

Early Literacy Achievements, Population Density and the Transition to Modern Growth

Raouf Boucekkine

Dept. of economics & CORE

Univ. cath. Louvain

David de la Croix

Dept. of economics & CORE

Univ. cath. Louvain

Dominique Peeters

Dept. of geography & CORE

Univ. cath. Louvain

This version: March 18, 2005

¹Boucekkine and de la Croix acknowledges 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 P5/10, “Equilibrium theory and optimization for public policy and industry regulation”).

We thank H. d’Albis, M. Doepke, J. Finlay, F. Heylen, J.-O. Hairault, E. Lehman, O. Leukhina, T. Lindh, B. Malmberg, F. Maniquet, M. Perez-Nievas, A. Pommeret, G. Rayp, A. Schaefer, J. Thisse, E. Toulemonde, D. van de Gaer, and seminar participants at IIASA (Vienna), Lausanne, Ghent, Louvain-la-Neuve (CORE), Minneapolis FED, Paris I and Namur Universities, and participants at SED 2004 and T2M 2005 for their comments on an earlier draft. We thank David Cressy for providing useful information on English schools prior to the Industrial Revolution.

Contact addresses: Dept. of Economics, Place Montesquieu 3, B-1348 Louvain-la-Neuve, Belgium.
E-mail addresses: boucekkine@ires.ucl.ac.be, delacroix@ires.ucl.ac.be, peeters@geog.ucl.ac.be.

Abstract

The transition from economic stagnation to sustained growth is often modelled 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, while one sixth is linked to higher longevity and one half to technical progress. 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.

Keywords: Human Capital, Population Density, Education Investment, School Location, Technical Progress.

JEL Classification Numbers: O41, I21, R12, J11.

Introduction

Human capital theory aims at accounting for income growth in terms of the characteristics of the labor force. This theory is useful in addressing many pressing issues, as shown by the large literature on the subject. It is also relevant in an historical perspective to understand the link between education and the take-off of Western-Europe. Indeed, many authors advocate that there is a causality link between the rise in literacy in the pre-industrial era and the process leading to the Industrial Revolution. For example, Cipolla (1969) argues that literacy favored the Industrial Revolution in more than one way. It 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 a comparative advantage because they assimilate new ideas more readily (Bartel and Lichtenberg 1987).

Figure A-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.¹ Over the nineteenth century, however, they also witness a steady rise from 1580 to 1760. 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 a fifth to a 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 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.

Higher education achievements might have been triggered by several economic factors. We distinguish three of them (see Galor (2004) for an exhaustive list). First, technical progress increased labor productivity and wage rates in the modern sector (see Hansen and Prescott (2002) and Bar and Leukhina (2005)), and thereby the return to investment in education (Doepke 2004). Facing better income perspectives in this sector, households would engage in education to benefit from the higher skill premium. However, as illustrated by Figure A-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.

Secondly, another interesting candidate for explaining the rise in literacy is longevity

¹The 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.

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)). Boucekine, 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. This could only be part of the story however, as shown by English data. Figure A-3 presents the survival rate of five years old individuals. It therefore abstracts from infant mortality swings to concentrate on mortality during the active life. 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 of cities. During this period of high mortality, literacy rose continuously, as we shown on Figure A-1. Hence, for England, the mortality channel is of no help to explain the early rise in educational attainments.

A third possibility is drawn from various authors who stress that the rising density of population may have played a role in fostering the take-off. Higher density can lower the cost of education through facilitating the creation of schools (see Fujita and Thisse (2002) for a textbook treatment of this effect and Holmes (2005) for an application on contemporary data). There can also be externalities generated by denser population.² The view that 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)) and across US states (see Ciccone and Hall (1996)). Looking at English historical data, Figure A-4 shows that population³ rose rapidly in the sixteenth and nineteenth centuries, while the seventeenth century was one of demographic stagnation.⁴ The corresponding swings in crude birth rates are plotted in Figure A-5. We notice that rises in population in the sixteenth century correspond to the first wave of improvement in literacy.

In the papers with “population-induced” productivity growth, the effect of population on productivity is assumed instead of being derived from more primary assumptions. In this paper we propose a new framework of literacy and growth where the effect of population on growth is derived from maximizing behavior. By looking at the linkages between literacy, school establishment and income growth, we will be able to evaluate the exact role of population density. In our model, the length of schooling is chosen by individuals who maximize life-time income, which depends on future wages, longevity, and the distance to the nearest school. Then, the num-

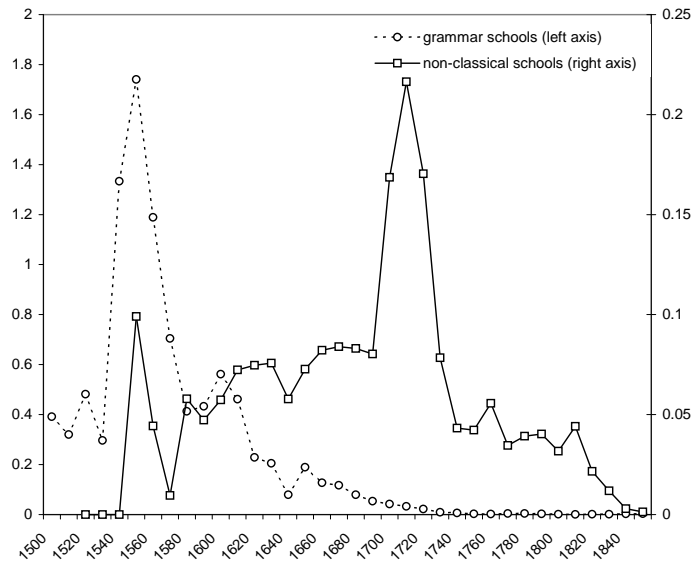
²Becker, Glaeser, and Murphy (1999) argue that larger populations encouraged greater specialization and increased investments in knowledge. For Kremer (1993), high population spurs technological change. For Galor and Weil (2000) and Lagerloef (2003), there is a “population-induced” technical progress which raised the return to human capital. Population needs to reach a threshold for productivity to take-off.

³We represent the population aged 6 and more, because it coincides with the concept of population of our model, disregarding the infants aged 0-5.

⁴Population growth in England was spectacular compared to other big European countries. According to Maddison (2001) numbers, English population was multiplied by 2.17 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.

ber and location of education facilities is determined, either chosen optimally by the State or following a free entry process. Higher population density makes it optimal to increase school density, opening the possibility to reach higher educational levels.

Figure 1: School creation rate



Source: Own estimation from Appendix B.

This effect of population on the density of schools is consistent with the available evidence that we have gathered from the reports of the British Parliamentary Papers (Schools Inquiry Commission (1868a) and Schools Inquiry Commission (1968b)) presented in Appendix B. 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.

The paper is organized as follows. The model economy is introduced in Section 1. In Section 2 we study the problem of the individual who has to choose the length of her studies. Section 3 is devoted to the problem of the school authority, who decides the number and the localization of schools. We also consider other types of school establishments, including free-entry and price making schools. A quantitative exercise is proposed in Section 4 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 5 contains robustness analyzes with respect to important parameters and assumptions on school institutions. Section 6 concludes.

1 The Model Economy

The model mixes aspects from growth theory with elements from the economic geography literature.

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 . We assume that abilities are distributed according to a probability density function $g(\mu)$ over the interval $[0, \bar{\mu}]$, where $\bar{\mu}$ is a strictly positive number (possibly infinite). This distribution is the same at every location. Moreover, we assume that g is unimodal and has the usual regularity properties. 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. Since, 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).

Space is modelled as a circle of unit length. 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. 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, since 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.

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: $x(i) \in [0, \frac{1}{2}] \cup \{+\infty\}$. 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 $x(i)$.

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

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, i.e., 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 (Blanchard 1985). 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 .

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:

$$Y_t = A_t H_t \quad \text{where} \quad A_t = e^{\gamma t}. \quad (4)$$

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 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.

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.

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 a 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 peoples willingness to pay across different social classes.

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

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

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 is a fixed cost repre-

senting 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 as the tuition fee. To ensure the existence of a balanced growth path, 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

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

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:

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

School Location and Policy

At each date a number E_t of classrooms is created to serve the newborn generation. These classrooms are assumed to be 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 have usually 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 four 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 the 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. In **M1**, we assume that the tuition fee k is exogenous. Each school possesses a catchment area. 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 and the tuition fee, let $R(E_t, k)$. The benefit drawn from the school built in a given area is equal to $kA_t\zeta_t R(E_t, k)$. The maximization problem can be written as:

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

Model M2: Another scenario supposes that 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:

$$\max_{E_t, k_t} A_t (k_t\zeta_t R(E_t, k_t) - f) E_t \quad (9)$$

Model M3: 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. In scenario **M3**, we suppose that the tuition cost k is exogenous. The equilibrium solution results from the free entry condition:

$$\text{Find } E_t \text{ such that } k\zeta_t R(E_t, k) - f = 0 \quad (10)$$

This model **M3** is equivalent to a model where a central authority maximizes aggregate attendance (for example, for religious reasons). In that case, it would create as

many schools as possible, subject to a non-negative profit condition, which would lead to equation (10).

Model M4: In the last scenario **M4**, 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 above. The problem can then be stated as:

$$\text{Find } (E_t, k_t) \text{ such that } E_t \text{ solves } k_t \zeta_t R(E_t, k_t) - f = 0 \text{ for a given } k_t \quad (11)$$

$$\text{and } k_t \text{ is the maximand of } k_t \zeta_t R(E_t, k_t) - f \text{ for a given } E_t. \quad (12)$$

Equilibrium

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

- A path of optimal education decision $\{S_t(\mu, i)\}_{t \geq 0}$ maximizing life-time resources (5), subject to (7);
- 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), (9), (10) or (11)-(12).

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.

2 Solution to the Individual's Problem

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.

Existence and Uniqueness of the Interior Solution

The first-order necessary condition is:

$$\mu \int_{t+S_t(\mu,i)}^{t+L_t} A_z m_t(z-t) dz = m_t(S_t(\mu,i)) A_{t+S_t(\mu,i)} (\mu S_t(\mu,i) + \xi x(i)). \quad (13)$$

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 since 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 an obvious variable change) can be re-written as:

$$\mu \int_{S(\mu,i)}^L e^{\gamma a} m(a) da = (\mu S(\mu,i) + \xi x(i)) e^{\gamma S(\mu,i)} m(S(\mu,i)). \quad (14)$$

Observe that L does check the integral equation because of $m(L) = 0$. However, since 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 (14). 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 (14) such that $0 < S(\mu,i) < L$ if and only if $\mu > \underline{\mu}(i)$, with*

$$\underline{\mu}(i) = \frac{\xi x(i)}{\int_0^L e^{\gamma a} m(a) da}. \quad (15)$$

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

Proof: see Appendix C. ■

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

Corollary 1 *The threshold $\underline{\mu}(i)$ is an increasing function of ξ , $x(i)$ and β . It is decreasing in α and γ .*

Proof: see Appendix D. ■

The main implication of the Corollary 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. These basic mechanisms are also clear in the comparative statics of the interior solution.

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 $x(i) = 0$.*

Proof: see Appendix E. ■

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 (27) to replace the integral appearing in the left hand side of the inequality, and rearranging terms, one gets the following condition:

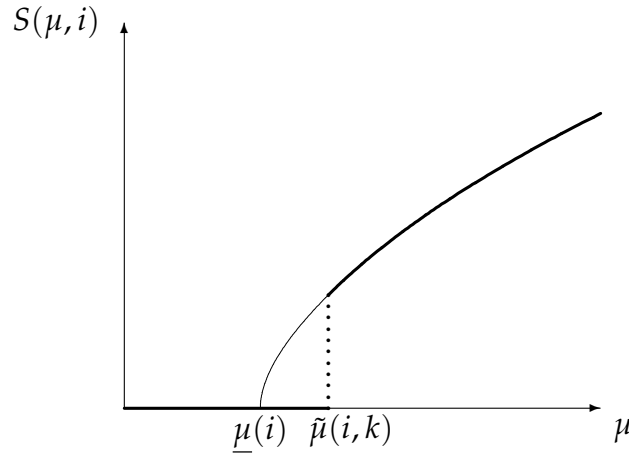
$$\hat{S}^2 e^{\gamma \hat{S}} m(\hat{S}) \mu + \xi x(i) \left[\hat{S} m(\hat{S}) e^{\gamma \hat{S}} - \int_0^{\hat{S}} e^{\gamma z} m(z) dz \right] > k + e^{-\gamma t} W^h.$$

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

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

The following optimality proposition holds.

Figure 2: Schooling and ability



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

$$\tilde{\mu}(i, k) = \Psi^{-1} \left(k + e^{-\gamma t} W^h \right), \quad (16)$$

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.

Proof: see Appendix F. ■

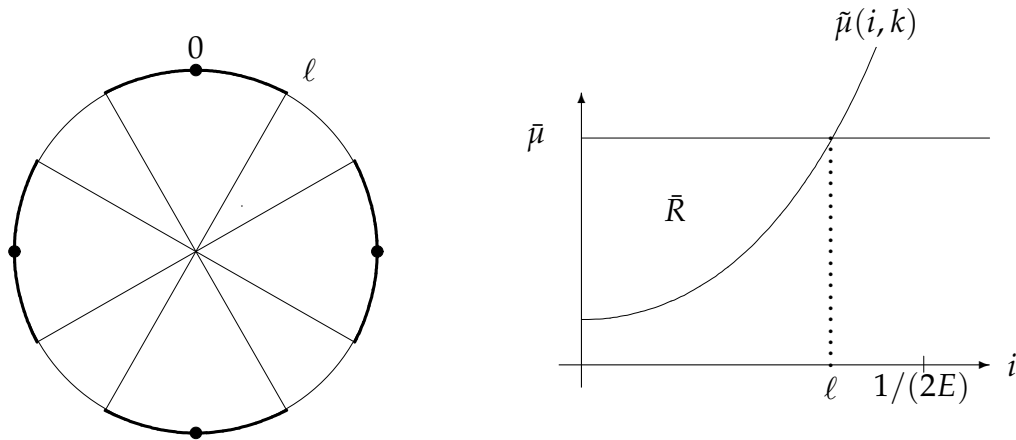
Figure 2 shows the respective role of $\underline{\mu}(i)$ and $\tilde{\mu}(i, k)$. The interior solution for schooling is positive for $\mu > \underline{\mu}(i)$ but is still dominated by the corner solution until the threshold $\tilde{\mu}(i, k)$ is reached. The optimal solution corresponds to the heavy line.

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 .

Proof: see Appendix G. ■

Figure 3: Separated catchment areas



3 Solution to the School Location Problem

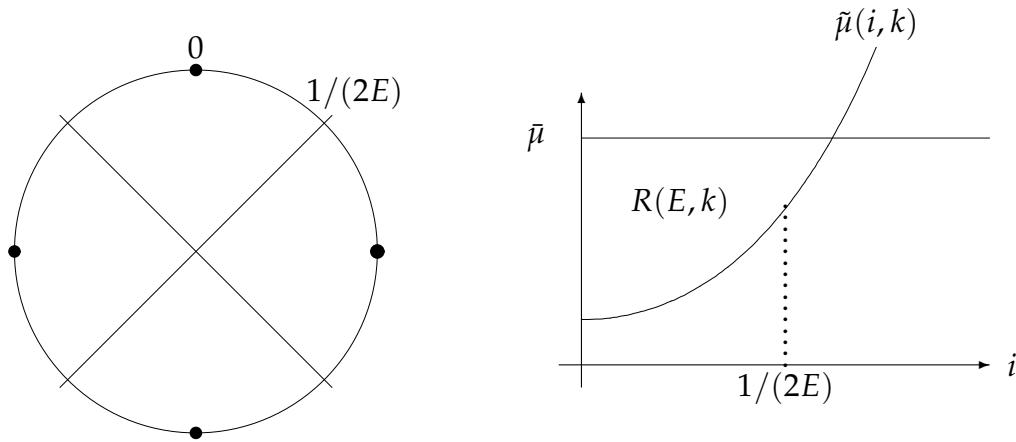
We now consider the choice of the density of schools E_t , denoted without time subscript in this section. The framework is inspired by Bos (1965), but our reaction function $\tilde{\mu}(i, k)$ is much more involved than in the original setting. Given our hypothesis on the dispersion of the population, 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 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 the arc length between location i and the closest school, hence in the catchment area of 0, $x(i) = |i|$. 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, while for very distant populations, only the most skilled do. Let us denote the attendance rate of population located at i by $r(i, k)$, where the index k will be useful later on. It is given by

$$\begin{aligned} r(i, k) &= \int_{\tilde{\mu}(i, k)}^{\bar{\mu}} g(\mu) d\mu && \text{if } \tilde{\mu}(i, k) \leq \bar{\mu} \\ &= 0 && \text{otherwise} \end{aligned}$$

If we multiply this quantity by ζ , we obtain the density of pupils at i . Furthermore,

Figure 4: Contiguous catchment areas



multiplying the latter by the tuition fee k gives the benefit accruing from the pupils at i :

$$b(i, k) = k\zeta r(i, k) \quad (17)$$

The question is now to determine the number of schools that will be created at time t .

Model M1

We assume the existence of a central authority which determines the optimal density of schools so as to maximize the profit of the whole schooling system. Two cases may occur. In the first one, illustrated in Figure 3, the effective catchment areas (i.e., the areas with a strictly positive demand) of the schools do not touch each other. This occurs when $\tilde{\mu}(\ell, k) = \bar{\mu}$ for some point ℓ such that $\ell \leq 1/2E$. By convention, we set $\ell = 1/2$ when $\bar{\mu} \geq \tilde{\mu}(1/2, k)$, i.e., a single school is able attract pupils from every location. The effective catchment area is thus the circular segment $[-\ell, \ell]$. The proportion of children attending school within the effective catchment area is then

$$\bar{R} = 2 \int_0^\ell r(i, k) di.$$

Notice that, in this case where catchment areas do not touch each other, \bar{R} is independent of the density of schools. The benefit drawn from a school built in this area is equal to the tuition fee k times the attendance:

$$k\zeta\bar{R}$$

The school can be sustained provided that the benefit outweighs the set-up cost f , hence:

$$k\zeta\bar{R} - f \geq 0 \quad (18)$$

The global profit when implementing E schools is

$$B(E, k) = [k\zeta\bar{R} - f] E \quad (19)$$

It follows from the previous discussion that if the term between brackets is negative, no schools will be built. If it is positive, the central authority has an incentive to create as many schools as possible. Now, it is obvious that the potential catchment areas of the schools shrink as E grows. This implies there is some minimal value \tilde{E} beyond which these areas come to contact:

$$\tilde{E} = \left\lfloor \frac{1}{2\ell} \right\rfloor$$

To simplify the presentation, we will assume that the number of schools is large enough, so that \tilde{E} can be assimilated to a continuous variable. Hence

$$\tilde{E} = \frac{1}{2\ell}$$

We now consider the second case where the potential catchment areas are in contact, i.e., where $E \geq \tilde{E}$. It is illustrated in Figure 4. The attendance rate in the catchment area of the school at 0 is

$$R(E, k) = 2 \int_0^{\frac{1}{2E}} r(i, k) di = 2 \int_0^{\frac{1}{2E}} \int_{\bar{\mu}(i, k)}^{\bar{\mu}} g(\mu) d\mu di \quad (20)$$

It can be easily checked that R is 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 location remains the same.

The profit function of the decision maker 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 E schools:

$$B(E, k) = [k\zeta R(E, k) - f] E \quad (21)$$

We now study the properties of the function B .

Lemma 1 *At given k , the benefit function $B(E, k)$ is a continuous concave function of E .*

Proof: see Appendix H. ■

It follows from the previous discussion that some schools will be opened provided that condition (18) is met. As stated in Appendix H, the partial derivative of B with respect to E is continuous at \tilde{E} provided that $\ell < 1/2$. Proposition 5 follows:

Proposition 5 1. If $\zeta < f/(\bar{R}k)$, then no schools are created.

2. If $\zeta = f/(\bar{R}k)$, then $\tilde{E} = 1/(2\ell)$ is the optimal density of schools, with ℓ such that $\tilde{\mu}(\ell, k) = \bar{\mu}$ (or $\ell = 1/2$ when $\tilde{\mu}(1/2, k) \geq \bar{\mu}$).
3. If $\zeta > f/(\bar{R}k)$ and $\ell < 1/2$, then E^* determined by the first order condition $B'(E^*, k) = 0$ is the optimal density of schools.
4. If $\zeta > f/(\bar{R}k)$, $\ell = 1/2$, and $B(\tilde{E}, k) \leq k\zeta \int_{\tilde{\mu}(\ell, k)}^{\bar{\mu}} g(\mu) d\mu$, then \tilde{E} is the optimal density of schools.
5. If $\zeta > f/(\bar{R}k)$, $\ell = 1/2$, and $B(\tilde{E}, k) > k\zeta \int_{\tilde{\mu}(\ell, k)}^{\bar{\mu}} g(\mu) d\mu$, then E^* determined by the first order condition $B'(E^*, k) = 0$, is the optimal density of schools.

Proposition 5 shows that, when the newborn population is low, set-up costs are not covered, hence no schools are created. Once the population reaches the threshold $f/(\bar{R}k)$, \tilde{E} schools are created at once. The process by which illiteracy is eliminated is initiated by a jump.

The next step is to examine the influence of the exogenous parameters on the density of schools. First, let us notice that (30) allows one to rewrite B' as follows:

$$\begin{aligned} B'(E, k) &= 2 \int_0^{\frac{1}{2E}} \left[b(i, k) - b\left(\frac{1}{2E}, k\right) \right] di - f \\ &= 2k\zeta \int_0^{\frac{1}{2E}} \int_{\tilde{\mu}(i, k)}^{\tilde{\mu}(\frac{1}{2E}, k)} g(\mu) d\mu di - f \end{aligned}$$

B' is a decreasing function of f , hence $dE^*/df < 0$. Also B' is an increasing function of ζ and thus $dE^*/d\zeta > 0$. The effect of the other parameters on E^* depends on the density function $g(\mu)$, and it is impossible to have general results. Let p denote any other parameters of the model. We have:

$$\frac{\partial B'}{\partial p}(E, k) = 2 \int_0^{\frac{1}{2E}} \left[\frac{\partial b}{\partial p}(i, k) - \frac{\partial b}{\partial p}\left(\frac{1}{2E}, k\right) \right] di \quad (22)$$

It then follows:

1. If $\frac{\partial^2 b}{\partial p \partial x} = 0$, then $dE^*/dp = 0$.
2. If $\frac{\partial^2 b}{\partial p \partial x} > 0$, then $dE^*/dp < 0$.
3. If $\frac{\partial^2 b}{\partial p \partial x} < 0$, then $dE^*/dp > 0$.

Model M2

In the alternative scenario **M2**, the central authority determines the optimal number of facilities E_t and tuition fee k_t in order to maximize the profit of the whole school system. Again here we drop the time index for the sake of clarify. The maximization problem can be expressed as follows

$$\max_{E,k} (k\zeta R(E,k) - f)E$$

We shall study briefly the properties of this problem. The following lemma will prove most helpful.

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

Proof: see Appendix I. ■

Let us now have a look at our optimization problem for a given $E > 0$. Notice that the profit $B(E,k) = (k\zeta R(E,k) - f)E$ is equal to $-fE$ for $k = 0$. By Lemma 2, we know that there exists a k -value, say K , which is independent of E and f , such that $\tilde{\mu}(i,k) \geq \bar{\mu}$. Hence, for $k \geq K$, $R(E,k) = 0$ and $B(E,k) = -fE$. 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

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

This maximum is also a global maximum since $B(E,k) = B(E,0) = -fE, \forall k > K$. Denote by \tilde{k} , the corresponding maximizer. Since 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:

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

Since $\tilde{k}(E)$ 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 since $\tilde{k}(E) < K$. 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 M2 has a unique and global optimal solution for (E, k) . The maximized profit is necessarily non-negative.*

Proof: see Appendix J. ■

Model M3

In the scenario **M3**, we consider that the tuition fee k is fixed, but we assume a free-entry process in which school creation decisions are totally decentralized. This yields an equilibrium density of schools such that the benefit gained from opening a school matches the cost it raises. This number, denoted E^\diamond , is such that

$$k\zeta R(E^\diamond, k) - f = 0 \quad (23)$$

It follows from Appendix H (Equation (30)) and from $B(E^\diamond, k) = 0$ that $B'(E^\diamond) < 0$, hence the density of schools with free entry is larger than the density chosen by a central authority.

Establishing the impact of the exogenous parameters on E^\diamond is much easier than it was for E^* . First, it is obvious that E^\diamond is a decreasing function of f and an increasing function of ζ . Next, let p denote any other parameter besides k . We have:

$$\frac{dB}{dp} = -k\zeta \int_0^{\frac{1}{2E}} \frac{d\tilde{\mu}}{dp}(i, k) g(\tilde{\mu}(i, k)) di$$

It then follows from Proposition 3 that E_t^\diamond is a strictly decreasing function of w^h , ξ , and β and a strictly increasing function α and γ . Finally, note that

$$\frac{dB}{dk} = \frac{B(E, k)}{k} - k\zeta \int_0^{\frac{1}{2E}} \frac{d\tilde{\mu}}{dk}(i, k) g(\tilde{\mu}(i, k)) di$$

It can easily be established that E_t^\diamond is a strictly decreasing function of k .

Model M4

The third scenario M4 mixes the two preceding cases of free entry process and optimal tuition fee. The problem can be stated as a two-stage procedure. Suppose that E

schools have been created. Next, each school chooses k to maximize its profit. This leads to an optimal tuition function of school density $\tilde{k}(E)$, the same as in model M2. 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 :

$$\max E \text{ subject to: } B(E, \tilde{k}(E)) \geq 0 \text{ and } B(E, \tilde{k}(E)) \geq B(E, k) \forall k > 0 \quad (24)$$

4 Baseline Scenario

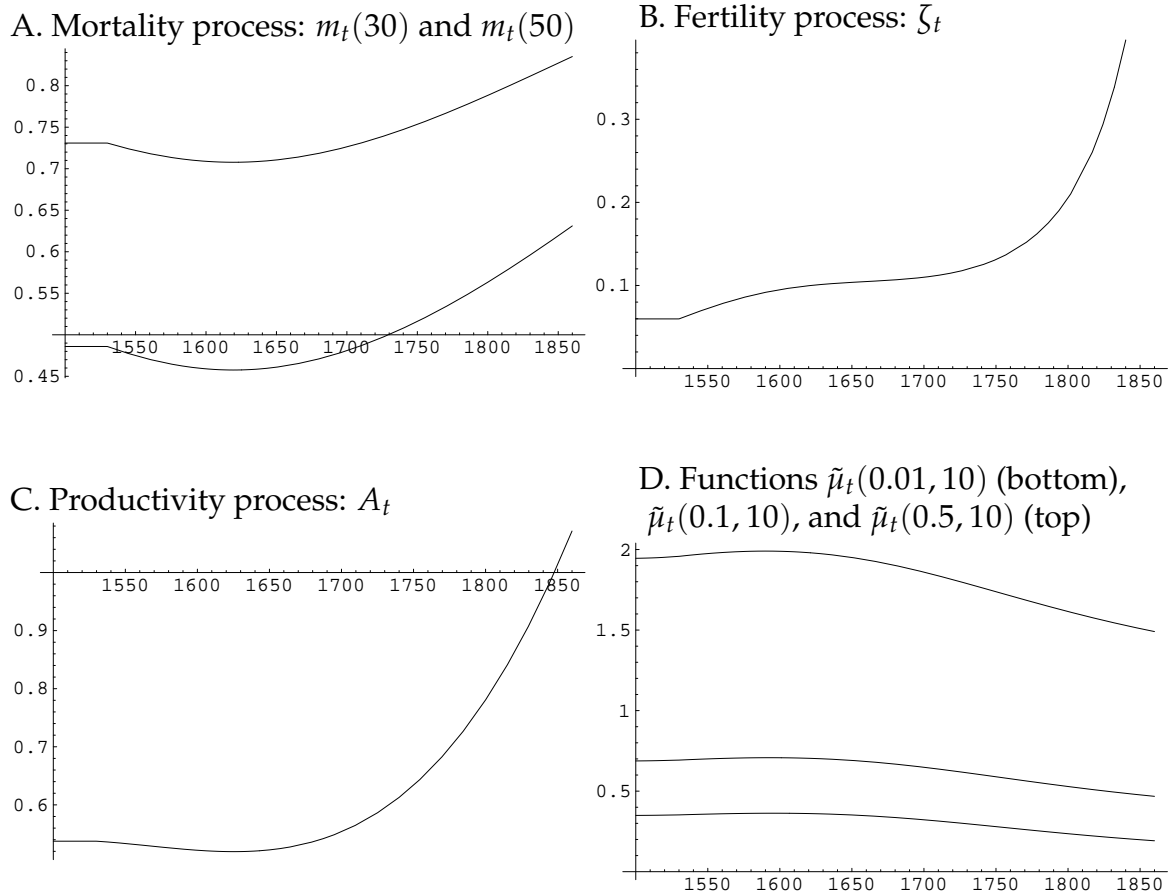
In this section, we are interested in the ability of our model to capture some of the early increase in literacy. School creation follows model **M1**. A robustness analysis to this assumption will be carried over in section 5. We consider the period 1530-1860, which corresponds to the period considered by Wrigley and Schofield (1989) (followed by Wrigley, Davies, Oeppen, and Schofield (1997)) for which they build a consistent 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.

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 distance with the survival functions estimated by Wrigley, Davies, Oeppen, and Schofield (1997). These survival functions applies to the age 5-85, and have been accordingly normalized to 1 at age 5, abstracting from infant mortality. The parameters of the process for ζ_t are chosen so that the 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 A-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). Panel A of Figure 5 presents the evolution of survival probabilities as implied by the calibrated α_t and β_t . All variables are assumed constant before 1530. The survival probabilities first drop, then increase. The level of 1530 is reached and surpassed in 1720. Panel B of Figure 5 represents the process ζ_t needed to generate the adequate level of population. We see that ζ_t increases in two distinct periods: 1530-1600 and 1730-. We use the label "birth density" for the process ζ_t because the main driving force behind consists in

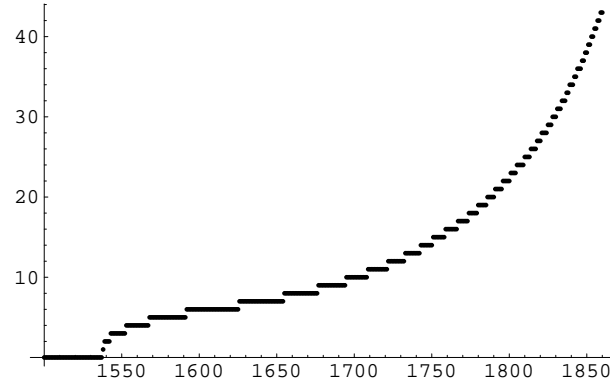
Figure 5: Exogenous processes



fluctuations of crude birth rate (net of infant mortality). The value of ζ_t should be interpreted as a density, since it depends on the normalization of the circle. If we let the circle shrink, this would have the same effect as increasing ζ_t . Finally, Panel C of Figure 5 represents the estimated productivity process A_t ; it is relatively constant until 1700, with a marked acceleration thereafter.

In a second step, we need to select a distribution function for abilities, $g(\mu)$, and choose the five remaining parameters: θ , k , ξ , f , and w^h . Since we have little information to calibrate those parameters, we chose values which give a reasonable baseline scenario, and carry out extensive sensitivity analysis in the next section. For $g(\mu)$ we have selected a log-normal distribution, which is commonly used to approach the actual distribution of abilities. The parameters of the lognormal are set to $-3/2$ and $1/2$ respectively. For the risk free interest rate, we choose $\theta = 0.05$ which lies in the range provided in Epstein (2000). The wage in the home production sector w^h is set to 0.6, which will deliver a reasonable literacy rate at the end of the sixteenth

Figure 6: Baseline simulation - density of schools E_t



century. The transportation cost $\bar{\xi}$ is equal to 40. This value has no precise meaning *per se*, since distances are not measured in actual terms. The set-up cost parameter f and the tuition fee parameter are determined jointly. The set-up cost f is chosen so as to meet the equality condition of Proposition 5 for the year 1539. Doing this, schools will be created from 1540 onwards. The tuition fee is set to correspond to the profit maximizing level of the school system in 1540. This yields $f = 0.0133$ and $k = 11.0256$.

Baseline Simulation

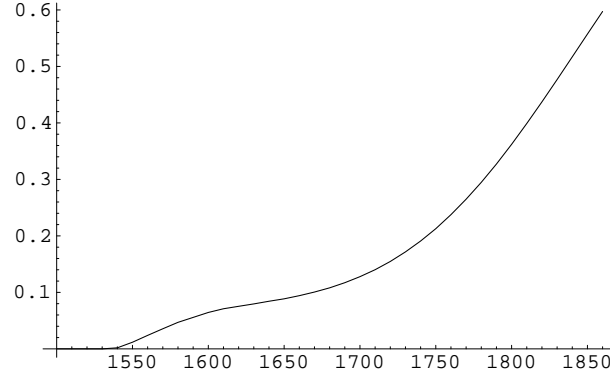
Panel D of Figure 5 reports the function $\tilde{\mu}_t(i, k)$ for three different values of the distance, $i = 1/100$ (bottom curve), $i = 1/10$ and $i = 1/2$ (top curve). The drop in the curves reflects higher incentives to go to schools, either because life expectations are increasing, and/or higher technical progress makes wages in the modern sector more attractive. The fact that the two curves are closer to each other at the end of the period means that distance matters less at this stage of development.

Figure 6 reports the optimal density of schools, such as it results from the maximizing process of the central school authority. In 1540 the optimal number goes from zero to 2 (\tilde{E} of Proposition 5) and then 3 in 1550. Thereafter, the density of schools increases monotonically, with a marked slowdown in the seventeenth century followed by an acceleration in the eighteenth century.

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

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

Figure 7: Baseline simulation - literacy rate - total population



The results are reported in Figure 7. 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.

The level of literacy is globally consistent with the data provided in the introduction (Figure A-1) although a precise mapping would be meaningless since literacy data cover the ability to sign on marriage register, not school attendance. Notice also the role of expectations: the sharp acceleration in the end of the period is related to the anticipation by households of strong productivity gains in the modern sector in the nineteenth century.

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:

$$H_t = \int_{t-\bar{L}}^t \zeta_z m_z(t-z) \int_0^1 \int_{\bar{\mu}_z(i,k)}^{\bar{\mu}} \delta[t-z-S_z(\mu,i)] h_z(\mu,i) g(\mu) d\mu di dz.$$

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

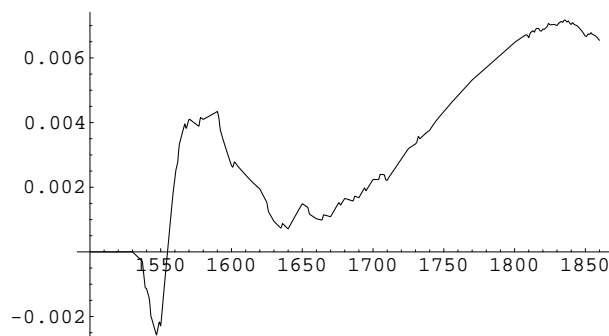
$$\Xi_t = -\xi \int_{t-\bar{L}}^t \zeta_z m_z(t-z) \int_0^1 \int_{\bar{\mu}_z(i,k)}^{\bar{\mu}} \delta[S_z(\mu,i) - (t-z)] x(i) g(\mu) d\mu di dz.$$

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

$$I_t = w^h(1 - \Lambda_t)P_t + A_t(H_t - \Xi_t - kE_t). \quad (25)$$

Figure 8 plots the growth rate of I_t/P_t over 10 years. It is by definition equal to 0 before the creation of the first school. Then, there is a period of negative growth after

Figure 8: Baseline simulation - growth of income per capita



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 very 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. With the displayed growth rates, GDP per capita is multiplied by 3.15 from 1500 to 1860 (it goes from 0.6 in 1500 to 1.89 in 1860). Maddison (2001) estimates the UK GDP per capita as being equal to 714 in 1500 and 2817 in 1860, which delivers a factor of 3.94. The difference between the two can be interpreted as the role played by physical capital accumulation.

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 1.

In the first simulation we have considered that birth density and technical progress are constant over the period; this allows to isolate the role of mortality. Since mortality drops very late in England (see other data on Geneva and Venice in Boucekine, de la Croix, and Licandro (2003)), it does not exert positive influence before the eighteenth century. If mortality improvements were the only driving force of the industrial revolution, the first school should have been created in 1730 and the literacy rate in 1850

Table 1: Counterfactual analysis - importance of each factor

	First school created	School creation	Change in Literacy	Average growth
Mortality	1730	13.2%	15.4%	12.2%
Birth density	1540	65.8%	31.5%	20.7%
Technical progress	1650	28.9%	59.7%	77.6%
Interaction terms		-7.9%	-6.6%	-10.6%

would have been 8.59%. Compared to the baseline simulation, mortality improvements explain 13.2% (5/38) of total school creations over the period 1500-1850, 15.4% of improvements in literacy and 12.2% of growth of income per capita.

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. In the seventeenth century, however, population stagnates, and school creation drops to zero. In the end, the rise in birth density explains 65.8% (25/38) of total school creations over the period 1500-1850, 31.5% of improvements in literacy and 20.7% of income growth per capita.

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 the first school is created in 1650. In the end, technical progress explains 28.9% (11/38) of total school creations over the period 1500-1850, 59.7% of improvements in literacy and 77.6% of the growth of income per capita.

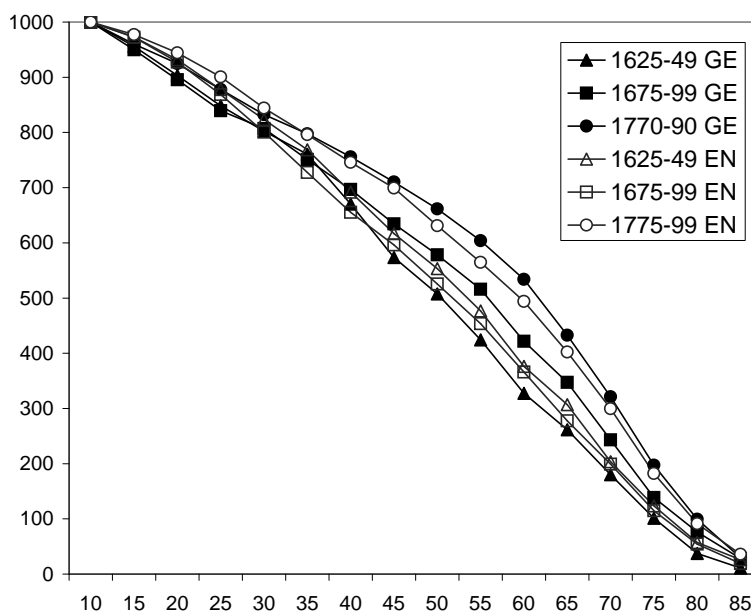
From the counterfactual analysis we conclude that neither productivity increases nor mortality improvements can explain the establishment of schools at a high rate in the sixteenth century documented in Table A-1 and Figure 1. Only the rise in population density can.

Further Discussion of the Results

The above 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 for only one sixth 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

Figure 9: Shifts in the survival function



Source: Geneva: Perrenoud (1978). England: Wrigley, Davies, Oeppen, and Schofield (1997).

Europe as a whole. Looking first at the raw data, Figure 9 compares the survival function in England as estimated by Wrigley, Davies, Oeppen, and Schofield (1997) to the one in Geneva (Perrenoud 1978). We have normalized the series to 1000 at the age of 10; doing so, we concentrate on the mortality of adults. 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.

During the early seventeenth century (data for 1625-49), surviving was more likely in England than in Geneva, specially for the ages 35-65. English longevity was remarkably high for this period. Fifty years later, mortality in Geneva has dropped substantially, becoming lower compared to England. More surprisingly, English mortality has in fact increased over this period. 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 thousands in 1500 to 948 thousands in 1800, while it grew only from 12 thousands to 25 thousands 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.

This view is in line with standard neo-classical growth theory starting with Solow (1956). 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 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:

$$A_t = e^{(1-\rho)\gamma t} H_t^\rho, \quad (26)$$

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

5 Robustness Analysis

Robustness to the Choice of Parameters

We now provide a comprehensive robustness analysis to changes in the parameters. For each simulation, we recalibrate the fixed cost f such that the first school is created in 1540.

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 have run a simulation where the transportation cost is kept not indexed; as a consequence, the relative importance of this cost will diminish in the eighteenth century and the rise in literacy should be more important. Quantitative results are presented in Table 2. 28 schools are created over the period 1500-1850, instead of 38 in the baseline simulation. Since transport is cheaper it is less useful to install many schools to attract pupils. The literacy rate increases by 57.97% instead of 55.73%. Income per capita reaches 1.84 in 1850 instead of reaching 1.77. This small difference reflects that lower transportation costs imply a smaller loss for the economy (the transport sector is not modelled and it is not part of value added). Hence, when the transportation cost is modified, the density of schools

Table 2: Robustness analysis - results for 1500-1850

	School creation	Change in Literacy	GDP per cap. in 1850
Baseline	38	55.73%	1.77
No indexation of transportation costs	28	57.97%	1.84
Risk free interest rate 3%	17	79.46%	1.96
Lower variance $g(\mu)$	80	66.19%	1.71
Higher home productivity w^h	46	48.62%	1.77
w^h indexed on A_t	30	29.57%	1.77

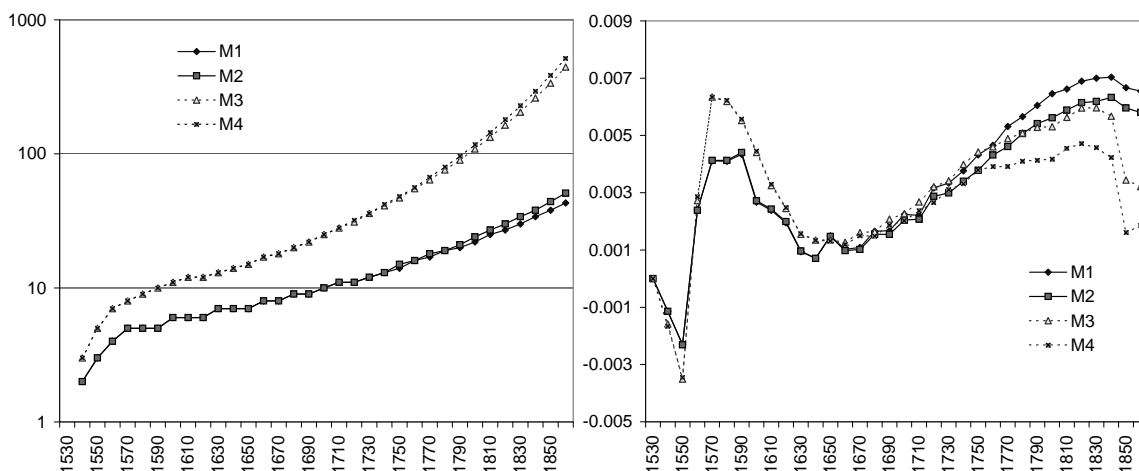
adjusts, and both literacy and growth are almost unaffected. This simulation shows that our previous results are robust to the choice of ξ .

We next analyze the robustness of the result to a change in the risk free interest rate θ . It was set to 5% in the baseline; we now assume a value of 3%. This should increase the incentive to be educated. Indeed, we observe from Table 2 that both literacy and growth reach higher levels. This improvement in performance is obtained with fewer schools; given the higher demand for education, authorities have an incentive to reduce school density compared to the baseline.

Lowering the variance parameter of the distribution $g(\mu)$ (1/3 instead of 1/2) has the opposite effect. There are fewer high skilled people ready to walk long distances to go to school, and consequently it is optimal to increase the density of schools compared to the baseline (from 38 to 80). The results on literacy and growth are slightly affected, with a positive impact on literacy and a negative impact on growth.

Until now, the robustness analysis indicates that the result on literacy and growth are little affected by changes in the parameters, essentially because the density of schools adjust endogenously. This conclusion is however not valid when the parameter w^h is concerned. We have first increased w^h by 25% from 0.6 to 0.75, making the modern sector less attractive. The density of schools adjusts slightly upwards. This has a strong negative influence on literacy, and almost no effect on GDP per capita in 1850. Since in this simulation we start with a higher GDP per capita (0.75 instead of 0.60), growth is slowed down. Secondly, we have indexed w^h on productivity A_t . Again, this prevents literacy from rising as much as in the baseline simulation. 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 explain about half the observed rise in literacy.

Figure 10: School density (log) and growth with models M1-M4



Robustness to the Assumptions on School Establishment

The simulations result for the four models are presented in Figure 10 for both the school density (in log) and growth of GDP per capita. There is an important difference between the models with a central authority (M1 and M2) and those with a free entry process (M3 and M4). Assuming a constant tuition fee k or assuming that the tuition fee is chosen so as to maximize profits does not matter much for school density and growth. In model M2, the optimal tuition fee k is about constant ($= 11$) over the period 1540-1650, then rises steadily to reach 17.7 in 1860 (18.2 with free entry). This rise in the tuition fees leads to a slightly higher density of schools and a slightly lower growth rate at the end of the period. Notice also that the timing of the take-off for school creation does not vary across models; for the same set-up cost parameter in the four cases, the first schools are created in 1540, at the point where schools make zero profit.

With free entry, the density of schools increases much faster than with a central authority. Indeed, the evolution of school density follows an exponential pattern (linear in the log-scale of the top panel of Figure 10) with M1-M2 (central authority) while it follows a double exponential with M3-M4. This very fast rise entails important fixed costs for the economy, which slow growth compared to the central authority case (bottom panel of Figure 10). 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.

6 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 (Hansen and Prescott 2002). (B) Lower mortality increased the return to education and induced higher investment in human capital (see Boucekkine, de la Croix, and Licandro (2003) and Nicolini (2004)). (C) Rising density of population induced economic gains which in turn raised the incentive to educate (Galor and Weil 2000).

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, while one sixth is linked to higher longevity and one half to exogenous total factor productivity growth. 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 1867-8. These data show a peak in school creation at the end of the sixteenth century. Our counterfactual analyzes conclude that neither productivity increases nor mortality improvements can explain this peak. But the rise in population density can.

We have finally carried out extensive sensitivity analysis of the results, both to the parameters of the model and to the assumptions made on school establishment. When we modify the value of the parameters, school density bears the major part of the adjustment, while the pattern of both literacy and growth is less sensitive. This robustness result holds for all parameters for which we have little information except for the productivity of the traditional sector relative to the modern one. We have also analyzed different models of school establishment, distinguishing centralized decisions from free entry processes, and exogenous versus endogenous tuition fee. Simulations show that the model where schools are created through free entry is not as good as a model with a central authority at reproducing the acceleration in growth during the eighteenth century.

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.

References

- Bairoch, Paul, Jean Batou, and Pierre Chèvre. 1988. *The population of European Cities from 800 to 1850*. Genève: CIEH.
- Bar, Michael and Oksana Leukhina. 2005. "A model of historical evolution of output and population." 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.
- Blanchard, Olivier. 1985. "Debts, deficits and finite horizon." *Journal of Political Economy* 93 (2): 223–247.
- Bos, Henricus Cornelius. 1965. *Spatial dispersion of economic activity*. Rotterdam: 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.
- . 2003. "Early mortality declines at the dawn of modern growth." *Scandinavian Journal of Economics* 105 (3): 401–418.
- 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*. Baltimore: Penguin Books.
- Clark, Gregory. 2001. "The secret history of the Industrial Revolution." UC Davis.
- . 2003. "The great escape: the industrial revolution in theory and history." UC Davis.
- Clark, Gregory and Gillian Hamilton. 2004. "Survival of the Fittest: Human Capital and Selection in the Malthusian World." Working Paper, UC Davis.
- Cressy, David. 1980. *Literacy and the social order – reading and writing in Tudor and Stuart England*. Cambridge: 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*. London: 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, 71–114. Stockholm: Almqvist and Wiksell International.
- Fujita, Masahisa and Jacques-François Thisse. 2002. *Economics of Agglomeration*. Cambridge: Cambridge University Press.
- Galor, Oded. 2004. "From Stagnation to Growth: Unified Growth Theory." Working Paper 4581, CEPR.
- 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.
- Holmes, Thomas. 2005. "The Location of Sales Offices and the Attraction of Cities." *Journal of Political Economy*, no. forthcoming.
- Houston, R A. 2002. *Literacy in Early Modern Europe*. Harlow: 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.
- 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*. Paris: 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. London and New York: Longman.
- Perrenoud, Alfred. 1978. "La mortalité à Genève de 1625 à 1825." *Annales de démographie historique*, pp. 209–233.
- Sanderson, Michael. 1995. *Education, Economic Change and Society in England 1780-1870*. Cambridge: Cambridge University Press.

Schools Inquiry Commission. 1868a. *Report of the Commissioners*. Volume 1. London: George E Eyre and William Spottiswoode.

———. 1968b. *Tables*. Volume 21. London: George E Eyre and William Spottiswoode.

Solow, Robert. 1956. "A contribution to the theory of economic growth." *Quarterly Journal of Economics* 70 (1): 65–94.

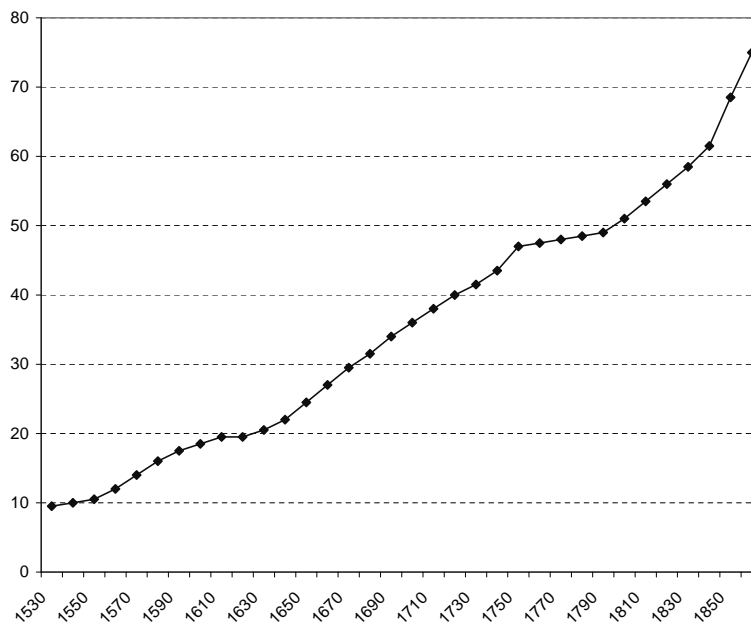
Wrigley, Edward and Roger Schofield. 1989. *The population history of England, 1541–1871: a reconstruction*. Cambridge, Mass: Cambridge University Press.

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

Appendices

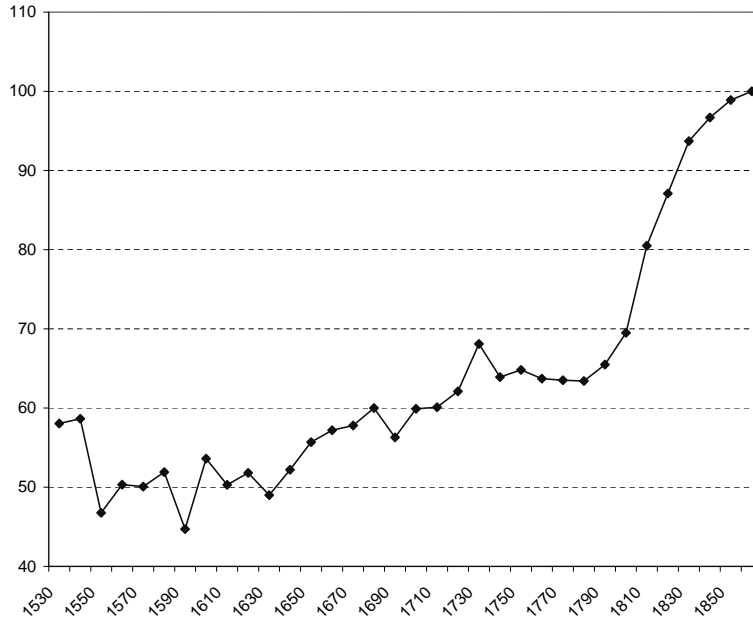
A England data: 1530-1860

Figure A-1: Literacy achievements (% population)



Estimation:Cressy (1980).

Figure A-2: Total factor productivity



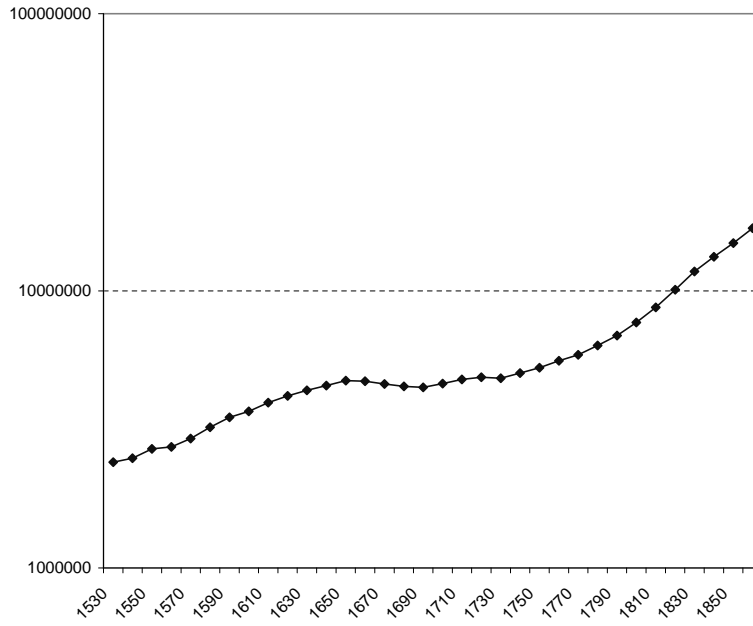
Estimation: Clark (2001).

Figure A-3: Mortality: number of survivors at age 40 from 1000 individuals at age 5



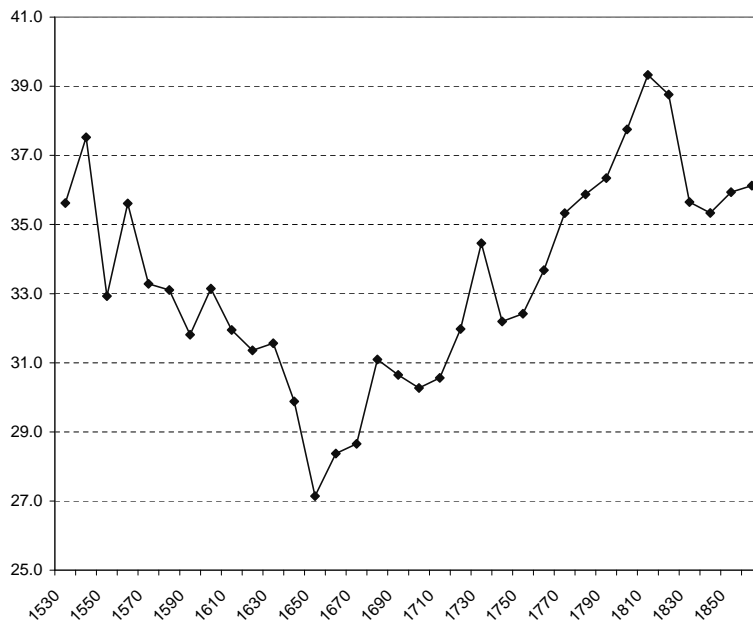
Source: Wrigley, Davies, Oeppen, and Schofield (1997).

Figure A-4: Population of England aged 5+



Estimation: Wrigley, Davies, Oeppen, and Schofield (1997).

Figure A-5: Crude birth rate



Estimation: Wrigley, Davies, Oeppen, and Schofield (1997).

B School establishments

The reports of the School Inquiry Commission written in 1867-8 documents the creation of schools.⁵ 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 (1968b)). 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. These data are presented in Table A-1.

Table A-1: School establishments

		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

⁵A special thank to Françoise Canart for her help in identifying the relevant British Parliamentary Papers.

Table A-2: School establishments (continued)

		grammar	non-classical
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

C 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 $S < L$):

$$\frac{\int_S^L e^{\gamma(z-S)} m(z) dz}{m(S)} - S = \phi, \quad (27)$$

where $\phi = \xi x(i)/\mu$. We shall denote by $\Phi(S)$ the function on the left hand side of Equation (27), 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 trivially 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 .⁶ It follows that $\Phi(S)$ tends to $-L$ when S tends to L , so that (27) 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 $\gamma = 0$.

⁶Just use $e^x - 1$ is almost equal to x when x is close to 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:

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

since 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 the sign of the simpler expression

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

Now, it is easy to 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 seen above, 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.

D Proof of Corollary 1

The proof is trivial except for the demographic parameters α and β . Notice that since $\underline{\mu}(i)$ is given by Equation (15), 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 $m(L) = 0$. Since $\partial m(a) / \partial p > 0$ when $p = \alpha$, and $\partial m(a) / \partial p < 0$ when $p = \beta$, we get the announced results.

E 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 (27).

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 's values small enough. Denoting by $\Phi^1(S) = \int_S^L m(z) dz / m(S)$, and by total differentiation of (27) 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

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

Since

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

and since 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 always bigger than 1 for $0 \leq z \leq L$, and the inequality holds.

F 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:

$$\gamma m(\hat{S}) + m'(\hat{S}) \geq \frac{-2 m^2(\hat{S})}{\int_{\hat{S}}^L e^{\gamma(z-\hat{S})} m(z) dz}. \quad (28)$$

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

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

Since, from Equation (27), \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

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

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 (28). 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 trivially 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 $\mu \underline{\mu}(i)$ tends to $\underline{\mu}(i)$, \hat{S} goes to zero by Proposition 1. Hence $\tilde{\mu}(i, k) = \Psi^{-1}(k + e^{-\gamma t} W^h) > \Psi^{-1}(0) = \underline{\mu}(i)$.

G 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:

$$\frac{\partial \tilde{\mu}(i, k)}{\partial p} = - \frac{\frac{\partial \Psi}{\partial \hat{S}} \frac{\partial \hat{S}}{\partial p} + \frac{\partial \Psi^1}{\partial p}}{\frac{\partial \Psi(\hat{S})}{\partial \mu}}, \quad (29)$$

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 (29) is always 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 (29). 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: $\partial\hat{S}/(\partial\xi x(i)) = 1/(\mu \Phi'(\hat{S})) < 0$, since 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.⁷ Since $\partial\Psi/\partial\hat{S} > 0$ by the proof of Proposition 3, it follows that the numerator in (29) 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 (29), 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

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

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, (29) implies that $\tilde{\mu}(i, k)$ decreases when β is raised.

⁷Put $\gamma = 0$ 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 $m(\hat{S})$.

At 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:

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

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

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

It follows that $\partial \Psi / \partial \gamma > 0$, and by (29), one concludes that $\tilde{\mu}(i, k)$ is a decreasing function of γ as expected.

H 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

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

The first order partial derivatives are

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

and

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

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 less children from every location.

Let us turn to the second order partial derivatives.

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

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.

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

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

$$\begin{aligned} \frac{\partial B}{\partial E} &= k\zeta R(E, k) - f + Ek\zeta \frac{\partial R}{\partial E}(E, k) \\ &= k\zeta R(E, k) - f - \frac{k\zeta}{E} \int_{\tilde{\mu}(\frac{1}{2E}, k)}^{\bar{\mu}} g(\mu) d\mu \\ &= \frac{1}{E} \left[B(E, k) - k\zeta \int_{\tilde{\mu}(\frac{1}{2E}, k)}^{\bar{\mu}} g(\mu) d\mu \right] \end{aligned} \quad (30)$$

and

$$\begin{aligned} \frac{\partial B}{\partial k} &= \zeta ER(E, k) + k\zeta E \frac{\partial R}{\partial k}(E, k) \\ &= 2\zeta E \int_0^{\frac{1}{2E}} \int_{\tilde{\mu}(i, k)}^{\bar{\mu}} g(\mu) d\mu di - 2k\zeta E \int_0^{\frac{1}{2E}} \frac{\partial \tilde{\mu}}{\partial k}(i, k) g[\tilde{\mu}(i, k)] di \end{aligned} \quad (31)$$

Comparing (19) and (30), we easily see that for a fixed k , the function $\frac{\partial B}{\partial E}$ is continuous at \tilde{E} provided that $\ell < 1/2$. Otherwise, there is a jump of magnitude

$$\frac{k\zeta}{\tilde{E}} \int_{\tilde{\mu}(\ell, k)}^{\bar{\mu}} g(\mu) d\mu$$

and a kink at \tilde{E} .

We turn to the second order partial derivatives.

$$\begin{aligned} \frac{\partial^2 B}{\partial E^2} &= 2k\zeta \frac{\partial R}{\partial E}(E, k) + k\zeta E \frac{\partial^2 R}{\partial E^2}(E, k) \\ &= -\frac{k\zeta}{2E^3} \frac{\partial \tilde{\mu}}{\partial i} \left(\frac{1}{2E}, k \right) g[\tilde{\mu}(\frac{1}{2E}, k)] < 0 \end{aligned} \quad (32)$$

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

$$\begin{aligned} \frac{\partial^2 B}{\partial k^2} &= -4\zeta E \int_0^{\frac{1}{2E}} \frac{\partial \tilde{\mu}}{\partial k}(i, k) g[\tilde{\mu}(i, k)] di \\ &\quad - 2k\zeta E \int_0^{\frac{1}{2E}} \left\{ \frac{\partial^2 \tilde{\mu}}{\partial k^2}(i, k) g[\tilde{\mu}(i, k)] + \left[\frac{\partial \tilde{\mu}}{\partial k}(i, k) \right]^2 g'[\tilde{\mu}(i, k)] \right\} di \end{aligned} \quad (33)$$

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 (30), the first order condition give

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

The derivative with regard to k gives:

$$\frac{dBE}{dk}(k) = \zeta \int_{\tilde{\mu}(\frac{1}{2E^*}, k)}^{\bar{\mu}} g(\mu) d\mu - k\zeta g \left[\tilde{\mu} \left(\frac{1}{2E^*}, k \right) \right] \left[-\frac{\partial \tilde{\mu}}{\partial i} \left(\frac{1}{2E^*}, k \right) \frac{1}{(2E^*)^2} \frac{dE^*}{dk} + \frac{\partial \tilde{\mu}}{\partial k} \left(\frac{1}{2E^*}, k \right) \right]$$

The first term is positive, while 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.

I 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 $\Psi(\mu)$ just above, this can only happen if $\tilde{\mu}(i, k)$ itself goes to infinity, since \hat{S} is bounded from above (by L for example).

J 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 set of \mathbb{R}^2 , and since $B(E, k)$ 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 $\theta(f)$.