Demographic Transitions Across Time and Space

Matthew J. Delventhal*

Jesús Fernández-Villaverde†

Nezih Guner‡

Version: October 2020.

Abstract

The demographic transition, i.e., the move from a regime of high fertility/high mortality into a regime of low fertility/low mortality, is a process that almost every country on Earth has undergone or is undergoing. Are all demographic transitions equal? Have they changed in speed and shape over time? And, how do they relate to economic development? To answer these questions, we put together a data set of birth and death rates for 186 countries that spans more than 250 years. Then, we use a novel econometric method to identify start and end dates for transitions in birth and death rates. We find, first, that the average speed of transitions has increased steadily over time. Second, we document that income per capita at the start of these transitions is more or less constant over time. Third, we uncover evidence of *demographic contagion*: the entry of a country into the demographic transition is strongly associated with its neighbors having already entered into the transition, even after controlling for other observables. Next, we build a model of demographic transitions that can account for these facts. The model economy is populated by different locations. In each location, parents decide how many children to have and how much to invest in their human capital. There is skill-biased technological change that diffuses slowly from the frontier country, Britain, to the rest of the world.

Keywords: Demographic Transition, Skill-Biased Technological Change, Diffusion.

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

*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, Casado del Alisal 5, 28014 Madrid, Spain nezih.guner@cemfi.es

1 Introduction

The world population will experience an unprecedented transformation during the coming decades. After increasing continuously since pre-modern times, and turning sharply upwards at the turn of the 20th century, the growth rate of world population reached a maximum around late 1960s around 2% per year. Since then, the growth rate of the world population has been declining and the U.N. expects it to be around just 0.1% by 2100 (Figure 1).

Figure 1: World Pop. Growth, 1600-2100

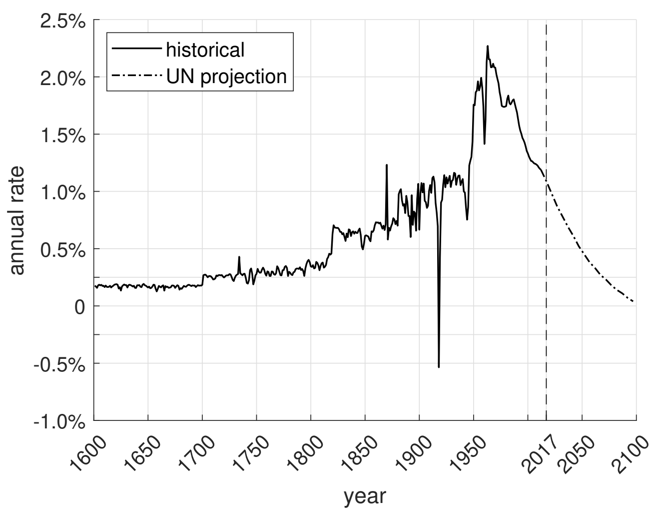
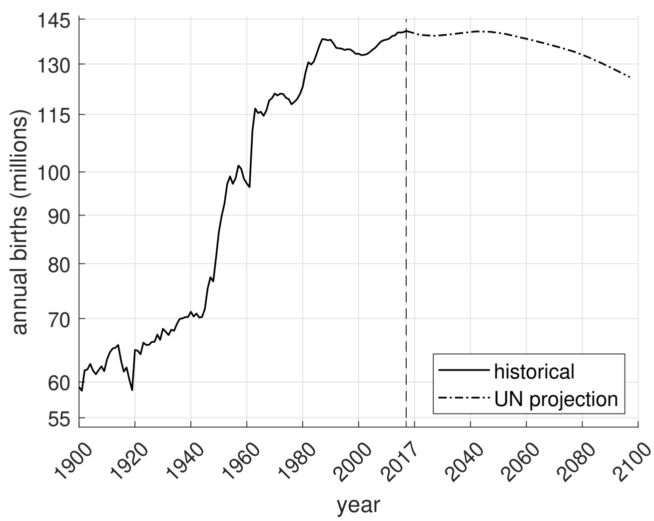


Figure 2: Annual Births, World, 1900-2100



Note: 1600-2016: authors' calculations; see Appendix D. 2017-2100: medium scenario of UN World Population Prospects: The 2017 Revision, (United Nations, 2017).

Another way to look at the ongoing transformation is to consider the total number of children born in the world (Figure 2). After increasing rapidly throughout the first part of the 20th century, this number has barely increased at all since 1980. According to U.N. projections, the number of births in the world is expected to reach a "peak" sometime between 2015 and 2020–indeed, it is most likely that "peak child" already passed in 2017-2018.[1] After that date, the number of children in the world is expected to stop growing.[2]

What is behind the dramatic slowdown in world population growth? The answer is the demographic transition: the move from a regime of high fertility/high mortality into a regime of low fertility/low mortality, is a process that almost every country on Earth has undergone or is undergoing.

[1] The term "peak child" was coined by Hans Rosling, the co-founder and chairman of the Gapminder Foundation, https://www.gapminder.org/tag/hans-rosling/. The U.N. expects a small rebound in total births in the 2040s due to the echo effects of large cohorts of women entering into fertile age at that time. The arguments we will discuss in the rest of the paper make us believe such rebound will not occur and that "peak child" is, indeed, in our past.

[2] Note that even if the number of children stops growing or decreases, population can still grow because of longer life expectancy. But this mechanism is usually much weaker in increasing population than larger birth cohorts.

Indeed, the demographic transition constitutes one of the most powerful ideas in economics and demography. The text book description of demographic transition is as follows:

> "The recent period of very rapid demographic change in most countries around the world is characteristic of the Central phases of a secular process called the demographic transition. Over the course of this transition, declines in birth rates followed by declines in death rates bring about an era of rapid population growth. This transition usually accompanies the development process that transforms an agricultural society into an industrial one. Before the transition's onset, population growth (which equals the difference between the birth and death rate in the absence of migration) is near zero as high death rates more or less off set the high birth rates typical of agrarian societies before the industrial revolution. Population growth is again near zero after the completion of the transition as birth and death rates both reach low levels in the most developed societies." (Boongaarts 2009, page 2985).

Motivated by these observations, in this paper we do two things: First, we put together and analyze data set on crude death rates (CDR) and crude birth rates (CBR) for 186 countries that spans more than 250 years. Following the text book description of the demographic transition, we then 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, iv) final (post-transition) levels of the CDR and CBR. This procedure also allows us to estimate the length and the speed of each transition.

Looking at demographic transitions across time and space, we show that: 1) transitions are becoming faster, 2) the average level of GDP per capita at the start of a transition is more or less constant, 3) demographic transitions are contagious; an important predictor of a country's transition is the prior transition of other countries which are "close" to it in a geographical and a cultural sense.

Next, we build a model economy that can account for these facts. We consider an economy with multiple locations. 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 the modern technology is more intensive in skilled labor. Mortality is exogenous and is determined by the level of an aggregate medical technology.

Economy is initially in a Malthusian steady state with high and constant mortality fertility. The economy do not grow since the total factor productivity levels of both ancient and modern technologies as well as the level of medical knowledge is constant. At a certain point in time, the TFP in both sectors and the medical technology start growing. This occurs first in the frontier country, Britain in our analysis, and then diffuses slowly to other locations. Higher demand for skilled labor and rising skill premium makes investment in children more valuable and parents react by reducing the number of children but educating them better. We first calibrate the model economy to replicate the demographic and economic transition in Britain. We then show that a simple mechanism of diffusion where technological change travels from Britain to the rest of the world in a manner that depends on geographic and cultural/linguistic distances is able to generate sequences of demographic transitions, each happening faster than the previous one, exactly as we observe in the data. Furthermore, as countries embark on their demographic transitions, the educational attainment of their population also increases, again as we observe in the data. Together with demographic transition, the world also experience economic transition and global GDP per capita increases by more than 10 fold between the middle of the 19th century and today and the inequality across countries in GDP per capita first increase sharply until the 1980 and then declines.

Understanding the relationship between income and population is one of the oldest challenges in economics, going back to Malthus (1803) who developed a powerful model that links better technology with constant living standards. In a Malthusian world, technological change allows a higher income per capita which leads to higher population through higher fertility and lower mortality. In the presence of a fixed input such as land, this higher population translates into lower marginal productivities that decrease per capita income back to the stationary level previous to the technological advance. Malthus' model is quite successful at accounting for the main facts that prevailed until the nineteenth century, but it fails to explain the coexistence of growth in per capita income and low fertility. Becker (1960) and Becker and Lewis (1973) develop the idea of a trade-off between quantity and quality of children to show that higher per capita incomes and low fertility can go together. The interest in this mechanism was revived with the presentation of an operational dynastic model of fertility in Barro and Becker (1989) and Becker and Barro (1988).

Building on this initial work, Becker, Murphy, and Tamura (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. Several recent papers, e.g., Fernandez-Villaverde (2001), Kalemli-Ozcan (2003), 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, Schoonbroodt, and Tertilt (2011) and Greenwood, Guner, and Vandenbroucke (2017) provide recent reviews of this literature.

Few recent papers study the historical evolution of fertility. Spolaore and Wacziarg (2014) document that genetic and linguistic distance from France was associated with the onset of the fertility transition in Europe. De la Croix and Perrin (2017) focus on the fertility and education transition in France during the 19th century, and show that a simple quality-quantity model can do a decent job in explaining variations of fertility across time and counties in France. De Silva and Tenreyro (2017) focus on post-1960 transitions and emphasize the role of social norms and family planning programs in recent declines in fertility rates in developing countries. Our paper is also related to recent studies that provide an empirical analysis of demographic transitions across countries. 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) also constructs a panel and .finds evidence for a robust effect of early childhood education on fertility decline. Building on these earlier 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.

Finally, by proposing technology diffusion as a mechanism linking the process of the demographic transition in different countries, our analysis also borrows from recent studies on technology diffusion, such as Lucas (2009), Comin and Hobijn (2010), and Comin and Mestieri (2018).

2 Demographic transitions: a methodology

In this section, we propose a methodology for documenting the shape and speed of demographic transitions across time and space. For that purpose, we compile country-level vital statistics, in particular crude birth rates (CBR) and the crude death rates (CDR), across a broad panel.[3] We focus on the CBR and the CDR instead of the other statistics such as the total fertility rate (TFR) rate or life expectancy because the CBR and CDR are more reliably measured in the available data: a researcher only needs an accurate count of births, deaths, and total population. Thus, CBRs and CDRs are available for long periods of time 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 future age-specific fertility and mortality rates. These additional data are not available or are very imprecisely measured for most countries during the pre-modern era and many countries even today; which furthermore provides little reliable basis for making the additional assumptions about future age-specific fertility and mortality which are essential to current TFR and life expectancy calculations.

In the textbook case, a demographic transition has four stages (Chesnais, 1992):

- In Stage 1, both the CBR and the CDR are high and stationary.
- In Stage 2, the CDR starts to decline while the CBR stays high.

[3]Recall that 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.

- In Stage 3, the CBR also starts to decline.
- In Stage 4, both the CDR and the CBR stop falling and become stationary at a lower level.

We take this 4-stage demographic transition as a benchmark model of the evolution of the CBR and CDR and try to fit it to available data for each country. More concretely, for the CBR and the CDR, we estimate, for each rate, the following variables:

1. an initial (pre-transition) average level;
2. the start date of the decline;
3. the end date of the decline; and
4. a final (post-transition) average level.

We do not impose that, either before or after the demographic transition, the average level of the CBR and CDR are equal to each other. Pre-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 anything about the relative ordering of the start dates of CDR and CBR declines: CDR may begin declining before CBR, as in Chesnais' configuration, or CBR may be the first to decline.

2.1 Econometric model

Consider a dependent variable y_t observed for periods $t \in \{1, ..., T\}$. We will assume that y_t can be represented as 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, $s \in \{1, ..., S+1\}$, the dependent variable is defined by:

$$y_{\tau_s} = x'_{\tau_s} \alpha_s + \sigma_s \nu_{s,\tau_s}, \tag{1}$$

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:

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

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 will 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.,

$$f_s(\alpha_s, \alpha_{s+1}, t) = \frac{1}{\tau_{s+1} - \tau_s} \left[(\tau_{s+1} - t)\alpha_s + (t - \tau_s)\alpha_{s+1} \right], \tag{2}$$

and

$$g_s(\sigma_s, \sigma_{s+1}, t) = \frac{1}{\tau_{s+1} - \tau_s} \left[(\tau_{s+1} - t)\sigma_s + (t - \tau_s)\sigma_{s+1} \right]. \tag{3}$$

To apply this theoretical framework to the specific context under study, 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 it is declining, and a final stage it is stationary again). Furthermore, 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 equations (2) and (3), we can write y_t as

$$\begin{aligned}
y_t \;=\; & d_{1t}[x_t'\alpha_1 + \varepsilon_{1t}\sigma_1] \\
& + d_{2t}[x_t' \frac{1}{\tau_3 - \tau_2} \left[(\tau_3 - t)\alpha_1 + (t - \tau_2)\alpha_3 \right] \\
& + d_{2t}[\frac{1}{\tau_3 - \tau_2} \left[(\tau_3 - t)\sigma_1 + (t - \tau_2)\sigma_3 \right] \varepsilon_{2t} \\
& + d_{3t}[x_t'\alpha_3 + \varepsilon_{3t}\sigma_3],
\end{aligned} \tag{4}$$

where $\{d_{st}\}_{s=1}^{3}$ are indicator functions given by

$$d_{1t} = 1\{t \leq \tau_2\}, \quad d_{2t} = 1\{\tau_2 < t < \tau_3\}, \text{ and } d_{3t} = 1\{t \geq \tau_3\}.$$

Equation (4) can then be rearranged as

$$
\begin{aligned}
y_t &= \left[d_{1t} + d_{2t}\left(\frac{\tau_3 - t}{\tau_3 - \tau_2}\right)\right] x_t'\alpha_1 + \left[d_{3t} + d_{3t}\left(\frac{t - \tau_3}{\tau_3 - \tau_2}\right)\right] x_t'\alpha_3 \\
&\quad + \left[d_{1t}\varepsilon_{1t} + d_{2t}\left(\frac{\tau_3 - t}{\tau_3 - \tau_2}\right)\varepsilon_{2t}\right]\sigma_1 + \left[d_{3t}\varepsilon_{3t} + d_{2t}\left(\frac{\tau_3 - t}{\tau_3 - \tau_2}\right)\varepsilon_{3t}\right]\sigma_3, \qquad (5)
\end{aligned}
$$

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 we 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. This means that for a given (τ_2, τ_3) pair, 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 the possible values.

To this end, we define the scalars

$$z_{1t} \equiv d_{1t} + d_{2t}\left(\frac{\tau_3 - t}{\tau_3 - \tau_2}\right),$$

and

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

Then given

$$\underset{1\times T}{y'} \equiv [y_1 \ldots y_T],$$

and

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

the OLS estimators of (α_1, α_3) given (τ_2, τ_3) have the following 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

$$e_t \equiv y_t - [z_{11}x_1 \ z_{31}x_1]\begin{bmatrix} \hat{\alpha}_1 \\ \hat{\alpha}_3 \end{bmatrix},$$

$$\underset{1\times T}{e_z^{1\prime}} \equiv \left[z_{11}e_1 \ldots z_{1T}e_T\right],$$

and

$$\underset{1\times T}{e_z^{3\prime}} \equiv \left[z_{31}e_1 \ldots z_{3T}e_T\right].$$

We calculate the following estimators for σ_1 and σ_3 given (τ_2, τ_3), which are asymptotically equivalent to the OLS estimators:

$$\widehat{\sigma}_1^2 = \left(\sum_{t=1}^{T} z_{1t}\right)^{-1} e_z^{1\prime} e_z^1$$

and

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

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.

While in general it may be interesting to include a larger number of regressors in x_t, the only specification that we consider in the analysis that follows is the one where x_t contains only a constant term, $x_t' = 1$ for $\forall t$ and $k = 1$. Hence, before a transition start, 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-phase model is a useful characterization of the empirical evidence, one or more of the phases might not be observed, either because of the sample is too short or because the demographic transition is still on-going. In particular, we can have six different cases, as illustrated in Figure 3 (we plot the six cases of CBR transitions, but a comparable figure exists for the CDR transitions).

In the top left panel of Figure 3, we have Case 1: all three phases are observed. In the top right panel, we have Case 2: only phases 2 and 3 are observed. In the Middle row, we see Cases 3, only phases 1 and 2 are observed, and 4, just phase 2 is observed. In the bottom left panel, we see the rare Case 5, where only phase 1 is observed, and in the bottom right panel, Case 6, where only phase 3 is observed. To distinguish Case 5 from Case 6, as they are equivalent econometrically, we use external information about the levels of the CBR and CDR to classify the country either as Case 5 or as Case 6. As we will see later, in our sample, we only estimate 4 countries in Cases 5 for the CBR and none for the CDR. We have a few more observations of Case 6, 15 for the CDR and 3 for the CBR. Cases 2 and 6 will usually be associated with vital

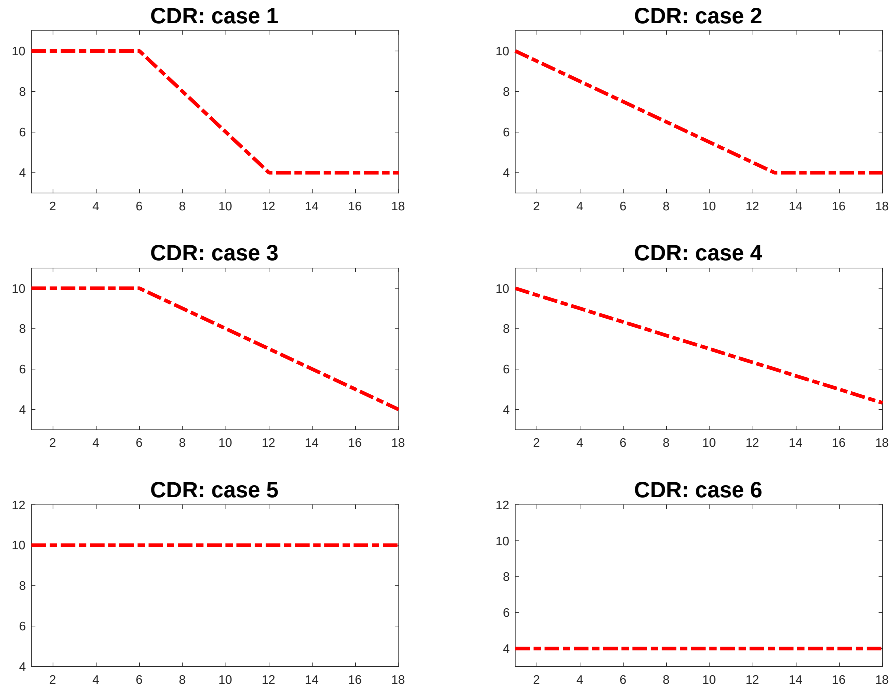


Figure 3: 6 cases of the CBR transition

statistics not going back in time for a sufficiently long period, while Cases 3 to 5 will be more often linked with ongoing transitions.

To discriminate among all these different possibilities, we estimate, for each country in the data, all six cases. Table 1 summarizes the nesting structure among cases.

Table 1: Different cases of the general model

	Parameter restriction	Explanation	Num. of parameters
Case 1	–	All 3 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

We select, among the five cases, 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. In the first pass of such selection, we use an F-test at the 95% confidence level:

$$\frac{\frac{SSE^b - SSE^a}{m^a - m^b}}{\frac{SSE^a}{T - m^a}} \tag{6}$$

where a nests b, and, as mentioned in the previous section, $m^I = 2k + 2$. We find that 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 this statistical method from over-fitting short-run anomalies in countries for which the time series is not as extensive, we also apply a set of simple auxiliary rules. For example, if he statistical method detects the end of a fertility transition at a final level of higher than 20 per thousand, with an end date less than 20 years before the end of the data series, we throw out this transition end date, moving the country from Case I to Case III, or from Case II to Case IV. A complete description of the auxiliary rules can be found in Appendix B.

3 Data

3.1 Vital statistics and GDP per capita

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. From 1960 onwards, we rely on the World Bank Development Indicators. For many countries, we fill in the period between 1950 and 1960 with data from the UNData service of the United Nations Statistics Division. To gather vital statistics before 1950, we start with data from Chesnais's (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 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 on CDRs and CBRs covers 186 countries from 1541 to 2016.

We take data on real GDP per capita (GDPpc), given in constant 2011 US Dollars purchasing power parity (PPP), from the 2018 version of Maddison's database.[4] Table 2 shows the means and the standard deviations of the CBR, the CDR, and the log GDP per capita in our sample.

[4]Bolt, Inklaar, de Jong, and van Zanden (2018). The database can be accessed here: https://www.rug.nl/ggdc/historicaldevelopment/maddison/releases/maddison-project-database-2018

The Madison data covers 165 countries from between the years 1 and 2016.[5]

Table 2: Summary statistics of demographics and GDPpc

Variable	sample mean	st. Dev.	N. Obs.
crude death rate *(CDR),* per 1000	14.1	8.0	16206
crude birth rate *(CBR),* per 1000	30.5	11.8	16198
ln GDP per capita *(lnGDPPC)*	8.3	1.1	16694

Table 3 shows the correlations across the three variables. We see i) a strong negative correlation between the CBR and the log GDP per capita; ii) a slightly less strong negative correlation between the CDR and the log GDP per capita; and iii) the positive comovement of the CBR and the CDR.

Table 3: Correlations among key variables

	CDR	CBR	lnGDPPC
crude death rate *(CDR),* per 1000	1	0.48	-0.56
crude birth rate *(CBR),* per 1000		1	-0.71
ln GDP per capita *(lnGDPPC)*			1

3.2 Projecting CDR backward

Vital statistics for only a few countries are available back into the 19th century and for a great many not until after 1950. As a result, there are numerous countries for which the start of either the CBR or the CDR transition is not observed (cases 2 and 4 in Figure 3). Since the CDR transition starts, on average, earlier than the CBR transition, we have many more "missing starts" for CDR transitions than for CBR transitions. In all, there are 89 countries for which our estimation procedure indicates that the start of the CBR transition is observed but the beginning of the CDR transition is not.

We extend our set of estimated CDR transition start dates for a subset of these countries; those for which we observe a CBR transition start and a downward trend in death rates. For these countries (107 in total), we project the downward trend in death rates backwards and impute a transition start date. We do this by assuming that the pre-transition gap between birth and death rates is equal to 8.86, which is the unweighted arithmetic mean across the 23 countries for which we observe the start of both transitions, and for which fertility transitions start prior to 1950. Hence, for these countries we assign the start of the CDR transition at

[5]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, as for the majority of the covered period Slovakia was part of Czechoslovakia.

a point when the level of the CDR is 8.86 higher than the pre-transition level of the CBR. Using this procedure, we are able to more than double the number of countries for which some estimate of the CDR transition start date is available, from 46 to 143.[6]

4 Results

Figure 4 displays time series of the CBRs and CDRs, along with the fitted 3-phase transitions for each of these rates, for six countries. Each country is a representative example of a form of demographic transition. The top left panel is the demographic transition of Great Britain/UK, 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, later than the Great Britain/UK's, but representative of many Western European countries that followed the Great Britain/UK's lead with only a few decades delay.

The right Middle panel is the demographic transition of Spain, a late but completed transition, with the CBR stabilizing as recently as 1999. The left Middle panel is the demographic transition for Chile, a typical case of late and on-going 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 of CBR has even started. Table A in the Appendix documents the start and end dates of the demographic transition for each country in our sample.

Table 4: Summary statistics

	CDR	CBR
mean initial level	27.05	42.87
mean lnGDPpc at transition start	7.59	7.91
N	65	123
mean final level	8.02	13.02
mean lnGDPpc at transition end	8.63	9.51
N	79	54

Table 4 presents some summary statistics of the CDR and CBR at the start and end of the transitions as well as log GDPpc the for those countries that we observe starts (or ends) of such transitions. We can see, in particular, the large drop of around 66% in both mean CDR and

[6] In order to make sure we are left with a set of reasonable estimates, we throw out the 10 imputed start dates found using this method which are more than 100 years before the first year of CDR data for their respective countries.

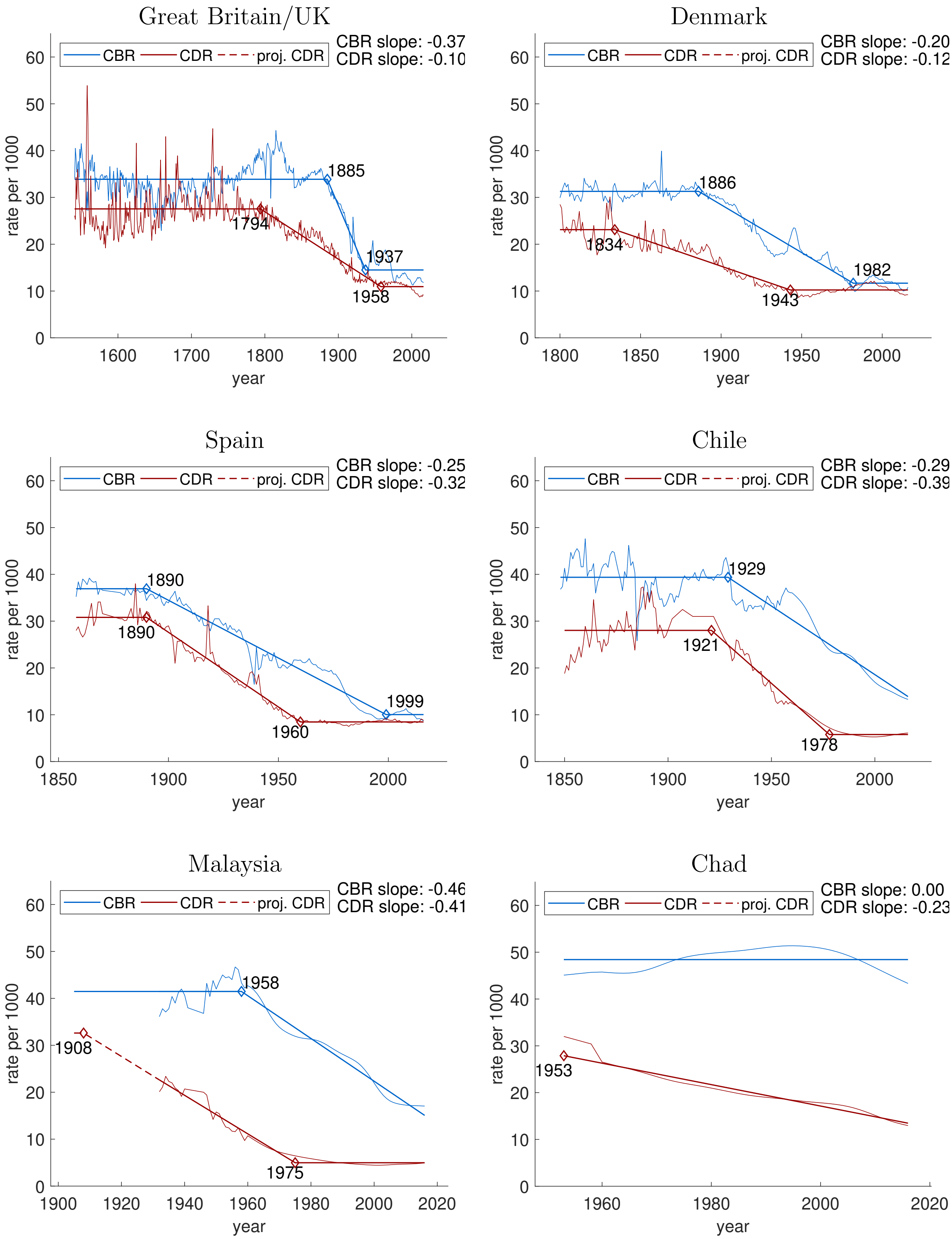


Figure 4: Six examples of demographic transitions.

CBR, with a difference between both of them much small at the end of the transition than at the start.

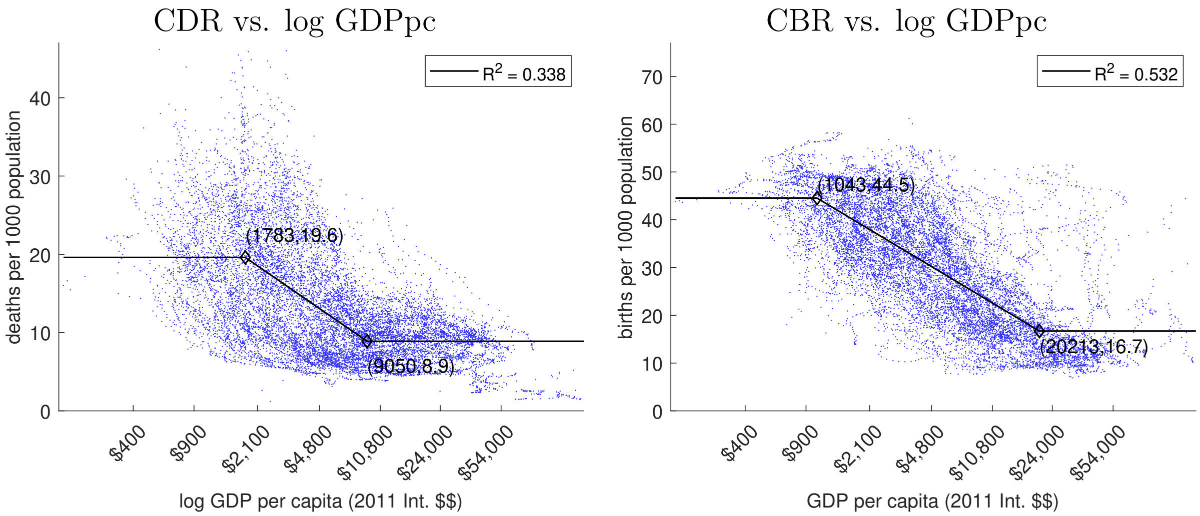


Figure 5: CDR and CBR vs. log GDPpc.

Figure 5 displays scatter plots of CDR and CBR, using data for every country in every year in the sample, against log GDP per capita. Superimposed onto the plots is the best fit for a 3-phase transition as specified previously, but with log GDP per capita taking the place of time. While admittedly a crude first exercise, this structure provides a reasonably good fit for the panel data with R^2 coefficients of 0.338 and 0.532, respectively. According to this estimation, the "average" pre-transition CDR for the entire panel is 19.6 per year per 1000 people, and the pre-transition CBR for the entire sample is 44.5. The estimated post-transition CBR and CDR for the entire sample are 8.9 and 16.7, respectively. The CDR transition is estimated to start, on " average," when a country achieves a real GDP per capita of $1,783 constant 2011 constant US dollars PPP. The "average" start of the CBR transition is estimated to be at the lower level of $1,043. The end of the CDR and CBR transitions are placed at $9,050 and $20,213, respectively.

Table 5 documents the distribution of all countries in our sample according to different cases outlined in Table 1. Out of 186 countries, we have 175 countries that have completed the mortality transition (Cases 1 and 2) and 80 that have completed the fertility transition. This shows how the global drop of death rates is considerably more advanced than the decline of birth rates: most of the planet has finished the drop in CDRs, but there is still much space to cover in the fall of CBRs. Notice how for the CDR, we have large count (131) of countries where stages 2 and 3 of the transition are observed, but not stage 1, most likely because data does not go back enough in time. We do not find any country where the start of 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 7 countries in Case 6 of the CDR. These are typically Eastern European countries that started their demographic transitions earlier than the availability of data.

Table 5: Case counts

CDR \CBR	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Total
Case 1	27	0	17	0	0	0	44
Case 2	26	20	79	6	0	0	131
Case 3	0	0	1	0	1	0	2
Case 4	0	0	2	0	0	0	2
Case 5	0	0	0	0	0	0	0
Case 6	0	7	0	0	0	0	7
Total	53	27	99	6	1	0	186

Figure 6 plots the empirical frequency of log GDP per capita at the start of each type of transition. These distributions are roughly uni-modal, which may be adequately approximated by a normal distribution.

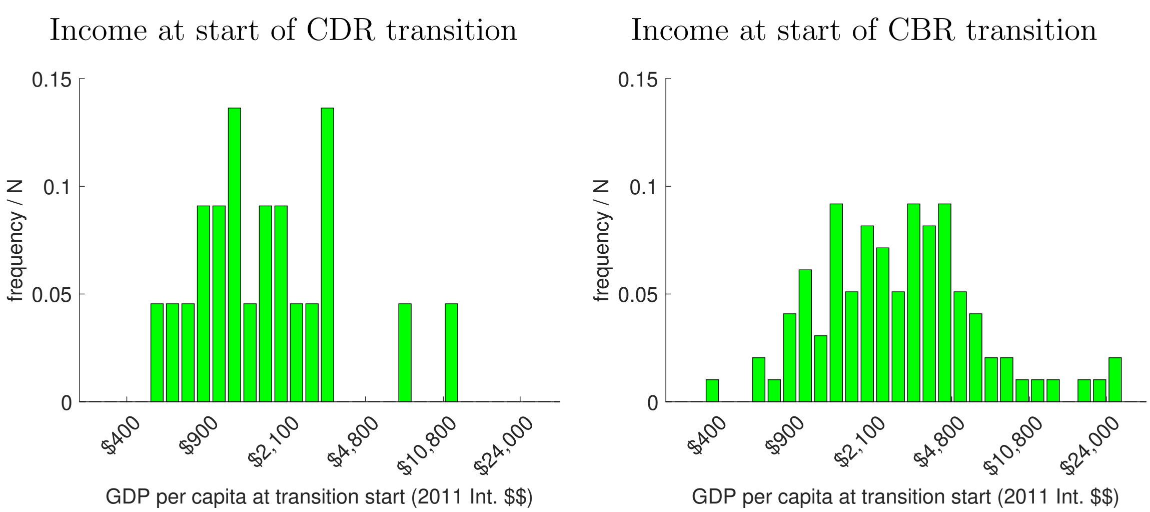


Figure 6: log GDP per capita at the start of each transition.

4.1 Are demographic transitions getting faster?

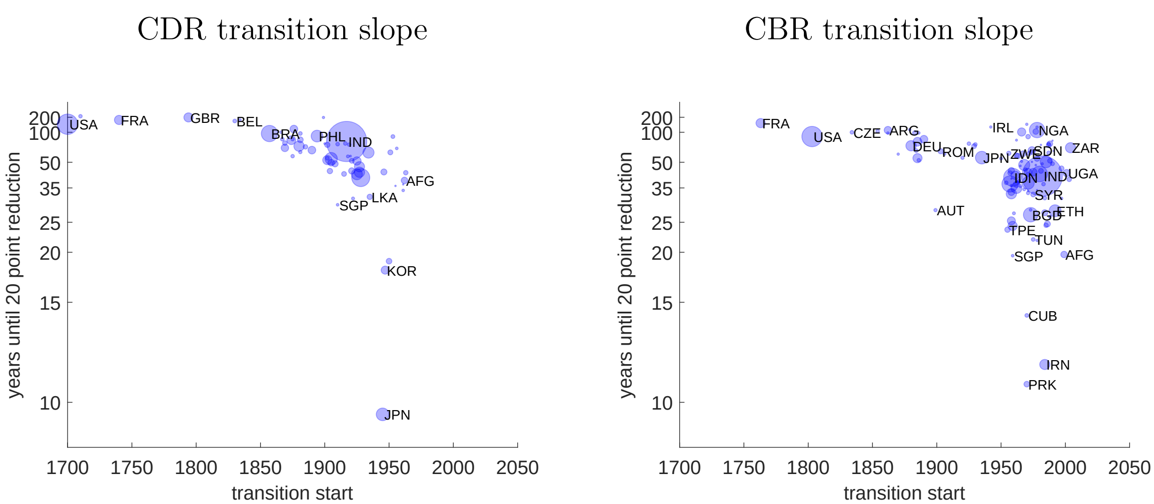
Table 6 reports summary statistics for some features of countries as they enter the CDR and CBR transitions, broken into groups according to the period in which their transition started. Table 6 reveal three patterns of interest. The first pattern is that start dates of the CDR transitions are more dispersed over time than the start dates of the CBR transitions. The former also peak sooner, with many starts clustered between 1900 and 1960. In comparisons, most CBR transitions start between 1960 and 1990, with 9 transitions starting since 1990.

Table 6: Countries entering transitions

	before 1870	1870-1900	1900-1930	1930-1960	1960-1990	after 1990	All
mean initial lnGDPpc	7.72	7.75	7.55	7.38	7.58	–	7.59
mean initial 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

	before 1870	1870-1900	1900-1930	1930-1960	1960-1990	after 1990	All
mean initial lnGDPpc	7.52	8.39	7.70	7.94	7.97	7.33	7.91
mean initial 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	-0.57	-0.50	-0.51
N	6	11	5	19	71	11	123

The second pattern in Table 6 is that later transitions are faster. The slope of the reduction in CDR and CBR during the transition (i.e., the decline in the rates per year) is much larger for later transitions. Figure 7 shows this pattern graphically for all the countries in our sample with complete transitions. An alternative way to make the same point is to plot, in Figure 8, the measured transition length from plateau to plateau.[7]

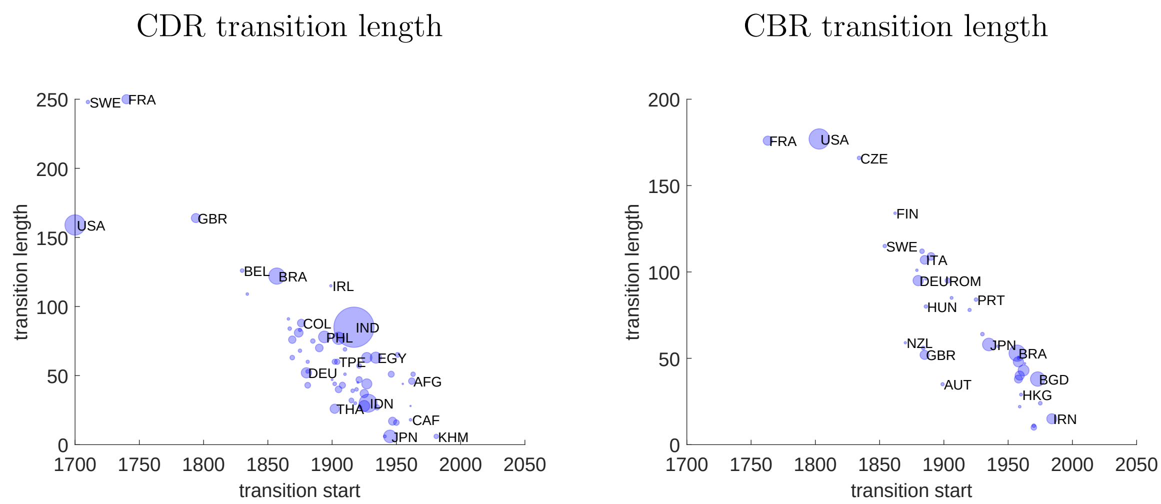


NOTE: The size of the circle for each country is proportional to its share of the 2016 world population.

Figure 7: Transition slopes.

To measure the strength of this downward trend more precisely, we use a linear regression, which allows us to control for additional factors that may affect transition speed and length beside timing. We hypothesize that, in addition to the timing of the transition start, the speed of the transition may also be affected by the level of GDP per capita at the transition start

[7] The circle for each country in these plots is proportional to its share of the 2016 world population.



NOTE: The size of the circle for each country is proportional to its share of the 2016 world population.

Figure 8: Transition lengths.

and by how high crude birth rates were initially.[8] Table 7 displays the results of the linear regressions for the slope and length of the transition speeds. In each case, the transition start date is significantly related to transition speed.

Table 7: Transition Speed

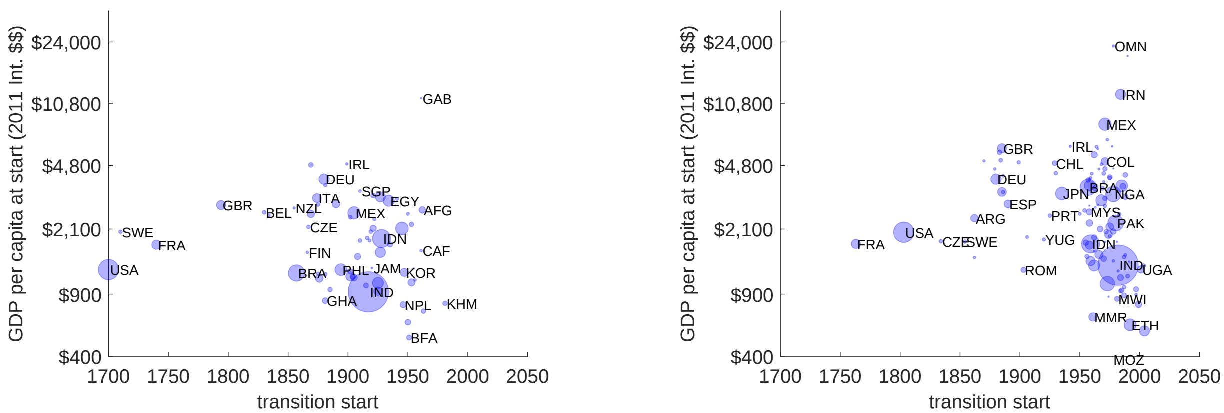
Dependent variable	CDR slope	CBR slope	CDR length	CBR length
Cons	159.33	0.21	-0.48	0.16
	(4.57)	(0.54)	(-0.32)	(2.35)
ln GDPPC at start	-4.90	-0.05	0.04	-0.02
	(-1.49)	(-1.26)	(0.25)	(-3.24)
starting CBR /10	-0.75	0.04	0.17	-0.08
	(-0.19)	(0.93)	(1.20)	(-1.30)
start date /10	-3.36	-0.03	-0.05	-0.00
	(-6.12)	(-4.24)	(-3.28)	(-2.04)
N. Obs.	110	110	63	102
R^2	0.285	0.149	0.156	0.139

The third pattern in Table 6 is that, while the GDP per capita at the start of the CDR transition is lower for later transitions, there is no clear trend in the GDP per capita at the beginning of the CBR transitions. The GDP per capita is remarkably similar, for example, for

[8]The initial level of the CBR is highly correlated with the initial level of CDR. Thus, including the latter in the regression does not significantly affect the results.

the CBR transitions that started during the 1870-1900 period and the 1960-90 period. Figure 9 shows scatter plots of log GDP per capita in each country at the start of its CDR and CBR transition, respectively.

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



NOTE: The size of the circle for each country is proportional to its share of the 2016 world population.

Figure 9: Log GDPpc at the start of transitions.

5 An empirical analysis of demographic transitions

In the previous section, we saw that the distributions of log GDP per capita levels at the start of transitions in crude birth or death rates are fairly stable over time and possess unimodal distributions. This suggests that a modeling strategy that links the level of log GDP per capita to transition takeoffs may have some explanatory power. One possible approach is to model the start of each transition as a random event whose probability of occurring depends on log GDP per capita and possibly other variables. Let T represent the time at which a one-off event, such as the start of a CDR or CBR transition, occurs. Suppose that the probability of the event occurring at time t in country i, conditional on not having occurred previously, can be expressed as:

$$\Pr(T^i = t | T^i \geq t) = G\left(\sum_{l=0}^{k-1} x_{l,it}\beta_l\right), \tag{7}$$

where $G(.)$ is a function bounded between 0 and 1. In the exercise that follows, we will assume that $G(.)$ is the logistic cumulative distribution function and that $(x_{0,it}, x_{1,it}, ..., x_{k-1,it})$ are a set of k explanatory variables.

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 time for $t \in \{1, 2, 3, ...T\}$. Let T^i represent the time at which a given one-off event occurs in country i, and let $\mathcal{I}_{it}$ be an indicator function taking the value 1 if the event occurs in country i at time t and 0 otherwise. Let the conditional probability of a transition be given by equation (7). The parameters of this model can then be estimated by maximizing the log-likelihood:

$$\log L_N = \sum_{i=1}^{N} \sum_{t=1}^{T_i} \log \left[\mathcal{I}_{it} G \left(\sum_{l=0}^{k-1} x_{l,it} \beta_l \right) + (1 - \mathcal{I}_{it}) \left(1 - G \left(\sum_{l=0}^{k-1} x_{l,it} \beta_l \right) \right) \right]. \tag{8}$$

Table 8: GDPpc and CBR transition, Logit results

Variable	Estimates
Cons	-55.79
	(17.22)
lnGDPPC	10.31
	(4.26)
lnGDPPC ^{2}	-0.49
	(0.26)
LLn	-254.1
Pseudo-R^2	0.184
N	19230

We want the estimates we obtain to be informed by the fact that every country in the world existed for a very long time without experiencing a demographic transition. For this purpose, we construct a balanced panel with yearly, interpolated values for real GDP per capita and transition status, starting in the year 1500, more than 250 years before the first observed CBR transition start. 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 conservative imputations for a small set of additional countries. These are countries which have some pre-modern GDP per capita data in the Maddison dataset, though not for the year 1500 specifically. We make imputations for 37 countries.[9] After excluding countries for which we do not observe the start of the CBR transition, this gives us a panel of 42 countries between 1500 and 2016.

Table 8 reports the Logit estimation for the CBR (the results for the CDR are reported in the Appendix E) when the only explanatory variable is log GDP per capita. As shown in Figure 10, this specification replicates well the distribution of log GDP per capita at the start of 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 simplest specification is sufficient to generate the observed aggregate timing of transition starts across levels of GDP per capita. It does not does not do as well, however, in matching the observed

[9] These imputations are described in Appendix C.

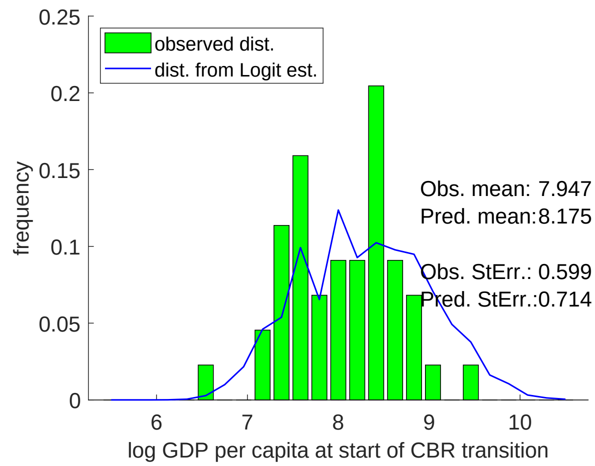


Figure 10: Distribution of log GDPpc at the start of the CBR transitions

timing of transition starts across actual time.

Figure 11: Within Sample Predictions

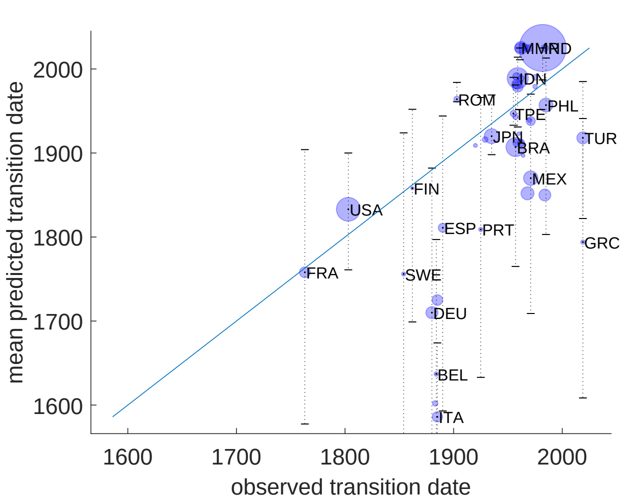


Figure 12: Distribution of Transition Dates

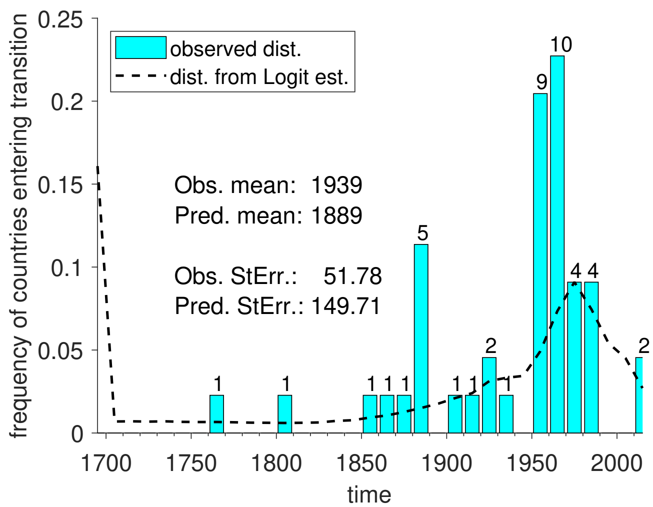


Figure 11 plots observed and predicted start dates for individual countries. Three-letter country abbreviations and 60% confidence intervals are plotted for a subset of countries. As we can see in Figure 11, the mean predicted transition dates for the majority of countries are close to the 45 degree line. The confidence intervals are in general quite large, though they are generally smaller for late transitions than for early ones. This can be accounted for by the fact that growth in GDP per capita was generally faster in the second half of the 20th century than the second half of the 19th, meaning that late transitioners, on average, pass through the critical "window" of GDP per capita levels over a shorter span of time. Figure 11 also clearly shows that the mean predicted start dates are also quite early for most of the early-transitioning European countries. This can be attributed to the fact that several of these European countries enjoyed levels of GDP per capita throughout the 17th and 18th centuries which, while low by today's standards, were higher than what most of the late-transition countries would achieve until the latter half of the 20th century. The inclusion of country fixed effects would, obviously, be able to bring each country's mean predicted transition start in line with the observed date, but would do nothing to narrow the confidence intervals.

The early predicted transitions show up as a large mass of predicted transitions prior to the year 1700 in Figure 12, which plots the observed distribution of start dates across decades against the distribution generated by the model. The bars represent observed transition starts, with the number just above each bar representing the number of transitions observed starting during that decade. The dotted line represents the predicted density of transition starts over time. Aside from the large mass of early starts, the predictions match the remainder of the distribution fairly well, replicating, in particular, the peak of transition starts in the 1960s and 1970s.

5.1 Demographic contagion

The shortcomings of the simplest specification to account for key features of the aggregate distribution of transition start dates across time and the large confidence intervals for predicted start dates for individual countries motivate us to explore additional plausible drivers of demographic change. One such plausible driver, documented by Spolaore and Wacziarg (2014) inter alia, is spillover effects across country borders. We model spillover effects by adding $\mathcal{A}_{it}$ as an explanatory variable, representing country i's access to countries that have started their transition prior to period t:

$$\mathcal{A}_{it} \equiv \left[\sum_{j=1}^{N} g_{ij} \mathcal{I}_{j,t-1} \right]^{\psi} . \tag{9}$$

Access to transitions is calculated as a weighted sum of all countries which have already begun their transitions, where weights are determined by distance. In the above expression, $\mathcal{I}_{j,t}$ is an indicator function taking a value of 1 if country j started its transition before period t, and 0 otherwise. The inverse bilateral distance between countries i and j is represented by g_{ij} . If country j is very far from country i, then g_{ij} is close to zero, and whether or not country j has already started its transition has little effect on the probability that country i starts its transition. On the other hand, if country j is close to country i, then g_{ij} is close to 1, and if country j has already started its transition this could increase the probability of a transition in country i considerably. The parameter $\psi > 0$ adds curvature. If $\psi < 1$, then each additional transition start has a small marginal impact on the probability of future transition starts. If $\psi > 1$, then each additional transition start has a larger marginal impact.

We parameterize inverse bilateral distance g_{ij} as a function of geographical and possibly other types of barriers to contact between countries.

$$g_{ij} = \exp\{\mathbf{z}_{ij}'\gamma\}, \tag{10}$$

where $\mathbf{z}_{ij}$ is a column vector of bilateral distance measures and γ is a vector of coefficients. Given (9) and (10), we estimate the following equation:

$$\Pr(T^i = t | T^i \geq t) = G\left(\sum_{l=0}^{k-1} x_{l,it}\beta_l + \beta_k \mathcal{A}_{it}\right), \tag{11}$$

In spite of the fact that this is no longer a linear model, the parameter vectors β and γ and the parameter ψ can be still be estimated by minimizing the log-likelihood function given by (8).

We take data on geographic distances between countries from Mayer and Zignago (2011). In particular, we make use of the $distw_{ij}$ weighted distance measure, which is calculated by taking the average great circle distance between each of country i's cities and each of country j's cities, weighted by the share of each city in the national population. We borrow data on linguistic and cultural barriers to contact from Melitz and Toubal (2013).[10] To infer linguistic barriers, we use Melitz and Toubal's (2013) "CL" measure of linguistic proximity, which they construct using data on the distribution of spoken languages and Bakker, Müller, Velupillai, Wichmann, Brown, Brown, Egorov, Mailhammer, Grant, and Holman's (2009) calculation of linguistic similarity.[11] To reflect connections that may exist between countries for historical reasons independently of shared language, we also consider Melitz and Toubal's (2013) index of shared religion and a dummy variable for common legal origins. Table 9 displays summary statistics for these variables, and the correlation table is given in Table 10.

Table 9: Distance measures, summary statistics

Variable	sample mean	st. Dev.	N. Obs.
ln Distance, km *(ldi)*	7928.3	4521.4	33856
Linguistic proximity *(lp2)*	0.1	0.1	33856
Common religion *(cmr)*	0.1	0.3	34596
Common legal system *(cml)* $\in \{0,1\}$	0.1	0.2	34596

[10] Melitz and Toubal (2013) investigate the importance of these non-geographic factors as barriers to trade. They build on a large literature in international trade that estimates gravity equations where the distance between countries considers both geographical measures and the effects of language and other related factors. Egger and Lassmann (2012) provide an overview.

[11] Melitz and Toubal (2013) construct and test several alternative measures of the degree of linguistic commonality between countries, ranging from the narrowest definition, simply recording whether the two countries share an official language or not, to more nuanced definitions based on the shares of the population in each country that speak the same or similar languages. "LP2" is comprehensive yet relatively parsimonious.

Table 10: Distance measures, correlations

	lp2	*ldi*	*cmr*	*cml*
ln Distance, km *(ldi)*	1	-0.10	-0.08	-0.14
Linguistic proximity *(lp2)*		1	-0.13	0.15
Common religion *(cmr)*			1	0.57
Common legal system *(cml)* $\in \{0, 1\}$				1

The linguistic, religious, and legal *proximity* measures (*cl*, *cmr*, and *cml*) are transformed into *distance* measures by calculating *distance* = 1 - *proximity*. Missing bilateral distances are imputed to take the maximum theoretical value for that distance–1 in the case of $1 - cl$, $1 - cmr$, and $1 - cml$, and (the natural log of) 20,015 km in the case of great-circle distance *(ldi)* between capital cities.[12] Finally, log geographical distance *ldi* is divided by $ln(20,015)$ so that this distance measure, too, is normalized to fall between 0 and 1.

For estimation, we use the same balanced panel of 42 countries as before. Both own GDP per capita and the sequence of other countries transitions are taken as completely exogenous. A series of access to transitions $\mathcal{A}_{it}$ is constructed for each country using the observed transition start dates of a broader sample of 152 countries. We estimate several specifications of equation (11), the results of which are shown in Table 11. Column (1) reports again the results shown in Table 8. Specification (2) is the simplest specification that includes some sort of spillover effects. In this case they are global, and represented by a simple, unweighted count of the number of countries that have begun the transition. Specification (3) is slightly more sophisticated, adding curvature to this global sum. The estimated value of ψ, being less than 1, implies that there are diminishing returns–the more countries have already entered the transition, the smaller the effect of each additional country on other countries' odds of entering the transition.

Specifications (4) through (9) include measures of access to transition that are local–the influence of one transitioned country on other countries according to the inverse distance between them. In specification (4), the only distance measure is the log geographic distance. As it turns out, the discrete, non-parametric measure of distance used in specification (5) has more success in replicating the data. For both formulations, the association of shorter geographic distances with stronger spillover effects is important and statistically significant.

In specification (6) and (7), no significant association with either linguistic or religious distance is detected, after accounting for GDP per capita. In both cases, the estimated coefficient on distance is negative, and the coefficient on access to transitions is very small and not statistically significant. The results for specification (8) show a statistically significant association with legal distance at the 95% level. Finally, specification (9) includes geographic distance and

[12]The circumference of the Earth is 40,030 kilometers, and so the maximum great-circle distance between any two points on the globe is approximately 20,015 kilometers (the Earth not being perfectly spherical).

legal distance in the same estimation. Geographic and legal distance are both found to have coefficients which are statistically significant at the at the 95% level.

In these estimations we have not done anything to account for the endogeneity of past transition starts, and these results should not be interpreted causally. We believe, however, that taken together, the results of these non-linear Logit regressions strongly suggest the existence of some type of demographic spillover effects.

In Figure 13, we look at the access to transitions measure implied by specification (9) (the distributions displayed in all of these figures are smoothed using a Gaussian kernel). Using the estimated parameters, access is calculated as

$$\mathcal{A}_{it} \equiv \left[\sum_{j=1}^{N} \exp[\mathcal{D}_{ij} + 0.90 \cdot \mathrm{cml}_{ij}] \mathcal{I}_{j,t-1} \right]^{0.41},$$

where

$$\mathcal{D}_{ij} \equiv 1.80 \cdot \mathbf{1}\{\mathrm{ldi}_{ij} < \ln 800\} + 0.57 \cdot \mathbf{1}\{\ln 800 \leq \mathrm{ldi}_{ij} < \ln 2000\}.$$

is the step variable for distance.

The top left panel of Figure 13 shows the distribution of this measure at different points in time. Not surprisingly, as more countries transition, this distribution moves steadily to the right. The top right panel of Figure 13 plots the transition probabilities implied if each country is assigned its actual access to transitions value and GDP per capita equal to \$2000. Here we can see that in 1850, 1900, and 1950, "Access to transitions" in the great majority of countries was such that their probability of transition at \$2000 GDP per capita would have been relatively small. In the year 2000, this situation changes dramatically, and the lowest yearly probability of transition for any country with \$2000 GDP per capita would be 10%.

The bottom left panel of Figure 13 shows the evolution of the distribution of GDP per capita over time. This distribution shifts right as time passes and more countries enjoy higher levels of GDP per capita. The bottom right panel of Figure 13 shows the distribution of the probability of transition, given the observed GDP per capita for each country, assuming they have the mean level of "Access to transitions" existing in the year 2000. This panel demonstrates the importance of the complementarity between a country's level of development and the influence of its neighbors. In 1850, even countries with relatively high log GDP per capita had a low transition probability. In comparison, by 2000, a country with the relatively low level of GDP per capita (\$2000) has a probability of transition close to 1 if enough of their neighbors have already started the transition.

In Appendix E.1 we repeat all the exercises described in this section for the CDR. The lessons are very similar except that the neighborhood effect is weaker for mortality transitions.

Finally, Figures 14 and 15 shows the improvement of specification (9) in matching the

Table 11: Determinants of the start of the CBR transition

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
cons	-55.79	-73.97	-61.95	-55.27	-49.10	-61.71	-68.59	-53.69	-35.95
	(17.22)	(18.40)	(18.58)	(17.96)	(18.09)	(20.48)	(18.97)	(14.33)	(17.84)
lnGDPPC	1.03	1.61	1.29	10.97	9.40	1.23	14.57	10.76	6.19
	(0.43)	(0.46)	(0.47)	(4.45)	(4.49)	(0.51)	(4.79)	(3.74)	(4.46)
lnGDPPC ^{2}	-0.00	-0.01	-0.01	-0.63	-0.54	-0.07	-0.86	-0.63	-0.35
	(0.00)	(0.00)	(0.00)	(0.28)	(0.28)	(0.03)	(0.29)	(0.25)	(0.28)
access		0.13	0.75	6.77	3.18	0.12	0.03	1.35	5.14
		(0.01)	(0.44)	(1.75)	(0.45)	(0.08)	(0.15)	(0.26)	(0.72)
geo dist.				4.39					
				(0.83)					
$<$ 800km					1.83				1.80
					(0.18)				(0.33)
800-2000km					0.53				0.57
					(0.19)				(0.29)
ling. dist						-5.87			
						(0.15)			
relig dist							-5.04		
							(0.00)		
legal dist								0.71	0.90
								(0.19)	(0.41)
ψ, curv.			0.57	0.47	0.51	0.41	0.61	0.51	0.50
			(0.10)	(0.36)	(0.09)	(0.01)	(0.01)	(0.06)	(0.37)
LLn	-254.1	-208.6	-206.2	-202.9	-198.5	-209.2	-204.6	-205.9	-196.9
Pseudo-R^2	0.184	0.330	0.338	0.349	0.363	0.328	0.343	0.339	0.368
N. Obs.	19230	19230	19230	19230	19230	19230	19230	19230	19230

Note: Standard errors of the estimated coefficients are given in parentheses.

observed distribution of transition starts across time. Aside from calling France and the United States too late on average, the specification with demographic contagion does appreciably better

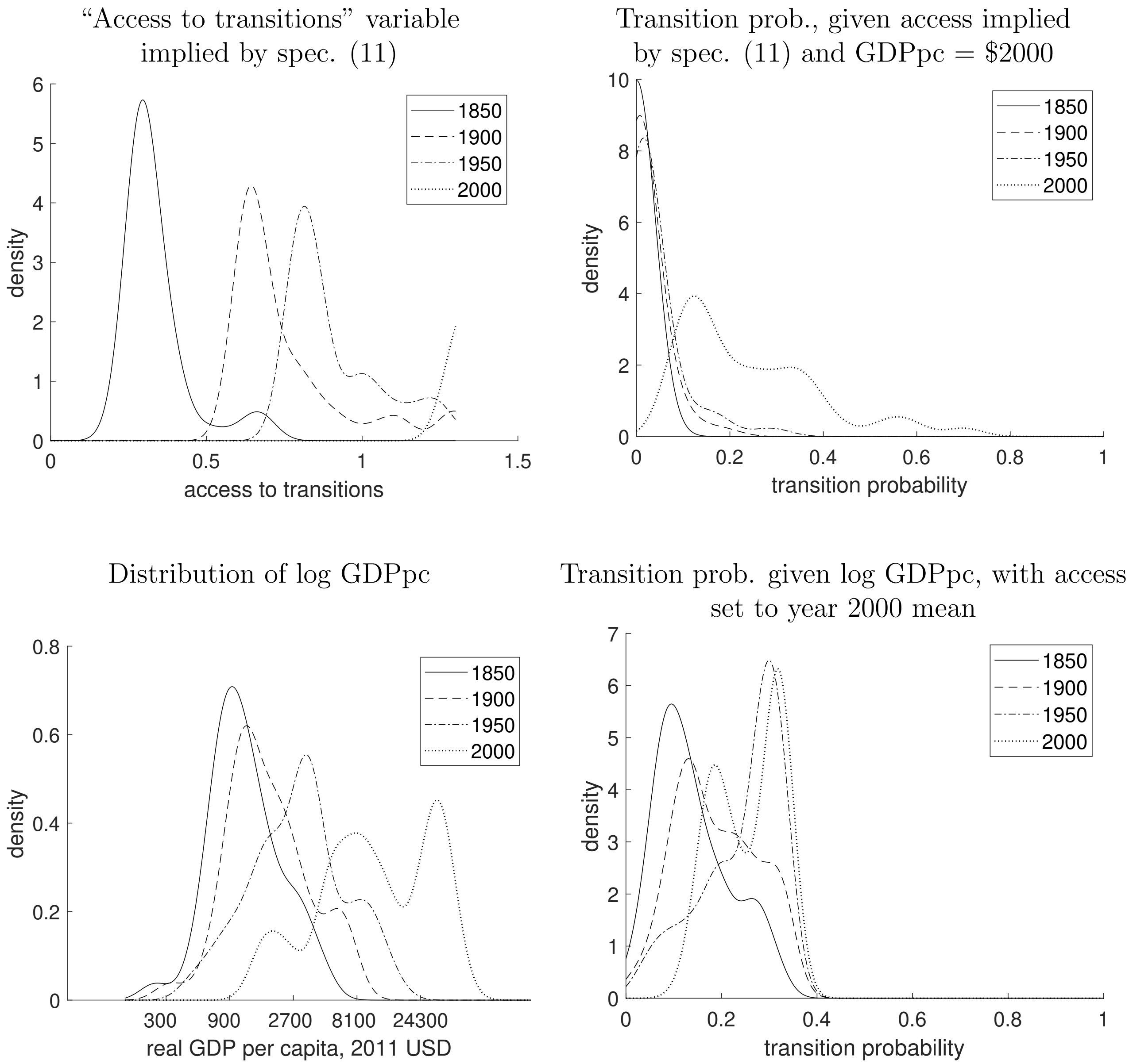


Figure 13: Demographic contagion.

on all dimensions.

5.2 A recap

In this section and the previous one, we have documented three findings. First, transitions in both fertility and mortality have been getting faster over time. Second, in spite of this increase in the speed of the transitions, there is no clear trend in the level of GDP per capita at which countries enter the fertility transition. Finally, we have found suggestive evidence for a kind of “demographic contagion,” whereby a transition in one country is statistically associated with following transitions in countries which are close to it geographically and linguistically and have

Figure 14: Within Sample Predictions, Spec. (9)

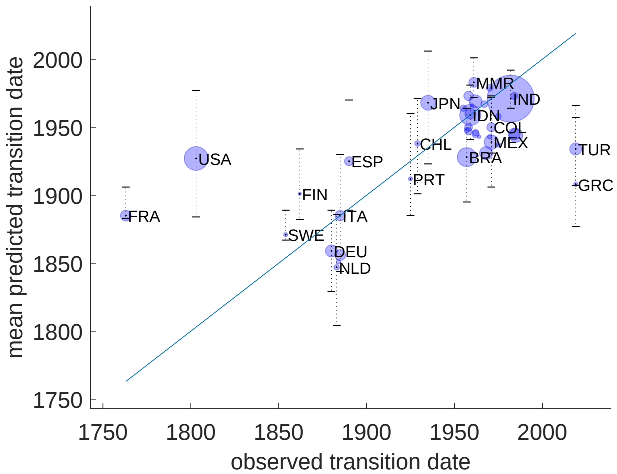
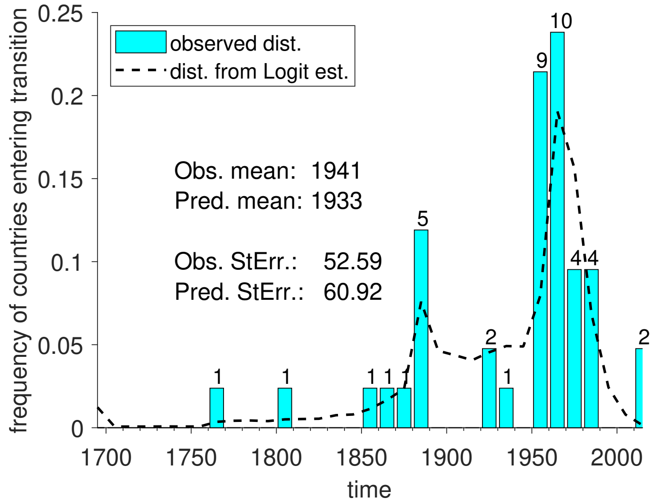


Figure 15: Distribution of Transition Dates, Spec. (9)



similar legal systems.

6 Model

This section builds a model of endogenous fertility, education, and technology diffusion to understand the trends we have documented. In the model, parents face a quantity-quality trade-off between how many children to have and how much to educate them, following classic work by Barro and Becker (1989). We propose an economy with traditional and modern sectors, as in Hansen and Prescott (2002). With economic growth, economic resources move from traditional to modern sector and the skill premium increases. The economic growth starts first in Britain and then diffuses to other countries like Lucas (2009). The catch-up of countries depends on how far they are from Britain, geographically, and culturally.

6.1 Consumer preferences, fertility, and education decisions

Consider a world that consists of many locations, which will correspond to countries in our analysis. Consumers in each location 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. For ease of exposition, below we drop country index i whenever this does not cause any confusion. 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 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 the equilibrium unskilled and skilled wages, w_t^U and w_t^S, and their level of human capital, h_t. In exchange for each unit of labor supplied, adults receive

$$y_{t+j-1} \equiv \zeta^j \left(w_{t+j-1}^U + h_{t+j-1} w_{t+j-1}^S\right) \text{ for } j = 1, 2, 3$$

Hence, we assume that all workers at time t get w_t^U for one unit of raw labor they have and are paid an additional w_t^S for their skills.

Young adults choose how many children to have, n_t, and how much education, e_t, to provide for each of them. The probability of an infant, who is born at time t, surviving birth and becoming a child is s_t^0. Hence, if a young adult chooses to have n_t births in period t, she will have $s_t^0 n_t$ surviving children to educate. The probability of a child, born in period 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.

Each birth requires a time commitment of τ_1. The education a young adult chooses to give to each child that survives infancy is denoted by e_t. Educating each of these surviving children requires a time investment of τ_2. To achieve a level of education e_t for each child, parents must pay a total time cost of $n_t e_t \tau_2$. 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, given by

$$h_{t+1} = h_t^{\upsilon} e_t^{\xi}, \text{ with } \upsilon, \xi \in (0,1).$$

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

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

and

$$h_{t+1} = h_t^{\upsilon} e_t^{\xi}.$$

Adults derive utility from the number of children who survive infancy, $s_t^0 n_t$. The parameter $\gamma \geq 0$ represents the strength of this preference, and $\overline{n}$ is a parameter representing the minimum desired number of descendants who survive to adulthood. Adults 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. Finally, parents and each child that survives infancy must also be provided with $\bar{c}$ units of sustenance.

The first order conditions for an interior solution for n_t is given by

$$\frac{1}{c_t^1}\left[(w_t^u + w_t^s h_t)(\tau_1 + \tau_2 s_t^0 e_t) + s_t^0 \bar{c}\right] = \gamma \frac{s_t^0}{s_t^0 n_t - \overline{n}}, \tag{12}$$

where the left and right hand sides represent the marginal cost and marginal benefit. The marginal cost is increasing in the time cost of children, τ_1, and, if $e_t > 0$, in the time cost of education, τ_2. It is also increasing e_t, leading to a quality-quantity trade-off. Similarly, the first order condition for $e_t > 0$ is given by

$$\frac{1}{c_t^1}\left[(w_t^u + w_t^s h_t) n_t s_t^0 \tau_2\right] = \frac{\phi \xi s_t^1 h_t^{\upsilon} w_{t+1}^s}{w_{t+1}^u + h_t^{\upsilon} w_{t+1}^s e_t^{\xi}} e_t^{\xi - 1}, \tag{13}$$

where the marginal cost is increasing in the number of children and the marginal benefit is increasing in the skill premium at time $t+1$.

6.2 Vital Statistics

Given the numbers of age-1 and age-2 adults at time t, denoted by N_t^1 and N_t^2, the number of adults at time $t+1$ is given by

$$N_{t+1}^1 = n_t s_t^0 s_t^1 N_t^1,$$

and

$$N_{t+1}^2 = s_t^2 N_t^1,\ \ N_{t+1}^3 = s_t^3 N_{it}^2.$$

The crude birth rate, the number of births per person, at time t is given by

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

and the crude death rate is given by

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

We assume that the survival probabilities are determined as a function of medical technology at time t, denoted by M_t

$$s_t^j = \sigma_j\left(M_t\right) = 1 - (1 - s_0^j)\frac{1+e^{1-\delta}}{1+e^{M_t-\delta}},\ \text{ for } j = 0,1,2,3, \text{with } M_0 = 1 \text{ and } \delta \geq 0. \tag{14}$$

Hence, when $M_0 = 1$, the survival rates are given by s_0^j. As the medical technology M_t increases, the survival rates also increase. The parameter δ determines the how fast the survival rates increase with M_t. As $\delta \to \infty$, s_t^j remains constant at its initial value independent of M_t, while as $\delta \to 0$, it converges more quickly to 1.

6.3 Production

For any country i, the economy consists of two sectors: ancient and modern. Ancient sector production is carried out by a representative firm with the following decreasing returns to scale production function,

$$Y_{t,a} = A_t L_{t,a}^{\alpha} H_{t,a}^{\rho_a - \alpha} T_{t,a}^{1-\rho_a},$$

where A_t represents the total factor productibity (TFP) in the ancient sector, $K_{t,a}$ represents the aggregate ancient capital stock, $L_{t,a}$ represents the amount of unskilled labor used in the ancient sector, $H_{t,a}$ represents the quantity of skilled labor used in the ancient sector, and $T_{t,a}$ represents the land used in the ancient sector.

Modern sector production is carried out by a representative firm using the following production function:

$$Y_{t,m} = B_t L_{t,m}^{\beta} H_{t,m}^{\rho_b - \beta} T_{t,m}^{1-\rho_m},$$

where A_t represents time-t TFP in the modern sector, L_t represents the amount of unskilled labor employed in the modern sector, and H_t represents the amount of skilled labor employed in the modern sector. It is assumed that $\alpha > \beta$, i.e. traditional sector has a greater unskilled-labor intensity. of the traditional sector. We assume that there is a fixed amount of land used in each sector and it is normalized to $T_{t,a} = T_{t,m} = 1$. Then the representative firm in each sector solves the standard profit-maximization problem. This, combined with the perfect mobility of skilled and unskilled labor, implies that in equilibrium

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

and

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

6.4 The History of the World

Until now we have specified a model of how production happens in this economy and how parents make the decisions that generate the human capital resources of each country. We will now show how we use this model to describe the timeline of world economic and demographic changes in a stylized way. In the beginning all countries are identical, and all growth rates are zero. Then, at time τ, medical and productive technologies in a frontier country begin to grow. At the same time, the costs of geographic and cultural distance, which a barrier to technological diffusion, begin to fall. As the barriers to diffusion are reduced, the technological progress of the frontier country diffuses, first to its near neighbors, and eventually to the whole world. Every country eventually takes off into economic growth and demographic change.

Initially, prior to τ, the realized growth rates of ancient and modern technology as well as the growth rate of the medical technology are, in every country of the world, equal and zero. Hence, for any country i

$$\frac{A_{i,t+1} - A_{i,t}}{A_{i,t}} = \frac{B_{i,t+1} - B_{i,t}}{B_{i,t}} = \frac{M_{i,t+1} - M_{i,t}}{M_{i,t}} = 0 \text{ for all } t < \tau.$$

Let A_0, B_0 and M_0 denote the level of these technologies in pre-growth period. We provide a characterization of the steady state equilibrium in Appendix G.

The growth rates of technology in the frontier country after period τ are constant and given by

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

Each country has a fixed distance to the frontier country, which in our quantitative analysis will be Great Britain. This time-invariant distance is denoted by $d_{i,0}$. After period τ, the take-off period for Great Britain, the costs imposed by these distances distances start to fall at a constant rate, reflecting improvements in transport and communication technology.

$$d_{i,t+1} = d_{i,t}(1 - \varphi)$$

where φ represents the speed at which the cost of distance is reduced after period τ.

Once Great Britain starts growing at $t = \tau$, the other countries follow her lead. For a country i that is at a distance d_{it} from Great Britain, the growth rates of A_t, B_t and M_t are given by

$$\mu^b_{t,i} = \mu^b_f \times \exp(-d_b \times d^\lambda_{it}) \times \left(\frac{B_{t-1,f}}{B_{t-1,i}}\right)^\theta,$$

$$\mu^a_{t,i} = \mu^a_f \times \exp(-d_a \times d^\lambda_{it}) \times \left(\frac{A_{t-1,f}}{A_{t-1,i}}\right)^\theta,$$

and

$$\mu^m_{t,i} = \mu^m_f \times \exp(-d_m \times d^\lambda_{it}) \times \left(\frac{B_{t-1,f}}{B_{t-1,i}}\right)^\theta.$$

$d_b, d_a, d_m > 0$ represent the strength of distance as a barrier to the diffusion of technology in the modern and ancient sectors, and the medical sector, respectively. The parameter $\lambda > 0$ represents the elasticity of kilometers of geographic distance to technology transmission barrier, which is common to all types of technology. Hence, at time $t = \tau$, all countries have the potential to begin growing, but those countries that are further away from Great Britain, geographically and culturally, will initially experience effectively zero growth. The functional form of growth rates is such that, when costs imposed by distance are high, they will be effectively zero, but as the cost of distance falls, one country after another will cross the threshold where it starts to benefit from frontier technology and starts to grow.

Countries which begin growing later have the advantage of backwardness and experience catch-up growth, shown in the term $\left(\frac{X_{t-1,f}}{X_{t-1,i}}\right)^\theta$ term for $X = A, B, M$. The parameter $\theta > 0$ represents the elasticity of catch-up growth to backwardness. If θ is greater, so is the advantage of backwardness.

7 Quantitative Analysis

To estimate the model we proceed in three steps. First, we normalize a few parameters or borrow them from other studies. Next, we select preference and technology parameters to match demographic and economic transition in Britain. In the final step, we choose the parameters that determine cross-country technology diffusion to generate a global demographic and economic transition.

7.1 Parameters set exogenously

Following Desmet and Rappaport (2017), we set $\rho_a = 0.7$ (a land share of 0.3 in ancient sector) and $\rho_m = 0.9$ (a land share of 0.1 in modern sector). The labor endowment for old adults ζ^2, is set to 1, and the labor endowment for the elderly, ζ^3, is set to 0.5. The initial level of agricultural technology, A_0, is normalized to 1, as is the initial level of medical technology, M_0. Finally, the initial mortality levels s_0^j are calculated from the mortality rates for the 1675-1699 period reported by Schofield and Wrigley (1989).[13] These parameters and theire values are summarized in Table 12.

Table 12: Parameters set exogenously

Description	Parameter	Value
Technology		
Returns to scale, ancient	ρ_a	0.7
Returns to scale, 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

7.2 Calibration, first stage

In the first stage of the calibration, we select thirteen parameters to match the demographic and economic transition in Britain as closely as possible. These thirteen parameters, $\{\gamma, \phi, B_0, \tau_1, \tau_2, \bar{c}, \bar{n}, \mu_f^a, \mu_f^b, \mu_f^m, \delta, \alpha, \beta, \xi\}$, determine preferences for and cost of children, human

[13] The mapping between age-specific mortality rates in the data and their model counterparts is detailed in Appendix F.

capital production, the shares of skilled and unskilled labor in each sector, and the growth rates of TFP in the ancient and modern sectors, A_t and B_t, and the medical technology, M_t, in post-Malthusian period. In order to determine these parameters we simulate the demographic and economic transition in the model for a single country and match its experience to Britain's demographic and economic transformation.

A model period is 20 years and economic growth start in $\tau = 1690$. In order to compare model 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 $t - 10$ and $t + 9$. Hence the data targets for 1690 is an average of the data from 1680 to 1699 and so on. Starting from $\tau = 1690$, we simulate the model economy moving forward and choose these thirteen parameters to match the following set of moments:

1. The levels of the CDR and CBR between 1690 and 2010 (Figure 16). For the pre-transition period, we assume that CBR and CDR are constant and equal to their 1690 values.[14]

2. The share of labor employed in ancient sector (Figure 17). We an ancient sector share of 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).

3. The GDP per capita between 1690 and 2010 (Figure 18).

4. The years of education between 1870 and 2010 (Figure 19). The data on educational attainment is taken from the Lee and Lee (2016) dataset. The average total years of enrollment in Great Britain was only 1.0 in 1870, which grow very rapidly after that and reach 6.6 by 1950 and 11.4 by 2010.

The model does a good job matching these targets. The parameters that are calibrated in the first stage are given in Table 13.

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

Table 13: Calibrated parameters, first stage

Description	Parameter	Value
Utility Function		
Utility weight for fertility	γ	0.715
Parental altruism, warm glow	ϕ	1.353
Minimum consumption as fraction of wage	$\tilde{c}$	0.141
Minimum fertility	$\overline{n}$	0.300
Cost of Children		
Quantity	τ_1	0.113
Quality	τ_2	0.022
Technology		
Unskilled labor share, ancient	α	0.690
Unskilled labor share, modern	β	0.302
Growth rate of A_t	μ_f^a	0.891% (yearly)
Growth rate of B_t	μ_f^b	0.893% (yearly)
Growth rate of M_t	μ_t^m	0.614% (yearly)
Medical technology lag	δ	3.29
Dynamic complementarity of human capital	υ	0.394
Elasticity of education effort to human capital	ξ	0.644

Figure 16: UK CBR/CDR, sim vs. data

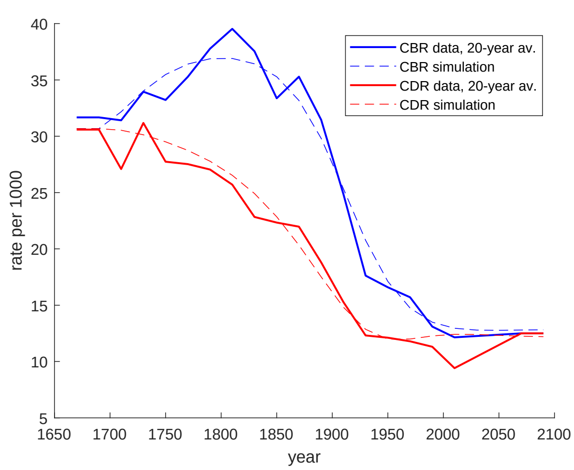


Figure 17: UK ag. sector share, sim

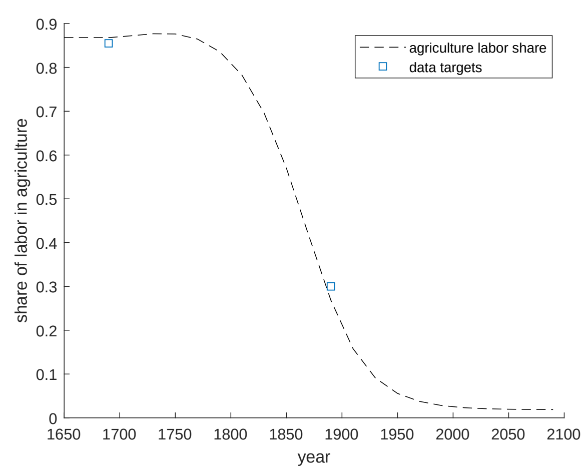


Figure 18: UK GDP per capita, sim vs. data

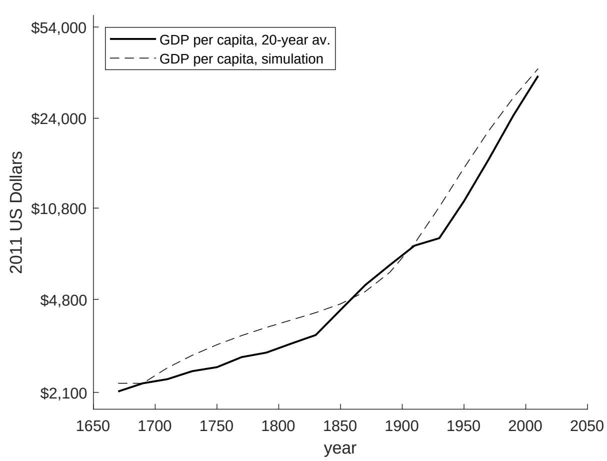
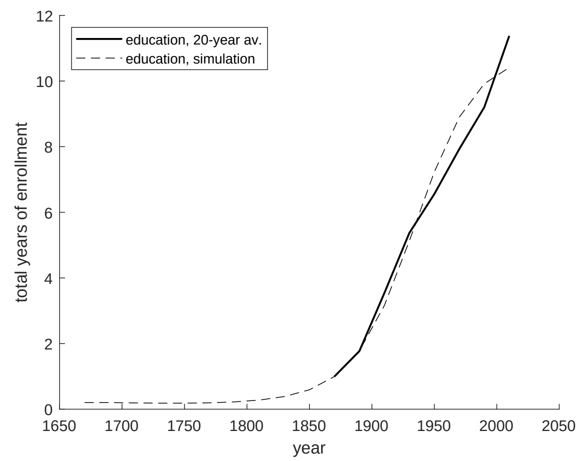


Figure 19: UK years of education, sim vs. data



7.3 Calibration, second stage

Taking the parameters calibrated in the first stage as given, in the second stage we choose eleven parameters that govern the process of technology diffusion: d_j for $j = a, b, m$ that govern the cost of distance as a barrier to agricultural, non-agricultural, and medical technology; θthat represents the elasticity of catch-up growth to backwardness; and φ which determines how fast distances shrink.

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. Figure 24 shows how well the model does for a set of countries (the same set of countries that appear in Figure 4). While the procedure does not target any specific country, the model does a very nice job capturing demographic transition in different countries. It is also not surprising that it fails for some, such as Chad. We then choose these eleven parameters to match:

1. The global average CBR and CDR between 1950 and 2017 (Figure 20).
2. The global average GDP per capita between 1950 and 2017 (Figure 21).
3. The standard deviation of log GDP per capita for the whole world between 1950-2017 (Figure 22). Note that the level of inequality is much smaller in the model economy since we assume that countries are identical in 1690. It would not be very difficult to add initial differences in the level of technology to account for cross country income differences in 1690. The main message of Figure 22 is, however, the increase in inequality in the model and the data.

We construct these targets by weighting each country by its population according to World Bank data. The calibrated parameter values are given in Table 14.

Table 14: Calibrated parameters, second stage

Description	Parameter	Value
Distance		
elasticity of km to effective distance	λ	1.72
cost of distance for ag. tech. diffusion	d_b	0.0019
cost of distance for non-ag. tech. diffusion	d_a	0.033
cost of distance for medical tech. diffusion	d_m	0.0010
growth rate of cost of physical distance	φ	-8.1% (yearly)
elasticity to backwardness	θ	1.04

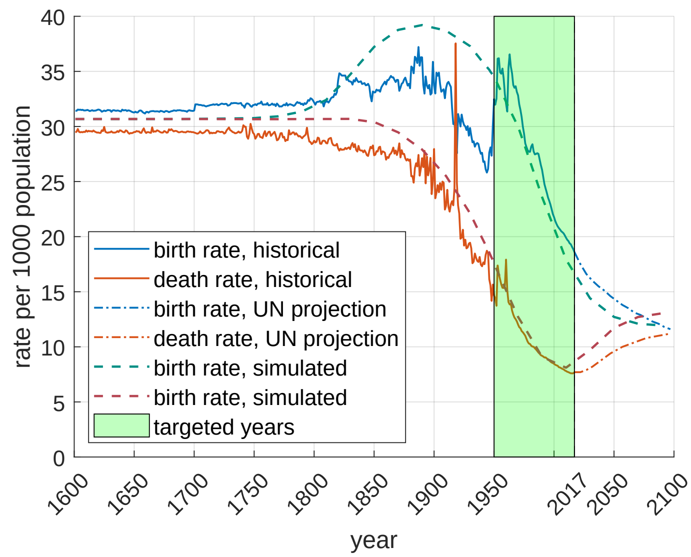


Figure 20: World CBR and CDR, 1600-2100

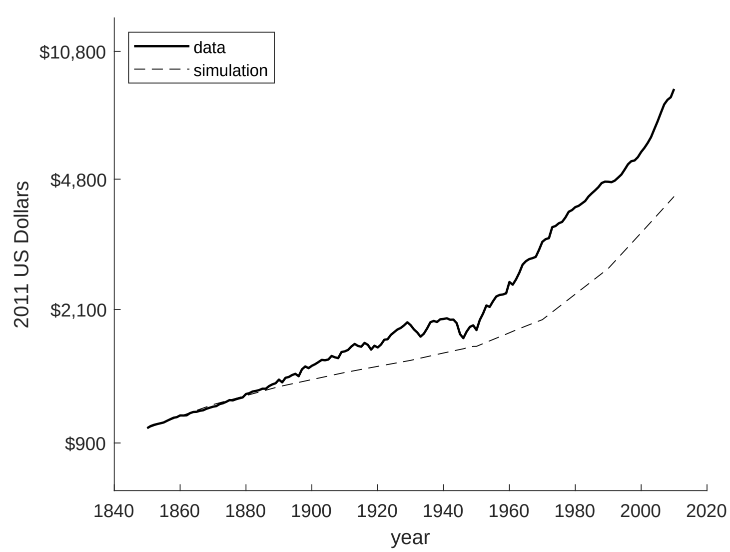


Figure 21: Global mean GDP per capita

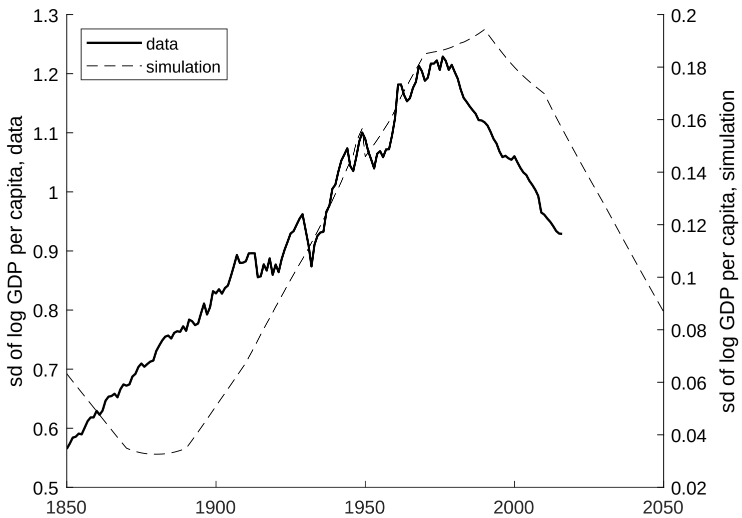


Figure 22: Global standard deviation of log GDP per capita

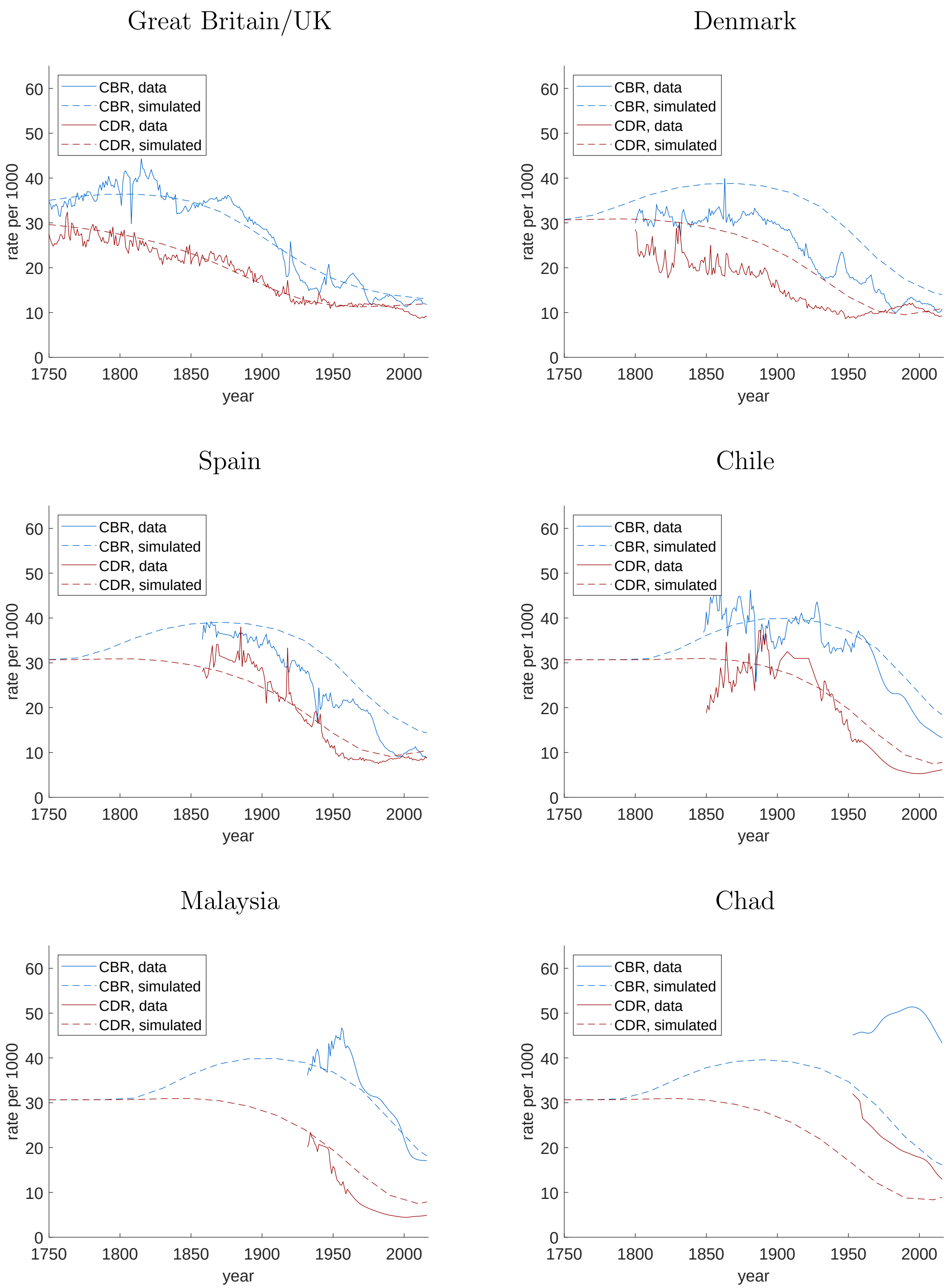
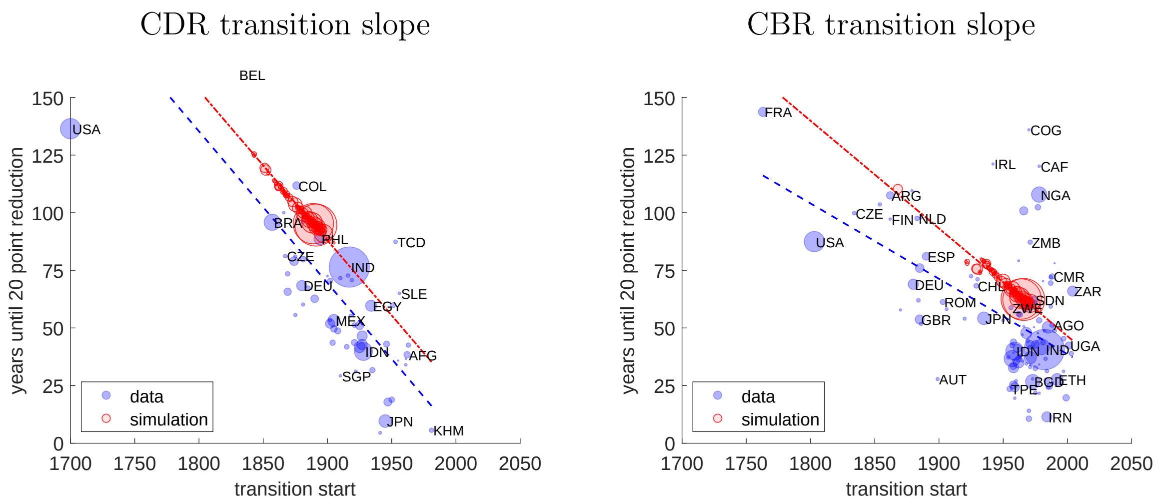


Figure 23: Six examples of demographic transitions, compared with simulation

8 Demographic Transitions, past and present

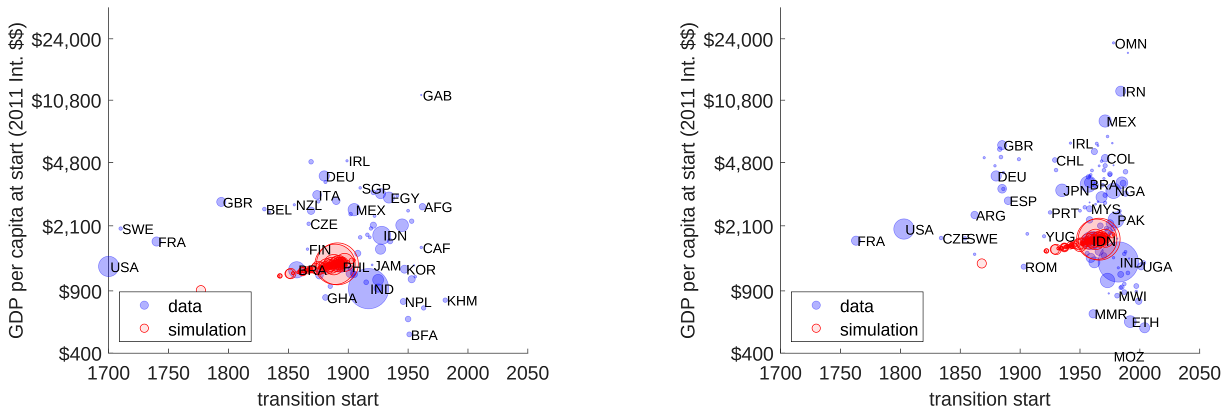
How do demographic transitions look in the model? In particular, does the model economy is able to generate two facts that emerged from our analysis of cross country data? First, transitions in both fertility and mortality have been getting faster over time. Second, in spite of this increase in the speed of the transitions, there is no clear trend in the level of GDP per capita at which countries enter the fertility transition. Figures 24 and 25 answers these questions. As Figure 24 shows, the transitions are indeed are getting faster in the model. As the cost of distance falls in the model, each country experiences a growth take-off, with the closer countries taking of first. The catch-up growth is, however, faster in countries that join the growth process later. Since the rise skill premium and the associated rise in education levels is sharper in later-transitioning countries, so the fall in fertility is also more rapid, and the overall transition period shorter. Figure 25 shows the level of GDP per capita at the start of the simulated CDR and CBR transitions exhibits a slightly increasing trend, but is still nearly constant. This is also consistent with what is observed in the data.



NOTE: The size of the circle for each country is proportional to its share of the 2016 world population.

Figure 24: Transition slopes.

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



NOTE: The size of the circle for each country is proportional to its share of the 2016 world population.

Figure 25: Log GDPpc at the start of transitions.

9 Conclusions

In this paper we have constructed a dataset consisting of birth rates and death rates, and GDP per capita for a panel of 186 countries and spanning from 1735 until 2014. We have proposed a way of measuring demographic transitions which lets the data pick likely start and end dates for fertility and mortality transitions, and used our results to show that: 1) transitions are becoming faster, 2) the average level of GDP per capita at the start of a transition is more or less constant, and 3) an important predictor of a country's transition is the prior transition of other countries which are "close" to it in a geographical and a linguistic sense, and which have similar legal systems.

We then build a model in the tradition of Barro, Becker and Lucas that can account for these facts. In addition to the standard quantity-quality trade-off between how many children to have and how much to educate them, there is also technological diffusion between countries. We conduct a quantitative exercise to show that a simple mechanism of diffusion where skill-biased technological change travels from Britain to the rest of the world in a manner that depend on geographic and cultural distance is able to generate sequences of demographic transitions, each happening faster than the previous one, as we observe in the data, and the account for roughly half of the observed reduction in total transition time. The model we build also predicts a positive relationship between the speed of the fertility transition and the speed of increase in years of education due to the quantity-quality trade-off. We confirm the existence of this pattern in the data, and find that our quantitative exercise produces a quantitatively similar trend.

References

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. Mailhammer, 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. (1960): "An Economic Analysis of Fertility," *Demographic and Economic Changes in Developed Countries*, 11, 209–231.

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., K. Murphy, and R. Tamura (1990): "Human Capital, Fertility, and Economic Fertility," *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," *Maddison Project Working Paper*, (10).

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

Comin, D., and B. Hobijn (2010): "An Exploration of Technology Diffusion," *American Economic Review*, 100(5), 2031–59.

Comin, D., and M. Mestieri (2018): "If Technology Has Arrived Everywhere, Why Has Income Diverged?," *American Economic Journal: Macroeconomics*, 3(10), 137–178.

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

De la Croix, D., and F. Perrin (2017): "French Fertility and Education Transition: Rational Choice vs. Cultural Diffusion," Working Paper.

De Silva, T., and S. Tenreyro (2017): "The Fall in Global Fertility: A Quantitative Model," Working Paper.

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.

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

Egger, P. H., and A. Lassmann (2012): "The Language Effect in International Trade: A Meta-analysis," *Economics Letters*, 116(2), 221–224.

Fernandez-Villaverde, J. (2001): "Was Malthus Right? Economic Growth and Population Dynamics," *Working Paper*.

Galor, O., and D. N. Weil (1996): "The Gender Gap, Fertility, and Growth," *American Economic Review*, 86(3), 374–387.

——— (1999): "From the Malthusian Regime to Modern Growth," *American Economic Review*, 89, 150–154.

——— (2000): "Population, Technology, and Growth: From Malthusian Stagnation to the Demographic Transition and Beyond," *American Economic Review*, 90(4), 806–828.

Greenwood, J., N. Guner, and G. Vandenbroucke (2017): "Family Economics Writ Large," *Journal of Economic Literature*, 55(4), 1346–1434.

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

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 (2011): "Fertility Theories: Can They Explain the Negative Fertility-Income Relationship?," *Demography and the Economy*, pp. 43–100.

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.

LUCAS, R. E. (1988): "On the Mechanics of Economic Development," *Journal of Monetary Economics*, 22(1), 3 – 42.

——— (2002): *Lectures on Economic Growth*. Harvard University Press.

——— (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.

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," *Cambridge University Press*, 93(2), 351–363.

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

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*. National Central Bureau of Statistics, Stockholm, 2 edn.

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.

SHORTER, F., AND M. MACURA (1982): *Trends in Fertility and Mortality in Turkey, 1935-1975*. National Academy Press.

Spolaore, E., and R. Wacziarg (2014): "Fertility and Modernity," *Working Paper*.

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.

United Nations (2017): *World Population Prospects: The 2017 Revision*. Department of Economic and Social Affairs, Population Division, https://population.un.org/wpp/. Accessed 16 March 2019.

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

A 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	1962	2008	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
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	1830*	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

Calculated Transition Start and End Dates				
	CDR		CBR	
Country	Start	End	Start	End
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
Croatia	n/a	n/a	n/a	2002
Cuba	1902⋆	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	1914⋆	1975	1969	1976
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

Calculated Transition Start and End Dates				
	CDR		CBR	
Country	Start	End	Start	End
Ghana	1881⋆	1996	1967	n/a
Greece	1916	1955	1930	1994
Grenada	1883⋆	1973	1957	2004
Guam	1946⋆	1950	1963	n/a
Guatemala	1917	1997	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
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

Calculated Transition Start and End Dates				
	CDR		CBR	
Country	Start	End	Start	End
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
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	1735⋆	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

Calculated Transition Start and End Dates				
	CDR		CBR	
Country	Start	End	Start	End
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
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

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

B Auxiliary Rules for Model Selection

B.1 Auxiliary Rules of Transition Starts

A statistically-detected Crude Death Rate transition start date is removed, moving from Case I to Case II, or Case III to Case IV, if one or more of the following conditions holds:

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 Crude Death Rate transition start date is added, moving from Case II to Case I, or Case IV to Case III, if both of the following conditions holds:

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

A statistically-detected Crude Birth Rate transition start date is removed, moving from Case I to Case II, or Case III to Case IV, if one or more of the following conditions holds:

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 Crude Birth Rate transition start date is added, moving from Case II to Case I, or Case IV to Case III, if both of the following conditions holds:

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 Crude Death Rate transition end date is removed, moving from Case I to Case III, or Case II to Case IV, if one or more of the following conditions holds:

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

A Crude Death Rate transition end date is added, moving from Case III to Case I, or Case IV to Case II, if both of the following conditions holds:

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 Crude Birth Rate transition start date is removed, moving from Case I to Case III, or Case II to Case IV, if one or more of the following conditions holds:

1. Estimated initial CBR level of greater than 20, less than 20 years after the start of the series.

2. Estimated initial CBR level of greater than 25, regardless of timing.

A Crude Birth Rate transition start date is added, moving from Case III to Case I, or Case IV to Case II, if both of the following conditions holds:

1. Estimated final CBR less than 12.

2. CBR end date has not been previously removed by the first set of rules.

C Extension of GDP per capita data

Recall that the main source for GDP per capita data that we use 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 Logit analysis in section 5, we make a small number of conservative 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: for these, our work has been done for us. There are also 11 countries in Category 2. For these countries, we impute for the year 1500 the value of GDP per capita for 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 this 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. We believe that there can be little harm in assuming that this set of countries was in the pre-modern regime of (a lack of) economic growth, and that their GDP per capita was the same in 1500 as it was in the first year we observe it. Therefore, 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 total. The remaining 138 countries in our dataset belong to Category 4. Some of these countries have estimates of GDP per capita estimates dating back to the 18th or 19th centuries, but these estimates are too high for us to safely presume that they predate 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 feel comfortable projecting their initial first GDP per capita observation all the way back from, say, 1950 or 1975 to the year 1500.

D Historical Estimates of World Vital Statistics

We construct world average crude birth rates and crude death rates for the year 1600 through the year 2016 using two sources of information:

- Data on birth rates and death rates by country from the various sources detailed in Section 3.1
- Data on population by country from the Maddison 2018 database (Bolt, Inklaar, de Jong, and van Zanden, 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 project CBR backwards from the start of the data to 1600 by setting it equal to its pre-transition mean.
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 and an observation, an interpolated value, or a backward-projected value for the CBR in that year.

Following the exact same process for crude death rates as we did for crude birth rates would lead to an implied rate of pre-modern world population growth that is much higher than all available historical estimates. The reason for this is that following the exact same process would project rates of natural increase from the start of the data for each country back into history, when all available evidence indicates that rates of natural increase were in fact much lower. To maintain consistency with the available data on pre-modern population growth, we follow a slightly modified process for crude death rates, described below in five steps.

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 backwards 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.
3. Then, for the 93 countries for which we do not observe the start of the mortality transition but which for which we are able to impute a transition start date using the method described in section 3.2, we project CDR backwards from the start of the data until the imputed start of the CDR transition by assuming it is equal to transition mean.

4. Then, for each of these 93 countries, we project CDR backwards 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 crude birth rate 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.

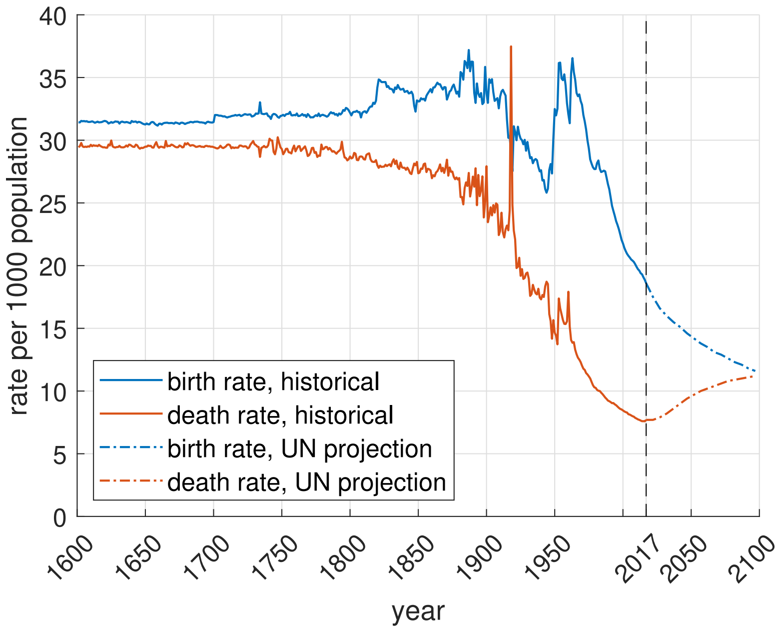


Figure 26: World CBR and CDR, 1600-2100

Figure 26 shows the results of these calculations, and combines them with post-2016 projections from the United Nations (United Nations, 2017). Looking at the period from 1850 to 1900, the lower volatility of these rates in comparison to the series for individual countries (see Figure 4) can partly be interpreted as the result of local shocks in different parts of the world canceling each other out. During the 20th century the world was becoming more and more connected, and shocks more correlated. We can clearly see the effects of the global influenza pandemic of 1918, and the global baby boom of the 1950s and 60s. Prior to 1850, it is less clear how we should interpret the relative "smoothness" of the rates shown on the graph. The farther we go back in time, the fewer countries have "live" data available–so some part of this smoothness must be due to an increasing share of back-projected pre-transition means being included in the average.

The world average rate of population growth is calculated as the difference between births and deaths–as there is no space travel yet, on a world level, the rate of natural increase equals the population growth rate. The total number of annual births is calculated by multiplying

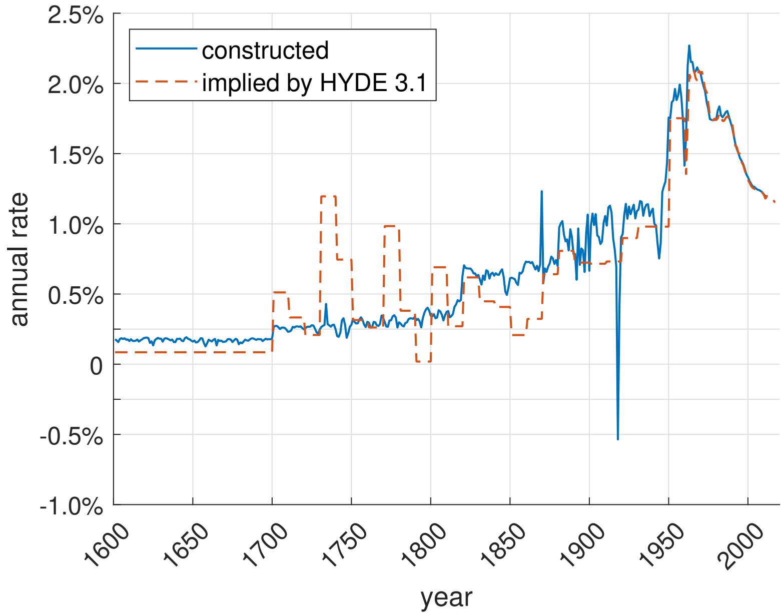


Figure 27: World population growth, comparison

the world average crude birth rate by the total world population, taken from the HYDE 3.1 database (Klein Goldewijk, Beusen, van Drecht, and de Vos, 2011). Figure 27 compares the constructed annual population growth rates to those implied by the world population data in the HYDE 3.1 database.

E An empirical analysis of CDR transitions

Table A1: GDPpc and CDR transition, Logit results

Variable	Estimates
Cons	9.28
	(57.42)
lnGDPPC	-6.34
	(15.24)
lnGDPPC ^{2}	0.57
	(1.01)
LLn	-128.0
Pseudo-R^2	0.079
N	7680

Table A1 reports the Logit estimation for CDR when the only explanatory variable is log GDP per capita. As shown in Figure 28, this specification replicates well the distribution of log GDP per capita at the start of the CDR transition. This specification does not perform well, however, in replicating the distribution of CDR transition starts over time or in predicting transition start dates for individual countries, as seen in Figures 29 and 30.

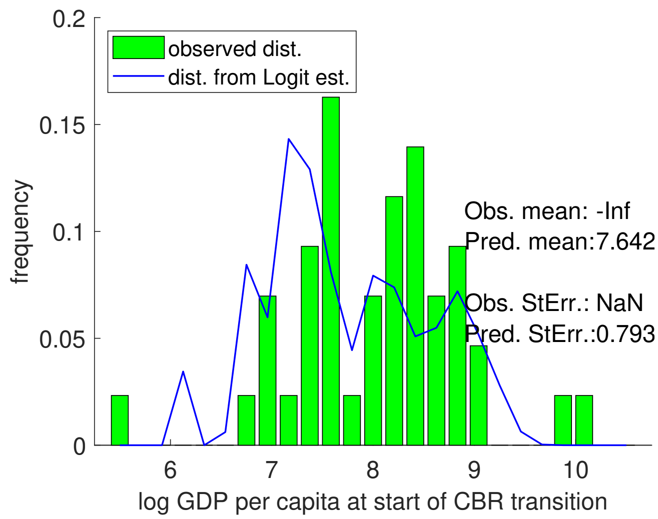


Figure 28: Distribution of log GDPpc at the start of the CDR transitions

Figure 29: Within Sample Predictions

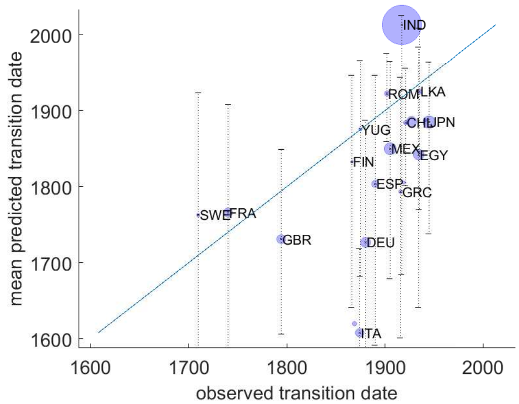
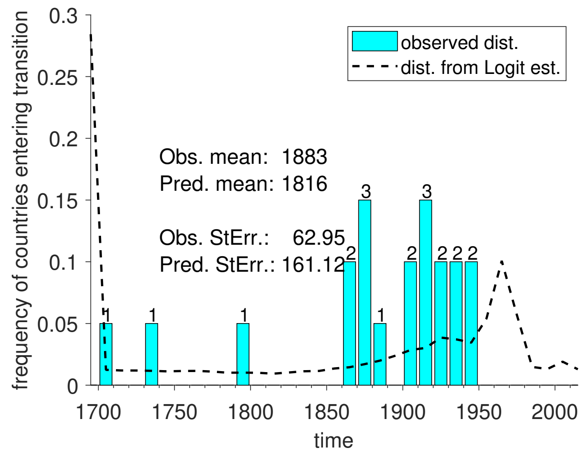


Figure 30: Distribution of Transtion Dates



E.1 Demographic Contagion for CDR

Table A2 shows the results of the Logit regression described in Section 5 for CDR. Specification (1) shows the results of the regression without including any inter-country influence. Specification (2) adds a global count of the number of countries that have begun the transition, and specification (3) adds some curvature to that sum, which is still global. The estimated value of ψ, being less than 1, implies that there are diminishing returns–the more countries have already entered the transition, the smaller the effect of each additional country on other countries' odds of entering the transition. Specifications (4) through (11) weight the influence of one transitioned country on other countries according to the inverse distance between them, as determined by various measures of distance. When included by themselves, all 4 measures of distance (geographic, linguistic, religious and legal) have highly significant estimated coefficients, with geographic distance having somewhat more explanatory power (as reflected in the log likelihood sum) than the others. Religious distance has the wrong sign, which means that it is probably correlated with some excluded factor and, thus, the coefficient does not reflect the real effect of religious distance. Specification (9) includes more than one measure of distance simultaneously. In this specification both geographic and legal distance have significant coefficients.

In Figure 13, we look at the access to transitions measure implied by specification (9) (the distributions displayed in all of these figures are smoothed using a Gaussian kernel). Using the estimated parameters, access is calculated as

$$\mathcal{A}_{it} \equiv \left[\sum_{j=1}^{N} \exp[\mathcal{D}_{ij} + 1.37 \cdot \text{cml}_{ij}] \mathcal{I}_{j,t-1} \right]^{0.45},$$

Table A2: Determinants of the start of the CDR transition

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
cons	9.28	-16.00	1.27	13.13	-36.07	1.81	1.36	-8.60	-30.40
	(57.42)	(56.19)	(54.76)	(57.08)	(62.46)	(54.52)	(54.72)	(56.57)	(57.90)
lnGDPPC	-0.63	0.15	-0.34	-6.42	6.81	-0.32	-3.44	0.02	6.19
	(1.52)	(1.49)	(1.45)	(15.14)	(16.47)	(1.44)	(14.52)	(15.20)	(15.29)
lnGDPPC ^{2}	0.01	-0.00	0.00	0.47	-0.41	0.02	0.29	-0.00	-0.44
	(0.01)	(0.01)	(0.01)	(1.00)	(1.08)	(0.10)	(0.96)	(1.02)	(1.01)
access		0.13	1.22	9.68	3.09	2.13	1.29	2.41	3.96
		(0.02)	(0.85)	(2.28)	(0.55)	(0.55)	(0.51)	(0.45)	(0.70)
geo dist.				5.08					
				(0.79)					
$<$ 800km					1.92				0.94
					(0.23)				(0.47)
800-2000km					0.88				0.92
					(0.23)				(0.34)
ling. dist						1.81			
						(0.28)			
relig dist							0.18		
							(0.63)		
legal dist								2.08	1.37
								(0.85)	(0.63)
ψ, curv.			0.42	0.60	0.70	0.50	0.43	0.55	0.86
			(0.20)	(0.29)	(0.18)	(0.14)	(0.03)	(0.13)	(0.20)
LLn	-128.0	-106.7	-104.1	-98.6	-94.3	-103.0	-104.0	-100.6	-92.5
Pseudo-R^2	0.079	0.232	0.251	0.291	0.321	0.259	0.251	0.277	0.334
N. Obs.	7680	7680	7680	7680	7680	7680	7680	7680	7680

Note: Standard errors of the estimated coefficients are given in parentheses.

where

$$\mathcal{D}_{ij} \equiv 0.94 \cdot \mathbf{1}\{\text{ldi}_{ij} < \ln 800\} + 0.92 \cdot \mathbf{1}\{\ln 800 \leq \text{ldi}_{ij} < \ln 2000\}.$$

The top left panel of Figure 31 shows the distribution of this measure at different points in time. Not surprisingly, as more countries transition, this distribution moves steadily to the right. The top right panel of Figure 31 plots the transition probabilities implied if each country is assigned its actual access to CDR transitions value and GDP per capita equal to $2000. Here we can see that in 1850 and 1900 "Access to CDR transitions" in the great majority of countries was such that their probability of transition at $2000 GDP per capita would have been relatively small. In 1950 and the year 2000, the distributions shift outward somewhat. In each of these two years, there are still some countries that would have zero probability of transition at $2000 GDP per capita, and the majority of countries have less than 20% yearly probability of transition at this income level. would be 10%.

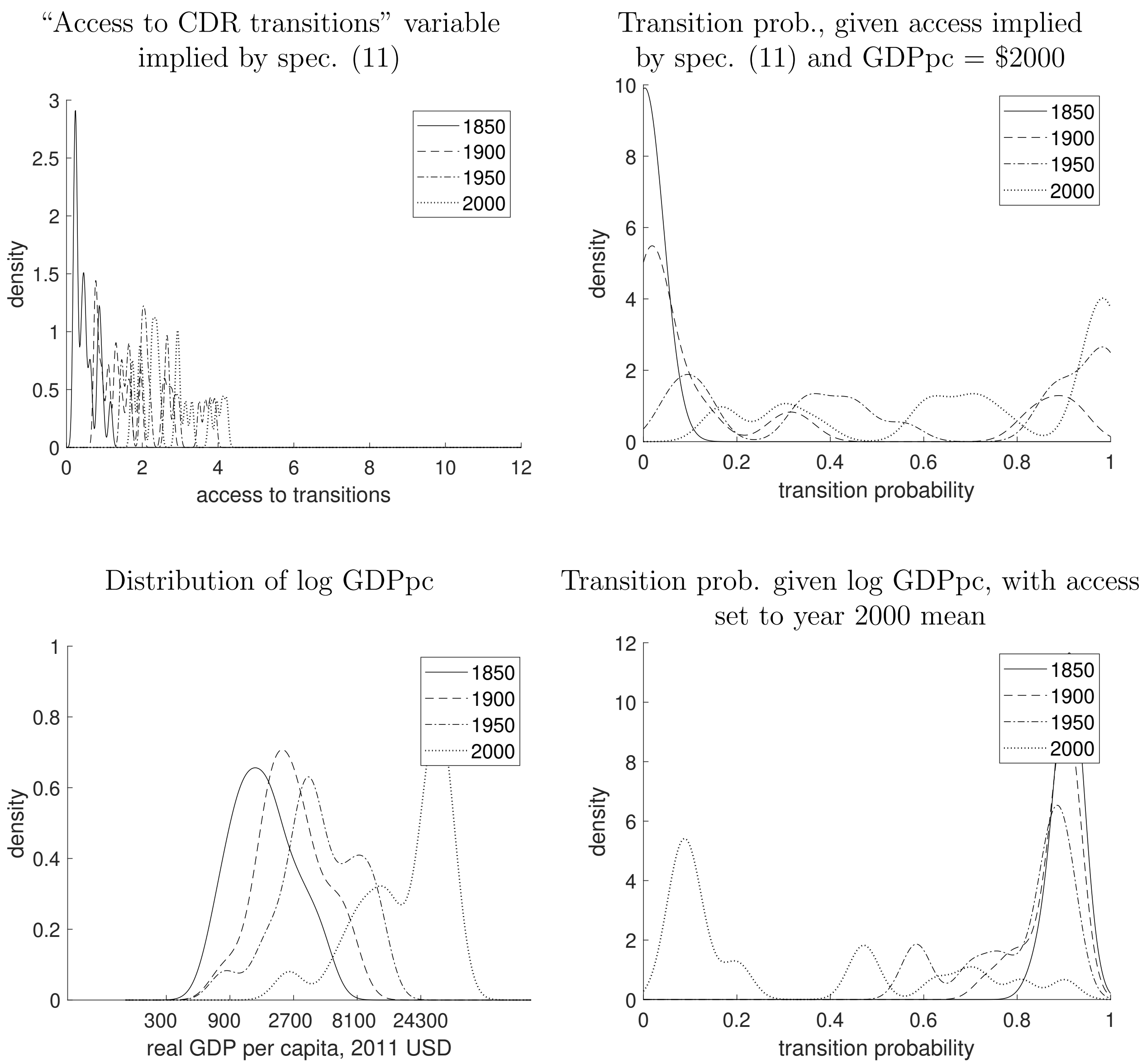


Figure 31: Demographic contagion.

The bottom left panel of Figure 31 shows the evolution of the distribution of GDP per capita over time. This distribution shifts right as time passes and more countries enjoy higher levels of GDP per capita. The bottom right panel of Figure 31 shows the distribution of the probability of CDR transition, given the observed GDP per capita for each country, assuming they have the mean level of "Access to CDR transitions" existing in the year 2000. This panel demonstrates the importance of the complementarity between a country's level of development and the influence of its neighbors. In 1850, even countries with relatively high log GDP per capita had a low transition probability. In comparison, by 2000, a country with the relatively low level of GDP per capita ($2000) has a greater than 40% probability of starting the CDR transition if enough of their neighbors started before them.

F Accounting for Intra-Period Mortality

A proper definition of s_0, s_1, s_2, and s_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. To that end, let us define $\tilde{s}_x$ as the average fraction of those alive at the beginning of age x 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 $x+1$; where x takes values in $\{1, 2, 3, 4\}$. Then,

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

Futhermore, 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

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

F.1 Proof of integral for continuous deaths over discrete periods

Let s_t represent the fraction of a cohort alive at the start of a given period t that is alive at the end of the period. Suppose that instead of all dying in a single moment at the end of the period, that 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 given by

$$\tilde{s}_t = \frac{1 + s_t^{\frac{1}{x}} + s_t^{\frac{2}{x}} + \cdots + s_t^{\frac{x-1}{x}}}{x} = \frac{1}{x}\sum_{z=0}^{x-1} s_t^{\frac{z}{x}}$$

The sum in the above expression has a convenient closed form:

$$\begin{aligned}
J &= \sum_{z=0}^{x-1} s_t^{\frac{z}{x}} \\
J - 1 + s_t &= s_t^{\frac{1}{x}} + s_t^{\frac{2}{x}} + \cdots + s_t^{\frac{x-1}{x}} + s_t^{\frac{x}{x}} \\
J - 1 + s_t &= s_t^{\frac{1}{x}} J \\
J(1 - s_t^{\frac{1}{x}}) &= 1 - s_t \\
J &= \frac{1 - s_t}{1 - s_t^{\frac{1}{x}}}
\end{aligned}$$

Applying this closed form, we can write

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

If we wish to know $\tilde{s}_t$ when mortality is a continuous process, we must take the limit of $\tilde{s}_t$ as the number of sub-periods x approaches infinity.

$$\tilde{s}_t = \lim_{x\to\infty} \frac{1 - s_t}{x(1 - s_t^{\frac{1}{x}})}$$

$$\begin{aligned}
&= (1 - s_t)\frac{1}{\lim_{x\to\infty} \frac{1 - s_t^{\frac{1}{x}}}{x^{-1}}} \\
&= (1 - s_t)\frac{1}{\lim_{x\to\infty} \frac{\frac{s_t^{\frac{1}{x}} \log s_t}{x^2}}{-x^{-2}}} \\
&= -(1 - s_t)\frac{1}{\lim_{x\to\infty} s_t^{\frac{1}{x}} \log s_t} \\
&= \frac{s_t - 1}{\log s_t}
\end{aligned}$$

F.2 UK Mortality Data

Table A3 shows the English mortality data that initial, pre-modern survival probabilities are based upon. All these data are derived from Wrigley, Davies, Oeppen, and Schofield (1997), Chapter 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.

Table A3: Mortality in the 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

Initial survival probabilities are then calculated using the method described in section F, where $S_t(x)$ is a stepwise function consistent with the data in Table A3.

G Equilibrium allocations

We assume that there is a fixed amount of land used in each sector and it is normalized to $T_{t,a} = T_{t,m} = 1$. Then the representative firm in each sector solves the standard profit-maximization problem. This, combined with the perfect mobility of skilled and unskilled labor, implies that in equilibrium

$$w_t^u = \alpha A_t L_{t,a}^{\alpha-1} H_{t,a}^{\rho_a-\alpha} = \beta B_t L_{t,m}^{\beta-1} H_{t,m}^{\rho_b-\beta}, \tag{15}$$

and

$$w_t^s = (\rho_b - \alpha) A_t L_{t,a}^{\alpha} H_{t,a}^{\rho_a-\alpha-1} = (\rho_b - \beta) B_t L_{t,m}^{\beta} H_{t,m}^{\rho_b-\beta-1} \tag{16}$$

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

$$L_{t,a} + L_t = N_t^1 + \zeta^2 N_t^2 + \zeta^3 N_t^3 \equiv \bar{L}_t, \tag{17}$$

and

$$H_{t,a} + H_t = h_t N_t^1 + \zeta^2 h_{t-1} N_t^2 + \zeta^3 h_{t-2} N_t^3 \equiv \bar{H}_t. \tag{18}$$

Combine (15) with (17) and (18) to obtain

$$\alpha A_t \left(\bar{L}_t - L_{t,m}\right)^{\alpha-1} \left(\bar{H}_t - H_{t,m}\right)^{\rho_a-\alpha} = \beta B_t L_{t,m}^{\beta-1} H_{t,m}^{\rho_m-\beta} \tag{19}$$

Similarly, combine (16) with (17) and (18) to get

$$(\rho_a - \alpha) A_t \left(\bar{L}_t - L_{t,m}\right)^{\alpha} \left(\bar{H}_t - H_{t,m}\right)^{\rho_a-\alpha-1} = (\rho_m - \beta) B_t L_{t,m}^{gb} H_{t,m}^{\rho_m-\beta-1}. \tag{20}$$

The last two equations imply

$$\frac{H_{t,m}}{\bar{H}_t} = \frac{1}{\frac{\beta}{\alpha}\frac{\rho_a-\alpha}{\rho_m-\beta}\frac{\bar{L}_t}{L_t} + 1 - \frac{\beta}{\alpha}\frac{\rho_a-\alpha}{\rho_m-\beta}}. \tag{21}$$

Equation (19) can be developed into

$$\frac{\alpha}{\beta}\frac{A_t}{B_t}\frac{\bar{L}_t^{\alpha-\beta}}{\bar{H}_t^{\alpha-\beta+\rho_m-\rho_a}} = \frac{\left(1 - \frac{L_{t,m}}{\bar{L}_t}\right)^{1-\alpha}}{\left(\frac{L_{t,m}}{\bar{L}_t}\right)^{1-\beta}} \frac{\left(\frac{H_{t,m}}{\bar{H}_t}\right)^{\rho_m-\beta}}{\left(1 - \frac{H_{t,m}}{\bar{H}_t}\right)^{\rho_a-\alpha}} \tag{22}$$

Equations (21) and (22) are two equations in two unknowns, $\frac{H_{t,m}}{\bar{H}_t}$ and $\frac{L_{t,m}}{\bar{L}_t}$.

Combining (21) and (22) and rearranging, we can derive $z\left(\frac{L_{t,m}}{\bar{L}_t}\right)$, defined as

$$
\begin{aligned}
z\left(\frac{L_{t,m}}{\bar{L}_t}\right) \equiv\; & 1-\frac{L_{t,m}}{\bar{L}_t} \\
& -\left(\frac{L_{t,m}}{\bar{L}_t}\right)^{\frac{1-\rho_m}{1-\rho_a}}\left(\left(1-\frac{\beta}{\alpha}\frac{\rho_a-\alpha}{\rho_m-\beta}\right)\frac{L_{t,m}}{\bar{L}_t}+\frac{\beta}{\alpha}\frac{\rho_a-\alpha}{\rho_m-\beta}\right)^{\frac{-\rho_a+\rho_m-\beta+\alpha}{1-\rho_a}} \times \\
& \times\left(\frac{\alpha}{\beta}\frac{A_t}{B_t}\frac{\bar{L}_t^{\alpha-\beta}}{\bar{H}_t^{\alpha-\beta+\rho_m-\rho_a}}\right)^{\frac{1}{1-\rho_a}}\left(\frac{\beta}{\alpha}\frac{\rho_a-\alpha}{\rho_m-\beta}\right)^{\frac{\rho_a-\alpha}{1-\rho_a}}
\end{aligned}
\tag{23}
$$

The equilibrium value of $\frac{L_{t,m}}{\bar{L}_t}$ can be characterized as the unique point at which $z\left(\frac{L_{t,m}}{\bar{L}_t}\right)=0$. The first derivative is given below for the purpose of implementing Newton's Method:

$$
\begin{aligned}
z'\left(\frac{L_{t,m}}{\bar{L}_t}\right) =\; & -1 \\
& -\left(\frac{L_{t,m}}{\bar{L}_t}\right)^{\frac{1-\rho_m}{1-\rho_a}}\left(\left(1-\frac{\beta}{\alpha}\frac{\rho_a-\alpha}{\rho_m-\beta}\right)\frac{L_{t,m}}{\bar{L}_t}+\frac{\beta}{\alpha}\frac{\rho_a-\alpha}{\rho_m-\beta}\right)^{\frac{-\rho_a+\rho_m-\beta+\alpha}{1-\rho_a}} \times \\
& \times\left(\frac{\alpha}{\beta}\frac{A_t}{B_t}\frac{\bar{L}_t^{\alpha-\beta}}{\bar{H}_t^{\alpha-\beta+\rho_m-\rho_a}}\right)^{\frac{1}{1-\rho_a}}\left(\frac{\beta}{\alpha}\frac{\rho_a-\alpha}{\rho_m-\beta}\right)^{\frac{\rho_a-\alpha}{1-\rho_a}} \times \\
& \times\left[\frac{1-\rho_m}{1-\rho_a}\left(\frac{L_{t,m}}{\bar{L}_t}\right)^{-1}+\right. \\
& \left.\left(1-\frac{\beta}{\alpha}\frac{\rho_a-\alpha}{\rho_m-\beta}\right)\frac{-\rho_a+\rho_m-\beta+\alpha}{1-\rho_a}\left(\left(1-\frac{\beta}{\alpha}\frac{\rho_a-\alpha}{\rho_m-\beta}\right)\frac{L_{t,m}}{\bar{L}_t}+\frac{\beta}{\alpha}\frac{\rho_a-\alpha}{\rho_m-\beta}\right)^{-1}\right]
\end{aligned}
\tag{24}
$$

G.1 Dynamic equilibrium

Define the vector of time-t state variables

$$
x_t \equiv \begin{bmatrix} A_t & B_t & M_t & N_t^1 & N_t^2 & N_t^3 & h_t^1 & h_t^2 & h_t^3 \end{bmatrix}'.
$$

Furthermore let the law of motion $x_{t+1} = m_t(x_t)$ be defined by the following set of equations:

$$\begin{aligned}
N^1_{t+1} &= s^0_t s^1_t n_t N^1_t \\
N^2_{t+1} &= s^2_t N^1_t \\
N^3_{t+1} &= s^3_t N^2_t \\
h^1_{t+1} &= e_t \\
h^2_{t+1} &= h^1_t \\
h^3_{t+1} &= h^2_t
\end{aligned} \tag{25}$$

and by given series of A_t, B_t, and M_t for $t \in \{0, 1, 2, ..., T\}$. Note that s^0_t, s^1_t, s^2_t and s^3_t are all determined by M_t according to (14).

Labor allocations $L_{t,m}$ and $H_{t,m}$ are implicit functions of x_t characterized by (21), (23), (17), (18). Given labor allocations, wages w^u_t and w^s_t are also implicit functions of x_t characterized by (15) and (16).

Define $\tilde{w}_{t+1} \equiv \frac{w^s_{t+1}}{w^u_{t+1}}$. According to the solution characterized by (**??**) and (**??**), n_t and e_t are both implicit functions of x_t and $\tilde{w}_{t+1}$. From (25) we see that x_{t+1} is determined by x_t and the choices n_t and e_t, so we can reformulate $L_{t,m}$ and $H_{t,m}$ as functions of x_t, n_t and e_t. Finally, employing (15) and (16) once more, define

$$g_t(\tilde{w}) \equiv \frac{\rho_b - \beta}{\beta} \frac{L_{t+1,m}[x_t, n_t(x_t, \tilde{w}), e_t(x_t, \tilde{w})]}{H_{t+1,m}[x_t, n_t(x_t, \tilde{w}), e_t(x_t, \tilde{w})]} - \tilde{w}. \tag{26}$$

Given a set of initial cohort sizes and education levels N^1_0, N^2_0, N^3_0, h^1_0, h^2_0, and h^3_0, and a series of technology levels $\{A_t, B_t, M_t\}^T_{t=0}$, a dynamic equilibrium consists of a series $\{\tilde{w}_{t+1}\}^T_{t=0}$ such that $g_t(\tilde{w}_{t+1}) = 0$ and $x_{t+1} = m_t(x_t)$ for all t.

It can be shown that the function $g_t(\tilde{w})$ is monotonically decreasing and is 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 (26) 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 have explored in this study, however, this condition has never occurred, and a solution has always existed.

G.2 Solution for steady state

In steady state, fertility is constant at replacement level. In other words, $n_t = \tilde{n} = \frac{1}{s_0 s_1}$. The size of each cohort is constant over time. The size of the old adult and elderly cohorts is related to the size of the young adult cohort, N_1, as follows:

$$N_2 = s^2 N_1$$
$$N_3 = s^3 N_1$$

Education e, human capital h, technologies A and B, and wages w^u and w^s are likewise constant over time. The stock of unskilled and skilled labor are constant and given by the following two expressions:

$$\bar{L} = (1 + s^2\zeta^2 + s^2 s^3 \zeta^3) N^1$$

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

Skilled and unskilled wages are given by the following expressions:

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

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 in closed form as a function of model primitives. The following four equations characterize the solution:*

1.

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

2.

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

3.

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

4.

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

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

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

Proof: In the steady state, the stock of unskilled and skilled labor are given by the following two expressions:

$$\bar{L} = (1 + s^2\zeta^2 + s^2 s^3\zeta^3)N^1$$

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

Furthermore, skilled and unskilled labor are given by the following expressions:

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

and

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

It then follows that the steady-state skilled and unskilled wage are related in the following way:

$$\frac{w^s}{w^u} = \frac{\tilde{z}}{h}. \tag{27}$$

where $\tilde{z}$ is the following function of the share of unskilled labor in the modern sector and factor shares:

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

The choice of education in the steady state is characterized by the following expression.

$$-e^{1+\frac{\xi}{1-v}}w^s\frac{1+\phi s^1}{\phi s^1}-ew^u\frac{2+\phi s^1}{\phi s^1}+e^{\frac{\xi}{1-v}}\frac{w^s(1-\tau_1\tilde{n})}{\tilde{n}s^0\tau_2}-e^{1-\frac{\xi}{1-v}}\frac{1}{\phi s^1}\frac{w^u}{w^s}w^u+\frac{w^u(1-\tau_1\tilde{n})}{\tilde{n}s^0\tau_2}-\frac{\bar{c}(1+s^0\tilde{n})}{\tilde{n}s^0\tau_2}\leq 0$$

As a first step, substitute $w^s=\frac{\tilde{z}}{h}w^u$, $\bar{c}=\tilde{c}w^u$, and $h=e^{\frac{\xi}{1-v}}$, then divide everything by w^u.

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

Finally, rearrange to obtain

$$e\geq\frac{\phi s_1}{\tilde{n}s^0\tau_2}\frac{(1+\tilde{z})(1-\tau_1\tilde{n})-\tilde{c}(1+s^0\tilde{n})}{\tilde{z}(1+\phi s^1)+2+\phi s_1+\frac{1}{\tilde{z}}} \tag{28}$$

In the steady state, fertility must be at replacement level: $n=\tilde{n}\equiv\frac{1}{s^0s^1}$. Applying this condition to (**??**), and applying (27), we obtain

$$\begin{aligned}\tilde{n}&=\frac{\gamma}{1+\gamma}\left[\frac{1-\frac{\tilde{c}w^u}{w^u+\tilde{z}w^u}}{\tau_1+\tau_2s^0e+\frac{s^0\tilde{c}w^u}{w^u+\tilde{z}w^u}}+\frac{1}{\gamma}\frac{\overline{n}}{s^0}\right]\\&=\frac{\gamma}{1+\gamma}\left[\frac{1-\frac{\tilde{c}}{1+\tilde{z}}}{\tau_1+\tau_2s^0e+\frac{s^0\tilde{c}}{1+\tilde{z}}}+\frac{1}{\gamma}\frac{\overline{n}}{s^0}\right]\end{aligned}$$

$$\tilde{n}-\frac{1}{1+\gamma}\frac{\bar{n}}{s^0}=\frac{\gamma}{1+\gamma}\frac{1-\frac{\tilde{c}}{1+\tilde{z}}}{\tau_1+\tau_2s^0e+\frac{s^0\tilde{c}}{1+\tilde{z}}}$$

Next, we substitute in for e using (28), and solve for $\tilde{z}$. Manipulation yields the following expression:

$$\tilde{z}^2\left[1-\frac{\phi s^1(1-\bar{n}s^1)}{\gamma}-\tau_1\tilde{g}\right]+\tilde{z}\left[1-\tilde{g}(2\tau_1+s^0\tilde{c})+(1-\tilde{c})\left(1-\frac{\phi s^1(1-\bar{n}s^1)}{\gamma}\right)\right]+1-\tilde{c}-\tilde{g}(\tau_1+s^0\tilde{c})=0 \tag{29}$$

Where $\tilde{g}$ is the following function of model primitives:

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

If the solution for $\tilde{z}$ exists, then, it can must be consistent with the quadratic equation:

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

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

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

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

If $\frac{L_m}{\bar{L}}$ implied by the 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. In case of existence, steady-state education can be recovered thus,

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

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

Now that we know the share of unskilled labor in the modern sector and the level of human capital, we can solve for the steady-state level of population. From (23):

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

Substituting in for $\bar{L}$ an $\bar{H}$ and manipulating yields:

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