Demographic Transitions Across Time and Space*

Matthew J. Delventhal†

Jesús Fernández-Villaverde‡

Nezih Guner§

January 20, 2024

Abstract

The demographic transition –the move from a high fertility/high mortality regime into a low fertility/low mortality regime– is one of the most fundamental transformations that countries undertake. To study demographic transitions across time and space, we compile a data set of birth and death rates for 186 countries spanning more than 250 years. We document that (i) a demographic transition has been completed or is ongoing in nearly every country; (ii) the speed of transition has increased over time; and (iii) having more neighbors that have started the transition is associated with a higher probability of a country beginning its own transition. To account for these observations, we build a quantitative model in which parents choose child quantity and educational quality. Countries differ in geographic location, and improved production and medical technologies diffuse outward from Great Britain, the technological leader. Our framework replicates well the timing and increasing speed of transitions. It also produces a strong correlation between the speeds of fertility transition and increases in schooling similar to the one in the data.

Keywords: Demographic transition, skill-biased technological change, diffusion.

JEL codes: J13, N3, O11, O33, O40

*We would like to thank for their comments seminar participants at Arizona State University, Brown University, CEPR Macro Group Workshop (2020), EIEF (Rome), Statistical Institute-Delhi, NBER SI 2019, NBER Growth Meetings (2018), and U. of Mannheim. Florian Fiaux, Claudio Luccioletti, and Yongkun Yin provided excellent research assistance. The companion web page for the paper with our database is https://sites.google.com/view/demographic-transitions.

†The Robert Day School of Economics and Finance, Claremont McKenna College, 500 E. 9th Street, Claremont, CA 91711 mdelventhal@cmc.edu

‡Department of Economics, 133 S 36th St, Philadelphia, PA 19104, University of Pennsylvania, Philadelphia, PA 19104-6297 jesusfv@econ.upenn.edu

§CEMFI, Calle Casado del Alisal 5, 28014 Madrid, Spain. nezih.guner@cemfi.es

1 Introduction

The demographic transition, a critical global phenomenon, involves the shift from high fertility and mortality rates to low levels. It began in late 18th-century Northern Europe and has since affected every country. Initially, mortality rates decline, followed by a decrease in fertility a few decades later, resulting in a surge in global population growth. This transition has elevated world population growth rates from historical lows to over 2% in the early 1960s, but growth has since declined to below 1% (Figure 1). According to U.N. population projections, the world population will peak in 2086 and start falling afterward.[1]

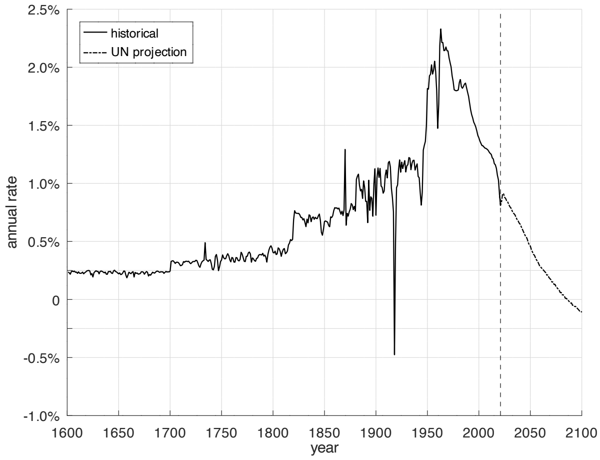


Figure 1: World population growth, 1600-2100

Note:1600-2016: authors' calculations; see Appendix A. 2017-2100: estimates and medium scenario of UN 2022 World Population projections.

Another way to look at this transformation is to consider the total number of children born in the world. After increasing rapidly throughout most of the 20th century, total world births barely increased from 1980 to 2012 and have been falling since then. With current world fertility at 2.31 children per woman and naively extrapolating the drop in fertility from the last 25 years, the world will fall below the conventional replacement rate of 2.1 children per woman before 2030.[2]

How can we account for this momentous transformation of the world's population? Have demographic transitions changed in speed and shape over time? Were the transitions of the 18^{th} and 19^{th} centuries in Western Europe similar to the ongoing transitions in Sub-Saharan Africa, or are these two inherently different? What mechanisms explain the timing of the demographic transition? And what is the link between demographic transformation and economic development?

[1]Section 3 indicates that, based on our estimates, virtually every country has initiated or completed the mortality transition, with Chad being the sole exception regarding the fertility transition. Even in Chad, birth rates have declined consistently over the past 25 years, and with further data, our econometric approach will probably confirm the onset of the fertility transition in Chad as well.

[2]The replacement rate is higher than 2.1 when the sex ratio at birth is biased against women, as it occurs in many Asian countries, and mortality among women before completing their fertility age is relatively elevated, as it happens in Africa. Thus, humanity might already be at or below replacement level.

While the literature on the demographic transition is enormous, these questions have not been fully answered. Our paper advances this line of inquiry in two ways: first, we document demographic patterns across a broad array of countries in a data set stretching back more than 250 years. Second, we propose a simple quantitative model that can account for the emergence of these cross-country patterns during the global demographic transition.

Our first step is to put together and analyze a data set of crude death rates (CDR) and crude birth rates (CBR) for 186 countries that spans more than 250 years. Following the textbook description of the demographic transition, we propose an econometric method to estimate, for each country in our sample: i) initial (pre-transition) levels of the CDR and CBR, ii) the start dates of the mortality and fertility transitions, iii) the end dates of the mortality and fertility transitions, and iv) final (post-transition) levels of the CDR and CBR. This procedure allows us to estimate the length and the speed of each transition.

Looking at demographic transitions across time and space, we show that: i) the start dates of the CDR transitions are more dispersed over time than the start dates of the CBR transitions, ii) transitions are becoming faster, iii) the average level of GDP per capita at the start of a transition is roughly constant, and iv) demographic transitions are contagious: an important predictor of a country's transition is the prior transition of other countries that are "close" to it geographically or culturally. To the best of our understanding, we are the first to systematically document the first three facts, which are central aspects of the demographic transformation of humanity.[3] The fourth fact has been documented within specific regions (e.g., 19$^{\text{th}}$-century Europe in Coale and Watkins, 1986, and Spolaore and Wacziarg, 2021), but we are the first to report it for a comprehensive set of countries across the planet.

Next, we build a quantitative general equilibrium model that can account for these facts across all countries using a single piece of country-specific information–each country's great circle distance from Great Britain. Again, we are the first to conduct such an exercise in a cross-country context. We consider a model with four overlapping generations and multiple locations, each representing one country. Each location is populated by a representative household that decides how many children to have and how much to invest in their education. Having and educating children is costly for parents. There are two production technologies: ancient and modern. Both technologies use unskilled labor, skilled labor, and land, but modern technology uses skilled labor more intensively. Survival rates at each age depend on an aggregate medical technology, which improves over time.

The economy is initially in a Malthusian steady state with high and constant fertility and mortality. The economy does not grow since the total factor productivity (TFP) in both the ancient and modern technologies, as well as the level of medical knowledge, is constant. At some moment, the TFP in both sectors and medical technology start growing. This occurs first in the frontier country, Great Britain in our calibration, and then diffuses

[3] Other researchers have noted that fertility and mortality declines in some developing countries happened more quickly than in more developed ones. See, e.g., Lee (2003).

slowly to other locations. Higher demand for skilled labor and a rising skill premium make investment in children more valuable, and parents react by reducing the number of children but educating them better.

We bring our framework to the data in two steps. First, we calibrate the model economy to replicate the demographic and economic transition in Great Britain, where population growth together with industrialization and urbanization occurred first (Morland, 2019). Next, we show that a diffusion mechanism where technological change travels from Great Britain to the rest of the world depending on geographic distance can generate sequences of demographic transitions, each happening faster than the previous one, exactly as we observe in the data.

As a country embarks on its demographic transition, the educational attainment of its population increases. Thus, together with the demographic transition, the world also experiences an economic transition, and global GDP per capita increases by more than tenfold between the middle of the 19$^{\text{th}}$ century and today. Inequality across countries in GDP per capita first increases sharply until the 1980s and then declines as more and more countries experience demographic and economic transformation, in line with the data. Faster transitions are associated with faster increases in educational attainment, and the model can account for the changes in education levels across countries and their correlation with the fall in fertility.

Understanding the relationship between income and population is one of the oldest challenges in economics, going back to Malthus (1993). Becker (1960) and Becker and Lewis (1973) postulated a trade-off between quantity and quality of children that can account for the co-emergence of growing per capita incomes and low fertility. The interest in this mechanism was revived with the presentation of a dynastic model of fertility in Barro and Becker (1989) and Becker and Barro (1988).

Building on this work, Becker et al. (1990), Lucas (1988, 2002), Jones (2001), and, in particular, Galor and Weil (1996, 1999, 2000) present models that try to capture the historical evolution of population and output. Fernández-Villaverde (2001), Greenwood and Seshadri (2002), Kalemli-Ozcan (2003), Tamura (2006), Doepke (2017), and Bar and Leukhina (2010) present quantitative versions of these models that can account for historical evidence on demographic transitions for specific countries. Jones et al. (2010) and Doepke et al. (2022) provide excellent reviews.

Within this literature, De Silva and Tenreyro (2020) and Cavalcanti et al. (2021) emphasize the role of contraception and population planning programs, Vogl (2016) highlights a historical shift, from positive to negative, in the relationship between family size and educational attainment of children, and Manuelli and Seshadri (2009) explore productivity and taxes to explain the differences in fertility between Europe and the U.S. Córdoba et al. (2020) decompose the 2013 cross-section of fertility and education differences, and find that TFP accounts for a large part of the cross-county variation, while differences in public funding of education also play an important role. Cervellati and Sunde (2015) calibrates a growth model with fertility and education choices to Sweden, and then explores whether exogenous differences in initial mortality rates can account for differences between Sweden's demographic evolution and the evolution of other countries post-1960. Cervellati et al. (2023) build a unified growth model that incorporates the endogenous accumulation of physical capital, population, human capital, and technology and studies how the timing of demographic transitions matters for convergence patterns in the data. Finally, Vogl (2020) argues that intergenerational associations of fertility vary over the fertility transition due to a reversal of fertility differences by skill.

Following the influential Princeton study (Coale and Watkins, 1986), researchers have emphasized the importance of cultural factors in the diffusion of fertility behavior across time and countries. Spolaore and Wacziarg (2021) document that linguistic distance from France was associated with the onset of the fertility transition in Europe. Focusing on the fertility and education transition in France during the 19th century De la Croix and Perrin (2018) show that a simple quality-quantity model can explain some of the variations in fertility across time and French counties, but cultural barriers are likely to interact with economic incentives. Building on these contributions, our paper is the first to detect empirically a "demographic contagion" effect at a global scale and to investigate it within a quantitative framework.

Our paper is also related to studies in demography and economics that provide empirical analyses of demographic transitions across countries. Focusing on transitions since 1950, Casterline (2001) highlights the importance of cross-country differences in the speed of transitions. Dyson and Murphy (1985) show that at the start of fertility transitions, there is an initial increase in fertility followed by a sustained decline. Reher (2004) looks at a broad panel of countries and compares earlier with later demographic transitions, with a particular focus on the role of mortality in driving fertility changes. Murtin (2013) finds evidence for a robust effect of early childhood education on fertility decline.

Lastly, by proposing technology diffusion as a mechanism linking the process of the demographic transition across countries, our analysis borrows from the influential work on technology diffusion by Lucas (2009), Comín and Hobijn (2010), and Comín and Mestieri (2018). Along these lines, Hejkal et al. (2022), model the mortality transition as a diffusion process, where adoption becomes cheaper as more individuals acquire the modern technology. Our emphasis on the role of a country's neighbors is shared by Buera et al. (2011), who model how neighbors' past experiences influence the implementation of market-oriented policies. Fogli and Veldkamp (2011) and Fernández (2013) study how information diffusion impact female labor force participation during the 20th century.

The rest of the paper is organized as follows. Section 2 presents our methodology for measuring demographic transitions. Sections 3 and 4 describe the data and empirical results. Section 5 introduces our model, which is analyzed quantitatively in Sections 6 and 7. Section 8 concludes. Several appendices add further details.

2 Measuring demographic transitions

This section proposes a methodology for documenting the shape and speed of demographic transitions across time and space. In a textbook demographic transition, mortality and fertility go through three stages (Chesnais, 1992): In stage 1, both the crude birth rate (CBR) and crude death rate (CDR) are high and stationary.[4] In stage 2, they decline. In stage 3, both the CBR and CDR stop falling and become stationary at a lower level.

Taking this three-stage demographic transition as a benchmark, we fit it to available data for each country by estimating: i) an initial (pre-transition) average level of the CBR and CDR; ii) the start date of the decline of each rate; iii) the end date of the decline of each rate; and iv) a final (post-transition) average level of the CBR and CDR. We do not require that, either before or after the demographic transition, the average level of the CBR and CDR be equal to each other. Pre- or post-transition, the population of a country may be growing (the average CBR is higher than the average CDR) or declining (the average CBR is lower than the average CDR). We also do not impose a relative ordering of the start dates of CBR and CDR declines: the CDR may begin declining before the CBR, as in a typical textbook configuration, or the CBR may decline first.[5]

2.1 Econometric model

Consider a dependent variable y_t observed for $t \in \{1, ..., T\}$. We assume that y_t is a linear function of a vector x_t of k regressors and a residual. Furthermore, suppose that the relationship between y_t and x_t evolves over time and can be broken into S distinct stages $s \in \{1, 2, ..., S\}$ connecting $S + 1$ distinct endpoints represented by $\{\tau_1, \tau_2, ..., \tau_{S+1}\}$, such that $\tau_1 = 1$, $\tau_{S+1} = T$, $\tau_s \in \{2, ..., T-1\}$ for $s \in \{2, ..., S\}$, and $\tau_s < \tau_{s+1}$ for all $s \in \{1, ..., S\}$.

At each endpoint τ_s, the dependent variable is defined by:

$$y_{\tau_s} = x'_{\tau_s}\alpha_s + \sigma_s \nu_{s,\tau_s},$$

where $\nu_{s,t} \sim \mathcal{N}(0, 1)$ for all s, α_s is a $k \times 1$ vector of regression coefficients, and σ_s is a scalar that determines the volatility of the residual at point τ_s.

Now suppose that in each stage s, i.e., when $\tau_s < t < \tau_{s+1}$, the dependent variable is defined by:

$$y_t = x'_t f_s(\alpha_s, \alpha_{s+1}, t) + \varepsilon'_{s,t} g_s(\sigma_s, \sigma_{s+1}, t),$$

[4] The CBR is the number of live births per year per 1,000 in a population. The CDR is the number of deaths per year per 1,000 in a population.

[5] We focus on the CBR and the CDR instead of statistics such as the total fertility rate (TFR) or life expectancy because CBRs and CDRs are more reliably measured in the data: a researcher only needs an accurate count of births, deaths, and total population. Thus, CBRs and CDRs are available for long periods and are comparable across many different countries. In contrast, estimating current TFR or life expectancy requires both additional data, such as exact current age-specific fertility rates, and additional assumptions, in particular about mortality rates. These additional data are not available or are imprecisely measured for most countries during the pre-modern era and many countries today.

where $\varepsilon_{s,t} \sim \mathcal{N}(0,1)$ for all s, and f_s and g_s are continuous functions $f_s : \mathbb{R}^k \times \mathbb{R}^k \times \mathbb{R} \to \mathbb{R}^k$, $g_s : \mathbb{R}^+ \times \mathbb{R}^+ \times \mathbb{R} \to \mathbb{R}^+$ such that $f_s(\alpha_s, \alpha_{s+1}, \tau_s) = \alpha_s$, $f_s(\alpha_s, \alpha_{s+1}, \tau_{s+1}) = \alpha_{s+1}$, $g_s(\sigma_s, \sigma_{s+1}, \tau_s) = \sigma_s$, and $g_s(\sigma_s, \sigma_{s+1}, \tau_{s+1}) = \sigma_{s+1}$.

While it is possible to analyze the more general class of transition functions we just defined, we restrict our attention to the simplest case where f_s and g_s are linear transitions with respect to time between the parameters at τ_s and τ_{s+1} for all $s \in \{1, ..., S\}$, i.e.,

$$f_s(\alpha_s, \alpha_{s+1}, t) = \frac{(\tau_{s+1} - t)\alpha_s + (t - \tau_s)\alpha_{s+1}}{\tau_{s+1} - \tau_s} \quad \text{and} \quad g_s(\sigma_s, \sigma_{s+1}, t) = \frac{(\tau_{s+1} - t)\sigma_s + (t - \tau_s)\sigma_{s+1}}{\tau_{s+1} - \tau_s}. \tag{1}$$

To apply this framework to demographic transitions, suppose that the dependent variable y_t is either the CBR or the CDR for a particular country and that $S = 3$ (i.e., there is a stage where y_t is stationary, another stage where it is declining, and a final stage where it is stationary again). Also, we are interested in transitions between two stable regimes (high vs. low CBR and CDR), so assume that $\alpha_s = \alpha_{s+1}$, $\sigma_s = \sigma_{s+1}$, and $\nu_{st} = \nu_{s+1,t} = \varepsilon_{st}$ for $s \in \{1, 3\}$.

Substituting in for f_1 and g_1 as given by equation (1), we can write y_t as:

$$\begin{aligned} y_t \;=\; & d_{1t}[x_t'\alpha_1 + \varepsilon_{1t}\sigma_1] + d_{2t}x_t'\frac{1}{\tau_3 - \tau_2}\left[(\tau_3 - t)\alpha_1 + (t - \tau_2)\alpha_3\right] \\ & + d_{2t}\frac{1}{\tau_3 - \tau_2}\left[(\tau_3 - t)\sigma_1 + (t - \tau_2)\sigma_3\right]\varepsilon_{2t} + d_{3t}[x_t'\alpha_3 + \varepsilon_{3t}\sigma_3], \end{aligned} \tag{2}$$

where $\{d_{st}\}_{s=1}^3$ are indicator functions given by $d_{1t} = 1\{t \leq \tau_2\}$, $d_{2t} = 1\{\tau_2 < t < \tau_3\}$, and $d_{3t} = 1\{t \geq \tau_3\}$.

Equation (2) can then be rearranged as:

$$\begin{aligned} y_t \;=\; & \left[d_{1t} + d_{2t}\left(\frac{\tau_3 - t}{\tau_3 - \tau_2}\right)\right] x_t'\alpha_1 + \left[d_{3t} + d_{3t}\left(\frac{t - \tau_3}{\tau_3 - \tau_2}\right)\right] x_t'\alpha_3 \\ & + \left[d_{1t}\varepsilon_{1t} + d_{2t}\left(\frac{\tau_3 - t}{\tau_3 - \tau_2}\right)\varepsilon_{2t}\right]\sigma_1 + \left[d_{3t}\varepsilon_{3t} + d_{2t}\left(\frac{\tau_3 - t}{\tau_3 - \tau_2}\right)\varepsilon_{3t}\right]\sigma_3, \end{aligned} \tag{3}$$

where $\tau_2 \in \{1, ..., T-1\}$ and $\tau_3 \in \{\tau_2 + 1, ..., T\}$, with $\tau_2 \leq \tau_3$.

2.2 Estimation

The model, as specified above, has $2k + 2$ free parameters: the k parameters in α_1, the k parameters in α_3, plus τ_2 and τ_3. We choose these parameters to minimize the unweighted sum of squared errors. Thus, for a given (τ_2, τ_3) pair, the estimation of (α_1, α_3) reduces to ordinary least squares (OLS). The optimal (τ_2, τ_3) can then be located by a search algorithm across possible values. To this end, we define the scalars:

$$z_{1t} \equiv d_{1t} + d_{2t}\left(\frac{\tau_3 - t}{\tau_3 - \tau_2}\right) \qquad \text{and} \qquad z_{3t} \equiv d_{3t} + d_{2t}\left(\frac{t - \tau_2}{\tau_3 - \tau_2}\right).$$

Then, given $\underset{1\times T}{y'} \equiv [y_1 \dots y_T]$ and

$$\underset{2k\times T}{Z'} \equiv \left[\begin{bmatrix} z_{11}x_1 \\ z_{31}x_1 \end{bmatrix} \dots \begin{bmatrix} z_{1T}x_T \\ z_{3T}x_T \end{bmatrix} \right],$$

the OLS estimators of (α_1, α_3) given (τ_2, τ_3) have a closed-form expression:

$$\begin{bmatrix} \widehat{\alpha}_1 \\ \widehat{\alpha}_2 \end{bmatrix} = [Z'Z]^{-1}Z'y.$$

Estimating σ_1 and σ_3 in this configuration is straightforward, except for the fact that the contribution of each variance to the total variance differs across periods and so the errors must be weighted accordingly. To this end, define:

$$e_t \equiv y_t - [z_{11}x_1 \ z_{31}x_1] \begin{bmatrix} \hat{\alpha}_1 \\ \hat{\alpha}_3 \end{bmatrix}, \qquad \underset{1\times T}{e_z^{1\prime}} \equiv [z_{11}e_1 \dots z_{1T}e_T], \qquad \text{and} \qquad \underset{1\times T}{e_z^{3\prime}} \equiv [z_{31}e_1 \dots z_{3T}e_T].$$

Given (τ_2, τ_3), we calculate the estimators for σ_1 and σ_3:

$$\widehat{\sigma}_1^2 = \left(\sum_{t=1}^{T} z_{1t} \right)^{-1} e_z^{1\prime} e_z^1 \qquad \text{and} \qquad \widehat{\sigma}_2^2 = \left(\sum_{t=1}^{T} z_{3t} \right)^{-1} e_z^{3\prime} e_z^3.$$

These estimators are asymptotically equivalent to the OLS estimators.[6]

While in general it may be interesting to include a larger number of regressors in x_t, we only consider the specification where x_t contains only a constant term, $x_t' = 1$ for $\forall t$ and $k = 1$. Hence, before a transition starts, i.e., while $t < \tau_2$, $y_t = \alpha_1$ (stage 1); between τ_2 and τ_3, y_t declines linearly (stage 2); and at τ_3, $y_t = \alpha_3$ (stage 3).

2.3 Restricted cases

A challenge in estimating the econometric model described above is data limitations. Even if the three-stage model of the demographic transition is a valuable characterization of the empirical evidence, one or more stages might not be observed, either because the sample is too short or because the demographic transition is still ongoing. In particular, we can have six different cases, as illustrated in Figure 2 for CBR transitions.

In the top left panel of Figure 2, we have case 1: all three stages are observed. In the top middle panel, we have case 2: only stages 2 and 3 are observed. In the top right panel, we see case 3, where only stages 1 and 2 are observed. On the bottom row, we see case 4, where just stage 2 is observed, case 5, where only stage 1 is observed, and case 6, where only stage

[6] When $\sum_{t=1}^{T} d_{st} = 1$ and $\sum_{t=1}^{T} d_{2t} = 0$ for $s \in \{1, 3\}$, σ_s is not identified, but this is of little consequence as none of the estimators for the other parameters depend on the variance estimates.

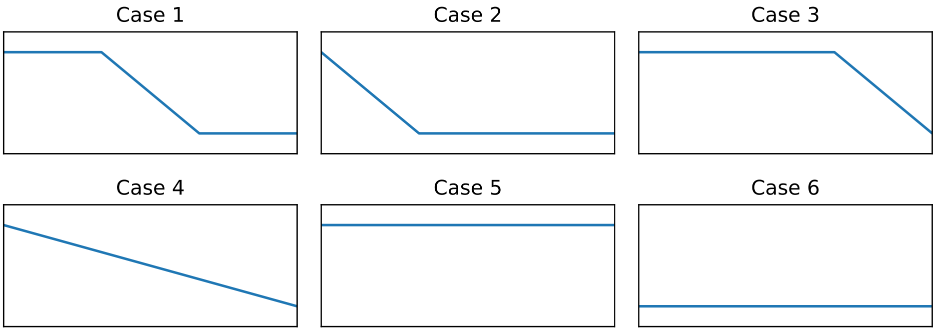


Figure 2: Six cases of the CBR transition

3 is observed. To distinguish case 5 from case 6, as they are equivalent econometrically, we look at the pre- and post-transition levels of the CBR and the CDR in comparison with historical averages across countries.

To discriminate between these possibilities, we estimate, for each country in the data, all six cases. Table 1 summarizes the nesting structure among cases. We select the version of the model that has the best trade-off between fitting the data and fewer restrictions. That is, we select a less restricted case only if it does a significantly better job of fitting the data. To this end, we use as our primary guide an F-test at the 95% confidence level: $\frac{\frac{SSE^b - SSE^a}{m^a - m^b}}{\frac{SSE^a}{T - m^a}}$, where $a, b \in \{1, 2, 3, 4, 5, 6\}$ and a nests b.

Table 1: Different cases of the general model

	Parameter restriction	Explanation	Num. of parameters
Case 1	–	All three stages are observed	$2k + 2$
Case 2	$\tau_2 = 1$	Only stages 2 and 3 are observed	$2k + 1$
Case 3	$\tau_3 = T$	Only stages 1 and 2 are observed	$2k + 1$
Case 4	$\tau_2 = 1, \tau_3 = T$	Only stage 2 is observed	$2k$
Case 5	$\tau_2 = 1, \tau_3 = T, \alpha_1 = \alpha_3$	Only stage 1 is observed	k
Case 6	$\tau_2 = 1, \tau_3 = T, \alpha_1 = \alpha_3$	Only stage 3 is observed	k

This statistical test performs best for countries with a long series of observations extending both before and after the transition in birth rates and/or death rates. To prevent our statistical method from over-fitting short-run anomalies in countries for which the time series is short, we also apply a set of simple auxiliary rules, all of them with a clear and simple intuitive interpretation (see Appendix B for a complete description of the auxiliary rules). For example, suppose the statistical method picks case 2 but detects the end of a fertility transition at a final level of higher than 20 per 1,000, with an end date less than 20 years before the end of the data series. Since this conflicts with the wide consensus of demographers that the fall in fertility does not stop until the CBR is at least around 10 per 1,000, we throw out this transition end date, moving the country from case 2 to case 4. The purpose of these auxiliary rules is, therefore, to ensure that the starts and ends of the transitions are clearly present in the data and are not the consequence of random fluctuations within short data series. Also, most importantly, our auxiliary rules do not change the dates selected for each case, only which case we select as the "best" description of the data.

3 Data and Results

We merge data from different sources to obtain time series for CBRs and CDRs that go back as long as possible for the greatest possible number of countries.[7] From 1960 onward, we rely on the World Bank Development Indicators. For many countries, we fill in the period between 1950 and 1960 with data from the U.N. data service of the United Nations Statistics Division. To gather vital statistics before 1950, we start with data from Chesnais' (1992) classic book on the demographic transition and augment them with observations from Mitchell's (2013) International Historical Statistics. We also use additional sources for a few countries: State Statistical Institute of Turkey (1995) and Shorter and Macura (1982) for Turkey; Swiss Federal Statistics Office (1998) for Switzerland; Maines and Steckel (2000) for the U.S.; Schofield and Wrigley (1989) for Great Britain/United Kingdom; Edvinsson (2015) and National Central Bureau of Statistics (1969) for Sweden; and Davis (1946) for India. The resulting data set of CDRs and CBRs covers 186 countries from 1541 to 2016. There are 16,206 country$\times$year observations for CDRs and 16,198 for CBRs. We take data on real GDP per capita (GDPpc), given in constant 2011 U.S. dollars purchasing power parity (PPP), from the 2018 version of Maddison's database.[8] The Madison data cover 165 countries between the years 1 and 2016, with 16,694 country$\times$year observations.[9]

Vital statistics from the 19$^{\text{th}}$ century and before are available for only a few countries. Many countries do not have data until after 1950. As a result, there are numerous countries where the start of either the CBR or the CDR transition is not observed (cases 2 and 4 in Figure 2). CDR transitions start, on average, earlier than CBR transitions, so we have more "missing starts" for the former than for the latter. We observe a CBR start, and no CDR start but a downward trend in CDR, for 109 countries. We project a CDR start date for 96 of them by drawing a line straight backward on the CDR trend until it hits the average observed starting gap between CBR and CDR, 8.86 per 1,000, the unweighted arithmetic mean across the 23 countries for which we observe the start of both transitions, and the fertility transition starts before 1950.[10] Using this procedure, the number of countries with

[7] Our database can be accessed interactively through the following web page: https://sites.google.com/view/demographic-transitions.

[8] Bolt et al. (2018). The database can be accessed here: https://www.rug.nl/ggdc/historicaldevelopment/maddison/releases/maddison-project-database-2018.

[9] There are 31 countries, most of them tiny island territories, for which we have data on CDRs and CBRs, but which are not included in Maddison's database. Maddison's database has data for Slovakia, but we exclude it to avoid double-counting, since for the majority of the covered period, Slovakia was part of Czechoslovakia.

[10] We dropped ten out of 109 countries because backward projection implies adding more than 100 years to the timeline, and three more countries because their estimated initial CBR-CDR gap is already smaller than the average.

a CDR transition start date more than doubles, from 46 to 142. The empirical analysis of mortality transitions is based on this extended data set.

Table 2 documents the distribution of all countries in our sample according to the six cases in Table 1. For each cell, we report two numbers: first, the number of cases before we make projections for the CDR, and then (in parenthesis) the change due to CDR projections. Hence, while the raw data have only 27 countries that have completed CDR and CBR transitions (case 1), this number increases to 52 after the adjustment, as we can pinpoint the start of CDR transitions for an additional 25 countries. Similarly, the number of countries with a complete CDR transition (case 1) and ongoing CBR transition (case 3) increases by 70.

Table 2: Case counts

CDR \CBR	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Total
Case 1	29 (+19)	0	18 (+70)	0	0	0	47 (+89)
Case 2	24 (-19)	20	77 (-70)	7	0	0	128 (-89)
Case 3	0 (+0)	0	1 (+2)	0	0	0	1 (+2)
Case 4	0 (+0)	0	2 (-2)	0	1	0	3 (-2)
Case 5	0	0	0	0	0	0	0
Case 6	0	7	0	0	0	0	7
Total	53	27	98	7	1	0	186

Out of 186 countries, we have 175 countries that have completed the mortality transition, and 80 that have completed the fertility transition (cases 1 and 2). This shows how the global drop in death rates is considerably more advanced than the decline in birth rates: most of the planet has finished the drop in CDRs, but there is still space to cover in the fall of CBRs. We do not find any country where the drop in the CDR has not started. We find one country, Chad, where we do not detect the beginning of a CBR transition. Finally, we have seven countries in case 6 of the CDR. These are typically Eastern European countries that started their demographic transitions earlier than the availability of data.

Figure 3 displays the time series of the CBRs and CDRs, along with the fitted transitions, for six representative countries. The top left panel is the demographic transition of Great Britain, a typical instance of an early demographic transition. The CDR started falling in 1794 and stabilized by 1958, while the CBR began dropping in 1885 and stabilized around 1937. The top right panel is the demographic transition of Denmark, a representative of many Western European countries that followed Great Britain's lead with a few decades delay. The left middle panel shows the demographic transition of Spain, a late but completed transition, with the CBR stabilizing in 1999. The right middle panel is the demographic transition for Chile, a typical case of late and ongoing transitions, where the CBR still has not stabilized. Finally, in the bottom row, we have Malaysia, a late demographic transition for which we calculate a projected start date for the fall of the CDR, and Chad, the one remaining country in our sample where it is not clear whether the fall in the CBR has started. Table D in the Appendix reports the start and end dates of the demographic transition for

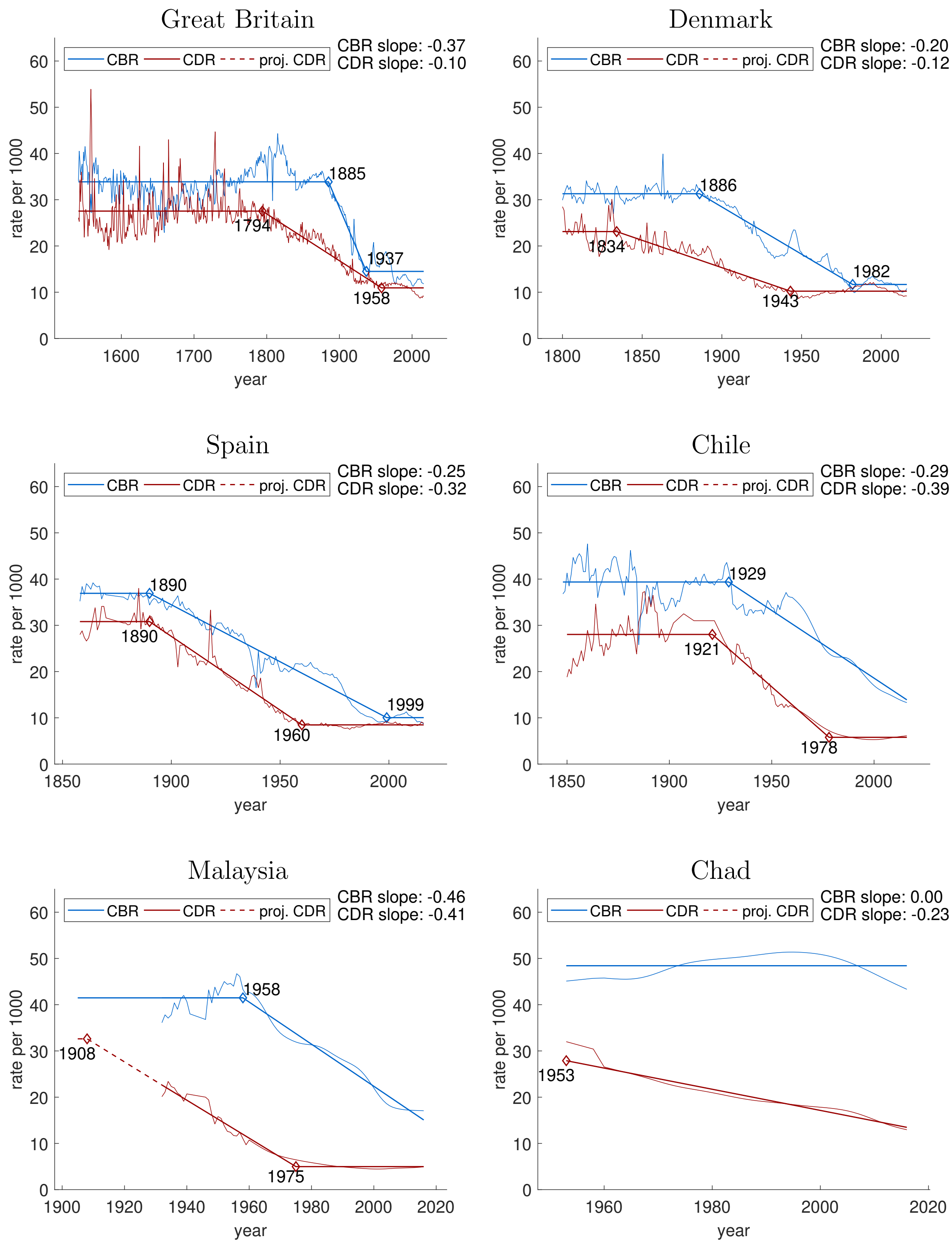


Figure 3: Six examples of demographic transitions

each country in our sample.[11]

[11]There are alternative ways to date the start of mortality and fertility transitions. Chesnais (1992), for example, suggests life expectancy exceeding 50 and the crude birth rate below 30 per 1000. Figure 26 in Appendix C shows how our start dates for fertility transitions compare with the date at which the CBR declines below 30 (where we have a large number of observations to compare two measures). The correlation between the two measures is high, 0.65. A simple 30 per 1000 rule assigns earlier start dates for many countries since our measure requires a sustained decline in the CBR, not a one-time decline below 30 that might be reversed later on.

3.1 Demographic transitions and GDP per capita

The average observed mortality transition starts at 27.05 deaths per 1,000, and ends when the CDR is 8.06 per 1,000. The average observed fertility transition begins with 42.87 births per 1,000 and ends with the CBR at 7.91 per 1,000. GDP per capita is equal to $1,938 at the start of the average mortality transition, and $2,724 at the start of the average fertility transition.

Table 3: Countries entering transitions

	bef. 1870	1870-1900	1900-1930	1930-1960	1960-1990	after 1990	All
exp{mean init. lnGDPpc}	$2,231	$2,322	$1,882	$1,604	$1,808	–	$1,939
mean init. CDR	29.92	26.37	25.08	28.01	29.94	–	27.05
mean slope CDR	-0.18	-0.26	-0.40	-1.01	-1.13	–	-0.51
N	11	12	25	12	5	0	65

	bef. 1870	1870-1900	1900-1930	1930-1960	1960-1990	after 1990	All
exp{mean init. lnGDPpc}	$1,845	$4,403	$2,208	$2,807	$2,893	$1,525	$2,724
mean init. CBR	42.53	35.90	37.87	41.08	44.26	46.40	42.87
mean slope, CBR	-0.19	-0.32	-0.32	-0.55	-0.54	-0.50	-0.49
N	6	11	5	19	71	11	123

Table 3 reports initial levels (α_1 from equation 3), slopes, and initial GDP per capita levels for CDR and CBR transition starts, grouped into six time periods. Here, three facts are revealed. First, the start dates of the CDR transitions are more dispersed over time than the start dates of the CBR transitions. Figure 4, which shows scatter plots of log GDP per capita in each country at the start of its CDR and CBR transition, also illustrates this fact. The circle for each country in these plots (and all the other plots in the paper) is proportional to its share of the 2016 world population. The start dates of the CDR transition peak sooner, with many starts clustered between 1900 and 1960. In comparison, most CBR transitions start between 1960 and 1990, with nine transitions starting since 1990.[12]

Log GDPpc at the start of the CDR transition Log GDPpc at the start of the CBR transition

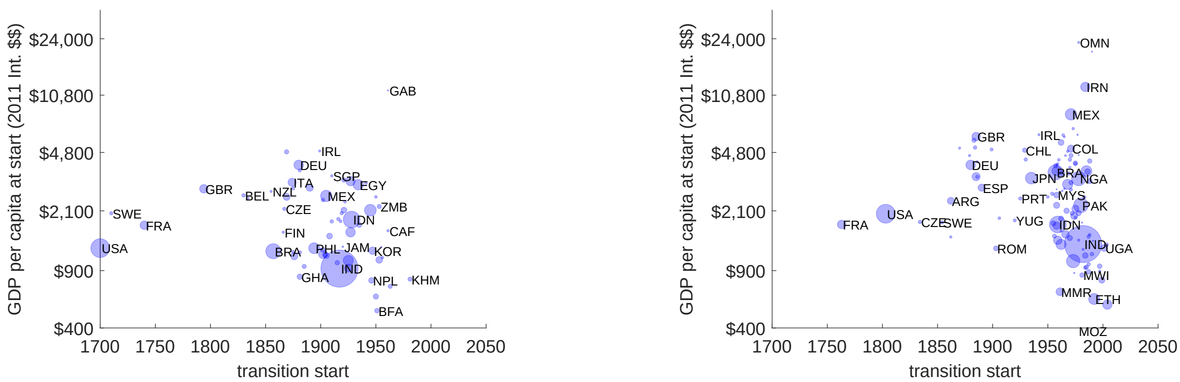


Figure 4: Log GDPpc at the start of transitions

[12]None of these qualitative facts are affected by excluding imputed mortality transition start dates. Also, note that while some of the mortality transition start dates are imputed, none of the GDP per capita numbers are imputed.

Second, later transitions are faster. The slope of the reduction in the CDR and the CBR during the transition is much larger for later transitions. Figure 5 shows this pattern for all the countries in our sample with complete transitions. Figure 6 makes the same point alternatively by plotting the measured transition length from plateau to plateau.[13] A linear regression for the slope and length of the transition speeds as a function of the start date (controlling for the level of GDP per capita at the transition start and the initial level of the CBR) shows a large and statistically significant negative coefficient: the later the start of the transition, the shorter it lasts.

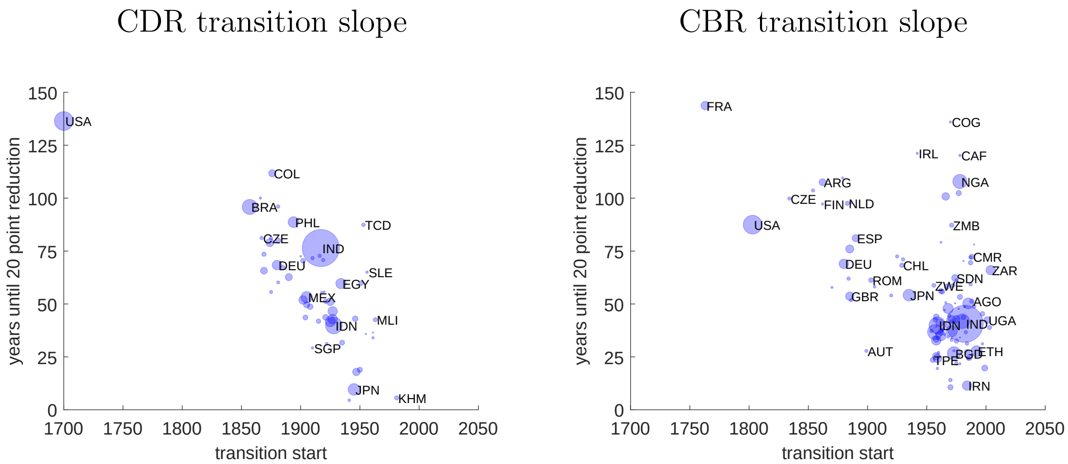


Figure 5: Transition slopes

Third, the average GDP per capita at the start of CDR and CBR transitions is roughly similar across time (see, again, Figure 4). For the CDR transitions that began in 1870-1900, GDP per capita at the start of the transition was \$2,322. Despite the larger variance in GDP per capita levels for transitions that start in the second half of the 20$^{\text{th}}$ century, for the 1960-90 period, GDP per capita at the start of the transition was not much lower, \$1,808.

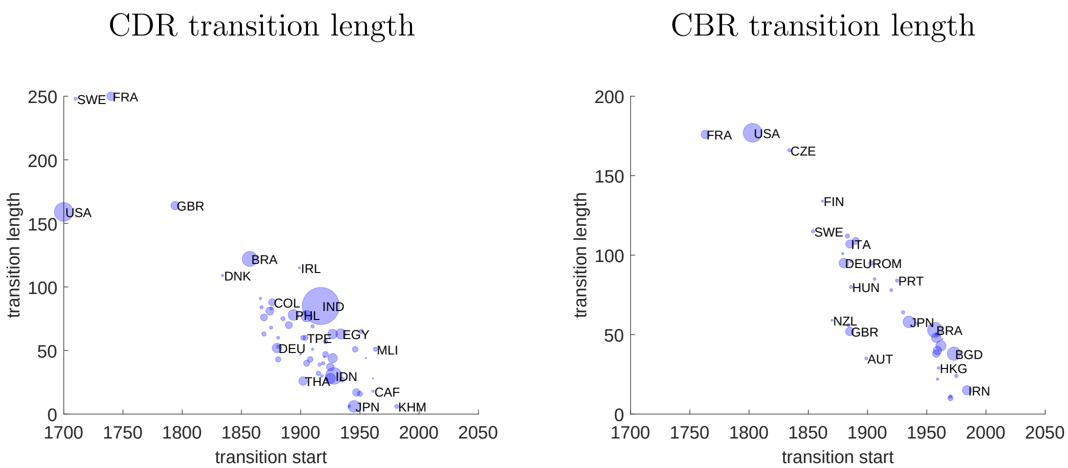


Figure 6: Transition lengths

[13]Note that the plot for CDR in the first panel of Figure 6 is populated mostly by countries with imputed mortality transition start dates. The exclusion of these countries does not significantly affect the trend line. Also, notice that the trend lines for the first panels of Figures 5 and 6 are nearly identical, and that Figure 5 does not depend on imputations.

Together with Great Britain, our estimated transition start dates single out France, Sweden, and the United States as demographic early starters.

4 A statistical model of demographic transitions

In this section, we model the demographic transition start dates as a time-varying hazard, investigating factors associated with fertility transition onset. Three frameworks are considered: the first, based solely on national GDP per capita, reproduces the distribution of GDP per capita at transition onset well but does not explain the timing of transitions. The second framework examines contagion effects: a transition in one country may make a transition in another country more likely if the two are geographically and linguistically close to each other. We find a statistically significant association for geographic and linguistic links and that this network-diffusion framework successfully explains the timing. The third framework uses a simplified diffusion model based on geographic and linguistic distance from a single country, with Great Britain as the origin, and provides a better fit for the timing of fertility transitions worldwide.

4.1 Hazard Model

Consider a world populated with N different countries indexed by $i \in \{1, 2, ..., N\}$ for which a set of variables $x_{it} \in X$ is observed at time $t \in \{1, 2, 3, ...T\}$. Let T^i represent the period at which a one-off event, such as the start of a CDR or CBR transition, occurs in country i.

Next, suppose that the probability of the event occurring at period t in country i, conditional on not having occurred previously, can be expressed as:

$$\Pr(T^i = t | T^i \geq t) = G\left(\sum_{l=0}^{k-1} x_{l,it}\beta_l\right), \tag{4}$$

where $G(\cdot)$ is a function bounded between 0 and 1, and $(x_{0,it}, x_{1,it}, ..., x_{k-1,it})$ is a set of k explanatory variables with coefficients β_l. We will assume that $G(\cdot)$ is the logistic CDF. Then, if the conditional probability of a transition is given by equation (4), the parameters of this model can be estimated by maximizing the log-likelihood:

$$\log L_N = \sum_{i=1}^{N}\sum_{t=1}^{T_i} \log\left[\mathcal{I}_{it} G\left(\sum_{l=0}^{k-1} x_{l,it}\beta_l\right) + (1-\mathcal{I}_{it})\left(1 - G\left(\sum_{l=0}^{k-1} x_{l,it}\beta_l\right)\right)\right], \tag{5}$$

where $\mathcal{I}_{it}$ is an indicator function taking the value 1 if the event occurs in country i at time t and 0 otherwise.

4.2 Transitions as a function of GDP

The fact that the distributions of log GDP per capita levels at the start of transitions in CBRs or CDRs are i) fairly stable over time and ii) roughly unimodal suggests a link between the level of log GDP per capita and transition takeoffs. In the first specification of our hazard model, we will consider it as the only factor. To this end, we construct a balanced panel with yearly interpolated values for real GDP per capita and transition status, starting in 1500. The 2018 version of the Maddison database assigns GDP per capita values for 11 countries in the year 1500. We expand our panel by making cautious imputations for 37 additional countries that have some pre-modern GDP per capita data, though not for 1500 specifically (see Appendix E). After excluding 4 countries for which we do not observe the start of the CBR transition and 1 for which we lack bilateral distance data, this gives us a panel of 43 countries between 1500 and 2016.

Table 4: Determinants of the start of the CBR transition

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
constant	-84.127**	-108.569**	-95.319**	-110.627**	-98.492**	-7.131	-122.044***	-17.692
	(33.170)	(37.111)	(36.823)	(37.595)	(39.009)	(54.371)	(36.654)	(57.358)
ln GDP pc	10.237**	13.186**	11.457**	13.455**	11.843**	6.331	6.504	5.733
	(4.315)	(4.840)	(4.804)	(4.903)	(5.091)	(4.494)	(4.519)	(4.628)
$(\ln \text{GDP pc})^2$	-0.484*	-0.772**	-0.658**	-0.786**	-0.693**	-0.361	-0.359	-0.337
	(0.266)	(0.302)	(0.300)	(0.305)	(0.316)	(0.280)	(0.279)	(0.288)
global transitions		0.217***	0.203***	0.219***	0.241***			
		(0.050)	(0.052)	(0.051)	(0.055)			
$(\text{global trans.})^2$		-0.002*	-0.002*	-0.002*	-0.003**			
		(0.001)	(0.001)	(0.001)	(0.001)			
year						-0.021	0.035***	-0.014
						(0.022)	(0.006)	(0.024)
physically close transitions			0.083**					
			(0.040)					
linguistically close transitions				-0.020				
				(0.064)				
phys. *and* ling. close trans.					0.686***			
					(0.198)			
phys. close, ling. far trans.					0.018			
					(0.051)			
ling. close, phys. far trans.					-0.092			
					(0.083)			
ln GC dist from GBR						-3.849**		-4.257**
						(1.347)		(1.404)
ln GC dist from GBR × year						0.008**		0.009**
						(0.003)		(0.003)
ling. prox. to GBR							11.283*	22.522**
							(5.920)	(7.791)
ling. prox. to GBR × year							-0.016	-0.044*
							(0.020)	(0.023)
log-likelihood	-249.3	-208.7	-206.6	-208.6	-202.4	-200.9	-204.3	-195.2
Pseudo-R^2	0.181	0.314	0.321	0.315	0.335	0.340	0.329	0.359
N. obs.	18,775	18,775	18,775	18,775	18,775	18,775	18,775	18,775

The first column of Table 4 reports the results for this specification.[14] Figure 7 shows how well this specification replicates the distribution of log GDP per capita at the start of

[14] For consistency with the specifications discussed in Section 4.3, this estimation excludes from its sample the few countries that do not have a full set of bilateral distance data. An estimation including the excluded countries yields almost the same coefficients.

the transition. The predicted mean and standard error are 8.2 and 0.70, versus an observed mean and standard error of 7.9 and 0.63, a remarkably close fit. In other words, this simple specification is sufficient to generate the observed aggregate distribution of transition starts across levels of GDP per capita.

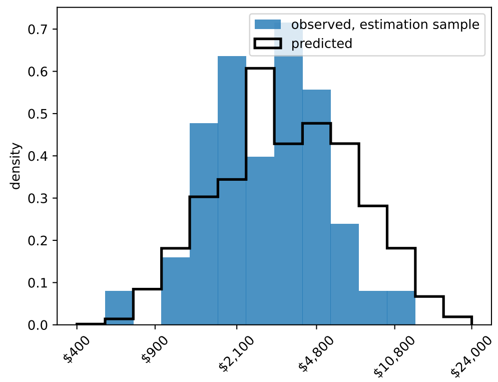


Figure 7: Distribution of log GDPpc at the start of the CBR transitions

Panel (a) of figure 8 plots observed and predicted start dates for individual countries. Three-letter country abbreviations and 60% confidence intervals are plotted for a subset of countries. The confidence intervals are large, and predicted transition dates are generally too early, reflecting the significant number of countries with relatively high GDP per capita levels for many decades before starting their transitions.[15]

Panel (b) of figure 8 plots the observed distribution of start dates (solid bars) across decades against the distribution generated by the model (empty bars). The model fails to generate the clustering of CBR transitions around the turn of the 20th century and the very large peak around the 1960s. These clusters of start dates suggest the presence of a demographic contagion: the possibility that prior transitions by a country's neighbors may push forward its demographic transition. Now, we will extend the model to study whether the data statistically support this reading of the evidence.

4.3 Demographic contagion

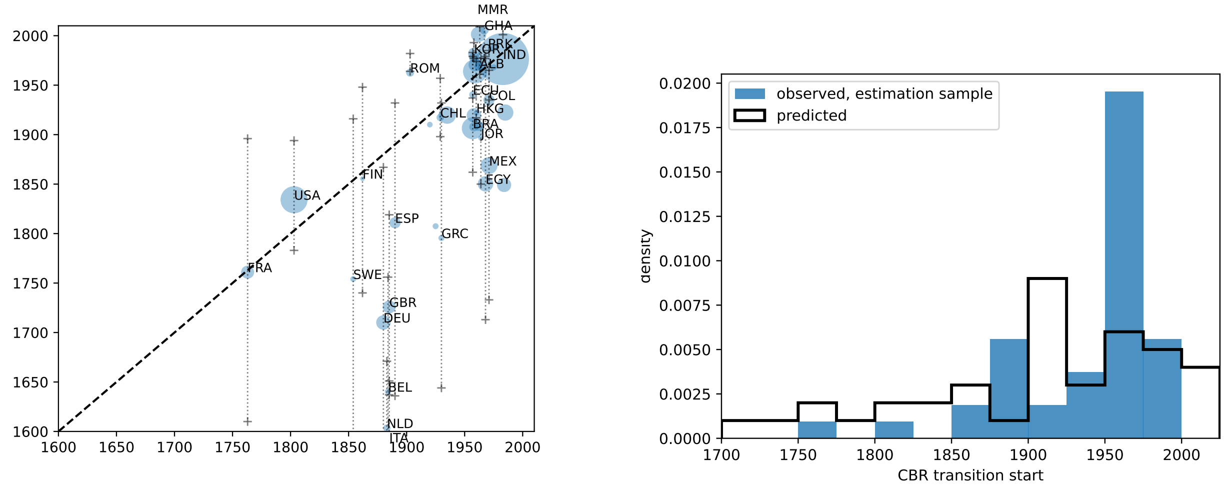
To allow for the possibility of demographic contagion, we construct access to transitions $\mathcal{A}_{it}$ in the following way:

$$\mathcal{A}_{it} = \underbrace{\delta_0 \sum_{j=1}^{N} \mathcal{I}_{j,t-1} + \delta_1 \left(\sum_{j=1}^{N} \mathcal{I}_{j,t-1} \right)^2}_{\text{global}} + \underbrace{\sum_{z \in Z \setminus z_0} \gamma_z \sum_{j|(i,j) \in z} \mathcal{I}_{j,t-1}}_{\text{network-mediated}} \tag{6}$$

In (6), $\mathcal{I}_{j,t}$ is an indicator function taking a value of 1 if country j started its transition

[15]Country fixed effects would bring each country's mean predicted transition start date in line with the observed date but would not narrow the confidence intervals.

Figure 8: Specification (1), Start of CBR Transitions



Panel (a): within-sample predictions   Panel (b): distribution of transition dates

before period t, and 0 otherwise. Access has a global component and a network-mediated component. In the global component, the parameter δ_0 determines the average impact of one additional transition, anywhere in the world, on the probability of every other country transitioning. The parameter δ_1 is applied to the square of the global sum of transitions, allowing for the impact of an individual transition to be attenuated (or, possibly, amplified) as the overall number of past transitions adds up.

To construct the network-mediated component of (6), we segment all pairs of countries (i, j) into mutually-exclusive proximity groups, defined below:

$$\begin{aligned} z_0 &: \text{ all pairs } (i,j) \text{ s.t. } i \text{ and } j \text{ are neither physically nor linguistically close} \\ z_1 &: \text{ all pairs } (i,j) \text{ s.t. } i \text{ and } j \text{ are physically but not linguistically close} \\ z_2 &: \text{ all pairs } (i,j) \text{ s.t. } i \text{ and } j \text{ are not physically but are linguistically close} \\ z_3 &: \text{ all pairs } (i,j) \text{ s.t. } i \text{ and } j \text{ are both physically and linguistically close} \end{aligned} \tag{7}$$

In (6), γ_z for $z \in Z \setminus z_0$ represent the additional effect, over and above the average effect, of a transition in one country on other countries that are close to it in the way defined by proximity group z. Country pairs having no close relationship, z_0, is the excluded category.[16]

With access thus defined, the extended model is given by

$$\Pr(T^i = t | T^i \geq t) = G\left(\sum_{l=0}^{k-1} x_{l,it}\beta_l + \mathcal{A}_{it}\right). \tag{8}$$

To implement this specification, we must define our proximity categories empirically. We take physical distance to be the average great circle distance between the largest 25 cities

[16] By including linguistic proximity, we follow the international trade literature highlighting the importance of non-geographic factors in gravity equations. See Egger and Lassmann (2012) and Melitz and Toubal (2013).

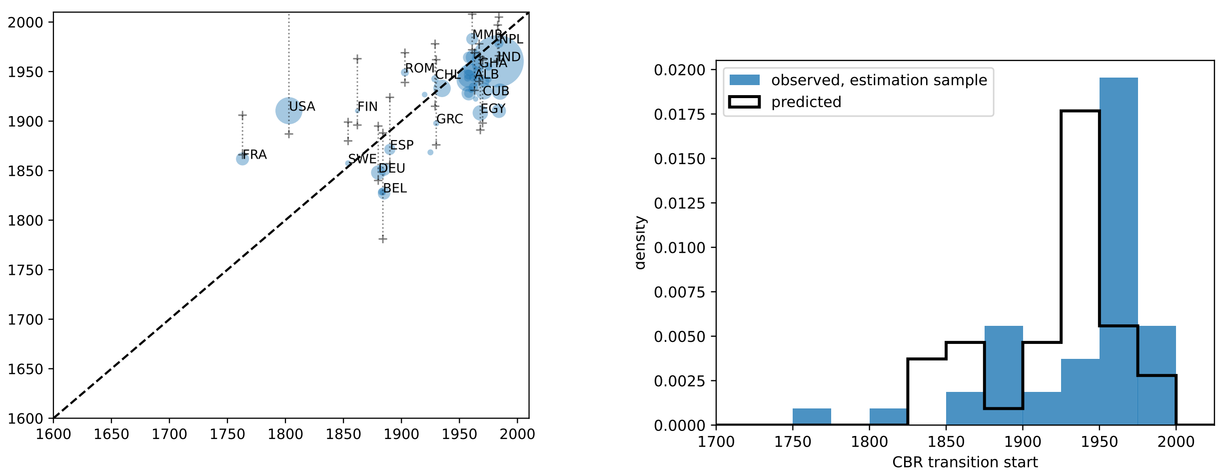
in each country, weighted by population, as provided by Mayer and Zignago (2011). Two countries are defined as being physically close if their physical distance is at or below the 10^{th} percentile of bilateral distances, which is 2,263 km. For linguistic proximity, we use the "PROX2" measure constructed by Melitz and Toubal (2013).[17] Two countries are defined as linguistically close if their proximity is at or above the 90^{th} percentile of bilateral linguistic proximities–a "PROX2" value of 0.195.

Column (2) in Table 4 shows results for a specification in which all countries have the same impact on transition probability: $Z = \{\}$. We find a strong, significant positive effect, i.e., previous transitions are strongly correlated with future ones, beyond what is captured by GDP growth over time. A weakly significant negative coefficient on the quadratic term suggests some attenuation in the cumulative effect.

Next, in columns (3) and (4), we allow countries to have a different bilateral effect on each other, using a single measure of closeness: $Z = \{z_0, z_1 \cup z_3\}$ for column (3), $Z = \{z_0, z_2 \cup z_3\}$ for column (4). Physical proximity has a positive impact, i.e., previous transitions by countries that are geographically close matter more, while no significant effect is detected for linguistic proximity.

Finally, Columns (5) of 4 shows the results of the full specification, in which bilateral effects are allowed to differ according to both proximity measures. Now, being physically and linguistically close is estimated to have a significant positive effect. This is despite the weak or insignificant effects found for linguistic proximity individually, indicating a complementarity between the two types of proximity.

Figure 9: Specification (5), Start of CBR Transitions



Panel (a): within-sample predictions Panel (b): distribution of transition dates

Panels (a) and (b) of 9 show the improvement of specification (5) in matching the ob-

[17] Melitz and Toubal (2013) test several alternative measures of the degree of linguistic commonality between countries, ranging from the narrowest definition (i.e., whether the two countries share an official language), to more nuanced definitions based on the shares of the population in each country that speak the same or similar languages. "PROX2" is comprehensive yet parsimonious. See also Bakker et al. (2009).

served transition starts across time. Except for calling France and the US too late, the specification with demographic contagion does much better, as most of the observations are now more closely aligned along the 45-degree line. The estimation also comes close to producing the twin peaks of demographic transitions in the late 19th century and in the 1960s, though the first it produces a decade early and the second a decade late.[18]

4.4 Single-source diffusion

One weakness of the network diffusion model we have specified above is that it does not allow the strength of the influence between two countries to vary continuously–the countries are either close or not. If we wished to identify the effect of distance on contagion more precisely, uncertainty about the appropriate specification and the relatively small sample of countries would represent formidable barriers.

Another approach is suggested by the Lucas (2002) model of economic growth take-offs, and echoed by Spolaore and Wacziarg's (2021) study of the diffusion of the French fertility transition. We can posit that the network influences on a country's likelihood to transition can be summarized by the strength of its connection to a single leading country. One reason this specification is likely to embed a large part of the information contained in the full network specification is that countries that are the same distance from a single lead country are likely to be close to each other. As we are here taking a view of the demographic transition as a simultaneous and possibly complementary phenomenon to the Industrial Revolution, it is natural for us to place Great Britain in this leading role. Let us then define an alternative access measure $\tilde{\mathcal{A}}_{it}$ as

$$\tilde{\mathcal{A}}_{it} = \nu_P d_i^P + \nu_L p_i^L + \eta_P d_i^P \times t + \eta_L d_i^L \times t, \tag{9}$$

...where d_i^P and d_i^L represent country i's log physical distance to and linguistic proximity from the lead country, respectively. Parameters ν_P and ν_L represent the time-invariant base effect of proximity, while η_P and η_L represent the time-varying interaction effects. We then supply $\tilde{A}_{it}$ in place of $\mathcal{A}_{it}$ in equation (8).

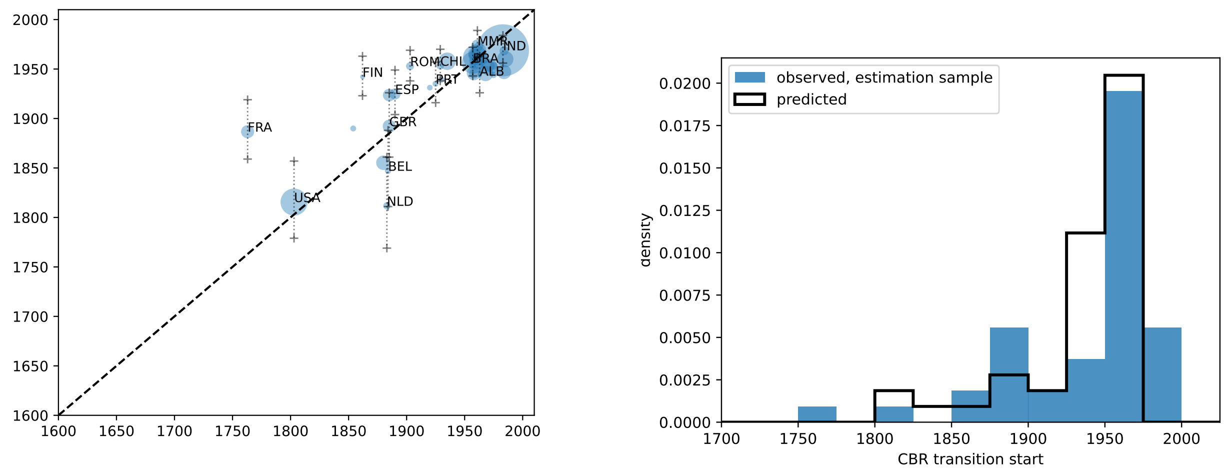
Column (8) of table 4 presents the results for this specification, with columns (6) and (7) presenting the results of restricted versions with only physical distance or linguistic proximity. The estimated coefficients have the expected signs and are significant in all cases: ν_P is negative while η_P is positive, and ν_L is positive while η_P is negative, indicating a significant effect for both proximity measures, which is declining over time.

Panels (a) and (b) of Figure 10 show the estimated distribution of demographic transition starting dates for the full specification. Interestingly, we can see that it achieves a better fit with the data than any of the other specifications that were tried, hitting both the late-19th

[18] In Appendix G, we show that weighting each country by its population (i.e., having a neighbor with a larger population might have a bigger impact on contagion) does not affect our results.

and mid-20th century peak of transitions in the correct decades.[19]

Figure 10: Specification (8), Start of CBR Transitions



Panel (a): within-sample predictions | Panel (b): distribution of transition dates

4.5 Crude Death Rate contagion

In Appendix section H, we report the results of running the same set of specifications described above to a model of crude death rate transition starts. Unlike the estimates for crude birth rates, GDP per capita appears to explain little of transition timing. Physical and linguistic distances from Great Britain have great explanatory power again, supporting the model presented in Section 5.

4.6 Taking stock

In the preceding two sections, we have documented four findings. First, the start dates of the CDR transitions are more dispersed over time than the start dates of the CBR transitions. Second, transitions in both fertility and mortality have been getting faster over time. Third, despite this increase in the speed of the transitions, the average GDP per capita at the start of the CDR and CBR transition is similar across time. Finally, we have significant demographic contagion, whereby a transition in one country is statistically associated with following transitions in countries close to it geographically and linguistically. We show that this contagion framework explains the timing of transitions well when formulated in terms of diffusion through a time-invariant bilateral network and even better when formulated as time-varying diffusion from a single lead country, Great Britain.

[19]In table A2 in Appendix F, we also estimate the full specification with the place of Great Britain taken by France and Sweden, two other countries that started their demographic transitions quite early. Although the specification with Great Britain fits the data slightly better, the results are rather similar. This is not surprising, given the closeness of the estimated fertility start dates, and the geographic proximity of France and Sweden to Great Britain.

5 An economic model of demographic transitions

We now build a model of endogenous fertility, education, and technology diffusion to understand our previous four empirical findings. Following Barro and Becker (1989), parents face a quantity-quality trade-off between how many children to have and how much to educate them. The economy has an ancient and a modern sector, as in Hansen and Prescott (2002). Both sectors use land and skilled and unskilled labor, but the modern sector uses skilled labor more intensively. With economic growth, total factor productivity (TFP) in both sectors increases, and the skill premium rises as resources move from the ancient to the modern sector. Economic growth also brings improvements in life expectancy. As in Lucas (2009), economic growth starts in Britain and then diffuses to other countries. In our model, there is an advantage to backwardness: the later economic growth starts in a country, the faster it occurs.

5.1 Preferences, fertility, and education decisions

Consider a world that consists of many locations, which will correspond to countries in our analysis. For ease of exposition, we drop the country index i whenever this does not cause any confusion. Households in each country i live for four periods: period 0 as children, period 1 as young adults, period 2 as middle-aged adults, and period 3 as elders. The probability of an infant born at time t surviving birth and becoming a child is s_t^0. The probability of a child born at time t surviving to adulthood is s_t^1. Finally, the probability that a young adult and a middle-aged adult survive to middle and old age are given by s_t^2 and s_t^3, respectively.

Children are provided with basic sustenance by adults and do not earn an income or make independent decisions. Young adults are endowed with 1 unit of time, denoted by $\zeta^1 = 1$, which they divide between market work, caring for children, and educating them. Middle-aged adults and elders are endowed with $\zeta^2 \geq \zeta^3$ units of time, which they provide inelastically to the labor market. The income that parents receive per unit of labor depends on unskilled and skilled wages, w_t^U and w_t^S and their human capital h_t. The total income of an age j worker born at time t is given by $\zeta^j \left(w_{t+j-1}^U + h_t w_{t+j-1}^S\right)$ for $j = 1, 2, 3$.

Young adults choose how many children to have, n_t, and how much education, e_t, to provide for each child that survives infancy (a concave utility ensures it is optimal to give each child the same level of education). The level of education that children receive, e_t, and parental human capital, h_t, determine their level of human capital when they are adults: $h_{t+1} = h_t^{\upsilon} e_t^{\xi}$, with $\upsilon, \xi \in (0, 1)$.

Each birth requires a time commitment of τ_1, and giving each surviving child a unit of education requires a time investment of τ_2. Since a young adult who chooses to have n_t births at time t will have $s_t^0 n_t$ surviving children to educate, the total time cost of education is $n_t \tau_2 s_t^0 e_t$. Parents must provide themselves, and each child who survives infancy, at least

$\bar{c}$ units of the consumption good as sustenance.

Let c_t^j be the consumption of an age-j adult at time t. Parents choose $\left\{c_{t+j}^{j+1}\right\}_{j=0}^{2}$, e_t, and n_t to maximize

$$\log\left(c_t^1\right) + \gamma \log\left(s_t^0 n_t - \overline{n}\right) + \phi s_t^1 \log\left(w_{t+1}^u + w_{t+1}^s h_{t+1}\right) + s_t^2 \log\left(c_{t+1}^2\right) + s_{t+1}^3 \log\left(c_{t+2}^3\right),$$

subject to

$$\begin{aligned} c_t^1 &= (w_t^u + w_t^s h_t)(1 - n_t(\tau_1 + \tau_2 s_t^0 e_t)) - (1 + s_t^0 n_t)\bar{c}, \\ c_{t+j}^{j+1} &= \zeta^{j+1}(w_{t+j}^u + w_{t+j}^s h_t) - \bar{c} \text{ for } j = 1, 2, \\ h_{t+1} &= h_t^v e_t^\xi. \end{aligned}$$

Aside from their consumption, adults derive utility from the number of children who survive infancy, $s_t^0 n_t$. The parameter $\gamma \geq 0$ encodes the strength of this preference, and $\overline{n}$ is a parameter representing the minimum desired number of descendants who survive to adulthood. They also derive "warm glow" utility from anticipating the future wage income of their children, $w_{t+1}^u + w_{t+1}^s h_{t+1}$. The parameter $\phi \geq 0$ represents the strength of this preference.

The first-order conditions for an interior solution for n_t is:

$$\frac{1}{c_t^1}\left[(w_t^u + w_t^s h_t)(\tau_1 + \tau_2 s_t^0 e_t) + s_t^0 \overline{c}\right] = \gamma \frac{s_t^0}{s_t^0 n_t - \overline{n}}. \tag{10}$$

The marginal cost of additional children (i.e., the left-hand side of equation 10) is increasing in the opportunity cost of parents' time, $w_t^u + w_t^s h_t$, the time cost of children, τ_1, and, if $e_t > 0$, in the time cost of education, τ_2. The marginal benefit of additional children (i.e., the right-hand side of equation 10) is decreasing in n_t. Similarly, the first-order condition for $e_t > 0$ is:

$$\frac{1}{c_t^1}\left[(w_t^u + w_t^s h_t) n_t s_t^0 \tau_2\right] = \phi \frac{\xi s_t^1 h_t^v \left(\frac{w_{t+1}^s}{w_{t+1}^u}\right)}{1 + h_t^v \left(\frac{w_{t+1}^s}{w_{t+1}^u}\right) e_t^\xi} e_t^{\xi - 1}, \tag{11}$$

where the marginal cost of e_t is increasing in $(w_t^u + w_t^s h_t)$ and n_t while the marginal benefit (i.e., the right-hand side) is increasing in the skill premium $\frac{w_{t+1}^s}{w_{t+1}^u}$ at time $t+1$.[20]

Notice that e_t increases the marginal cost of a child in equation (10) and the marginal benefit in equation (11). These two opposite forces embody the quality-quantity trade-off that parents face.

[20]See Vogl (2016) for the importance of corner solutions in the fertility/education decisions of poorer parents. Since we worked with a representative household within each location, we ignore that possibility.

5.2 Population dynamics

Survival probabilities are determined as a function of M_t, the medical technology at time t:

$$s_t^j = 1 - (1 - s_0^j)\frac{1 + e^{1-\delta}}{1 + e^{M_t - \delta}}, \text{ for } j = 0, 1, 2, 3, \tag{12}$$

where s_0^j is the initial survival probability for age j. We assume that the initial level of medical technology M_0 is equal to 1. The parameter δ determines how slowly survival rates increase as M_t increases. A higher value of δ implies slower increases in survival. In the limit, as δ approaches infinity, survival probabilities will never change, regardless of the value of M_t.

To calculate model-implied crude birth rates and death rates, we take the total population to be equal to the number of age-1 adults, plus the number of age-2 adults, plus the number of age-3 adults, plus the total number of children born: $(1 + n_t)N_t^1 + N_t^2 + N_t^3$. The total number of births is given by $n_t N_t^1$, and therefore, the crude birth rate is given by

$$CBR_t \equiv \frac{n_t N_t^1}{(1 + n_t)N_t^1 + N_t^2 + N_t^3}$$

The total number of deaths that occur during a model period is equal to the sum of infant and childhood deaths, plus those young adults who do not make it to middle age, plus those middle-aged adults who do not survive to be elders, plus the entire elderly cohort. The crude death rate, therefore, is given by

$$CDR_t \equiv \frac{(1 - s_t^0 s_t^1)n_t N_t^1 + (1 - s_t^2)N_t^1 + (1 - s_t^3)N_t^2 + N_t^3}{(1 + n_t)N_t^1 + N_t^2 + N_t^3}.$$

The law of motion of the population is determined by the size of the young adult and middle-aged cohorts (N_t^1, N_t^2), the fertility rate n_t, and survival rates. Given these time-t objects, time $t+1$ adult population levels are given by $N_{t+1}^1 = n_t s_t^0 s_t^1 N_t^1$, $N_{t+1}^2 = s_t^2 N_t^1$, and $N_{t+1}^3 = s_t^3 N_t^2$.

5.3 Production

For any country i, the economy consists of two sectors: ancient and modern, which produce the same homogeneous good. Ancient sector production $Y_{a,t}$ (i.e., agriculture, servants, small-scale low-skill artisans, etc.) is carried out by a representative firm with a production function $Y_{a,t} = A_t L_{a,t}^{\alpha} H_{a,t}^{\rho_a - \alpha} T_{a,t}^{1-\rho_a}$, using unskilled labor, $L_{a,t}$, skilled labor, $H_{a,t}$, and land, $T_{a,t}$ and given a TFP level A_t. We will assume that $1 > \rho_a > \alpha > 0$.

In comparison, modern sector production $Y_{b,t}$ (i.e., mechanized industry, professional services) is carried out by a representative firm with a production function $Y_{b,t} = B_t L_{b,t}^{\beta} H_{b,t}^{\rho_b - \beta} T_{b,t}^{1-\rho_b}$, using unskilled labor, $L_{b,t}$, skilled labor, $H_{b,t}$, and land, $T_{b,t}$ and given a TFP level B_t. We will assume that $1 > \rho_b > \beta > 0$.

The main difference between the sectors is that the ancient sector is more unskilled-labor intensive than the modern sector, that is, $\alpha > \beta$. We assume that there is a fixed amount of land used in each sector, normalized to $T_{a,t} = T_{b,t} = 1$.[21] Thus, total production Y_t is

$$Y_t = Y_{a,t} + Y_{b,t} = A_t L_{a,t}^{\alpha} H_{a,t}^{\rho_a - \alpha} + B_t L_{b,t}^{\beta} H_{b,t}^{\rho_b - \beta},$$

and it is divided between wages paid for unskilled and skilled labor, which workers consume, and rents paid for land, which is consumed by absentee landlords. We assume that landlords are a distinct class with a negligible population size who play no economic or demographic role other than to receive land rents.

There is a representative firm in each sector that solves a standard profit-maximization problem under conditions of perfect competition. This, combined with the perfect mobility of skilled and unskilled labor, implies that:

$$w_t^u = \alpha A_t L_{a,t}^{\alpha-1} H_{a,t}^{\rho_a - \alpha} = \beta B_t L_{b,t}^{\beta-1} H_{b,t}^{\rho_b - \beta}, \tag{13}$$

and

$$w_t^s = (\rho_a - \alpha) A_t L_{a,t}^{\alpha} H_{a,t}^{\rho_a - \alpha - 1} = (\rho_b - \beta) B_t L_{b,t}^{\beta} H_{b,t}^{\rho_b - \beta - 1}. \tag{14}$$

Together, these conditions imply that the skill premium, $\frac{w_t^s}{w_t^u}$, must have a particular relationship with the relative importance of each skill type, and the ratio of unskilled to skilled labor in each sector:

$$\frac{w_t^s}{w_t^u} = \frac{\rho_b - \beta}{\beta} \frac{L_{b,t}}{H_{b,t}} = \frac{\rho_a - \alpha}{\alpha} \frac{L_{a,t}}{H_{a,t}}.$$

This, with market clearing conditions $H_{a,t} + H_{b,t} = \overline{H}_t$ and $L_{a,t} + L_{b,t} = \overline{L}_t$, implies:

$$\frac{w_t^s}{w_t^u} = \left[\frac{\rho_a - \alpha}{\alpha} \frac{L_{a,t}}{\overline{L}_t} + \frac{\rho_b - \beta}{\beta} \frac{L_{b,t}}{\overline{L}_t} \right] \frac{\overline{L}_t}{\overline{H}_t}. \tag{15}$$

Equation (15) reveals that the equilibrium skill premium depends on two components. First, it depends on the supply of skilled relative to unskilled labor, $\frac{\overline{L}_t}{\overline{H}_t}$. The relatively scarcer the skill is, the better it will be paid. Second, it depends on the relative importance of skilled versus unskilled labor in production. Because we have two production functions, this mechanism depends on a weighted average of the factor-share ratios for the ancient and modern sectors, $\frac{\rho_a - \alpha}{\alpha}$ and $\frac{\rho_b - \beta}{\beta}$, where the weights are the fraction of unskilled labor in each sector, $\frac{L_{a,t}}{\overline{L}_t}$ and $\frac{L_{b,t}}{\overline{L}_t}$. If $\frac{\rho_b - \beta}{\beta} > \frac{\rho_a - \alpha}{\alpha}$, an increase in the share of output produced in the modern sector will lead to a rise in the skill premium unless we see a sufficiently sizeable counterbalancing increase in the relative supply of skilled labor.

[21] An alternative, isomorphic specification would have no land and decreasing returns to scale in production.

5.4 An economic-demographic history of the world

We now use our model to describe the timeline of the world's economic and demographic changes in a stylized way. From the start of time until a time $t = \tau$, the growth rates of ancient, modern, and medical technologies are, in every country of the world, zero (the "Malthusian era"). For any country i, until $t = \tau$, $M_{i,t} = M_0$, $A_{i,t} = A_0$, $B_{0,t} = B_0$ (we characterize this Malthusian steady state analytically in Appendix L).

There exists one country, $i = f$, which we call the "frontier country". At time $t = \tau$, ancient, modern and medical technologies in the frontier country begin to grow at constant, possibly different, rates:

$$\frac{A_{f,t+1} - A_{f,t}}{A_{f,t}} = \mu_f^a, \ \frac{B_{f,t+1} - A_{f,t}}{B_{f,t}} = \mu_f^b, \ \frac{M_{f,t+1} - M_{f,t}}{M_{f,t}} = \mu_f^m \text{ for all } t \geq \tau.$$

This can be interpreted as the dawn of the Industrial Revolution.

Each country $i \neq f$ in the world exists at a certain effective distance $d_{i,t}$ from the frontier country. This distance represents a combination of geographical distance and other possible barriers to trade and the exchange of ideas, such as differences in language.[22] For $t < \tau$, it takes a constant value $d_{i,t} = d_{i,0}$. At time $t = \tau$, when frontier technology begins to grow, effective distances simultaneously begin to shrink at a constant rate $\varphi > 0$, reflecting concurrent improvements in transportation and communication. Thus, for $t \geq \tau$,

$$d_{i,t} = d_{i,0}(1 - \varphi)^{t-\tau}.$$

As $d_{i,t}$ falls, the technological progress of the frontier country diffuses, first to its near neighbors and eventually to the whole world. More concretely, for a country i that is at a distance $d_{i,t}$ from the frontier country, the growth rates of $A_{i,t}$, $B_{i,t}$, and $M_{i,t}$ are given by

$$\mu_{i,t}^x = \mu_f^x \times \exp(-d_x \times d_{i,t}^{\lambda}) \times \left(\frac{X_{f,t-1}}{X_{i,t-1}}\right)^{\theta}, \tag{16}$$

for $x \in \{a, b, m\}$, having abused notation slightly. Parameters $d_b, d_a, d_m > 0$ represent the strength of distance as a barrier to the diffusion of technology in the modern, ancient, and medical sectors, respectively. The parameter $\lambda > 1$ captures the convex relationship between geographic distance and the diffusion of technology.

Hence, all countries have the potential to grow, but in the first periods following time τ, those countries that are farther away from Great Britain will see close to zero growth, while nearby countries grow at rates close to μ_f^j, $j \in \{a, b, m\}$. Effective distances keep shrinking; eventually, even the most remote countries experience economic growth and demographic change.

Countries that begin to grow later have the advantage of backwardness. After time τ,

[22] A potential link between distance and diffusion might work through trade. Much new technology is embodied in traded capital goods.

any distant country growing at a slow rate will see its technology gap with the frontier, $\frac{X_{f,t-1}}{X_{i,t-1}}$ for $X = A, B, M$, widen. But this has a positive effect on the growth rate, captured by the $\left(\frac{X_{f,t-1}}{X_{i,t-1}}\right)^{\theta}$ term in (16). The parameter $\theta > 0$ controls the elasticity of catch-up growth to backwardness.

In our simulations, we choose Great Britain as the frontier country. The widespread consensus among economic historians is that Great Britain was the technological leader of the Industrial Revolution that started during the 18th century (Broadberry et al., 2015). This is why it is commonly posed as the leader in simulations of economic growth (e.g., Lucas, 2002). Furthermore, the availability of long, high-frequency, high-quality British historical data makes it a convenient target for calibrated parameters.[23]

6 Quantitative analysis

To take the model to the data, we proceed in two steps. First, we select preference and technology parameters so that a single simulated country matches demographic and economic observations in Great Britain, our frontier country. Then, we choose the parameters that determine cross-country technology diffusion so that a simulation of all countries matches key features of the global demographic and economic transition.

6.1 Great Britain

We start by normalizing some parameters or borrowing them from the literature, as summarized in Table 5. Following Desmet and Rappaport (2017), we set $\rho_a = 0.7$ (a land share of 0.3 in the ancient sector) and $\rho_m = 0.9$ (a land share of 0.1 in the modern sector). The labor endowment for middle-aged adults ζ^2, is set to 1, and the labor endowment for the elderly, ζ^3, is set to 0.5.[24] The initial levels of agricultural technology, A_0, and medical technology, M_0, are normalized to 1. The initial mortality levels s_0^j are calculated from the mortality rates for the 1675-1699 period reported by Schofield and Wrigley (1989). Appendix M explains the mapping between age-specific mortality rates in the data and their model counterparts.

Next, we select 15 parameters so that the simulation of a single country matches the observed demographic and economic transitions in Great Britain. The parameters

$$\{\gamma, \phi, \tau_1, \tau_2, \bar{c}, \bar{n}, \upsilon, \xi, \alpha, \beta, \mu_f^a, \mu_f^b, \mu_f^m, B_0, \delta\}$$

determine preferences for and the cost of children, human capital production, the shares of

[23] As highlighted by Cummins (2013) and Gay et al. (2022), fertility decline in early modern France, another frontrunner, can be due to idiosyncratic cultural and legal factors such as changes in the inheritance law distinct from, and possibly complementary to, the quantity-quality trade-off.

[24] Historically, the labor force participation of the elderly was low and most worked less than full time, and typically in lower-paid jobs. For example, in 1891, 64.8% of males above 65 were in the labor force (Boyer and Schmidle, 2009).

Table 5: Parameters set exogenously

Description	Parameter	Value
Technology		
Complement of land share, ancient	ρ_a	0.7
Complement of land share, modern	ρ_b	0.9
Initial level of A_t	A_0	1
Initial level of M_t	M_0	1
Labor endowment of old adults	ζ^2	1
Labor endowment of elderly	ζ^3	$\frac{1}{2}$
Initial prob. that infants survive to be children (age 1-20)	s_0^0	0.685
Initial prob. that children survive to be young adults (age 21-40)	s_0^1	0.752
Initial prob. that young adults survive to be old adults (age 41-60)	s_0^2	0.620
Initial prob. that old adults survive to be elderly (age 61-80)	s_0^3	0.344

skilled and unskilled labor in each sector, and the growth rates of A_t, B_t, and M_t in the post-Malthusian era, respectively.

A model period is 20 years and, motivated by the evidence in Broadberry et al. (2015), we set $\tau = 1690$. To compare model and data targets, we take a moving average of the data where for any variable X_t, the data target is calculated as a moving average of values from $t - 10$ to $t + 9$. Hence, the data targets for 1690 are an average of the data from 1680 to 1699.

Starting from $\tau = 1690$, we simulate the model economy moving forward and choose these fifteen parameters to match the following moments:

1. The levels of the CDR and CBR between 1690 and 2010 (Figure 11). For the pre-transition period, we assume that CBR and CDR are constant and equal to their 1690 values.[25]

2. The share of labor employed in the ancient sector: 85.5% in 1690 and 30% in 1890. These shares correspond to the fraction of England's population living in rural areas in these two years, according to Bairoch (1991). Figure 12 plots our simulated ancient sector share (dashed line) against Bairoch's (1991) data for the fraction of the population living in rural areas. The simulated path is broadly consistent with these data even though we only target two points.

3. The GDP per capita between 1690 and 2010 (Figure 13).

4. The years of education between 1870 and 2010 (Figure 14). The data on educational attainment are taken from the Lee and Lee (2016) data set. The average years of enrollment in Great Britain was only 1.0 in 1870, after which the average grew very rapidly, reaching 6.6 by 1950 and 11.4 by 2010.

[25]For the end of transitions, we observe CDR_{2010} and CBR_{2010}. We assume that CDR and CBR are 12.5 after 2070. Between 2010 and 2070 (three model periods), we asume that they decline linearly.

Table 6: Calibrated parameters, first stage

Description	Parameter	Value
Utility Function		
Utility weight for fertility	γ	0.583
Parental altruism, warm glow	ϕ	0.640
Minimum consumption as fraction of wage	$\tilde{c}$	0.128
Minimum fertility	$\overline{n}$	0.172
Cost of Children		
Quantity	τ_1	0.092
Quality	τ_2	0.032
Technology		
Unskilled labor share, ancient	α	0.69
Unskilled labor share, modern	β	0.098
Growth rate of A_t	μ_f^a	0.54% (yearly)
Growth rate of B_t	μ_f^b	0.61% (yearly)
Growth rate of M_t	μ_t^m	0.67% (yearly)
Medical technology lag	δ	3.87
Dynamic complementarity of human capital	υ	0.483
Elasticity of education effort to human capital	ξ	0.671

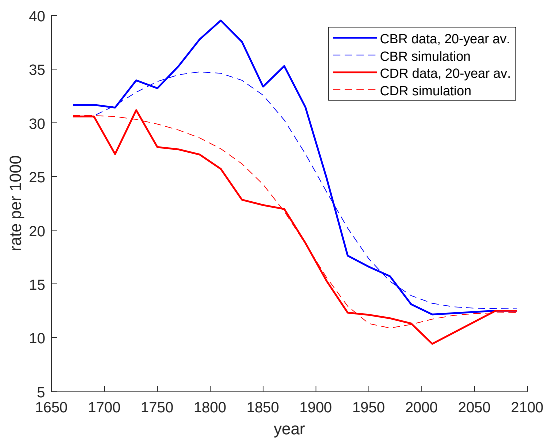


Figure 11: GB CBR/CDR, sim vs. data

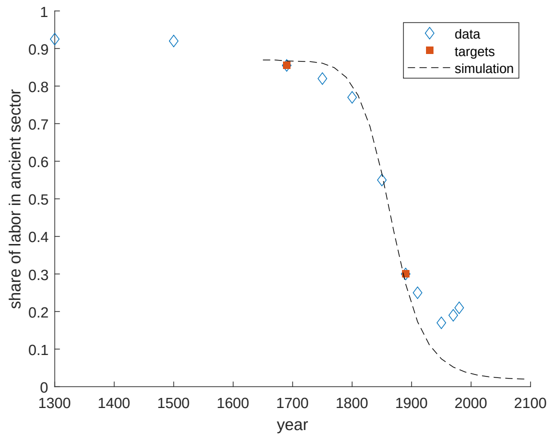


Figure 12: GB ancient sector share, sim vs. data

Figures 11 to Figure 14 show that the model does an excellent job matching these targets. Table 6 reports the parameter values calibrated in this first stage. Our estimates imply that each child reduces the available time for work by around 9%, and each year of education for a surviving child reduces it by another 3%. For example, having five surviving children and providing them with ten years of schooling would leave parents with almost no time for work. Our estimates for exogenous technological change in the ancient and modern sectors are 0.54% and 0.61% per year. Yet, the model generates the observed growth in GDP per capita due to endogenous human capital accumulation. The estimated value for δ implies that if the initial survival rate was around 0.6 (recall that a model period is 20 years), doubling the medical technology from its initial value of 1 would increase the survival rate by 50% to around 0.9. With an annual growth rate of 0.67%, this would take around 100 years. Finally, the model detects a large gap between α and β, the shares of unskilled labor in the ancient and modern sectors. The calibrated value for α, 0.69, implies that the ancient sector uses very little skilled labor, an intuitive result if we interpret this sector as (mostly) farmers, servants, and small-scale low-skill artisans such as bakers and bricklayers.

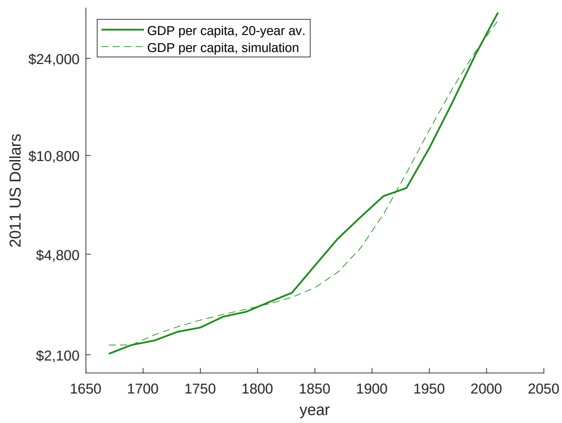


Figure 13: GB GDP per capita, sim vs. data

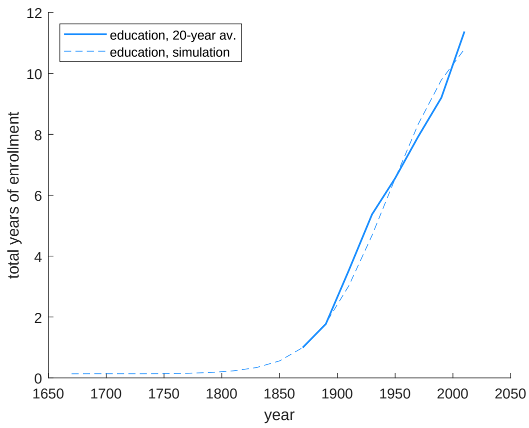


Figure 14: GB years of education, sim vs. data

The calibrated differences between α and β are key to understanding why our model replicates the demographic transition of Great Britain (Figure 11). As TFP starts to grow, it does so faster in the modern sector than in the ancient sector. Thus, the modern sector grows more quickly (Figure 12) and, because of its higher skill intensity, the skill premium rises (equation 15). A higher skill premium leads, through the quality-quantity trade-off, to an increase in education (Figure 14), a fall in births, and sustained growth in GDP per capita (Figure 13).[26]

We test the importance of the change in the skill premium by running a counterfactual simulation in which the premium does not change. We achieve this by assuming that the two sectors are identical: $\alpha = \beta = 0.69$, and $\rho_a = \rho_b = 0.7$.[27] Figures 15 and 16 illustrate the result. With no change in the skill premium, there is no increase in education. The CBR only falls to around 25 instead of 12. Interestingly, we also see that growth in GDP per capita levels off in the middle of the 19$^{\text{th}}$ century and then reverses. This is consistent with Malthusian theory and with the importance many scholars have placed on education in the escape from the "Malthusian trap" (Becker et al., 1990).

It is also instructive to analyze the reasons behind the small reduction in fertility seen in Figure 16. This is partly due to better medical technology reducing the need for "extra" births to hit parents' preference to have a minimum number of surviving children, represented

[26] The increase in medical technology also lowers births, as fewer births are required to obtain the same level of surviving children, but this effect is smaller.

[27] For simplicity, we also assume $A = B$, but this is immaterial for our point.

in our model by the parameter $\bar{n}$. Another part of the effect is Malthusian. Lower death rates mean a higher population growth rate, which causes the land/labor ratio to fall faster. In this simulation, decreasing returns to labor overwhelm TFP growth from the middle of the 19^{th} century. This leads to a fall in GDP per capita back toward the minimum consumption constraint $\bar{c}$, contributing further to the fall in fertility.

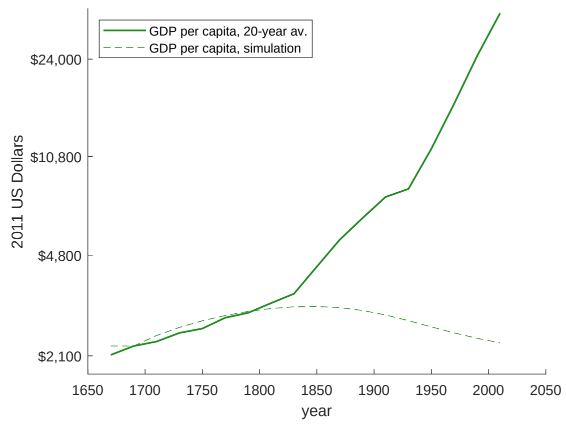


Figure 15: GB GDP per capita, no rise in skill premium

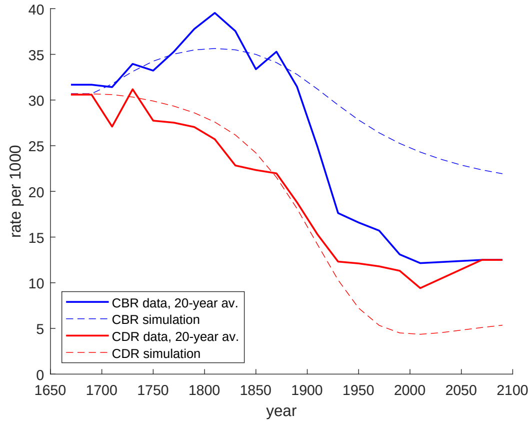


Figure 16: GB vital statistics, no rise in skill premium

6.2 Global demographic transition

Taking the parameters calibrated in the first stage as given, in the second stage we choose six parameters that govern the process of technology diffusion, $\{\lambda, d_b, d_a, d_m, \varphi, \theta\}$. Among these parameters, φ determines how fast the distances between countries shrink once Great Britain starts to grow, while λ determines the elasticity of the effective distance to kilometers of the great-circle distance between each country's capital city and London. The values d_j for $j = a, b, m$ determine the cost of distance as a barrier to technological diffusion in the ancient, modern, and medical sectors, respectively. Finally, θ is the elasticity of catch-up growth to backwardness.

To discipline these parameters, we simulate the model and calculate CBR and CDR levels as well and GDP per capita for all countries in our sample after 1690. We choose these six parameters to match the following targets:

1. The global average CBR and CDR between 1950 and 2017 (Figure 17).
2. The global average GDP per capita between 1950 and 2017 (Panel (a) of figure 18).
3. The cumulative fraction of the world's population that lives in countries that have permanently crossed below $CBR = 25$ per 1000, representing the definitive start of the fertility transition (Panel (b) of figure 18).

We construct the targets of the calibration by weighting each country by its population in 2016, according to World Bank data. The measure of effective distance only accounts for geographic distance. While other determinants of technology diffusion may be relevant (such as linguistic distance), we lack obvious targets to calibrate additional parameters, and the calibrated model conforms well to the data with this single distance measure. The calibrated parameter values are given in Table 7.

Table 7: Calibrated parameters, second stage

Description	Parameter	Value
Distance		
elasticity of effective distance to km of geographic distance	λ	5.13
log cost of distance for anc. tech. diffusion	$\ln d_b$	-26.7
log cost of distance for non-anc. tech. diffusion	$\ln d_a$	-26.4
log cost of distance for medical tech. diffusion	$\ln d_m$	-23.1
growth rate of cost of physical distance	φ	-10.3% (yearly)
elasticity to backwardness	θ	1.83

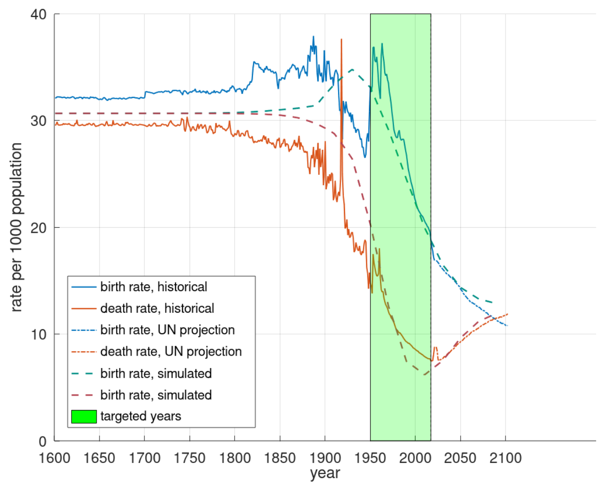
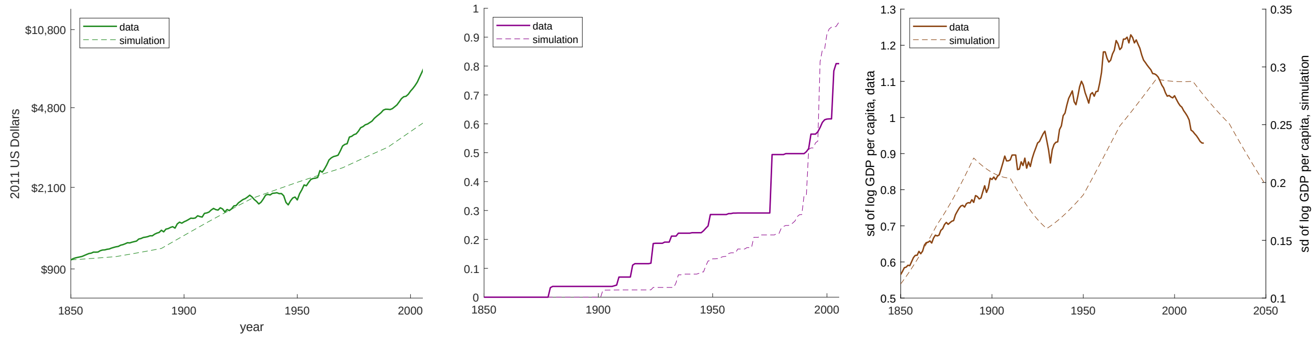


Figure 17: World CBR and CDR, 1600-2100

Figure 17 and panel (c) of 18 show that the model replicates the world demographic transition and the evolution of the global mean of GPD per capita reasonably well. But those were targeted moments to learn about the parameters controlling technological diffusion. How well does the model match non-targeted moments?

We assess the performance of the model by looking, in panel (c) of 18, at the standard deviation of log GDP per capita for the whole world between 1950-2017. Despite this not being a targeted moment, the model does a nice job of replicating the initial increase and subsequent decline in inequality across countries. The level of inequality is smaller in the model economy since we assume that countries were identical in 1690. It would not be difficult to add initial differences in the level of technology to account for cross-country income differences in 1690.

Figure 18: Model vs. targets



Panel (a); Global mean GDP per cap. Panel (b): Frac. of world pop. in countries <25 births/1000 Panel (c): Global st. dev. of log GDP per capita

The most interesting test, however, is to gauge how well the model replicates the demographic transition for individual countries. This is a powerful measure of the strength of the model as we do not use data from any particular country (except Great Britain) beyond its distance to the frontier and that *all* the calibrated parameters –e.g., for parental altruism, cost of children, elasticity to backwardness– are the same across countries.

Figure 19 shows the model vs. the data for the same set of countries that appear in Figure 3 (in the next section, we will explore all countries more systematically). The model does surprisingly well accounting for the demographic transitions of Denmark and Spain, a fairly good job with Chile and Malaysia, but it fails with Chad.

We take Figure 19 as a strong validation of the model. For countries like Denmark and Spain, close to Great Britain geographically and culturally, a simple quantity-quality trade-off together with technological diffusion can account for a large part of the observed demographic transitions. The model, obviously, does not fit the data perfectly. Other mechanisms (culture, social norms, legislation, taxes, migration, labor market regulations) play a role in fertility decisions and mortality. Therefore, it is not surprising that we miss an important part of the action in Chad, a country with economic and social institutions very different from Great Britain's. For example, cultural norms might cause different levels of parental altruism across countries. But Figure 19 strongly suggests that the trade-offs our model highlights are of first-order importance.

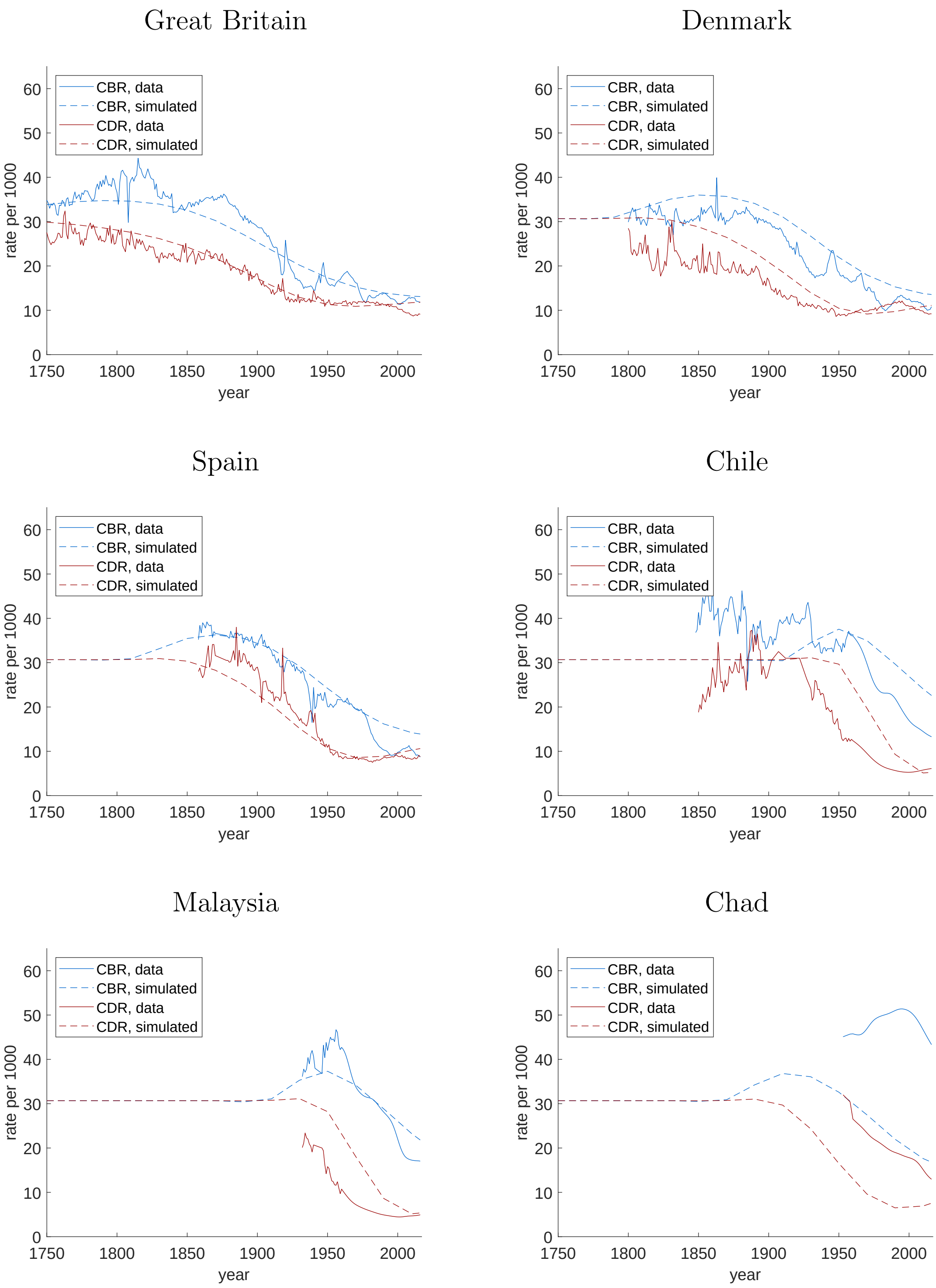


Figure 19: Six examples of demographic transitions, compared with simulation

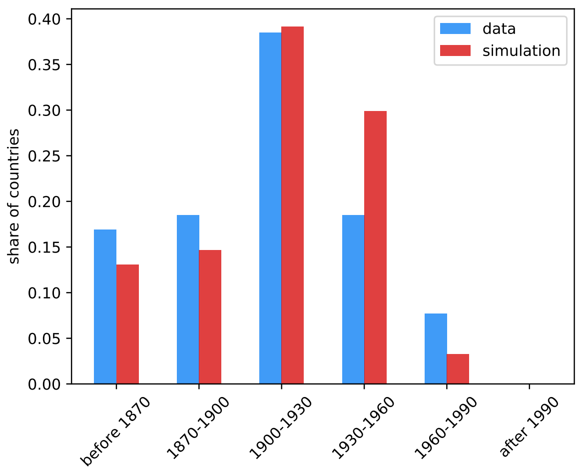
7 Assessing the model

How general are the results of Figure 19? Does our model generate demographic transitions that look like those in the data for the cross-section of countries? Does the model generate the four findings that emerged from Section 3? And does the model generate changes in education levels that resemble those in the data? This section assesses the model along these dimensions.

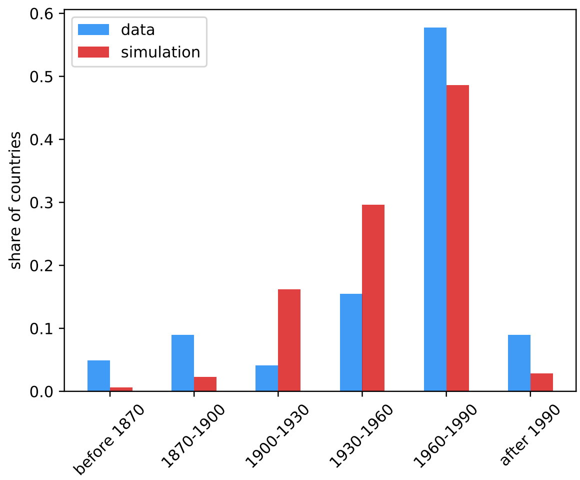
7.1 The past and present of demographic transitions

The first finding from Section 3 was that the start dates of the CDR transitions are more dispersed over time than the start dates of the CBR transitions. Panels (a) and (b) of 20 compare the start dates of the mortality and fertility transitions for the last 300 years in the data and the simulation. The model does an excellent job of matching the data in terms of timing and range: the CDR transitions' start dates are more dispersed than the start dates of the CBR transitions. The model, however, underestimates the number of very early transitions, which is not a surprise since some of those might have been triggered by factors we do not consider.

Figure 20: Transition starts over time



Panel (a): Crude Death Rate



Panel (b): Crude Birth Rate

The second finding from Section 3 was that transitions in both fertility and mortality have been getting faster over time. Figures 21 and 22 corroborate that our model replicates this observation (the plotted line is a linear regression, blue for data and red for simulation). We match the regression intercept and slope of the regression nearly perfectly for the CDR transition and the slope of the CBR transition while missing the intercept by around 25 years. Each country experiences a growth take-off as effective distance falls, with countries closer to Great Britain taking off first. Catch-up growth is, however, faster in countries that join the growth process later, as they have a wider gap in TFP to close. Since the increase in the skill premium and the associated rise in education levels are sharper in later-transitioning countries, the fall in fertility is also more rapid, and the overall transition period is shorter. For transitions starting in the 19th century, the CDR and CBR took more than 100 years to fall by 20 points (the pre-transition levels were around 30 per 1000 for the CDR and 40 for the CBR). For the transitions in the 20th century, on the other hand, from similar pre-transition levels, the required time for a 20-point decline was around 50 years.

The third finding from Section 3 was that average GDP per capita at the start of the CDR and CBR transitions is similar across time. Figure 23 shows the level of GDP per capita at the beginning of the observed (blue bars) and simulated (red bars) CDR and

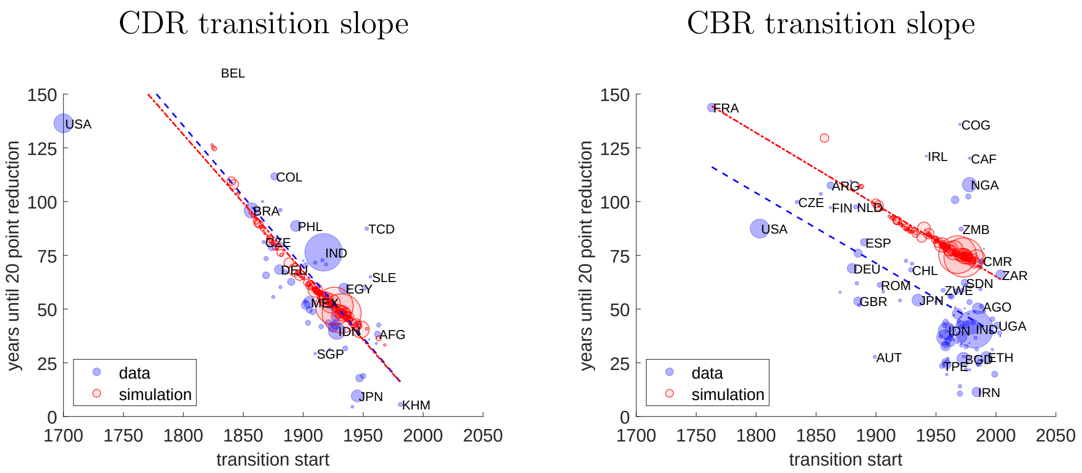


Figure 21: Transition slopes

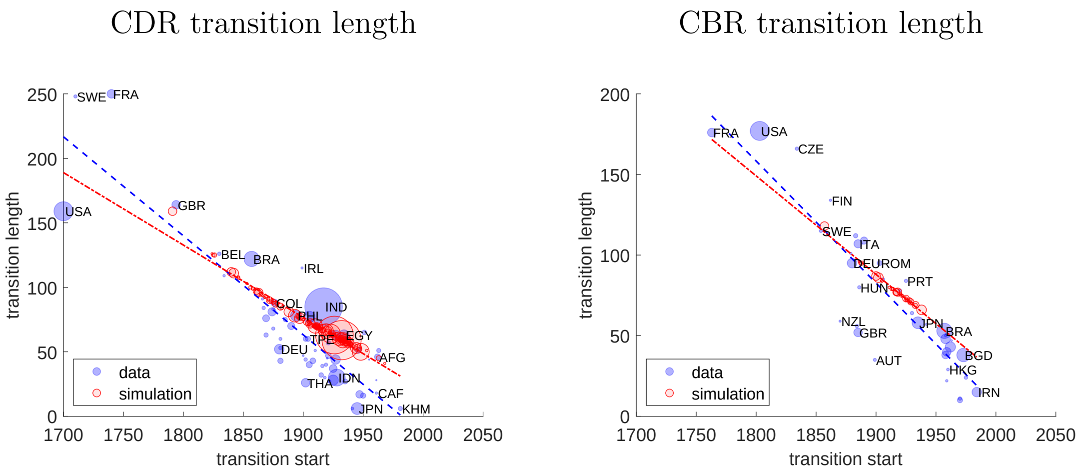


Figure 22: Transition lengths

CBR transitions. There is a more significant variance in GDP per capita at the start of the transitions for the later years, particularly for the CBR transitions of the 1960-2000 period. The model performs quite well at capturing the observed distribution of transitions, although the match is not perfect. This is not a surprise, as there are many country-specific factors behind the demographic transitions, which we purposefully left out of the analysis. For example, oil revenues for Oman and Iran allowed these countries to reach high levels of GDP per capita at a relatively lower level of the technology adoption that drives our model. Soviet-style regimes in countries like Romania also experienced lower fertility for reasons we do not model. Furthermore, as highlighted by De Silva and Tenreyro (2020), several low-income countries achieved lower fertility rates due to population control policies introduced in the 1960s and 1970s.

The fourth fact, demographic contagion, was explicitly incorporated into the model, and so is replicated by construction.

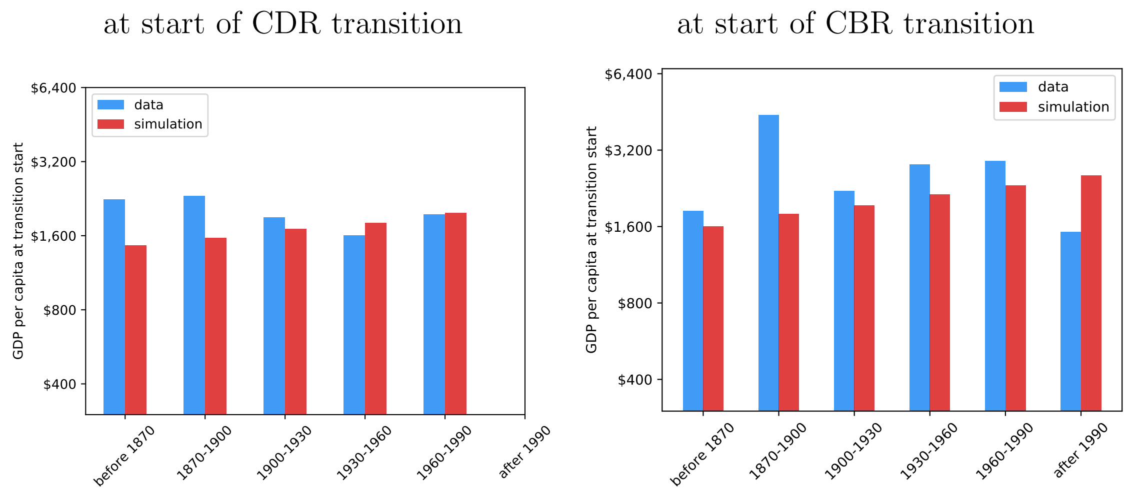


Figure 23: Log GDPpc at the start of transitions

7.2 Education and fertility

Next, we take a closer look at the relationship between education and fertility. Our model predicts that countries that reduce fertility faster will also increase years of education more quickly. As we have not used any information on trends in education to inform our model or its calibration so far, this presents two excellent tests of the model's performance: first, whether the qualitative pattern predicted by the model exists in the data, and second, how well the model matches the data on education and fertility. Lee and Lee (2016) provide data on average years of schooling for 110 countries at 5-year intervals from 1870 to 2010. For each country, we calculate a measure of speed in educational attainment by dividing the total increase in years of schooling during the CBR transition by the total number of years that the CBR transition is observed. A higher number indicates a faster gain in years of schooling during the CBR transition. We compare the speed of increase in educational attainment to the speed of the demographic transition over the same period, measured as the number of years per 20 point decline in CBR.

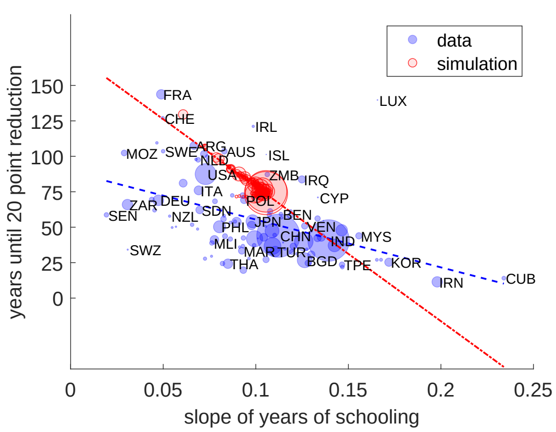


Figure 24: Fertility and education, initial rates of change

Figure 24 plots the speed of the change in educational attainment against the speed of CBR decline, both for the data and our model. The data do indeed show a robust positive relationship between the speed of the fertility transition and the speed of educational progress, consistent with the quantity-quality trade-off at the center of our model. Quantitatively, the slope of the best-fit relationship is a bit steeper in our simulation than in the data, and our model does not produce examples of very rapid education increases such as those observed for many late-developing countries such as South Korea and India. That may be partly because these countries embarked on aggressive government-sponsored education programs that may have changed the costs faced by parents (e.g., by redistributing the cost of education across income levels) or even pushed some parents away from their optimal quantity-quality mix.[28]

Overall, we judge our model to be successful at replicating the main patterns the reduction in fertility and the increases in education in the data. Despite the lack of any country-specific details regarding educational policy, our model captures much of the co-movement in the data.

8 Conclusions

In this paper, we have constructed a data set consisting of birth rates, death rates, and GDP per capita for a panel of 186 countries spanning from 1735 until 2014. We have proposed a way to measure demographic transitions that lets the data pick start and end dates for fertility and mortality transitions. Our method documented several important findings regarding when and how quickly countries go through the demographic transition. We highlighted, in particular, the existence of a "demographic contagion" across countries that are close to each other (either geographically or linguistically). Finally, we argued that a simple model where parents choose child quantity and educational quality and where technology diffuses from a frontier country to the periphery can account for all of these observations.

Bibliography

Ayyar, R. (2017). *History of Education Policymaking in India, 1947–2016*. OUP India.

Bairoch, P. (1991). *Cities and Economic Development: From the Dawn of History to the Present*. University of Chicago Press.

Bakker, D., A. Müller, V. Velupillai, S. Wichmann, C. Brown, P. Brown, D. Egorov, R. Mail-

[28]For South Korea, see Seth (2002). For India, and in particular, the role of the Kothari Commission (1964-1966), see Ayyar (2017). Our data set also includes cases such as Cuba, where wages (and hence the skill premium) and educational forces are not determined by the decisions of private agents.

hammer, A. Grant, and E. Holman (2009). Adding typology to lexicostatistics: a combined approach to language classification. *Linguistic Typology 13*(1), 167–179.

Bar, M. and O. Leukhina (2010). Demographic transition and industrial revolution: A macroeconomic investigation. *Review of Economic Dynamics 13*(2), 424–451.

Barro, R. and G. Becker (1989). Fertility choice in a model of economic growth. *Econometrica 57*(2), 481–501.

Becker, G. and R. Barro (1988). A reformulation of the theory of fertility. *Quarterly Journal of Economics 103*(1), 1–25.

Becker, G. and H. Lewis (1973). On the interaction between the quantity and the quality of children. *Journal of Political Economy 81*(2), S279–S288.

Becker, G. S. (1960). An Economic Analysis of Fertility. In *Demographic and Economic Change in Developed Countries*, pp. 209–240. National Bureau of Economic Research.

Becker, G. S., K. M. Murphy, and R. Tamura (1990). Human capital, fertility, and economic growth. *Journal of Political Economy 98*(5), S12–S37.

Bolt, J., R. Inklaar, H. de Jong, and J. L. van Zanden (2018). Rebasing 'Maddison': New income comparisons and the shape of long-run economic development. Technical Report 10, Maddison Project Working Paper.

Boyer, G. R. and T. P. Schmidle (2009). Poverty among the elderly in late Victorian England. *Economic History Review 62*, 249–78.

Broadberry, S., B. Campbell, A. Klein, M. Overton, and B. van Leeuwen (2015). *British Economic Growth, 1270-1870.* Cambridge University Press.

Buera, F. J., A. Monge-Naranjo, and G. E. Primiceri (2011). Learning the wealth of nations. *Econometrica 79*(1), 1–45.

Casterline, J. B. (2001). The pace of fertility transition: National patterns in the second half of the twentieth century. *Population and Development Review 27*, 17–52.

Cavalcanti, T., G. Kocharov, and C. Santos (2021). Family planning and development: Aggregate effects of contraceptive use. *Economic Journal 131*(634), 624–657.

Cervellati, M., G. Meyerheim, and U. Sunde (2023). The empirics of economic growth over time and across nations: a unified growth perspective. *Journal of Economic Growth 28*(2), 173–224.

Cervellati, M. and U. Sunde (2015). The economic and demographic transition, mortality, and comparative development. *American Economic Journal: Macroeconomics 7*(3), 189–225.

Chesnais, J.-C. (1992). *The Demographic Transition: Stages, Patterns, and Economic Implications.* Oxford University Press.

Coale, A. J. and S. C. Watkins (1986). *The Decline of Fertility in Europe.* Princeton University Press.

Comín, D. and B. Hobijn (2010). An exploration of technology diffusion. *American Economic Review 100*(5), 2031–59.

Comín, D. and M. Mestieri (2018). If technology has arrived everywhere, why has income diverged? *American Economic Journal: Macroeconomics 3*(10), 137–178.

Córdoba, J. C., X. Liu, and M. Ripoll (2020). Accounting for the international quantity-quality trade-off. Technical report, Working Paper, Iowa State University.

Cummins, N. (2013). Marital fertility and wealth during the fertility transition: rural France, 1750-1850. *Economic History Review 66*(2), 449–476.

Davis, K. (1946). Human fertility in India. *American Journal of Sociology 52*(3), 243–254.

De la Croix, D. and F. Perrin (2018). How far can economic incentives explain the French fertility and education transition? *European Economic Review 108*, 221–225.

De Silva, T. and S. Tenreyro (2020). The fall in global fertility: A quantitative model. *American Economic Journal: Macroeconomics 12*(3), 77–109.

Desmet, K. and J. Rappaport (2017). The settlement of the United States, 1800 to 2000: The long transition towards Gibrat's law. *Journal of Urban Economics 98*, 50–68.

Doepke, M. (2017). Accounting for fertility decline during the transition to growth. *Journal of Economic Growth 9*(3), 347–383.

Doepke, M., A. Hannusch, F. Kindermann, and M. Tertilt (2022). The economics of fertility: A new era. Working Paper 29948, NBER.

Dyson, T. and M. Murphy (1985). The onset of fertility transition. *Population and Development Review 11*(3), 399–440.

Edvinsson, R. B. (2015). Recalculating Swedish pre-census demographic data: Was there acceleration in early modern population growth? *Cliometrica 9*(2), 167–191.

Egger, P. H. and A. Lassmann (2012). The language effect in international trade: A meta-analysis. *Economics Letters 116*(2), 221–224.

Fernández, R. (2013, February). Cultural change as learning: The evolution of female labor force participation over a century. *American Economic Review 103*(1), 472–500.

Fernández-Villaverde, J. (2001). Was Malthus right? Economic growth and population dynamics. Technical report, Working Paper, University of Pennsylvania.

Fogli, A. and L. Veldkamp (2011). Nature or nurture? learning and the geography of female labor force participation. *Econometrica 79*(4), 1103–1138.

Galor, O. and D. N. Weil (1996). The gender gap, fertility, and growth. *American Economic Review 86*(3), 374–387.

Galor, O. and D. N. Weil (1999). From the Malthusian regime to modern growth. *American Economic Review 89*, 150–154.

Galor, O. and D. N. Weil (2000). Population, technology, and growth: From Malthusian stagnation to the demographic transition and beyond. *American Economic Review 90*(4), 806–828.

Gay, V., P. E. Gobbi, and M. Goñi (2022). Revolutionary transition: Inheritance change and fertility decline. Technical report, Working Paper.

Greenwood, J. and A. Seshadri (2002). The U.S. demographic transition. *American Economic Review 92*(2), 153–159.

Hansen, G. D. and E. C. Prescott (2002). Malthus to Solow. *American Economic Review 92*(4), 1205–1217.

Hejkal, J., B. Ravikumar, and G. Vandenbroucke (2022). Technology adoption, mortality, and population dynamics. Working paper.

Jones, C. (2001). Was an industrial revolution inevitable? Economic growth over the very long run. *Advances in Macroeconomics 1*(2), 1–45.

Jones, L. E., A. Schoonbroodt, and M. Tertilt (2010, November). Fertility theories: Can they explain the negative fertility-income relationship? In J. B. Shoven (Ed.), *Demography and the Economy*, pp. 43–100. University of Chicago Press.

Kalemli-Ozcan, S. (2003). A stochastic model of mortality, fertility, and human capital investment. *Journal of Development Economics 70*(1), 103–118.

Klein Goldewijk, K., A. Beusen, G. van Drecht, and M. de Vos (2011). The HYDE 3.1 spatially explicit database of human-induced global land-use change over the past 12,000 years. *Global Ecology and Biogeography 20*(1), 73–86.

Lee, J.-W. and H. Lee (2016). Human capital in the long run. *Journal of Development Economics 122*, 147–169.

Lee, R. (2003, December). The demographic transition: Three centuries of fundamental change. *Journal of Economic Perspectives 17*(4), 167–190.

Lucas, R. E. (1988). On the mechanics of economic development. *Journal of Monetary Economics 22*(1), 3 – 42.

Lucas, R. E. (2002). *Lectures on Economic Growth.* Harvard University Press.

Lucas, R. E. (2009). Trade and the diffusion of the industrial revolution. *American Economic Journal: Macroeconomics 1*(1), 1–25.

Maines, M. and R. H. Steckel (2000). *A Population History of North America.* Cambridge University Press.

Malthus, T. (1993). *An Essay on the Principle of Population.* Oxford University Press.

Manuelli, R. E. and A. Seshadri (2009). Explaining International Fertility Differences. *Quarterly Journal of Economics 124*(2), 771–807.

Mayer, T. and S. Zignago (2011). Notes on CEPII's distances measures: The geodist database. Working Papers 2011-25, CEPII.

Melitz, J. and F. Toubal (2013). Native language, spoken language, translation and trade. *Journal of International Economics 93*(2), 351–363.

Mitchell, B. R. (2013). *International Historical Statistics: 1750-2010.* Palgrave MacMillan.

Morland, P. (2019). *The Human Tide: How Population Shaped the Modern World.* John Murray Publishers.

Murtin, F. (2013). Long-run determinants of the demographic transition. *Review of Economics and Statistics 95*(2), 617–631.

National Central Bureau of Statistics (1969). *Historical Statistics of Sweden. Part 1: Population* (2 ed.). National Central Bureau of Statistics, Stockholm.

Reher, D. (2004). The demographic transition revisited as a global process. *Population, Space, and Place 10*(1), 19–41.

Schofield, R. S. and E. A. Wrigley (1989). *The Population History of England 1541-1871.* Cambridge University Press.

Seth, M. (2002). *Education Fever: Society, Politics, and the Pursuit of Schooling in South Korea.* University of Hawaii Press.

Shorter, F. and M. Macura (1982). *Trends in Fertility and Mortality in Turkey, 1935-1975.* National Academy Press.

Spolaore, E. and R. Wacziarg (2021, 11). Fertility and modernity. *Economic Journal 132*(642), 796–833.

State Statistical Institute of Turkey (1995). *The Population of Turkey, 1923-1994: Demographic Structure and Development : with Projections to the Mid-21st Century.* State Institute of Statistics, Prime Ministry, Republic of Turkey.

Swiss Federal Statistics Office (1998). *Two Centuries of Swiss Demographic History: Graphic Album of the 1860-2050 Period.* Federal Statistical Office.

Tamura, R. (2006, February). Human capital and economic development. *Journal of Development Economics 79*(1), 26–72.

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

Vogl, T. S. (2020). Intergenerational Associations and the Fertility Transition. *Journal of the European Economic Association 18*(6), 2972–3005.

Wrigley, E. A., R. S. Davies, J. E. Oeppen, and R. S. Schofield (1997). *English Population History from Family Reconstitution 1580-1837.* Cambridge University Press.

A Historical estimates of world vital statistics

We construct world average CDRs and CBRs from 1600 to 2016 using:

- Data on birth rates and death rates by country from the sources detailed in Section 3
- Data on population by country from the Maddison 2018 database (Bolt et al., 2018).

For the world average birth rate, we then proceed in three steps:

1. First, we linearly interpolate gaps in birth rate and population data for each country.
2. Then, for each of the 152 countries for which we observe the start of the fertility transition, we assume CBR is equal to the pre-transition mean between 1600 and the start of the transition.
3. Finally, we calculate the world average crude birth rate for each year as the population-weighted average of all countries that have both population data from the Maddison 2018 database and an observation, an interpolated value, or a backward-projected value for the CBR in that year.

Following the same process for CDRs as we did for CBRs would lead to an implied rate of pre-modern world population growth that is much higher than all available historical estimates. To avoid this problem, we follow a slightly modified process for CDRs:

1. First, we linearly interpolate gaps in the death rate and population data for each country.
2. Then, for each of the 44 countries for which we observe the start of the mortality transition, we project CDR backward from the start of the data to 1600 by assuming that it is equal to the CBR minus the annual population growth rate implied by the population data.[29]
3. Then, for the 96 countries for which we do not observe the start of the mortality transition but for which we can impute a transition start date using the method described in Section 3, we project CDRs backward from the start of the data until the imputed start of the CDR transition by assuming it is equal to the transition mean.
4. Then, for each of these 96 countries, we project CDRs backward from the imputed transition start date to 1600, by assuming that it is equal to the CBR minus the annual population growth rate.
5. Finally, we calculate the world average CBR for each year as the population-weighted average of all countries that have both population data and an observation, an interpolated value, or a backward-projected value for the CDR in that year.

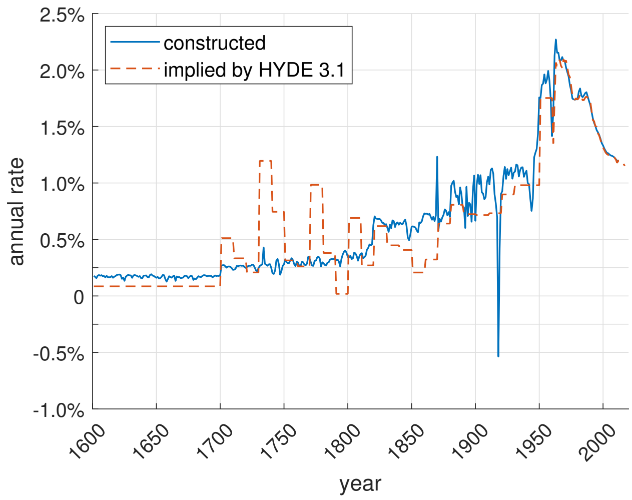


Figure 25: World population growth, comparison

The world average rate of population growth is calculated as the difference between CBRs and CDRs. The total number of annual births is calculated by multiplying the world average CBR by the total world population, taken from the HYDE 3.1 database (Klein Goldewijk et al., 2011). Figure 25 compares the constructed annual population growth rates to those implied by the world population data in the HYDE 3.1 database.

B Auxiliary rules for model selection

B.1 Auxiliary rules of transition starts

A statistically detected CDR transition start date is removed, moving the country from case 1 to case 2, or case 3 to case 4, if one or more of the following conditions hold:

1. Estimated initial CDR level of less than 25, less than 20 years after the start of the series.
2. Estimated initial CDR level of less than 15, regardless of timing.
3. Estimated initial CDR level more than 20 points below the initial level of CBR, regardless of timing.

A CDR transition start date is added, moving the country from case 2 to case 1, or case 4 to case 3, if both of the following conditions hold:

1. Estimated initial CDR level greater than 35.
2. CDR start date has not been previously removed by the first set of rules.

[29] In the pre-modern era, net migration was pretty close to zero everywhere and, in any case, we are computing a global average CDR. Earth has a net migration of zero.

A statistically detected CBR transition start date is removed, moving the country from case 1 to case 2, or case 3 to case 4, if one or more of the following conditions hold:

1. Estimated initial CBR level of less than 30, less than 20 years after the start of the series.
2. Estimated initial CBR level of less than 20, regardless of timing.

A CBR transition start date is added, moving from case 2 to case 1, or case 4 to case 3, if both of the following conditions hold:

1. Estimated initial CBR level greater than 50.
2. CBR start date has not been previously removed by the first set of rules.

B.2 Auxiliary rules of transition ends

A statistically detected CDR transition end date is removed, moving the country from case 1 to case 3, or case 2 to case 4, if one or more of the following conditions hold:

1. Estimated final CDR level of greater than 20, less than 20 years after the start of the series.
2. Estimated final CDR level greater than 25, regardless of timing.

A CDR transition end date is added, moving the country from case 3 to case 1, or case 4 to case 2, if both of the following conditions hold:

1. Estimated final CDR level less than 12.
2. CDR end date has not been previously removed by the first set of rules.

A statistically detected CBR transition end date is removed, moving the country from case 1 to case 3, or case 2 to case 4, if one or more of the following conditions hold:

1. Estimated final CBR level of greater than 20, less than 20 years before the end of the series.
2. Estimated final CBR level of greater than 25, regardless of timing.

A CBR transition end date is added, moving the country from case 3 to case 1, or case 4 to case 2, if both of the following conditions hold:

1. Estimated final CBR less than 12.
2. CBR end date has not been previously removed by the first set of rules.

C Comparison with alternative fertility transition start

Figure 26 shows the correlation between our estimated start of the crude birth transition and a method close to that propsed by (Chesnais, 1992).

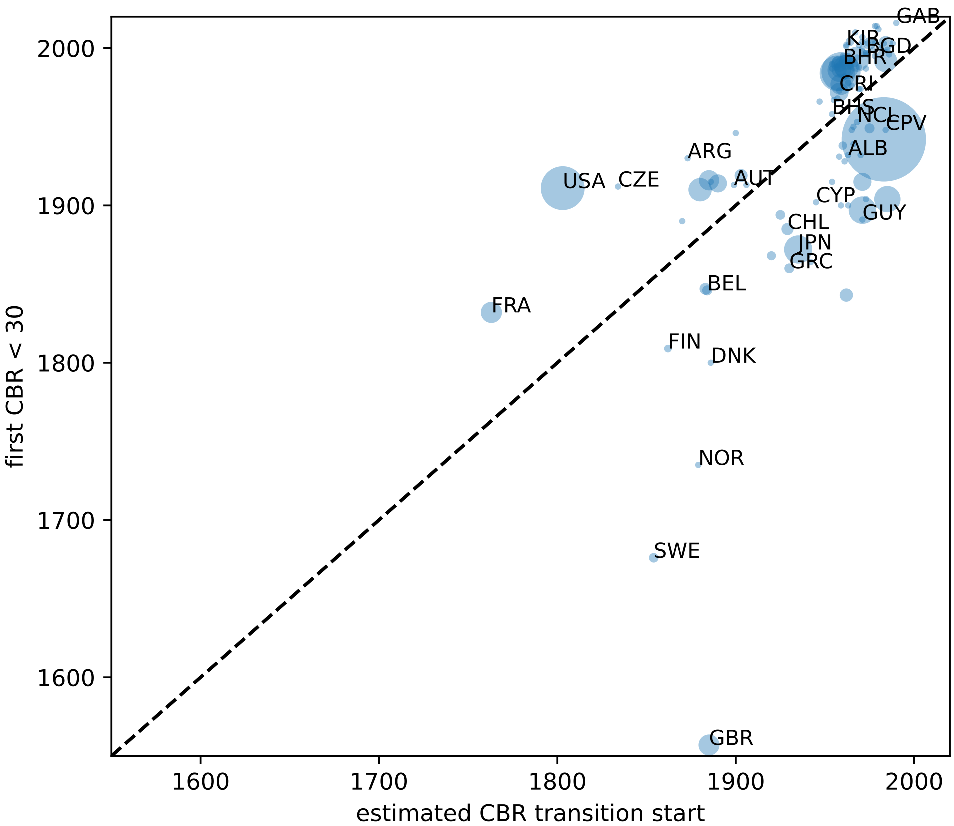


Figure 26: Comparison of estimated crude birth transition starts with Chesnais rule

Note: From authors' calculations. Circles proportion to 2016 population.

D Supplementary tables

A CDR calculated by projecting backward using the method described in Section 2 is indicated by ⋆.

Calculated Transition Start and End Dates				
	CDR		CBR	
Country	Start	End	Start	End
Afghanistan	1941⋆	2009	1999	n/a
Albania	1900⋆	1977	1963	2010
Algeria	1919⋆	1993	1965	n/a
Angola	1930⋆	2016	1988	n/a
Argentina	1869	1945	1862	n/a

Calculated Transition Start and End Dates				
	CDR		CBR	
Country	Start	End	Start	End
Armenia	n/a	n/a	n/a	2001
Australia	n/a	1961	n/a	1987
Austria	1881	1941	1899	1934
Azerbaijan	n/a	1988	n/a	1999
Bahamas, The	1918*	1967	1954	n/a
Bahrain	1918*	1979	1960	2011
Bangladesh	1910*	2004	1973	2011
Barbados	1923	1957	1954	1987
Belarus	n/a	n/a	n/a	1998
Belgium	n/a	1956	1884	1940
Belize	1910*	1972	1981	n/a
Benin	1939*	2001	1987	n/a
Bhutan	1938*	2004	1977	2012
Bolivia	1910*	2011	1969	n/a
Bosnia and Herzegovina	n/a	1964	n/a	2000
Botswana	1913*	1977	1971	n/a
Brazil	1857*	1994	1957	2010
Brunei Darussalam	1904*	1974	1954	2007
Bulgaria	1918	1948	1906	1991
Burkina Faso	1951	2016	1997	n/a
Burundi	1880*	2016	1987	n/a
Cambodia	1981	1987	1985	n/a
Cameroon	1888*	2016	1988	n/a
Canada	n/a	1955	n/a	2009
Cape Verde	1893*	2000	1984	n/a
Central African Republic	1961	1979	1978	n/a
Chad	1953	n/a	n/a	n/a
Channel Islands	n/a	2016	n/a	2013
Chile	1921	1978	1929	n/a
China	n/a	1972	n/a	2005
Colombia	1876*	1990	1971	n/a
Comoros	1921*	1999	1980	n/a
Congo, Dem. Rep.	1892*	2016	2004	n/a
Congo, Rep.	1930*	1974	1970	n/a
Costa Rica	1878*	1982	1958	2008
Cote d'Ivoire	1927*	1981	1963	n/a

Calculated Transition Start and End Dates				
	CDR		CBR	
Country	Start	End	Start	End
Croatia	n/a	n/a	n/a	2002
Cuba	n/a	1946	1970	1981
Cyprus	1922	1955	1945	2010
Czechoslovakia	1867	1951	1834	2000
Denmark	1834	1943	1886	1982
Djibouti	1935⋆	1979	1978	n/a
Dominica	1915⋆	1975	1960	n/a
Dominican Republic	1903⋆	1981	1954	n/a
Ecuador	1885⋆	1992	1957	n/a
Egypt, Arab Rep.	1934	1997	1968	n/a
El Salvador	1877⋆	1996	1968	n/a
Equatorial Guinea	1947⋆	2009	1997	n/a
Eritrea	1914⋆	2015	1967	n/a
Estonia	n/a	n/a	n/a	2001
Ethiopia	1919⋆	2016	1992	n/a
Fiji	1866⋆	1976	1964	n/a
Finland	1866	1957	1862	1996
France	1740	1990	1763	1939
French Polynesia	1861⋆	1987	1956	n/a
Gabon	1961	1989	1990	n/a
Gambia, The	1955	1999	1981	n/a
Georgia	n/a	1967	n/a	2000
Germany	1880	1932	1880	1975
Ghana	1881⋆	1996	1967	n/a
Greece	1916	1955	1930	1994
Grenada	1883⋆	1973	1957	2004
Guam	1946⋆	1950	1966	n/a
Guatemala	1902⋆	1999	1971	n/a
Guinea	1941⋆	2014	1990	n/a
Guinea-Bissau	1923⋆	2012	1991	n/a
Guyana (British Guiana)	1919	1962	1971	n/a
Haiti	1922⋆	2004	1983	n/a
Honduras	1913⋆	1992	1971	n/a
Hong Kong SAR, China	1941	1947	1960	1989
Hungary	1875	1943	1886	1966
Iceland	n/a	2006	1963	n/a

Calculated Transition Start and End Dates				
	CDR		CBR	
Country	Start	End	Start	End
India	1917	2002	1982	n/a
Indonesia	1928⋆	1983	1959	n/a
Iran, Islamic Rep.	1927⋆	1997	1984	1999
Iraq	n/a	1992	n/a	n/a
Ireland	1899	2014	1942	1999
Israel	n/a	1945	n/a	n/a
Italy	1874	1955	1885	1992
Jamaica	1920	1965	1965	n/a
Japan	1945	1951	1935	1993
Jordan	1922⋆	1980	1964	n/a
Kazakhstan	n/a	1971	n/a	1996
Kenya	1914⋆	1983	1975	n/a
Kiribati	1910⋆	1996	1962	n/a
Korea, Dem. Rep.	1950⋆	1969	1970	1980
Korea, Rep.	1947⋆	1970	1958	1996
Kuwait	n/a	1985	1968	n/a
Kyrgyz Republic	n/a	1992	n/a	n/a
Lao PDR	1915⋆	2012	1988	n/a
Latvia	n/a	n/a	n/a	2002
Lebanon	n/a	1972	n/a	2008
Lesotho	1924⋆	1981	1974	n/a
Liberia	1925⋆	2016	1982	n/a
Libya	1930⋆	1983	1967	n/a
Lithuania	n/a	n/a	n/a	2004
Luxembourg	n/a	2016	n/a	1978
Macao SAR, China	n/a	1970	n/a	1969
Macedonia, FYR	n/a	1967	n/a	2005
Madagascar	1916⋆	2012	1978	n/a
Malawi	1912⋆	2016	1981	n/a
Malaysia	1908⋆	1975	1958	n/a
Maldives	1936⋆	2000	1986	2001
Mali	1963	2014	2003	n/a
Malta	n/a	2000	n/a	2001
Mauritania	1916⋆	1989	1962	n/a
Mauritius	1930	1965	1958	2009
Mexico	1905	1982	1971	n/a

Calculated Transition Start and End Dates				
	CDR		CBR	
Country	Start	End	Start	End
Micronesia, Fed. Sts.	n/a	1986	1971	n/a
Moldova	n/a	1963	n/a	2007
Mongolia	1895*	2002	1965	n/a
Morocco	1905*	1993	1958	n/a
Mozambique	1924*	2016	1977	n/a
Myanmar	1925*	1990	1961	n/a
Namibia	1926*	1982	1977	n/a
Nepal	1946*	2004	1984	n/a
Netherlands	1869	1932	1883	1995
New Caledonia	1861*	1992	1968	2008
New Zealand	n/a	2016	1870	1929
Nicaragua	1900*	1996	1973	n/a
Niger	1917*	2016	1987	n/a
Nigeria	1897*	n/a	1978	n/a
Norway	n/a	1954	1879	1980
Oman	1934*	1991	1978	n/a
Pakistan	1918*	1994	1980	n/a
Panama	1859*	1982	1966	n/a
Papua New Guinea	1938*	1986	1967	n/a
Paraguay	n/a	1994	1950	n/a
Peru	1921*	1989	1962	n/a
Philippines	1894*	1981	1985	n/a
Poland	n/a	1957	n/a	2004
Portugal	1919	1959	1925	2009
Puerto Rico	1905*	1961	1947	2008
Qatar	n/a	1970	n/a	2013
Romania	1902	1962	1903	1998
Russian Federation	1891	1951	1900	1990
Rwanda	1881*	n/a	1984	n/a
St. Lucia	1899*	1978	1969	2010
St. Vincent and the Grenadines	1884*	1977	1961	2002
Samoa	n/a	1992	n/a	n/a
Saudi Arabia	1932*	1988	1974	n/a
Senegal	1931*	2001	1972	n/a
Serbia (Yugoslavia from 1900)	1875	1958	1920	1998
Seychelles	1874*	1980	1965	2001

Calculated Transition Start and End Dates				
	CDR		CBR	
Country	Start	End	Start	End
Sierra Leone	1956	n/a	1997	n/a
Singapore	1910	1961	1959	1981
Slovenia	n/a	2011	n/a	1998
Solomon Islands	1861*	2014	1979	n/a
Somalia	1915*	2016	2004	n/a
South Africa	n/a	1972	n/a	n/a
Spain	1890	1960	1890	1999
Sri Lanka	1935	1962	1962	n/a
Sudan	1862*	2010	1974	n/a
Suriname	n/a	1985	1963	n/a
Swaziland	1922*	1982	1978	n/a
Sweden	1710	1958	1854	1969
Switzerland	n/a	1953	n/a	1996
Syrian Arab Republic	1915*	1985	1975	n/a
Taiwan	1904*	1966	1955	n/a
Tajikistan	n/a	2012	1962	n/a
Tanzania	1870*	2016	1966	n/a
Thailand	1902*	1979	1959	1999
Togo	1928*	1987	1975	n/a
Tonga	n/a	1974	1963	n/a
Trinidad and Tobago	1897	1966	1961	2002
Tunisia	1881*	1999	1975	1999
Turkey	1927	1990	1958	2006
Turkmenistan	1869*	1992	1960	n/a
Uganda	n/a	2016	2001	n/a
Ukraine	n/a	n/a	n/a	1999
United Arab Emirates	n/a	1977	n/a	2010
United Kingdom	1794	1958	1885	1937
United States	1700*	1954	1803	1980
Uruguay	n/a	1939	n/a	1941
Uzbekistan	1861*	1995	1960	n/a
Vanuatu	n/a	1998	n/a	n/a
Venezuela, RB	1915	1975	1973	n/a
Vietnam	1925*	1981	1962	2005
Yemen, Rep.	1938*	1996	1986	n/a
Zambia	n/a	2016	1971	n/a

Calculated Transition Start and End Dates				
	CDR		CBR	
Country	Start	End	Start	End
Zimbabwe	1925*	1968	1956	n/a

E Extension of GDP per capita data

Our main source for GDP per capita data is the 2018 version of Maddison's database. While this database provides us with estimates for some countries going as far back as the year 1 CE, the time series for most countries does not start until the early 19th century or later, which is after many countries entered the CBR and CDR transitions. To allow the construction of a balanced panel for the empirical analysis in Section 3, we make a small number of cautious imputations of GDP per capita values for the year 1500. The set of countries in the Maddison database can be divided into four categories:

1. Countries that have a GDP per capita value for the year 1500.

2. Countries that do not have a GDP per capita value for the year 1500, but which have some value given between the years 1 and 1650.

3. Countries that do not have any GDP per capita value between the years 1 and 1650, but which have a value given between 1650 and 1900, which is not greater than $1,176.

4. All other countries.

There are 11 countries in category 1. There are also 11 countries in category 2. For these countries, we assign for the year 1500 the value of GDP per capita from the closest year prior to 1650. In doing so, we are taking advantage of the historical consensus that GDP per capita changed very slowly and exhibited close to zero long-run growth during the pre-modern era.

Category 3 is comprised of 26 countries. These countries have some data available for GDP per capita prior to the 20th century. Furthermore, based on these data, they were not at this point any richer than was England in the 13th century–the mean GDP per capita that the Maddison database gives for England from 1262-1312 is $1,176. There is little harm in assuming that these countries were in the pre-modern regime of no economic growth, and that their GDP per capita was the same in 1500 as it was in the first year we observe it. While this may not be exactly true, it is approximately so. Thus, for these countries we impute the earliest available value for GDP per capita to the year 1500.

Categories 1 through 3 are comprised of 48 countries. The remaining 138 countries in our data set belong to category 4. Some of these countries have GDP per capita estimates dating back to the 18th or 19th centuries, but these estimates are too high to presume that they pre-date the advent of modern economic growth. Some countries do not have any data for GDP per capita until well into the 20th century. For these countries, even if they appear quite poor during the first year of observation, we do not project their initial first GDP per capita observation all the way back from, say, 1950 or 1975 to the year 1500.

F Demographic contagion: Alternate epicenters

Table A2: Determinants of the start of the CBR transition (III)

	(1)	(2)	(3)
constant	-17.692	41.874	47.128
	(57.358)	(55.960)	(61.409)
ln GDP pc	5.733	6.507	6.833
	(4.628)	(4.612)	(4.667)
$(\ln \text{GDP pc})^2$	-0.337	-0.375	-0.398
	(0.288)	(0.286)	(0.290)
year	-0.014	-0.047**	-0.051*
	(0.024)	(0.023)	(0.027)
rln GC dist from GBR'	-4.257**		
	(1.404)		
r'ln GC dist from GBR × year'	0.009**		
	(0.003)		
ling. prox. to GBR	22.522**		
	(7.791)		
r'ling. prox. to GBR × year'	-0.044*		
	(0.023)		
ln GC dist from FRA		-5.302***	
		(1.528)	
ln GC dist from FRA × year		0.011**	
		(0.003)	
ling. prox. to FRA		-13.301	
		(12.142)	
ling. prox. to FRA × year		0.031	
		(0.029)	
ln GC dist from SWE			-5.997***
			(1.681)
ln GC dist from SWE × year			0.012**
			(0.004)
rling. prox. to SWE'			15.847
			(12.968)
ling. prox. to SWE × year			-0.027
			(0.035)
log-likelihood	-195.2	-199.1	-197.7
Pseudo-R^2	0.359	0.346	0.351
N. obs.	18,775	18,775	18,775

G Demographic contagion: Population-weighted

Here we show results for an alternative version of the demographic contagion model presented in Section 4.3, in which the spillover effect is weighted by a country's share of the population. Formally, equation (6) is now:

$$\mathcal{A}_{it} = \underbrace{\delta_0 \sum_{j=1}^{N} p_j \mathcal{I}_{j,t-1} + \delta_1 \left(\sum_{j=1}^{N} p_j \mathcal{I}_{j,t-1} \right)^2}_{\text{global}} + \underbrace{\sum_{z \in Z \setminus z_0} \gamma_z \sum_{j|(i,j) \in z} p_j \mathcal{I}_{j,t-1}}_{\text{network-mediated}} \tag{17}$$

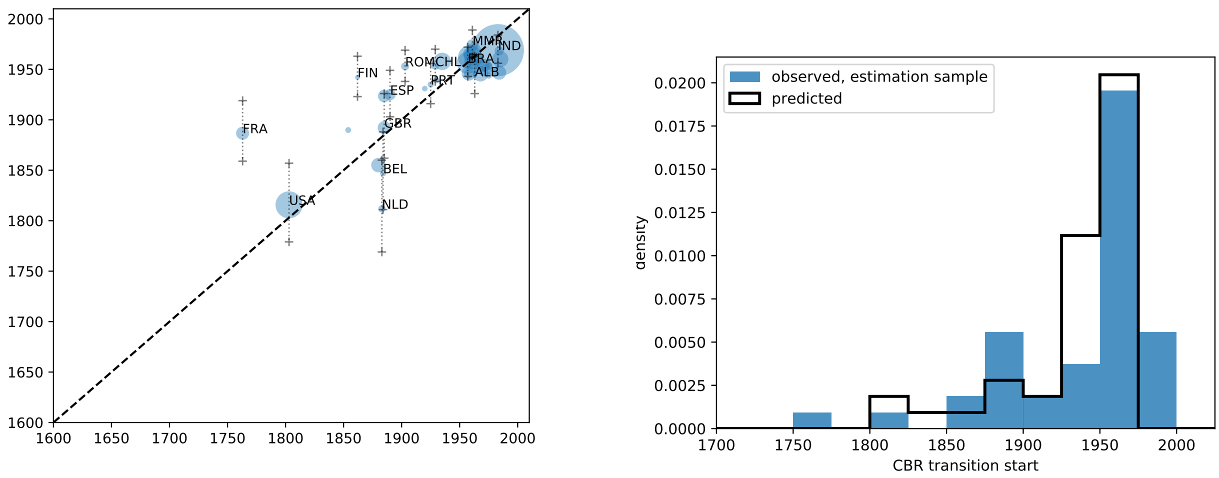
where p_j is country j's share of the global population at time t. In order to fill in some gaps in the pre-1950 data, we assume constant population levels prior to the first observation of population.

Table A3: Determinants of the start of the CBR transition, population-weighted

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
constant	-84.127**	-128.839**	-113.862**	-122.092**	-107.404**	-7.131	-122.044***	-17.692
	(33.170)	(39.617)	(39.187)	(40.369)	(39.746)	(54.371)	(36.654)	(57.358)
ln GDP pc	10.237**	16.026**	14.041**	15.144**	13.197**	6.331	6.504	5.733
	(4.315)	(5.148)	(5.096)	(5.248)	(5.171)	(4.494)	(4.519)	(4.628)
$(\ln \text{GDP pc})^2$	-0.484*	-0.850**	-0.756**	-0.804**	-0.705**	-0.361	-0.359	-0.337
	(0.266)	(0.318)	(0.316)	(0.324)	(0.320)	(0.280)	(0.279)	(0.288)
global transitions		0.016***	0.014***	0.015***	0.013**			
		(0.004)	(0.004)	(0.004)	(0.004)			
$(\text{global trans.})^2$		-0.000**	-0.000**	-0.000**	-0.000**			
		(0.000)	(0.000)	(0.000)	(0.000)			
year						-0.021	0.035***	-0.014
						(0.022)	(0.006)	(0.024)
physically close transitions			0.157***					
			(0.031)					
linguistically close transitions				0.036				
				(0.032)				
phys. *and* ling. close trans.					0.111			
					(0.084)			
phys. close, ling. far trans.					0.175***			
					(0.041)			
ling. close, phys. far trans.					0.023			
					(0.043)			
ln GC dist from GBR						-3.849**		-4.257**
						(1.347)		(1.404)
ln GC dist from GBR × year						0.008**		0.009**
						(0.003)		(0.003)
ling. prox. to GBR							11.283*	22.522**
							(5.920)	(7.791)
ling. prox. to GBR × year							-0.016	-0.044*
							(0.020)	(0.023)
log-likelihood	-249.3	-239.7	-229.3	-239.1	-229.0	-200.9	-204.3	-195.2
Pseudo-R^2	0.006	0.007	0.008	0.007	0.008	0.011	0.011	0.012
N. obs.	18,775	18,775	18,775	18,775	18,775	18,775	18,775	18,775

Table A3 and Figure 27 show the results. The direction, magnitude, and significance of all estimated parameters are essentially unchanged.

Figure 27: Specification (8), Start of CBR Transitions, population-weighted



Panel (a): within-sample predictions

Panel (b): distribution of transition dates

H An empirical analysis of CDR transitions

Table A4 reports the results of estimating the same model that was described for crude birth rates in Section 3. It is apparent that unlike for crude birth rates, GDP per capita does not appear to have an explanatory role to play in the timing of death rate transitions. While the global accumulation of transitioning countries (i.e., global transitions) has robust predictive power, no significant effect is measured for any of our bilateral measures of distance. Distance from Great Britain has strong predictive power, validating our choice of specification for the quantitative exercise in Section 5.

Figures 28 and 29 visualize the results from the right-most column of Table A4.

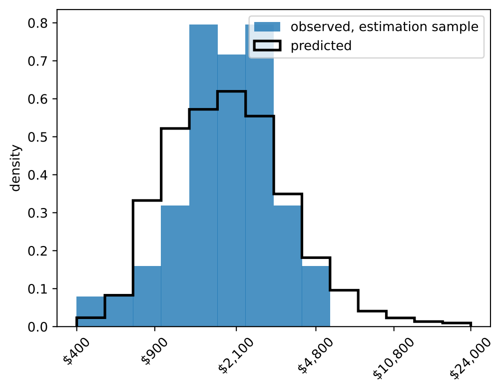
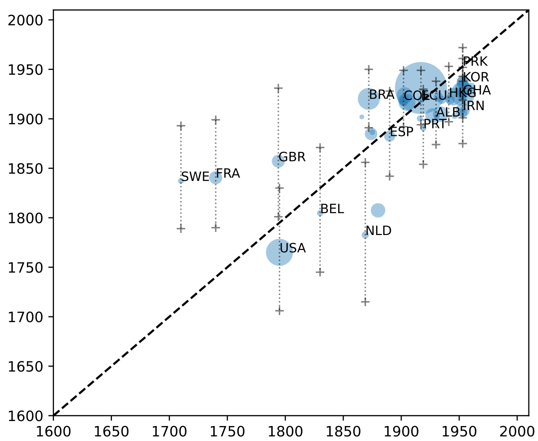


Figure 28: Distribution of log GDPpc at the start of the CDR transitions, Specification (8)

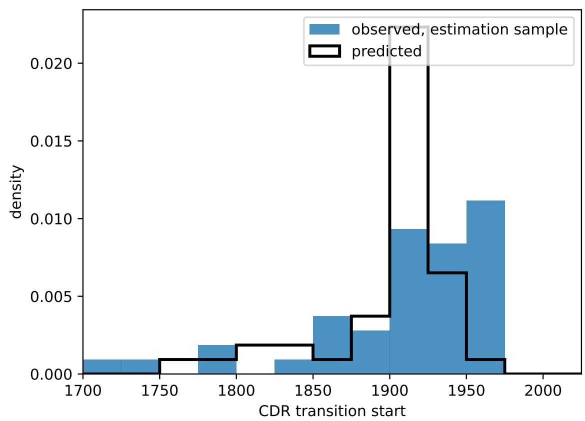
Table A4: Determinants of the start of the CDR transition

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
constant	-51.528	-35.696	-27.200	-40.133	-39.357	-51.843	-103.618*	-63.845
	(65.277)	(57.773)	(57.793)	(57.999)	(58.866)	(65.016)	(58.649)	(64.473)
ln GDP pc	6.032	3.686	2.578	4.251	4.132	7.509	6.788	8.167
	(8.489)	(7.517)	(7.521)	(7.546)	(7.659)	(7.560)	(7.595)	(7.642)
$(\ln \text{GDP pc})^2$	-0.293	-0.187	-0.113	-0.236	-0.235	-0.475	-0.415	-0.535
	(0.575)	(0.511)	(0.511)	(0.514)	(0.522)	(0.512)	(0.513)	(0.519)
global transitions		0.286***	0.280***	0.276***	0.283***			
		(0.045)	(0.045)	(0.046)	(0.047)			
$(\text{global trans.})^2$		-0.005***	-0.005***	-0.005***	-0.005***			
		(0.001)	(0.001)	(0.001)	(0.001)			
year						-0.004	0.025***	-0.000
						(0.017)	(0.004)	(0.018)
physically close transitions			0.035					
			(0.041)					
linguistically close transitions				0.064				
				(0.057)				
phys. *and* ling. close trans.					0.311			
					(0.201)			
phys. close, ling. far trans.					0.007			
					(0.050)			
ling. close, phys. far trans.					0.006			
					(0.081)			
ln GC dist from GBR						-1.724**		-1.855**
						(0.845)		(0.893)
ln GC dist from GBR × year						0.003		0.004
						(0.002)		(0.002)
ling. prox. to GBR							7.790**	10.765**
							(3.912)	(4.679)
ling. prox. to GBR × year							-0.014	-0.022
							(0.014)	(0.017)
log-likelihood	-284.3	-238.7	-238.3	-238.1	-237.1	-231.7	-231.7	-228.0
Pseudo-R^2	0.057	0.208	0.209	0.210	0.213	0.231	0.231	0.243
N. obs.	17,517	17,517	17,517	17,517	17,517	17,517	17,517	17,517

Figure 29: Specification (8), Start of CDR Transitions



Panel (a): within-sample predictions



Panel (b): distribution of transition dates

I A note on British data

Since a large part of our calibration relies on British data, which come in different territorial units (e.g., England and Wales vs. Great Britain vs. the United Kingdom), this appendix provides further details.

For vital statistics, 1541 through 1839 correspond to England only, while 1840 through 1854 correspond to England and Wales. From 1855 through 1921, these variables correspond to Great Britain: England, Wales, and Scotland. From 1922 onward, they correspond to the United Kingdom (Great Britain plus Northern Ireland).

For GDP per capita, 1500 through 1700 corresponds to England alone. From 1700 to 1850, the numbers correspond to Great Britain: England, Wales, and Scotland. After 1850, they correspond to the United Kingdom, which exists in two phases: from 1850 to 1921 including all of Ireland, and from 1922 onward including only Northern Ireland.

Nonetheless, given the demographic and economic weight of England within the British Islands, adjusting these data to slightly different territorial units (if we had access to the raw microdata, which, unfortunately, we do not) would have only a minor impact on our results.

J Optimality conditions

First, notice that no solution for the household problem exists if the income of the current adults cannot cover the subsistence requirement $\bar{c}$. Also, any $n_t \leq \frac{\bar{n}}{s_t^0}$ cannot be part of an optimal solution. Combining these two requirements, a necessary and sufficient condition for the existence of an optimal solution is that

$$1 - \frac{(1+\bar{n})\bar{c}}{w_t^u + w_t^s h_t} > 0.$$

If this condition is met, we can turn our attention to the optimality conditions. The first-order condition for n_t is:

$$\frac{1}{c_t^1}\left[(w_t^u + w_t^s h_t)(\tau_1 + \tau_2 s_t^0 e_t) + s_t^0 \bar{c}\right] \geq \gamma \frac{s_t^0}{s_t^0 n_t - \bar{n}}. \tag{18}$$

This equation must hold with strict equality if $n_t > 0$, which will always be the case, as long as $\bar{n} > 0$, as we assume in our calibration. Thus, we get an expression for n_t in terms of e_t:

$$n_t = \frac{\gamma}{1+\gamma}\left[\frac{1 - \frac{\bar{c}}{w_t^u + w_t^s h_t}}{\tau_1 + \tau_2 s_t^0 e_t + \frac{s_t^0 \bar{c}}{w_t^u + w_t^s h_t}} + \frac{1}{\gamma}\frac{\bar{n}}{s_t^0}\right] \tag{19}$$

The first-order condition for $e_t \geq 0$ is given by

$$\frac{(w_t^u + w_t^s h_t) n_t s_t^0 \tau_2}{c_t^1} \geq \frac{\phi s_t^1 h_t^v w_{t+1}^s}{w_{t+1}^u + h_t^v w_{t+1}^s e_t^\xi} e_t^{\xi-1}. \tag{20}$$

Since $\xi \in (0,1)$, $e_t = 0$ can never be optimal. Therefore, this equation holds with strict equality and we get another closed-form expression for n_t in terms of e_t:

$$n_t = \frac{\frac{\phi s_t^1}{1+\phi s_t^1}\left(1 - \frac{\bar{c}}{w_t^u + w_t^s h_t}\right)}{\frac{\phi s_t^1}{1+\phi s_t^1}\left(\tau_1 + \frac{s_t^0 \bar{c}}{w_t^u + w_t^s h_t}\right) + \frac{s_t^0 \tau_2}{1+\phi s_t^1} \frac{w_{t+1}^u}{h_t^v w_{t+1}^s} e_t^{1-\xi} + \tau_2 s_t^0 e_t} \tag{21}$$

Combining equations (19) and (21), we can derive $\frac{\partial u(n_t(e_t), e_t)}{\partial e_t} \equiv f(e_t)$:

$$f(e_t) = \underbrace{\frac{\frac{\phi s_t^1}{1+\phi s_t^1}\left(1 - \frac{\bar{c}}{w_t^u + w_t^s h_t}\right)}{\frac{\phi s_t^1}{1+\phi s_t^1}\left(\tau_1 + \frac{s_t^0 \bar{c}}{w_t^u + w_t^s h_t}\right) + \frac{s_t^0 \tau_2}{1+\phi s_t^1} \frac{w_{t+1}^u}{h_t^v w_{t+1}^s} e_t^{1-\xi} + \tau_2 s_t^0 e_t}}_{\tilde{b}(e_t)}$$

$$- \underbrace{\frac{\gamma}{1+\gamma}\left[\frac{1 - \frac{\bar{c}}{w_t^u + w_t^s h_t}}{\tau_1 + \tau_2 s_t^0 e_t + \frac{s_t^0 \bar{c}}{w_t^u + w_t^s h_t}} + \frac{1}{\gamma}\frac{\overline{n}}{s_t^0}\right]}_{\tilde{k}(e_t)} \tag{22}$$

Next, we can examine the labor-market clearing conditions. The total number of young and middle-aged adults and elders working in the economy at time t is given by N_t^1, N_t^2, and N_t^3 respectively. Market clearing for labor requires that

$$L_{a,t} + L_t = N_t^1 + \zeta^2 N_t^2 + \zeta^3 N_t^3 \equiv \overline{L}_t, \tag{23}$$

and

$$H_{a,t} + H_t = h_t N_t^1 + \zeta^2 h_{t-1} N_t^2 + \zeta^3 h_{t-2} N_t^3 \equiv \overline{H}_t. \tag{24}$$

Combining equation (13) with equations (23) and (24), we get

$$\alpha A_t \left(\overline{L}_t - L_{m,t}\right)^{\alpha-1} \left(\overline{H}_t - H_{m,t}\right)^{\rho_a - \alpha} = \beta B_t L_{m,t}^{\beta-1} H_{m,t}^{\rho_m - \beta}. \tag{25}$$

Similarly, combining equation (14) with equations (23) and (24)), we get

$$(\rho_a - \alpha) A_t \left(\overline{L}_t - L_{m,t}\right)^{\alpha} \left(\overline{H}_t - H_{m,t}\right)^{\rho_a - \alpha - 1} = (\rho_m - \beta) B_t L_{m,t}^{gb} H_{m,t}^{\rho_m - \beta - 1}. \tag{26}$$

The last two equations imply:

$$\frac{H_{m,t}}{\overline{H}_t} = \frac{1}{\frac{\beta}{\alpha}\frac{\rho_a - \alpha}{\rho_m - \beta}\frac{\overline{L}_t}{L_t} + 1 - \frac{\beta}{\alpha}\frac{\rho_a - \alpha}{\rho_m - \beta}}. \tag{27}$$

Equation (25) can be developed into

$$\frac{\alpha}{\beta}\frac{A_t}{B_t}\frac{\overline{L}_t^{\alpha-\beta}}{\overline{H}_t^{\alpha-\beta+\rho_m-\rho_a}} = \frac{\left(1-\frac{L_{m,t}}{\overline{L}_t}\right)^{1-\alpha}}{\left(\frac{L_{m,t}}{\overline{L}_t}\right)^{1-\beta}}\frac{\left(\frac{H_{m,t}}{\overline{H}_t}\right)^{\rho_m-\beta}}{\left(1-\frac{H_{m,t}}{\overline{H}_t}\right)^{\rho_a-\alpha}} \tag{28}$$

Equations (27) and (28) are two equations in two unknowns, $\frac{H_{m,t}}{\overline{H}_t}$ and $\frac{L_{m,t}}{\overline{L}_t}$.

Combining equations (27) and (28), we can derive $z\left(\frac{L_{m,t}}{\overline{L}_t}\right)$:

$$\begin{aligned} z\left(\frac{L_{m,t}}{\overline{L}_t}\right) \equiv\ & 1-\frac{L_{m,t}}{\overline{L}_t} \\ & -\left(\frac{L_{m,t}}{\overline{L}_t}\right)^{\frac{1-\rho_m}{1-\rho_a}}\left(\left(1-\frac{\beta}{\alpha}\frac{\rho_a-\alpha}{\rho_m-\beta}\right)\frac{L_{m,t}}{\overline{L}_t}+\frac{\beta}{\alpha}\frac{\rho_a-\alpha}{\rho_m-\beta}\right)^{\frac{-\rho_a+\rho_m-\beta+\alpha}{1-\rho_a}}\times \\ & \times\left(\frac{\alpha}{\beta}\frac{A_t}{B_t}\frac{\overline{L}_t^{\alpha-\beta}}{\overline{H}_t^{\alpha-\beta+\rho_m-\rho_a}}\right)^{\frac{1}{1-\rho_a}}\left(\frac{\beta}{\alpha}\frac{\rho_a-\alpha}{\rho_m-\beta}\right)^{\frac{\rho_a-\alpha}{1-\rho_a}} \end{aligned} \tag{29}$$

The equilibrium value of $\frac{L_{m,t}}{\overline{L}_t}$ can be characterized as the unique point at which $z\left(\frac{L_{m,t}}{\overline{L}_t}\right) = 0$.

K Equilibrium

Define the vector of time-t state variables $x_t \equiv [A_t, B_t, M_t, N_t^1, N_t^2, N_t^3, h_t^1, h_t^2, h_t^3]'$. Let the law of motion $x_{t+1} = m_t(x_t)$ be given by:

$$\begin{aligned} N_{t+1}^1 &= s_t^0 s_t^1 n_t N_t^1 \\ N_{t+1}^2 &= s_t^2 N_t^1 \\ N_{t+1}^3 &= s_t^3 N_t^2 \\ h_{t+1}^1 &= e_t \\ h_{t+1}^2 &= h_t^1 \\ h_{t+1}^3 &= h_t^2 \end{aligned} \tag{30}$$

and a series of technology levels $\{A_t, B_t, M_t\}_{t=0}^T$. The survival probabilities s_t^0, s_t^1, s_t^2 and s_t^3 are determined by M_t according to equation (12).

Labor allocations $L_{m,t}$ and $H_{m,t}$ are implicit functions of x_t characterized by equations (23), (24), (27), and (29). Given labor allocations, the wages w_t^u and w_t^s are also implicit functions of x_t characterized by equations (13) and (14).

Define $\tilde{w}_{t+1} \equiv \frac{w_{t+1}^s}{w_{t+1}^u}$. According to the solution characterized by (19) and (22), n_t and e_t are both implicit functions of x_t and $\tilde{w}_{t+1}$. From equation (30), we see that x_{t+1} is determined by x_t and the choices n_t and e_t, so we can reformulate $L_{m,t}$ and $H_{m,t}$ as functions of x_t, n_t, and e_t. Finally, employing equations (13) and (14) once more, define

$$g_t(\tilde{w}) \equiv \frac{\rho_b - \beta}{\beta} \frac{L_{t+1,m}[x_t, n_t(x_t, \tilde{w}), e_t(x_t, \tilde{w})]}{H_{t+1,m}[x_t, n_t(x_t, \tilde{w}), e_t(x_t, \tilde{w})]} - \tilde{w}. \tag{31}$$

Given a set of initial cohort sizes and education levels N_0^1, N_0^2, N_0^3, h_0^1, h_0^2, and h_0^3, and a series of technology levels $\{A_t, B_t, M_t\}_{t=0}^T$, an equilibrium consists of a series $\{\tilde{w}_{t+1}\}_{t=0}^T$ such that $g_t(\tilde{w}_{t+1}) = 0$ and $x_{t+1} = m_t(x_t)$ for all t.

The function $g_t(\tilde{w})$ is monotonically decreasing and continuous almost everywhere. For certain parameter values, it may have a single point of discontinuity at $\tilde{w}_0$, the level of the wage ratio where the optimal choice of education shifts from positive to zero. The only way for a solution not to exist for equation (31) is if $\lim_{\tilde{w}\to\tilde{w}_0^+} g_t(\tilde{w}) > 0$ and $\lim_{\tilde{w}\to\tilde{w}_0^-} g_t(\tilde{w}) < 0$. With the subset of the parameter space that we work with, however, this condition never holds and we always find a solution.

L Steady state

In the steady state, fertility is constant at replacement level: $n_t = \tilde{n} = \frac{1}{s_0 s_1}$. The size of each cohort is constant over time, with $N_2 = s^2 N_1$ and $N_3 = s^3 N_1$. Education e, human capital h, technologies A and B, and wages w^u and w^s are likewise constant over time.

Then, the stock of unskilled and skilled labor is constant and given by:

$$\overline{L} = (1 + s^2\zeta^2 + s^2 s^3 \zeta^3)N^1$$
$$\overline{H} = (1 + s^2\zeta^2 + s^2 s^3 \zeta^3)N^1 h.$$

Skilled and unskilled wages are:

$$w^u = \alpha A L_a^{\alpha-1} H_a^{\rho_a-\alpha} = \beta B L_m^{\beta-1} H_m^{\rho_b-\beta},$$

and

$$w^s = (\rho_b - \alpha) A L_a^{\alpha} H_a^{\rho_a-\alpha-1} = (\rho_b - \beta) B L_m^{\beta} H_m^{\rho_b-\beta-1}.$$

For the purposes of characterizing the steady state, it is convenient to define $\tilde{c} \equiv \frac{\bar{c}}{w^u}$.

Lemma 1 *If a steady state exists, it can be characterized by the following four equations:*

$$h = \left(\frac{\phi s_1}{\tilde{n} s^0 \tau_2} \frac{(1+\tilde{z})(1-\tau_1\tilde{n}) - \tilde{c}(1+s^0\tilde{n})}{\tilde{z}(1+\phi s^1) + 2 + \phi s_1 + \frac{1}{\tilde{z}}} \right)^{\frac{\xi}{1-v}}$$

$$N^1 = \left[\frac{\left(1-\frac{L_m}{\bar{L}}\right)^{1-\rho_a} \left(1+s^2\zeta^2+s^2s^3\zeta^3\right)^{\rho_m-\rho_a} h^{\alpha-\beta+\rho_m-\rho_a}}{\left(\frac{L_m}{\bar{L}}\right)^{1-\rho_m} \left(\left(1-\frac{\beta}{\alpha}\frac{\rho_a-\alpha}{\rho_m-\beta}\right)\frac{L_m}{\bar{L}}+\frac{\beta}{\alpha}\frac{\rho_a-\alpha}{\rho_m-\beta}\right)^{-\rho_a+\rho_m-\beta+\alpha} \frac{\alpha}{\beta}\frac{A}{B}\left(\frac{\beta}{\alpha}\frac{\rho_a-\alpha}{\rho_m-\beta}\right)^{\rho_a-\alpha}} \right]^{\rho_a-\rho_m}$$

$$\frac{L_m}{\bar{L}} = \frac{\tilde{z}-\frac{\rho_a-\alpha}{\alpha}}{\frac{\rho_b-\beta}{\beta}-\frac{\rho_a-\alpha}{\alpha}}$$

$$\tilde{z} = \frac{-\hat{b}\pm\sqrt{\hat{b}^2-4\hat{a}\hat{c}}}{2\hat{a}},$$

with $\hat{a}$, $\hat{b}$, and $\hat{c}$ defined as:

$$\begin{aligned}
\hat{a} &\equiv 1-\frac{\phi s^1(1-\bar{n}s^1)}{\gamma}-\frac{\tau_1}{s^0s^1}\left(1+\frac{1-\bar{n}s^1}{\gamma}\right)\\
\hat{b} &\equiv 1-\left(1+\frac{1-\bar{n}s^1}{\gamma}\right)\frac{2\tau_1+s^0\tilde{c}}{s^0s^1}+(1-\tilde{c})\left(1-\frac{\phi s^1(1-\bar{n}s^1)}{\gamma}\right)\\
\hat{c} &\equiv 1-\tilde{c}-\left(1+\frac{1-\bar{n}s^1}{\gamma}\right)\frac{\tau_1+s^0\tilde{c}}{s^0s^1}
\end{aligned}$$

Proof: In the steady state, the stock of unskilled and skilled labor is given by:

$$\begin{aligned}
\bar{L} &= (1+s^2\zeta^2+s^2s^3\zeta^3)N^1\\
\overline{H} &= (1+s^2\zeta^2+s^2s^3\zeta^3)N^1h
\end{aligned}$$

Furthermore, skilled and unskilled wages are:

$$w^u = \alpha A L_a^{\alpha-1}H_a^{\rho_a-\alpha} = \beta B L_m^{\beta-1}H_m^{\rho_b-\beta},$$

and

$$w^s = (\rho_b-\alpha)AL_a^{\alpha}H_a^{\rho_a-\alpha-1} = (\rho_b-\beta)BL_m^{\beta}H_m^{\rho_b-\beta-1}$$

Thus, the skill premium is

$$\frac{w^s}{w^u} = \frac{\tilde{z}}{h}. \tag{32}$$

where $\tilde{z} \equiv \frac{\rho_a-\alpha}{\alpha}\left(1-\frac{L_m}{\bar{L}}\right)+\frac{\rho_b-\beta}{\beta}\frac{L_m}{\bar{L}}$.

The choice of education is characterized by:

$$\begin{aligned}
&-e^{1+\frac{\xi}{1-v}}w^s\frac{1+\phi s^1}{\phi s^1}-ew^u\frac{2+\phi s^1}{\phi s^1}+e^{\frac{\xi}{1-v}}\frac{w^s(1-\tau_1\tilde{n})}{\tilde{n}s^0\tau_2}\\
&\qquad\qquad -e^{1-\frac{\xi}{1-v}}\frac{1}{\phi s^1}\frac{w^u}{w^s}w^u+\frac{w^u(1-\tau_1\tilde{n})}{\tilde{n}s^0\tau_2}-\frac{\overline{c}(1+s^0\tilde{n})}{\tilde{n}s^0\tau_2}\leq 0.
\end{aligned}$$

To analyze this expression, substitute $w^s = \frac{\tilde{z}}{h} w^u$, $\overline{c} = \tilde{c} w^u$, and $h = e^{\frac{\xi}{1-v}}$, and then divide everything by w^u:

$$-e\tilde{z}\frac{1+\phi s^1}{\phi s^1} - e\frac{2+\phi s^1}{\phi s^1} + \tilde{z}\frac{1-\tau_1\tilde{n}}{\tilde{n}s^0\tau_2} - e\frac{1}{\tilde{z}}\frac{1}{\phi s^1} + \frac{1-\tau_1\tilde{n}}{\tilde{n}s^0\tau_2} - \frac{\tilde{c}(1+s^0\tilde{n})}{\tilde{n}s^0\tau_2} \leq 0.$$

Rearranging:

$$e \geq \frac{\phi s_1}{\tilde{n}s^0\tau_2}\frac{(1+\tilde{z})(1-\tau_1\tilde{n}) - \tilde{c}(1+s^0\tilde{n})}{\tilde{z}(1+\phi s^1) + 2 + \phi s_1 + \frac{1}{\tilde{z}}}. \tag{33}$$

Since in the steady state, fertility must be at replacement level, $n = \tilde{n} \equiv \frac{1}{s^0 s^1}$, we use equations (19) and (32) to obtain

$$\begin{aligned}
\tilde{n} &= \frac{\gamma}{1+\gamma}\left[\frac{1-\frac{\tilde{c}w^u}{w^u+\tilde{z}w^u}}{\tau_1+\tau_2 s^0 e + \frac{s^0\tilde{c}w^u}{w^u+\tilde{z}w^u}} + \frac{1}{\gamma}\frac{\overline{n}}{s^0}\right] \\
&= \frac{\gamma}{1+\gamma}\left[\frac{1-\frac{\tilde{c}}{1+\tilde{z}}}{\tau_1+\tau_2 s^0 e + \frac{s^0\tilde{c}}{1+\tilde{z}}} + \frac{1}{\gamma}\frac{\overline{n}}{s^0}\right]
\end{aligned}$$

or:

$$\tilde{n} - \frac{1}{1+\gamma}\frac{\overline{n}}{s^0} = \frac{\gamma}{1+\gamma}\frac{1-\frac{\tilde{c}}{1+\tilde{z}}}{\tau_1+\tau_2 s^0 e + \frac{s^0\tilde{c}}{1+\tilde{z}}}.$$

Next, we substitute in for e using equation (33), and solve for $\tilde{z}$:

$$\tilde{z}^2\left[1 - \frac{\phi s^1(1-\overline{n}s^1)}{\gamma} - \tau_1\tilde{g}\right] + \tilde{z}\left[1 - \tilde{g}(2\tau_1 + s^0\tilde{c}) + (1-\tilde{c})\left(1 - \frac{\phi s^1(1-\overline{n}s^1)}{\gamma}\right)\right] + 1 - \tilde{c} - \tilde{g}(\tau_1 + s^0\tilde{c}) = 0 \tag{34}$$

where $\tilde{g} \equiv \tilde{n}\left(1 + \frac{1-\overline{n}s^1}{\gamma}\right)$.

If a solution for $\tilde{z}$ exists, then it solves:

$$\tilde{z} = \frac{-\hat{b} \pm \sqrt{\hat{b}^2 - 4\hat{a}\hat{c}}}{2\hat{a}}$$

where:

$$\begin{aligned}
\hat{a} &= 1 - \frac{\phi s^1(1-\overline{n}s^1)}{\gamma} - \frac{\tau_1}{s^0 s^1}\left(1 + \frac{1-\overline{n}s^1}{\gamma}\right) \\
\hat{b} &= 1 - \left(1 + \frac{1-\overline{n}s^1}{\gamma}\right)\frac{2\tau_1 + s^0\tilde{c}}{s^0 s^1} + (1-\tilde{c})\left(1 - \frac{\phi s^1(1-\overline{n}s^1)}{\gamma}\right) \\
\hat{c} &= 1 - \tilde{c} - \left(1 + \frac{1-\overline{n}s^1}{\gamma}\right)\frac{\tau_1 + s^0\tilde{c}}{s^0 s^1}.
\end{aligned}$$

The fraction of unskilled labor used in the modern sector can then be recovered as:

$$\frac{L_m}{\overline{L}} = \frac{\tilde{z} - \frac{\rho_a - \alpha}{\alpha}}{\frac{\rho_b - \beta}{\beta} - \frac{\rho_a - \alpha}{\alpha}}$$

If $\frac{L_m}{\overline{L}}$ implied by one of the two quadratic solutions is between 0 and 1, then it is a solution. If neither quadratic solution meets this criterion, then no steady state exists.[30] In the case of its existence, the steady-state education is:

$$e = \frac{\phi s_1}{\tilde{n} s^0 \tau_2} \frac{(1+\tilde{z})(1-\tau_1 \tilde{n}) - \tilde{c}(1+s^0 \tilde{n})}{\tilde{z}(1+\phi s^1) + 2 + \phi s_1 + \frac{1}{\tilde{z}}},$$

and steady-state human capital is simply $h = e^{\frac{\xi}{1-v}}$.

Next, we can solve for the steady-state level of population. From equation (29):

$$\left(\frac{L_m}{\overline{L}}\right)^{\frac{1-\rho_m}{1-\rho_a}} \left(\left(1 - \frac{\beta}{\alpha}\frac{\rho_a - \alpha}{\rho_m - \beta}\right) \frac{L_m}{\overline{L}} + \frac{\beta}{\alpha}\frac{\rho_a - \alpha}{\rho_m - \beta} \right)^{\frac{-\rho_a + \rho_m - \beta + \alpha}{1-\rho_a}}$$
$$\times \left(\frac{\alpha}{\beta} \frac{A}{B} \frac{\overline{L}^{\alpha - \beta}}{\overline{H}^{\alpha - \beta + \rho_m - \rho_a}} \right)^{\frac{1}{1-\rho_a}} \left(\frac{\beta}{\alpha} \frac{\rho_a - \alpha}{\rho_m - \beta} \right)^{\frac{\rho_a - \alpha}{1-\rho_a}} = 1 - \frac{L_m}{\overline{L}}.$$

Substituting in for $\overline{L}$ an $\overline{H}$ yields:

$$N^1 = \left[\frac{\left(1 - \frac{L_m}{\overline{L}}\right)^{1-\rho_a} \left(1 + s^2\zeta^2 + s^2 s^3 \zeta^3\right)^{\rho_m - \rho_a} h^{\alpha - \beta + \rho_m - \rho_a}}{\left(\frac{L_m}{\overline{L}}\right)^{1-\rho_m} \left(\left(1 - \frac{\beta}{\alpha}\frac{\rho_a - \alpha}{\rho_m - \beta}\right) \frac{L_m}{\overline{L}} + \frac{\beta}{\alpha}\frac{\rho_a - \alpha}{\rho_m - \beta} \right)^{-\rho_a + \rho_m - \beta + \alpha} \frac{\alpha}{\beta}\frac{A}{B} \left(\frac{\beta}{\alpha}\frac{\rho_a - \alpha}{\rho_m - \beta}\right)^{\rho_a - \alpha}} \right]^{\rho_a - \rho_m}.$$

M Accounting for intra-period mortality

This section describes the initial values of the survival probabilities that we use in our calibration of the model. A proper definition of s_0^0, s_0^1, s_0^2, and s_0^3 will take into account both the probability of survival until the next period and the average number of years alive during the following period (our period lasts 20 years). To that end, we define $\tilde{s}_x$ as the average fraction of those alive at the beginning of age x that are alive during age x; and $\bar{s}_x$ as the fraction of those alive at the beginning of age x that are alive at the beginning of age

[30] We checked, numerically, that for our calibration (and a wide range of robustness values of parameters), we have one and only one solution to this quadratic equation between 0 and 1. We conjecture that, if a solution exists between 0 and 1, such a solution is unique.

$x + 1$, where x takes values in $\{1, 2, 3, 4\}$. Then:

$$\begin{aligned} s_0 &\equiv \tilde{s}_1 \\ s_1 &\equiv \bar{s}_1 \tilde{s}_2 \\ s_2 &\equiv \bar{s}_2 \tilde{s}_3 \\ s_3 &\equiv \bar{s}_3 \tilde{s}_4. \end{aligned}$$

Furthermore, let the function $S_t(x)$, mapping $\mathbb{R}_+ \to [0, 1]$, represent the survival probability of birth cohort t to age x, where x is measured in years. Let $S_t(0) = 1$ and $S_t(80) = 0$. We can then define:

$$\begin{aligned} \bar{s}_1(t) &\equiv S_t(20) \\ \bar{s}_2(t) &\equiv \frac{S_t(40)}{S_t(20)} \\ \bar{s}_3(t) &\equiv \frac{S_t(60)}{S_t(40)} \\ \tilde{s}_1(t) &\equiv \frac{1}{20} \int_0^{20} S_t(x)dx \\ \tilde{s}_2(t) &\equiv \frac{1}{20} \frac{1}{\bar{s}_1(t)} \int_{20}^{40} S_t(x)dx \\ \tilde{s}_3(t) &\equiv \frac{1}{20} \frac{1}{\bar{s}_2(t)} \int_{40}^{60} S_t(x)dx \\ \tilde{s}_4(t) &\equiv \frac{1}{20} \frac{1}{\bar{s}_3(t)} \int_{60}^{80} S_t(x)dx. \end{aligned}$$

Finally, let s_t represent the fraction of a cohort alive at the start of a given time t that is alive at the end of the period. Suppose that instead of everyone dying in a single moment at the end of the period, mortality is spread out across x distinct sub-periods and that the mortality hazard is constant across sub-periods. Then, the constant mortality hazard is $s_t^{\frac{1}{x}}$, and the average fraction of people alive during the entire period is:

$$\tilde{s}_t = \frac{1 + s_t^{\frac{1}{x}} + s_t^{\frac{2}{x}} + \cdots + s_t^{\frac{x-1}{x}}}{x} = \frac{1 - s_t}{x(1 - s_t^{\frac{1}{x}})}.$$

N English mortality data

Table A5 shows the English mortality data that we use to pin down the pre-modern survival probabilities in our model. The data are derived from Wrigley et al. (1997, ch. 6). Deaths during the first year of life are derived from Table 6.4 and represent averages over the period from 1675 to 1699. Deaths between the ages of 1 and 14 are derived from Table

6.1 and represent averages over the period 1680-1699. Deaths between the ages of 25 and 79 are derived from Table 6.19 and represent averages over the period 1680-1699. Deaths between the ages of 15 and 24 are imputed by assuming a linear progression between the 10-14 years age range and the 25-29 years age range. All deaths are given as rates per 1000 living members of the cohort. Initial survival probabilities are calculated using the method described in Section M, where $S_t(x)$ is a stepwise function consistent with the data in Table A5.

Table A5: Mortality in England, 1680-1699

age range	deaths per 1000
0 days	50.90
1-6 days	28.90
7-29 days	34.10
30-59 days	17.80
60-89 days	13.10
90-179 days	24.60
180-274 days	16.00
275-364 days	16.60
1-4 years	108.65
5-9 years	45.05
10-14 years	26.15
15-19 years	46.83
20-24 years	67.52
25-29 years	88.20
30-34 years	87.20
35-39 years	97.20
40-44 years	93.95
45-49 years	119.15
50-54 years	145.85
55-59 years	191.15
60-64 years	244.70
65-69 years	269.70
70-74 years	411.85
75-79 years	524.00