and the history of social mobility*. Princeton University Press.

Clark, G., & Cummins, N. (2014). Surnames and social mobility in england, 1170–2012. *Human Nature*, *25*, 517–537.

Clark, G., & Cummins, N. (2015). Intergenerational wealth mobility in england, 1858–2012: surnames and social mobility. *The Economic Journal*, *125*(582), 61–85.

Clark, G., Cummins, N., Hao, Y., & Vidal, D. D. (2015). Surnames: A new source for the history of social mobility. *Explorations in Economic History*, *55*, 3–24.

Cobban, A. (1964). *The social interpretation of the french revolution.* Cambridge university press.

Collado, M. D., Ortuño-Ortín, I., & Stuhler, J. (2023, 05). Estimating intergenerational and assortative processes in extended family data. *The Review of Economic Studies*, *90*(3), 1195-1227. Retrieved from https://doi.org/10.1093/restud/rdac060 doi: 10.1093/restud/rdac060

Collège de France. (2007). *Liste des professeurs depuis la fondation du Collège de France en 1530.* Paris. Retrieved from https://www.college-de-france.fr/media/en-college/UPL50600_LISTE_DES_PROFESSEURS.pdf (Conception et réalisation: Danièle Quérehen. Affaires culturelles et Relations extérieures)

Comninel, G. C. (1991). *Rethinking the french revolution: Marxism and the revisionist challenge.* Verso.

Cummins, N. (2024). Ethnic wealth inequality in england and wales, 1858–2018. *Explorations in Economic History*, *94*, 101617.

Cummins, N. J., & Gráda, C. Ó. (2025). The irish in england. *The Journal of Economic History*, *85*(1), 180–214.

Darnton, R. (1982). *The literary underground of the old regime.* Harvard University Press.

Daumard, A. (1973). *Les fortunes françaises au xixe siècle: enquête sur la répartition et la composition des capitaux privés à paris, lyon, lille, bordeaux et toulouse d'après l'enregistrement des déclarations de succession* (Vol. 27). Walter de Gruyter GmbH & Co KG.

de Barrau, H. (1853). *Documents historiques et généalogiques sur les familles et les hommes remarquables du rouergue.* Rodez. (4 vols., 1853–1860)

de Foras, A., & de Mareschal de Luciane, F.-C. (1863). *Armorial et nobiliaire de l'ancien duché de savoie.* Grenoble: Allier Frères. (5 vols., 1863–1910, BNF 30881392)

Del Pizzo, A., Nybom, M., & Stuhler, J. (2026). Indirect estimators of intergenerational mobility. *IZA Discussion Paper*, *18641*.

de Rivoire de La Bâtie, G. (1867). *L'armorial de dauphiné.* Lyon: L. Perrin.

de Saint-Houardon, J. (2007). *Noblesse de bretagne: histoire et catalogue de la noblesse bretonne subsistante.* Versailles: Mémodoc.

de Séréville, E., & de Saint-Simon, F. (1975). *Dictionnaire de la noblesse française* (Vol. 1). (BNF 34584509)

de Séréville, E., & de Saint-Simon, F. (1977). *Dictionnaire de la noblesse française* (Vol. 2). (BNF

34605975)

de Vaulchier, J., de Saulieu, J.-A., & de Bodinat, J. (2004). *Armorial de l'association d'entraide de la noblesse française*. Éditions du Gui.

de Woelmont de Brumagne, H. (1928). *La noblesse française subsistante*. Paris: Champion.

Dupraz, Y., & Simson, R. (2024). Elite persistence in sierra leone: What can names tell us? *Journal of Development Economics*, *171*, 103333.

du Puy de Clinchamps, P. (2005). *Dictionnaire et armorial de la noblesse*. Patrice du-Puy-Éditeur. (5 vols., 2005–2009)

Eeckhout, T. (2023). *Measuring corruption, intergenerational mobility and human capital using trace data: evidence from russia* (Doctoral Dissertation). Ghent University. Faculty of Economics and Business Administration, Ghent, Belgium.

Eeckhout, T., Markevich, A., & Schoors, K. (2023a). *Multiple-outcome data on surnames for intergenerational mobility estimation* (Ph.D. dissertation). Ghent University, Ghent.

Eeckhout, T., Markevich, A., & Schoors, K. (2023b). *Social mobility in times of revolutions and regime changes: Persistence of elites in russia from tsarism to putinism* (Ph.D. dissertation). Ghent University, Ghent.

Farcy, J.-C. (2010). *Annuaire rétrospectif de la magistrature xixe-xxe siècles.* LIR3S (Université Bourgogne Europe). Retrieved from https://annuaire-magistrature.fr/ (Accessed: 2026-08-25)

Finley, T., Franck, R., & Johnson, N. D. (2021). The effects of land redistribution: Evidence from the french revolution. *The Journal of Law and Economics*, *64*(2), 233–267.

Forster, R. (1967). The survival of the nobility during the french revolution. *Past & Present*(37), 71–86.

Franck, R., & Michalopoulos, S. (2017). *Emigration during the french revolution: consequences in the short and longue durée* (Tech. Rep.). National Bureau of Economic Research.

Frotier de La Messelière, H. (1912). *Filiations bretonnes 1650–1912*. Saint-Brieuc: Imprimerie Prudhomme. (5 vols., 1912–1926, BNF 30469358)

Galli, S., Theodoridis, D., & Rönnbäck, K. (2024). Elite persistence and inequality in the danish west indies, 1760-1914. *Explorations in Economic History*, 101636.

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.

Güell, M., Mora, J. V. R. M., & Telmer, C. I. (2015). Intergenerational mobility and the informational content of surnames. *The Review of Economic Studies*, *82*(2), 693–735.

Güell, M., Rodríguez Mora, J. V., & Solon, G. (2018). New directions in measuring intergenerational

mobility: Introduction. *The Economic Journal*, *128*(612), F335–F339.

Güell, M., Pellizzari, M., & Rodríguez Mora, J. V. (2018). Correlating social mobility and economic outcomes. *The Economic Journal*, *128*(612), 353–403.

Higgs, D. (1987). *Nobles in nineteenth-century france: The practice of inegalitarianism*. Johns Hopkins University Press.

Jaurès, J. (1924). *Histoire socialiste de la révolution française* (Vol. 7). Éditions de la Librairie de l'humanité.

Jougla de Morenas, H., & de Warren, R. (1934). *Grand armorial de france: catalogue général des armoiries des familles nobles de france*. Société du Grand armorial de France. (7 vols., 1934–1952, BNF 34209165)

La Chesnaye des Bois, F., & De Badier, A. (1783). *Recueil généalogies, pour servir de suite ou dictionnaire de la noblesse*. Paris: Lamy Badiez.

Lefebvre, G. (1939). *The coming of the french revolution* (R. R. Palmer, Trans.). Princeton University Press.

Lersch, P. M., Longmuir, M., & Schnitzlein, D. D. (2024). Intergenerational persistence of wealth. In *Research handbook on intergenerational inequality* (pp. 86–99). Edward Elgar Publishing.

Li, M., Lv, T., Chen, J., Cui, L., Lu, Y., Florencio, D., ... Wei, F. (2023). Trocr: Transformer-based optical character recognition with pre-trained models. In *Proceedings of the aaai conference on artificial intelligence* (Vol. 37, pp. 13094–13102).

Lindahl, M., Palme, M., Massih, S. S., & Sjögren, A. (2015). Long-term intergenerational persistence of human capital: An empirical analysis of four generations. *The Journal of Human Resources*, *50*(1), 1–33. Retrieved 2024-04-30, from http://www.jstor.org/stable/24735407

Long, J., & Ferrie, J. (2018). Grandfathers matter (ed): occupational mobility across three generations in the us and britain, 1850–1911. *The Economic Journal*, *128*(612), F422–F445.

Mathiez, A. (1922). *La révolution française*. Colin, Paris.

Noble, A., & Sutter, N. (tba). From records to riches: An automated pipeline for digitising french tax records, 1790-1870.
(Forthcoming)

Petitpierre, R., Kramer, M., & Rappo, L. (2023). An end-to-end pipeline for historical censuses processing. *International Journal on Document Analysis and Recognition (IJDAR)*, *26*(4), 419–432.

Piketty, T. (2001). *Les hauts revenus en france au xxe siècle*. Hachette.

Piketty, T. (2011). On the long-run evolution of inheritance: France 1820-2050. *The Quarterly Journal of Economics*, *126*(3), 1071—1131.

Piketty, T. (2014). *Capital in the twenty-first century*. Harvard University Press.

Piketty, T., Postel-Vinay, G., & Rosenthal, J.-L. (2006). Wealth concentration in a developing economy: Paris and france, 1807–1994. *American economic review*, *96*(1), 236–256.

Piketty, T., Postel-Vinay, G., & Rosenthal, J.-L. (2014). Inherited vs self-made wealth: Theory & evidence from a rentier society (paris 1872–1927). *Explorations in economic history*, *51*, 21–40.

Postel-Vinay, G., & Rosenthal, J.-L. (2026). *A capital's capital: Wealth and inequality in paris 1807-2023*. Princeton University Press.

Quérard, J.-M. (1827–1839). *La france littéraire, ou dictionnaire bibliographique*. Paris: Firmin Didot, père et fils, libraires. Retrieved from https://fr.wikisource.org/wiki/La_France_litt%C3%A9raire,_ou_Dictionnaire_bibliographique (10 volumes published between 1827 and 1839. Source: Bibliothèque nationale de France.)

Révérend, A. (1895). *Armorial du premier empire: Titres, majorats et armoires concédé par napoléon i^er^*. Paris: Bureau de "l'Annuaire de la Noblesse".

Révérend, A. (1901). *Titres, anoblissements et pairies de la restauration 1814–1830*. Paris: Albert Révérend Honoré Champion.

Soboul, A. (1953). Classes and class struggles during the french revolution. *Science & Society*, 238–257.

Soboul, A. (1988). *Understanding the french revolution*. International Publishers, New York.

Solon, G. (1999). Intergenerational mobility in the labor market. In *Handbook of labor economics* (Vol. 3, pp. 1761–1800). Elsevier.

Souibgui, M. A., & Kessentini, Y. (2020). De-gan: A conditional generative adversarial network for document enhancement. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *44*(3), 1180–1191.

Stuhler, J. (2024). Multigenerational inequality. *Research Handbook on Intergenerational Inequality*, 100–121.

Tarride, S., Maarand, M., Boillet, M., McGrath, J., Capel, E., Vézina, H., & Kermorvant, C. (2023). Large-scale genealogical information extraction from handwritten quebec parish records. *International Journal on Document Analysis and Recognition (IJDAR)*, 1–18.

Valette, R. (1959). *Catalogue de la noblesse française*. (Six editions, 1959–2007, BNF 31521922)

Valette, R. (1989). *Catalogue de la noblesse française*. Paris: Robert Laffont. (BNF 36638580)

Valette, R. (2002). *Catalogue de la noblesse française au xxi^e^ siècle*. Paris: Éditions Robert Laffont.

Valette, R. (2007). *Catalogue de la noblesse française*. Paris: Éditions Robert Laffont.

Van Leeuwen, M. H., Maas, I., Rébaudo, D., & Pélissier, J.-P. (2016). Social mobility in france 1720–1986: Effects of wars, revolution and economic change. *Journal of Social History*, *49*(3),

585–616.

Vosters, K., & Nybom, M. (2017). Intergenerational persistence in latent socioeconomic status: Evidence from sweden and the united states. *Journal of Labor Economics*, *35*(3), 869–901.

Appendices

A Data Appendix

The digitisation of the wealth data has been collaborative work with Aurelius Noble (Noble & Sutter, tba).

A.1 Wealth Data Coverage

To capture how much of French top wealth is captured by the newly digitised *Tables des Successions et Absences* wealth data, I consult the *Bulletin de Statistique et Législation Comparée* (BSLC). This source records the number of estates and the total bequeathed wealth by *département* and wealth bracket. It is the source that Piketty et al. (2006) and Piketty (2001, 2014) use to construct aggregate top-wealth-share series for France in the 20th — combined with the data compiled by Postel-Vinay and Rosenthal (2026) for Paris for the 19th century. I have digitised the BSLC for the years 1902, 1903 and 1910. The BSLC was published starting in the late 19th century. 1902 is the first year, however, that reported the number of estates and the total bequeathed wealth by *département* and wealth bracket. I apply generalised Pareto interpolation (Blanchet, Fournier, & Piketty, 2022) to the BSLC data to determine how much of the country's top wealth is covered by the 16 *départements* for which this paper has data. The 16 *départements*[20], as they contain Paris and other cities with a lot of wealth concentration like Marseille and Lille, capture 73% of the top 0.1% of the wealth distribution, 62% of the top 1%, and 55% of the top 5% of the wealth distribution in 1902.[21] No comparable data for the entirety of France exists for the 19th century. Looking at the data in Daumard (1973) and again employing generalised Pareto interpolation (Blanchet et al., 2022), shows, however, that the wealth concentration in the early 19th century was lower than during the *Belle Époque*. Between the 1820s and 1911, the top 1% wealth share rises from 31% to 35% in Paris (1820-1911), from 17% to 46% in Bordeaux (1824-1911), and from 22% to 33% in Toulouse (1826-1911).

[20] In the BSLC, I rely on the pre-World War I *départements*. The old *département de la Seine* encompasses today's Paris, Hauts-de-Seine, Seine-Saint-Denis and Val-de-Marne. The old Seine-et-Oise encompasses today's Yvelines, Essonne, Val-d'Oise and parts of the other three above. Our TSA data covers Paris, Val-de-Marne and Seine-Saint-Denis but not Hauts-de-Seine, and Yvelines and Essonne but not Val-d'Oise.

[21] In 1910, the numbers are 72% of the top 0.1%, 63% of the top 1%, 55% of the top 5%, and 52% of the top 10%.

Figure 7: Wealth Concentration in three Cities according to Daumard (1973)

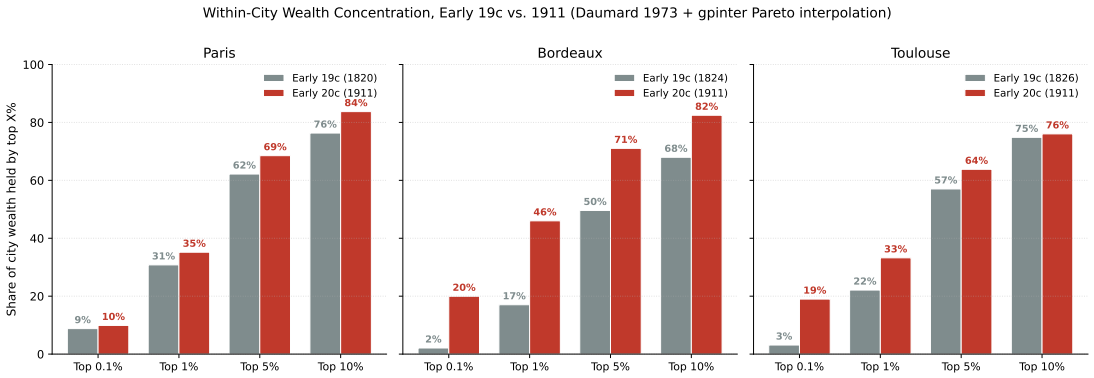


Figure 8: Data Coverage by Decade

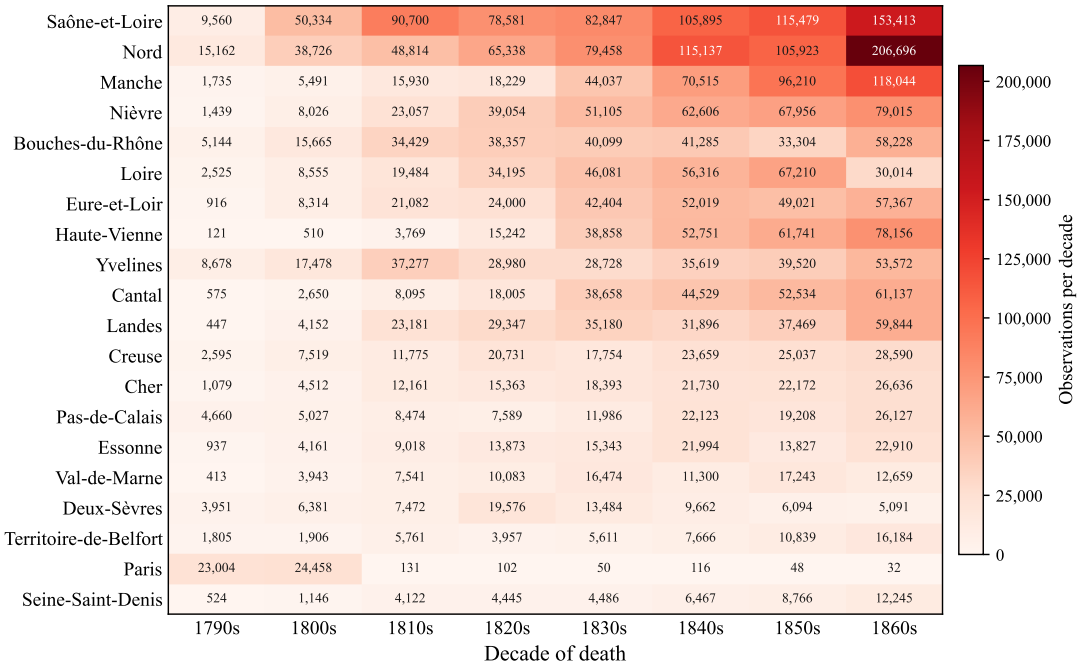


Figure 8 reports the number of death records per *département* and decade in the current release. Coverage grows over time, as fewer early registers have survived, and the 1790s are thinly populated for most *départements*. Paris is covered only until 1806; the later Paris registers are being processed.

A.2 Comparison with Existing Wealth Datasets

Table 9 compares the dataset with existing datasets. Its main advantage over existing datasets is its size and that it includes both the top of the wealth distribution and rural areas. Its main shortcoming

is that it is based on the *Tables des Successions et Absences*, which record real estate wealth at the location of the real estate instead of the location of death, and only record revenue of real estate instead of its total value. Therefore does not always record individuals' full real estate holdings, it gives a good estimate of individuals' mobile wealth, however.

Table 9: Comparison of Our Dataset with Existing Datasets

Noble and Sutter (tba)	**Enquête TRA**	**Postel-Vinay and Rosenthal (2026)**
Sources		
Tables des Successions et Absences	Combination of all available sources, notably the Tables des Successions et Absences and the Régistres des Mutations par Décès for wealth	Same as Enquête TRA
Coverage		
All individuals dying in one of 20 départements	All individuals with a surname starting with *Tra-*	Paris in years ending in -2 or -7
Number of Observations		
4,561,426	56,110	27,928
Inclusion of Elites		
Captures both urban and rural elites	Top of the wealth distribution not captured; includes only 8 noble families	Includes urban wealth elites but not rural elites

A.3 Wealth Data Digitisation Pipeline

Recently, important advances were made in the automated transcription of French-language handwritten documents, most notably Petitpierre, Kramer, and Rappo (2023), Tarride et al. (2023), and the Socface project. The pipeline utilized to digitise the data for this paper employs state-of-the-art open-source software at every step. It will be introduced in detail in Noble and Sutter (tba).

The first step is the mass downloading of the scanned *Tables des Successions et Absences* from the website of the departmental archives where these records are stored. The files are downloaded and stored as jpg files. To do this, I use the webcrawler chromedriver, which accesses the website using a chrome browser. In a next step, these images are rescaled and binarized using DE-GAN (Souibgui & Kessentini, 2020). Then Doc-UFCN (Boillet, Kermorvant, & Paquet, 2021), fully convolutional network, is used to detect individual text lines, as well as to segment the page into rows and columns. We used a fine-tuned version of Doc-UFCN that was trained on 100 manually annotated *TSA* pages. The optical character recognition of the handwritten text lines is performed using TrOCR (Li et al., 2023), a pre-trained transformer-based optical character recognition model. We again used a fine-tuned version of the model that had been trained on 41,621 text lines. This transformer-based approach contrasts with traditional OCR/HTR algorithms that are based on convolutional neural networks (CNN).

A.4 Wealth Data Evaluation

In order to be able to evaluate the end-to-end performance of our pipeline, we have fully transcribed 20 randomly selected pages from the *TSA* corpus of the *Département de la Nièvre*. These serve as the ground truth against which we can evaluate the output of our end-to-end pipeline.

Tables 10, 11, and 12 give a detailed overview of the accuracy and precision of the methods employed for text-line detection, column assignment, and row assignment.

The results from the handwritten text recognition step are very promising. Tables 13 and 14 contain the mean and median word error rates (WER) and character error rates (CER) of the end-to-end pipeline for the different columns/variables in our dataset.

From the validation of the training on the 200 transcribed pages, we know the generic CER of our model. This generic CER is 1.4%. This is lower than the generic character error rate of 3.44% achieved by Petitpierre et al. (2023). We have identified 37 different variables in the tables of succession[22]. Of these 37, only a limited amount are of paramount importance to us — mainly names and wealth

[22] Article Number, Surname, First Name, Profession, Residence, Date of Death, Names, occupations and residences of each of the heirs, donees, legatees and others, Details and situation of the objects transmitted by succession or by the event of dispositions, Estimation of the amount of wealth, Dates for declarations and payment of proportional registration duties, Value of furniture, money, annuities and debts, Income from real estate, Location of real estate, Date of apposition

Table 10: Text-line Metrics

Metric	Value
Pixel Level	
Intersection over Union (IoU)	0.78
Precision	0.92
Recall	0.84
F-score	0.88
Object Level	
Average Precision @[.5]	0.93
Average Precision @[.75]	0.75
Average Precision @[.95]	0.01
Average Precision @[.5,.95]	0.62

Table 11: Column Assignment Metrics

	Accuracy (%)	
Metric	**Hough Transform**	**Doc-UFCN**
Mean	95.2	81.7
Median	100.0	77.3
Min	81.0	63.6
Max	100.0	100.0
Std. Dev.	6.3	14.9

Note: This is the percentage of columns where all cells are correctly assigned.

Table 12: Row Assignment Metrics

	Accuracy (%)
Metric	**Hough Transform & Doc-UFCN**
Mean	85.6
Median	83.7
Min	61.9
Max	100.0
Std. Dev.	11.7

Note: This is the percentage of rows where all text-lines that are in that row are assigned to it, and no incorrect text-lines.

estimates, but also age at death, civil status, and domicile. Others — for example the date of the unsealing of the will — are of limited importance. For this reason CER and WER disentangled by column contains more useful information.

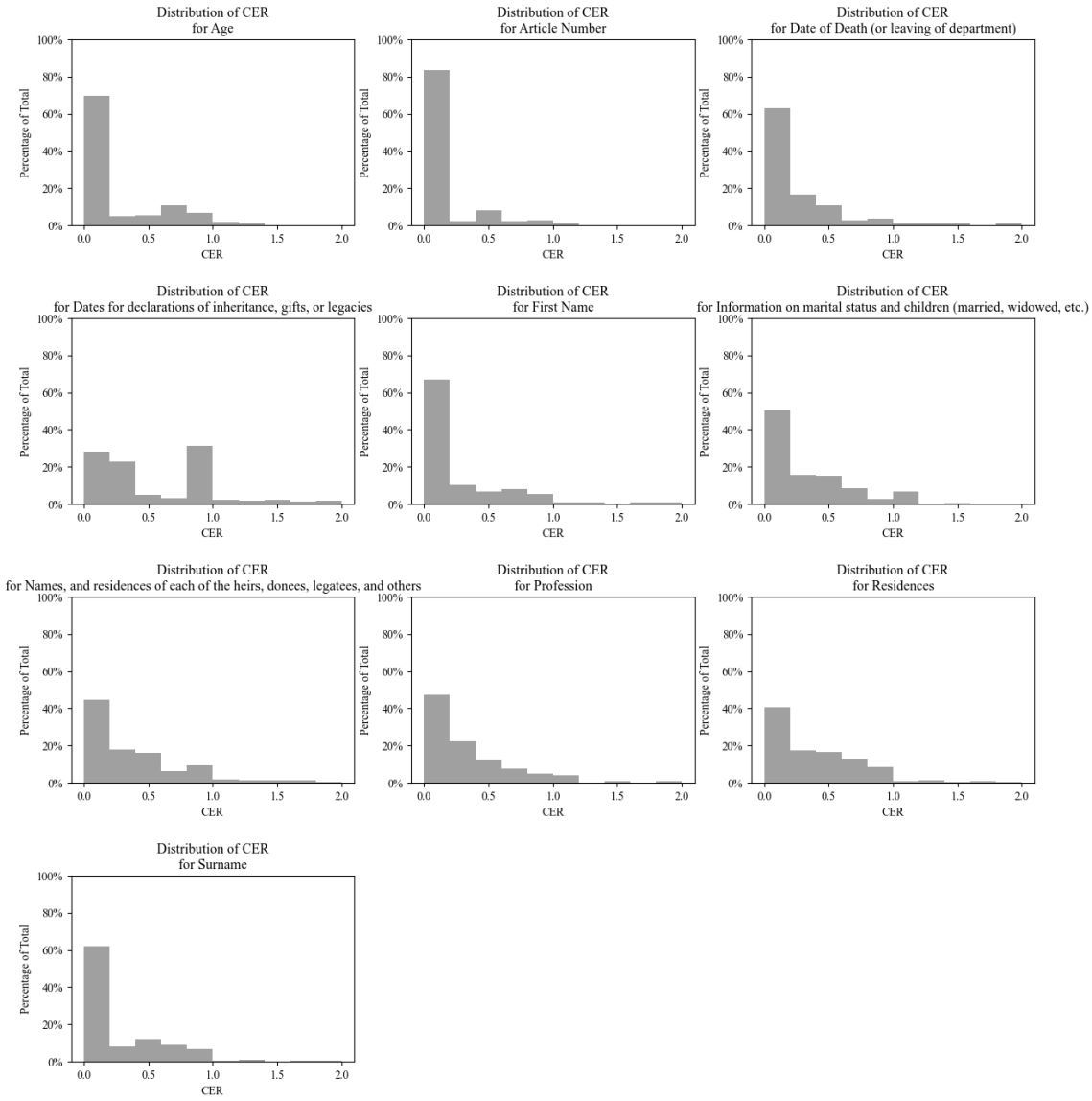
Table 13 lists the mean and median character error rate across all 20 pages for all variables for which we have more than ten observations. The results are encouraging. The median character error rate for surnames, the key variable, is 0%. For other variables like profession or the different wealth measures, it is higher and ranges from 0% to 50%. Mean CER and WER are naturally higher. Figure 9 shows the distribution of character error rates for each variable for which we have more than 100 observations. We can see that the vast majority of pages, notwithstanding errors in page segmentation, have very low character error rates. The distribution for almost all variables has a mode in the lowest bin (0%-20% CER).

of seals and date of unsealing, Acceptances subject to inventory, Renunciation of inheritances, legacies or community property, Date and place of registration, Date of notice to heirs, legatees, etc. Referrals, and Observations.

Table 13: Word Error Rate (WER) and Character Error Rate (CER) for all Variables with 10 or more Observations

Variable	**Mean**		**Median**		
	WER	CER	WER	CER	Observations
Age	0.437	0.215	0.000	0.000	260
Article Number	0.211	0.103	0.000	0.000	213
Date of Death	0.540	0.292	0.333	0.100	278
Date of declaration	1.551	0.553	1.000	0.382	168
Income from real estate	1.087	0.649	1.000	0.714	75
Situation of real estate	1.375	0.449	1.000	0.333	54
Value of furniture, money, annuities and debts	0.978	0.618	1.000	0.477	100
First Name	0.563	0.230	0.000	0.000	296
Guardianships and trusteeships	0.574	0.392	0.333	0.167	17
Marital status	0.874	0.387	1.000	0.194	154
Inventory: Date of Registration	0.577	0.388	0.542	0.167	28
Inventory: Valuation amount	1.130	0.283	1.000	0.200	23
Location of Death	0.821	0.434	1.000	0.348	13
Heirs, donees, legatees	0.869	2.106	0.750	0.333	196
Number of the dubious base	1.313	0.673	1.000	0.667	67
Observations	0.986	0.567	0.800	0.431	64
Profession	0.871	0.312	1.000	0.200	211
Residence	1.134	0.490	1.000	0.296	295
Surname	0.624	0.246	0.000	0.000	306

Figure 9: Distribution of Character Error Rates (CER) by Column (only for columns for which we have more than 100 data points in our evaluation sample)



For this reason, we have calculated the CER and WER exclusively for pages on which the page segmentation worked without error. These mean and median character error rates and word error rates are displayed in table 14.

Table 14: Word Error Rate (WER) and Character Error Rate (CER) for all Pages with Correctly Recognized Rows and Columns

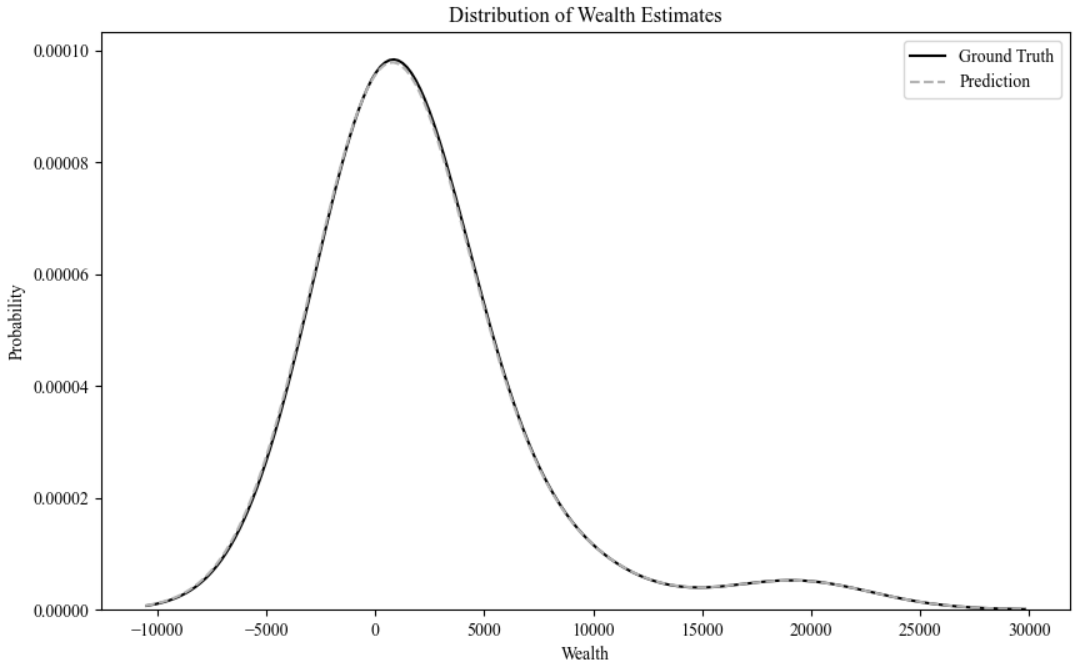
Variable	**Mean**		**Median**		
	WER	CER	WER	CER	Observations
Age	0.073	0.048	0.000	0.000	48
Article Number	0.167	0.067	0.000	0.000	42
Date of Death	0.513	0.165	0.333	0.091	62
Date of declaration	0.918	0.230	0.500	0.200	40
Income from real estate	0.727	0.652	1.000	0.667	11
Value of furniture, money, annuities and debts	0.864	0.658	1.000	0.667	22
First Name	0.342	0.101	0.000	0.000	60
Marital status	0.635	0.163	0.750	0.135	32
Heirs, donees, legatees	0.656	0.574	0.750	0.150	49
Number of the dubious base	1.000	0.792	1.000	0.667	12
Observations	0.774	0.493	0.800	0.414	14
Profession	0.569	0.196	0.000	0.000	51
Residence	0.814	0.296	0.800	0.200	65
Surname	0.285	0.113	0.000	0.000	65

Estimates of wealth and, to the extent available, income, are crucial for studies of social mobility. They are, however, prone to high character error rates. This is for two reasons. First, the transformer-based text recognition algorithm performs very well with natural language but not necessarily with numbers. While not every possible sequence of alphabetic characters is a word in the English or French language, every possible sequence of digits is an existing number. Secondly, because the strings in the wealth column tend to be short, they are prone to having high character and word error rates.

What matters is whether the data reflects the true distribution of wealth given in the ground truth data. For this reason, we have cleaned the output of the text recognition step for one of the wealth measures — the inventory valuation amount, for which most data was available in the 20 pages —

and have plotted it against the ground truth wealth estimate data from the *TSA*. The results can be seen in figure 10. As we can see, the predicted values — the dashed grey line tracks the black ground truth almost perfectly. This indicates that we *do indeed* capture the true distribution of wealth.

Figure 10: Distribution of Wealth Estimates in *Francs* — Ground Truth vs. Prediction



To illustrate the meaning of different error rates better, table 15 shows the transcription corresponding to the mean character error rate for every variable. As can be seen, in many cases the CER is driven by case sensitivity — the meaning of the words is not changed due to the mistranscribed characters. The meaning of the word, however, has not been affected by the transcription.

Table 15: Transcription Result with the Median CER for each Variable

Variable	**Ground Truth**	**Prediction**
Age	4 mois	4 mois
Article Number	48	48
Date of Death	17 aout 1866	17 aout 1860
Seals Apposition	Michot	Michot
Declarations Date	17 Xbre	17 10bre
Income from Buildings	143	143
Building Situation	Asnois	asnois.
Furniture Value	361/0	361.5
Wealth Estimation	1230	1230
First Name	Gilbert	Gilbert
Furniture Sale Date	5 8er 1849	5 8e. 1849
Furniture Valuation	258.55	258.55
Trusteeships	27 arvil 1850	27 avril 1850
Marital Status	ep. Marie	ep. Marie
Inventory Date	26 fevrier	26 Fevrier
Inventory Value	714.9	714.9
Location of Death	Revenue C	Revenier
Heirs' Names	Devoucout, antoine	Devoncous, metoine
Dubious Base Number	225	225
Observations	Sans Droits acquis	Sans Edriot et acquis
Profession	mineur	Mineur
Residences	Alluy	Allay
Surname	Louvrier	Louvrier

A.5 Elite Datasets

Table 16: All Elite Datasets

Source	Type	Role	Src	No Ind.	Surname groups
Panel A. Ancien régime elites (a priori surname lists)					
Dictionnaire/Annuaires de la noblesse[a]	Nobility	Elite	D	55,132	13,783
of which noblesse d'épée	Nobility	Elite	D	12,956	3,239
of which noblesse de robe	Nobility	Elite	D	5,820	1,455
Ferme Générale	Economic	Elite	D	228	57
Avocats au parlement de Paris, 1770	Judicial	Elite	D	1,972	493
Sénat conservateur[b]	Political	Elite	N	1,081	93
Conseil des Anciens[b]	Political	Elite	N	524	319
Députés, Constituante 1789	Political	Elite	N	207	44
Députés, Législative 1791	Political	Elite	N	78	12
Corps des Mines, entrants ≤1815	Technical	Elite	N	–	31
Mémoire des Hommes, base 61[c]	Military	Elite	N	5,000	–
Panel B. Outcomes (registers of realised elite position)					
Parliament and cabinet	Political	Outcome	D	8,767	4,997
Légion d'honneur[d]	Societal	Outcome	D	350,434	3,069
Book authors (Quérard)[e]	Cultural	Outcome	D	13,566	8,210
Napoleonic ennoblements	Nobility	Outcome	D	2,690	2,148
Census voters (*censitaires*)	Political	Outcome	D	32,739	9,394
Industrialists, 1840	Economic	Outcome	D	1,833	1,221
Colonial governors	Admin.	Outcome	D	626	435
Economic elites (Wikidata)	Economic	Outcome	D	309	260
Collège de France/Académie française	Cultural	Outcome	D	105	101
Magistrature	Judicial	Both	N	59,717	–
CTHS, *La France savante*	Cultural	Both	N	31,617	–
Sénat, four chambers	Political	Both	N	2,790	–
École des chartes	Educ.	Outcome	N	3,081	–
Corps des Mines	Technical	Both	N	2,529	–
SIPROJURIS (law professors)	Educ.	Outcome	N	567	–
Académie des sciences	Cultural	Outcome	N	8,030	–
Académie de médecine	Cultural	Outcome	N	4,301	–
Académie française (Wikidata)	Cultural	Outcome	N	3,529	–
Pairs de France	Political	Outcome	N	1,881	–
Conseillers d'État	Admin.	Outcome	N	1,194	–
Préfets	Admin.	Outcome	N	572	–
INPI patents[f]	Economic	Outcome	N	44,962	–

Notes: *Role* follows Eeckhout (2023): "Elite" denotes an *a priori* surname list defining a group k; "Outcome" a register in which relative representation is measured; "Both" a source whose early cohort defines a group and whose later cohorts serve as an outcome (a group is never measured in the register that defines it). *In window* counts records datable to 1791–1902; *Matched* is the share of those linked to an INSEE surname group. Surnames are matched on the particled (prefix) form where a particle is recorded, and on the first component for hyphenated compounds. [a]Conseil des Anciens, Sénat conservateur, Second Empire, Third Republic. [b]1,316 register entries plus 1,213 biographies; the biographies carry parents' names. [c]Harvest incomplete: parsed and reported but excluded from all estimates (`include_in_analysis` = 0). [d]Base 61, *registres de contrôle et matricules* 1683–1815; detail records carry the father's name, so this source supports direct father–son validation rather than representation counts. [e]Wikidata coverage is notability-weighted and is used only for numerators and for *a priori* group definitions, never as a population frame; the INSEE birth register is the population frame throughout.

B Derivation Theoretical Model

This section contains a derivation of the model employed in section 5. All variables are expressed as deviations from the population mean at time t. This means that intercepts are dropped.

B.1 Model Setup

Each family holds a wider capital endowment consisting of two capital stocks. K_t is traditional economic capital, z_t are social and cultural capital, forms of capital that decay slower and are harder to be expropriated. As stated, z_t evolves as follows:

$$z_t = b_z \, z_{t-1} + \eta \, I_t + \varepsilon_t^z, \tag{9}$$

Cultural and social capital are accumulated through the investment I_t of time and financial resources and decay at a slow rate b_z. Windfalls are not possible.

$$K_t = b_e \, K_{t-1} + r \, z_{t-1} - I_t + \varepsilon_t^K, \tag{10}$$

Wealth at death that we observe is a noisy projection of economic capital with a measurement error:

$$w_t = K_t + u_t. \tag{11}$$

In the derivation of observed wealth persistence in section B.2 investments I_t are set to zero and returns r_t are assumed to be constant for the sake of tractability.

B.2 Derivation of Observed Wealth Persistence

Observed wealth persistence can be written as:

$$\operatorname{plim} \hat{\beta}_w = \frac{Cov(w_{it}, w_{it-1})}{Var(w_{it-1})} \tag{12}$$

Combining equations 10 and 11 wealth at time t can be written as:

$$w_{it} = b_e K_{it-1} + r z_{t-1} - I_t + \varepsilon_t^K + u_t \tag{13}$$

Which allows us to write out the variance and covariance terms:

$$\text{Var}(w_{it-1}) = \text{Var}(b_c K_{it-1} + r z_{t-1} - I_t + \varepsilon_t^K + u_t) = \sigma_K^2 + \sigma_u^2 \tag{14}$$

$$\text{Cov}(w_{it}, w_{it-1}) = \text{Cov}(b_c K_{it-1} + r z_{t-1} - I_t + \varepsilon_t^K + u_t) = b_e \sigma_K^2 + rS \tag{15}$$

Where $S \equiv \text{Cov}(K_t, z_t)$. This is the cross-sectional covariance, across families at time t between economic capital K_t and broader, slow-moving capital z_t. Because of stationarity it has no time subscript t. I assume that the error terms u_t are serially uncorrelated and independent of K_t and that ε_t^K does not enter these expressions, as it has a time subscript.

We can then split observed persistence into an attenuation term ϕ, a feature of all latent factor models (Clark & Cummins, 2015), a *"reserve bonus"* Δ for families with a higher broader capital stock generating an economic return each period t, and the persistence of economic capital in isolation:

$$\text{plim}\, \hat{\beta}_w = \frac{b_e \sigma_K^2 + rS}{\sigma_K^2 + \sigma_u^2} = \underbrace{\frac{\sigma_K^2}{\sigma_K^2 + \sigma_u^2}}_{\phi} (b_e + \underbrace{\frac{rS}{\sigma_K^2}}_{\Delta}) \tag{16}$$

It remains to define σ_K^2 and $S \equiv \text{Cov}(K_t, z_t)$ in terms of the primitives of the model.

I start by deriving the variance of broad capital stock z_t in equation 9, assuming $I_t = 0$, and using the fact that ε_t^z is uncorrelated with z_{t-1}:

$$\sigma_z^2 = b_z^2 \sigma_z^2 + \text{Var}(\varepsilon^z) \quad \Longrightarrow \quad \sigma_z^2 = \frac{\text{Var}(\varepsilon^z)}{1 - b_z^2}, \qquad \text{Cov}(z_t, z_{t-k}) = b_z^k \sigma_z^2. \tag{17}$$

No we can turn directly to S, the covariance between economic capital and cultural/social capital by substituting 10 and 9 into S, and using the stationarity of $\text{Cov}(K_{t-1}, z_{t-1})$:

$$S = b_e b_z S + r\, b_z\, \sigma_z^2 \quad \Longrightarrow \quad S = \frac{r\, b_z\, \sigma_z^2}{1 - b_e b_z} \; > \; 0. \tag{18}$$

What this means is that families who hold a lot of social and cultural capital also persistently hold more economic capital because an economic return from z_t has paid into their economic estates.

In a next step, variance of of economic capital σ_K^2 needs to be defined in terms of primitives of the model. Taking variances in 10, we get:

$$\sigma_K^2 = \frac{r^2 \sigma_z^2 + 2\, b_e\, r\, S + \text{Var}(\varepsilon^K)}{1 - b_e^2}. \tag{19}$$

This means that variance in wealth has two sources: Windfall profits and losses $\text{Var}(\varepsilon^K)$, as well as variation in the rate of return on cultural and social capital r_t.

We now turn ot the first autocovariance of economic capital $\gamma_1 = \text{Cov}(K_t, K_{t-1})$. Substituting 10 into γ_1 and making use of the fact that ε_t^K is not correlated with K_{t-1}, we get:

$$\gamma_1 = \text{Cov}(K_t, K_{t-1}) = b_e \, \sigma_K^2 + r \, S. \tag{20}$$

This shows that economic capital predicts economic capital in the next generation via two different channels: first, the persistence of economic capital itself b_e, and the fact that cultural and social capital will refill economic estates in each generation.

As we only observe measured wealth, which is a noisy reading of underlying economic capital, we now need to step from economic capital to wealth. As the error term u_t is serially uncorrelated and is independent of economic capital K, we can write:

$$\text{Var}(w) = \sigma_K^2 + \sigma_u^2, \qquad \text{Cov}(w_t, w_{t-1}) = \gamma_1. \tag{21}$$

We can now write the broad capital bonus Δ in terms of the now defined primitives of the model:

$$\underbrace{\Delta}_{\text{broad capital bonus}} = \frac{rS}{\sigma_K^2} = \frac{r^2 b_z \sigma_z^2}{(1 - b_e b_z)\sigma_K^2} \tag{22}$$

As established in equation 16, measured wealth persistence is the product of an attenuation factor $\phi = \frac{\sigma_K^2}{\sigma_K^2 + \sigma_u^2}$ and the sum of the faster decay rate b_e and a broad capital bonus Δ that pushes the observed persistence upwards for families with more social and cultural capital. This Δ increases in the persistence rate of social and cultural capital b_z and decreases in the variance of windfalls profits and losses σ_K^2, meaning it is lower in times of economic uncertainty and radical institutional change.

B.3 Reproduction Strategies

In section 5, it was suggested that families with a dynastic motive choose a portfolio of financial, social, and cultural capital that maximizes persistence and intergenerational total capital. This implies a optimisation problem that determines the optimal investment or *reproduction* strategy (P. Bourdieu, 2011). In the previous section, I have abstracted from investments I_t to make the derivation of the broad capital bonus Δ more tractable. Now I_t has to be restored. If we define investment I_t as a proportion of ι of financial capital K_t, equations 9 and 10 can be written as a a first-order vector autoregression:

$$\begin{pmatrix} K_t \\ z_t \end{pmatrix} = \begin{pmatrix} b_e - \iota & r \\ \eta\,\iota & b_z \end{pmatrix} \begin{pmatrix} K_{t-1} \\ z_{t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_t^K \\ \varepsilon_t^z \end{pmatrix}, \tag{23}$$

Reintroducing I_t does not change the basic form of plim $\hat{\beta}_w$. It remains a product of the attenuation factor ϕ. The broad capital bonus Δ, however, now increases in $\eta\iota$, meaning both the conversion factor η, how directly investments are translated into social and cultural capital, and the investment rate ι, while b_e is replaced by $b_e - \iota$, as investments into z_t can be seen as a transfer of economic capital into the future.

The optimal investment/reproduction strategy solves the following optimisation problem:

$$\{I_t\} = \arg \max_{I_t \geq 0,\, I_t \leq Y_t} \mathbb{E}_t\left[\sum_{\tau=0}^{\infty} \beta^\tau K_{t+\tau}\right], \qquad Y_t \equiv K_{t-1} + r\, z_{t-1}, \tag{24}$$

A family at time t tries to maximise the present value of discounted future economic capital $\sum_{\tau=0}^{\infty} \beta^\tau K_{t+\tau}$. The do so through investments I_t into cultural and social capital z_t. These investments, either in the form of time or in the form of resources, have to be financed out of current economic resources K_t. The constraint is a resource constraint and not a budget that needs to be divided up between different capital forms.[23]

B.4 Version of the Model with separate Cultural and Social Capital Terms

C Multiple Outcome Approach

C.1 Multiple outcome Measurement — From Overrepresentation to a Measure of Status

The registers of elite membership described in section 2.2 allow me to observe an outcome of the very high upper tail of the distribution of social and cultural capital z_t directly. Eeckhout et al. (2023a) have developed a methodology to use elite over-representation to measure the latent variable of an surname group.

Elite over-representation is defined as:

$$RR_{k,tj} = \frac{\pi'_{k,tj}}{\pi'_{pop,tj}} \tag{25}$$

[23] P. Bourdieu (2011) is explicit in pointing out that this is not a conscious profit maximisation but happens habitually.

Where $\pi'_{k,tj}$ is the observed relative representation of elite k — for example the *ancien régime* nobility — among elite outcome j — for example parliament — and $\pi'_{pop,tj}$ is the relative representation of the general population among elite outcome j.

Eeckhout et al. (2023a) assume that cultural and social capital z_t is normally distributed with equal variance in every group, and that the distribution of elite group k is shifted by its mean $s_{k,t}$. Any elite outcome j is attained by individuals with relevant cultural or social capital above a certain threshold. Under these assumptions, I can estimate how far the distribution of z_t sits above the population's (normalised at 0) in terms of standard deviations:

$$\tilde{z}_{k,tj} := \Phi^{-1}(1 - \pi_{pop,tj}) - \Phi^{-1}(1 - \pi_{k,tj}) = \beta_{tj} \cdot z_{k,t} \tag{26}$$

Where Φ^{-1} is the cumulative normal distribution of z_t. A measure of $z_{k,tj}$ scaled by β_{tj}, the sensitivity of outcome j to cultural and/or social capital can thus be derived from the over-representation measure $\pi'_{k,tj}$. By averaging over n different outcomes, outcome-specific sensitivity/noise is replaced by its period mean $\bar{\beta}_t$:

$$\frac{1}{n_t} \sum_j \tilde{z}_{k,tj} = \overline{\beta}_t + \epsilon_{k,t} \tag{27}$$

Under the assumption that $\overline{\beta}_t$ is constant over time, the persistence of z_t can now be calculated. Specifically, if I am looking at generations of a 30-year length,

$$\tilde{z}_{k,t} = \frac{1}{n_t} \sum_j \tilde{z}_{k,tj} = \overline{\beta}_t + \epsilon_{k,t} \tag{28}$$

If z_t decays by a factor γ per generation, after n generations:

$$z_{k,t} = \gamma^n z_{k,t_1}, \qquad n = \frac{t - t_1}{30} \tag{29}$$

This becomes linear when taking logs:

$$\log(z_{k,t}) = \log(z_{k,t_1}) + n \cdot \log(\gamma) \tag{30}$$

Substitute in equation 28:

$$\log(\tilde{z}_{k,t}) = \underbrace{\log(\overline{\beta})}_{=\alpha} + \underbrace{\log(\gamma)}_{=\kappa} \cdot \frac{t - t_1}{30} + \epsilon_{k,t} \tag{31}$$

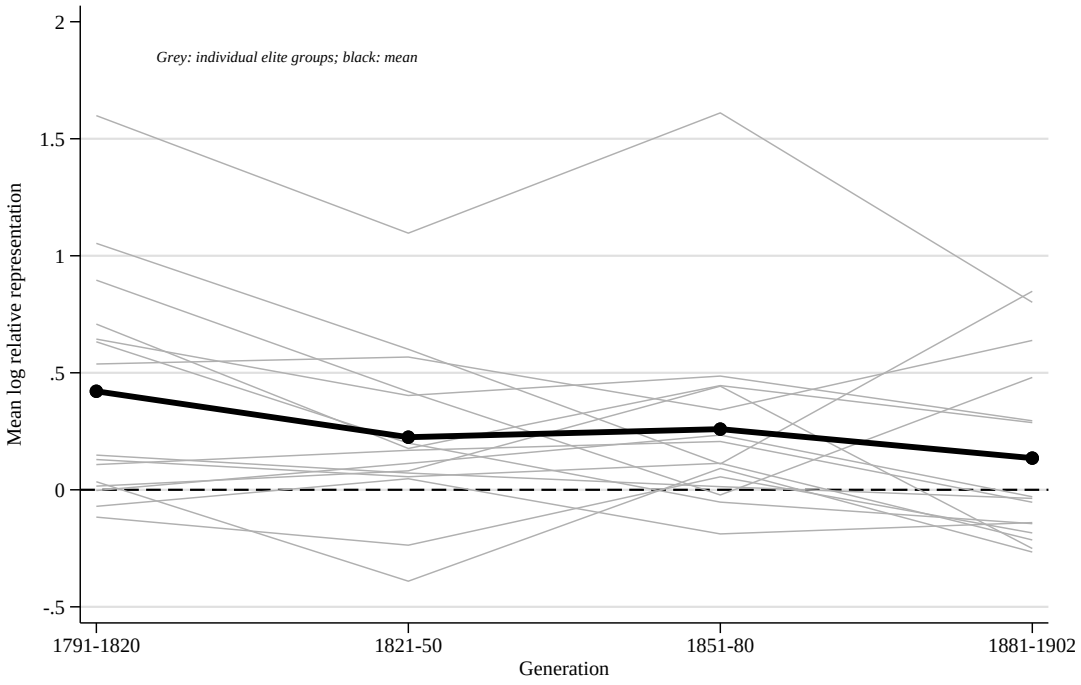
I can therefore run the following regression to uncover γ:

$$\log(\tilde{z}_{k,t}) = \alpha + \kappa \cdot \frac{t - t_1}{30} + \epsilon_{k,t} \tag{32}$$

As $\gamma = e^{\kappa}$.

C.2 Multiple Outcome Approach — Alternative Results

Figure 11: Trajectory Elite Over-Representation — All *Ancien Régime* Elite Groups



D Broad Capital Stock and Multiplicity

Recently, Del Pizzo et al. (2026) proposed that indirect estimators of social mobility, such as IV-based estimators or surname-based measures, are structurally different from direct estimators of social mobility, because intergenerational transmission operates through multiple pathways.

Del Pizzo et al. assume that a socioeconomic outcome y_{it} such as income or wealth is transmitted through multiple pathways X_{it}^j, for example skills, regional advantages, race, health, etc.:

$$y_{it} = \beta_0 + \sum_{j=0}^{J} \rho_j X_{it}^j + u_{it} \tag{33}$$

Each of these pathways has a specific (economic) return ρ_j and a specific heterogeneous rate of persistence λ_j:

$$X_{it}^j = \lambda_j X_{it-1}^j + v_{it}^j \tag{34}$$

Direct estimators of intergenerational persistence are weighted averages of all of these persistence rates λ_j:

$$\beta_{direct} = \frac{\text{Cov}(y_{it}, y_{it-1})}{\text{Var}(y_{it-1})} = \frac{\sum_{j=0}^{J} \lambda_j \omega_j}{\sum_{j=0}^{J} \omega_j + \omega_u} \tag{35}$$

The weights ω_j depend on the variance and returns ρ_j of each characteristic.

Indirect estimators of intergenerational persistence by contrast, weight the persistence rates of different pathways differently. The weights ω_j^{Sur} still depend on the return on X_{it}, ρ_j, but crucially also on the variance of the *expected value* of X_{it} *given* Surname Sur_{it-1}. As Del Pizzo et al. (2026) show, they therefore *"overweight"* pathways with higher persistence rates, as surnames only carry information about these characteristics. The longer the distance to the last common ancestor, the more weight the estimator puts on these characteristics. Surnames carry almost no information about traits which are inherited with low persistence and fade within one or two generations and wash out before the common ancestor of the surname group. One example is health, which clearly has an effect on income but is not strongly persistent across generations. Geographical location or race by contrast are very persistent across generations and influence income strongly, the former through geographical access to economic opportunity, the latter through potential discrimination based on race.

As group-based indirect estimators put more emphasis on more persistent characteristics and I assume in the model that broader capital has a high persistence, these indirect measures are better at uncovering the persistence of underlying broad capital, unbiased by more temporary transmission pathways like health.

This can also explain the remarkable wealth persistence of the groups with names associated with *ancien régime* nobility. The social capital institutionalised in a noble title is associated directly with the surname and is therefore highly persistent *as a characteristic* — by construction $\lambda_j \approx 1$ in the male line. Surname-based measures will overweight this characteristic heavily in the measured intergenerational persistence rate of wealth. Our *ancien régime* nobility-specific persistence rate even sorts on it. The high measured persistence therefore mostly is evidence for the persistent ability of this social capital pathway to generate an economic return ρ_j that increases wealth in each period t.

The theoretical framework presented in this paper is an operationalisation of the framework proposed by P. Bourdieu (1984, 2011). With its focus on multiple transmission channels — social and cultural

capital — which can generate an economic return r_{it}, it is nested in this more general model that also nests the classic latent factor model (Clark & Cummins, 2015).