Lewis Lab Working Paper 2024-05

Demographic Transitions, Rural Flight, and Intergenerational Persistence: Evidence From Crowdsourced Genealogies

Guillaume Blanc
The University of Manchester

April 10, 2024

Abstract

This paper draws on a novel dataset crowdsourced from publicly available online genealogies to study demographic change and development in Europe before modern censuses became available. Using millions of publicly available family trees, I reconstruct fertility from horizontal lineages and identify migration to and from urban centers. Then, I systematically compare the data to a range of representative sources in thirty countries and show that selection is limited after the mid-seventeenth century. Finally, I document novel stylized facts on the rural flight, the demographic transition, and the intergenerational persistence of migration, fertility, and longevity; providing suggestive evidence that substantial changes took hold in the eighteenth century, in the early stages of the transition from stagnation to growth.

JEL Codes: J10, N33, O10

Keywords: fertility, demography, migration, development

Demographic transitions, rural flight, and intergenerational persistence: evidence from crowdsourced genealogies

Guillaume Blanc, *The University of Manchester*[*]

April 10, 2024

Link to the latest version

Abstract

This paper draws on a novel dataset crowdsourced from publicly available online genealogies to study demographic change and development in Europe before modern censuses became available. Using millions of publicly available family trees, I reconstruct fertility from horizontal lineages and identify migration to and from urban centers. Then, I systematically compare the data to a range of representative sources in thirty countries and show that selection is limited after the mid-seventeenth century. Finally, I document novel stylized facts on the rural flight, the demographic transition, and the intergenerational persistence of migration, fertility, and longevity; providing suggestive evidence that substantial changes took hold in the eighteenth century, in the early stages of the transition from stagnation to growth.

JEL codes: *J*10, *N*33, *O*10
Keywords: *fertility, demography, migration, development*

1 Introduction

The transition to modern economic growth brought dramatic transformations and upheavals in the eighteenth and nineteenth centuries. Yet there is hardly any detailed data available on populations at the time since modern censuses only became available after the mid-nineteenth century, and often in aggregate form only.[1]

This research studies demographic change and development in the past, using a novel dataset crowdsourced from publicly available online genealogies, with emphasis on migrations, fertility transitions, and intergenerational associations, in Europe. The raw

[*] Arthur Lewis Building, Oxford Road, Manchester M13 9PL, UK, **guillaume.blanc@manchester.ac.uk**.
I warmly thank Oded Galor, Romain Wacziarg, Stelios Michalopoulos, and David Weil for their kind support and dedicated supervision. This research has benefited from insightful comments and suggestions of Philipp Ager, Diego Alburez-Gutierrez, Greg Clark, Anne McCants, Nathan Nunn, Sheilagh Ogilvie, Sarah Patterson, Giovanni Peri, Faustine Perrin, Leigh Shaw-Taylor, Joachim Voth, and members of the Growth Lab at Brown University. I also thank seminar and conference participants at Brown, IESEG Migration and Family Workshop, Max Planck Institute for Demographic Research, NBER Summer Institute, Oxford, Population Association of America, SDU, Toulouse School of Economics, Zurich, and World Economic History Congress for helpful comments and discussions. I am grateful to the Population Studies and Training Center at Brown for training support and for general support. All remaining errors are my own.

[1] See for an estimate of the timing of introduction of the modern census across countries.

data, the *familinx* dataset, consists of all genealogical records from geni.com, scraped, cleaned, and made publicly available by Kaplanis et al. (2018). It contains information on eighty-six million individuals, with ancestral (vertical) lineages, intergenerational links, and locations and dates of birth and death. Following Blanc (2024), who uses this data to study the cultural origins of the demographic transition in France, I trace horizontal lineages to reconstruct fertility rates in thirty countries and merge the genealogies with municipality-level data on historical urbanization over time to identify the rural-urban status of individuals' municipalities of birth and death.

To comprehensively and systematically evaluate the quality of the data across such a large number of countries, I gather a wide range of data from modern censuses and a number of other sources on fertility, urbanization, and longevity in the thirty countries under study, and assess selection into the sample by comparing the genealogies to the best available representative series in these countries. The data suggests that selection is limited from the mid-seventeenth century to the early twentieth century and that the millions of individuals in the genealogies in Europe form a representative sample of the overall population in their respective countries: by and large, the genealogies consist of ordinary individuals.[2] I also show that, in periods without censuses, new estimates of urbanization and fertility can be constructed for the first time while available estimates can be significantly improved upon.

Finally, I generate novel stylized facts about the transition from stagnation to growth in Europe to illustrate what that can be done using this data. I comprehensively document: the rise in human mobility and the rural flight into urban centers in the early nineteenth century, by looking at the geographical distance from birth to death location or the share of rural-born individuals who died in a place that was a city at the time of their birth or in the same place as their place of birth; the exceptionally early decline in fertility in France, more than a century earlier than in any other country (already documented in earlier work, see Blanc, 2024); the gradual cultural diffusion of the historical fertility transition from France to the rest of Europe, from before the French transition began to after it spread to all countries; and a rapid change, as early as the eighteenth century, in the degree of intergenerational persistence of fertility, while intergenerational elasticities in migration and longevity remained stable.[3]

This research adds to a large literature in economic history on the (automated) linking of historical data (Abramitzky et al., 2021*b*; Abramitzky, Mill and Pérez, 2020; Bailey et al., 2020), mostly across censuses and using data from the US, which boasts some of the most comprehensive census records dating back to the 1850s. Bailey et al. (2023);

[2] Throughout the paper, all series focus on the period 1675–1900 for urbanization, 1700–1925 for fertility, and 1750–1975 for longevity—plotted against years of birth, marriage, and death—since this is when, on average, selection into the data is most limited. The series for periods outside of this window are available in the supplementary data files.

[3] Some of these facts, e.g. the rural flight into urban centers, are already known but have never been comprehensively measured.

Black et al. (2023); Price et al. (2021) combine instead vital records with census data to improve census linking, while Collado, Ortuño-Ortín and Stuhler (2023) combine full count censuses with a range of other dataset, but such data is not available in most countries. My contribution is to use genealogical data to study demographic change and development in Europe. I also comprehensively assess selection into the sample in the countries under study by comparing the genealogical data to the best available representative data in those countries.

Second, it adds to a large literature in historical demography on family reconstitutions from parish records, dating back to the 1970s (Henry, 1972*a*,*b*, 1978; Henry and Houdaille, 1973; Houdaille, 1976; Wrigley et al., 1997; Wrigley and Schofield, 1981). However, this literature suffers from a number of issues, in particular that of representativeness, since it is so time intensive that it can only cover a handful of small villages. More recently, Galor and Klemp (2019); Shiue (2017) trace the descendants of specific founding populations, clans or migrants within the Anhwei Province of China and in Quebec, Canada. My contribution is to use big crowdsourced genealogical data across a large number of regions, encompassing individuals born in both rural and urban places, as shown in the evaluation of representativeness.

Third, it adds to an emerging literature using genealogies to study historical populations. Clark (2023); Cummins (2017) study lineages of persons with rare surnames or of elites. Following Kaplanis et al. (2018), who made data from online family trees on geni.com publicly available, a number of studies have used this data to study longevity, in Germany, the Netherlands, and the US (Colasurdo and Omenti, 2024; Minardi, Corti and Barban, 2023; Stelter and Alburez-Gutierrez, 2022).[4] Blanc (2024) was the first research to use the familinx data to study fertility, providing a new methodology to reconstruct fertility from crowdsourced genealogies, later adopted by Gay, Gobbi and Goñi (2023). My contribution is to reconstruct fertility from horizontal lineages, to merge the genealogical data with data on historical urbanization to study migrations, and to evaluate selection at the country level, in a *systematic* manner and for such a large set of countries, and to make this data publicly available for others to use.

Finally, it adds to different literatures on fertility transitions (e.g. Beach and Hanlon, 2022; Spolaore and Wacziarg, 2021), migrations (e.g. Abramitzky and Boustan, 2017; Beck Knudsen, 2024), and intergenerational associations (e.g. Abramitzky et al., 2021*a*; Chetty et al., 2014). My contribution is to document novel stylized facts on demographic change over time in the past, using novel measures or data not previously available. For example, studying fertility before census data became available; tracing changes in the geographical distance from birth to death location or in the proportion of rural-born individuals who died in a city or in their birthplace; and documenting the evolution of

[4] Charpentier and Gallic (2020) also use crowdsourced genealogical data from French website Geneanet, to study selection. However, this data is not publicly available.

intergenerational associations for fertility, migration, and longevity.

2 Data, Measurement, Selection

2.1 Online family trees

The data. The genealogies used in this paper rely on the crowdsourced work of persons who traced their lineage back through history to reconstruct their family tree by searching church registers of baptisms, burials, and marriages, which "life consists only of" (Wrigley et al., 1997, p. 12). In most of Europe, these records are available online with unrestricted access from the mid-seventeenth century onward and are categorized by town and year. Users of the genealogical website geni.com search for their ancestors in the parish records and, for each generation, find precise information regarding birth, marriage, and death. Then, they upload their family tree online. Appendix Figure A4 shows a screenshot of the home page of geni.com (Panel A) and the profile of an individual, Guillaume Blanc, born in 1668 (Panel B). We can also see the type of information recorded (date and location of birth and death, parents' and siblings' names).

I use the raw data from Kaplanis et al. (2018) in the main database. Observations were scraped from the universe of public profiles on geni.com. The data consists of 86,124,645 individuals along with direct intergenerational links. About 16 million individuals were placed into latitude/longitude coordinates. In the main sample, I select all 9,426,965 individuals who were born, had their first child, or died in thirty countries in Europe (see Appendix Figure A3 for the list). Additionally, I improve the geocoding of places of birth and death by merging different coordinates for each place into the coordinates of the 2016 NUTS statistical regions of Europe. This allows for classification into provinces (NUTS 3), regions (NUTS 2), and countries, which can be used to match the data to other sources at the provincial or regional level – for example to capture within-country differences along dimensions such as income per capita or schooling. Finally, in most of the figures in this paper, to simplify my presentation of the data, I aggregate countries into broader European regions: the British Isles, central Europe and the Low Countries, France, northern Europe, and southern Europe.

Link to historical urbanization. To account for rural-urban differences and to proxy for within-province variation in prosperity, I merge municipalities of birth and death in the genealogies to historical data on urbanization at the town level from Bairoch, Batou and Chèvre (1988). The data is available every fifty years from 1500 to 1850, and places are coded as urban if they were home to more than five thousand inhabitants at the time of the birth of the individual. This further allows me to reconstruct demographic series by historical urban status and, more importantly, to evaluate the representativeness of the sample.

Reconstructing fertility. I reconstruct fertility from the genealogies, adopting the methodology first introduced by Blanc (2024).

Two methods have been traditionally employed to study historical fertility. First, family reconstitutions from parish records, essentially case-studies, have gathered data on a limited number of villages in preindustrial Europe (more details below in 'Genealogies and family reconstitutions'). Second, aggregated administrative data from historical censuses or population counts have been used. In particular, Coale and Watkins (1986) reconstructed fertility in the past, starting from the mid-nineteenth century in Europe. Their preferred measure is the index of marital fertility I_g, defined as the ratio of actual births to potential births to married women. However, this data is not available before modern censuses became available in the 1830's. Additionally, it is only available at the aggregate level. This measure also does not capture well declining fertility since the marital-fertility index is sensitive to the age distribution of the population (see Brown and Guinnane, 2007).

In this paper, fertility is defined at the individual level as the total number of children ever born.[5] The main concern is the recording of horizontal lineages, which are needed to gather full rates of fertility, but are often missing. By definition, only vertical lineages are complete in genealogies. To deal with this issue, following Blanc (2024) and Gay, Gobbi and Goñi (2023), I define the fertility sample, the sample with a recorded horizontal lineage, by retaining only observations for which at least one parent in any of the four generations preceding an individual's observation is recorded as having a fertility rate that is strictly greater than one. Appendix Figure A5 plots an individual's family tree to illustrate this restriction. There are thirty parents in the four generations preceding individual i.[6] If at least one (excluding i) was recorded as having strictly more than one child, then i is included in the fertility sample.

Other issues could arise. In particular, individuals with fewer children may have fewer descendants and therefore may be less likely to appear in the data. And it is possible that only wealthy and educated individuals reconstruct their family trees. These concerns cannot be addressed directly, but they are unlikely to significantly confound the analysis. First, an individual born in 1700 likely has many thousands or even millions of descendants even with rates of fertility persistently below replacement level. Second, there is imperfect intergenerational transmission of fertility and wealth: an individual with low fertility may have children with high fertility; and the distant ancestors of wealthy people today were not necessarily wealthy, even with a high degree of persistence.[7] Third,

[5] Hence it is a measure of completed gross fertility. Because the vertical lineages of the horizontal branches of family trees are not recorded, child mortality is unfortunately under-recorded; therefore net fertility cannot be assessed.

[6] The number of ancestors is computed as follows: $\sum_{k=1}^{4} 2^k = 30$.

[7] Regarding fertility, see Galor and Klemp (2019) or this paper's estimates of the intergenerational transmission of fertility. Regarding wealth or income, conventional estimates of historical intergenerational wealth or income elasticity are usually in the 0.2-0.5 range (Clark et al., 2014), while Clark and Cummins (2015) find a higher

it is possible to compare series in the crowdsourced sample to representative series based on census data to evaluate whether selection with respect to these issues is significant (see the following subsection).

The resulting fertility sample comprises 759, 824 individuals. A few observations are in order. First, because of the individual-level nature of the data, fertility in the genealogies is a different measure from most other available measures used by demographers, who rely on aggregate data to reconstruct marital-fertility indexes or age-specific fertility rates. Second, marriage is not observed and therefore is proxied by year of birth of the first child. Third, childless individuals are not recorded in the fertility sample since it would require a perfect recording of the vertical lineages of the horizontal branches, which is highly unlikely and cannot be evaluated. Finally, not all individuals have a spouse in the data, and sometimes only one of the two parents is recorded. To deal with this issue, standard errors are always clustered at the couple level and both spouses are included in time series and regression analyses.[8] Finally, Appendix Table A2 displays summary statistics for the main and fertility samples.

Weighing. In the analyses at the European and regional levels, observations are reweighed to account for the over-representation of some countries in the genealogies.[9]

Crowdsourced genealogies and family reconstitutions. We used to know very little about historical populations before modern censuses became available. A large literature, in history, has tried to reconstruct population dynamics in the past at the town level, usually for a single or at most a few dozen rural villages, following the path-breaking work of Henry (1972*a*,*b*, 1978); Henry and Houdaille (1973); Houdaille (1976) in France and Wrigley et al. (1997); Wrigley and Schofield (1981) in England and Wales.[10] However this approach raises the important issues discussed in Alter (2019); Ruggles (1999); Schofield (1972); Séguy (2001) among many other studies. First, because of the cost and difficulties associated with searching for records in populated areas, only a small number of small rural villages can be studied, which raises the issue of selection and representativeness. Second, migration cannot be accounted for in historical reconstitutions by demographers because they can only search through parish records one place at a time. Third, reconstructing fertility requires full knowledge of horizontal lineages (all births from a parent or couple), with limited hints on where or

degree of wealth persistence with estimates of roughly 0.7. Even in the most conservative estimates, wealth or income differences would fade away for individuals born before 1900, which is what I find in the data.

[8]The couple associated with individual i is defined as the pair comprising individual i and their "spouse," where the spouse is the individual with whom i had most of their children.

[9]The weighing scheme does so using the total number of observations in the dataset for a given country divided by the population in 1820 (the earliest year for which all countries have an estimate for population) from Bolt and van Zanden (2020). Weights are provided in the data file. It is unfortunately impossible to use time-varying weights since historical population estimates are not available yearly, and not for most countries, at least until the second half of the twentieth century.

[10]With recent contributions by Blanc and Wacziarg (2020); Cummins (2012).

when to look. Moreover, records often provide limited information (names can change, age or date of birth is not always provided, dates are rounded up), and significant investment is required to cross-check their information. Finally, there are the issues of poor handwriting in early registers of dubious quality (for example, see Figure A2).

Online genealogies allow to crowdsource such tedious, difficult, and uncertain work to family members and provide greater incentives to overcome these issues and thoroughly track the entire universe of birth, marriage, and death registers related to a single ancestor and their children since their descendants today possess a knowledge of family history that may help them with family reconstitution of their direct ancestors despite poor record quality and past migrations. Additionally, and more importantly, crowdsourced genealogical data provides substantial spatial variation that can be exploited by researchers interested in studying the determinants of urbanization, migration, fertility, and longevity. Instead of capturing the full population of a small number of villages using family reconstitutions, the genealogies allow to study small numbers of individuals originating from each particular place, but (almost) the entire universe of municipalities, including both villages and cities.

2.2 Measurement and sample selection

To evaluate the degree of selection in the crowdsourced data, I compare observables in the genealogies to representative country-level series relying on census data and on a wide range of representative data. For the representative series, I leverage various estimates of historical rates of urbanization based on estimations of urban and total population from Bairoch (1988); Bairoch and Goertz (1986); Bairoch, Batou and Chèvre (1988), Bairoch, Batou and Chèvre (1988) combined with McEvedy and Jones (1978) and de Vries (1984); and I generate new estimates combining the data on urban population from Bairoch, Batou and Chèvre (1988) and data on total population from Bolt and van Zanden (2020). Additionally, I use data from census and population counts in the Human Mortality Database (2019) for life expectancy and in Coale and Watkins (1986) for marital fertility. Blanc (2024) leverages some of these sources for France, while I systematically and comprehensively use the best available representative data from a variety of historical sources, for thirty different countries in Europe. Appendix Section 1 and Appendix Table A1 provide more details on the representative series. Throughout the paper, all series focus on the period 1675–1900 for urbanization, 1700–1925 for fertility, and 1750–1975 for longevity, plotted against years of birth, marriage, and death, since this is when selection into the data is most limited. The series for periods outside of this window are available in the supplementary data files.

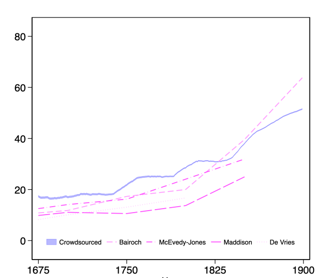
Figure 1, Panel A, displays correlation coefficients between the crowdsourced data and representative country-level series across European regions for urbanization, fertility, and longevity. Across European regions, there is a particularly high correlation between the genealogies and representative series when both are available, suggesting that selection into the sample is limited and that individuals in the genealogies are a representative sample of the overall population—except perhaps in Southern Europe, which has less observations. To compare changes over time across regions, Panels B-G, H-M, and N-S, display urbanization, fertility, and longevity in the crowdsourced genealogies and in representative data over time in the British Isles, Central Europe and the Low Countries, Eastern Europe, France, Northern Europe, and Southern Europe. During the period under study, when the representative data is available, the series coincide almost perfectly, further suggesting that the degree of selection is very limited, and that the genealogies are a representative sample for the period under study. Appendix 2 presents the corresponding ninety country-level series, for urbanization, fertility, and longevity; while Appendix Figure A6 shows the data at the aggregated level, for all of Europe.

Figure 1: Correlation between genealogies and representative data

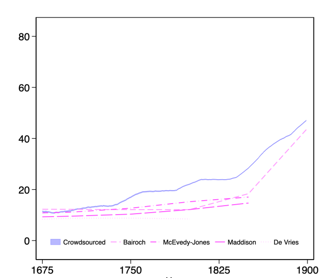
Note: This figure displays a summary of my evaluation of selection into the sample, comparing the genealogical data to representative series. Panel A displays the correlation between the crowdsourced sample and representative data for urbanization, fertility, and longevity. Panels B-G display urbanization in the crowdsourced sample and in representative series over time in Europe. Panels H-M display fertility in the crowdsourced sample and in representative series over time in Europe. Panels N-S display longevity in the crowdsourced sample and in representative series over time in Europe. Urbanization is defined as the share of the population born in a given year in a town coded as urban at the time of an individual's birth (in the genealogies) or as the share of the population living in an urban town in a given year (in representative data). Fertility is defined as the average number of children born to individuals who had their first child in a given year (in the genealogies) or as the marital-fertility index in a given year (in representative data). Longevity is defined as the average age at death minus 30 years for individuals who died aged 30 years or older in a given year (in the genealogies) or as the life expectancy at 30 in a given year (in representative data). All observations were reweighed to account for country population and differential selection into the sample. Appendix Section 1 and Appendix Table A1 provide more details on the representative series.

	Urbanization	Fertility	Mortality
British Isles	0.98	0.86	0.97
Central Europe and Low Countries	0.92	0.94	0.88
Eastern Europe	0.90		
France	0.98	0.93	0.86
Northern Europe	0.95	0.97	0.87
Southern Europe	0.46	0.50	

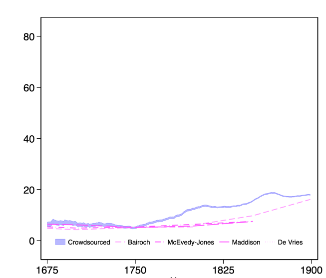
(Panel A) Correlation between genealogies and representative data



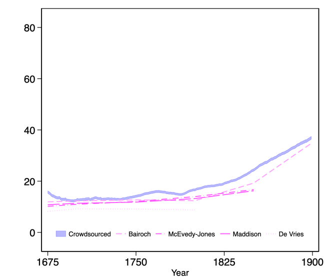
(Panel B) (Urbanization) British Isles



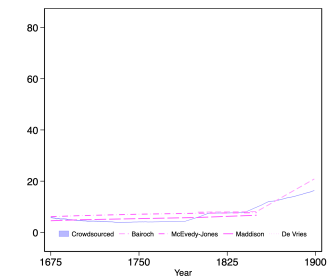
(Panel C) (Urbanization) Central Europe and Low Countries



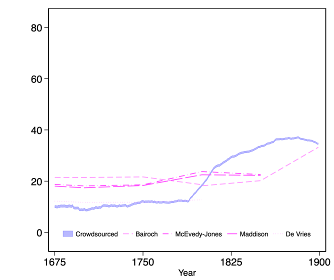
(Panel D) (Urbanization) Eastern Europe



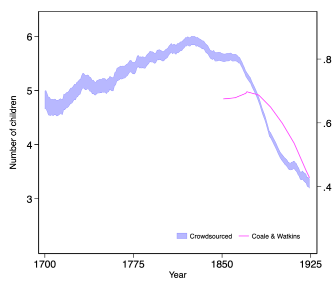
(Panel E) (Urbanization) France



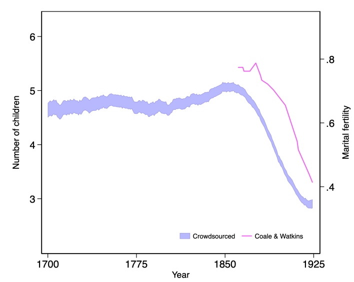
(Panel F) (Urbanization) Northern Europe



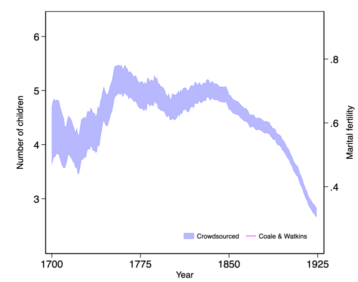
(Panel G) (Urbanization) Southern Europe



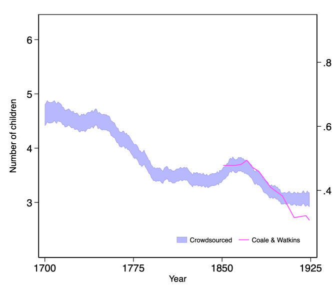
(Panel H) (Fertility) British Isles



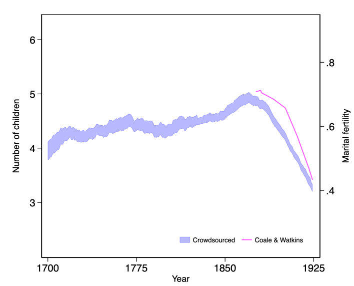
(Panel I) (Fertility) Central Europe and Low Countries



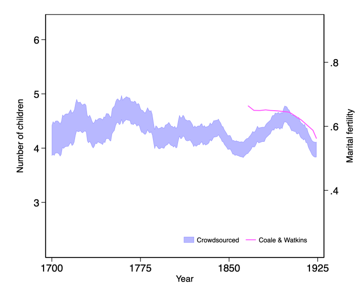
(Panel J) (Fertility) Eastern Europe



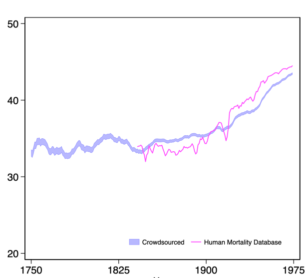
(Panel K) (Fertility) France



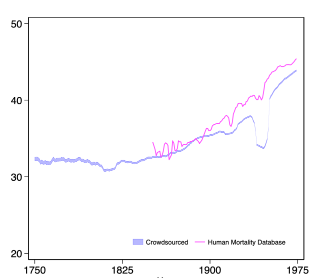
(Panel L) (Fertility) Northern Europe



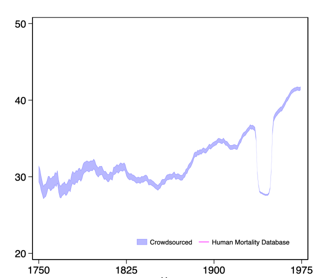
(Panel M) (Fertility) Southern Europe



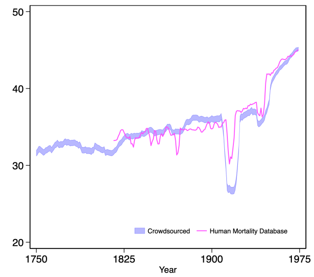
(Panel N) (Longevity) British Isles



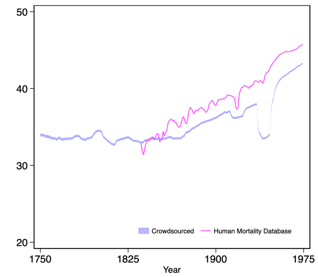
(Panel O) (Longevity) Central Europe and Low Countries



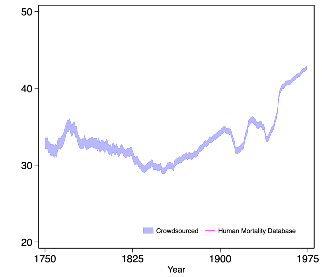
(Panel P) (Longevity) Eastern Europe



(Panel Q) (Longevity) France



(Panel R) (Longevity) Northern Europe



(Panel S) (Longevity) Southern Europe

Using the genealogies, it is also possible to improve the measurement of urbanization, fertility, and longevity before modern censuses became available. While the representative series are not always available throughout the period under study, the genealogical data is. It is possible to extend the available series for all countries, and even to generate new series, otherwise unavailable, for a number of countries, especially in Eastern Europe.

Typically, data on urbanization is available throughout the period, while data on longevity and fertility is only available in the nineteenth century. For urbanization, although the difference between the genealogies and representative series is small, there are sometimes important differences across the different representative series. However, Blanc (2024) shows that, when comparing the genealogies to population counts from censuses in France, the genealogies match almost perfectly the rate of urbanization in the census, while the historical rate of urbanization estimated in the available series by Bairoch and others is consistently and significantly lower. Indeed, to estimate urbanization at the country level, the available series rely on two estimates: an estimation of the urban population and an estimation of the total population. While the population of urban centers is well estimated by Bairoch, Batou and Chèvre (1988) (see Appendix Figure A7), population at the country level is famously poorly estimated before census data became widely available (see Guinnane, 2021).

For fertility, representative data is typically not available until the mid-nineteenth century. However, trends in the genealogical data before the availability of census data look visually consistent with what we know about historical demography (Galor, 2011). Additionally, using aggregate statistics from a comprehensive sample of 404 parishes in England from Wrigley and Schofield (1981), Blanc (2024) shows that the genealogical data also tracks particularly well the representative series from Wrigley and Schofield (1981) in the eighteenth century.

3 Using the Genealogies to Document Novel Stylized Facts

In this section I present novel stylized facts, some of which already known but not well measured, while others are entirely new. Importantly, this is not a comprehensive study of causes and consequences; rather, only an illustration of the aggregate-level patterns that can be documented using this data.

> **Stylized fact # 1.** *The rural flight into urban centers took hold in the late eighteenth century in Europe.*

Using the genealogical data, I first study patterns of migration and find important increases in human mobility in the first half of the nineteenth century in Europe in Figure 2. The historical urbanization of Europe and its causes are known, but migrations from rural to urban had never been observed in a historical setting. Panel A plots the share of rural-born individuals who were born and died in the same place. Almost two thirds of those born before 1750 were *stayers*, who died in their municipality of birth. The likelihood of migration increased over time, and in particular in the nineteenth century. Panel B shows the rural flight into urban centers, measured using the share of rural-born individuals who died in a place that was urban at the time of their birth. Less than 5 percent of rural-born individuals born before 1800 died in a place that was a city at the time of their birth. This number increased to almost 30 percent only a century later.

Panel C answers the question of how far these individuals migrated by showing the hyperbolic sine of the average, median, and 75th percentile of the inverse hyperbolic sine of distance (in kilometers) between places of birth and death for rural-born individuals.[11] While more than half of individuals born in Europe before 1820 died in their place of birth, the median individual born in Europe in 1900 died about fifty kilometers away from their place of birth, reflecting not only urbanization and rural flight but also the fact that people migrated farther than ever before. The average distance also significantly increased in the early nineteenth century.

> **Stylized fact # 2.** *The demographic transition took hold in the late nineteenth century in Europe.*

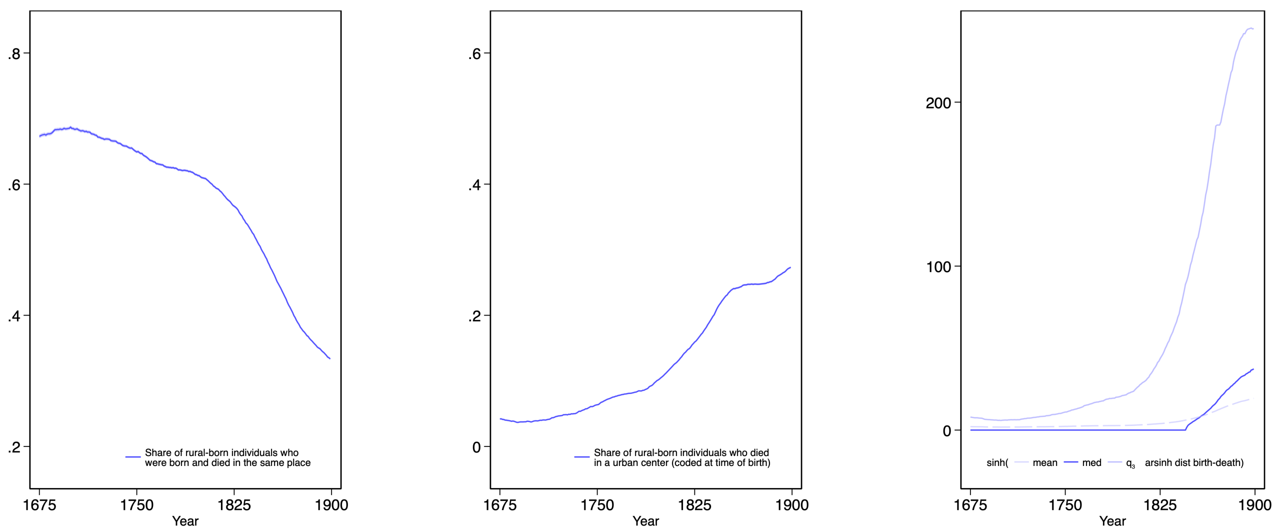
Then, I study the historical fertility transition during the demographic transition.[12] Although the final stage of the transition is well documented using census data, there is only very limited data before modern censuses became available in the second half

[11] Note that the values for the hyperbolic sine of the median and 75th percentile of the inverse hyperbolic sine are exactly equal to the median and 75th percentile of distance since the inverse hyperbolic sine is a monotonic transformation.

[12] Appendix Figure A8 also shows the rise in adult life expectancy in the nineteenth and twentieth centuries, capturing the historical adult mortality transition.

Figure 2: Rural flight

Note: This figure displays evidence on the rural flight and the rise in human mobility in Europe.Panel A displays the share of rural-born individuals who were born and died in the same place.Panel B displays the share of rural-born individuals who died in a place that was urban at the time of their birth.Panel C displays the hyperbolic sine of the median, mean, and seventy-fifth percentile of the inverse hyperbolic sine of distance (in kms) from birth to death location for rural-born individuals, plotted against the year of birth of individuals.In all panels, only rural-born individuals who were born and died in Europe are included. The shaded area represents 90% confidence intervals. Geographical distance is generated using the Stata command `geodist` (Picard, 2019). All observations were reweighed to account for country population and differential selection into the sample. The shaded area represents 90% confidence intervals.



(Panel A) Stayers

(Panel B) Rural flight

(Panel C) How far?

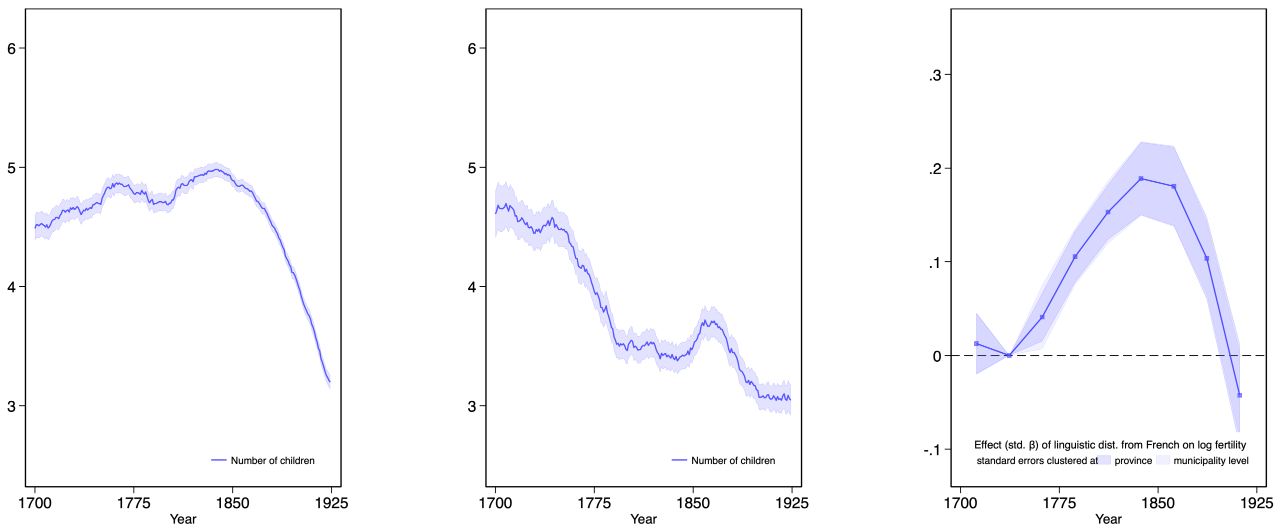
of the nineteenth century. In the Malthusian epoch, the positive relationship between income and population condemned societies to stagnation in income per capita and high fertility. As technological progress accelerated, returns to human capital rose and fertility decisions were altered, gradually triggering the onset of the decline in fertility and the transition toward modern economic growth as parents substituted quantity for quality. This process especially affected elites, with a reversal in the income-fertility gradient over the course of the historical fertility transition (Galor, 2011; Galor and Weil, 2000; Galor and Moav, 2002).

Figure 3, Panel A shows that the historical fertility transition and escape from Malthusian stagnation took hold in the late nineteenth century in Europe. Before 1800, the average woman gave birth to more than four children. In the first half of the nineteenth century, as the Industrial Revolution took hold, this number increased to more than five (suggesting a dominating income effect in the quantity-quality trade-off) before declining to about three in 1925. The effect is particularly marked in England (see in Appendix A2.1).

Stylized fact # 3. *The historical fertility transition first took hold in the mid-eighteenth century in France, more than a century earlier than in the rest of*

Figure 3: The demographic transition

Note: This figure displays evidence on the historical fertility transition during the demographic transition in Europe. Panel A displays average fertility over time for individuals in the fertility sample who were born, had their first child, or died in Europe (excluding France), plotted against year of birth of first child. Panel B displays average fertility over time for individuals in the fertility sample who were born, had their first child, or died in France, plotted against year of birth of first child. Fertility is defined as the average number of children born to individuals who had their first child in a given year. Panel C displays the results (standardized beta coefficients) of the regression of log fertility on linguistic distance from French, plotted against period of birth of first child and compared to the period 1725-1750, immediately before the onset of the decline in fertility in France. The following controls are included: urbanization status of the town of birth, gender, and period fixed effects, a quadratic term in age at birth of the first child (and its interaction with the male dummy), and urbanization status of the town of birth-by-period fixed effects. Linguistic distance from French is taken from Spolaore and Wacziarg (2021). Periods are defined over twenty-five years, and coefficients are centered in the middle of each period. Standard errors are two-way clustered at the couple level and at the town or province (NUTS 3) level. Appendix Figure A8, Panel B plots the raw coefficients over time and finds similar results. All observations in Panels A and C were reweighed to account for country population and differential selection into the sample. The shaded area represents 90% confidence intervals.



(Panel A) The historical fertility transition in Europe

(Panel B) The early decline in fertility in France (Blanc, 2024)

(Panel C) The diffusion of the historical fertility transition in Europe

Europe.

Where and when did the onset of the historical fertility transition take place? Figure 3, Panel B shows that France experienced a sustained decline in fertility more than a century earlier than in the rest of Europe. The early demographic transition in France is known, yet its timing has long been uncertain. The work of demographers, with family reconstitutions, brought attention to this event—which can be estimated to have taken place in the 1770s using the family reconstitutions data (Blanc, 2024). Cummins (2012), using estimates from Weir (1994), also argues that fertility started to decline in 1776. However, as discussed in Section 2 and as portrayed in Figure 1, Panel E, the data suffers from important limitations; in particular, the fact that it only selects a small number of small, rural villages. Using the genealogical data, Blanc (2024) estimates the onset of the decline in fertility to the 1760s, about fifteen years earlier than was previously thought, and discusses the role of the loss of influence of the Catholic Church in France.

Stylized fact # 4. *After the onset of the decline in fertility in France, low fertility spread to the rest of Europe along cultural lines.*

Since the historical fertility transition had cultural origins, one might ask whether cultural proximity or barriers mattered for its spread. Figure 3, Panel C shows that provinces linguistically closer to France adopted limitations on fertility earlier, consistent with a process of social influence. One of the most well known example is that of Belgium, where the French-speaking region of Wallonia experienced a decline in fertility rates several decades earlier than the Dutch-speaking region of Flanders (Lesthaeghe, 1977). Spolaore and Wacziarg (2021) document this pattern using modern census data on fertility at the province level from Coale and Watkins (1986). I also find this pattern in the data, using a similar specification; however, I document the entire process of diffusion, while they can only document it after 1851.

Using the genealogies, I find that fertility limitations started spreading immediately after the onset of the historical fertility transition in France. The size of the effect gradually increased and reached its maximum at the time of the aggregate decline in fertility in Europe in the second half of the nineteenth century. After most societies have successfully made the transition to low fertility, the diffusion is complete and linguistic distance to France does not have any effect anymore in Europe; if anything, the effect is slightly negative.[13]

Stylized fact # 5. *The intergenerational persistence of fertility collapsed in the late eighteenth and late nineteenth centuries in Europe, while intergenerational elasticities in migration and longevity remained stable.*

Finally, I study the intergenerational persistence of demographic behaviors and migration over the long-run, through periods of important changes. I estimate intergenerational elasticities and their changes over time by regressing some monotonic transformation of the outcome of an individual on the same transformation of the average outcome of its parents (e.g., a log-log specification) interacted with period fixed effects, controlling for a range of controls.[14] In particular, I control for period, province, and province-by-period fixed effects to account for cross-sectional province-level differences in each period.[15] As a result, any changes in the estimated elasticities over time will

[13] Blanc (2024), using data from Blanc and Kubo (2024) on linguistic distance from French, also finds that the effect holds within France.

[14] To study fertility and longevity, I rely on a log-log specification. For migration, since most individuals never moved, I rely on an arsinh-arsinh specification, using the inverse hyperbolic sine of distance to account for zeros. Formally, I estimate the following equation, with y the individual's outcome, x the average outcome of the parents, and f some monotonic transformation: $f(y) = \alpha + \beta f(x) + \varepsilon$. With a log transformation, the estimated coefficient captures the intergenerational elasticity of the outcome. In the case of an inverse hyperbolic sine transformation, Bellemare and Wichman (2020) show that the elasticity is $\hat{\beta}\sqrt{(y^2+1)x^2}/\sqrt{(x^2+1)y^2}$ and propose to report it at the means of x and y.

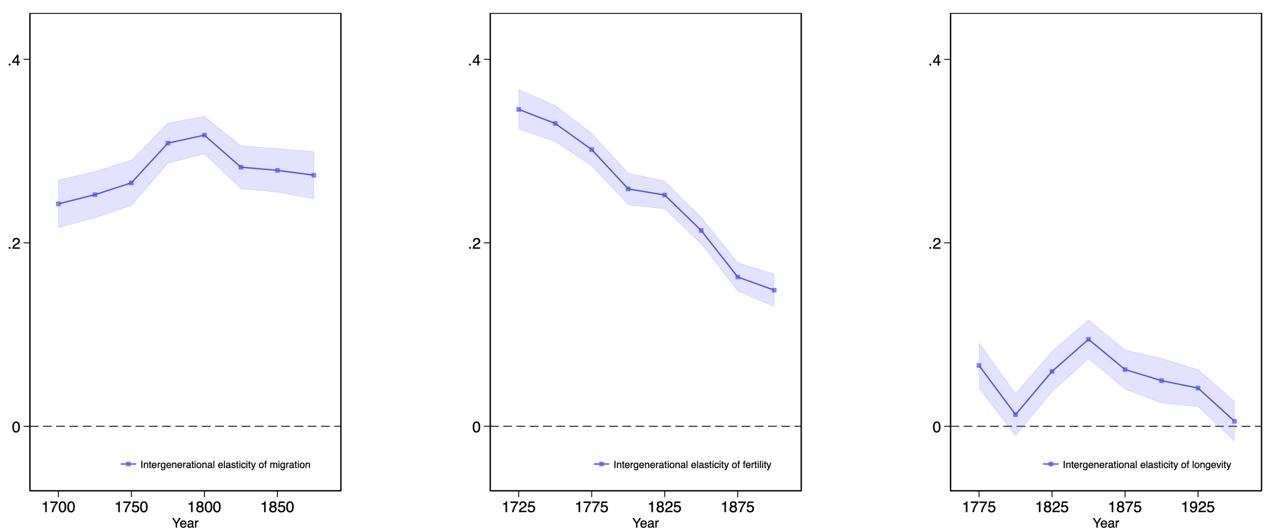
[15] Additionally, I control for rural-urban status, and its interaction with period FEs.

exclusively reflect distributional shifts, not secular transformations. Figure 4 plots the intergenerational elasticities in migration (Panel A), fertility (Panel B), and longevity (Panel C) over time. The data reveals a set of surprising results.[16]

First, the intergenerational elasticity in migration remained stable over time, at around 0.3, in line with what traditional estimates find for income or wealth (Charles and Hurst, 2003; Chetty et al., 2014). This result, suggesting that the same families kept moving while others stayed, and that individuals from these families migrated increasingly farther away (in other words, the absence of distributional changes), is particularly surprising since we know that a rural flight took place in the nineteenth century.[17]

Figure 4: Intergenerational persistence

Note: This figure displays the intergenerational elasticity coefficients for migration (Panel A), fertility (Panel B), and longevity (Panel C), plotted against (respectively) period of birth, period of birth of first child, and period of death. Panel A includes individuals in the main sample who were born and died in Europe; Panel B includes individuals in the fertility sample who were born, had their first child, or died in Europe; Panel C includes individuals in the main sample who died in Europe aged 30 years or older. Migration is defined as the geographical distance (in kms) from birth to death. Fertility is defined as the total number of children born to individuals. Longevity is defined as the age at death minus 30 years. The following controls are included: male, province (NUTS 3) of birth (Panel A and B) or death (Panel C), urban status of town of birth (Panel A and B) or death (Panel C), and period fixed effects, as well as province-by-period and urban status of town-by-period fixed effects. Periods are defined over twenty-five years. Robust standard errors are reported. Geographical distance is generated using the Stata command `geodist` (Picard, 2019). I focus on the period 1700–1900 for migration, 1725–1925 for fertility, and 1775–1975 for longevity (excluding the first 25 years of the period with limited selection each time). All observations were reweighed to account for country population and differential selection into the sample. The shaded area represents 90% confidence intervals.



(Panel A) Migration **(Panel B)** Fertility **(Panel C)** Longevity

Second, I find that fertility became increasingly less intergenerationally persistent. In

[16] I also generate a scale-invariant measure of persistence (similar to a rank-rank specification), the intergenerational correlation, by multiplying the estimated elasticity with the ratio of the standard deviation in the outcome for the individual to that for the parents (see Chetty et al., 2014, for a more detailed discussion). Results are displayed in Appendix Figure A10.

[17] I find some evidence for this in Appendix Figure A9, Panel A, which compares the estimated elasticities with and without period fixed effects. Following the rural flight into urban centers, the intergenerational elasticity in migration increased significantly, to more than .5 (in the specification without any controls).

the first half of the eighteenth century, it was highly persistent, with an intergenerational elasticity in fertility near .4. The elasticity rapidly and significantly decreased on the eve of the nineteenth century, before weakening further during the historical fertility transition in the late nineteenth century. The decline in intergenerational persistence after the onset of the decline in fertility is not surprising since, because of the period fixed effects, it can only be explained by distributional changes, such as a reversal in the income-fertility gradient (see Galor and Weil, 2000; Galor and Moav, 2002), or changes in the distribution of human capital and education, which we know took hold during the historical fertility transition in Europe.[18] In recent research, Vogl (2020) also shows intergenerational associations weakened after the onset of the transition in a sample of developing countries.[19]

However, I find that the intergenerational elasticity also decreased before the historical fertility transition, in the late eighteenth century. Since the masses were not educated until the nineteenth century (Buringh and van Zanden, 2009), and since there is no documented changes in the income-fertility gradient before the transition, this result suggests that important distributional changes took hold in Europe as early as the eighteenth century, during the Age of Enlightenment.[20]

Last, I find a significantly smaller intergenerational elasticity for longevity, with a coefficient of less than 0.1, declining slowly over time to become null by the end of the period. Black et al. (2023) find similar results (although slightly larger elasticities) using US data. This could reflect the fact that mortality only has a small inheritable component and is influenced by a variety of factors largely independent from individual-level characteristics. Additionally, the small decline in the twentieth century could reflect improvements in access to health for the masses, but more research is needed to confirm this.

4 Concluding Remarks

This paper used a novel individual-level dataset from publicly available crowdsourced genealogies to study development and demographic change in the past in Europe. I reconstructed fertility by tracing horizontal lineages and linked observations to data on historical urbanization. By comparing the genealogical data to representative data on urbanization, fertility, and longevity in thirty countries, I showed that there was limited

[18] Appendix Figure A9, Panel B shows that the raw (without controls) intergenerational elasticity was stable before the eighteenth century, further suggesting that distributional changes played a significant role in both periods, while secular changes only played a role at the time of the historical fertility transition, with the decline in fertility.

[19] See also Murphy (1999, 2012) for an overview of the literature, which mostly focuses on contemporary trends—with weak coefficients near 0.

[20] Note that, because I control for province and province-by-period fixed effects, it is impossible to test if this result can be accounted for by the cultural diffusion of the early decline in fertility in France, as the data on linguistic distance in Europe is only available at the regional level.

selection into the sample.

Then, I generated novel stylized facts suggesting that a wave of important changes gave rise to new behaviors in the nineteenth century and that the origin of many of these can be traced back to the eighteenth century, in the early stages of the transition from stagnation to growth. This paper should open avenue for future research to better understand these changes, focusing on the behavior of populations in periods without modern censuses.

References

Abramitzky, Ran, and Leah Boustan. 2017. "Immigration in American Economic History." *Journal of Economic Literature*, 55(4): 1311–45.

Abramitzky, Ran, Leah Boustan, Elisa Jacome, and Santiago Perez. 2021*a*. "Intergenerational Mobility of Immigrants in the United States over Two Centuries." *American Economic Review*, 111(2): 580–608.

Abramitzky, Ran, Leah Boustan, Katherine Eriksson, James Feigenbaum, and Santiago Pérez. 2021*b*. "Automated Linking of Historical Data." *Journal of Economic Literature*, 59(3): 865–918.

Abramitzky, Ran, Roy Mill, and Santiago Pérez. 2020. "Linking individuals across historical sources: A fully automated approach*." *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 53(2): 94–111.

Alter, George C. 2019. "The Evolution of Models in Historical Demography." *The Journal of Interdisciplinary History*, 50(3): 325–362.

Bailey, Martha J., Connor Cole, Morgan Henderson, and Catherine Massey. 2020. "How Well Do Automated Linking Methods Perform? Lessons from US Historical Data." *Journal of Economic Literature*, 58(4): 997–1044.

Bailey, Martha, Peter Z. Lin, A. R. Shariq Mohammed, Paul Mohnen, Jared Murray, Mengying Zhang, and Alexa Prettyman. 2023. "The creation of LIFE-M: The Longitudinal, Intergenerational Family Electronic Micro-Database project." *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 56(3): 138–159.

Bairoch, Paul. 1988. *Cities and Economic Development: From the Dawn of History to the Present, translated by Christopher Braider.* Chicago:University of Chicago Press.

Bairoch, Paul, and Gary Goertz. 1986. "Factors of Urbanisation in the Nineteenth Century Developed Countries: A Descriptive and Econometric Analysis." *Urban Studies*, 23(4): 285–305.

Bairoch, Paul, Jean Batou, and Pierre Chèvre. 1988. *The Population of European Cities, 800-1850: Data Bank and Short Summary of Results.* Geneve:Librairie Droz.

Beach, Brian, and W Walker Hanlon. 2022. "Culture and the Historical Fertility Transition." *The Review of Economic Studies*.

Beck Knudsen, Anne Sofie. 2024. "Those Who Stayed: Selection and Cultural Change in the Age of Mass Migration." Working paper.

Bellemare, Marc F., and Casey J. Wichman. 2020. "Elasticities and the Inverse Hyperbolic Sine Transformation." *Oxford Bulletin of Economics and Statistics*, 82(1): 50–61.

Black, Sandra E, Neil Duzett, Adriana Lleras-Muney, Nolan G Pope, and Joseph Price. 2023. "Intergenerational Correlations in Longevity." National Bureau of Economic Research Working Paper 31034.

Blanc, Guillaume. 2024. "The Cultural Origins of the Demographic Transition in France." Lewis Lab Working Paper 2024-02.

Blanc, Guillaume, and Masahiro Kubo. 2024. "French." Lewis Lab Working Paper 2024-01.

Blanc, Guillaume, and Romain Wacziarg. 2020. "Change and Persistence in the Age of Modernization: Saint-Germain-d'Anxure, 1730-1895." *Explorations in Economic History.*

Bolt, Jutta, and Jan Luiten van Zanden. 2020. "Maddison Project Database: Maddison style estimates of the evolution of the world economy. A new 2020 update."

Brown, John C., and Timothy W. Guinnane. 2007. "Regions and time in the European fertility transition: problems in the Princeton Project's statistical methodology." *The Economic History Review*, 60(3): 574–595.

Buringh, Eltjo, and Jan Luiten van Zanden. 2009. "Charting the "Rise of the West": Manuscripts and Printed Books in Europe, a Long-Term Perspective from the Sixth through Eighteenth Centuries." *The Journal of Economic History*, 69(2): 409–445.

Charles, Kerwin Kofi, and Erik Hurst. 2003. "The Correlation of Wealth across Generations." 111(6): 1155–1182.

Charpentier, Arthur, and Ewen Gallic. 2020. "La démographie historique peut-elle tirer profit des données collaboratives des sites de généalogie." *Population.*

Chetty, Raj, Nathaniel Hendren, Patrick Kline, and Emmanuel Saez. 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.

Clark, Gregory. 2023. "The inheritance of social status: England, 1600 to 2022." *Proceedings of the National Academy of Sciences*, 120(27).

Clark, Gregory, and Neil Cummins. 2015. "Intergenerational Wealth Mobility in England, 1858-2012: Surnames and Social Mobility." *The Economic Journal*, 125(582): 61–85.

Clark, Gregory, Neil Cummins, Yu Hao, Daniel Diaz Vidal, Tatsuya Ishii, Zach Landes, Daniel Marcin, Firas Abu-Sneneh, Wilfred Chow, Kuk Mo Jung, Ariel M. Marek, and Kevin M. Williams. 2014. *The Son Also Rises: Surnames and the History of Social Mobility.* Princeton University Press.

Coale, Ansley J., and Susan C. Watkins. 1986. *The Decline of Fertility in Europe.* Princeton:Princeton University Press.

Colasurdo, Andrea, and Riccardo Omenti. 2024. "Using Online Genealogical Data for Demographic Research: An Empirical Examination of the FamiLinx Database." Working paper.

Collado, M Dolores, Ignacio Ortuño-Ortín, and Jan Stuhler. 2023. "Estimating Intergenerational and Assortative Processes in Extended Family Data." *The Review of Economic Studies*, 90(3): 1195–1227.

Cummins, Neil. 2012. "Marital Fertility and Wealth During the Fertility Transition: Rural France, 1750-1850." *The Economic History Review*, 66(2): 449–476.

Cummins, Neil. 2017. "Lifespans of the European Elite, 800–1800." 77(2): 406–439.

de Vries, Jan. 1984. *European urbanization 1500–1800.* London:Metheun.

Galor, Oded. 2011. *Unified Growth Theory.* Princeton:Princeton University Press.

Galor, Oded, and David Weil. 2000. "Population, Technology, and Growth: From the Malthusian Regime to the Demographic Transition and Beyond." *American Economic Review*, 90(4): 806–828.

Galor, Oded, and Marc Klemp. 2019. "Human genealogy reveals a selective advantage to moderate fecundity." *Nature Ecology & Evolution*, 3(5): 853–857.

Galor, Oded, and Omer Moav. 2002. "Natural Selection and the Origin of Economic Growth." *The Quarterly Journal of Economics*, 117(4): 1133–1191.

Gay, Victor, Paula E. Gobbi, and Marc Goñi. 2023. "Revolutionary Transition: Inheritance Change and Fertility Decline." Working paper.

Guinnane, Timothy W. 2021. "We Do Not Know the Population of Every Country in the World for the past Two Thousand Years." Working paper.

Henry, Louis. 1972*a*. "Fécondité des Mariages dans le Quart Sud-Ouest de la France de 1720 à 1829 (I)." *Annales*, 27(3): 612–640.

Henry, Louis. 1972*b*. "Fécondité des Mariages dans le Quart Sud-Ouest de la France de 1720 à 1829 (suite)." *Annales*, 27(4): 977–1023.

Henry, Louis. 1978. "Fécondité des Mariages dans le Quart Sud-Est de la France de 1670 à 1829." *Population*, 33(4): 855–883.

Henry, Louis, and Jacques Houdaille. 1973. "Fécondité des Mariages dans le Quart Nord-Ouest de la France de 1670 à 1829." *Population*, 28(4): 873–924.

Houdaille, Jacques. 1976. "La Fécondité des Mariages de 1670 à 1829 dans le Quart Nord-Est de la France." *Annales de Démographie Historique*, 1(1): 341–391.

Human Mortality Database. 2019. *HMD.* University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at **www.mortality.org** or **www.humanmortality.de** (data downloaded on November 30, 2019).

Kaplanis, Joanna, Assaf Gordon, Tal Shor, Omer Weissbrod, Dan Geiger, Mary Wahl, Michael Gershovits, Barak Markus, Mona Sheikh, Melissa Gymrek, Gaurav Bhatia, Daniel G. MacArthur, Alkes L. Price, and Yaniv Erlich. 2018. "Quantitative Analysis of Population-Scale Family Trees With Millions of Relatives." *Science*, 360(6385): 171–175.

Lesthaeghe, Ron J. 1977. *The Decline of Belgian Fertility, 1800-1970.* Princeton:Princeton University Press.

McEvedy, C., and R. Jones. 1978. *Atlas of world population history.* Harmondsworth, Middlesex, England:Penguin Books Ltd.

Minardi, Saverio, Giulia Corti, and Nicola Barban. 2023. "Historical Patterns in the Intergenerational Transmission of Lifespan and Longevity: Evidence from the United States, 1700-1900." Working paper.

Murphy, M. 1999. "Is the relationship between fertility of parents and children really weak?" *Social Biology*, 46(1-2): 122–145.

Murphy, Michael. 2012. "Intergenerational fertility correlations in contemporary developing counties." *American Journal of Human Biology*, 24(5): 696–704.

Picard, Robert. 2019. "GEODIST: Stata module to compute geographical distances."

Price, Joseph, Kasey Buckles, Jacob Van Leeuwen, and Isaac Riley. 2021. "Combining family history and machine learning to link historical records: The Census Tree data set." *Explorations in Economic History*, 80: 101391.

Ruggles, Steven. 1999. "The limitations of English family reconstitution: English population history from family reconstitution 1580–1837." 14(1): 105–130.

Schofield, Roger S. 1972. "Representativeness and family reconstitution." *Annales de Démographie Historique*, Techniques et méthodes. Actes du colloque de Florence, 1er-3 octobre 1971.

Séguy, Isabelle. 2001. *La population de la France de 1670 à 1829: l'enquête Louis Henry et ses données. Classiques de l'économie et de la population*, PAris:INED.

Shiue, Carol H. 2017. "Human capital and fertility in Chinese clans before modern growth." *Journal of Economic Growth*, 22(4): 351–396.

Spolaore, Enrico, and Romain Wacziarg. 2021. "Fertility and Modernity." *The Economic Journal*, 132(642): 796–833.

Stelter, Robert, and Diego Alburez-Gutierrez. 2022. "Representativeness is crucial for inferring demographic processes from online genealogies: Evidence from lifespan dynamics." *Proceedings of the National Academy of Sciences*, 119(10): e2120455119.

Vogl, Tom S. 2020. "Intergenerational Associations and the Fertility Transition." *Journal of the European Economic Association.*

Weir, David R. 1994. "New Estimates of Nuptiality and Marital Fertility in France, 1740-1911." *Population Studies*, 48(2): 307–331.

Wrigley, E. A., R. S. Davies, J. E. Oeppen, and R. S. Schofield. 1997. *English Population History from Family Reconstitution 1580–1837. Cambridge Studies in Population, Economy and Society in Past Time*, Cambridge:Cambridge University Press.

Wrigley, Edward A., and Roger S. Schofield. 1981. *The Population History of England, 1541-1871: A Reconstruction.* London:Edward Arnold.

The online appendix for this paper can be found **here**.