Demographic Change and Development from Crowdsourced Genealogies in Early Modern Europe

Guillaume Blanc

To cite this version:

Guillaume Blanc. Demographic Change and Development from Crowdsourced Genealogies in Early Modern Europe. 2020. hal-02922398v2

HAL Id: hal-02922398
https://hal.science/hal-02922398v2

Preprint submitted on 1 Sep 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Demographic Change and Development from Crowdsourced Genealogies in Early Modern Europe

Guillaume Blanc, *Brown University**

August 2020 (**latest version**)
VERY PRELIMINARY VERSION

Abstract

This paper draws on a novel historical dataset crowdsourced from publicly available genealogies to study demographic change and development at the individual level in the distant past. I reconstruct fertility series, identify migration in and out of urban centers, and provide novel measures and stylized facts in a period without census and with millions of ordinary individuals observed in thirty countries. For each country, I carefully show that selection is limited in the data. Then, I document patterns of human mobility, fertility, and adult mortality in Early Modern Europe, through the Industrial Revolution and demographic transition. Finally, I present several findings at a disaggregated level suggesting that substantial and rapid changes in preferences took hold with the Age of Enlightenment and played an important role in the transition from stagnation to growth. In particular, I estimate the onset of the decline in fertility in France in the 1760s, a hundred years before the rest of Europe and earlier than previously thought, and I find a weaker intergenerational persistence of fertility behavior in Europe as early as in the late eighteenth century.

JEL codes: *J*10, *N*33, *O*10
Keywords: *demographic, development, migration, health*

1 Introduction

The emergence of sustained economic growth and the escape from malthusian stagnation in Europe generated substantial changes and sweeping transformations for populations and places. Aggregate statistics exist for some dimensions and in some places after the late nineteenth century, but demographic change and development cannot be studied without more detailed data before, in the distant past, in a period without censuses or representative statistics, and without data on populations and ordinary individuals.

* Department of Economics and PSTC, Robinson Hall, 64 Waterman Street, Providence RI 02912, **guillaume_blanc@brown.edu**. I would like to warmly thank Oded Galor, Romain Wacziarg and David Weil for their kind support and dedicated supervision. This research has benefited from insightful comments and suggestions of members of the Growth Lab at Brown. I also thank seminar and conference participants at Brown University and PAA for helpful comments and discussions. I am grateful to the Population Studies and Training Center at Brown University, which receives funding from the NIH, for training support (T32 HD007338) and for general support (P2C HD041020). Claude Motte and Julien Perret kindly shared data. All remaining errors are my own.

This research uses a novel dataset crowdsourced from publicly available genealogies to do this. I study populations in the past in Europe, focusing on changes in human mobility and demographic behaviors in particular. I present and clean the genealogical dataset crowdsourced by Kaplanis et al. (2018), reconstruct fertility from family trees, and link towns of birth, marriage, and death to historical urbanization data in Europe. Then, I leverage all known, reliable, and representative country-level series of urbanization, fertility, and life expectancy, mostly relying on censuses, in order to thoroughly assess selection into the sample, country by country, by comparing the series to the crowdsourced genealogies. In all, I find that the genealogies are a representative sample from the late seventeenth century to the early twentieth century.

In the paper, I show that estimates of historical urbanization, fertility behavior, and life expectancy can be improved upon or reconstructed for the first time in Early Modern Europe. Estimates of historical urbanization at the country level can be improved from Bairoch, Batou and Chèvre (1988). Similarly, estimates of adult life expectancy and fertility are in line with the estimates of the Human Mortality Database (2019) and Coale and Watkins (1986) but available much earlier than 1831, the earliest date with representative data on fertility.

I show that crowdsourcing publicly available genealogical data provides a new way to study demographic change and development in populations in the past without census, and at the individual level. In particular, the genealogies perform very well in countries such as France or England, or in Northern Europe, allowing to present novel findings on the long run evolution of demographic behavior over the course of development. I also provide novel measures and novel stylized facts for known but not well measured phenomenons.

Using the genealogies, I systematically study human mobility, fertility, and adult mortality in Europe as a whole and at a more disaggregated level, with a particular emphasis on fertility decisions. I find that the transition away from malthusian stagnation and into sustained economic growth generated substantial changes in the late nineteenth century. For example, I quantify the rise in human mobility and rural flight into urban centers following the Industrial Revolution, and estimate completed rates of fertility over time with hundreds of thousands of observations.

Yet, at a more disaggregated level, I identify the roots of changes a century earlier during the Age of Enlightenment, suggesting that cultural changes played an important role. In particular, I estimate the onset of the demographic transition and decline in fertility in France in the 1760s, about fifteen years earlier than previously estimated by Louis Henry (Cummins, 2012). In Europe, I also find, like Spolaore and Wacziarg (2019), that norms of limited fertility spread from France along cultural and linguistic lines, in a process that developped in the second half of the eighteenth century. Finally, I look at intergenerational associations and find that substantial and rapid changes in preferences took hold in the eighteenth century, resulting in weaker persistence of behavior across generations.

The results for fertility represent significant improvements from the series reconstructed in the early modern period with family reconstitutions techniques and the historical search and link of baptism, marriage, and death records at the town level by Henry (1972*a*,*b*, 1978); Henry and Houdaille (1973) in France and Wrigley et al. (1997); Wrigley and Schofield (1981) in England and Wales. Thanks to variation in both space and time, the study of fertility from genealogies provides avenue for future research regarding the understanding of early declines in fertility in France (Blanc, 2020) or in the US (Haines, Hacker and Jaremski, 2020), and the evolution of the fertility-development gradient and changes in intergenerational associations along the escape from the malthusian trap.[1] Most of the current research using individual-level data with substantial variation across space and time (Chatterjee and Vogl, 2018; Vogl, 2016), or with the possibility to study the intergenerational transmission of fertility (Vogl, 2020), has relied on developing countries and not on societies that have successfully industrialized in the past. Using crowdsourced genealogies allows to rely on a novel source of data at the individual level through periods of unprecedented changes.

The results for human mobility, with estimates of both the likelihood of migration and migratory distance, suggest that significant advances can be made regarding the determinants of migration, both within country at the time of the rural flight and urbanization of Europe, and internationally with migration to the Americas. In particular, the genealogies allow to identify the rural flight into urban centers but also migrations from urban centers at the time of the market for ideas in Europe. Moreover, it is possible to track within country of origin migration to the new world from all over Europe, while the best available evidence so far has only used migration from Norway (thanks to particularly detailed censuses) in order to study the origins and self-selection of migrants (Abramitzky, Boustan and Eriksson, 2012).

These things are only a subset of what can be studied. Altogether, this paper is not a comprehensive study or review of these phenomenons, but rather aims to describe improvements in measurement, show that some series can be reconstructed for the first time in order to study demographic change and development in the distant past, and finally selects some findings and stylized facts of interest in a period of rapid and important changes.

Last but not least, this research emphasizes the role played by populations rather than places, following a recent and important literature that has documented deep rooted factors in development and persistence through populations and some of their characteristics (Ashraf and Galor, 2013; Putterman and Weil, 2010; Spolaore and Wacziarg, 2013;

[1]See Ashraf and Galor (2011); Galor and Weil (2000); Galor and Moav (2002) for theoretical contributions.

Voigtländer and Voth, 2012). However, the lack of intermediate outcomes between events in the distant past and current outcomes has been criticized. Using crowdsourced genealogies allows to study some of these intermediate outcomes before the escape from Malthus, and the paper itself should become a useful ressource for researchers hoping to leverage big data in order to answer questions involving these variables in the past, at the individual level and with millions of observations.

2 Data, measurement, and sample selection

In this section, I first provide an overview of the data and its construction, describe the cleaning process and how observations were linked to data on historical urbanization, and finally I describe the reconstruction of fertility through crowdsourced family reconstitution (Section 2.1). Then, I discuss the improvements in measurement permitted by the data and assess sample selection using available observable series at the country level (Section 2.2).

2.1 The data: crowdsourced genealogies

The data The genealogies used in the paper rely on the crowdsourced work of persons tracing 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 some places (e.g. France, or the UK), records are available online with unrestricted access from the mid-17th century onwards and sometimes before (e.g. in Sweden, starting from the early 17th century), and are categorized by town and year. The users of the genealogical website search for their ancestors in the parish records, and for each generation find precise information regarding birth, marriage, and deaths, and trace all subsequent births of children in the lineage (vertical, and sometimes horizontal).

Following the path-breaking work of Henry (1972*a*,*b*, 1978); Henry and Houdaille (1973) in France and Wrigley et al. (1997); Wrigley and Schofield (1981) in England and Wales, a large literature has tried to reconstruct population dynamics in the past at the town level, often to study the population of a single village or at most of a few dozens.[2] However there are important issues, discussed in Alter (2019); Schofield (1972); Séguy (2001) among many others. First, because of the cost and difficulties associated with searching for records in populated areas, only small rural villages can be studied (and only a small number) raising 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 of one place at a time. Third, reconstructing fertility requires a full knowledge of horizontal lineages (all births from a parent or couple), with limited hints on where or

[2] 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 require a significant investment to cross-check informations. Finally, there is the issue of poor handwriting and dubious quality in early registers (for example see Figure A2.1.1).

Crowdsourcing such tedious, difficult, and uncertain work to family members allows for greater incentives to overcome these issues and thoroughly track the entire universe of birth, marriage and death registers involving a single person and its children since descendants possess a knowledge of family history and identity that may help them in the process of family reconstitution of their direct ancestors, through poor record quality and past migrations.

I rely on raw data from Kaplanis et al. (2018) in order to construct the main database. Each observation was scrapped from the universe of public profiles on geni.com, a genealogy website. The raw dataset consists of 86, 124, 645 single individuals worldwide along with a second dataset consisting only in the direct intergeneration links. About 16 millions individuals were placed into latitude/longitude coordinates. In the *main sample*, I select all 9, 426, 965 individuals who were born, married (had their first child), or died in thirty countries in Europe. In order to simplify the presentation of the data, countries are aggregated into country groups: British Isles, Central Europe and Low Countries, France, Northern Europe, and Southern Europe (see Table A2.1.1).

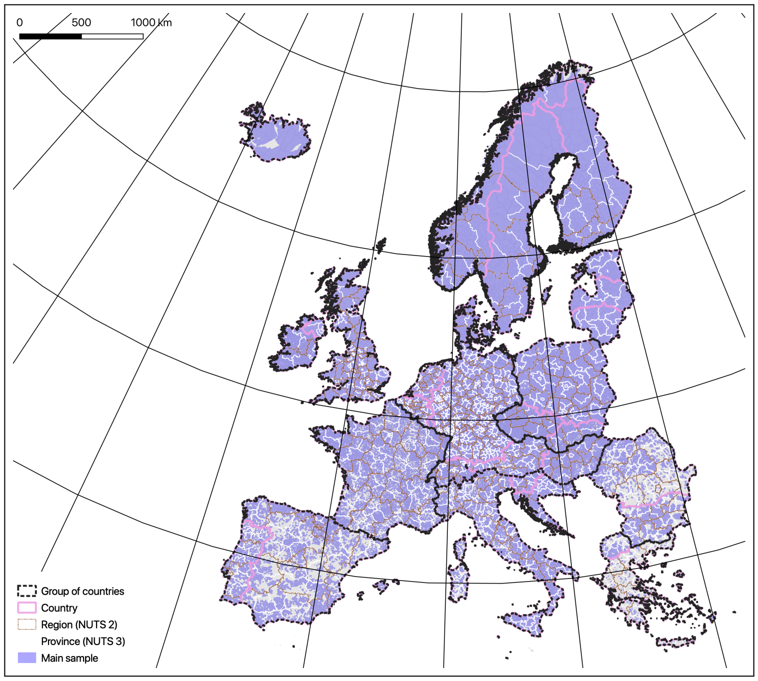
Figure 1 displays the spatial distribution of places of birth in the main sample. I improve the geo-coding by merging different coordinates for each place into the coordinates of the 2016 NUTS statistical regions of Europe. Moreover, it allows a classification into provinces (NUTS 3, the smallest administrative unit in Europe), regions (NUTS 2), and countries, which can be used to match the data to other sources at the provincial or regional level, or to study population dynamics in the past whithin administrative unit, allowing to capture within-country differences in dimensions such as income per capita or schooling. A recent and emerging literature (see Charpentier and Gallic (2020*a*,*b*); Kaplanis et al. (2018); Lussier and Keinan (2018) in particular) has used crowdsourced genealogies to study demographic change and development in the past, yet none has merged the genealogical data with historical urbanization status, has evaluated selection at the country level in a systematic manner using all the available representative or census data, or has leveraged family trees to reconstruct fertility in the distant past.[3]

Link to historical urbanization In order to proxy for within-province variation in prosperity, I merge towns to historical data on urbanization at the town level from Bairoch,

[3]Kaplanis et al. (2018) study migration and life expectancy at the global level from the crowdsourced genealogies, in a dataset that is not cleaned and includes many individuals without geocoding or without date of birth or death. In the paper, the authors do not study or reconstruct fertility, they do not reconstruct series or evaluate selection at the country level, their analysis of human mobility and migratory distance is biased by the inclusion of people moving to the new world who are over-represented in the data, and there is no variable proxying economic development (e.g. urbanization).

Figure 1: Main sample

Note: This Figure displays the geographic coverage of towns of birth of individuals in the main sample in Europe.



Batou and Chèvre (1988). The data is available every 50 years from 1500 to 1850, and places are coded as urban if they reached more than 5,000 inhabitants at the time of the birth of the individual.s It also allows to reconstruct series for human mobility, fertility, and life expectancy by historical urban status, and to evaluate the representativeness of the sample.

Reconstructing fertility Until now, primarily two methods have been employed to study historical fertility. First, family reconstitutions from parish records tried to gather data in a limited number of villages in some countries of pre-industrial Europe, with many issues (see above). Second, aggregated administrative data from historical censuses or population counts. In particular, Coale and Watkins (1986) reconstructed fertility in the past, starting from 1830 in all of 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, on top of problems associated with the aggregate nature of the data, it suffers from many issues (Brown and Guinnane, 2007). In particular, there is no administrative record prior to 1830 and the index does not necessarily capture a decline in the rate of fertility. When it does, it does so with a lag since the marital fertility index is sensible to the age distribution of the population.

Hence, I reconstruct historical rates of fertility from the genealogies. In this paper, fertility is defined as the total number of children ever born, hence it is a measure of completed gross fertility.[4] The main concerns regards the recording of horizontal lineages, which are needed in order to gather full rates of fertility. By definition, only vertical lineages are complete in genealogies and horizontal lineages can be missing. In order to deal with this issue, 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 A2.1.2 plots an individual's family tree in order to preview this restriction. There are 30 parents in the four generations preceding individual i.[5] If at least one was recorded as having strictly more than one child (excluding the fertility rate of i), then i is included in the fertility sample.

Other issues could arise, in particular individuals with fewer children may have less descendants and therefore would be less likely to appear in the data, or it is possible to imagine that only wealthy and educated individuals would reconstruct their family tree. 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 millions descendants even with rates of fertility persistently below replacement level. Second, there is imperfect intergenerational transmission of fertility or wealth, and an individual with low fertility may have children with high fertility while wealthy people today do not necessarily have wealthy ancestors in the distant past, even with a high degree of persistence.[6] Third, it is possible to compare series in the crowdsourced sample to representative series based on census data in order to evaluate if there is an important degree of selection with respect to these issues (see following sub-section).

The resulting fertility sample comprises $759,824$ individuals (Figure A2.1.3 displays the spatial distribution of towns of births in the fertility sample). A few observations, or additional limitations, stand. 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 in order to reconstruct marital fertility indexes or age-specific rates. Marriage is not observed and therefore is proxied by year of birth of the first child. Moreover, 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. In order to deal with this

[4] Because the vertical lineages of the horizontal branches of family tree are not recorded, child mortality is unfortunately under-recorded - therefore net fertility cannot be assessed.

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

[6] Regarding fertility, see Galor and Klemp (2019) or this paper's estimates of the intergenerational transmission of fertility in Section 3.2. Regarding wealth or income, conventional estimates of historical intergenerational wealth or income elasticity are usually in the .2-.5 range (Clark et al., 2014), while Clark and Cummins (2015) find a higher degree of wealth persistence with estimates of roughly .7. Even in the case of the most conservative estimates, wealth or income differences would fade away for individuals born before 1900 - which is what I find in the data.

issue, standard errors are always clustered at the couple level and both spouses included in time series or regression analyses.[7] Finally, Appendix Table A2.1.2 displays summary statistics for the main and the fertility sample.

Weighing In the data, some countries are more represented than others, especially in Northern Europe (see Appendix Table A2.1.1). Hence, in all the analysis at the European level or at the group of country level, observations are re-weighed in order to account for the over-representation of some countries in the crowdsourced sample. The weighing scheme accounts for this by weighing each individual by the total number of observations in the dataset for his or her country divided by that country's population in 1820 (the earliest year for which all countries have an estimate for population, from Bolt and van Zanden (2014)). Unfortunately, it is not possible to use time-varying weighs since historical population estimates are not available yearly, at least until the second half of the twentieth century. Weighs are provided in the data file.

2.2 Measurement and sample selection

Assessing sample selection In order to evaluate the degree of selection in the crowdsourced data, I compare observables in the genealogies to representative country-level series relying mostly on census data. In particular, I use data from census and population count estimates in the Human Mortality Database (2019) for life expectancy and in Coale and Watkins (1986) for marital fertility. Historical rates of urbanization are taken from Bairoch, Batou and Chèvre (1988) and based on estimations of population for cities and estimations of population at the country-level. The purpose of this exercise is not to claim that selection is inexistant, since there are many potential sources of bias, but rather to show that it is limited and that the crowdsourced data can be used to reconstruct population dynamics in the past, when censuses or individual-level data were not available.

I find that the crowdsourced genealogies are a roughly representative sample at the country level in most of Europe from 1675 to 1900 (corresponding to 1750 to 1975 for mortality, which is evaluated at the year of death, and 1700 to 1925 for fertility, evaluated at the year of birth of the first child).[8] Appendix Figure A2.2.1 displays urbanization,

[7] The couple associated with individual i is defined as a pair comprised of individual i and its 'spouse' such that the spouse is the individual with whom i had most of its children.

[8] Note that these dates correspond to an average, and are deduced by comparing the crowdsourced series to representative country-level series. It is also possible to evaluate the extent of selection by noting that, in the seventeenth century (or much later), one can often observe the rate of urbanization or the log distance between place of birth and death declining at some point, likely corresponding to the moment the data becomes representative. Similarly, fertility can be either too low, because of the unavailability of data or because of the poor quality of parish records, or too high, because only wealthy individuals are included in the sample, and then adjusts as selection becomes limited. There is no way to compare these to representative country level series since, often, these do not exist, but the degree of selection can be further assessed with logics and deduction thanks to the Supplementary Figures at the country level in Appendix 1. Finally, some countries can have data that is representative before the

adult mortality, and fertility series in the crowdsourced genealogies and in representative data in Europe. The series coincide almost perfectly, suggesting that the degree of selection is very limited. Only the figure for urbanization displays a substantially different (higher) rate in the crowdsourced sample. However, it is likely due to the fact that the historical data on urbanization from Bairoch, Batou and Chèvre (1988) is an estimate itself and likely under-estimates the rate of urbanization since it does not always account for small cities who are not always in the sample as urban (more details on that in the following sub-section).

Country-level series for **urbanization**, **fertility**, and **adult mortality** are available online as Supplementary Figures in Appendix 1. From these series it is possible to evaluate which countries have the most reliable and representative data.

Table 1: Selection on observables, by region

Note: This Table displays the correlation between the crowdsourced sample and representative data for urbanization, life expectancy, and fertility by regions in Europe. The representative sample series are taken from Bairoch, Batou and Chèvre (1988) for urbanization, Human Mortality Database (2019) for life expectancy at 30, and census data and Coale and Watkins (1986) for fertility. All observations in the crowdsourced genealogies are re-weighed to account for country population and differential selection into the sample, and observations in the representative samples are re-weighed to account for country population. The correlation coefficients are computed only for years with available representative data. Regions are defined in Table A2.1.1. Appendix Figure A2.2.1 displays the crowdsourced and representative series over time for all Europe, while Appendix Figures A2.2.2, A2.2.4, and A2.2.3 display the series region by region.

	Urbanization	**Life expectancy**	**Fertility**
British Isles	0.98	0.96	0.89
Central Europe and Low Countries	0.91	0.88	0.93
Eastern Europe	0.91		
France	0.97	0.83	0.93
Northern Europe	0.93	0.84	0.93
Southern Europe	0.92	0.91	0.97

Finally, Table 1 displays correlation coefficients between the crowdsourced data and representative country-level series, by group of countries, for adult life expectancy, urbanization, and fertility. Appendix Figures A2.2.2, A2.2.4, and A2.2.3 plot the corresponding series by group of country. There is a particularly tight fit, both in correlation and in levels, in most places - although the crowdsourced genealogies in Southern Europe seem to be of poor quality overall due to the small number of observations.

Improvements in measurement Country-level representative data from census is almost never available in early modern Europe. In particular, the figures at the group of country level and the country-level supplementary figures show that series for urbanization, life expectancy, and fertility can be reconstructed when no other data is available. Data on urbanization in Bairoch, Batou and Chèvre (1988) is not always available in the eighteenth century and requires knowledge of the total population in both urban and rural towns, therefore can be significantly biased. Data on fertility is only available after 1831 in Coale

dates listes above (for example, see Blanc (2020) for France).

and Watkins (1986), and is only available in a small subset of countries. Data on adult life expectancy in the Human Mortality Database (2019) is not available in many countries and is almost never available in the early modern period. In what follows, I show some selected improvements that are made possible by the use of crowdsourced genealogies.

In Panel A and B of Figure 2, I display urbanization and adult life expectancy over time in France. In Panel A, the historical rate of urbanization estimated by Bairoch, Batou and Chèvre (1988) is consistently and significantly lower than in the crowdsourced genealogies. In order to compare urbanization in the crowdsourced data to a more representative series than Bairoch data, I leverage population counts from French censuses (BDCassini, 2017). I find that the share of individuals born in a city in the crowdsourced data matches almost perfectly the rate of urbanization in the census. In order to estimate urbanization at the country level, the Bairoch, Batou and Chèvre (1988) estimate require a knowledge of the total number of people in both urban and rural areas and may misestimate the former or the latter. In the crowdsourced genealogies, only the urban status of towns (and the count in both statuses) is required. While the population counts are only available after 1793, the crowdsourced genealogies allow to reconstruct plausible estimates of urbanization a hundred years before. Moreover, the geocoding of places of birth and death allow a granular study of human mobility in terms of likelihood of migration and migratory distance, and it is possible to decompose the human mobility series by urban status of place of birth, as in the following section. In Panel B, I extend the adult life expectancy series before 1825 using estimates of Blayo (1975). Life expectancy in the crowdsourced genealogies follows both series very closely. The genealogies allow not only to have a consistent estimate of life expectancy over time but also to use the individual-level data and geocoding to study the determinants of mortality in the distant past.

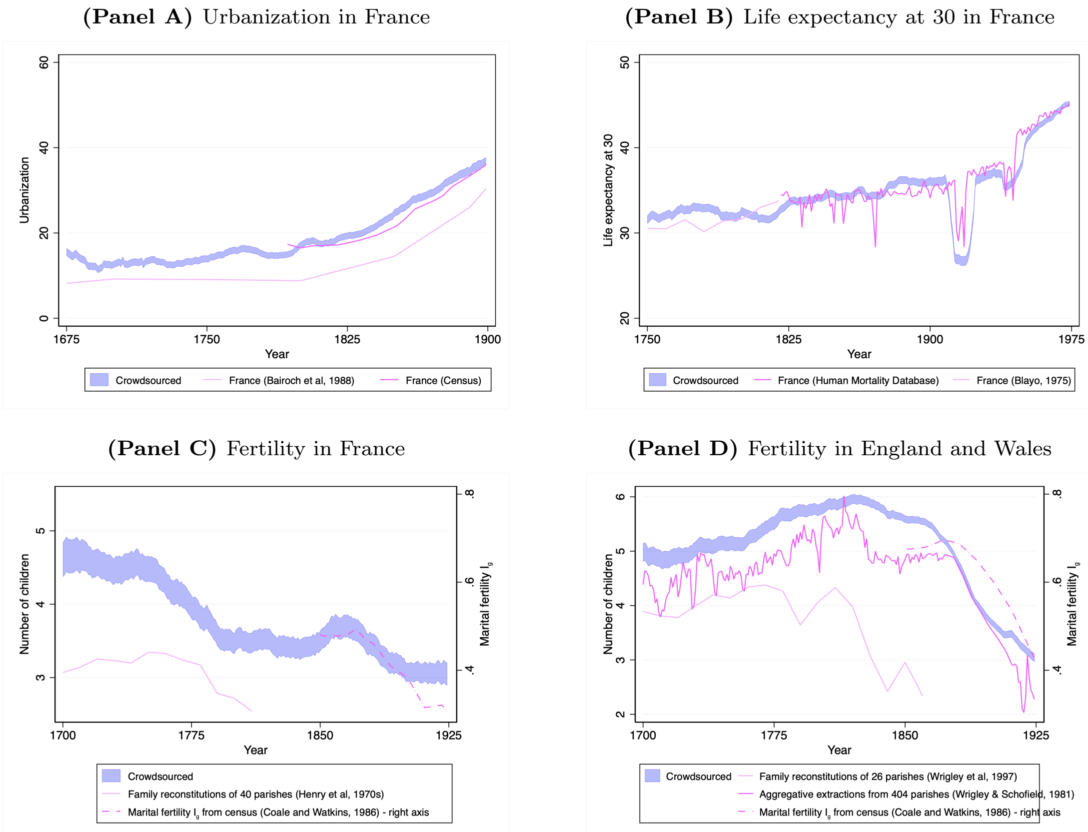
In Panel C and D of Figure 2, I focus on improvements in the measurement of fertility. As already discussed above, the marital fertility index of Coale and Watkins (1986), relying on census data, is only available after 1831 at the very earliest.[9] The only other method available, family reconstitution from parish records, as been extensively developed in France and in England and Wales in order to reconstruct fertility in the eighteenth century. In France, Henry (1972*a*,*b*, 1978); Henry and Houdaille (1973), who invented the method of family reconstitutions, undertook a colossal effort of data collection in 40 villages, with data available before 1810. In England and Wales, Wrigley et al. (1997) used similar data and methods in 26 parishes (before 1860), and later on Wrigley and Schofield (1981) used aggregate statistics at the parish level from a sample of 404 parishes (available throughout) in order to reconstruct representative population dynamics at the country level.

I display total fertility in the crowdsourced genealogies along with the census-based

[9] Paris is only included in the data after 1851 therefore the series for France is only provided after 1851.

Figure 2: Improving the measurement of urbanization, mortality, and fertility

Note: This Figure displays examples of improvements in measurement allowed by the use of crowdsourced genealogies. Panel A displays urbanization in France in the crowdsourced sample and using Bairoch, Batou and Chèvre (1988) and population counts (BD-Cassini, 2017). Panel B displays life expectancy at age 30 in France in the crowdsourced sample and using Human Mortality Database (2019) and Blayo (1975) (before 1825). Panel C displays fertility in France in the crowdsourced sample and using marital fertility from Coale and Watkins (1986) after 1851 (representative data from census) and the total fertility rate in the family reconstitutions of rural parishes by Henry (1972*a*,*b*, 1978); Henry and Houdaille (1973) in the 18th century. Panel D displays fertility in England and Wales in the crowdsourced sample and using marital fertility from Coale and Watkins (1986) after 1851 and the total fertility rate in the family reconstitutions of rural parishes by Wrigley et al. (1997) and aggregative extractions by Wrigley and Schofield (1981).



marital fertility index and the results of family reconstitutions and aggregative extractions in Panel C for France and Panel D for England and Wales. First, one can observe the limits of family reconstitutions discussed in Section 2.1, with the series displaying important variance and correlating very imperfectly with the best available representative data. Second, the aggregative extraction of data from 404 parishes by Wrigley and Schofield (1981) yields results that align well with the marital fertility index in England and Wales. Third, the crowdsourced genealogies allow significant improvements in measurement, with an estimated rate of fertility very similar to both the marital fertility index and the total fertility rate from aggregative extractions, but available systematically over time and space. In particular, the results are striking for France, which did not have any aggregate statistics at the parish level available in the eighteenth century.

These are only some of the improvements made possible by the use of crowdsourced genealogies. I also estimate fertility in the US in Appendix Figure A2.2.5. The US are known for being one of the countries with the earliest decline in fertility, and relying on crowdsourced genealogies allows to reconstruct fertility for individuals born in the US and those who just migrated, with results similar to that in Haines, Hacker and Jaremski (2020). Finally, in Appendix 1, I present country-level series for **human mobility** (**by urban status of town of birth**), **fertility** (**by urban status of town of birth**), **adult mortality** (**by urban status of town of death**). These supplementary figures are available online in separate documents for the sake of readability and simplicity.

3 Main empirical findings

In this section, I present novel empirical findings in the form of stylized facts at the European level (Section 3.1). Some of these are known but not well measured facts, in a period of unprecedented changes. Then, I trace many of these changes to the eighteenth century and to the period surrounding the Age of Enlightenment, using disaggregated data (Section 3.2). In each sub-section, I systematically focus on human mobility, fertility, and adult mortality, with a particular emphasis on fertility since it was not measured before in the eighteenth century. In all, this section does not aim to be a comprehensive study or review of these phenomenons. I only provide some selected findings for each dimension, and Appendix 1 provides a total of 270 figures at the country level.

3.1 Demographic change and development in Europe

Human mobility and rural flight I study patterns of human mobility for individuals who were born and died in Europe in Figure 3.[10] I find large increases in human mobility in the first half of the nineteenth century, summarized in Stylized fact 1.

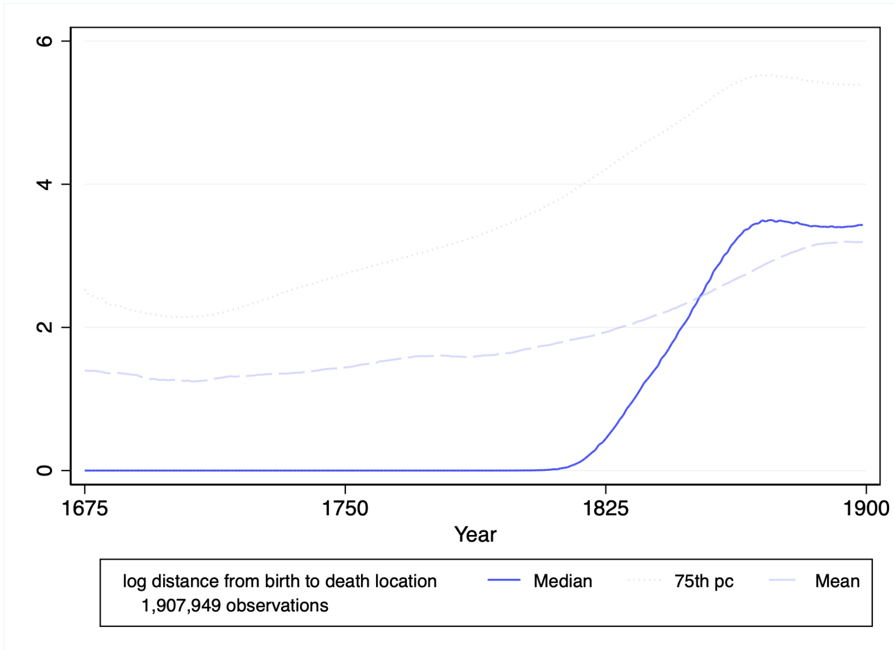
> **Stylized fact # 1.** *In the early nineteenth century, human mobility significantly increased in Europe with a rural flight into urban centers after the Industrial Revolution.*

While the majority of individuals born in Europe before 1820 died in their place of birth, the median individual born in Europe after 1850 died about 50 kilometers away from his or her place of birth. This likely reflects the process of urbanization and a large rural flight into urban centers that affected most of Europe through the second half of the century when the Industrial Revolution set the course for sustained economic growth (Bairoch, Batou and Chèvre, 1988).

[10] Note that the data also allows to study individuals who migrated to the new world.

Figure 3: Human mobility and rural flight in Europe

Note: This Figure displays the median, mean, and 75th percentile of log 1 + distance from place of birth to place of death over time in Europe in the main sample, for individuals who were born and died in Europe. All observations are re-weighed to account for country population and differential selection into the sample.



The urbanization of Europe was a known phenomenon, but the crowdsourced genealogies allow to leverage granular individual level data on human mobility, both regarding migratory distance and the likelihood of migration. It may also allow to reach a better understanding of family structures, through migration for the purpose of marriage, and agglomeration economies, by understanding the determinants of migration in the course of history through characteristics of both places of departure and arrival. In recent research, Clemens (2020) shows that emigration rises with GDP per capita in poor countries, then slows down, and rises again at high levels of GDP per capita. In particular, the genealogies allow to study migration within country, between countries, and at different stages of development.

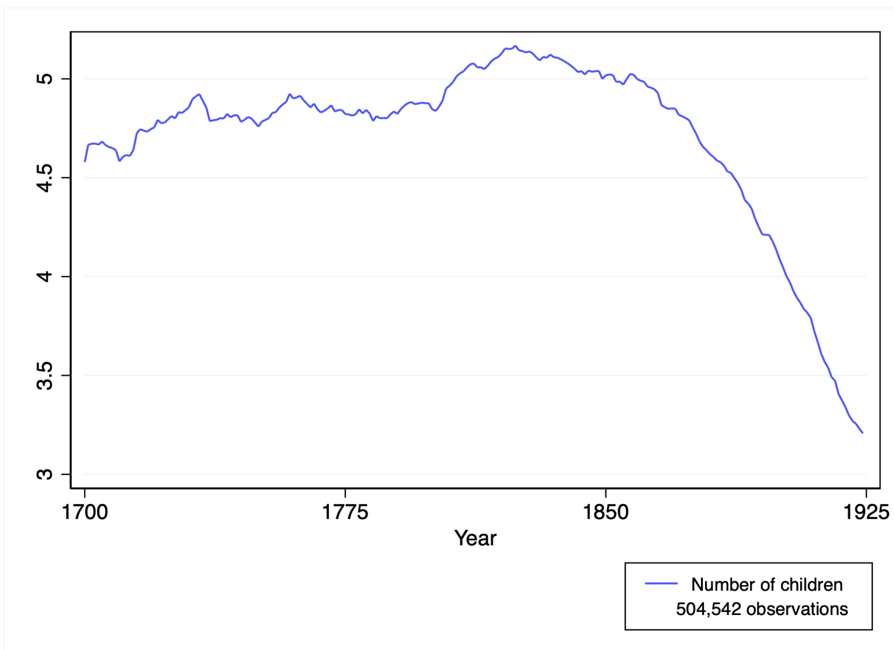
Fertility and demographic change I study fertility in Europe in Figure 4. In the late nineteenth century, most of Europe escaped the Malthusian trap. Before 1800, the average individual gave birth to more than four offsprings. This figure increased to more than five children in the first half of the nineteenth century before declining to about three children in 1925. This is summarized in Stylized fact 2.

Stylized fact # 2. *In the late nineteenth century, fertility significantly declined in Europe.*

The decline in fertility marked one of the most important events in human history (McCloskey, 2008, Chapter 23), the transition to a regime of sustained economic growth through human capital accumulation. In the Malthusian trap, income per capita fluctuates around subsistence level because of the positive relationship between income and fertility. With industrialization and the acceleration of the pace of technological progress, returns to human capital rose and fertility decisions were altered, gradually triggering the onset of the transition (Galor, 2011; Galor and Weil, 2000). Before the onset of the demographic transition, the gradual increase in fertility is explained by a dominating income effect in the quantity quality trade-off. The increase is particularly important in England and Wales, where the Industrial Revolution came first.

Figure 4: Fertility in Europe

Note: This Figure displays the number of children over time for individuals who were born, had their first child, or died in Europe, excluding France, in the fertility sample. All observations are re-weighed to account for country population and differential selection into the sample.



The historical demographic transition and decline in fertility have been extensively studied, in particular by Coale and Watkins (1986), who highlighted the importance of cultural forces, but the data is only available after 1831 and at the regional level. Moreover, it is impossible to study decisions of fertility at the individual level with this data. With family reconstitutions, it is only possible to study the demographic behaviors of individuals within a village, with virtually no spatial variation (e.g. there are only 40 villages in the Louis Henry dataset).

With the demographic transition, an unprecedented increase in life expectancy also affected human societies in the course of development. In Panel A of Appendix Figure A3.1.1 and Stylized fact 3, I show that adult life expectancy significantly increased in the second half of the nineteenth century.

> **Stylized fact # 3.** *In the mid-nineteenth century, Europeans experienced an unprecedented increase in adult life expectancy.*

The increase in adult life expectancy was a known phenomenon, described in the work of Deaton (2013) in particular, and data of very good quality is available at the country level in the Human Mortality Database (2019), but crowdsourced genealogies allow to use data at the individual level with millions of observations in order to evaluate the determinants of the transition to low mortality.[11] Also known but not empirically quantified at the level of Europe, I find that adult mortality significantly increased in urban centers after the Industrial Revolution. Appendix Figure A3.1.1, Panel B, compares the life expectancy of people who died in a city versus those who died in a rural town. After the Industrial Revolution, urban living conditions were detrimental to adult mortality. In England and Wales, which industrialized before other countries, the differential between urban and rural mortality was already important at the start of the period, consistent with the results in Clark and Cummins (2009); Voigtländer and Voth (2013).

3.2 Progress and change in the Age of Enlightenment

In this section, I focus on changes in human mobility and demographic behaviors at a disaggregated level. In particular, I compare urban centers to rural places, I focus on some countries specifically (France and England in particular), and I evaluate the importance of distributional changes in behaviors over time.

Human mobility and rural flight In Europe, human mobility increased in the first half of the nineteenth century. Was the same pattern observable everywhere in Europe, in every country, and did mobility increased more or earlier for people born in rural places? In Figure 5, I show that median human mobility from rural towns increased everywhere around the same time, with the rural flight into urban centers. Moreover, I split the sample by place of birth and show that, in France and in England and Wales, mobility from urban centers increased significantly earlier (Stylized fact 4).

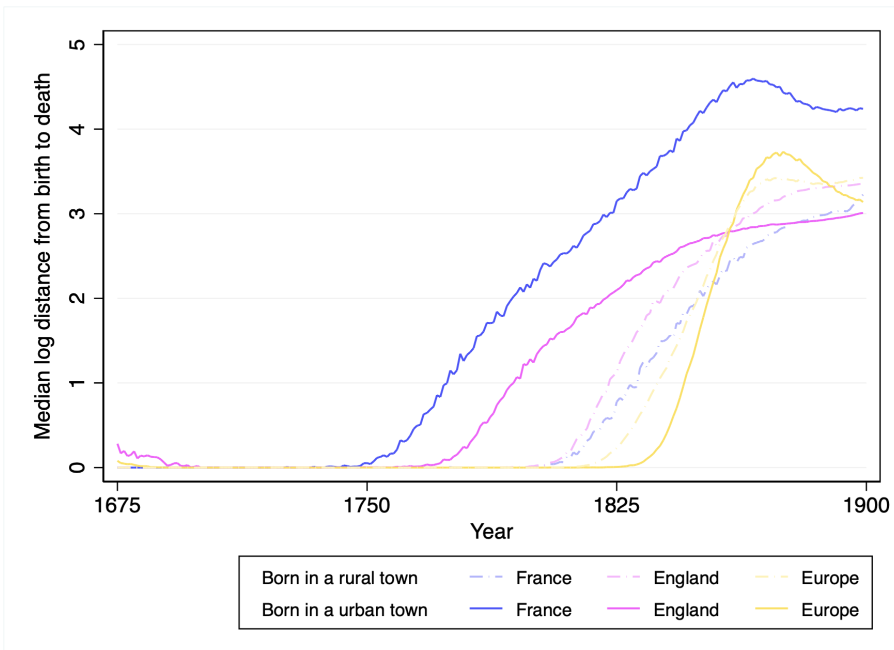
> **Stylized fact # 4.** *In France and in England and Wales, human mobility from urban centers substantially increased in the mid-eighteenth century, 75 years before the rest of Europe.*

[11] Cummins (2017) studies mortality in the distant past, but only of elites, not of ordinary individuals.

What is particularly interesting is that there was only limited variation in the timing of the onset of the rural flight and urbanization of Europe. However, it seems that in Western Europe, in places that were centers of the Enlightenment and of the Industrial Revolution, there was important migration from urban into urban centers for individuals born after 1750. This finding indicates that changes were already taking place as early as in the middle of the eighteenth century, whether those changes were economic, educational, or societal, in the form of reallocation of human capital between cities or accumulation of knowledge in a European market for ideas (Mokyr, 2016).

Figure 5: Human mobility by urban status

Note: This Figure displays the median of the log 1 + distance from place of birth to place of death over time for individuals who were born in France, in England & Wales, or in the rest of Europe, and died in Europe, in the main sample. I break down the series by urban status of the town of birth. All observations in the 'Europe' sample are re-weighed to account for country population and differential selection into the sample.



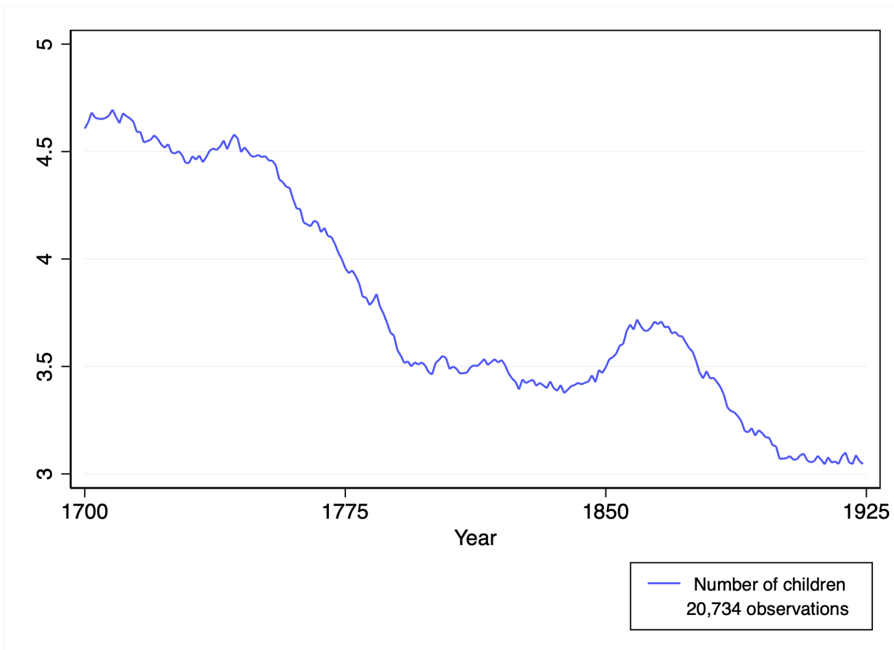
Fertility and demographic change The historical decline in fertility took place in the late nineteenth century. However, in Figure 6, I show that France experienced a much earlier decline in fertility, more than a hundred years before the rest of Europe. This is summarized in Stylized fact 5.

> **Stylized fact # 5.** *In France, the decline in fertility started a hundred years before the rest of Europe, in the mid-eighteenth century.*

The early demographic transition in France was a known phenomenon, partly due to the work of Henry (1972*a*,*b*, 1978); Henry and Houdaille (1973) with family reconstitutions in French villages. In the villages of Louis Henry, Cummins (2012) estimates the onset of the decline to 1776. Yet, as already discussed in Section 2 and with Figure 2, Panel C, the data suffers from important limitations. With crowdsourced genealogies, which leverage data on ordinary individuals all over France, I estimate the onset of the decline in fertility roughly fifteen years earlier that what was previously thought. For the first time, I provide a measure of fertility that is consistent over time and space, and available in the eighteenth century.

Figure 6: Decline in fertility in France

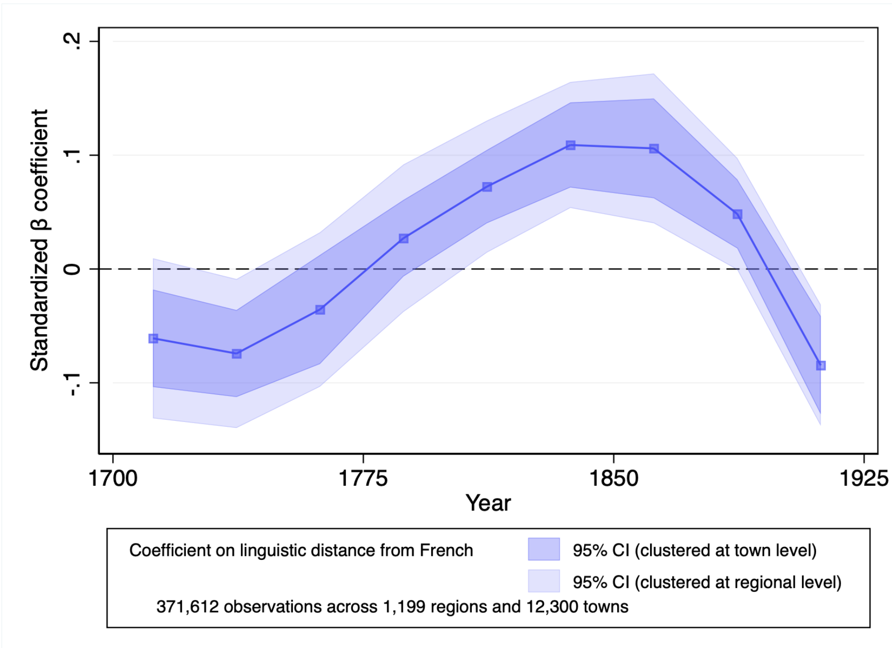
Note: This Figure displays the number of children over time for individuals who were born, had their first child, or died in France, in the fertility sample.



In recent research, Blanc (2020) show that dechristianization was likely the main driving force behind the early decline in fertility in France. At the European level, Spolaore and Wacziarg (2019) show that norms of limited fertility spread along cultural and linguistic lines from France into Europe. I also find this pattern in Figure 7. However, as opposed to Spolaore and Wacziarg (2019) who use data after 1831 from Coale and Watkins (1986), I show the entire process of diffusion. Indeed, I show that norms of limited fertility slowly started spreading right after the onset of the demographic transition in France. The size of the effect gradually increases, and reaches 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 process of diffusion is complete and linguistic distance to France does not have any effect anymore in Europe and, if anything, the effect is slightly negative. In all, these facts highlight the importance of cultural changes along the course of development and in the transition from stagnation to growth.

Figure 7: Diffusion of the decline in fertility: effect of linguistic distance from French

Note: This Figure displays the results of the regression of log fertility on linguistic distance from French, interacted with period fixed effects. All observations are re-weighed to account for country population and differential selection into the sample. The following controls are included: country of birth fixed effects, a quadratic in age at birth of the first child interacted with a male dummy, urbanization status of the town of birth by period fixed effect. Linguistic distance from French is taken from Spolaore and Wacziarg (2019) and is defined as the number of different linguistic nodes to French. Periods are defined over 25 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 regional (NUTS 3) level. This figure was generated with the reghdfe command in Stata (Correia, 2014).



Finally, I study the persistence of fertility behavior across generations over the long run in Europe. The data allows to study for the first time the intergenerational transmission of fertility behavior over the long run, through periods of important changes. Figure 8 plots the intergenerational transmission of fertility over time, from 1700 to 1925. I regress the log fertility of individuals in the fertility sample on the log fertility of parents and plot the estimated coefficients. Importantly, I control for region (NUTS 3) by period fixed effect. Therefore, in each period, all trends and cross-sectional regional level differences are accounted for. Hence, the results do not capture any aggregate-level trends such as the increase in fertility until the late nineteenth century, or the decline in fertility after the onset of the demographic transition.

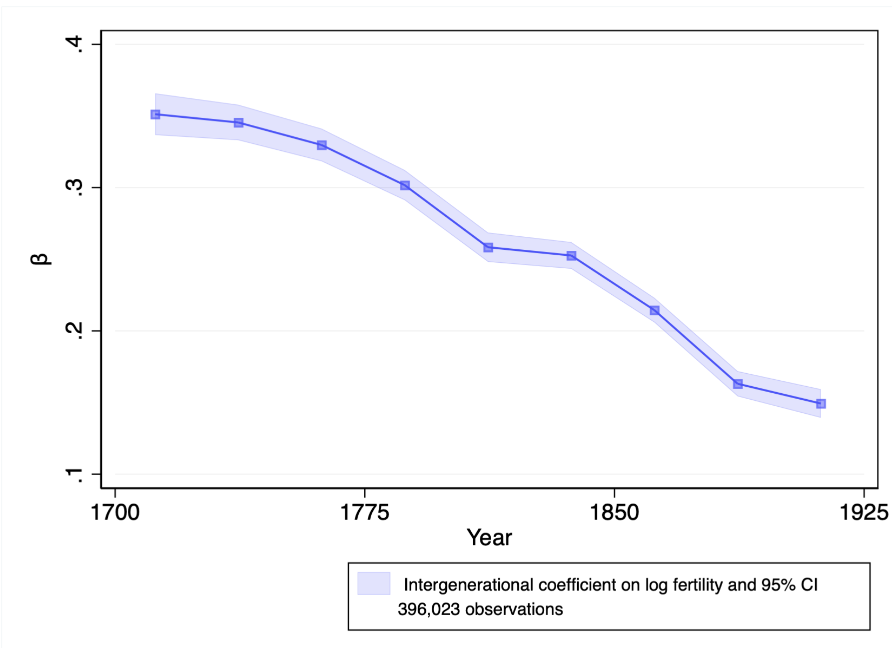
Stylized fact # 6. *In Europe, fertility became increasingly less inter-*

generationally persistent during the Age of Enlightenment, suggesting substantial distributional changes in preferences.

In Stylized fact 6, I find that fertility became increasingly less intergenerationally persistent in Europe over time. In the first half of the eighteenth century fertility behavior was highly persistent, yet the intergenerational association rapidly and significantly decreased in the second half of the century, before weakening further after the aggregate decline in fertility in Europe. The degree of intergenerational persistence is not explained by factors such as aggregate-level changes in fertility. Because of the time-varying regional level fixed effect, the weaker intergenerational associations are only due to distributional changes over time, whether these are changes in the income-fertility gradient, changes in preferences for children, or changes in the distribution of human capital over time.

Figure 8: Intergenerational transmission of fertility behavior

Note: This Figure displays the intergenerational transmission coefficient for fertility in Europe over time, in the fertility sample. All observations are re-weighed to account for country population and differential selection into the sample. This is the coefficient for the log-log regression of an individual's number of children on parental fertility. The following controls are included: region of birth (NUTS 3) by period fixed effects, a male dummy, urbanization status of the town of birth by period fixed effects. Periods are defined over 25 years and coefficients are centered in the middle of each period. Standard errors are two way clustered at the couple level. This figure was generated with the reghdfe command in Stata (Correia, 2014).



That intergenerational persistence was lower following the decline in fertility in Europe in the late nineteenth century is no surprise, and is likely due to changes in the income-fertility gradient or in the cross-sectional distribution of human capital and education. Indeed, after the pace of technological progress accelerated with the Industrial Revolution, incentives to accumulate human capital generated a reversal in the gradient Galor and Weil (2000); Galor and Moav (2002). Similarly, in recent research, Vogl (2020) shows that intergenerational associations weakened after the onset of the demographic transition in a sample of developing countries.[12]

However, I find that intergenerational persistence also broke down before the escape from malthusian stagnation, during the Age of Enlightenment. Since education for the masses did not happen until later in the nineteenth century (Buringh and van Zanden, 2009), and since there was no change in the income-fertility gradient before the escape from the malthusian trap (Ashraf and Galor, 2011), this result suggests that important changes in preferences took hold in Europe as early as in the eighteenth century.[13]

4 Conclusion

This paper uses a novel individual-level dataset crowdsourced from publicly available genealogies to study development and demographic change in the past. I reconstruct fertility from the data and link observations to data on historical urbanization. Relying mostly on censuses, I leverage representative data on urbanization, fertility, and life expectancy in thirty European countries in order to show that there was limited selection into the sample, by comparing the representative data to the series reconstructed from the crowdsourced genealogies.

I generate new measures for known facts and novel empirical findings in Europe for human mobility, fertility behavior, and adult mortality. In particular, I show 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 Age of Enlightenment.

In all, this research focuses on populations rather than places and it should hopefully become a useful ressource for researchers hoping to leverage big data in order to answer questions involving human mobility, fertility, or adult mortality in the past, with millions of ordinary individuals observed in the data.

[12] My results represent significant improvement to the extent that I study completed fertility for both generations, and I study individuals in the distant past in Europe, the first continent in history to have experienced a sustained decline in fertility.

[13] In Appendix Figure A3.2.1, I plot the intergenerational transmission of life expectancy and human mobility over time. Mortality arguably has a very small inheritable component, and is influenced by many factors. Indeed, I find that the intergenerational association is significantly smaller than for fertility throughout, with a coefficient of roughly .1. I estimate that the coefficient for the intergenerational persistence of mobility is about .3. There is not any important changes over time, which is likely due to the fact that the region by period fixed effects account for the rural flight and urbanization of Europe after the 1830s.

References

Abramitzky, Ran, Leah Platt Boustan, and Katherine Eriksson. 2012. "Europe's Tired, Poor, Huddled Masses: Self-Selection and Economic Outcomes in the Age of Mass Migration." *American Economic Review*, 102(5): 1832–56.

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

Ashraf, Quamrul, and Oded Galor. 2011. "Dynamics and Stagnation in the Malthusian Epoch." *American Economic Review*, 101(5): 2003–2041.

Ashraf, Quamrul, and Oded Galor. 2013. "The 'Out of Africa' Hypothesis, Human Genetic Diversity, and Comparative Economic Development." *American Economic Review*, 103(1): 1–46.

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.

BDCassini. 2017. *Geohistoricaldata.* Paris:LaDéHiS, CRH, EHESS.

Blanc, Guillaume. 2020. "Modernization Before Industrialization: Cultural Roots of the Demographic Transition in France." Working paper (hal-02318180v3).

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

Blayo, Yves. 1975. "La mortalité en France de 1740 à 1829." *Population*, 1: 123–142.

Bolt, Jutta, and Jan Luiten van Zanden. 2014. "The Maddison Project: collaborative research on historical national accounts." *The Economic History Review*, 67(3): 627–651.

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.

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

Charpentier, Arthur, and Ewen Gallic. 2020*b*. "Using collaborative genealogy data to study migration: a research note." *The History of the Family*, 25(1): 1–21.

Chatterjee, Shoumitro, and Tom S. Vogl. 2018. "Escaping Malthus: Economic Growth and Fertility Change in the Developing World." *American Economic Review*, 108(6): 1440–67.

Clark, Gregory, and Neil Cummins. 2009. "Urbanization, Mortality, and Fertility in Malthusian England." *American Economic Review*, 99(2): 242–47.

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.

Clemens, Michael. 2020. "The Emigration Life Cycle: How Development Shapes Emigration from Poor Countries." Center for Global Development CGD Working Paper 539, Washington, DC.

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

Correia, Sergio. 2014. "REGHDFE: Stata module to perform linear or instrumental-variable regression absorbing any number of high-dimensional fixed effects." *Statistical Software Components, Boston College Department of Economics.*

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." *The Journal of Economic History*, 77(2): 406–439.

Deaton, A. 2013. *The Great Escape: Health, Wealth, and the Origins of Inequality.* Princeton:Princeton University Press.

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.*

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

Haines, Michael R, J. David Hacker, and Matthew S Jaremski. 2020. “Early Fertility Decline in the United States: Tests of Alternative Hypotheses using New Complete-Count Census Microdata and Enhanced County-Level Data.” National Bureau of Economic Research Working Paper 27668.

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.

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.

Lussier, Alexandre A., and Alon Keinan. 2018. “Crowdsourced Genealogies and Genomes.” *Science*, 360(6385): 153–154.

McCloskey, Deirdre N. 2008. “‘You know, Ernest, the rich are different from you and me’: A Comment on Clark’s A Farewell to Alms.” *European Review of Economic History*, 12(2): 138–148.

Mokyr, Joel. 2016. *A Culture of Growth: The Origins of the Modern Economy.* Princeton:Princeton University Press.

Putterman, Louis, and David N. Weil. 2010. “Post-1500 Population Flows and The Long-Run Determinants of Economic Growth and Inequality.” *The Quarterly Journal of Economics*, 125(4): 1627–1682.

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.

Spolaore, Enrico, and Romain Wacziarg. 2013. "How Deep Are the Roots of Economic Development?" *Journal of Economic Literature*, 51(2): 325–69.

Spolaore, Enrico, and Romain Wacziarg. 2019. "Fertility and Modernity." Working paper.

Vogl, Tom S. 2016. "Differential Fertility, Human Capital, and Development." *The Review of Economic Studies*, 83(1): 365–401.

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

Voigtländer, Nico, and Hans-Joachim Voth. 2012. "Persecution Perpetuated: The Medieval Origins of Anti-Semitic Violence in Nazi Germany." *The Quarterly Journal of Economics*, 127(3): 1339–1392.

Voigtländer, Nico, and Hans-Joachim Voth. 2013. "The Three Horsemen of Riches: Plague, War, and Urbanization in Early Modern Europe." *The Review of Economic Studies*, 80(2): 774–811.

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.

Online appendix to Demographic Change and Development

Appendix 1– Supplementary figures

A1.1 Selection at the country level

The supplementary country-level series are available online for **urbanization**, **fertility**, and **adult mortality**.

A1.2 Main empirical findings at the country level

The supplementary country-level series are available online for **human mobility** (**by urban status of town of birth**), **fertility** (**by urban status of town of birth**), **adult mortality** (**by urban status of town of death**).

Appendix 2– Data, measurement, and sample selection

A2.1 The data: crowdsourced genealogies

Figure A2.1.1: Parish records, Saint-Germain-d'Anxure (1760)

Note: sgasga

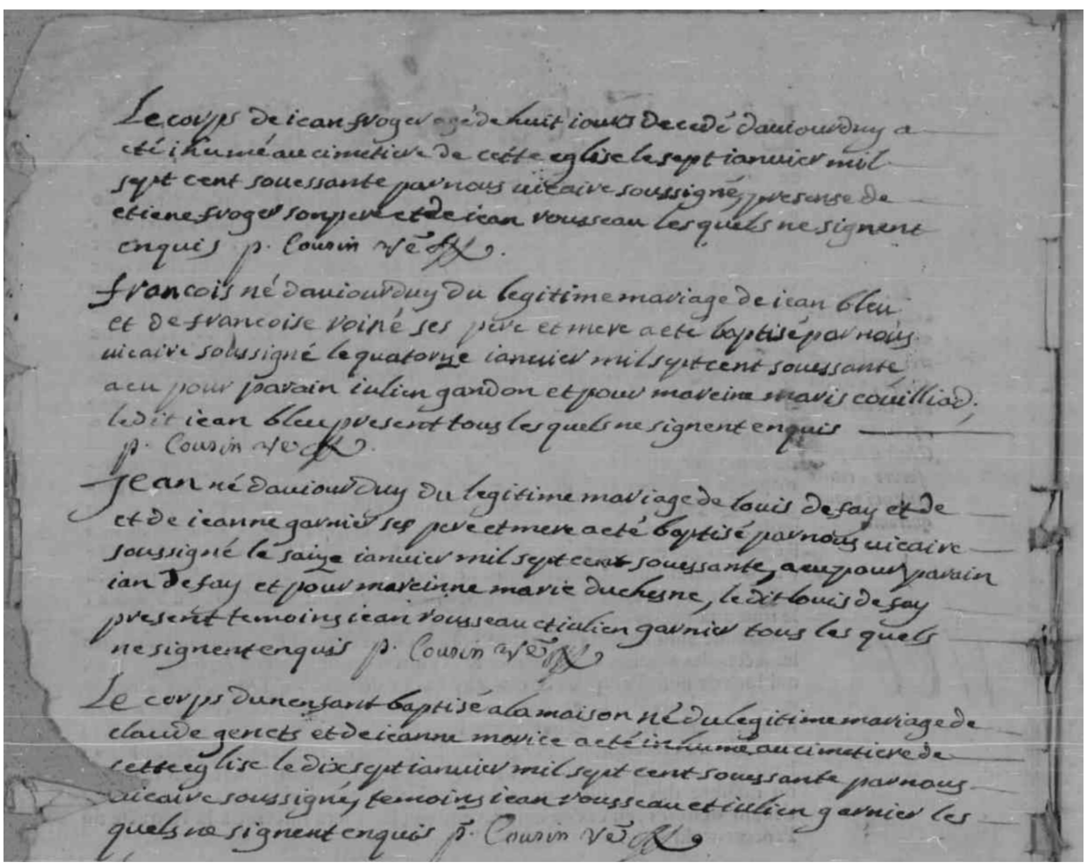

Le corps de iean Froger agé de huit iours decedé d'auiourd'huy a
eté inhumé au cimetiere de cette eglise le sept ianvier mil
sept cent soixante par nous vicaire soussigné presence de
etiene Froger son pere et de iean Rousseau les quels ne signent
enquis p. Cousin vre
françois né d'auiourd'huy du legitime mariage de iean bleu
et de françoise Roiné ses pere et mere a eté baptisé par nous
vicaire soussigné le quatorze ianvier mil sept cent soixante
a eu pour parain iulien gardon et pour mareine marie couillard.
ledit iean bleu present tous les quels ne signent enquis
p. Cousin vre
Jean né d'auiourd'huy du legitime mariage de louis defay et de
et de ieanne garnier ses pere et mere a eté baptisé par nous vicaire
soussigné le saize ianvier mil sept cent soixante, a eu pour parain
ian defay et pour mareinne marie duchesne, ledit louis defay
present temoins iean rousseau et iulien garnier tous les quels
ne signent enquis p. Cousin vre
Le corps d'un enfant baptisé a la maison né du legitime mariage de
claude genets et de ieanne morice a eté inhumé au cimetiere de
cette eglise le dix sept ianvier mil sept cent soixante par nous
vicaire soussigné, temoins iean rousseau et iulien garnier les
quels ne signent enquis p. Cousin vre

Table A2.1.1: Regions

Note: This Table presents our classifications of countries into regions. The number of observations comprises individuals born in a given region.

Country group	**Countries**	**Observations**
British Isles	Ireland, UK	1,778,304
Central Europe and Low Counties	Austria, Belgium, Germany, Hungary, Luxembourg, Netherlands, Switzerland	1,968,031
Eastern Europe	Bulgaria, Czechia, Poland, Romania, Slovakia	510,606
France	France	290,690
Northern Europe	Denmark, Estonia, Finland, Iceland, Lithuania, Latvia, Norway, Sweden	3,174,535
Southern Europe	Croatia, Greece, Italy, Portugal, Macedonia, Slovenia, Spain	484,183

Figure A2.1.2: Lineage and fertility sample

Note: This Figure displays a typical genealogical lineage, for individual i. The tree goes back four generations. The fertility sample includes all individuals with at least one ancestor in the four preceding generations recorded as having had more than one child. Individuals in the fertility sample are therefore more likely to have all their children (horizontal lineage) recorded.

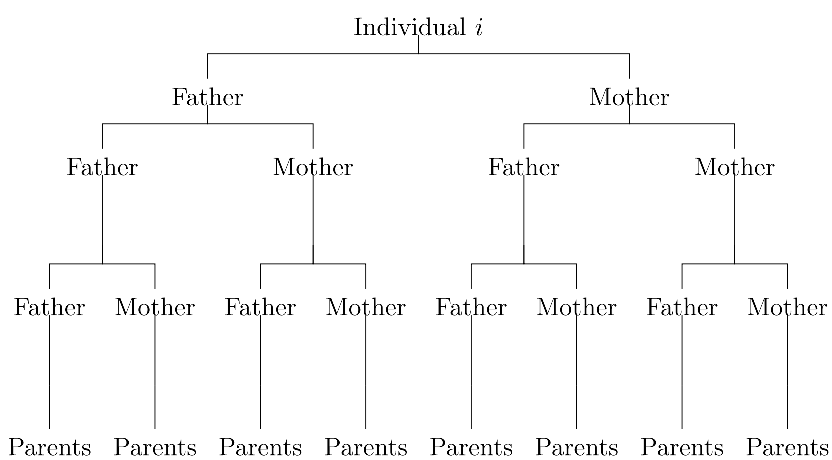


Figure A2.1.3: Fertility sample

Note: This Figure displays the geographic coverage of towns of birth of individuals in the fertility sample in Europe.

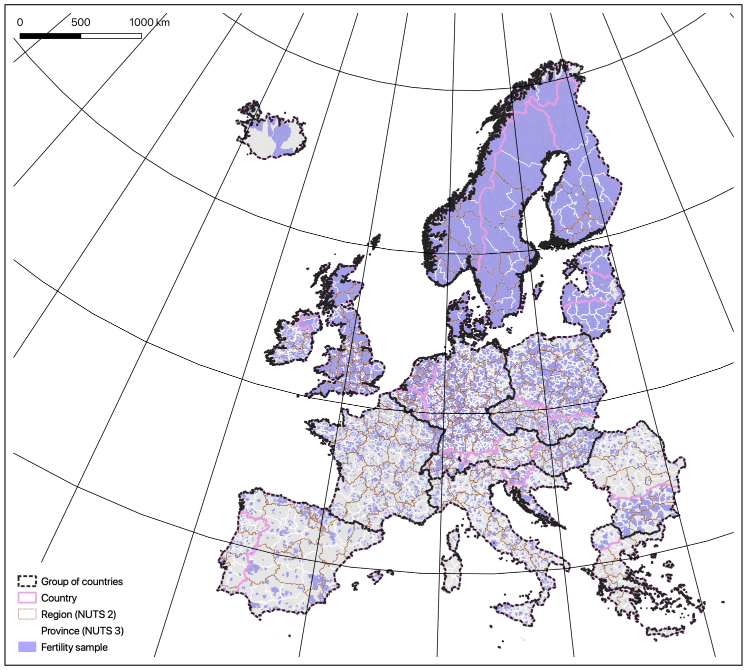


Table A2.1.2: Summary statistics

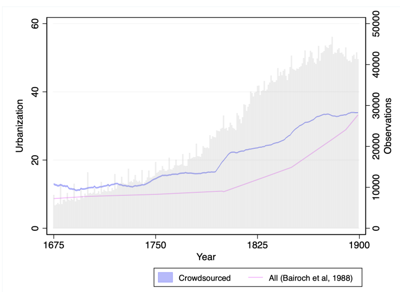
	Mean	St. dev.	Minimum	Maximum	N
Panel A: Main sample					
Main sample	1.00	0.00	1	1	9,426,965
Male	0.54	0.50	0	1	9,423,211
Age at death	65.45	18.31	15	125	5,287,540
log 1 + distance birth and death	3.14	3.33	0	9	3,412,174
Non-missing intergenerational link	0.26	0.44	0	1	9,426,965
Non-missing birth place	0.89	0.32	0	1	9,426,965
Born in Europe	0.99	0.10	0	1	8,343,300
Year of birth	1815.09	98.34	1400	2017	8,669,781
Birth place is urban	0.24	0.43	0	1	7,472,765
Non-missing death place	0.51	0.50	0	1	9,426,965
Died in Europe	0.96	0.20	0	1	4,779,632
Year of death	1874.33	110.74	1400	2017	5,526,520
Death place is urban	0.27	0.44	0	1	3,680,675
Panel B: Fertility sample					

Fertility sample	1.00	0.00	1	1	759,824
Year of birth of first child	1822.28	91.20	1400	2014	634,746
Age at birth of first child	28.46	6.69	15	65	615,259
Spouse is non-missing	0.62	0.49	0	1	759,824
Fertility	4.07	3.12	1	49	759,824
Fertility of parents	5.33	3.31	1	48	759,824

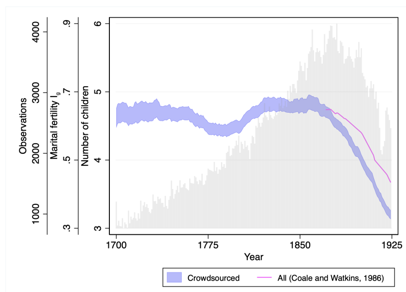
A2.2 Measurement and sample selection

Figure A2.2.1: Selection on observables in Europe, over time

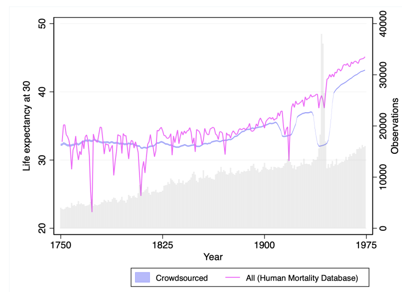
Note: This Figure displays urbanization, life expectancy, and fertility series over time in Europe. In each figure, the corresponding series are displayed from the crowdsourced sample as well as from representative data for Europe (Bairoch, Batou and Chèvre (1988) for urbanization, Human Mortality Database (2019) for life expectancy at 30, census data and Coale and Watkins (1986) for fertility). All observations in the crowdsourced genealogies are re-weighed to account for country population and differential selection into the sample, and observations in the representative samples are re-weighed to account for country population.



(Panel A) Urbanization



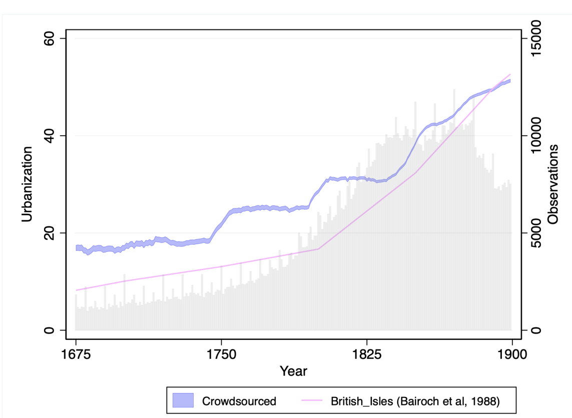
(Panel B) Fertility



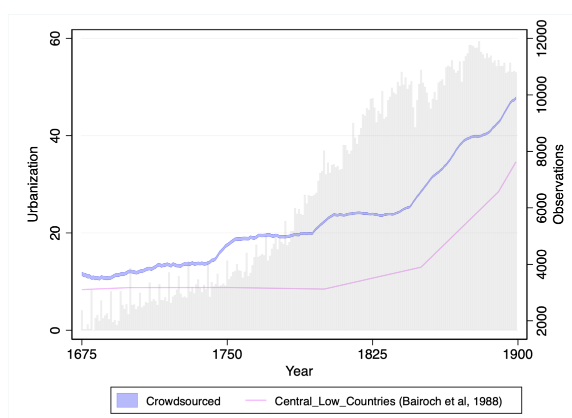
(Panel C) Mortality

Figure A2.2.2: Sample selection at the regional level: urbanization

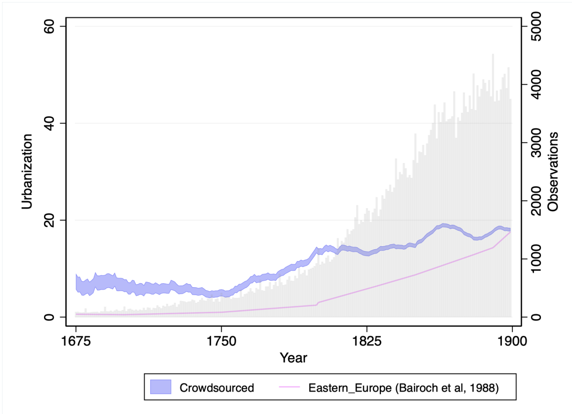
Note: This Figure displays urbanization over time in regions of Europe. In each figure, the series are displayed from the crowdsourced sample as well as from representative data for Europe (Bairoch, Batou and Chèvre, 1988). Regions are defined in Table A2.1.1. All observations in the crowdsourced genealogies are re-weighed to account for country population and differential selection into the sample, and observations in the representative sample are re-weighed to account for country population.



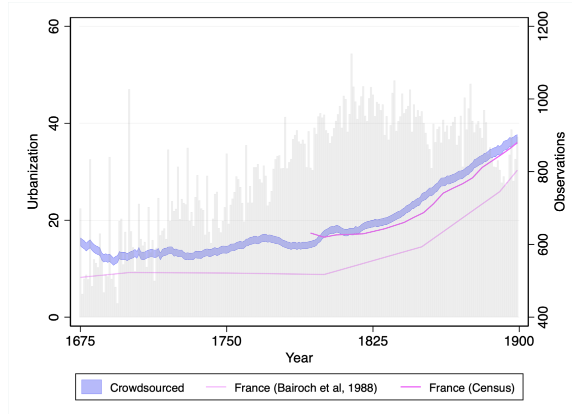
(Panel A) British Isles



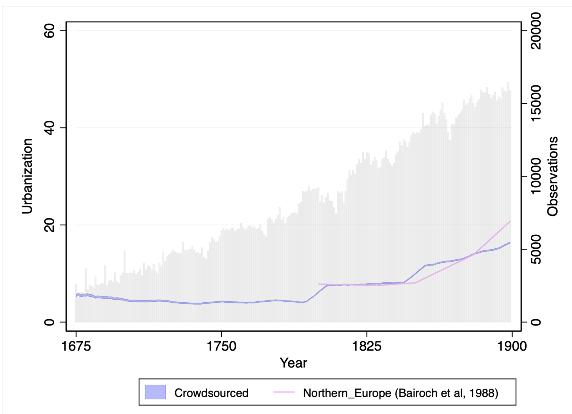
(Panel B) Central Europe and Low Countries



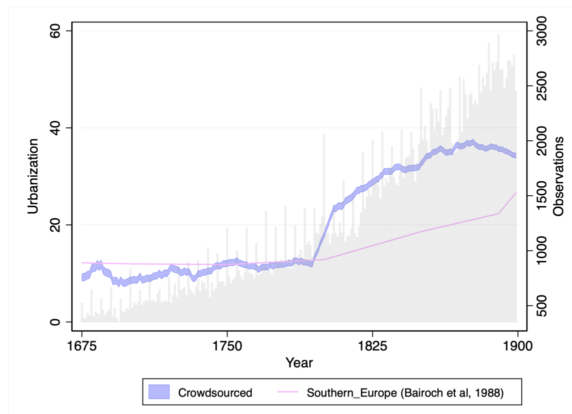
(Panel C) Eastern Europe



(Panel D) France



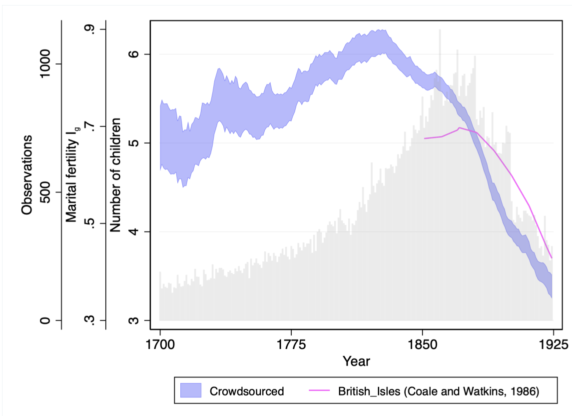
(Panel E) Northern Europe



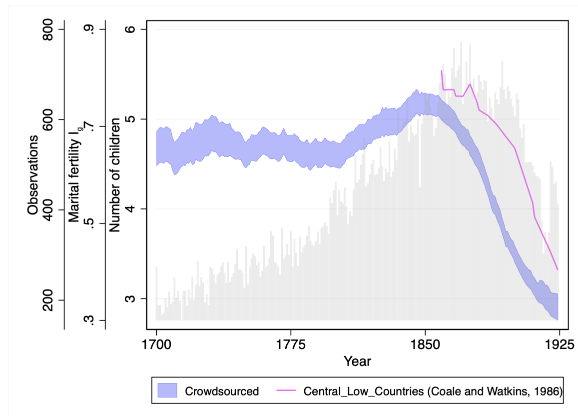
(Panel F) Southern Europe

Figure A2.2.3: Sample selection at the regional level: fertility

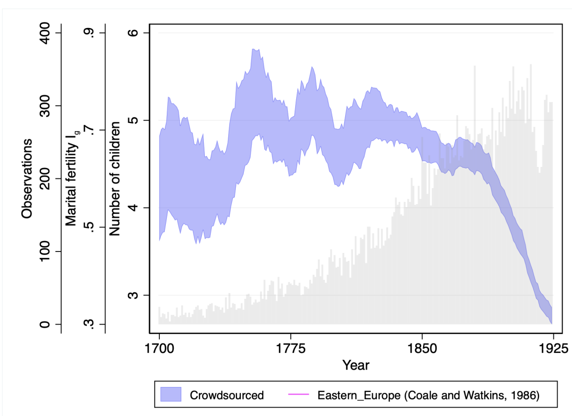
Note: This Figure displays fertility over time in regions of Europe. In each figure, the series are displayed from the crowdsourced sample as well as from representative data for Europe (census data and Coale and Watkins (1986)). Regions are defined in Table A2.1.1. All observations in the crowdsourced genealogies are re-weighed to account for country population and differential selection into the sample, and observations in the representative sample are re-weighed to account for country population.



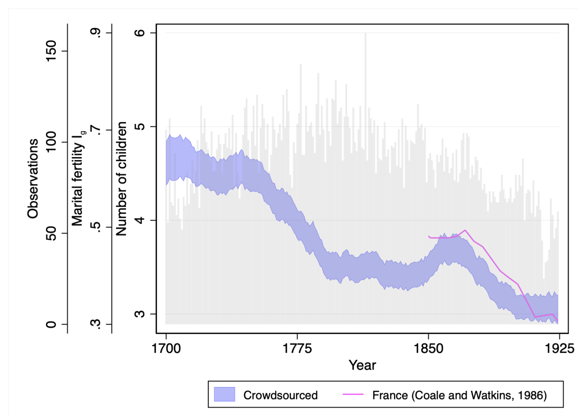
(Panel A) British Isles



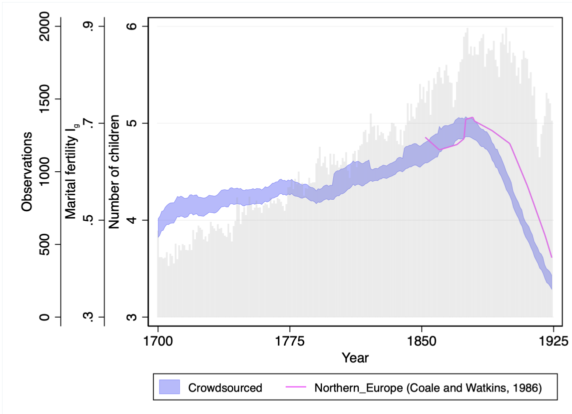
(Panel B) Central Europe and Low Countries



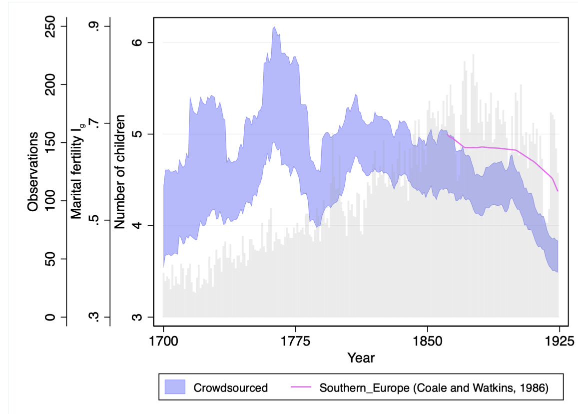
(Panel C) Eastern Europe



(Panel D) France



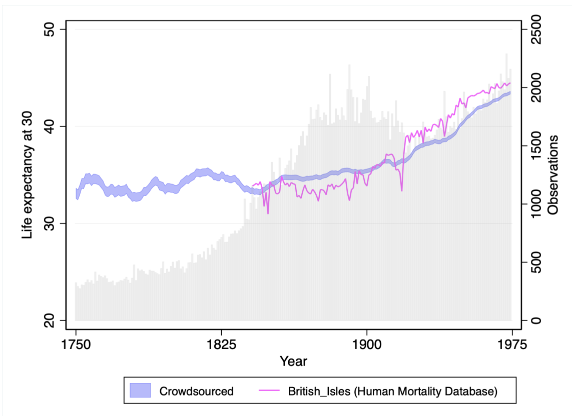
(Panel E) Northern Europe



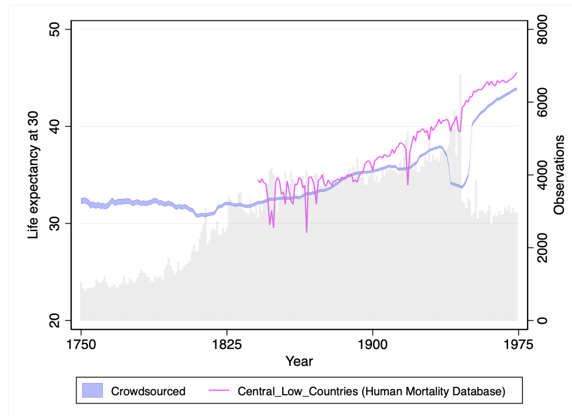
(Panel F) Southern Europe

Figure A2.2.4: Sample selection at the regional level: mortality

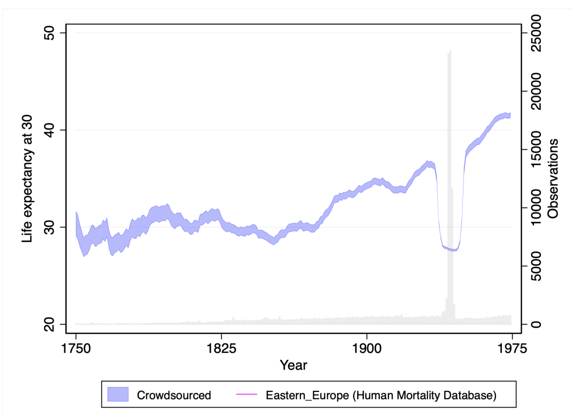
Note: This Figure displays mortality (life expectancy at 30) over time in regions of Europe. In each figure, the series are displayed from the crowdsourced sample as well as from representative data for Europe (census data and Human Mortality Database (2019)). Regions are defined in Table A2.1.1. All observations in the crowdsourced genealogies are re-weighed to account for country population and differential selection into the sample, and observations in the representative sample are re-weighed to account for country population.



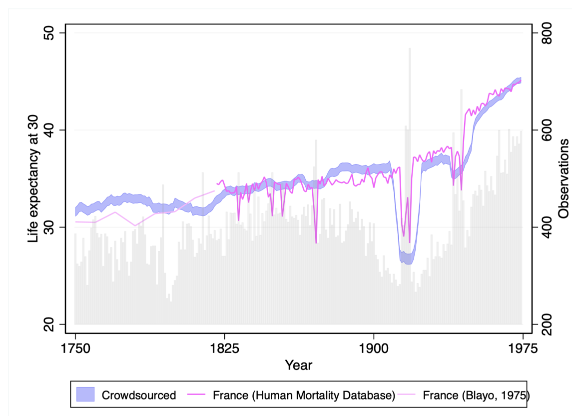
(Panel A) British Isles



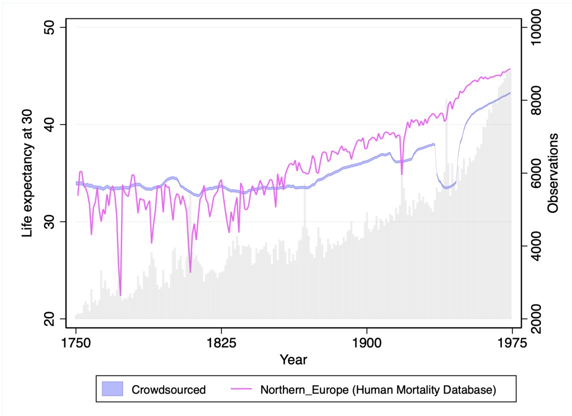
(Panel B) Central Europe and Low Countries



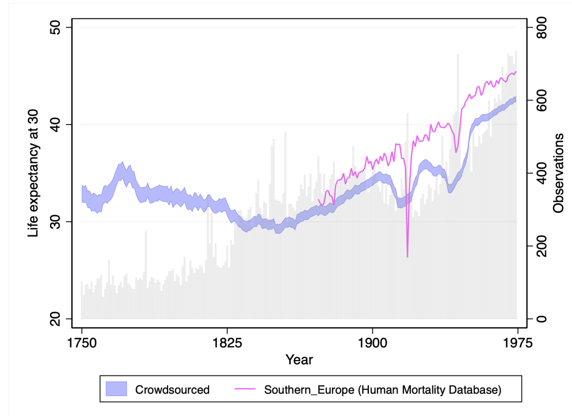
(Panel C) Eastern Europe



(Panel D) France



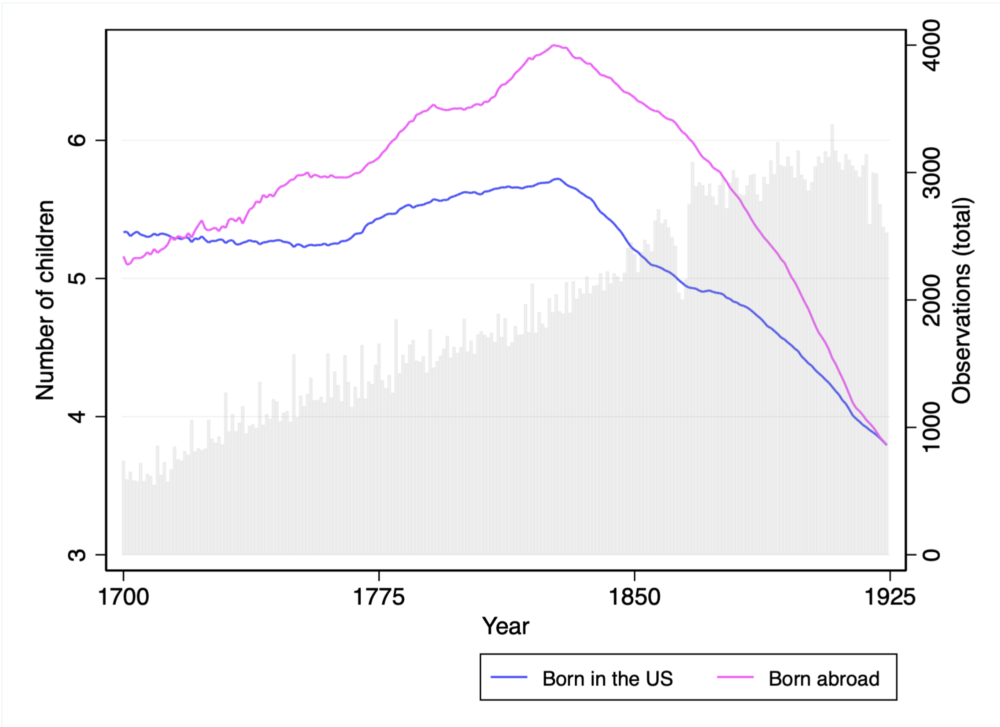
(Panel E) Northern Europe



(Panel F) Southern Europe

Figure A2.2.5: Fertility (US)

Note: This Figure displays the number of children over time for individuals who were born in the US or abroad, in the fertility sample.

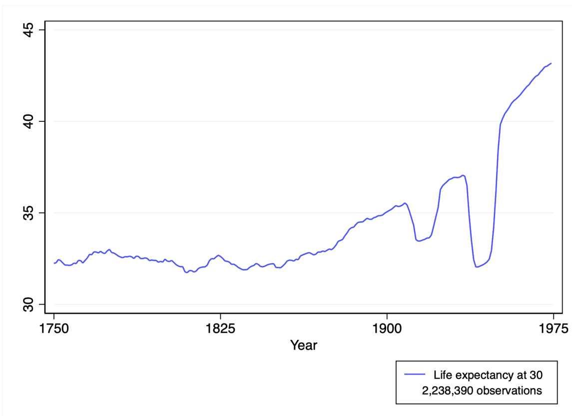


Appendix 3– Main empirical findings

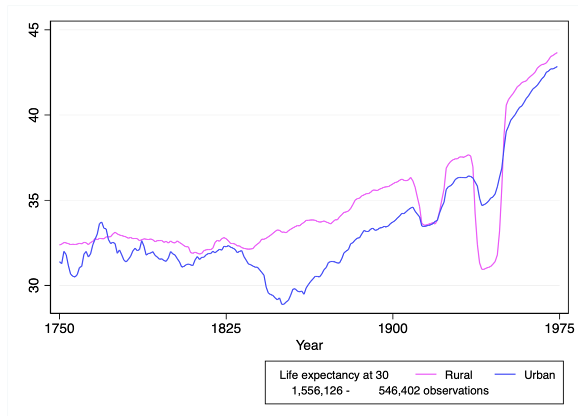
A3.1 Demographic change and development in Europe

Figure A3.1.1: Mortality in Europe

Note: This Figure displays life expectancy at 30 years old over time for individuals who died in Europe, in the main sample. Life expectancy is defined as the average age at death (in excess of 30 years) for individuals who died after 30 years old.



(Panel A) In Europe

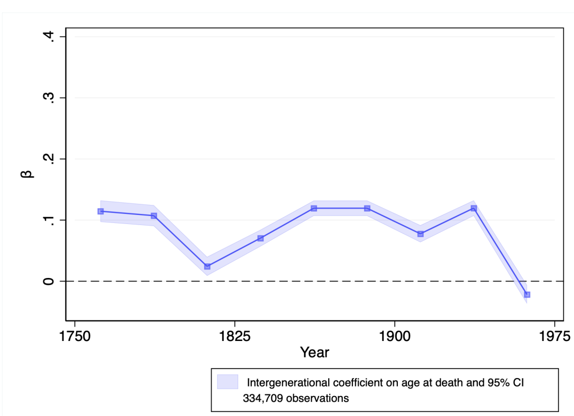


(Panel B) By urban status

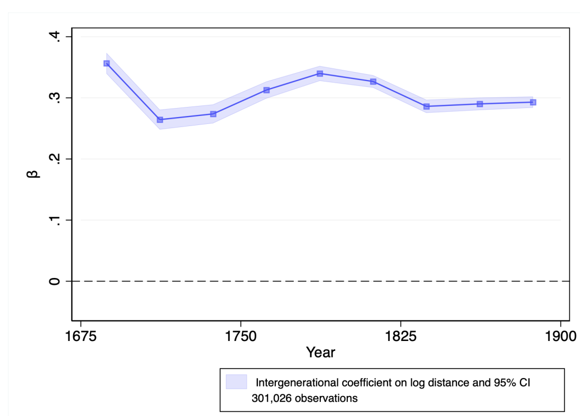
A3.2 Progress and change in the Age of Enlightenment

Figure A3.2.1: Intergenerational transmission of mortality and mobility in Europe

Note: This Figure displays the intergenerational transmission coefficient for adult mortality and human mobility in Europe over time. All observations are re-weighed to account for country population and differential selection into the sample. This is the coefficient for the regression of an individual's age at death or log 1 + distance from place of birth to place of death on its parents'. The following controls are included: region of birth (or death, for mortality) (NUTS 3) by period fixed effects, a male dummy, urbanization status of the town of birth and death by period fixed effects. Periods are defined over 25 years and coefficients are centered in the middle of each period. This figure was generated with the reghdfe command in Stata (Correia, 2014).



(Panel A) Mortality



(Panel B) Mobility