

EARLY LITERACY ACHIEVEMENTS, POPULATION DENSITY, AND THE TRANSITION TO MODERN GROWTH

Raouf Boucekkine

Université catholique de Louvain and
CORE

Dominique Peeters

Université catholique de Louvain and
CORE

David de la Croix

Université catholique de Louvain and
CORE

Abstract

The transition from economic stagnation to sustained growth is often modeled thanks to “population-induced” productivity improvements, which are assumed rather than derived from primary assumptions. In this paper the effect of population on productivity is derived from optimal behavior. More precisely, both the number and location of education facilities are chosen optimally by municipalities. Individuals determine their education investment depending on the distance to the nearest school, and also on technical progress and longevity. In this setting, higher population density enables the set-up costs of additional schools to be covered, opening the possibility to reach higher educational levels. Using counterfactual experiments we find that one-third of the rise in literacy can be directly attributed to the effect of density, and one-sixth is linked to higher longevity. Moreover, the effect of population density in the model is consistent with the available evidence for England, where it is shown that schools were established at a high rate over the period 1540–1620. (JEL: O41, I21, R12, J11)

1. Introduction

The transition from stagnation to growth has been the subject of intensive research in the growth literature in recent years. It is now recognized that the

Acknowledgments: Boucekkine and de la Croix acknowledge financial support from the Belgian French-speaking community (Grant ARC 03/08–235 “New Macroeconomic Approaches to the Development Problem”) and the Belgian Federal Government (Grant PAI P5/21, “Equilibrium Theory and Optimization for Public Policy and Industry Regulation”). We thank A. Ciccone, H. d’Albis, M. Doepke, J. Finlay, F. Heylen, J.-O. Hairault, E. Lehman, O. Leukhina, F. Maniquet, M. Perez-Nievas, A. Pommeret, G. Rayp, A. Schaefer, J. Thisse, E. Toulemonde, D. van de Gaer, seminar participants at Lausanne, Ghent, Louvain-la-Neuve (CORE), Minneapolis FED, Paris I and Namur Universities, participants at SED 2004 and T2M 2005, three anonymous referees, and the editor for their comments on an earlier draft. We thank David Cressy for providing useful information on English schools prior to the Industrial Revolution.

E-mail addresses: Boucekkine: boucekkine@ires.ucl.ac.be; de la Croix: delacroix@ires.ucl.ac.be, Peeters: peeters@geog.ucl.ac.be

understanding of the contemporary growth process needs to be based on proper micro-foundations that would reflect the central driving forces of the growth process. Moreover, a comprehensive understanding of the factors preventing less developed economies from growing in a sustained way would be futile unless the factors that prompted the take-off of the currently developed economies could be identified.

The unified theory of economic growth surveyed by Galor (2005) suggests that the transition from stagnation to growth is an inevitable byproduct of the process of development. In particular, Galor and Weil (2000) demonstrate that the involved Malthusian interaction between technology and population did speed up technological progress through a population density induced mechanism. This ultimately raised an industrial demand for human capital and further technological progress, leading to a demographic transition to modern growth.

This paper provides micro-foundations to this population density effect. In particular, we build up a theoretical framework featuring the effect of population density on human capital formation, and we evaluate quantitatively its importance in the industrialization process of England. Where Galor and Weil 2000 relies on a reduced form to model the positive effect of population density on technological progress, we point out that such an effect typically shows up through the increasing profitability of school foundations when population density rises.¹ The exploited mechanism is in line with a wider class of models where there are positive externalities generated by denser population. Becker, Glaeser, and Murphy (1999) argue that larger populations encouraged greater specialization and increased investments in knowledge. For Kremer (1993), high population spurs technological change.²

The proposed theory generates a fundamental testable prediction: the increase in population density prior to the Industrial Revolution induces a rise in human capital formation well before the Industrial Revolution that is critical for the emergence of the Industrial Revolution. The view that human capital formation was critical for the take-off was put forward by some economic historians, as Cipolla (1969), who argue that literacy favored the Industrial Revolution in more than one way. In particular, human capital accumulation avoided shortages of literate workers in those fields in which such workers were specifically required, and on a more general ground, it made people more adaptable to new circumstances and receptive to change. In times of innovations, educated workers have

1. See Fujita and Thisse (2002) for a textbook treatment of this effect and Holmes (2005) for an application on contemporary data.

2. Notice also that, the view according to which population density matters for growth is in accordance with the empirical literature which finds that density appears as a significant factor in growth regressions across countries (see, e.g., Kelley and Schmidt 1995), across US states (Ciccone and Hall 1996), and in large European states (Ciccone 2002). Employment density seems also important for hourly earnings of workers across districts of Great Britain (Anastassova 2006).

a comparative advantage because they assimilate new ideas more readily (Bartel and Lichtenberg 1987).

In contrast, most of the existing theories emphasizing the role of human capital formation in the transition from stagnation to modern growth (e.g., Galor and Weil 2000; Doepke 2004; Lagerloef 2003, 2006) are rather concerned with the increased industrial demand for human capital in the second phase of the industrial revolution, following the acceleration in the rate of technological progress. The only other theory that emphasizes the role of human capital formation prior to industrialization in the take-off is Galor and Moav (2002). They argue that a process of natural selection prior to the Industrial Revolution increased the representation of individuals with higher valuation for child quality, gradually increasing the average level of investment in human capital, stimulating technological progress and ultimately the take-off from stagnation to growth.

The paper is organized as follows. Data on England over the period 1540–1860 are introduced in Section 2 with a focus on education. The model economy is described in Section 3. In Section 4 we study the problem of the individual who has to choose the length of her studies. Section 5 is devoted to the problem of the school authority (who decides the number and the localization of schools). We also consider an alternative mechanism of school establishments based on free-entry. A quantitative exercise is proposed in Section 6 in order to evaluate the ability of the model to explain the rise of schooling over the period 1530–1860. In addition, we run several counter-factual experiments to disentangle the role of each factor in the development of education. Section 7 contains robustness analyzes with respect to important assumptions we made. Section 8 concludes.

2. The Rise in English Education: 1540–1860

In this section we present some important data on the early rise in education in England and on its possible factors. The left panel of Figure 1 shows literacy rates (average of men and women) for England as estimated by Cressy (1980). It suggests that improvements in literacy started as early as in the sixteenth century.³ This picture is consistent with the overall survey of Houston (2002) for the early modern period. According to Houston (2002), the percentage of children who were able to attend school at some stage during their youth lies between one-fifth to one-third at the end of the early modern period. Areas where half or more of the school-age children received instruction were educationally advanced. England belonged to these favored zones. O'Day (1982) stresses that a key determinant

3. The data from Cressy (1980) are globally consistent with the ones displayed by Clark and Hamilton (2004) from other sources. There is, however, an unresolved debate on literacy developments during the first phase of the industrial revolution. Sanderson (1995) claims that literacy rate was stagnant over the period 1760–1830.

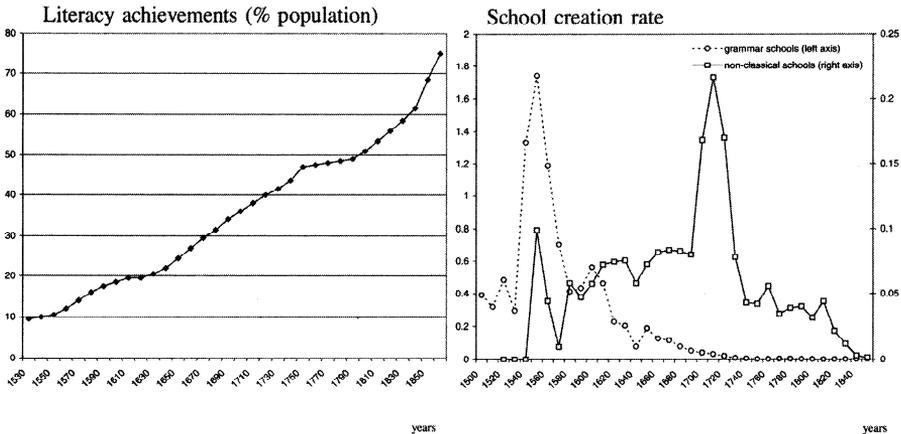


FIGURE 1. Education in England. Estimation (left): Cressy (1980). Source (right): Author estimation from Appendix A.

for this relative success was accessibility to schools: only a small proportion of the rural population was geographically distant from access to any kind of formal education provision.

We have also gathered from the reports of the British Parliamentary Papers evidence on the number of schools. The reports of the School Inquiry Commission written in 1867–1868 documents the creation of schools.⁴ Two lists of schools are provided, together with their date of establishment. The “endowed grammar schools” taught a mixture of Latin and practical skills to sons of the middling sort and lesser elite (list in Schools Inquiry Commission 1868a). The “endowed non-classical schools” were products of the Charity School Movement, offering protestant socialization and basic skills to the worthy poor (list in Schools Inquiry Commission 1868b). According to Cressy (1980), although short-lived private schools are omitted from the list, a check against other sources proves the commission’s work to be reliable. We use these lists to compute the number of school establishments per decade. The data are presented in Appendix A. The right panel of Figure 1 shows that the creation of grammar schools accelerated markedly over the period 1540–1620. The expansion of grammar schools was followed by the one of non-classical schools in the eighteenth century.

Higher education achievements might have been triggered by several economic factors. Beyond the non-economic factors, serving national, religious, and social goals, which may have promoted education before and during the first phase of the industrial revolution, we document the possible role of three economic factors (see Galor 2005 for an exhaustive list). First, technical progress

4. A special thank to Françoise Canart for identifying the relevant British Parliamentary Papers.

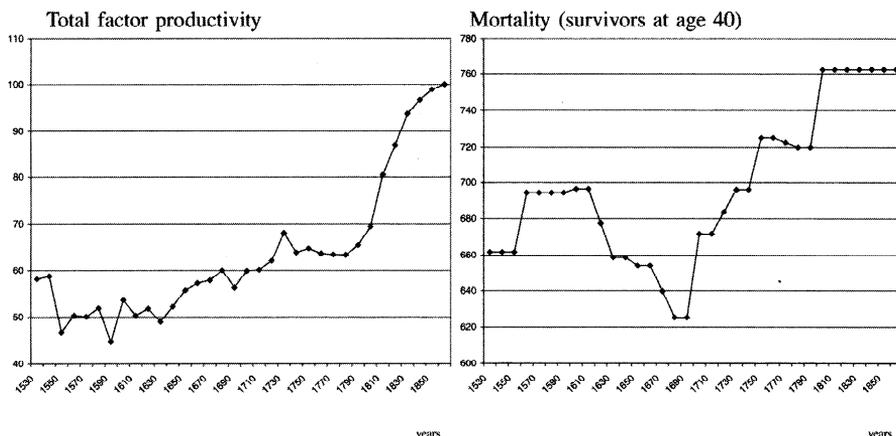


FIGURE 2. Pre-industrial England. Estimation (left): Clark (2001). Source (right): Wrigley et al. (1997).

increased labor productivity and wage rates in the modern sector, and consequently the return to investment in education. Facing better income perspectives in this sector, households would engage in education to benefit from the higher skill premium. However, as illustrated by the left panel of Figure 2, productivity gains started to accelerate in the beginning of the nineteenth century. The timing implied by this explanation is thus partly counterfactual, and cannot account for the fact that higher literacy rates were achieved two centuries before any significant gain in productivity.

Another interesting candidate for explaining the rise in literacy is longevity improvements. In Western Europe, adult mortality started to drop before the Industrial Revolution took place for reasons that are not yet fully understood (see, e.g., the discussion in Fridlitzius 1985). Boucekkine, de la Croix, and Licandro (2003) and Nicolini (2004) argue that lower mortality induced higher investment in human capital and/or in physical capital, therefore paving the way to future growth.⁵

This could only be part of the story however, as shown by English data. The right panel of Figure 2 presents the survival rate of five-year-old individuals.⁶ We see that adult longevity was first stagnant then declining over the period 1600–1700, probably because of the urban penalty associated with the fast growth

5. The effect of life expectancy on human capital formation is positive provided that the prolongation of life would affect the return to quality more than the return to quantity. Hazan and Zoabi (2005) show that if parents derive utility from the aggregate wage income of their children, prolongation of life would increase the return to quantity and quality symmetrically. Clearly, however, life expectancy affects directly the incentive of the individual himself to acquire education.

6. It therefore abstracts from infant mortality swings to concentrate on mortality during the active life.

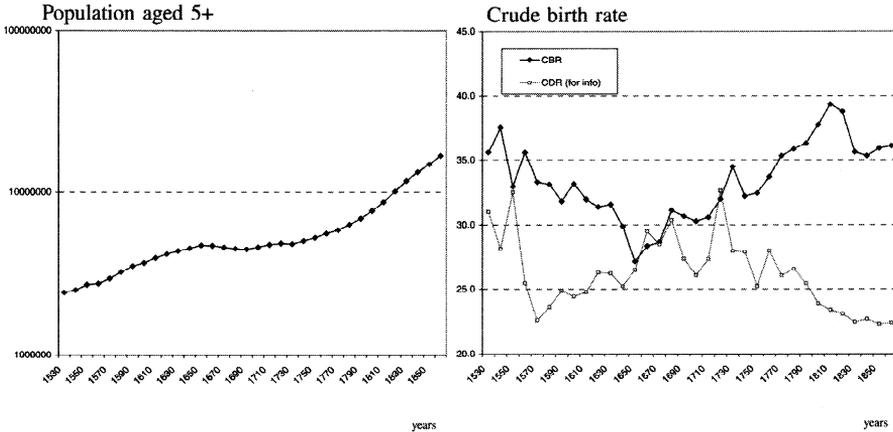


FIGURE 3. Population of England. Estimation: Wrigley et al. (1997).

of cities. During this period of high mortality, literacy rose continuously, as we show on Figure 1. Hence, for England, the mortality channel is of no help to explain the early rise in educational attainments.

A third factor introduced previously, on which our main argument will rely, is population density. Looking at English historical data, the left panel of Figure 3 shows that population⁷ rose rapidly in the sixteenth and nineteenth centuries, whereas the seventeenth century was one of demographic stagnation.⁸ The corresponding swings in crude birth rates (CBR) are plotted in the right panel of Figure 3 together with the crude death rate (CDR). We notice that rises in population in the sixteenth century correspond to the first wave of improvement in literacy.

3. The Model Economy

The model mixes aspects from growth theory with elements from the economic geography literature. The length of schooling is chosen by individuals who maximize lifetime income, which depends on future wages, longevity, and the distance to the nearest school. Then, the number and location of education facilities is determined, either chosen optimally by the state or following a free-entry process.

7. We represent the population aged 6 and more, because it coincides with the concept of population of our model, disregarding infants aged 0–5.

8. Population growth in England was spectacular compared to other big European countries. According to Maddison (2001) numbers, English population increased by 2.17 times between 1500 and 1700. This growth factor was about 1.43 for France, 1.25 for Germany, 1.27 for Italy, and 1.29 for Spain.

Higher population density makes it optimal to increase school density, opening the possibility to reach higher educational levels.

In this model we have three heterogeneity sources (age, skill, and location), and the basic structure is overlapping generations in continuous time with a “realistic” survival law (borrowed from Boucekkine, de la Croix, and Licandro, 2002). Despite of this complexity, which is required by the calibration and counterfactual analysis carried later, we obtain a fine analytical characterization of the solutions in the benchmark case.

3.1. Time and Space

Time is continuous. At each point in time t a new generation of size ζ_t is born. Individuals born at the same date have different innate abilities, μ , and location, i . Abilities are distributed according to a probability density function $g(\mu)$. We assume that g is log-normal, which is commonly used to approach the actual distribution of abilities. This distribution is the same at every location. Within generations individuals are thus indexed by (μ, i) .

We assume that a given household stays at their location i forever. In the pre-industrial era, the main reason for households to move was to reach regions with better employment opportunities or higher wages.⁹ Space is modeled as a circle of unit diameter. We suppose that each new generation is uniformly spread over the circle, hence ζ_t can also be interpreted as the density of the newborn population at any point, because it depends on the normalization of the circle. If we let the circle shrink, this would have the same effect as increasing ζ_t .¹⁰ We use the label “birth density” for the process ζ_t because the main driving force behind consists in fluctuations of crude birth rate (net of infant mortality). Given the dispersion of the population, the schools will be optimally located if they are evenly spaced, let us say at locations j/E , where $j = 0, \dots, (E - 1)$, E being the number of schools. $x(i)$ measures the distance between the individual located at i and the closest school. If no school is created, distance is infinite: $x(i) \in [0, 1/2] \cup \{+\infty\}$. For the sake of clarity, all individual variables will be expressed as functions of absolute location i , while they are actually functions of the distance $x(i)$.

9. Because, in our set-up, production can take place anywhere with the same set of technologies, the prime reason for migration is not present in the model. Introducing migration in a meaningful way would imply specifying employment basins and using core-periphery models à la Fujita and Thisse (2002).

10. As an alternative to the assumption of a uniform distribution of the population over the circle, we could have taken a negative exponential density function, reflecting the idea that there is a center, London, and the countryside. However, this would make little sense, because the model cannot explain why some people would have an interest to be concentrated and some other would stay in the countryside. Notice also that rural population was still more than 80% of total population by the end of the period we consider.

Notice finally that the number of schools E should be interpreted as a density, because it depends on the normalization of the circle. If we multiply the diameter of the circle by a factor C , there will be $C E$ schools.

3.2. Demographics

Each individual has an uncertain lifetime. The unconditional probability for an individual belonging to the cohort t of reaching age a , is given by the survival function introduced by Boucekkine, de la Croix, and Licandro (2002):

$$m_t(a) = \frac{e^{\beta_t a} - \alpha_t}{1 - \alpha_t}, \quad (1)$$

ASSUMPTION 1. The functions α_t and β_t satisfy: $\alpha_t > 1$ and $\beta_t > 0$.

Assumption 1 guarantees that the survival function is concave, that is, the probability of death increases with age, and that there is a maximum age L_t that an individual can reach. This parameter configuration allows the function $m_t(a)$ to represent with better accuracy the empirical survival laws compared to models with constant probability of death. The maximum age is obtained by solving $m_t(L_t) = 0$ and is equal to

$$L_t = \frac{\log(\alpha_t)}{\beta_t}. \quad (2)$$

The size of the generation born in t at any time $z \in [t, t + L_t]$ is given by $\zeta_t m_t(z - t)$, reflecting that the measure of each generation declines deterministically through time. The size of total population at time t is given by

$$P_t = \int_{t-\bar{L}_t}^t \zeta_z m_z(t - z) dz, \quad (3)$$

where \bar{L}_t is the age of the oldest cohort still alive at time t .

3.3. Technology

There is a unique material good, the price of which is normalized to 1, that can be used for consumption. This good can be produced through two different technologies. In the “modern sector,” which can represent both modern agriculture and industry, the technology employs human capital H_t with constant returns:

$$Y_t = A_t H_t \quad (4)$$

where

$$A_t = e^{\gamma t}.$$

The process $\gamma_t > 0$ models the rate of exogenous technological progress.

In equilibrium the wage per unit of human capital is equal to the marginal productivity A_t . In what follows, we already incorporate this result into the household decision problem.

The physical good can also be produced in the “traditional sector,” which can be seen as a home production sector. Individuals who work in this sector have a productivity w^h per unit of time, and this productivity is independent of their level of human capital. The equilibrium wage per unit of time in the traditional sector is simply w^h and is constant through time. There is no technical progress in this sector and, as a consequence, the modern sector will become more attractive over time provided that $\gamma_t > 0$. This captures the mechanism calibrated by Hansen and Prescott (2002) and Bar and Leukhina (2005), where technical progress in the modern sector is much faster than that in the traditional sector.

3.4. *Individuals*

We consider the case of an individual (μ, i) born at time t . We assume that this individual has the same survival function (and thus the same parameters α_t and β_t) as the other individuals of the cohort born at t . We also assume that the individual has no initial endowment neither in the final good nor in human capital. For simplicity, we assume perfect annuity markets, all the contracts by which they buy goods and sell labor are contingent on survival. Individuals are fully rational and perfectly anticipate future aggregate variables.

We will model the cost of schooling with two elements: transportation costs and tuition fees. Transportation costs are a function of distance. From the Schools Inquiry Commission (1868a) we learn that boys can attend a city school from distances up to 20 miles and a daily consumption of time amounting to more than one hour in the morning and in the evening. In the country, many pupils were farmers’ sons who came in from the neighborhood, some on foot, some on ponies or donkeys, for which stabling is provided at some cost. The report lists many examples, among which the one of an 11-year-old girl who walked five miles and back every day.

In the period considered, schools were funded through income from an endowment and through fees paid by the students’ parents. The Schools Inquiry Commission (1868a) noted that fees were imposed in order to supplement the endowment, and that parents were willing to pay fees, provided the fees were not excessive, and the education was suitable. The report also provides estimates of people’s willingness to pay across different social classes.

The problem of the individual (μ, i) is to maximize lifetime resources W :

$$W[S_t(\mu, i)] = \int_{t+S_t(\mu, i)}^{t+L_t} \omega_t(\mu, i, z) m_t(z-t) e^{-\theta(z-t)} dz - \int_t^{t+S_t(\mu, i)} \xi x(i) e^{\gamma z} m_t(z-t) e^{-\theta(z-t)} dz - k_t e^{\gamma t} \delta[S_t(\mu, i)], \quad (5)$$

where θ is the risk-free interest rate, $S_t(\mu, i)$ is the schooling length, $\omega_t(\mu, i, z)$ is the spot wage income at time z , $\xi x(i)$ is a transportation cost, and k_t is a fixed cost representing tuition fees to be paid only if the individual decides to go to school. δ is an indicator function equal to 1 if its argument is positive, and equal to 0 otherwise. In the sequel we will refer to k_t as the tuition fee. To prevent costs from diminishing automatically when productivity increases, both the tuition cost and the transportation cost are indexed on the rate of exogenous productivity growth.

The spot wage $\omega_t(\mu, i, z)$ is given by

$$\omega_t(\mu, i, z) = A_z h_t(\mu, i),$$

where $h_t(\mu, i)$ is the human capital of individual (μ, i) in cohort t . The individual's human capital is built according to the following technology:

$$h_t(\mu, i) = \mu S_t(\mu, i). \quad (6)$$

While the individual is still at school, we assume that his/her human capital is nil. Hence, productive human capital at age a can be expressed as $\delta[a - S_t(\mu, i)]h_t(\mu, i)$.

For education to be an optimal outcome, the resources obtained by spending the optimal time at school should be larger than the one obtained if the individuals stay in the "traditional sector," where no human capital is needed. This constraint writes

$$W[S_t(\mu, i)] > \int_t^{t+L_t} w^h m_t(z-t) e^{-\theta(z-t)} dz \equiv W_t^h. \quad (7)$$

3.5. School Location and Policy

At each date a number E_t of classrooms is created to serve the newborn generation. These classrooms only accept children born at time t and are scrapped when the last person of generation t graduates. We need to determine the objective pursued by the schools founders. The Schools Inquiry Commission (1868a) distinguishes three types of schools: endowed schools, private schools, and proprietary schools. Endowed schools usually have some income from funds permanently

appropriated to the school. Even in this category, there is a wide variation in their character and history. Some are part of large charitable foundations, others are run by the Church. Many endowed schools have no exclusive connection. The private schools are the property of the master or mistress who teaches in them. They “owe their origin to the operation of the ordinary commercial principle of supply and demand,” according to the Schools Inquiry Commission (1868a). They provide more individual care and teaching, but the School Enquiry Commission extensively complains about the quality of these schools. Commissioners noted that

A really large and flourishing school is of course a marketable commodity, and sometimes sells well. But it is always a dangerous purchase for a stranger... when the school declines the house is let for a shop or a private residence, and the master betakes himself elsewhere.

And also

Considered commercially, few descriptions of business seem to require less capital than the keeping of a private day school of the second order. A house is taken, a cane and a map of England bought, an advertisement inserted, and the master has nothing more to do but teach. It is not likely that schools established at so slight a cost should have buildings well adapted to purposes of education.

These two quotes stress the commercial nature of private schools. The last of the three classes of schools is composed of the proprietary schools who belong to a body of shareholders. They are alike private property. This type of school is more recent, not more than 40 years old in 1860.

Given the uncertainty around the objectives these schools were actually pursuing, we will consider two different types of institutional arrangements.

Model M1. Although there was not a single school in England over which the state actually exercised full control (with the exception of some military schools), endowed schools were subject to rules (“The State has allowed endowments to be scattered over the whole surface of England”; Schools Inquiry Commission 1868a), and the founders were themselves obedient to a superior authority. In scenario M1, we assume the existence of a central authority that determines each year the optimal number of classrooms to be built. We assume that the objective is to maximize aggregate profits, reflecting that “the purpose of schools was never to save those from paying who could afford to pay.”

The set-up cost for implementing a classroom is equal to f , which is indexed on the rate of exogenous productivity growth. This cost can be seen as being net of the possible endowment. The attendance rate is the proportion of children from the newborn generation in the catchment area who decide to attend school. It can be expressed as a function of the density of schools E_t and of the tuition fee k_t , let $R(E_t, k_t)$. The benefit drawn from the school built in a given area is equal to

$k_t A_t \zeta_t R(E_t, k_t)$. The central authority determines simultaneously at each date t the density of schools E_t and the tuition fee k_t to maximize the system's profit. This maximization problem is formulated as follows:

$$\max_{E_t, k_t} A_t(k_t \zeta_t R(E_t, k_t) - f)E_t \quad (8)$$

Model M2. Instead of assuming the existence of any central authority, we can suppose that the density of schools results from a free entry process: schools are created as long as they earn a positive profit. We consider a free entry process of schools together with an endogenously fixed tuition fee k_t . This can reflect the functioning of private schools discussed previously. The problem can then be stated as

$$\text{Find } (E_t, k_t) \text{ such that } E_t \text{ solves } k_t \zeta_t R(E_t, k_t) - f = 0 \text{ for a given } k_t, \quad (9)$$

and k_t is the maximand of

$$k_t \zeta_t R(E_t, k_t) - f \text{ for a given } E_t. \quad (10)$$

Comparing models M1 and M2 in terms of welfare seems a rather complicated task. Indeed, even in a static setting, the inclusion of the geographical dimension prevents the application of standard welfare theorems (see, e.g., the analysis of the Salop model by Fujita and Thisse [2002, pp. 119–124]).

3.6. Equilibrium

Given exogenous demographic and technological trends α_t , β_t , γ_t , and ζ_t , an equilibrium consists of

- (a) A path of optimal education decision $\{S_t(\mu, i)\}_{t \geq 0}$ maximizing life-time resources (5), subject to (7).
- (b) A path of optimal density of schools $\{E_t\}_{t \geq 0}$ and tuition fee $\{k_t\}_{t \geq 0}$ solution to the maximization problem (8) or (9)–(10).

To solve for the equilibrium, we first consider the problem of the individual given a distance to the closest school. This will determine the optimal reaction of households to the creation of new schools. Once this is known, we can solve the problem of the school authority.

4. Solution to the Individual's Problem

Any individual of cohort t seeks to maximize lifetime resources as given by (5) with respect to schooling time, subject to human capital production technology (6)

and to the inequality constraint (7) representing the home sector outside option. Clearly the optimal schooling decision $S_t(\mu, i)$ can be interior or corner. Among the possible corner solutions, $S_t(\mu, i) = 0$ and $S_t(\mu, i) = L_t$, we can already disregard the latter as it is always welfare-dominated by the first due to the strictly positive costs inherent to schooling. Now, notice that if $S_t(\mu, i) = 0$, then the human capital of individual (μ, i) is nil, which implies that his lifetime labor income is equal to home production. For a given technological pace, if the interior solution for schooling is strictly positive, thus inducing nonzero human capital and wages, and if the resulting wage is high enough to more than compensate the schooling costs, then the interior solution is very likely to welfare-dominate the corner solution $S_t(\mu, i) = 0$. Nonetheless, an explicit proof is required, and we will provide it later. Firstly, let us establish the existence and uniqueness of interior solutions. Notice that all the results of this section are obtained for $\theta = 0$, and are therefore valid for small risk free interest rates, by continuity.

4.1. Existence and Uniqueness of the Interior Solution

We maximize

$$\int_{t+S_t(\mu,i)}^{t+L_t} A_z \mu S_t(\mu, i) m_t(z - t)e^{-\theta(z-t)} dz - \int_t^{t+S_t(\mu,i)} \xi x(i) A_z m_t(z - t)e^{-\theta(z-t)} dz - k_t A_t \delta[S_t(\mu, i)].$$

with respect to $S_t(\mu, i)$. The first-order necessary condition is

$$\mu \int_{t+S_t(\mu,i)}^{t+L_t} A_z m_t(z - t) dz = m_t(S_t(\mu, i)) A_{t+S_t(\mu,i)} (\mu S_t(\mu, i) + \xi x(i)). \tag{11}$$

We study hereafter the existence and optimality of the interior solution in the case of a steady technological progress $\gamma_t = \gamma > 0, \forall t$. Also, for ease of presentation, we shall temporarily drop the index t because we are deriving the optimal decisions of a fixed generation t . Naturally, the time argument will be re-introduced when needed.

At first, notice that under a steady technological progress, the first-order condition with respect to schooling (after a variable change) can be rewritten as

$$\mu \int_{S(\mu,i)}^L e^{\gamma a} m(a) da = (\mu S(\mu, i) + \xi x(i)) e^{\gamma S(\mu,i)} m(S(\mu, i)). \tag{12}$$

Observe that L does check the integral equation because of $m(L) = 0$. However, because it is always dominated by the other corner solution, $S(\mu, i) = 0$,

we abstract from it from now on. The next proposition provides a necessary and sufficient condition for the existence of an interior solution to (12). This condition is

$$\int_0^L e^{\gamma a} m(a) da > \frac{\xi x(i)}{\mu},$$

which imposes a condition on μ .

PROPOSITION 1. *For γ small enough, under Assumption 1, there exists a solution to (12) such that $0 < S(\mu, i) < L$ if and only if $\mu > \underline{\mu}(i)$, with*

$$\underline{\mu}(i) = \frac{\xi x(i)}{\int_0^L e^{\gamma a} m(a) da}. \quad (13)$$

The solution is unique. This solution tends to zero as μ gets closer to $\underline{\mu}(i)$.

See Appendix B.1 for the proof.

Proposition 1 implies that for fixed demographic parameters α and β , there exists an interior strictly positive schooling decision provided: (i) the transport cost parameter ξ and the distance to school $x(i)$ are small enough; and/or (ii) the leaning ability is large enough. In other words, such an interior solution may not exist under huge transport costs and distances to schools or under a poorly efficient education sector. For fixed ξ , μ , and $x(i)$, this solution neither exists if the demographic parameters induce markedly low life expectancy and maximal age figures. The next corollary establishes the optimality of the above identified interior schooling rule.

COROLLARY 1. *Under the assumptions of Proposition 1, the interior solution for schooling is the global maximizer of the lifetime resources as given by (5).*

See Appendix B.2 for the proof.

Next, the threshold $\underline{\mu}(i)$ is finely characterized.

COROLLARY 2. *The threshold $\underline{\mu}(i)$ is an increasing function of ξ , $x(i)$, and β . It is decreasing in α and γ .*

See Appendix B.3 for the proof.

PROPOSITION 2. *Under the conditions of Proposition 1, the interior solution S is a strictly increasing function of γ and α , and a strictly decreasing function of β and $x(i)$. For $x(i) > 0$, it is strictly decreasing in ξ , and strictly increasing in μ . It is independent from ξ and μ when $x(i) = 0$.*

See Appendix B.4 for the proof.

4.2. Optimal Schooling under Fixed Tuition Fees

We now compare the welfare implication of the corner solution $S(\mu, i) = 0$ versus the interior solution. Let us denote the latter \hat{S} . Also, denote by \hat{W} (resp. W^h) the intertemporal utility in the interior case (resp. corner case). We have: $W^h = \int_0^L w^h m(z) dz > 0$. The interior solution is optimal if and only if $\hat{W} > W^h$:

$$\int_{t+\hat{S}}^{t+L} \omega(\mu, i, z) m(z-t) dz > \int_t^{t+\hat{S}} \xi x(i) e^{\gamma z} m(z-t) dz + k e^{\gamma t} + W^h,$$

which can be written using the expression of the spot wage, and after rearranging terms and a variable change as

$$\int_{\hat{S}}^L \mu \hat{S} e^{\gamma z} m(z) dz > \int_0^{\hat{S}} \xi x(i) e^{\gamma z} m(z) dz + k + e^{-\gamma t} W^h.$$

Finally, using the optimality condition (12) to replace the integral appearing in the left-hand side of the inequality, and rearranging terms, one gets the following condition:

$$\hat{S}^2 e^{\gamma \hat{S}} m(\hat{S}) \mu + \xi x(i) \left[\hat{S} m(\hat{S}) e^{\gamma \hat{S}} - \int_0^{\hat{S}} e^{\gamma z} m(z) dz \right] > k + e^{-\gamma t} W^h.$$

Denote by $\Psi(\mu, i)$ the function in μ given by the left-hand side of the inequality:

$$\Psi(\mu, i) = \hat{S}^2 e^{\gamma \hat{S}} m(\hat{S}) \mu + \xi x(i) \left[\hat{S} m(\hat{S}) e^{\gamma \hat{S}} - \int_0^{\hat{S}} e^{\gamma z} m(z) dz \right].$$

The following optimality proposition holds.

PROPOSITION 3. *Under Assumption 1, $\Psi(\mu, i)$ is strictly increasing in μ , for $\mu > \underline{\mu}(i)$. Moreover, for any tuition fee $k > 0$, there exist a threshold $\tilde{\mu}(i, k) > \underline{\mu}(i)$, given by*

$$\tilde{\mu}(i, k) = \Psi^{-1}(k + e^{-\gamma t} W^h), \tag{14}$$

checking:

- (i) *If $\mu > \tilde{\mu}(i, k)$, then the interior schooling solution is optimal: $S^* = \hat{S}$.*
- (ii) *If $\mu < \tilde{\mu}(i, k)$, then the interior schooling solution is dominated: $S^* = 0$.*
- (iii) *If $\mu = \tilde{\mu}(i, k)$, then the individual is indifferent between the interior solution and the corner solution.*

See Appendix B.5 for the proof.

The following comparative statics can then be established.

PROPOSITION 4. *For any t , the threshold $\tilde{\mu}$ is a strictly increasing function of k , w^h , ξ , $x(i)$ and β , and a strictly decreasing function of α and γ .*

See Appendix B.6 for the proof.

The main implication of Proposition 4 is that the higher life expectancy (via a rise in α or a drop in β), the likelier the decision to go to school. In contrast, a tougher geographical situation in terms of distance to schools and transport costs makes this decision unlikely.

5. Solution to the School Location Problem

We consider now the determination of the density of schools. Our framework is inspired by Bos (1965) and Salop (1979); see Fujita and Thisse (2002, pp. 119–124) for a discussion of these models. To keep notations simple, we will drop the time subscript in this section. Assume in a first stage that the number of schools E as well as the tuition cost k are fixed. Remember we assumed that k is the same for all the students of a given generation, is independent of the duration of the studies, and is paid once for all.

Given our hypothesis on the dispersion of the population, the schools will be optimally located if they are evenly spaced. To get rid of indeterminacy, we will assume that, provided schools are created, there will be one at 0. Hence, the schools are located at $(j - 1)/E$, with $j = 1, \dots, E$.

The potential catchment area of the school at 0 is represented in Figure 4. It is the circular segment $[-1/2E, 1/2E]$. The members of the new-born cohort who attend school are the persons with a high ability μ such that $\mu > \tilde{\mu}(i, k)$, where the function $\tilde{\mu}(i, k)$ was defined in Proposition 3. The distance function $x(i)$ is

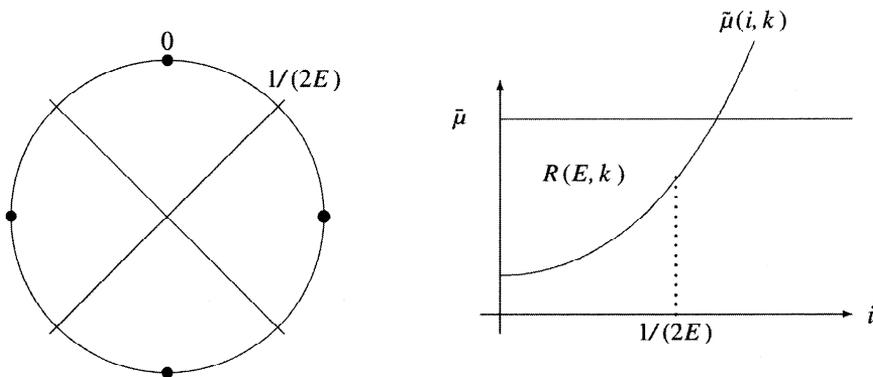


FIGURE 4. Contiguous catchment areas.

the arc length between location i and the closest school, hence in the catchment area of 0, $x(i) = |i|$. As stated in Proposition 4, the function $\tilde{\mu}(\cdot)$ is increasing in i , reflecting the idea that, for a population very close to the school, many children are likely to attend courses, whereas for very distant populations, only the most skilled do. Let us denote the attendance rate of population located at i by $r(i, k)$. It is given by

$$r(i, k) = \int_{\tilde{\mu}(i,k)}^{\infty} g(\mu) d\mu.$$

The attendance rate in the catchment area of the school at 0 is

$$R(E, k) = 2 \int_0^{\frac{1}{2E}} r(i, k) di = 2 \int_0^{\frac{1}{2E}} \int_{\tilde{\mu}(i,k)}^{\infty} g(\mu) d\mu di. \tag{15}$$

The benefit brought by a school is then

$$k\zeta R(E, k) - f. \tag{16}$$

The value of the decision variables E and k will be determined according to the institutional settings presented in Section 3.

5.1. Model M1

In model M1, we suppose the existence of a central authority that determines simultaneously the optimal density of schools and the tuition fee so as to maximize the profit of the whole schooling system. The global profit of the central authority when implementing E schools and imposing a tuition fee k is equal to the tuition paid by the total population of children of that generation who attend school, minus the set-up cost for opening the schools:

$$B(E, k) = [k\zeta R(E, k) - f]E.$$

The problem of the central authority is then to choose E and k so as to maximize global profit:

$$\text{M1: } \max_{E,k} B(E, k) = \max_{E,k} (k\zeta R(E, k) - f)E.$$

To characterize the solution of M1, we first establish the following property of the functions R and B .

LEMMA 1. *At given k , the attendance rate $R(E, k)$ is a decreasing function of E , whereas the benefit function $B(E, k)$ is a continuous concave function of E .*

See Appendix B.6 for the proof.

Given that $R(E, k)$ is decreasing in E , Proposition 5 is straightforward.

PROPOSITION 5. *A necessary and sufficient condition for schools to be created is that*

$$\zeta \geq \underline{\zeta} = \frac{f}{\max_k kR(1, k)} = \frac{f}{\tilde{k}R(1, \tilde{k})}.$$

The proposition says that when the newborn population is low, set-up costs are not covered, hence no schools are created. Only when the population density reaches the threshold $\underline{\zeta}$ do schools start to be created. The following lemma will prove most helpful for studying the properties of problem M1.

LEMMA 2. *$\tilde{\mu}(i, k)$ tends to infinity when k goes to infinity.*

See Appendix B.8 for the proof.

Notice that by (14) and the definition of function $\Psi(\cdot)$, $\tilde{\mu}(i, k)$ is a linear function of k given that the interior optimal schooling decision does not depend on k . Let us now look at our optimization problem. Observe that the profit $B(E, k) = (k\zeta R(E, k) - f)E$ is equal to $-fE$ for $k = 0$. We now state a lemma that proves that the profit from school foundation is also negative when k goes to infinity.

LEMMA 3. *There exists a constant K independent of k and E , such that the profit function $B(E, k) < 0$ for all $k > K$, and $E > 0$.*

See Appendix B.9 for the proof.

Now, consider the optimization problem of $B(E, k)$ for fixed $E > 0$. The problem admits necessarily a maximum on the compact $k \in [0; K]$. This maximum is necessarily interior because

$$\frac{\partial B(E, k)}{\partial k}(0) = 2\zeta E \int_0^{\frac{1}{2E}} \int_{\tilde{\mu}(i, 0)}^{\infty} g(\mu) d\mu di > 0.$$

This maximum is also a global maximum because $B(E, k) = B(E, 0) = -fE$, $\forall k > K$. Denote by \tilde{k} , the corresponding maximizer. Because it is interior, it satisfies the equation

$$\frac{\partial B(E, k)}{\partial k}(\tilde{k}) = 0,$$

or

$$\tilde{k} \frac{\partial R(E, k)}{\partial k}(\tilde{k}) + R(E, \tilde{k}) = 0.$$

Because the function $B(E, k)$ reaches its (strict) maximum at \tilde{k} , it should be strictly concave with respect to k at $k = \tilde{k}$. Henceforth, the derivative of the

left-hand side of the equality above should be strictly negative, and the implicit function theorem holds. For any E , we can associate a single $\tilde{k}(E)$. Moreover function $\tilde{k}(E)$ is differentiable by the implicit function theorem. Clearly, this maximizer is independent of f . Therefore, for the maximized $B(E, k)$, for fixed $E > 0$, to be non-negative, we need the condition

$$\tilde{k}(E) \zeta R(E, \tilde{k}(E)) = \Theta(E) \geq f.$$

Because $\tilde{k}(E)$ is differentiable, function $\Theta(E)$ is also differentiable (and thus continuous). Now notice that the previous non-negative profit condition cannot hold for E tending to infinity because, by definition, the attendance rate $R(E, k)$ goes to zero in such a case, and because $\tilde{k}(E) < K$. Therefore, the condition $\Theta(E) \geq f$ should correspond to a bounded E -set if non-empty. If this set is empty, then $E > 0$ cannot yield non-negative profits and the corner solution $E = 0$ applies. Otherwise, creating schools becomes optimal. We can now solve the free optimization problem with respect to (E, k) .

PROPOSITION 6. *The scenario M1 has a unique and global optimal solution for (E, k) . The maximized profit is necessarily non-negative.*

See Appendix B.10 for the proof.

5.2. Model M2

In the scenario M2, we assume a free-entry process in which school creation decisions are totally decentralized: schools will be created as long as profit is possible.

The problem can be stated as a two-stage procedure. In the first stage, E schools are created. In the second stage, each school chooses k to maximize its profit. Clearly, this leads to the same optimal tuition $\tilde{k}(E)$ for every location. If the profit is negative, there are too many schools and one will run out of business. If the profit is positive, there is potentially room for at least one more school. Formally the problem can be stated as follows:

$$\max E \text{ subject to: } B(E, \tilde{k}(E)) \geq 0 \text{ and } B(E, \tilde{k}(E)) \geq B(E, k) \forall k > 0 \quad (17)$$

It can be easily proved that, ceteris paribus, the number of schools opened under scenario M2 is always higher than under scenario M1.

6. Baseline Scenario

In this section, we calibrate the model in order to get a baseline simulation against which we can run counterfactual experiments. The counterfactual experiments

will allow us to isolate the effect of each of the important factors on schooling, literacy and growth. A robustness analysis to some of our assumptions will be carried over in Section 7. We consider the period 1530–1860, which corresponds to the period considered by Wrigley and Schofield (1989), followed by Wrigley et al. (1997), for which they build a comprehensive set of demographic data. To be consistent with our focus on adult mortality, we let the birth date in our model correspond to age 5 in the data. We first calibrate the exogenous processes and the parameters of the model.

6.1. Calibration

The four exogenous processes α_t , β_t , ζ_t , and γ_t should be made explicit. We assume that all these four processes follow a polynomial function of time. Polynomials of order 3 are sufficient to capture the main trends in the data. For the survival function processes α_t and β_t , the parameters of the polynomial are chosen by minimizing the Euclidian distance with the survival functions estimated by Wrigley et al. (1997). These survival functions apply to the age 5–85, and have been accordingly normalized to 1 at age 5, to abstract from infant mortality. The parameters of the process for ζ_t are chosen so that the Euclidian distance between total population P_t defined in equation (3) and the observed level of population aged 5 and over is minimized. Finally, the parameter of the exogenous technological progress is set to follow the estimated level of total factor productivity of Figure 2. For the period running after 1860 until 1920, which intervenes in the expected wages of the persons born in the nineteenth century, we assume a constant total factor productivity growth of one percent per annum (Clark 2003).

All variables are assumed constant before 1530. The survival probabilities as implied by the calibrated α_t and β_t first drop, then increase. The level of 1530 is reached and surpassed in 1720. Birth density ζ_t increases in two distinct periods: 1530–1600 and 1730–1860. The productivity process A_t is relatively constant until 1700, with a marked acceleration thereafter.

For the risk-free interest rate, we choose $\theta = 0.05$, which lies in the range provided in Epstein (2000). For the other parameters, ξ , f , w^h , and the standard error σ , we unfortunately cannot estimate them in the same way micro-econometricians unravel preference and technology parameters from contemporaneous data. We will therefore choose them to match a set of moments. Note that the mean of the distribution does not give an additional degree of freedom, and multiplying ξ , f , w^h , and the mean by the same factor is neutral to the result. The four moments used to select the parameters are: the level of schooling $E_{1820} = 120$ (there are 3,000 schools in our databank in 1820, so the chosen scale of our model economy is 1/25), the level of literacy (denoted Λ , see equation (18), subsequently), $\Lambda_{1820} = 0.55$, the change of literacy over the period 1540–1820 $\Delta\Lambda = 0.45$, and

TABLE 1. Calibrated parameters.

	M1	M2
ξ	75	9
f	0.002	0.01785
w^h	0.583256	0.59077
σ	0.30	0.34

a skill premium of 60% on average over the period for seven years of education (from van Zanden [2004], the 7-years’ apprenticeship period was the standard in English contracts for skilled craftsmen). The parameters needed to match those moments are presented in Table 1. We observe that we need higher fixed costs in model M2 to obtain the same number of schools as in M1. This requires a lower transportation cost in M2 to match literacy.

6.2. Baseline Simulation

Figure 5 reports the optimal density of schools resulting from the maximization process by the central school authority (M1) or from the free-entry process (M2). The model has been calibrated so as to match exactly the observed level in 1820. Both densities of schools increase monotonically, with a slowdown in

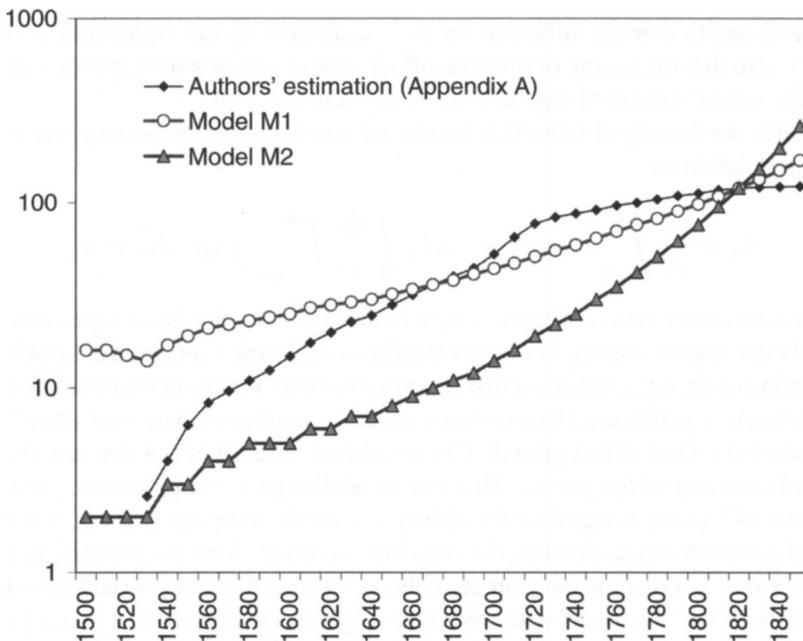


FIGURE 5. Baseline simulation—density of schools E_t .

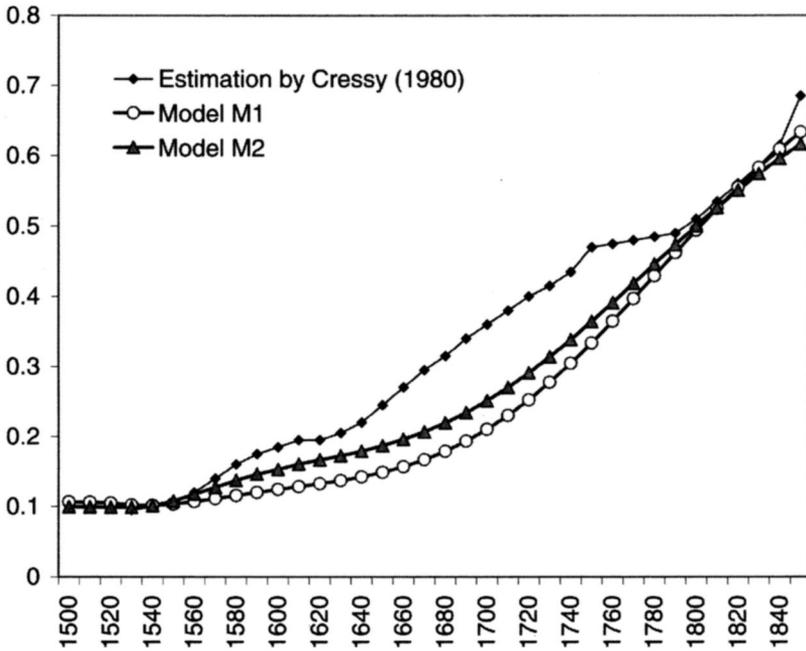


FIGURE 6. Baseline simulation—literacy rate, total population.

the seventeenth century followed by an acceleration in the eighteenth century. Notice also that the timing of the take-off for school creation does not vary across models; establishment of new schools really starts in 1540.

Once the density of schools is known we can compute the literacy rate of the total population as

$$\Lambda_t = \frac{1}{P_t} \int_{t-L}^t \zeta_z m_z(t-z) 2E_z \int_0^{\frac{1}{2E_z}} \int_{\bar{\mu}_z(i, k_z)}^{\infty} g(\mu) d\mu di dz. \quad (18)$$

The results are reported in Figure 6. Bear in mind that the simulated series matches exactly the overall change in literacy thanks to calibration procedure. Looking at the timing of changes, there is a first rise prior to 1600, thanks to the creation of the first schools. It is followed by a period with slower improvements, and, after 1700, by a second period of fast growth. Our simulated series does not rise fast enough in the beginning of the period. This can be attributed to many factors. First, the estimate of literacy is based on the ability to sign the marriage register, it is not a school attendance rate. Second, the simulated series are for total population; if we had reported a series for the younger cohorts, achieved literacy would have been higher. Note that the model with free entry achieves a higher level of literacy prior to 1820, with fewer schools; this is because the calibrated transportation cost in this model is lower.

Next we compute the total stock of human capital and total GDP. The total stock of human capital integrates all the generations which are currently at work in the modern sector:

$$\begin{aligned}
 H_t &= \int_{t-\bar{L}}^t \zeta_z m_z(t-z) 2E_z \\
 &\quad \times \int_0^{\frac{1}{2E_z}} \int_{\bar{\mu}_z(i, k_z)}^\infty \delta[t-z-S_z(\mu, i)] h_z(\mu, i) g(\mu) d\mu di dz.
 \end{aligned}$$

For those who go to school, we can compute the transportation costs as follows:

$$\begin{aligned}
 \Xi_t &= -\xi \int_{t-\bar{L}}^t \zeta_z m_z(t-z) 2E_z \\
 &\quad \times \int_0^{\frac{1}{2E_z}} \int_{\bar{\mu}_z(i, k_z)}^\infty \delta[S_z(\mu, i) - (t-z)] x(i) g(\mu) d\mu di dz.
 \end{aligned}$$

Finally we can compute total GDP I_t as the sum of the production of the traditional and modern sector minus the transportation cost and the set-up cost of schools:

$$I_t = w^h (1 - \Lambda_t) P_t + A_t (H_t - \Xi_t - f E_t). \tag{19}$$

Figure 7 plots the growth rate of I_t/P_t over 10 years together with the Maddison (2001) estimate. After a period of initial stagnation, there is a period of negative growth after 1530, because the economy has to pay the transportation costs of students and the set-up cost of schools, but does not yet benefit from better-educated persons. Thereafter there is a mini-boom when the newly educated generations start to work. Next, the seventeenth century is characterized by low growth. After this stagnation period, growth starts accelerating after 1700 to reach 0.7% per annum at the end of the eighteenth century.

Notice that our growth numbers should be interpreted as the growth generated by the accumulation of human capital and by TFP growth, without any effect from the accumulation of physical capital. This is why our model underestimates growth after 1820 compared to Maddison data.

With free entry, the density of schools increases faster than with a central authority. This fast rise entails important fixed costs for the economy, especially after 1820, which slows growth compared to the central authority case. The model with free entry is therefore not as good as the model with a central authority to reproduce the acceleration in growth during the early nineteenth century because it would imply too many school establishments.

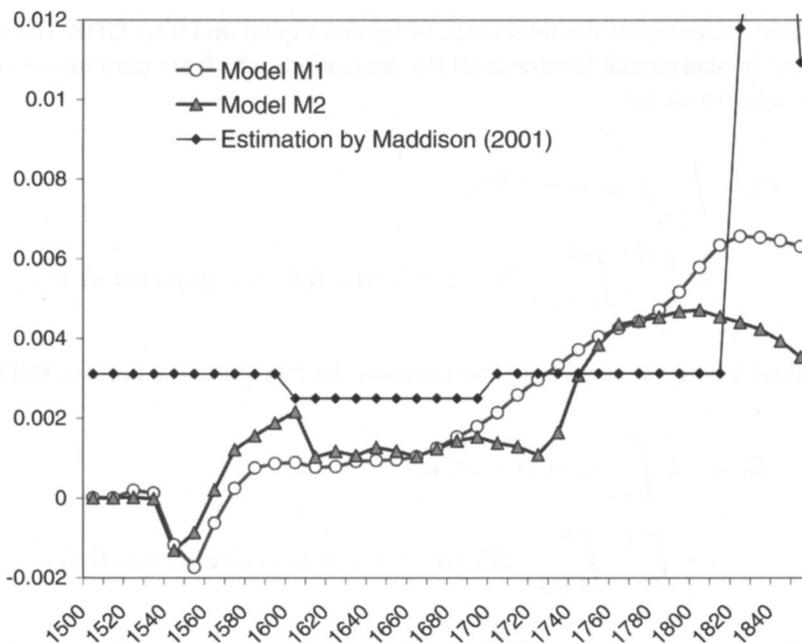


FIGURE 7. Baseline simulation—growth of income per capita.

6.3. Counterfactual Experiments

There are four exogenous processes in the model, representing three forces: mortality (with two processes), birth density, and exogenous technical progress. To evaluate the importance of each of them in accounting for literacy and growth, we have run experiments where we fix two of the exogenous variables to a constant path and allow the third to vary with the calibration. A summary of the results is provided in Table 2.

In the first simulation we have considered that birth density and technical progress are constant over the period; this allows us to isolate the role of mortality. Because mortality drops very late in England (see the following section), it does not exert positive influence before the eighteenth century. If mortality improvements were the only driving force of the industrial revolution, new schools would be established from 1650 onward, and the literacy rate would have increased by only 6.8% over 1500–1850, using model M1. In addition, GDP per capita would only have increased by 7.4% over 1500–1850. Compared to the baseline simulation, mortality improvements explains 6.5% of total school creations over the period 1500–1850, 12.8% of improvements in literacy and 7.5% of growth of income per capita. With model M2, the picture is slightly different. Now mortality explains 20% of the rise in literacy and 15.2% of income growth.

TABLE 2. Counterfactual analysis—importance of each factor.

Periods	Baseline			Mortality			Birth density			Productivity		
	ΔE_t	$\Delta \Lambda_t$	$\Delta I_t/P_t$	ΔE_t	$\Delta \Lambda_t$	$\Delta I_t/P_t$	ΔE_t	$\Delta \Lambda_t$	$\Delta I_t/P_t$	ΔE_t	$\Delta \Lambda_t$	$\Delta I_t/P_t$
Model M1												
1500–1549	3	-0.3%	-1.9%	-1	-0.2%	0.2%	7	1.0%	-1.5%	-4	-1.2%	-0.3%
1550–1599	6	2.1%	1.5%	-3	-1.8%	-1.7%	8	4.3%	4.8%	-3	-3.2%	-4.0%
1600–1649	7	2.4%	2.6%	0	-1.2%	-2.6%	2	1.9%	4.0%	8	2.0%	-0.8%
1650–1699	12	6.2%	5.9%	4	1.5%	0.9%	2	0.6%	1.0%	10	7.2%	7.3%
1700–1749	20	12.3%	16.0%	3	2.6%	2.9%	6	0.9%	-0.1%	11	11.4%	15.7%
1750–1799	34	16.0%	30.7%	4	2.9%	3.5%	15	2.2%	-0.2%	9	14.1%	31.2%
1800–1850	72	13.9%	44.3%	3	3.0%	4.1%	40	3.5%	-0.3%	6	12.8%	50.8%
1500–1850	154	52.7%	99.2%	10	6.8%	7.4%	80	14.5%	7.7%	37	43.0%	99.9%
<i>Share explained</i>				6.5%	12.8%	7.5%	51.9%	27.5%	7.8%	24.0%	81.6%	100.7%
Model M2												
1500–1549	1	0.8%	-1.5%	0	0.0%	0.1%	2	2.3%	-1.9%	0	-0.2%	-0.2%
1550–1599	2	4.5%	5.2%	-1	-1.6%	0.3%	3	7.3%	10.9%	-1	-4.1%	-4.4%
1600–1649	3	3.4%	4.5%	0	-3.2%	-6.7%	2	2.3%	5.3%	2	2.8%	-0.8%
1650–1699	6	6.4%	5.5%	1	3.7%	3.4%	0	0.8%	1.7%	3	9.4%	10.4%
1700–1749	16	11.3%	9.9%	1	3.8%	4.6%	3	0.9%	-0.6%	5	12.7%	17.8%
1750–1799	46	13.6%	24.4%	1	4.0%	6.1%	9	1.9%	-0.8%	6	13.7%	27.2%
1800–1850	185	11.6%	27.5%	2	4.0%	3.7%	34	2.4%	-2.8%	7	11.7%	47.0%
1500–1850	259	51.7%	75.4%	4	10.7%	11.4%	53	17.8%	11.7%	22	46.0%	97.0%
<i>Share explained</i>				1.5%	20.8%	15.2%	20.5%	34.5%	15.5%	8.5%	89.1%	128.6%

In a second step we run a simulation where both mortality and technical progress are constant. Only the birth density ζ_t varies, reflecting all changes in population which are not due to mortality. In this simulation we observe that the rise in population of the sixteenth century can be held responsible for school creation and increases in literacy during the sixteenth century. In the seventeenth century, however, population stagnates, and school creation drops. In the end, the rise in birth density explains the majority of total school creations over the period 1500–1850, 27.5% of improvements in literacy and 7.8% of income growth per capita. With model M2, birth density explains more literacy and more growth but less school foundations than with model M1.

In a third step we run a simulation where both mortality and birth density are constant. Only the technical progress A_t is variable. In this simulation we observe that technical progress is unable to explain the timing of school establishment and literacy. But, in the end, technical progress explains a major part of literacy changes and growth.

From the counterfactual analysis we conclude that neither productivity increases nor mortality improvements can explain the establishment of schools at a high rate in the sixteenth century documented in Appendix A and Figure 1. Only the rise in population density can. This view is in line with the numerical results of Lagerloef (2006) according to which the transition from the Malthusian regime to the post-Malthusian regime follows from a population externality, a more populated economy favoring the accumulation of skills. Still, the rise in productivity explains most of output growth over the period. Because these conclusions hold both in models M1 and M2, they are robust to the institutional set-up.

6.4. Further Discussion of the Results

These results merit further discussion, specially concerning two points: the small role played by mortality and the importance of technical progress.

As far as mortality is concerned, our counterfactual experiments estimate that mortality improvements account at best about 20% of the rise in literacy and growth. This is small especially compared to the important role it was supposed to play in Nicolini (2004) for England and Boucekkine, de la Croix, and Licandro (2003) for Western Europe as a whole. To better understand the timing of mortality drops in England, we have compared the survival function in England as estimated by Wrigley et al. (1997) to the one in Geneva (Perrenoud 1978). Perrenoud's data consists of a high-quality data set for the period 1600–1800, including mortality tables estimated from parish register information. It is the best we can find in continental Europe. Compared to Geneva, English longevity was remarkably high during the early seventeenth century. Then English mortality increased while mortality in Geneva dropped substantially, becoming lower compared to England. This different evolution in mortality is probably due to the sharp

rise in urbanization in England. At that time, big cities were unhealthy places. According to Bairoch, Batou, and Chèvre (1988), the population of London went from 50,000 in 1500 to 948,000 in 1800, whereas it grew only from 12,000 to 25,000 in Geneva over the whole period. In the eighteenth century, finally, survival probabilities increased at all ages in both England and Geneva. Over the two centuries, the improvements in longevity were steady in Geneva, while they came quite late in England probably because of the fast urbanization process. This is why, in our quantitative exercise, mortality reductions play a role only in the eighteenth century.

Concerning technical progress, we have assumed it follows a purely exogenous process as in standard neo-classical growth theory. The effect of population density passes uniquely through enhancing the quality of the labor force, via a better access to schooling. If, on the contrary, we had assumed that the stock of human capital exerts a positive influence on total factor productivity (TFP) through an externality as in Lucas (1988), the increase in total factor productivity would have been partly related to the density effect. An example of a formulation with externality consists in replacing equation (4), with:

$$A_t = e^{(1-\rho)\gamma t} H_t^\rho, \quad (20)$$

where the parameter ρ is related to the intensity of the externality. Such a formulation amounts to partially endogenizing technical progress, the extreme case $\rho = 1$ describing a model with endogenous growth. Unfortunately, including such an externality in the model would make the computations tremendously more complicated than a simple robustness check. For example, this would make the schooling decisions of the individuals depending on aggregate human capital, which in turn would involve that the ability thresholds, as featured in Proposition 3, depend on aggregate human capital. Because the computation of aggregate human capital (and thus of GDP) requires such thresholds, we have ultimately to solve a terrific fixed point problem. We can, however, be certain that, with this formulation, the estimation of the importance of the birth density effect would be larger than in our baseline simulation. Indeed, the initial effect would be amplified by the externality over time. We should therefore interpret the estimation according to which population density is responsible for one-third of the rise in literacy and one-fifth of the rise in growth as a lower bound, assuming that technical progress is not related to population density.¹¹

At last, notice that it would be interesting to introduce a feedback from the process of development to population growth. This would enable us to have a complete dynamical system in population and schooling that evolves endogenously till

11. This comment is also robust to the inclusion of urbanization. Modeling urbanization caused by increases in population would have reinforced our density effect by reducing distance (that is, shrinking the circle).

the threshold of profitable school construction is reached. In this set-up, the mechanism that assure the creation of schools becomes an inevitable by-product of the Malthusian interaction between technology and population. We are, however, not able to study the Malthusian mechanism either analytically nor through a robustness check exercise, and thus we prefer not to refer to such a mechanism. Introducing a Malthusian structure would require some simplifications elsewhere, and most likely swapping the latter for the realistic survival laws included in this paper.

7. Robustness Analysis

We now provide a robustness analysis by computing how the baseline simulation is modified when we change parameters and assumptions. We concentrate on model M1 for expositional purpose, but all the results carry over to M2. For each alternative assumption, we recalibrate the parameters ξ , f , w^h , and σ , and display how the baseline simulation is affected. Table 3 gives the calibration results, and Figures 8, 9, and 10 show the simulations.

In the baseline simulation, transportation costs are indexed on technical progress, see equation (5). This assumption is probably too pessimistic because transportation costs relative to other costs have probably been reduced thanks to urbanization. To evaluate the importance of this assumption we run a simulation where the transportation cost is kept not indexed. The second column of Table 3 provides the result of the calibration procedure. Only ξ and f need to be adjusted. Figures show that the baseline simulation with non-indexed transport costs yield the same results for schools, literacy and growth. Hence, when the assumption on transportation cost is modified, the calibration is adjusted to keep the same density of schools in 1820, and both literacy and growth are almost unaffected. Hence our previous results are robust to the assumption on transportation costs.

We next analyze the robustness of the result to a change in the method of calibration. In this paper, we have matched the change of literacy over the period 1540–1820 $\Delta\Lambda = 0.45$, in order to obtain the right magnitude in literacy improvements so as to generate counterfactuals. Alternatively, we can match the increase in school creation over the same period. This amounts to targeting a density of school of 4 (100/25) in 1540. This condition, however cannot be matched

TABLE 3. Robustness analysis—calibration procedure, M1.

	Baseline	Transportation costs not indexed	Calibration on schools rather than on literacy	Myopic foresight after 1860	$\theta = .03$
ξ	75	61	79.2	75	183
f	0.0020	0.0016	0.00217	0.0020	0.0064
w^h	0.583256	0.58537	0.583138	0.583256	0.596655
σ	0.30	0.30	0.30	0.30	0.368

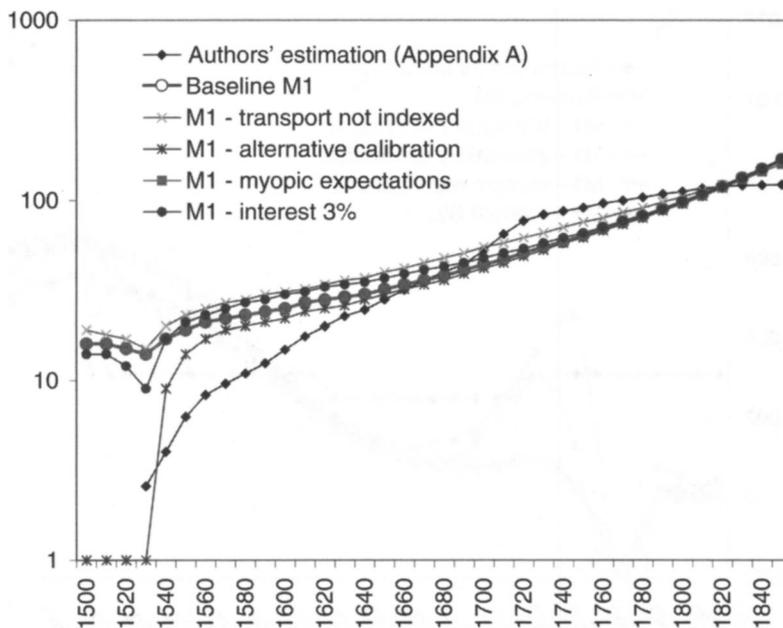


FIGURE 8. Robustness—density of schools E_t .

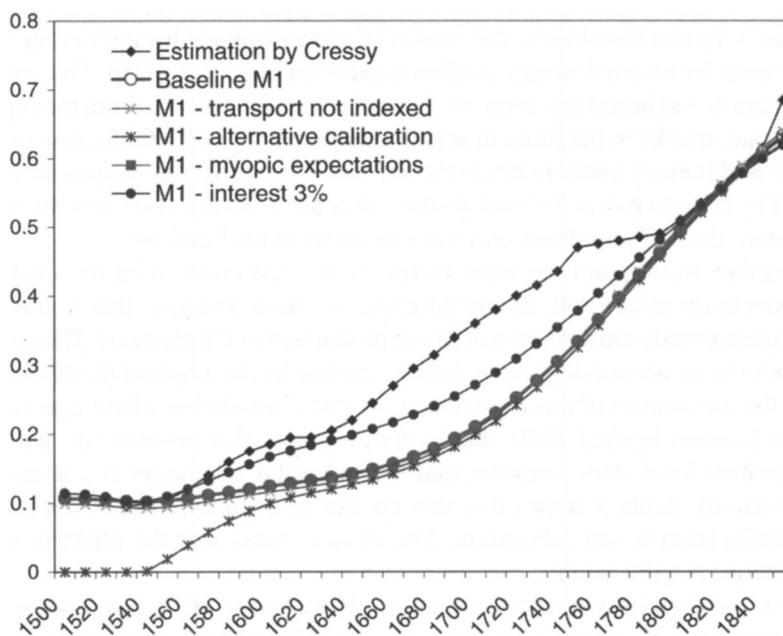


FIGURE 9. Robustness—literacy rate, total population.

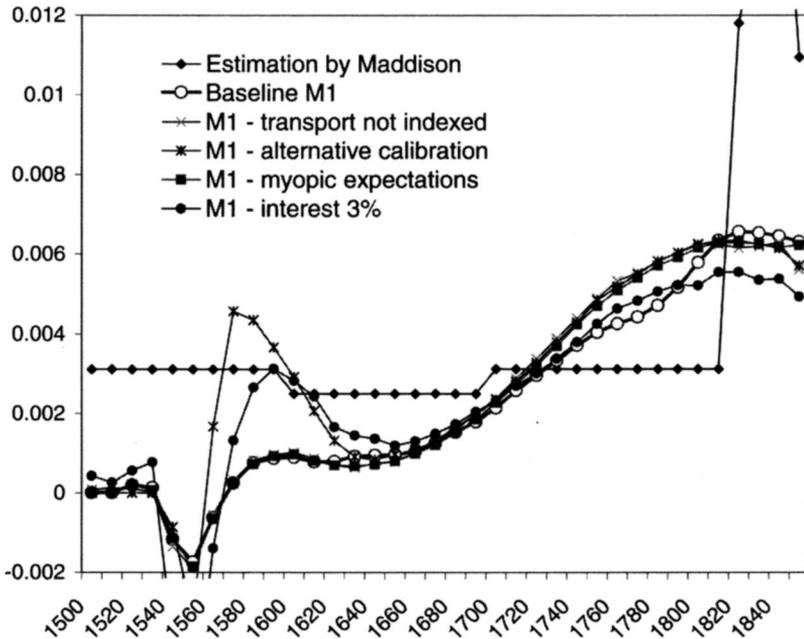


FIGURE 10. Robustness—growth of income per capita.

exactly. We still obtain a very close result with the parameters in the third column of Table 3. In that simulation, the density of schools jumps from 0 to 8 in 1540. The process by which illiteracy is eliminated is initiated by a jump. This jump is larger than 4, and hence that number cannot be reached exactly. From the Figures we see that, thanks to the jump in school establishment in 1540, the new path of schools and literacy better matches the estimated increase over the period 1540–1600. The price to pay is the total absence of schools before 1540. For the rest of the period, the new simulated series is very close to the baseline.

Another assumption we want to test is the one concerning the evolution of productivity after 1860. In the baseline, we have assumed that households anticipate correctly the evolution of future productivity (1% per year). This creates an incentive to accumulate more human capital in the nineteenth century. To assess the importance of this mechanism, we run a simulation where agents have myopic forecast beyond 1860, that is, they suppose that productivity will stay at a constant level (they consider that the industrial revolution is a temporary phenomenon). Table 3 shows that this change in assumption does not require any modification in the calibration. The figures shows that the effect of lower expectations is quite small.

In a fourth robustness test, we take a lower value of θ , assuming a risk-free interest rate of 3% per year instead of 5%. The other parameters need to be adjusted. A lower θ gives an incentive for households to get more education, and so

we need higher transportation costs to match the observed education investment. In the figures we observe that the number of schools is very close to the baseline, whereas literacy increases faster in the beginning of the period. Setting $\theta = 3\%$ would bring our simulated literacy closer to the estimate by Cressy.

Until now, the robustness analysis indicates that the result on literacy and growth are little affected by changes in the parameters. This conclusion is not valid however, when the parameter w^h is concerned. If, for example, we index w^h on productivity A_t , there is no way to choose the calibrated parameters so as to match the targeted moments, and in particular the rise in literacy over the period. In fact, the non-indexation of w^h is the main mechanism through which technical progress plays a role in the model. If we shut down this channel by indexing w^h , we reduce drastically the role of technical progress, and we are left with the two other factors, mortality and birth density, which together explain about 40% of the observed rise in literacy.

8. Conclusion

Literacy in England started to rise in the sixteenth century, two centuries before the official Industrial Revolution. In models of the transition from stagnation to growth, three economic factors can be held responsible for this fact. (a) Exogenous technical progress in the modern sector increased the return to investment in education. (b) Lower mortality increased the return to education and induced higher investment in human capital. (c) Rising density of population induced economic gains which in turn raised the incentive to educate.

The first contribution of this paper is to propose micro-foundations for the effect of population density on growth. We assume that the investment in education is chosen by individuals as a function of future wages, longevity, and the distance to the nearest school. The number and location of education facilities are chosen optimally either to maximize the profit of the whole school system or by assuming a free entry process. In this set-up, higher population density increases the number of schools, opening the possibility for individuals to reach higher educational levels.

Secondly, we use our model to measure the impact of the three factors on school density, literacy, and growth through a set of counterfactual experiments. We find that one-third of the rise in literacy over the period 1530–1850 can be directly related to the effect of density, and one-sixth is linked to higher longevity. The estimation of the effect of density is a lower bound, because we assume that density affects the productivity of labor through enhancement in human capital through schools, but we do not assume any externality from human capital to total factor productivity.

The predictions of the model are then compared to school establishments data that we have gathered from the reports of the School Inquiry Commission written

in 1867–1868. These data show a peak in school creation at the end of the sixteenth century. Our counterfactual analyses conclude that neither productivity increases nor mortality improvements can explain this peak. But the rise in population density can.

It is often claimed that the impressive performance of pre-industrial England in terms of education was related to cultural and religious factors (Protestantism). Our analysis shows that the surge in education provision as early as in the sixteenth century can also be understood as an optimal response to population density passing a given threshold.

Appendix A. School Establishments per Decade

	Grammar	Non-classical
<1500	23	2
1500–1510	9	2
1510–1520	8	2
1520–1530	13	0
1530–1540	8	0
1540–1550	36	0
1550–1560	47	10
1560–1570	44	7
1570–1580	31	2
1580–1590	19	14
1590–1600	26	13
1600–1610	41	18
1610–1620	42	27
1620–1630	27	33
1630–1640	31	38
1640–1650	15	33
1650–1660	42	45
1660–1670	34	58
1670–1680	38	67
1680–1690	31	75
1690–1700	25	81
1700–1710	23	188
1710–1720	24	287
1720–1730	23	279
1730–1740	12	152
1740–1750	9	91
1750–1760	4	93
1760–1770	3	128
1770–1780	7	84
1780–1790	8	99
1790–1800	6	106
1800–1810	3	87
1810–1820	1	125
1820–1830	3	64
1830–1840	4	36
1840–1850	6	9
1850–1860	9	4

Appendix B. Proofs

Appendix B.1. Proof of Proposition 1

We shall denote $S(\mu, i)$ by S . Therefore, we can rewrite the integral equation as follows (assuming $m(S) \neq 0$ or $S < L$):

$$\frac{\int_S^L e^{\gamma(z-S)} m(z) dz}{m(S)} - S = \phi, \tag{B.1}$$

where $\phi = \xi x(i)/\mu$. We shall denote by $\Phi(S)$ the function on the left-hand side of equation (B.1), and solve for $\Phi(S) = \phi$ for $S \in (0; L)$. First, notice that function $\Phi(S)$ can be expanded continuously to $S = L$. Indeed, simple Taylor first-order expansions allow to show that the limit of $\int_S^L e^{\gamma(z-S)} m(z) dz/m(S)$ is zero when S tends to L .¹² It follows that $\Phi(S)$ tends to $-L$ when S tends to L , so that (B.1) cannot have a solution in this neighborhood.

Next we prove that $\Phi(S)$ is decreasing for γ small enough and $S \in [0; L]$. By continuity of the problem with respect to γ , it is enough to prove this property for $\gamma = 0$. Differentiating $\Phi(S)$ and putting $\gamma = 0$, one gets

$$\Phi'(S) = -2 - \frac{m'(S)}{m^2(S)} \int_S^L m(z) dz.$$

To prove that function $\Phi(S)$ is decreasing, we shall demonstrate that

$$\frac{-m'(S)}{m^2(S)} \int_S^L m(z) dz \leq 1,$$

for $S \in [0; L]$. First, notice that

$$\frac{-m'(S)}{m^2(S)} \int_S^L m(z) dz = \frac{-m'(S)}{m(S)} \int_S^L \frac{m(z)}{m(S)} dz \leq \frac{-m'(S)}{m(S)} (L - S),$$

because function $m(S)$ is decreasing. We prove now that function

$$\rho(S) = -m'(S)/m(S)(L - S)$$

is strictly increasing on $[0; L]$. Indeed, the sign of $\rho'(S)$ is the one of $-\beta e^{\beta S} (\alpha - e^{\beta S}) + \alpha \beta^2 e^{\beta S} (L - S)$, which turns out to be the sign of the simpler expression

$$\alpha \beta (L - S) - (\alpha - e^{\beta S}).$$

12. $e^x - 1$ is almost equal to x when x is close to 0.

Now, we can see that the latter function should be positive: Its derivative is exactly $\beta(e^{\beta S} - \alpha) \leq 0$, for $0 \leq S \leq L$, and it is decreasing to 0 (its value when $S = L$). Henceforth, function $\rho(S)$ is increasing on the interval $[0; L]$. It is then straightforward, using the same kind of first-order Taylor expansion just shown, that the limit of $\rho(S) = \beta e^{\beta S} / (\alpha - e^{\beta S})(L - S)$ is exactly 1 when S tends to L . This in turn means that $\Phi'(S) < 0$ when $S \in [0; L]$. The necessary and sufficient condition for existence and uniqueness is then straightforward.

Appendix B.2. Proof of Corollary 1

In order to establish that the interior solution characterized in Proposition 1 is indeed a maximizer, we have to compute the second-order derivative of lifetime resources as given by (5) with respect to S . As in the proof of Proposition 1, we shall proceed under the simplifying assumption that $\gamma = 0$ (or by continuity, for γ small enough). The second-order derivative is computed as

$$-2 \mu m_t(S) - (\mu S + \xi x(i)) m_t'(S).$$

Now, use the first-order condition (12) to replace $\mu S + \xi x(i)$ by $\mu \int_S^L m(z) dz / m_t(S)$. Then the sign of the second-order derivative appears to be the one of function $\Phi'(S)$ seen in the proof of Proposition 1, which is unambiguously negative. Concavity of lifetime resources with respect to schooling follows.

Appendix B.3. Proof of Corollary 2

The proof is trivial except for the demographic parameters α and β . Notice that because $\mu(i)$ is given by equation (13), the effect on the threshold of an increase in α or β is the reverse of the effect on the integral $\int_0^L e^{\gamma a} m(a) da$. Differentiating this integral with respect to $p = \alpha$ or β yields $\int_0^L e^{\gamma a} \partial m(a) / \partial p da$, the effect via the integral limit L being zero as $m(L) = 0$. Because $\partial m(a) / \partial p > 0$ when $p = \alpha$, and $\partial m(a) / \partial p < 0$ when $p = \beta$, we get the announced results.

Appendix B.4. Proof of Proposition 2

Let us recall that the interior schooling decision is given by

$$\frac{\int_S^L e^{\gamma(z-S)} m(z) dz}{m(S)} - S = \phi,$$

with $\phi = \xi x(i) / \mu$. Again, the stated comparative statics properties are obvious except for the demographic parameters, α and β .

The last part of the proposition is straightforward: When $x(i) = 0, \phi = 0$, and the parameters ξ and μ vanish from the interior solution equation (B.1).

We shall prove the properties with respect to α and β for $\gamma = 0$, which by continuity of the problem with respect to γ , validate them for γ values small enough. Denoting by $\Phi^1(S) = \int_S^L m(z)dz/m(S)$, and by total differentiation of (B.1) with respect to S and p ($= \alpha$ or β), one gets

$$\frac{\partial S}{\partial p} = -\frac{\frac{\partial \Phi^1(S)}{\partial p}}{\Phi'(S)}.$$

We know by the proof of Proposition 1 that function $\Phi(S)$ is decreasing, hence the sign of $\partial S/\partial p$ is the sign of $\partial \Phi^1(S)/\partial p$. And the sign of the latter is the sign of

$$m(S) \int_S^L \frac{\partial m(z)}{\partial p} dz - \frac{\partial m(S)}{\partial p} \int_S^L m(z)dz.$$

For $p = \alpha$, one can see that the sign of latter expression is positive if and only if

$$\int_S^L \frac{\frac{\partial m(z)}{\partial \alpha}}{\frac{\partial m(S)}{\partial \alpha}} dz > \int_S^L \frac{m(z)}{m(S)} dz.$$

Because

$$\frac{\frac{\partial m(z)}{\partial \alpha}}{\frac{\partial m(S)}{\partial \alpha}} = \frac{e^{\beta z} - 1}{e^{\beta S} - 1},$$

and because the latter function is increasing from 1 while the integrand on the right-hand side of the inequality just above is obviously lower than 1 on $[0; L]$ as $m(S)$ is decreasing, it follows that $\partial \Phi^1(S)/\partial \alpha > 0$, which leads to $\partial S/\partial \alpha > 0$ as expected.

The result with respect to β is obtained with an identical reasoning. Notice that as $\partial m(z)/\partial \beta < 0, \forall z, 0 \leq z \leq L, \partial S/\partial \beta < 0$ if and only if

$$\int_S^L \frac{\frac{\partial m(z)}{\partial \beta}}{\frac{\partial m(S)}{\partial \beta}} dz > \int_S^L \frac{m(z)}{m(S)} dz.$$

Again, $(\partial m(z)/\partial \beta)/(\partial m(S)/\partial \beta) = ze^{\beta z}/(Se^{\beta S})$ is larger than 1 for $0 \leq z \leq L$, and the inequality holds.

Appendix B.5. Proof of Proposition 3.

Notice that properties (i) to (iii) hold if function Ψ is increasing in μ . By differentiation, we get

$$\Psi'(\mu) = \frac{\partial \Psi}{\partial \mu}(\mu, \hat{S}(\mu)) = \frac{\partial \Psi}{\partial \mu} + \frac{\partial \Psi}{\partial \hat{S}(\mu)} \frac{\partial \hat{S}(\mu)}{\partial \mu}.$$

Therefore, it remains to establish the positivity of the term $\frac{\partial \Psi}{\partial \hat{S}(\mu)}$. To unburden the presentation, we will denote it $\Psi'(\hat{S})$ hereafter.

When $x(i) = 0$, $\Psi(\cdot)$ becomes

$$\Psi(\hat{S}) = \hat{S}^2 e^{\gamma \hat{S}} m(\hat{S}) \mu.$$

Differentiating, one gets

$$\Psi'(\hat{S}) = \hat{S} e^{\gamma \hat{S}} [2m(\hat{S}) + \hat{S}(\gamma m(\hat{S}) + m'(\hat{S}))].$$

By the proof of Proposition 1, we know that the function $\Phi(S)$ is decreasing in S for γ small enough, which implies that at $S = \hat{S}$, we have:

$$\gamma m(\hat{S}) + m'(\hat{S}) \geq \frac{-2 m^2(\hat{S})}{\int_{\hat{S}}^L e^{\gamma(z-\hat{S})} m(z) dz}. \quad (\text{B.2})$$

Therefore, when $x(i) = 0$,

$$\Psi'(\hat{S}) \geq 2\hat{S} e^{\gamma \hat{S}} m(\hat{S}) \left[1 - \frac{\hat{S} m(\hat{S})}{\int_{\hat{S}}^L e^{\gamma(z-\hat{S})} m(z) dz} \right].$$

Because, from equation (B.1), \hat{S} is given by

$$\int_{\hat{S}}^L e^{\gamma(z-\hat{S})} m(z) dz = \hat{S} m(\hat{S})$$

when $\phi = \xi w(i)/\mu = 0$, we get $\Psi'(\hat{S}) \geq 0$.

If $x(i) > 0$, then, computing the derivative of Ψ with respect to \hat{S} , one concludes that the sign of this derivative is the one of

$$\frac{\mu}{\xi x(i)} \left(\gamma \hat{S} m(\hat{S}) + 2m(\hat{S}) + \hat{S} m'(\hat{S}) \right) + \gamma m(\hat{S}) + m'(\hat{S}).$$

By the proof of Proposition 1, we know that function

$$\Phi(S) = \int_S^L e^{\gamma(z-S)} m(z) dz / m(S) - S$$

is decreasing in S for γ small enough, which implies that at $S = \hat{S}$, we have equation (B.2). Therefore, Ψ is increasing in \hat{S} if and only if

$$\left(\frac{\mu}{\xi x(i)} \hat{S} + 1 \right) \frac{-2 m^2(\hat{S})}{\int_{\hat{S}}^L e^{\gamma(z-\hat{S})} m(z) dz} + \frac{2\mu}{\xi x(i)} m(\hat{S}) \geq 0.$$

We shall see that the latter large inequality is in fact checked with equality. Rearranging terms and simplifying, it is enough to prove

$$\frac{\mu}{\xi x(i)} \left[1 - \frac{\hat{S} m(\hat{S})}{\int_{\hat{S}}^L e^{\gamma(z-\hat{S})} m(z) dz} \right] - \frac{m(\hat{S})}{\int_{\hat{S}}^L e^{\gamma(z-\hat{S})} m(z) dz} = 0.$$

The reader can check that this equality is true because at the interior solution $\Phi(\hat{S}) = \xi x(i)/\mu$.

In order to see that $\tilde{\mu}(i, k) > \underline{\mu}(i)$, just notice that as $\mu\mu(i)$ tends to $\underline{\mu}(i)$, \hat{S} goes to zero by Proposition 1. Hence $\tilde{\mu}(i, k) = \Psi^{-1}(k + e^{-\gamma t} W^h) > \Psi^{-1}(0) = \underline{\mu}(i)$.

Appendix B.6. Proof of Proposition 4

The comparative statics with respect to k and w^h are obvious. For the remaining parameters, the comparative statics are obtained by total differentiation of $\Psi(\hat{S}(\mu)) = k + e^{-\gamma t} W^h = k + e^{-\gamma t} \int_0^L w^h m(z) dz$. For any parameter $p = \alpha, \beta, \gamma, \xi$, and $x(i)$, we have:

$$\frac{\partial \tilde{\mu}(i, k)}{\partial p} = - \frac{\frac{\partial \Psi}{\partial \hat{S}} \frac{\partial \hat{S}}{\partial p} + \frac{\partial \Psi^1}{\partial p}}{\frac{\partial \Psi(\hat{S})}{\partial \mu}}, \tag{B.3}$$

where

$$\frac{\partial \Psi(\hat{S})}{\partial \mu} = \frac{\partial \Psi}{\partial \hat{S}} \frac{\partial \hat{S}}{\partial \mu} + \frac{\partial \Psi}{\partial \mu},$$

and $\Psi^1 = \Psi - e^{-\gamma t} \int_0^L w^h m(z) dz$. We know that denominator in (B.3) is positive given that function $\Psi(\cdot)$ is strictly increasing by the proof of Proposition 3, \hat{S} is an increasing function of μ by Proposition 2, and $\partial \Psi / \partial \mu = \hat{S}^2 e^{\gamma \hat{S}} m(\hat{S}) > 0$. Therefore, we shall concentrate on the sign of the numerator in (B.3). Finally, notice that $\partial \Psi^1 / \partial p = \partial \Psi / \partial p$ for all parameters p except γ . Indeed, though α and β enter the integration bound L which appears in $\Psi - \Psi^1 = e^{-\gamma t} \int_0^L w^h m(z) dz$, the derivative of the integral just before with respect to either α and β is nil because $m(L) = 0$.

For $p = \xi x(i)$ (or equal to any of the parameters of the product), we have $\partial \hat{S} / (\partial \xi x(i)) = 1 / (\mu \Phi'(\hat{S})) < 0$, because function $\Phi(\cdot)$ is decreasing by Proposition 1, and

$$\frac{\partial \Psi}{\partial \xi x(i)} = \hat{S} m(\hat{S}) e^{\gamma \hat{S}} - \int_0^{\hat{S}} e^{\gamma z} m(z) dz < 0,$$

for γ small enough.¹³ Because $\partial \Psi / \partial \hat{S} > 0$ by the proof of Proposition 3, it follows that the numerator in (B.3) is positive. Therefore, $\tilde{\mu}(i, k)$ is an increasing function of ξ and $x(i)$, as expected.

We now move to the demographic parameters α and μ . For $p = \alpha$, notice that $\partial \hat{S} / \partial \alpha > 0$. We have to study the sign of $\partial \Psi / \partial \alpha$. We have

$$\frac{\partial \Psi}{\partial \alpha} = \hat{S}^2 e^{\gamma \hat{S}} m_{\alpha}(\hat{S}) \tilde{\mu} + \xi x(i) \left[\hat{S} m_{\alpha}(\hat{S}) e^{\gamma \hat{S}} - \int_0^{\hat{S}} e^{\gamma z} m_{\alpha}(z) dz \right],$$

where

$$m_{\alpha}(x) = \frac{\partial m(x)}{\partial \alpha} = \frac{e^{\beta x} - 1}{(\alpha - 1)^2} > 0.$$

The sign is trivial because the term between brackets is necessarily positive as function $e^{\gamma z} m_{\alpha}(z)$ is increasing from 0 to \hat{S} . Therefore, $\partial \Psi / \partial \alpha > 0$, and using (B.3), one can deduce that $\tilde{\mu}(i, k)$ is a decreasing function of α .

When $p = \beta$, one has to sign the derivative $\partial \Psi / \partial \beta$ since we already know by Proposition 2 that $\partial \hat{S} / \partial \beta < 0$. We have

$$\frac{\partial \Psi}{\partial \beta} = \hat{S}^2 e^{\gamma \hat{S}} m_{\beta}(\hat{S}) \tilde{\mu} + \xi x(i) \left[\hat{S} m_{\beta}(\hat{S}) e^{\gamma \hat{S}} - \int_0^{\hat{S}} e^{\gamma z} m_{\beta}(z) dz \right],$$

where

$$m_{\beta}(x) = \frac{\partial m(x)}{\partial \beta} = \frac{-x e^{\beta x}}{\alpha - 1} < 0.$$

And again, the sign is trivial by the same argument as just above: Because function $-e^{\gamma z} m_{\beta}(z)$ is increasing for positive z , it follows that the term between brackets is negative, and so is the derivative $\partial \Psi / \partial \beta$. Hence, (B.3) implies that $\tilde{\mu}(i, k)$ decreases when β is raised.

13. Put $\gamma = 0$ and explicit the integral in the inequality just above, the result is then trivial given that function $m(\hat{S})$ is decreasing: The area comprised between the curve $m(x)$ and the horizontal axis, and the vertical lines $x = 0$ and $x = \hat{S}$, is necessarily bigger than the area of the rectangle of dimensions \hat{S} and $m(\hat{S})$.

And last, for $p = \gamma$, we know that $\partial \hat{S} / \partial \gamma > 0$ by Proposition 2. We have to sign the partial derivative $\partial \Psi^1 / \partial \gamma$, which is equal to

$$\begin{aligned} \frac{\partial \Psi^1}{\partial \gamma} &= \hat{S}^3 e^{\gamma \hat{S}} m(\hat{S}) \tilde{\mu}(i, k) + \xi x(i) \left[\hat{S}^2 m(\hat{S}) e^{\gamma \hat{S}} - \int_0^{\hat{S}} z e^{\gamma z} m(z) dz \right] \\ &\quad + t e^{-\gamma t} \int_0^L w^h m(z) dz, \end{aligned}$$

which can be rewritten using the fact that $\tilde{\mu}(i, k) = \Psi^{-1}(k + e^{-\gamma t} W^h)$ by Proposition 3, as

$$\begin{aligned} \frac{\partial \Psi}{\partial \gamma} &= \hat{S} \left[k + e^{-\gamma t} W^h + \xi x(i) \int_0^{\hat{S}} e^{\gamma z} m(z) dz \right] \\ &\quad - \xi x(i) \int_0^{\hat{S}} z e^{\gamma z} m(z) dz + t e^{-\gamma t} W^h, \\ &= (k + e^{-\gamma t} W^h) \hat{S} + \xi x(i) \int_0^{\hat{S}} (\hat{S} - z) e^{\gamma z} m(z) dz + t e^{-\gamma t} W^h. \end{aligned}$$

It follows that $\partial \Psi / \partial \gamma > 0$, and by (B.3), one concludes that $\tilde{\mu}(i, k)$ is a decreasing function of γ as expected.

Appendix B.7. Properties of $R(\cdot)$ and $B(\cdot)$

Let us first study the properties of function $R(E, k)$. First, it tends asymptotically to 0 when either E or k grows infinitely. Moreover,

$$\lim_{E \rightarrow 0} ER(E, k) = 0, \quad \lim_{k \rightarrow 0} R(E, k) = 2 \int_0^{\frac{1}{2E}} \int_{\underline{\mu}(i, k)}^{\infty} g(\mu) d\mu di.$$

The first-order partial derivatives are

$$\frac{\partial R}{\partial E} = -\frac{1}{E^2} \int_{\underline{\mu}(\frac{1}{2E}, k)}^{\infty} g(\mu) d\mu < 0$$

and

$$\frac{\partial R}{\partial k} = -2 \int_0^{\frac{1}{2E}} \frac{\partial \tilde{\mu}}{\partial k}(i, k) g[\tilde{\mu}(i, k)] di < 0.$$

R is thus a decreasing function of E , which is not surprising as the catchment area of the school shrinks when E grows, while the attendance rate in each interior

location remains the same. It is also a decreasing function of k , which is obvious as the catchment area remains the same but the school attracts fewer children from every location.

Let us turn to the second-order partial derivatives.

$$\frac{\partial^2 R}{\partial E^2} = \frac{2}{E^3} \int_{\tilde{\mu}(\frac{1}{2E}, k)}^{\infty} g(\mu) d\mu - \frac{1}{2E^4} \frac{\partial \tilde{\mu}}{\partial i} \left(\frac{1}{2E}, k \right) g \left[\tilde{\mu} \left(\frac{1}{2E}, k \right) \right].$$

The sign of this expression is ambiguous. To study the behavior of the function, we assimilate E to a continuous variable. The second term dominates for values of E close to 0 and R is concave in a certain neighborhood of the origin. When E grows, the first term dominates and R is convex.

$$\frac{\partial^2 R}{\partial k^2} = -2 \int_0^{\frac{1}{2E}} \left\{ \frac{\partial^2 \tilde{\mu}}{\partial k^2}(i, k) g[\tilde{\mu}(i, k)] + \left[\frac{\partial \tilde{\mu}}{\partial k}(i, k) \right]^2 g'[\tilde{\mu}(i, k)] \right\} di.$$

The sign of this derivative is also ambiguous.

$$\frac{\partial^2 R}{\partial E \partial k} = \frac{1}{E^2} \frac{\partial \tilde{\mu}}{\partial k} \left(\frac{1}{2E}, k \right) g \left[\tilde{\mu} \left(\frac{1}{2E}, k \right) \right] > 0.$$

Only the sign of this cross-derivative is unambiguous.

We now turn to the properties of function $B(E, k)$. First, it is obviously continuous. Second, we have the following limits when either E or k grows infinitely:

$$\begin{aligned} \lim_{E \rightarrow 0} B(E, k) &= 0, & \lim_{k \rightarrow 0} B(E, k) &= -fE \\ \lim_{E \rightarrow \infty} B(E, k) &= -\infty, & \lim_{k \rightarrow \infty} B(E, k) &= -fE. \end{aligned}$$

The first-order partial derivatives are

$$\begin{aligned} \frac{\partial B}{\partial E} &= k\zeta R(E, k) - f + Ek\zeta \frac{\partial R}{\partial E}(E, k) \\ &= k\zeta R(E, k) - f - \frac{k\zeta}{E} \int_{\tilde{\mu}(\frac{1}{2E}, k)}^{\infty} g(\mu) d\mu \\ &= \frac{1}{E} \left[B(E, k) - k\zeta \int_{\tilde{\mu}(\frac{1}{2E}, k)}^{\infty} g(\mu) d\mu \right], \end{aligned} \quad (\text{B.4})$$

and

$$\begin{aligned} \frac{\partial B}{\partial k} &= \zeta ER(E, k) + k\zeta E \frac{\partial R}{\partial k}(E, k) \\ &= 2\zeta E \int_0^{\frac{1}{2E}} \int_{\tilde{\mu}(i, k)}^{\infty} g(\mu) d\mu di - 2k\zeta E \int_0^{\frac{1}{2E}} \frac{\partial \tilde{\mu}}{\partial k}(i, k) g[\tilde{\mu}(i, k)] di. \end{aligned} \quad (\text{B.5})$$

We turn to the second order partial derivatives.

$$\begin{aligned} \frac{\partial^2 B}{\partial E^2} &= 2k\zeta \frac{\partial R}{\partial E}(E, k) + k\zeta E \frac{\partial^2 R}{\partial E^2}(E, k) \\ &= -\frac{k\zeta}{2E^3} \frac{\partial \tilde{\mu}}{\partial i} \left(\frac{1}{2E}, k \right) g \left[\tilde{\mu} \left(\frac{1}{2E}, k \right) \right] < 0. \end{aligned} \tag{B.6}$$

Hence, for any fixed value of k , B is a concave function of E .

$$\begin{aligned} \frac{\partial^2 B}{\partial k^2} &= -4\zeta E \int_0^{\frac{1}{2E}} \frac{\partial \tilde{\mu}}{\partial k}(i, k) g[\tilde{\mu}(i, k)] di \\ &\quad - 2k\zeta E \int_0^{\frac{1}{2E}} \left\{ \frac{\partial^2 \tilde{\mu}}{\partial k^2}(i, k) g[\tilde{\mu}(i, k)] + \left[\frac{\partial \tilde{\mu}}{\partial k}(i, k) \right]^2 g'[\tilde{\mu}(i, k)] \right\} di. \end{aligned} \tag{B.7}$$

Obviously, the sign of this expression is ambiguous.

The envelope theorem can help us to characterize the solution to problem M2. Let $E^* = E^*(k)$ denote the optimal solution to problem M1 at given k . According to (B.4), the first-order condition gives

$$BE(k) = B(E^*, k) = k\zeta \int_{\tilde{\mu}(\frac{1}{2E^*}, k)}^{\infty} g(\mu) d\mu.$$

The derivative with regard to k gives

$$\begin{aligned} \frac{dBE}{dk}(k) &= \zeta \int_{\tilde{\mu}(\frac{1}{2E^*}, k)}^{\infty} g(\mu) d\mu - k\zeta g \left[\tilde{\mu} \left(\frac{1}{2E^*}, k \right) \right] \\ &\quad \times \left[-\frac{\partial \tilde{\mu}}{\partial i} \left(\frac{1}{2E^*}, k \right) \frac{1}{(2E^*)^2} \frac{dE^*}{dk} + \frac{\partial \tilde{\mu}}{\partial k} \left(\frac{1}{2E^*}, k \right) \right]. \end{aligned}$$

The first term is positive, and the coefficient of k in the second term is also always positive. It follows that the function BE is unimodal and, as $BE'(0) > 0$, there a unique interior to problem M2.

Appendix B.8. Proof of Lemma 2

$\tilde{\mu}(i, k)$ is given by the equality $\Psi(\tilde{\mu}(i, k)) = k + e^{-\gamma t} W_0$ with

$$\Psi(\tilde{\mu}(i, k)) = \hat{S}^2 e^{\gamma \hat{S}} m(\hat{S}) \tilde{\mu} + \xi x(i) \left[\hat{S} m(\hat{S}) e^{\gamma \hat{S}} - \int_0^{\hat{S}} e^{\gamma z} m(z) dz \right],$$

and $\hat{S} = \hat{S}(\tilde{\mu}(i, k))$. Therefore, $\Psi(\tilde{\mu}(i, k))$ goes to infinity when k is infinitely large. Given the expression of $\Psi(\mu)$, this can only happen if $\tilde{\mu}(i, k)$ itself goes to infinity, because \hat{S} is bounded from above (by L , for example).

Appendix B.9. Proof of Lemma 3

Denote by $G(\cdot)$ the cumulative density function of the long-normal density $g(\cdot)$, and recall that $\tilde{\mu}(i, k)$ is an increasing function in both arguments and that it is indeed upper- and lower-bounded by linear functions in k . Equation (15) implies the following bounds for the attendance rate $R(E, k)$:

$$0 \leq R(E, k) \leq \frac{1 - G(\tilde{\mu}(0, k))}{E},$$

using the fact the threshold is increasing with distance. It follows that the total amount of tuition fees paid to founded schools $E > 0$, that it is $k\zeta R(E, k)$; E , are bounded by

$$0 \leq kR(E, k); E \leq k; \zeta\{1 - G(\tilde{\mu}(0, k))\}.$$

Because G is the cumulative function of the log-normal law and because the ability thresholds are (upper and lower) bounded by linear functions of k , it is readily shown that the upper-bound $k, \zeta\{1 - G(\tilde{\mu}(0, k))\}$ goes to zero when k goes to infinity. It follows that there exists a $K > 0$, independent of E (and of course of k) so that the profit function is strictly negative for any $k > K$.

Appendix B.10. Proof of Proposition 6

Consider the set Ω given by

$$\Omega = \{(E, k) \in \mathbb{R}^2 : 0 \leq E; \Theta(E) \geq f; 0 \leq k \leq K\}.$$

This set is a compact subset of \mathbb{R}^2 , and because $B(E, k)$ is continuous in (E, k) , it must reach a maximum in Ω . The non-negativity of the profit comes immediately from the definition of K and $\Theta(E)$.

References

- Anastassova, Lubomira (2006). "Productivity Differences and Agglomeration across Districts of Great Britain." CERGE-EI Working Paper WP289, The Center for Economic Research and Graduate Education—Economic Institute, Prague.
- Bairoch, Paul, Jean Batou, and Pierre Chèvre (1988). *The Population of European Cities from 800 to 1850*. CIEH.
- Bar, Michael, and Oksana Leukhina (2005). "A Model of Historical Evolution of Output and Population." Working paper, Federal Reserve Bank of Minneapolis.

- Bartel, Ann, and Frank Lichtenberg (1987). "The Comparative Advantage of Educated Workers in Implementing New Technology." *Review of Economics and Statistics*, 69(1), 1–11.
- Becker, Gary, Edward Glaeser, and Kevin Murphy (1999). "Population and Economic Growth." *American Economic Review*, 89(2), 145–149.
- Bos, Henricus Cornelius (1965). *Spatial Dispersion of Economic Activity*. Rotterdam University Press.
- Boucekkine, Raouf, David de la Croix, and Omar Licandro (2002). "Vintage Human Capital, Demographic Trends and Endogenous Growth." *Journal of Economic Theory*, 104, 340–375.
- Boucekkine, Raouf, David de la Croix, and Omar Licandro (2003). "Early Mortality Declines at the Dawn of Modern Growth." *Scandinavian Journal of Economics*, 105(3), 401–418.
- Ciccone, Antonio (2002). "Agglomeration Effects in Europe." *European Economic Review*, 46(2), 213–227.
- Ciccone, Antonio, and Robert Hall (1996). "Productivity and the Density of Economic Activity." *American Economic Review*, 86(1), 54–70.
- Cipolla, Carlo (1969). *Literacy and Development in the West*. Penguin Books.
- Clark, Gregory (2001). "The Secret History of the Industrial Revolution." Working paper, University of California at Davis.
- Clark, Gregory (2003). "The Great Escape: The Industrial Revolution in Theory and History." Working paper, University of California at Davis.
- Clark, Gregory, and Gillian Hamilton (2004). "Survival of the Fittest: Human Capital and Selection in the Malthusian World." Technical Report, University of California at Davis.
- Cressy, David (1980). *Literacy and the Social Order: Reading and Writing in Tudor and Stuart England*. Cambridge University Press.
- Doepke, Matthias (2004). "Accounting for Fertility Decline During the Transition to Growth." *Journal of Economic Growth*, 9(3), 347–383.
- Epstein, Stephan (2000). *Freedom and Growth: The Rise of States and Markets in Europe, 1300–1750*. Routledge.
- Fridlitzius, Gunnar (1985). "The Mortality Decline in the First Phase of the Demographic Transition: Swedish Experiences." In *Pre-industrial Population Change*, edited by Tommy Bengtsson, Gunnar Fridlitzius, and Rolf Ohlsson, pp. 71–114. Almqvist and Wiksell International.
- Fujita, Masahisa, and Jacques-François Thisse (2002). *Economics of Agglomeration*. Cambridge University Press.
- Galor, Oded (2005). "From Stagnation to Growth: Unified Growth Theory." In *Handbook of Economic Growth*, edited by Philippe Aghion and Steven Durlauf Elsevier, 171–293.
- Galor, Oded, and Omer Moav (2002). "Natural Selection and the Origin of Economic Growth." *Quarterly Journal of Economics*, 117(4), 1133–1192.
- Galor, Oded, and David Weil (2000). "Population, Technology, and Growth: From the Malthusian Regime to the Demographic Transition and Beyond." *American Economic Review*, 90(4), 806–828.
- Hansen, Gary, and Edward Prescott (2002). "Malthus to Solow." *American Economic Review*, 92(4), 1205–1217.
- Hazan, Moshe, and Hosny Zoabi (2005). "Does Longevity Cause Growth." Technical Report 0507001, Economics Working Paper Archive, EconWPA.
- Holmes, Thomas (2005). "The Location of Sales Offices and the Attraction of Cities." *Journal of Political Economy*, 113(3), 551–581.
- Houston, R A (2002). *Literacy in Early Modern Europe*. Pearson Education.
- Kelley, Allan, and Robert Schmidt (1995). "Aggregate Population and Economic Growth Correlations: The Role of the Components of Demographic Changes." *Demography*, 32(4), 543–555.

- Kremer, Michael (1993). "Population Growth and Technological Change: One Million B.C. to 1900." *Quarterly Journal of Economics*, 108(3), 681–716.
- Lagerloef, Nils-Petter (2003). "From Malthus to Modern Growth: Can Epidemics Explain the Three Regimes?" *International Economic Review*, 44(2), 755–777.
- Lagerloef, Nils-Petter (2006). "The Galor–Weil Model Revisited: A Quantitative Exercise." *Review of Economic Dynamics*, 9, 116–142.
- Lucas, Robert (1988). "On the Mechanics of Economic Development." *Journal of Monetary Economics*, 22(1), 3–42.
- Maddison, Angus (2001). *The World Economy: A Millennial Perspective*. OECD, Development Centre Studies.
- Nicolini, Esteban (2004). "Mortality, Interest Rates, Investment, and Agricultural Production in 18th Century England." *Explorations in Economic History*, 41(2), 130–155.
- O'Day, Rosemary (1982). *Education and Society 1500–1800* (Themes in British Social History). Longman.
- Perrcnoud, Alfred (1978). "La mortalité à Genève de 1625 à 1825." *Annales de Démographie Historique*, 15, 209–233.
- Salop, S (1979). "A Model of the Natural Rate of Unemployment." *American Economic Review*, 69, 117–125.
- Sanderson, Michael (1995). *Education, Economic Change and Society in England 1780–1870*. Cambridge University Press.
- Schools Inquiry Commission. 1868a. *Report of the Commissioners*, Volume 1. George E. Eyre and William Spottiswoode.
- Schools Inquiry Commission. 1868b. *Tables*, Volume 21. George E. Eyre and William Spottiswoode.
- van Zanden, Jan Luiten (2004). "The Skill Premium and the Great Divergence." Working paper, IISG/University of Utrecht.
- Wrigley, E. A., R. S. Davies, J. E. Oeppen, and R. S. Schofield (1997). *English Population History from Family Reconstitution: 1580–1837*. Cambridge University Press.
- Wrigley, Edward, and Roger Schofield (1989). *The Population History of England, 1541–1871: A Reconstruction*. Cambridge University Press.