

**ELITE EDUCATION, MASS EDUCATION,
AND THE TRANSITION
TO MODERN GROWTH**

Holger Strulik, Katharina Werner

GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN

Elite Education, Mass Education, and the Transition To Modern Growth *

Holger Strulik[†]

Katharina Werner^{**}

March 2014

Abstract. For most of human history there existed a well-educated and innovative elite whereas mass education, market R&D, and high growth are phenomena of the modern period. In order to explain these phenomena we propose an innovation-driven growth model for the very long run in which the individual-specific return to education is conceptualized as an compound of cognitive ability and family background. This allows us to establish a locally stable steady state at which family background determines whether an individual experiences education and a locally stable steady state at which education is determined by cognitive ability. Compulsory schooling can move society from elite education to mass education. An interaction between education and life expectancy explains why the education period gets longer with ongoing economic development. Embedding this household behavior into a macro-economy we can explain different paths to modern growth: According to the Prussian way, compulsory education is implemented first and triggers later on the onset of market R&D and modern growth. According to the British way, market R&D and the take off to growth is initiated without mass education, which is triggered later by technical progress and economic development.

Keywords: long-run growth; elite education; compulsory education; longevity; R&D.

JEL: I24; J24; O30; O40.

* We would like to thank Lothar Banz for helpful comments.

[†] University of Goettingen, Department of Economics, Platz der Goettinger Sieben 3, 37073 Goettingen, Germany; email: holger.strulik@wiwi.uni-goettingen.de.

^{**} University of Goettingen, Department of Economics, Platz der Goettinger Sieben 3, 37073 Goettingen, Germany; email: kwerner@uni-goettingen.de.

1. INTRODUCTION

Technological innovations in contemporary highly developed economies are largely broad forward by costly R&D activities. From the historical viewpoint, however, market R&D activities are a relatively recent phenomenon. They are basically unobserved before the Industrial Revolution (Mokyr, 2002). Consequently, technological development in pre-industrial societies has so far been modeled in growth economics as either completely exogenous (e.g. Hansen and Prescott, 2002) or as learning-by-doing (e.g. Galor and Weil, 2000; Galor, 2005; Kremer, 1993; Strulik and Weisdorf, 2008). These learning-by-doing processes were conceptualized as being dependent on the size of the labor force and thus as driven by population growth. An education of parts of the workforce or the presence of scholars was not required in order to advance technological knowledge.

In this paper we propose a growth model with a refined view on pre-industrial innovations. We acknowledge the absence of market R&D and keep the idea of learning-by-doing, but we assume that innovations are brought forward by an educated elite. This view is inspired by the life history of famous pre-industrial inventors. Johann Gutenberg (c. 1390s–1468), Leonardo da Vinci (1452–1519), Otto von Guericke (1602–1686), Galileo Galilei (1564–1642), Christiaan Huygens (1629–1695), Isaac Newton (1642–1727), and Gottfried Wilhelm Leibniz (1646–1716) are some particularly popular examples of inventors coming from an educated family background who experienced a youth of education before they were able to advance technology. In some cases, the later inventors did not experience formal schooling but were trained by their educated parents or other relatives as, for example, Blaise Pascal (1623–1662). Little is known about the early life of most of the famous pre-industrial inventors from Arabia, China, and the classical era. But it is hard to believe that innovations like, for example, the waterpump, the calendar, or the sextant, were created out of the blue by uneducated workers. It seems more likely that they were developed by scholars, based on their knowledge, acquired through education. More is known of the early life of later innovators. Specifically, the innovators of the Industrial Revolution were educated men, “engineers”, trained to understand and apply scientific texts (Mokyr, 2002, 2005).

In order to establish a locally stable equilibrium of elite education we assume that the individual-specific return on education depends not only on cognitive ability but also on family

background such that, *ceteris paribus*, the return on education is higher for children with educated parents. The importance of family background for the return on education and years of schooling is a well-known result of labor economics of contemporaneous societies (e.g. Card, 1999; Meghir and Palme, 2005; Feinstein et al., 2008; Bjoerklund and Salvanes, 2010). There exist no specific estimate for pre-modern times, before the establishment of compulsory schooling, but the narrative is that education was mainly reserved for the children of an educated elite. The masses remained uneducated (Cipolla, 1969; Green, 1990; Mitch, 1993).

In this paper we show that a sufficiently long compulsory schooling period has the power to move society from elite education to mass education and to initiate the onset of market R&D and a take off to modern growth. The key mechanism is that education is conducive to a longer life. Higher life-expectancy then motivates the demand of even more education of the next generation and sets in motion a dynamic process of mutually self-enforcing improvements of education and life-expectancy. Eventually, the society is sufficiently well educated such that market R&D becomes worthwhile and economic growth really takes off. During the take-off to growth inequality first rises and then declines, as suggested by Kuznets (1955).

Our theory motivates two distinct paths to modern growth. According to the “Prussian way”, the setup of compulsory education precedes and enables the onset of market R&D (Becker et al., 2011). According to the “English way”, market R&D and technology growth precedes and triggers the onset of mass education (Mokyr, 2002, Galor, 2005). Which way to modern growth a society takes depends, among other things, on the productivity of market R&D. Furthermore, the model captures also the “Latin American way” according to which absent or insufficient compulsory schooling keeps the society at an equilibrium of elite education and inferior growth for an extended period of time (Sokoloff and Engerman, 2000).

Our study contributes to the unified growth literature and the literature on R&D-based growth.¹ These literatures are largely disconnected in the sense that unified growth theory neglects that modern growth is driven by R&D and that the R&D literature ignores a pre-modern state of economies and the gradual transition to long-run growth. A recent study unifying both literatures is Strulik et al. (2013). Here, we extend the literature by taking into

¹ For unified growth see Galor and Weil (2000), Kögel and Prskawetz (2001), Jones (2001), Lucas (2002), Galor and Moav (2002, 2006), Doepke (2004), Galor and Mountford (2008), Strulik and Weisdorf (2008) and many others. See Galor (2005, 2011) for surveys. For R&D-based growth see Romer (1990), Aghion and Howitt (1992), Jones (1995), Peretto (1998), Segerstrom (1998), and many others. See Gancia and Zilibotti (2005) for a survey.

account that the return of education depends on ability and family background. We use this fact in order to establish a pre-modern equilibrium of elite education and to explain the importance of (compulsory) mass education for the transition to modern growth.

The impact of compulsory education on income and income distribution in an overlapping generations model is also investigated by Eckstein and Zilcha (1994). There, the reason of heterogeneity in education originates from different preferences of parents. There is no endogenous long-run growth and the focus of the analysis is on welfare. There is an intergenerational externality in knowledge production such that a majority of the population favors compulsory schooling at some point of time after its introduction.

Particularly related to our paper are the studies of Boucekkine et al. (2002, 2003) and Cervellati and Sunde (2005). Boucekkine et al. observe a decline in adult mortality in the early modern period and argue that these improvements in longevity have triggered the incentive to invest in education and set in motion a transition towards long-run growth. Cervellati and Sunde integrate a feedback mechanism from education to longevity capturing, for example, the impact of knowledge about nutrition, sanitation, and diseases. The feedback mechanism creates a unified growth model according to which the share of educated persons and life expectancy are rising and mutually enforcing each other. Long-run growth in these studies is driven by human capital externalities and learning-by-doing. We built on this literature and advance it by integrating the notion that technology in modern societies is developed by market R&D and by investigating the distinct paths to modern growth. Moreover, we explore the role of family background for education, social mobility, and how compulsory education contributes to the transition to modern growth.

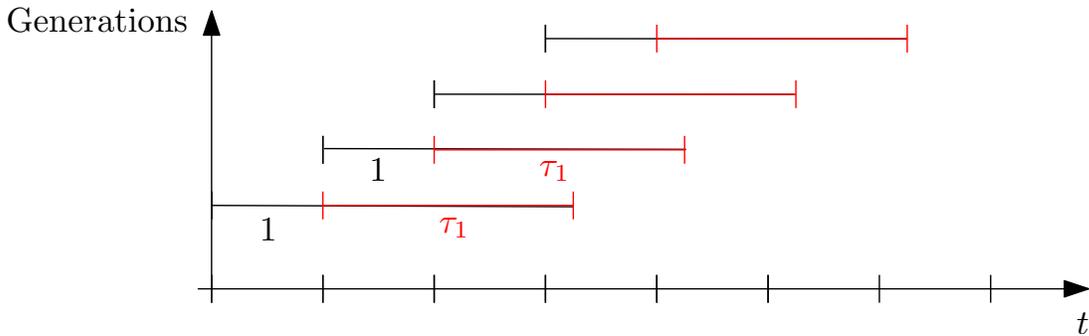
The idea that increasing longevity motivates education because it increases the period during which individuals are able to earn a return on education goes back to Ben-Porath (1967). It has recently been questioned by Hazan (2009) because it seems to be incompatible with a long-run trend of declining life time labor supply. Meanwhile, however, it has been shown that there are several extensions of the basic model possible that re-establish the Ben-Porath mechanism (Cervellati and Sunde, 2013; Strulik and Werner, 2013; Hansen and Loenstrup, 2012). Similarly the study of Acemoglu and Johnson (2009) could not find a positive impact of increasing life expectancy on education but later refinements of the empirical strategy were able to establish strongly supportive evidence (Cervellati and Sunde, 2001; Hansen, 2013). The

reverse channel running from education to health has never been controversial. A prominent study is Jayachandran and Lleras-Muney (2009); a recent survey is provided by Cutler et al. (2011).

2. THE MODEL

2.1. Demographic Structure. The household sector is represented by an overlapping generations model. Individuals live for two distinct periods, youth and (mature) adulthood. During youth individuals may or may not experience some schooling and spend the rest of their time working. During adulthood all time is used for labor supply. The length of the first period of life is constant and normalized to a unit of time of the dynamic macroeconomy ($\tau_0 \equiv 1$, which corresponds with, for example, 20 years). The length of the second period is given for the individual but potentially varying from one generation to the next. It is denoted by $\tau_1 \geq \tau_0$. Life expectancy is computed as $LE = 1 + \tau_1$. Fertility is assumed to be constant, i.e. at each unit of time a new generation is born. The demography of the population is visualized in Figure 1.

FIGURE 1. Demography



2.2. Households. Individuals experience utility from consumption in both periods of life. Imposing, for simplicity, a log-form of the period utility function, lifetime utility of an individual born at the beginning of period t is given by

$$U = \log(c_{t,0}) + \gamma\tau_1 \log(c_{t+1,1}), \quad (1)$$

in which γ is the time discount rate of the second period of life.

At each time increment of the first period individuals divide their time between education e_t and work $1 - e_t$, where e_t denotes the average amount of education per time increment of

an individual born in t . Besides voluntary education there may exist a period of compulsory education, $\sigma \in [0, 1)$. The introduction or extension of compulsory education will later on be our most interesting experiment of comparative statics and comparative dynamics. Let ϵ_t denote voluntary education, $\epsilon_t \in [0, 1 - \sigma)$. Total education is then given by $e_t = \sigma + \epsilon_t$.

The budget constraint of the first period is given by

$$c_{t,0} = (1 - e_t)w_{t,0}, \quad (2)$$

in which $w_{t,0}$ denotes the average wage rate per time increment and $c_{t,0}$ the average consumption per time increment of the first period.

During the second period of life individuals work and earn a wage, depending on their schooling level. Individual i is a child of parent j and earns a wage income per time increment $w_{t+1,1}h(e_t, \beta_j, \theta_i)$, $i, j \in \{L, H\}$, where $w_{t+1,1}$ denotes the wage per unit of human capital per time increment, and $h(e_t, \beta_j, \theta_i)$ measures the return to schooling. The parameters β_j and θ_i define the type of the individual. In particular, $\theta_i \in \{\theta_L, \theta_H\}$ is a measure of the innate ability, $\theta_L \leq \theta_H$. The parameter β_j takes into account that the intellectual background of parents influence the ability of children to learn. In particular, $\beta_j = \beta_L$ if parents experienced at most only compulsory education, i.e. for $e_{t-1} \leq \sigma$, and $\beta_j = \beta_H$ if parents invested voluntarily in education, i.e. for $e_{t-1} > \sigma$, with $\beta_H \geq \beta_L$. Generally, the size of β_j captures the importance of family background for human capital formation of the offspring. For linguistic ease we refer to the θ s simply as ability and to the β s as family background, knowing that the β 's capture ability as well, namely the part determined by educational background of the parents. See Feinstein et al. (2008) for a detailed overview on the impact of parental education on cognitive ability.

Human capital is accumulated according to Mincer (1974) with a constant return to schooling $\beta_j\theta_i$, i.e.

$$h(e_t, \beta_j, \theta_i) := \exp[\beta_j \cdot \theta_i \cdot e_t], \quad i, j \in \{L, H\}. \quad (3)$$

The interaction of cognitive ability θ_i and family background β_j introduces imperfect social mobility. Below it will explain the persistence of a steady state at which only a small elite experiences education, namely only children coming from a family background with education, and another steady state at which individuals educate according to ability.

The budget constraint of individuals in their second period of life is given by

$$c_{t+1,1} = h(e_t, \beta_j, \theta_i) \cdot w_{t+1,1}. \quad (4)$$

Individuals seek to maximize lifetime utility (1) such that the budget constraints (2) and (4), the human capital equation (3), and the non-negativity constraint $e_t \geq 0$ are fulfilled. In the Appendix we show that the solution for education is obtained as

$$e_t = \sigma + \epsilon_t \quad \text{with} \quad \epsilon_t = \max \left\{ 0, 1 - \sigma - \frac{1}{\gamma \tau_1 \beta_j \theta_i} \right\}, \quad i, j \in \{L, H\}. \quad (5)$$

The following proposition describes the comparative statics of the interior solution of the optimization problem.

PROPOSITION 1 (Education). *At the interior solution, voluntary education, $\epsilon_t = e_t - \sigma$, increases with cognitive ability θ_i , family background β_j , and the length of the period of adulthood τ_1 (i.e. with life expectancy). It decreases with the requested level of compulsory schooling σ .*

2.3. Aggregation. For the ease of exposition we refer to individuals with the highest level of education as scientists and engineers and denote their aggregate human capital by S_t . The remainder of the population is called workers and their aggregate supply is denoted by H_t . The first group is, of course, thought of comprising all occupations filled by individuals with high education and includes as well, lawyers, administrators, and other white collar employees. In general, there exist four types of individuals depending on ability and family background, namely $\{\beta_H, \theta_H\}, \{\beta_H, \theta_L\}, \{\beta_L, \theta_H\}, \{\beta_L, \theta_L\}$. The highest education level is achieved by individuals of high ability and well educated parents ($\{\beta_H, \theta_H\}$). We assume that only these individuals are able to become scientists. Let t denote the current time period as well as the birth year of a cohort. Furthermore, let $\tau_{t,1} \leq \bar{\tau}_1$ be the length of the adulthood period of the generation t , with upper bound $\bar{\tau}_1$. The aggregate human capital at time t can be calculated by summing up over all generations alive at time t . Since only individuals who finished their education (and moved further to the life period of adulthood) are able to become scientists, the aggregate human capital of scientists at time t is

$$S_t := \sum_{i=1}^{\lceil \bar{\tau}_1 \rceil} \Gamma_{t,t-i} \cdot h(\beta_H, \theta_H, e_{t-i}(\beta_H, \theta_H)) \cdot p \cdot q_{t-i} \quad (6)$$

where $\Gamma_{t,j}$ is an indicator function showing whether individuals born in j are still alive at time t , i.e.

$$\Gamma_{t,j} = \begin{cases} 1, & \text{if the generation born in } j \text{ is still alive in } t \\ 0, & \text{otherwise.} \end{cases}$$

The parameter $p \in (0, 1)$ is the population share of individuals with high cognitive ability θ_H , which is assumed to be constant over time. Furthermore, the variable $q_t \in [0, 1]$ is the population share of adults who had parents experiencing education beyond the compulsory level.

FIGURE 2. Genealogical Tree

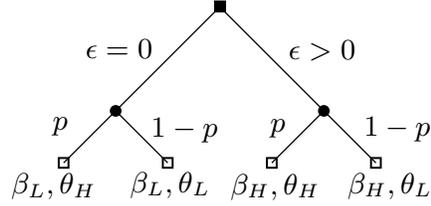


Figure 2 displays the genealogical tree. The parent generation experienced voluntary education or not ($\epsilon = 0$ or $\epsilon > 0$). With probability p the child generation has high cognitive ability and with probability $1-p$ it has low cognitive ability. The evolution of the 4 groups of the population can be determined as follows. Let α_t^{ij} , $i, j \in \{H, L\}$ be the population share of each group in period t . Furthermore, let

$$\chi(\epsilon^{ij}) = \begin{cases} 1, & \text{if } \epsilon^{ij} > \sigma \\ 0, & \text{otherwise} \end{cases}$$

denote the indicator function showing whether a group of individuals achieved more than compulsory education. The population shares of the four different types are obtained as

$$\alpha_{t+1}^{HH} = p [\alpha_t^{HH} \chi(\epsilon_t^{HH}) + \alpha_t^{HL} \chi(\epsilon_t^{HL}) + \alpha_t^{LH} \chi(\epsilon_t^{LH}) + \alpha_t^{LL} \chi(\epsilon_t^{LL})] \quad (7)$$

$$\alpha_{t+1}^{HL} = (1-p) [\alpha_t^{HH} \chi(\epsilon_t^{HH}) + \alpha_t^{HL} \chi(\epsilon_t^{HL}) + \alpha_t^{LH} \chi(\epsilon_t^{LH}) + \alpha_t^{LL} \chi(\epsilon_t^{LL})]$$

$$\alpha_{t+1}^{LH} = p [\alpha_t^{HH} (1 - \chi(\epsilon_t^{HH})) + \alpha_t^{HL} (1 - \chi(\epsilon_t^{HL})) + \alpha_t^{LH} (1 - \chi(\epsilon_t^{LH})) + \alpha_t^{LL} (1 - \chi(\epsilon_t^{LL}))]$$

$$\alpha_{t+1}^{LL} = (1-p) [\alpha_t^{HH} (1 - \chi(\epsilon_t^{HH})) + \alpha_t^{HL} (1 - \chi(\epsilon_t^{HL})) + \alpha_t^{LH} (1 - \chi(\epsilon_t^{LH})) + \alpha_t^{LL} (1 - \chi(\epsilon_t^{LL}))]$$

with α_0^{ij} given. The share of individuals with educated family background in period $t + 1$ is given by $q_{t+1} := \alpha_{t+1}^{HH} + \alpha_{t+1}^{HL}$. Aggregate human capital of workers is computed as

$$\begin{aligned}
H_t := & (1 - e_t(\beta_H, \theta_H)) \cdot p \cdot q_t + (1 - e_t(\beta_H, \theta_L)) \cdot (1 - p) \cdot q_t + (1 - e_t(\beta_L, \theta_H)) \cdot p \cdot (1 - q_t) + \\
& + (1 - e_t(\beta_L, \theta_L)) \cdot (1 - p) \cdot (1 - q_t) + \sum_{i=1}^{\lceil \bar{\tau}_1 \rceil} \Gamma_{t,t-i} \cdot [(1 - p) \cdot q_{t-i} \cdot h(\beta_H, \theta_L, e_{t-i,0}) \\
& + p \cdot (1 - q_{t-i}) \cdot h(\beta_L, \theta_H, e_{t-i,0}) + (1 - p)(1 - q_{t-i})h(\beta_L, \theta_L, e_{t-i,0})]. \quad (8)
\end{aligned}$$

2.4. Firms. The production side of the economy is a discrete time version of the well-known endogenous growth model from Romer (1990) and Jones (1995) augmented by learning-by-doing, as in Strulik et al. (2013). The production side of the economy is organized in three sectors, the final goods sector, the intermediate goods sector and the research sector.

2.4.1. Final Goods Sector. In the final goods sector perfectly competitive firms produce output Y_t given a Cobb-Douglas production technology, i.e. $Y_t = B_t H_{t,Y}^\alpha S_{t,Y}^\eta \int_{i=1}^{A_t} x_{i,t}^{1-\alpha-\eta}$. Here, $S_{t,Y}$ and H_t denote the human capital of scientists and workers employed in the final goods sector. A_t is the number of intermediate goods available at time t , B_t the technological level acquired through learning-by-doing. The variable $x_{i,t}$ describes the amount of the intermediate good i in the final goods production. The parameters α and η denote the share of workers and scientists, $\alpha, \eta \in (0, 1)$ and $\alpha + \eta < 1$ is the labor share in goods production. Profit maximization provides the indirect demand functions for workers $w_{t,H_t} = \alpha Y_t / H_t$, scientists $w_{t,S_{t,Y}} = \eta Y_t / S_{t,Y}$, and intermediate goods $p_{i,t} = (1 - \alpha - \eta) B_t H_t^\alpha S_{t,Y}^\eta x_{i,t}^{-\alpha-\eta}$.

2.4.2. Intermediate goods sector. In the intermediate goods sector firms produce under monopolistic competition. A firm that buys a blueprint from the R&D-sector is allowed to produce one specific intermediate good. Firms use a production technology that transfers one unit of the final good into one unit of the intermediate good. Hence, profits of an intermediate goods producer are $\pi_{i,t} = p_{i,t} x_{i,t} - x_{i,t}$. Profit maximization provides optimal quantities $x_{i,t} = [(1 - \alpha - \eta)^2 B_t H_t^\alpha S_{t,Y}^\eta]^{1/(\alpha+\eta)}$ and prices $p_{i,t} = 1/(1 - \alpha - \eta)$. Henceforth, the firm-specific index i is dropped, because all firms produce the same quantity and sell their products at the same price.

2.4.3. *Knowledge Production.* New knowledge is created by scientists in potentially two different ways. The learning-by-doing process is described by the function

$$B_{t+1} - B_t = S_t^\lambda, \quad \lambda \in (0, 1), \quad (9)$$

where S_t is the total amount of the human capital of the scientists. Taking learning-by-doing into account allows to explain how economies evolved through continuous innovations although market R&D was not yet profitable. It captures the long phase of human history in which knowledge was advanced by a small population share of educated tinkerers, but knowledge and the supply of scientists were too small to make large scale R&D activities profitable.

If the R&D-sector is operative, it employs scientists and produces new knowledge given by the production function

$$A_{t+1} - A_t = \delta_t S_{t,A}, \quad \delta_t \equiv \bar{\delta} A_t (1 + S_{t,A})^{-\phi}, \quad (10)$$

where $S_{t,A}$ denotes the scientists employed in the R&D-sector and $\bar{\delta} > 0$ is the general productivity of R&D, which is determined, for example, by the legal system and other institutions affecting the creation and diffusion of knowledge. The parameter ϕ controls for duplication externalities in research (stepping-on-toes effects). Using $1 + S_{t,A}$ instead of S_t (as, e.g. in Jones, 1995) for the computation of the duplication externality is a simple device to avoid the counterfactual prediction that productivity of the first person in R&D is infinitely large (i.e. when $S_{t,A} = 0$). It allows us to model a period in human history without market R&D. Maximization of profits in the R&D-sector implies wages $w_{t,S_{t,A}} = p_{t,A} \delta_t$, in which $p_{t,A}$ denotes the price of a blueprint for a new intermediate product. Following Aghion and Howitt (2009, Chapter 4) and Strulik et al. (2013) we assume that a patent holds for one generation (in the quantitative part this will be 20 years) and that afterwards, in any future period $t + 1$ the monopoly right is sold at price π_{t+1} to someone chosen at random from the currently active generation. The revenue is spent unproductively on public consumption. This simplification helps to avoid intertemporal (dynastic) problems of patent holding and patent pricing while keeping the basic incentive to create new knowledge intact.

2.4.4. *Life Expectancy.* In order to capture the idea that education exerts a positive influence on the evolution of life expectancy we follow Cervellati and Sunde (2005) and assume

a feedback effect from average education of the current generation to life expectancy of the next generation. In particular, we assume life expectancy of young adults to be determined by $\tau_{t,1} = \bar{\tau} \cdot (1 + \psi \tilde{e}_{t-1})$, $\psi > 0$, in which average education is computed as $\tilde{e}_t = \alpha_t^{HH} e_t^{HH} + \alpha_t^{HL} e_t^{HL} + \alpha_t^{LH} e_t^{LH} + \alpha_t^{LL} e_t^{LL}$. This completes the description of the economy.

3. STEADY STATES OF EDUCATION

A steady state is defined as a situation in which the level of education for each group of society and the associate population shares are constant and in which aggregate variables are either constant or grow at constant rates. The setup of the model allows for multiple steady states. Because *HH*-individuals always prefer the highest education level and *LL*-individuals prefer the lowest level, there are, in principle, five different combinations of education levels observable in the society, namely

- (1) $\epsilon^{HH} > 0$, $\epsilon^{HL} = \epsilon^{LH} = \epsilon^{LL} = 0$
- (2) $\epsilon^{HH} > 0$, $\epsilon^{HL} > 0$, $\epsilon^{LH} = \epsilon^{LL} = 0$
- (3) $\epsilon^{HH} > 0$, $\epsilon^{LH} > 0$, $\epsilon^{HL} = \epsilon^{LL} = 0$
- (4) $\epsilon^{HH} > 0$, $\epsilon^{LH} > 0$, $\epsilon^{HL} > 0$, $\epsilon^{LL} = 0$
- (5) $\epsilon^{HH} > 0$, $\epsilon^{LH} > 0$, $\epsilon^{HL} > 0$, $\epsilon^{LL} > 0$.

In case (1) only individuals with the highest return to schooling prefer education. The population share of this group evolves according to $\alpha_{t+1}^{HH} = p\alpha_t^{HH}$. It vanishes in the long run since the probability to have a high cognitive ability is smaller than one. This means that the first case is only observed in transition and not as a steady-state outcome.

In case (2) all individuals with educated parents are educated, which results in

$$\alpha_{t+1}^{HH} = (\alpha_t^{HH} + \alpha_t^{HL})p, \quad \alpha_{t+1}^{HL} = (\alpha_t^{HH} + \alpha_t^{HL})(1-p) \Rightarrow q_t = \alpha_t^{HH} + \alpha_t^{HL} \equiv \text{const} \quad (11)$$

In case (3) all individuals of high cognitive ability prefer education irrespective of family background. This implies $q_t = p$. In case (4) all but the individuals with the lowest return to schooling prefer education. With the same argument as for the first case these individuals die out, implying that in the long run everyone prefers education, i.e. $q_t \rightarrow 1$ for $t \rightarrow \infty$. Therefore, case (4) cannot be a steady state. In case (5) everyone prefers education, implying that after one generation everyone has a family background of more than voluntary education, $\beta = \beta_H$ and $q_t = 1$ from the second generation onwards. Finally there exists, of course, a trivial steady-state,

at which nobody is educated. Altogether this means that there are three types of non-trivial steady states observable, as summarized by the following proposition:

PROPOSITION 2 (Steady States of Education). *There are three possible non-trivial steady states for education, namely*

- (1) $\epsilon^{HH} > 0, \epsilon^{HL} > 0, \epsilon^{LH} = \epsilon^{LL} = 0$ for $\beta_H\theta_L > \beta_L\theta_H$ (*Elite Education*)
- (2) $\epsilon^{HH} > 0, \epsilon^{LH} > 0, \epsilon^{HL} = \epsilon^{LL} = 0$ for $\beta_L\theta_H > \beta_H\theta_L$ (*Education of Talented*)
- (3) $\epsilon^{HH} > 0, \epsilon^{LH} > 0, \epsilon^{HL} > 0, \epsilon^{LL} > 0$ for $\beta_H\theta_L \leq \beta_L\theta_H$ (*Mass Education*).

At a steady state both education and life expectancy have to be constant, i.e. $\epsilon_t = \epsilon_{t+1} = \epsilon$ and, hence, $\tau_{t,1} = \tau_{t+1,1} = \tau_1$. In the following discussion we focus on the behavior around the steady state and assume that small changes do not lead to a transition from one steady state to the next. In Section 3.4 we explicitly analyze the transition between different steady states. We call the first steady state, at which only children of educated parents prefer education, *Elite Education* and the second steady state, at which only individuals of high ability are educated, *Education of the Talented*. The third case, in which everybody prefers education irrespective of talent and family background, is called *Mass Education*.

3.1. Elite Education. Life expectancy at a steady state of elite education is obtained as $\tau_1 = \bar{\tau} (1 + \psi (\alpha^{HH}\epsilon^{HH} + \alpha^{HL}\epsilon^{HL} + \sigma))$. Inserting the expression into the preferred level of education (5) provides the education levels

$$\epsilon^{HH} = 1 - \sigma - \frac{1}{\gamma\theta_H\beta_H\bar{\tau} (1 + \psi (\alpha^{HH}\epsilon^{HH} + \alpha^{HL}\epsilon^{HL} + \sigma))} \quad (12a)$$

$$\epsilon^{HL} = 1 - \sigma - \frac{1}{\gamma\theta_L\beta_H\bar{\tau} (1 + \psi (\alpha^{HH}\epsilon^{HH} + \alpha^{HL}\epsilon^{HL} + \sigma))}. \quad (12b)$$

The system of two equations has two solutions, $\{\epsilon_1^{HH}, \epsilon_1^{HL}\}$ and $\{\epsilon_2^{HH}, \epsilon_2^{HL}\}$, which are computed in Appendix B. It holds that $\epsilon_1^{HH} \geq \epsilon_2^{HH}$ and $\epsilon_1^{HL} \geq \epsilon_2^{HL}$. Furthermore, it is trivial to show that $\epsilon^{LL} \leq \epsilon^{LH}$, because $\theta_H \geq \theta_L$. This allows us to compute the necessary and sufficient condition for existence of a steady state of elite education.

LEMMA 1. *Children of uneducated parents remain uneducated ($\epsilon^{LH} = 0$) iff*

$$\beta_L \leq \frac{1}{(1 - \sigma)\gamma\theta_H\tau_{t,1}}. \quad (13)$$

The condition is intuitive. If the return on education of talented children without a family background of education $\beta^L\theta^H$ is too low relative to life expectancy, these talented children remain uneducated and thus, naturally, untalented children of uneducated parents remain uneducated as well. In principle, the return on education could be either too low because the skill premium is too low (the return on high ability θ_H is too low) or because the missing family background of education has a sufficiently strong negative influence (β^L is too low). At an equilibrium of elite education, however, untalented children of educated parents receive an education. It is thus family background that inhibits education of talented children.

Taking the trivial steady state into account there is always more than one steady state and, hence, neither can be globally stable. For the analysis of local stability we follow Galor (2007, Chap. 4). We find that the Jacobian of the difference system linearized around the steady state has one eigenvector with eigenvalue zero and one eigenvector with a positive eigenvalue. Moreover, these two eigenvectors span the entire \mathbb{R}^2 . The first eigenvector corresponds to a situation in which the two education levels have been changed such that the life expectancy is unchanged. If the second eigenvalue is less than unity, the steady state is locally stable. This leads to the following proposition.

PROPOSITION 3 (Local Stability of Elite Education). *The steady state $(\epsilon_1^{HH}, \epsilon_1^{HL})$ is locally stable. The steady state $(\epsilon_2^{HH}, \epsilon_2^{HL})$ is unstable.*

PROPOSITION 4 (Comparative Statics of Elite Education). *The level of education of the elite increases in the importance of family background, β_H , in cognitive ability, θ_L and θ_H , and in compulsory schooling σ . Furthermore, increasing compulsory education σ leads to more voluntary education iff β_H is sufficiently low, i.e. $\beta_H < \bar{\beta}_{H,1}$ (for $\epsilon_{1,2}^{HH}$) and $\beta_H < \bar{\beta}_{H,2}$ (for $\epsilon_{1,2}^{HL}$). The thresholds $\bar{\beta}_{H,1}$ and $\bar{\beta}_{H,2}$ are defined in Appendix B.*

3.2. Education of the Talented. Analogously to the reasoning of Section 3.1 we compute the implied life expectancy at this steady state, $\tau_1 = \bar{\tau} (1 + \psi (\alpha^{HH}\epsilon^{HH} + \alpha^{LH}\epsilon^{LH} + \sigma))$, insert it into the preferred level of education (5) and obtain

$$\epsilon^{HH} = 1 - \sigma - \frac{1}{\gamma\theta_H\beta_H\bar{\tau}(1 + \psi(\alpha^{HH}\epsilon^{HH} + \alpha^{LH}\epsilon^{LH} + \sigma))}, \quad (14a)$$

$$\epsilon^{LH} = 1 - \sigma - \frac{1}{\gamma\theta_H\beta_L\bar{\tau}(1 + \psi(\alpha^{HH}\epsilon^{HH} + \alpha^{LH}\epsilon^{LH} + \sigma))}. \quad (14b)$$

Again, the system of equations exhibits two solutions $\{\epsilon_1^{HH}, \epsilon_1^{LH}\}$ and $\{\epsilon_2^{HH}, \epsilon_2^{LH}\}$, which are described in detail in Appendix B. It holds that $\epsilon_1^{HH} \geq \epsilon_2^{HH}$ and $\epsilon_1^{LH} \geq \epsilon_2^{LH}$, and it is trivial to show that $\epsilon^{LL} \leq \epsilon^{HL}$, because $\beta_H > \beta_L$. The following Lemma states the condition for existence of a steady state of Education of the Talented.

LEMMA 2. *Children of low cognitive ability remain uneducated ($\epsilon^{HL} = 0$) iff*

$$\theta_L \leq \frac{1}{(1 - \sigma)\gamma\beta_H\tau_{t,1}}. \quad (15)$$

The condition is structurally similar to (13). Here, however, high ability children of uneducated parents experience education such that low cognitive ability is identified as the driver of missing education. Local stability is proved analogously to Proposition 3.

PROPOSITION 5 (Local Stability of Education of the Talented). *The steady state $(\epsilon_1^{HH}, \epsilon_1^{LH})$ is locally stable, whereas the steady state $(\epsilon_2^{HH}, \epsilon_2^{LH})$ is unstable.*

PROPOSITION 6 (Comparative Statics of Education of the Talented). *The level of education of individuals with high cognitive ability increases in the importance of family background β_H, β_L , in cognitive ability θ_H and in compulsory schooling σ . Furthermore, increasing compulsory education σ leads to more voluntary education iff the level of θ_H is sufficiently low, that is iff $\theta_H < \bar{\theta}_{H,1}$ (for $\epsilon_{1,2}^{HH}$) and $\theta_H < \bar{\theta}_{H,2}$ (for $\epsilon_{1,2}^{LH}$). The thresholds $\bar{\theta}_{H,1}$ and $\bar{\theta}_{H,2}$ are defined in Appendix B.*

3.3. Mass Education. At the steady state of mass education everyone prefers education, which means that everyone has a family background of voluntary education, such that $\alpha^{LH} = \alpha^{LL} = 0$. This means that the steady state of mass education is similar to the steady state of elite education in that life expectancy is exclusively driven by ϵ^{HH} and ϵ^{HL} at both steady states. The steady states are structurally similar but, of course, the implied levels of population shares are very different. At the steady state of elite education only a few individuals experience education whereas here everybody experiences education. Because of structural similarity the stability analysis and the comparative statics are similar to the steady state the of elite education and the details are omitted. Results are summarized in the following proposition.

PROPOSITION 7. *There are two non-trivial steady states of mass education, one locally stable and the other unstable. The level of education increases in the importance of family background*

and in cognitive ability, and in compulsory schooling. Furthermore, increasing compulsory education leads to more voluntary education iff β_H is sufficiently low.

3.4. Transition from Elite Education to Mass Education. In this section we investigate how a compulsory schooling reform could initiate a transition from elite education to mass education. Notice that according to Proposition 2 it is impossible to transit from elite education to the steady state of education of the talented (without parametric changes in the returns to education $\beta_j\theta_i$). At the steady state of elite education, only individuals with a family background of education prefer education, whereas at the steady state of mass education all individuals prefer education. The transition starts when the LH -individuals prefer to get educated. Naturally, once individuals without a family background of education start education, the population share α^{LL} declines to zero, due to the inheritance of family background.

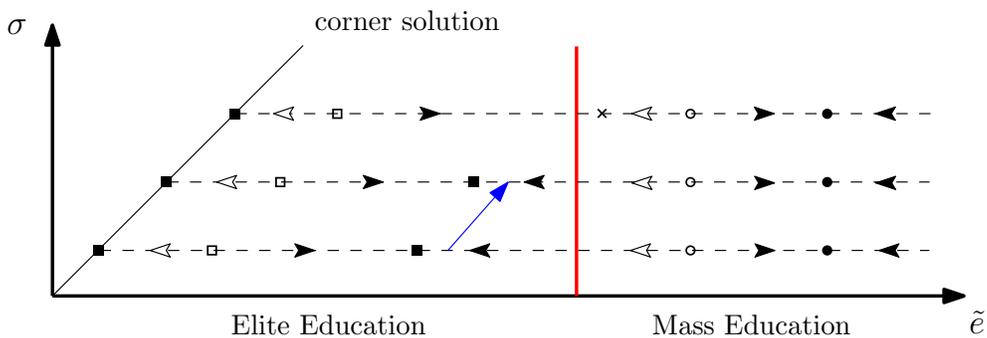
The effects of compulsory schooling on preferred education are potentially ambiguous. On the one hand, those individuals preferring anyway an education replace voluntary schooling by compulsory schooling. On the other hand, a higher level of average education improves life expectancy of the next generation and leads therewith indirectly to increasing demand of voluntary education, c.f. (5), which induces further rising life expectancy and education one period later. This means that there exist three different outcomes, depending on the size of the schooling reform and the economy's current situation. The three cases are visualized in Figure 3.

The first possibility, shown in the top panel of Figure 3, is that the triggered change of life expectancy is too insignificant such that the LH -individuals are still not motivated to obtain education beyond the compulsory level, c.f. Lemma 1. The economy remains in the domain of attraction of the steady-state of elite education.

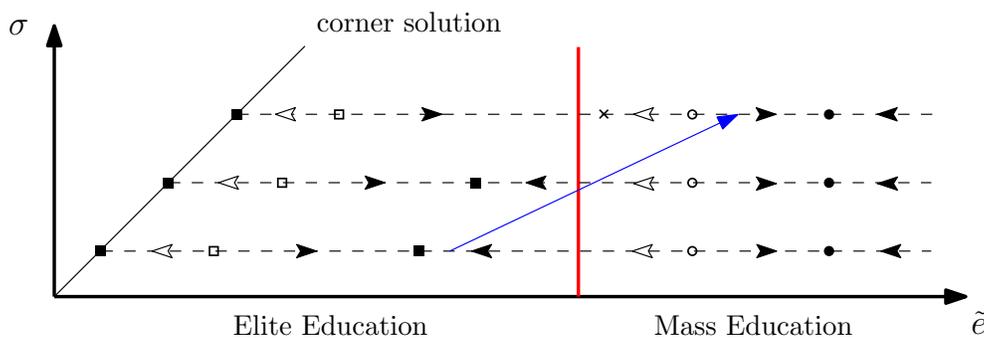
Secondly, it can be the case that the triggered increase of life expectancy is so strong that that LH -individuals of the next generation prefer voluntary education, according to Lemma 1. In this case, shown in the panel in the middle of Figure 3, the economy transits from elite to mass education from one generation to the next. Formally, the economy jumps in the domain of attraction of the steady-state of mass education.

Finally, there exists an intermediate case, shown in the bottom panel of Figure 3, in which compulsory schooling operates gradually through increasing life expectancy of the elite. Here, LH -individuals are not motivated immediately to take up schooling beyond the compulsory level

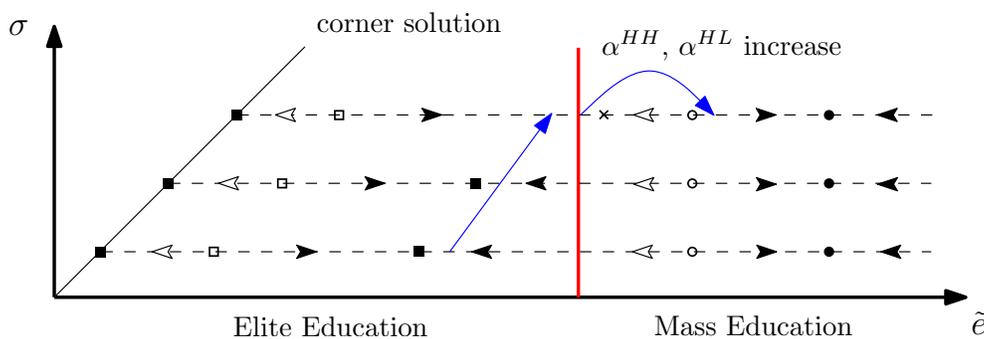
FIGURE 3. Effects of an Increase in Compulsory Schooling



(a) No Transition to Mass Education



(b) Immediate Transition to Mass Education



(c) Gradual Transition to Mass Education

- stable ss (elite) □ unstable ss (elite) • stable ss (mass)
- unstable ss (mass) ◀ transition to ss ◀ transition to corner solution
- × unattained stable ss (elite)

Notes: For average education \tilde{e} to the left of the red bar only HH - and HL -individuals prefer education. Dashed horizontal lines show the feasible set of average education \tilde{e} for given compulsory schooling level σ . Compulsory schooling σ is the lower bound of \tilde{e} such that the corner solutions lie on the 45° -line. At each side of the threshold there are two non-trivial steady states, indicated by circles (boxes): an unstable one (white) and a locally stable one (black). The blue arrows display the immediate effect of an increase in compulsory schooling. The slope depends inversely on the population share $\alpha^{LH} + \alpha^{LL}$.

but increased life expectancy motivates the *elite* to take up even more education, which further increases life expectancy and elite education one period later. As this process continues, life expectancy eventually crosses the threshold such that *LH*-individuals prefer to uptake education beyond the compulsory level. Formally, the first phase of the transition is characterized as movement towards the steady state “beyond the boundary”, marked in the figure by an “x”. During the transition towards this steady state, the economy passes the threshold for mass education and, hence, the steady state becomes unattainable. After crossing the threshold, *LH*-individuals prefer education. This leads to an additional jump of average education \bar{e} , which initiates the transition towards the steady-state of mass education. These insights are summarized by the following propositions.

PROPOSITION 8 (Fast Transition to Mass Education). *For an increase in compulsory schooling $\tilde{\sigma} \in (\tilde{\sigma}_3, \tilde{\sigma}_4) \cap (0, 1 - \sigma)$ the economy moves from elite education to mass education within one generation. The thresholds $\tilde{\sigma}_{3,4}$ are defined in Appendix C.*

PROPOSITION 9 (Gradual Transition to Mass Education). *As a response to the introduction of compulsory schooling $\tilde{\sigma} \in (\tilde{\sigma}_1, \tilde{\sigma}_2) \cap (0, 1 - \sigma)$ the economy undergoes a transition from elite education to mass education at some point in the future. The thresholds $\tilde{\sigma}_{1,2}$ are defined in Appendix C.*

Notice that the transition to mass education is uni-directional and irreversible by, for example, a reduction of compulsory schooling. At the steady state of mass education, everyone prefers voluntary education and thus a reduction of compulsory schooling would be harmless for life expectancy and education of the next generation.

4. THE MACRO-ECONOMY

4.1. Steady States. In equilibrium the labor market clears, i.e. $S_t = S_{t,Y} + S_{t,A}$. Furthermore, if the R&D-Sector is operative, wages of scientists are equalized across sectors, i.e.

$$w_{t,S_{t,Y}} = w_{t,S_{t,A}} \Leftrightarrow \eta \frac{Y_t}{S_{t,Y}} = \delta_t p_{t,A} = \bar{\delta} A_t (1 + S_{A,t})^{-\phi}. \quad (16)$$

Free entry to intermediate goods production implies that operating profits equal the price of a blueprint, i.e. $\pi_t^I = p_{t,A}$. Profits of intermediate goods producers can be restated as

$$\pi_t^I = p_t \cdot x_t - x_t = (\alpha + \eta)(1 - \alpha - \eta) \frac{Y_t}{A_t}. \quad (17)$$

From (16) and (17) we obtain an equation for the allocation of scientists

$$S_{t,Y} = \min \left\{ S_t, \frac{\eta(1 + S_{t,A})^\phi}{\bar{\delta}(\alpha + \eta)(1 - \alpha - \eta)} \right\}. \quad (18)$$

If there are too few scientists in society, they are all absorbed by goods production and there is no market R&D. Technology advances only through learning by doing activities. Using the labor market clearing condition and noting that $S_{t,A} \geq 0$, the growth rate of R&D output, $g \equiv (A_{t+1} - A_t)/A_t$, is computed as

$$g = \max \left\{ 0, \bar{\delta} S_{t,A} (1 + S_{t,A})^{-\phi} \right\}. \quad (19)$$

At a steady state the growth rates are constant and there is no population growth since life expectancy is constant as well. From equation (9) it follows that $g_{B,t} \rightarrow 0$ for $t \rightarrow \infty$. Hence, the impact of learning-by-doing on growth of the economy vanishes in the long-run and market R&D is the only driving force along the asymptotically attained balanced growth rate. Therefore, the steady state growth rate of the economy is (asymptotically) also given by equation (19).

The aggregate human capital of researchers, $S_{t,A}$, is implicitly determined by the aggregate human capital of scientists S_t (6). This leads to the following observations.

PROPOSITION 10.

- (1) *At the steady state the growth rate of GDP per capita equals the growth rate of technology.*
- (2) *At the steady state with R&D, the growth rate of technology is positive and depends positively on the aggregated human capital of researchers, $S_{t,A}$ and thus on the aggregated human capital of scientists S_t .*
- (3) *The steady state growth rate increases in the productivity of researchers $\bar{\delta}$ and the share of human capital of workers in production α . It decreases in the share of human capital of scientists in goods production η .*

These comparative statics of R&D and income growth are, in principle, well known since the Romer (1990) model. The new element here is that R&D-output is not driven by population size,

as in the conventional model, but by human capital supplied by the best educated individuals in society. Aggregate human capital of scientists, according to (6), is determined by the steady-state education of the HH -individuals and their life expectancy. Since there are several steady states of education levels (cf. Section 3) there are potentially several steady states of the macroeconomy. The following proposition establishes the comparative statics of aggregate human capital of scientists.

PROPOSITION 11 (Supply of Scientists). *Aggregate human capital of scientists increases in the education level of high-ability individuals ϵ^{HH} , life expectancy τ_1 , and in the population share of scientists α^{HH} .*

4.2. Transitional Dynamics: The Prussian Way to Modern Growth. The conventional model of unified growth theory argues that technological progress triggered the onset of mass education and the take-off to modern growth (Galor, 2005). The model approximates the English case, in which the Industrial Revolution preceded and triggered mass education. Our model suggests that there is also an alternative way to modern growth, according to which compulsory schooling and mass education was initiated first (for religious or political reasons) and then subsequently triggered market R&D and an Industrial Revolution (Becker and Woessmann, 2011). Since Prussia is regarded as the pioneer in compulsory schooling (implemented 1763, before the Industrial Revolution around 1830), we call this the Prussian way to modern growth.

We assume that the economy is situated in the neighborhood of (but not necessarily exactly at) the steady state of elite education. Furthermore, the education parameters and population shares are such that a well chosen increase in compulsory schooling leads to a transition to mass education, c.f. Proposition 9. As a consequence of compulsory schooling two things happen. Firstly, as a consequence of the generally better level of education in society, life-expectancy increases, such that the next generation of HH - and HL -individuals prefers more education. Secondly, with ongoing improvements in life-expectancy, eventually the high-ability children without a family background of education aspire more than just voluntary education. They provide their children with a background of education such that the share of HH -types increases in the next generation. This means that the supply of scientists S_t increases for two reasons: improving longevity and upward educational mobility. Eventually the supply of scientists becomes sufficiently large such that market R&D becomes worthwhile and economic growth takes

off. Compulsory schooling thus has triggered market R&D and growth, potentially with a delay of several generations.

In the next section we show that the model, after a mild extension, is also capable to explain the English way, i.e. the reverse sequence of events whereby the Industrial Revolution eventually triggers mass education and the take-off to growth.

5. NUMERICAL EXPERIMENTS: TWO PATHS TO MODERN GROWTH

5.1. Sweden. We illustrate the Prussian way to modern growth for the case of Sweden, because Swedish data is available for most of the time series we are interested in. Sweden implemented compulsory schooling in 1840, later than Prussia, but certainly before the Industrial Revolution, which took place between 1870 and 1914 (Koblik, 1974). We assume that a unit period of time equals 20 years. This means that every 20 years a new generation enters the economic system and experiences a period of youth and young adulthood of 20 years. We assume that individuals enter the economic system at age of five, approximating the onset of the period of education in most developed countries. Individuals are thus conceptualized as being 25 years old when they enter the second stage of life. We set $\bar{\tau}_1 = 25$ and $\psi = 1.6$ in order to fit the life expectancy at age 25 in Sweden from 1840 to 2000 (data from MPI, 2007).

We equally divide the population in high-ability and low-ability types, $p = 0.5$. Low ability thus means below average ability. We set the discount factor γ to 0.6, reflecting an annual discount rate of 0.025 over a discount period of 20 years. We set $\theta_H = 1.6$, $\theta_L = 1.2$ and $\beta_H = 1.5$, $\beta_L = 0.78$. These parameters imply that the return on education per year of education lies in the interval $[0.047, 0.12]$, that is in $[\beta_L\theta_L/20, \beta_H\theta_H/20]$. This range covers about the range of empirical estimates of the return on education when family background is not controlled for (Card, 1999). Notice that education is measured in per generation, i.e. per 20 years. The average return on education controlling for family background is 0.07, i.e. $(1.6 + 1.2)/2/20$, and it is approximately equal to the average point estimate of the return to education (Card, 1999). At the steady-state, family background explains about 60 percent of the variation in the years of schooling, which captures approximately the empirical contribution of family background to education (Bjoerklund and Salvanes, 2010). We set initial population shares $\alpha^{HH} = \alpha^{HL} = 0.05$ in order to match average education before the onset of mass education, implying $\alpha^{LH} = \alpha^{LL} = 0.45$ such that 90 percent of the pre-modern population remains uneducated.

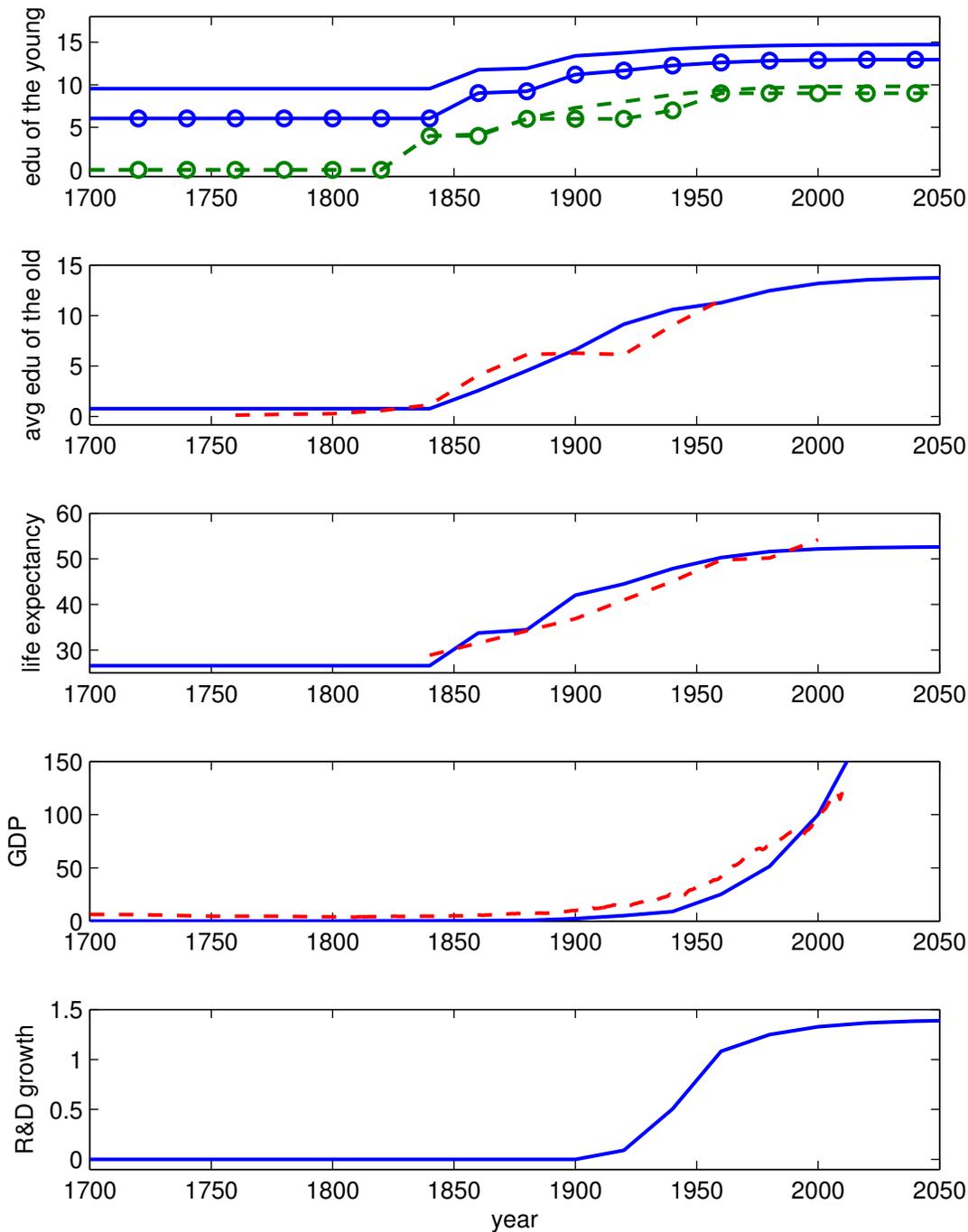
We set $B(0) = 1$ and $\lambda = 0.3$ such that there is almost no exponential growth in the pre-modern period (technology grows approximately arithmetically). We set the labor share of workers α and the labor share of scientists η to 0.25 such that the total labor share in goods production is 0.5. We set productivity in R&D ($\bar{\delta}$) and the stepping-on-toes externality (ϕ) such that (i) R&D based growth is triggered at the end of the 19th century and TFP growth during the second half of the 20th century is 1 to 1.5 percent annually. This provides the estimates $\bar{\delta} = 0.42$ and $\phi = 0.9$. A high value of ϕ is needed to keep the growth rate of R&D (i.e. TFP growth at the steady state) in a plausible range but is inconsequential for the other results. These values imply that at the end of the transition there are 8.6 percent of the population working in R&D.

Figure 4 shows the evolution of the economy from 1700 to 2050. The time before 1700 looks very much like 1700. The model rests at a state of elite education, the elite produces technological progress through learning-by-doing and there is mild arithmetic growth, which is hardly visible in the figure. The first panel from above shows the uptake of education for the four different classes of society: individuals with a family background of education are represented by blue solid lines and individuals without education in the family are represented by green dashed lines. High-ability types are indicated by plain lines and low-ability types are indicated by circled lines. Up to 1840 there is only elite education. Then, compulsory education is introduced in a piecemeal way: 4 years in 1840, 6 years in 1880, 7 years in 1940, and 9 years 1980 (Fischer et al., 2013).

The crucial outcome here is that compulsory education leads to a sufficient improvement of longevity such that the next generation of individuals with a family background of education increase their education even further. This enables the escape from the steady state of elite education and sets in motion the transition to modern growth. In 1860 the talented offspring of uneducated persons prefer for the first time more than just compulsory education. One generation later, their children are equipped with a background of education in the family (and be represented by the solid blue line). The virtuous cycle of upward mobility is set in motion.

Solid lines in the second panel from above show the implied average education of the adult workforce. These are individuals in the second period of their life and thus the take-off of education from the first panel (representing the young generation) is delayed by one period. We compare the model prediction with data on average schooling of the 16 to 65 years old population

FIGURE 4: TRANSITION TO MODERN GROWTH: SWEDEN



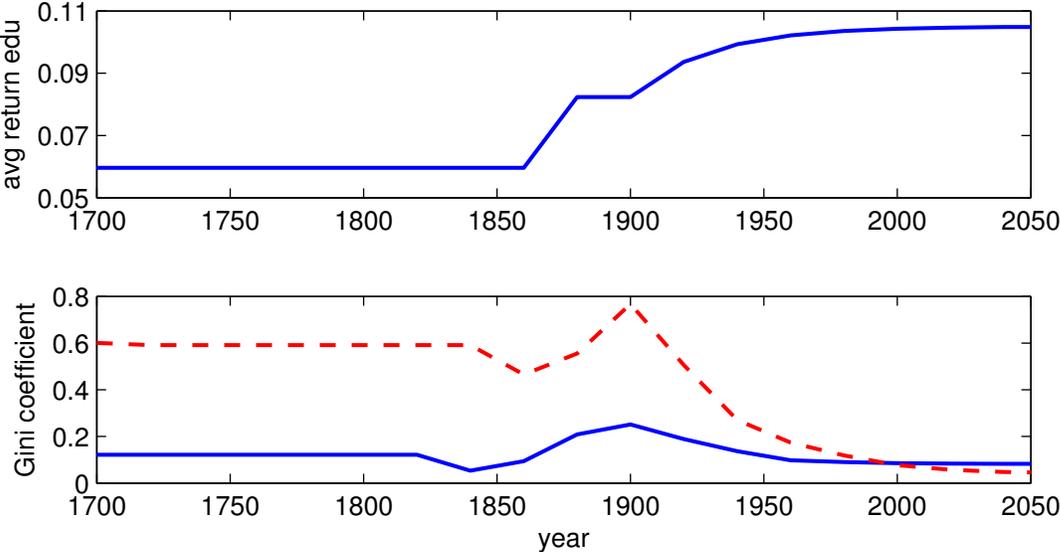
First panel: education of *HH*-types (solid), *HL*-types (solid-circled), *LH*-types (dashed) and *LL*-types (dashed-circled). The *LL*-line coincides with the evolution of compulsory schooling in Sweden. 2nd-5th panel: model (solid) and data (dashed). See Text for details.

in Sweden from 1760 to 1960 taken from de la Croix (2010) and represented by dotted lines. The model matches the data fairly closely. Average education is increasing from 0.15 years at the steady state of elite education to 11.5 years in 1960.

The third panel from above shows the evolution of life expectancy at age 25 in Sweden. Again the solid line represents the model prediction and the dashed line the historical data, which is available from 1850 to 2000 from MPI (2007). The model somewhat overestimates life expectancy during the first half of the 20th century but follows otherwise the data trajectory quite closely. The fourth panel shows the implied evolution of GDP per capita. The data series is from Maddison (2001). In order to compare both trajectories we normalized such that GDP per capita is 100 in the year 2000 according to both series. Again, the model traces the historical data reasonably closely.

The final panel in Figure 4 shows the time series for R&D output (which is equivalent to economic growth at the steady state). We see that compulsory schooling triggers the endogenous onset of R&D-based growth with a delay of 60 years. It takes 3 generations for the human capital of scientists to grow sufficiently large such that the corner solution for final goods production does no longer hold. The wage per unit of human capital of scientists becomes sufficiently low for market R&D to become profitable.

FIGURE 5: EVOLUTION OF RETURNS AND INEQUALITY



Bottom panel: solid lines: inequality in human capital, dashed lines: income inequality.

We next have a closer look at the implications for social mobility and inequality. The first panel in Figure 5 shows the evolution of the average return on education. The fact that individuals with a family background of only compulsory education or less are gradually dying out leads to a gradual increase of the average return in education in the population. It shows how

social mobility triggered by compulsory schooling may have contributed to the well known phenomenon of a gradually increasing return to education. Notice that the explanation here is a pure compositional effect. Innate abilities as well as the importance of family background remain constant over time but more and more individuals are coming from an educated family background.

The second panel in Figure 5 shows the evolution of education inequality (solid line) and income inequality (dashed line). Education inequality declines with the introduction of compulsory schooling in 1840 and a generation later income inequality declines as well. Subsequently, education inequality rises again because individuals from educated families increase their education as a response to the generally improved life expectancy while mass education remains (for now) at compulsory level. After the uptake of voluntary education of talented *LH*-individuals education inequality reaches a peak in 1900 and then declines as the pool of families with only compulsory education gradually empties out. The increase of income inequality is stronger because it is amplified by the onset of R&D in 1900 and the entailed increasing demand for scientists. Subsequently the increasing supply of scientists reduces inequality again such that the Kuznets curve emerges.

5.2. England. In order to capture the English case we modify the model slightly by integrating monetary costs of education. Although these costs depend on the length of the education period and not on income their presence makes the education decision income-dependent. At higher wages, individuals can more easily afford the monetary costs of education. Since these costs are independent of income their quantitative importance vanishes in a growing economy. They do not affect the long-run rate of modern growth. In principle we could have also integrated monetary costs in the Swedish case without changing any results. The complexity of the formal analysis of Section 3 and 4, however, would have been raised beyond tractability. Integrating monetary costs of education is relevant for the case of England because they are able to invert the sequence of impulse and response. In the English case, R&D is triggered by an education-independent shock, e.g. through changes in patent rules that improved the general accessibility of knowledge and thus the productivity of R&D. The mechanism is similar to the one developed by Galor and Weil (2000). The increase of the technology growth and income growth triggers eventually the endogenous onset of mass education. The difference here is that we take explicitly

into account that technology is advanced by R&D activities (and is not advanced just by learning-by-doing).

Specifically, monetary education costs change the budget constraint for the first period of life:

$$c_{t,0} = (1 - e_t)w_{t,0} - \xi e_t, \quad (20)$$

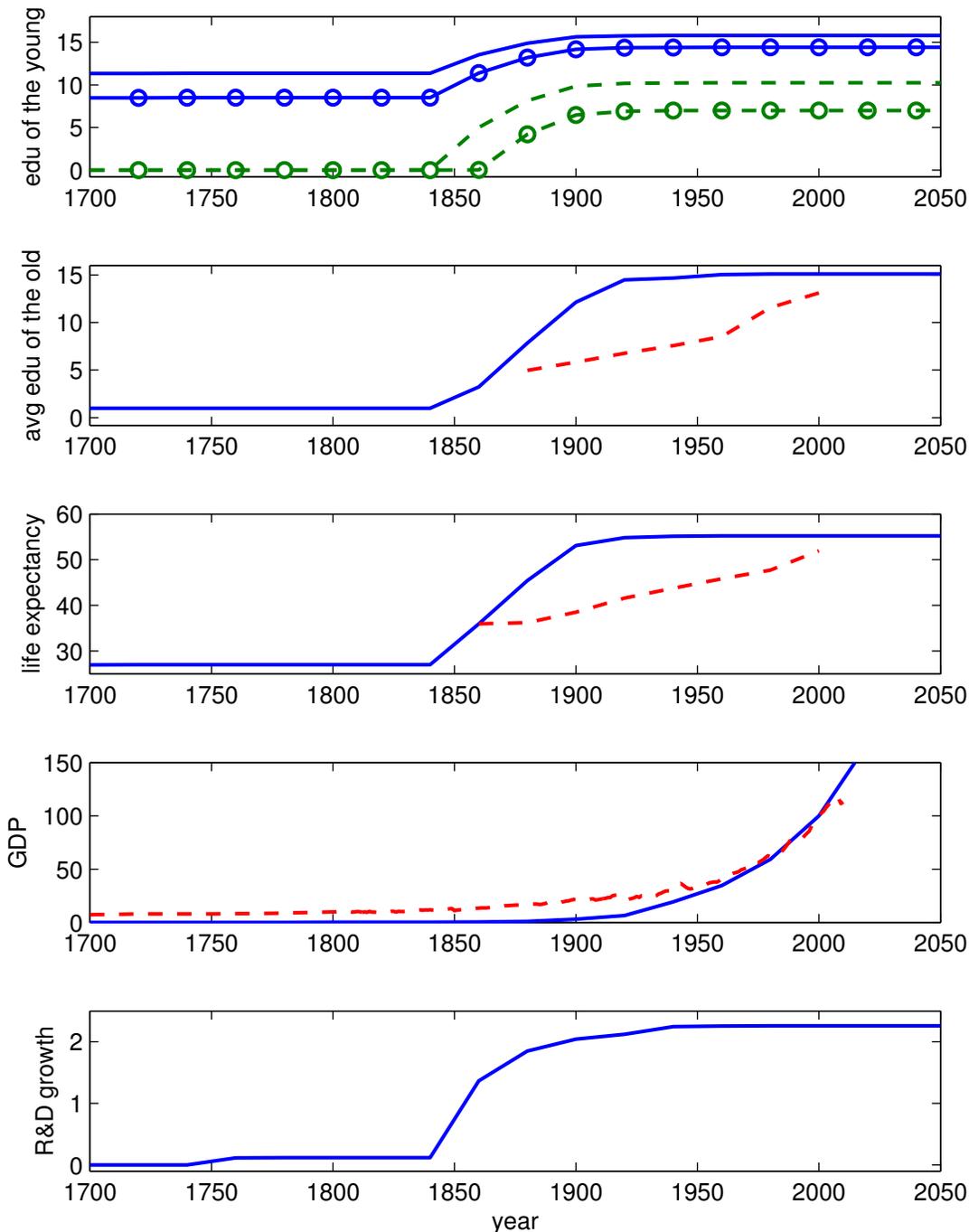
in which $\xi > 0$ denotes education costs per increment of education (per year of education). Since consumption during the first period of life has to be strictly positive, wages have to be sufficiently large for the uptake of education. Taking a potential corner solution into account, the optimal solution for voluntary education is obtained as

$$e_t = \max \left\{ 0, \frac{w_{t,0}}{w_{t,0} + \xi} - \frac{1}{\gamma\beta_j\theta_i\tau_{1,t}} - \sigma \right\}. \quad (21)$$

The fact that wages in turn depend on labor supply and on the education decision implies that the model turns into a non-linear system of equations that can no longer be solved analytically. The steady state of modern growth, however, is the same as before, since the importance of monetary costs of education vanishes asymptotically.

In setting up the model for England we tried to modify the Swedish case as little as possible. We set β_H to 1.8 (from 1.6), β_L to 0.774 (from 0.78), and the schooling cost parameter ξ to 38 in order to get the timing of the take-off of mass education about right. The higher difference between β_H and β_L captures the notion that family background is somewhat more important for education in England than in Sweden. Most importantly, we acknowledge that England in 1700 is about five times larger than Sweden (measured by aggregate GDP). We capture larger scale of the economy by assuming that every period a generation of mass 5 (rather than of mass 1) is born. Everything else is kept from the Swedish case. We assume that in 1760 productivity in R&D ($\bar{\delta}$) raises from 0.3, or some other low value, to 0.42, capturing, for example, the impact of the enlightenment or the revision of the patent law on diffusion of and access to knowledge (Mokyr, 2002). Such a change is ineffective for small Sweden (as shown above in Figure 4) but for large England it is sufficient to set in motion R&D-based growth. The reason is a scale effect. Given the larger workforce of scientists S_t , it is easier to set free scientists from goods production for R&D. Formally, $S_t > (1 + S_{t,A})^\phi \eta / [\bar{\delta}(\alpha + \eta)(1 - \alpha - \eta)]$ in (18) for England in 1760, but not for Sweden.

FIGURE 6: TRANSITION TO MODERN GROWTH: ENGLAND

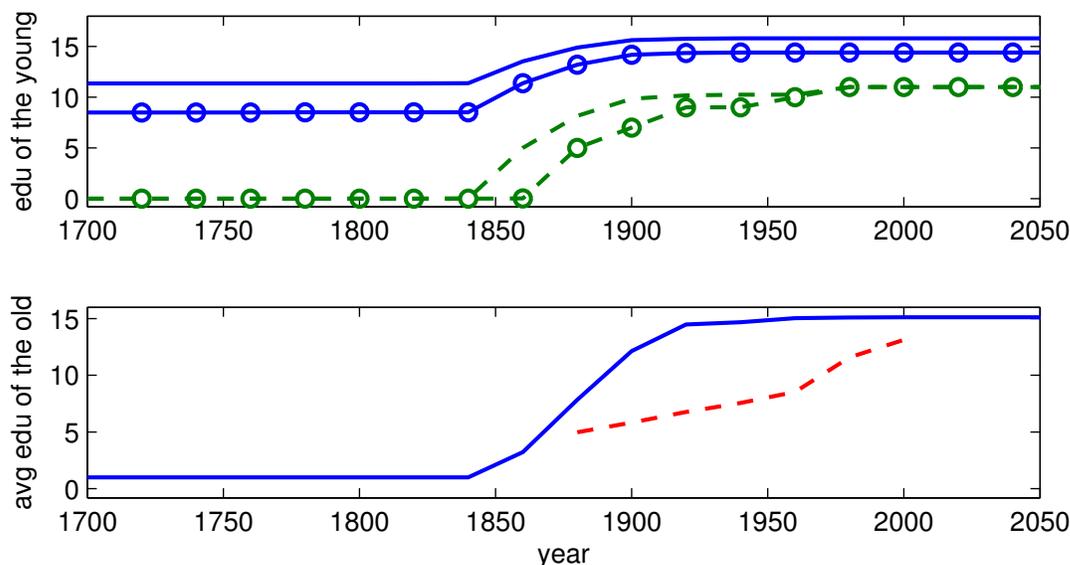


First panel: education o: *HH*-types (solid), *HL*-types (solid-circled), *LH*-types (dashed) and *LL*-types (dashed-circled). Exogenous Increase of R&D productivity in 1760 triggers endogenous mass education in 1840 (The numerical experiment ignores compulsory education).

For our first numerical experiment we ignore compulsory schooling in order to clearly demonstrate that technology growth has the power to trigger mass education. Results are shown in Figure 6. The impulse is visible in the bottom panel: R&D starts growing in 1760, albeit at

a small rate from today's perspective. R&D remains an elite activity for about one century when income increased enough to make education affordable for the masses. Now, from 1860 onwards with the rise of mass education, growth of R&D and income really take off and England experiences the Second Industrial Revolution. The second and third panel confront the model predictions of average education and life expectancy with the historical data, taken from MPI (2007) and Morrison and Murtin (2009), respectively. The model traces the data less well as in the Swedish case but gets the overall trends about right. Improvements are, of course, possible, by modifying further parameters of the model.

FIGURE 7: TRANSITION TO MODERN GROWTH: ENGLAND WITH COMPULSORY EDUCATION



First panel: education of *HH*-types (solid), *HL*-types (solid-circled), *LH*-types (dashed) and *LL*-types (dashed-circled). Exogenous Increase of R&D productivity in 1760 (as in Figure 6) triggers endogenous mass education in 1840 and onset of compulsory education in 1880.

Finally we introduce compulsory schooling. Results are shown in Figure 7 (the evolution of life expectancy, income, and R&D is not shown because it is virtually identical with Figure 6). Compulsory schooling is set to 5 years in 1880, 7 years in 1900, 9 years in 1920, 10 years in 1969, and to 11 years in 1980, approximating the historical implementation as closely as possible (Gillard, 2011). The main takeaway from Figure 7 is that the introduction of compulsory schooling has only minor effects on average education. It was introduced so lately that most of the children were anyway enrolled in school before it was made compulsory (Gillard, 2011, Chapter 3). The largest difference is observable for *LL* types (dashed-circled line) in the 20th century. The impact on average education however is relatively mild because the *LL*-types are

gradually vanishing from the population. The numerical experiment thus confirms for England the conventional view from unified growth theory (Galor, 2005): mass education was endogenously initiated by technology growth (here based on market R&D) rather than exogenously implemented. Notice that in the English case average schooling is low when the Industrial Revolution is initiated and remains low for almost another century. In the Swedish case, in contrast, average education is already quite high when the Industrial Revolution gets started.

CONCLUSION

In this paper we proposed a unified growth theory that takes into account family background as an important determinant, besides cognitive abilities, of the individual-specific return to education. We have shown that innovative pre-modern history can be conceptualized as a locally stable steady state of elite education. At the steady state, technological progress is brought forward by learning-by-doing activities of an educated elite (instead of by a growing uneducated workforce as in most related studies) but there are too few scholars and engineers for market R&D to be worthwhile. We have shown how compulsory education can instigate a transition to mass education, and trigger R&D-based growth. Industrialization, conceptualized as the gradual increase of R&D-based growth, is accompanied by first increasing and then decreasing income inequality and a generally increasing average return on education.

At the same time, the conventional mechanism from unified growth theory is also possible, namely that technological advances trigger mass education, which then accelerates R&D-based growth and leads to the convergence to a high and steady growth rate of income per capita. Which way to modern growth a society takes depends, among other things, on R&D productivity, population size, the composition of society, and the timing of compulsory schooling. In the Prussian case, compulsory schooling eventually initiates R&D-based growth whereas in the English case R&D-based growth initiates mass education. Moreover in the English case, average education can remain low for an extended period of time (e.g. a century) after the onset of R&D-based growth while it remains to be true that the high growth rates of innovation and income of the 20th century become only possible after technological growth has triggered mass education.

The insight that both ways to modern growth can be explained within one setup (with only little change of parameters and initial conditions) provides an inclusive, “unifying” view on unified

growth theory. The observation that mass education preceded and triggered industrialization (R&D-based growth) in some countries while in other countries the sequence of events was the other way round does not invalidate the conclusion that both market R&D and mass education are needed for a successful transition to modern growth.

APPENDIX A. OPTIMAL VALUE FOR EDUCATION

To obtain the globally optimal solution both the interior solution and the corner solutions have to be considered. The interior solution is derived from the first order condition of the optimization problem

$$\frac{\partial U}{\partial \epsilon_t} = -\frac{1}{1 - \epsilon_t - \sigma} + \theta_i \beta_j \gamma \tau_1 \stackrel{!}{=} 0 \Leftrightarrow \epsilon_t = 1 - \frac{1}{\gamma \tau_1 \theta_i \beta_j} - \sigma.$$

The second order condition reads

$$\frac{\partial^2 U}{\partial \epsilon_t^2} = -\frac{1}{(1 - \epsilon_t - \sigma)^2} < 0 \quad \forall \epsilon_t. \quad (22)$$

Hence, U is concave in $\epsilon_{t,0}$ and the local optimum is a local maximum. In addition to the interior solution there exists one possible corner solution with $\epsilon_{t,0} = 0$. Since U is concave and there exists a unique interior solution, we have

$$\frac{\partial U}{\partial \epsilon_{t,0}} \Big|_{\epsilon_{t,0}=0} \begin{cases} > 0 & \Rightarrow U(0) < U(\epsilon_{t,0}) \\ = 0 & \Rightarrow U(0) = U(\epsilon_{t,0}) \\ < 0 & \Rightarrow U(0) > U(\epsilon_{t,0}). \end{cases} \quad (23)$$

The interior solution is dominating the corner solution if and only if

$$\frac{\partial U}{\partial \epsilon_{t,0}} \Big|_{\epsilon_{t,0}=0} = -\frac{1}{1 - \sigma} + \gamma \tau_1 \theta_i \beta_j > 0 \Leftrightarrow 1 - \frac{1}{\gamma \tau_1 \theta_i \beta_j} - \sigma > 0. \quad (24)$$

Since U is concave in $\epsilon_{t,0}$, the global optimum is

$$\epsilon_{t,0} = \max \left\{ 0, 1 - \frac{1}{\gamma \tau_1 \theta_i \beta_j} - \sigma \right\}. \quad (25)$$

APPENDIX B. STEADY STATES OF EDUCATION

B.1. Elite Education. The two solutions of system (12) are

$$\begin{aligned} \epsilon_1^{HH} &= \frac{-\beta_H \theta_H \gamma \bar{\tau} (\theta_L + 2\alpha^{HL} \theta_H (-1 + \sigma) \psi + \theta_L (\alpha^{HL} + \alpha^{HH} (-1 + \sigma) + \sigma - \alpha^{HL} \sigma)) + \sqrt{\Delta}}{2\beta_H \gamma \theta_H \psi \bar{\tau} (\alpha^{HL} \theta_H + \alpha^{HH} \theta_L)} \\ \epsilon_1^{HL} &= \frac{-\beta_H \theta_L \gamma \bar{\tau} (\theta_H + 2\alpha^{HH} \theta_L (-1 + \sigma) \psi + \theta_H (\alpha^{HH} + \alpha^{HL} (-1 + \sigma) + \sigma - \alpha^{HH} \sigma)) + \sqrt{\Delta}}{2\beta_H \gamma \theta_L \psi \bar{\tau} (\alpha^{HL} \theta_H + \alpha^{HH} \theta_L)} \end{aligned}$$

and

$$\begin{aligned} \epsilon_2^{HH} &= \frac{-\beta_H \theta_H \gamma \bar{\tau} (\theta_L + 2\alpha^{HL} \theta_H (-1 + \sigma) \psi + \theta_L (\alpha^{HL} + \alpha^{HH} (-1 + \sigma) + \sigma - \alpha^{HL} \sigma)) - \sqrt{\Delta}}{2\beta_H \gamma \theta_H \psi \bar{\tau} (\alpha^{HL} \theta_H + \alpha^{HH} \theta_L)} \\ \epsilon_2^{HL} &= \frac{-\beta_H \theta_L \gamma \bar{\tau} (\theta_H + 2\alpha^{HH} \theta_L (-1 + \sigma) \psi + \theta_H (\alpha^{HH} + \alpha^{HL} (-1 + \sigma) + \sigma - \alpha^{HH} \sigma)) - \sqrt{\Delta}}{2\beta_H \gamma \theta_L \psi \bar{\tau} (\alpha^{HL} \theta_H + \alpha^{HH} \theta_L)} \end{aligned}$$

with

$$\Delta := \beta_H \theta_H \theta_L \gamma \bar{\tau} (-4\psi (\alpha^{HL} \theta_H + \alpha^{HH} \theta_L) + \beta_H \gamma \theta_H \theta_L \bar{\tau} (-1 + (-\alpha^{HH} - \alpha^{HL} + \sigma (-1 + \alpha^{HH} + \alpha^{HL}))) \psi^2). \quad (26)$$

Therefore, there exist up to two real, positive steady states if the discriminant Δ is non-negative, i.e.

$$\Delta \geq 0 \Leftrightarrow \beta_H \geq \frac{4\psi(\alpha^{HL}\theta_H + \alpha^{HH}\theta_L)}{\bar{\tau}\gamma\theta_H\theta_L(-1 + (-\alpha^{HH} - \alpha^{HL} + (-1 + \alpha^{HH} + \alpha^{HL})\sigma)\psi)^2}. \quad (27)$$

The critical values of β in Proposition 4 are given by

$$\bar{\beta}_{H,1} := \frac{(2\alpha^{HL}\theta_H + (\alpha^{HH} - \alpha^{HL})\theta_L)^2\psi}{\gamma\theta_H\theta_L(\alpha^{HL}(\theta_H - \theta_L) + \theta_L)\bar{\tau}(-1 - (1 - \sigma)(\alpha^{HH} + \alpha^{HL})\psi - \sigma\psi)^2} \quad (28)$$

$$\bar{\beta}_{H,2} := \frac{((-1 + \alpha^{HH} - \alpha^{HL})\theta_H - 2\alpha^{HH}\theta_L)^2\psi}{\gamma\theta_H\theta_L((1 - \alpha^{HH})\theta_H + \alpha^{HH}\theta_L)\bar{\tau}(-1 - (1 - \sigma)(\alpha^{HH} + \alpha^{HL})\psi - \sigma\psi)^2}. \quad (29)$$

B.2. Education of Talented. The two solutions of system (14) are

$$\epsilon_1^{HH} = \frac{-\beta_H\gamma\theta_H\bar{\tau}(\beta_L + 2\alpha^{LH}\beta_H(-1 + \sigma)\psi + \beta_L(\alpha^{LH} + \alpha^{HH}(-1 + \sigma) + \sigma - \alpha^{LH}\sigma)\psi) + \sqrt{\Delta}}{2\beta_H(\alpha^{LH}\beta_H + \alpha^{HH}\beta_L)\gamma\bar{\tau}\theta_H\psi}$$

$$\epsilon_1^{LH} = \frac{-\beta_L\gamma\theta_H\bar{\tau}(\beta_H + \alpha^{HH}\beta_L(-1 + \sigma)\psi + \beta_H(\alpha^{HH} + \alpha^{LH}(-1 + \sigma) + \sigma - \alpha^{HH}\sigma)\psi) + \sqrt{\Delta}}{2\beta_L(\alpha^{LH}\beta_H + \alpha^{HH}\beta_L)\gamma\theta_H\bar{\tau}\psi}$$

and

$$\epsilon_2^{HH} = \frac{-\beta_H\gamma\theta_H\bar{\tau}(\beta_L + 2\alpha^{LH}\beta_H(-1 + \sigma)\psi + \beta_L(\alpha^{LH} + \alpha^{HH}(-1 + \sigma) + \sigma - \alpha^{LH}\sigma)\psi) - \sqrt{\Delta}}{2\beta_H(\alpha^{LH}\beta_H + \alpha^{HH}\beta_L)\gamma\bar{\tau}\theta_H\psi}$$

$$\epsilon_2^{LH} = \frac{-\beta_L\gamma\theta_H\bar{\tau}(\beta_H + \alpha^{HH}\beta_L(-1 + \sigma)\psi + \beta_H(\alpha^{HH} + \alpha^{LH}(-1 + \sigma) + \sigma - \alpha^{HH}\sigma)\psi) - \sqrt{\Delta}}{2\beta_L(\alpha^{LH}\beta_H + \alpha^{HH}\beta_L)\gamma\theta_H\bar{\tau}\psi}$$

with

$$\Delta := \beta_H\beta_L\gamma\theta_H\bar{\tau}(-4\alpha^{HH}\beta_L\psi + \beta_H(-4\alpha^{LH}\psi + \beta_L\gamma\theta_H\bar{\tau}(-1 + (-\alpha^{HH} - \alpha^{LH} + (-1 + \alpha^{HH} + \alpha^{LH})\sigma)\psi)^2)) \quad (30)$$

Therefore, there exist up to two real, positive steady states if the discriminant Δ is non-negative, i.e.

$$\Delta \geq 0 \Leftrightarrow \theta_H \geq \frac{4(\alpha^{LH}\beta_H + \alpha^{HH}\beta_L)\psi}{\beta_H\beta_L\gamma\bar{\tau}(-1 + \alpha^{HH}(-1 + \sigma)\psi + \alpha^{LH}(-1 + \sigma)\psi - \sigma\psi)^2} \quad (31)$$

The critical values of θ in Proposition 6 are

$$\bar{\theta}_{H,1} := \frac{(2\alpha^{LH}\beta_H + \beta_L + \alpha^{HH}\beta_L - \alpha^{LH}\beta_L)^2\psi}{\beta_H\beta_L(\alpha^{LH}(\beta_H - \beta_L) + \beta_L)\gamma\bar{\tau}(-1 + \alpha^{HH}(-1 + \sigma)\psi + \alpha^{LH}(-1 + \sigma)\psi - \sigma\psi)^2} \quad (32)$$

$$\bar{\theta}_{H,2} := -\frac{((-1 + \alpha^{HH} - \alpha^{LH})\beta_H - 2\alpha^{HH}\beta_L)^2\psi}{\beta_H\beta_L((-1 + \alpha^{HH})\beta_H - \alpha^{HH}\beta_L)\gamma\bar{\tau}(-1 + \alpha^{HH}(-1 + \sigma)\psi + \alpha^{LH}(-1 + \sigma)\psi - \sigma\psi)^2}. \quad (33)$$

APPENDIX C. ADDITIONAL FORMULAS FROM SECTION 3.4

$$\begin{aligned} \tilde{\sigma}_{1,2} := & \frac{2(1 - \sigma)\beta_L^{3/2}\sqrt{\gamma\theta_H\bar{\tau}}\psi(\alpha^{HL}\theta_H + \alpha^{HH}\theta_L) + \beta_H\theta_L\sqrt{\beta_L\gamma\theta_H\bar{\tau}}(-1 + (1 - 2\alpha^{HL} + 2(-1 + \alpha^{HH} + \alpha^{HL})\sigma)\psi)}{2\psi\sqrt{\beta_L\gamma\theta_H\bar{\tau}}[(\beta_H(1 - \alpha^{HH}) + \alpha^{HH}\beta_L)\theta_L + \alpha^{HL}(\beta_L\theta_H - \beta_H\theta_L)]} \\ & \pm \frac{\sqrt{\beta_H\theta_L}\sqrt{-4\beta_L(\alpha^{HL}\theta_H + \alpha^{HH}\theta_L)\psi + \beta_H\theta_L(4(-1 + \alpha^{HH} + \alpha^{HL})\psi + \beta_L\gamma\theta_H\bar{\tau}(1 + \psi)^2)}}{2\psi\sqrt{\beta_L\gamma\theta_H\bar{\tau}}[(\beta_H(1 - \alpha^{HH}) + \alpha^{HH}\beta_L)\theta_L + \alpha^{HL}(\beta_L\theta_H - \beta_H\theta_L)]} \end{aligned}$$

$$\tilde{\sigma}_{3.4} = \frac{1}{2\beta_L\gamma\zeta_t\theta_H\bar{\tau}\psi} \left(-\beta_L\gamma\theta_H(\tau_{1,t} + \zeta_t(\sigma - 1)\bar{\tau}\psi) \pm \sqrt{\beta_L\gamma\theta_H(-4\zeta_t\bar{\tau}\psi + \beta_L\gamma\theta_H(\tau_{1,t} - \zeta_t(\sigma - 1)\bar{\tau}\psi)^2)} \right)$$

APPENDIX D. PROOF OF THE PROPOSITIONS

Proof of Proposition 1. The partial derivatives of the optimal value $\epsilon_{t,0} = 1 - \frac{1}{\gamma\tau_1\theta_i\beta_j} - \sigma$ are

$$\begin{aligned} \frac{\partial \epsilon_t}{\partial \sigma} &= -1 < 0, & \frac{\partial \epsilon_t}{\partial \tau_1} &= \frac{\gamma\theta_i\beta_j}{(\gamma\tau_1\theta_i\beta_j)^2} > 0, \\ \frac{\partial \epsilon_t}{\partial \beta_j} &= \frac{\gamma\tau_1\theta_i}{(\gamma\tau_1\theta_i\beta_j)^2} > 0, & \frac{\partial \epsilon_{t,0}}{\partial \theta_i} &= \frac{\gamma\tau_1\beta_j}{(\gamma\tau_1\theta_i\beta_j)^2} > 0. \end{aligned}$$

□

Proof of Proposition 3. The Jacobian of the right hand side of the difference system (12) in $(\epsilon^{HH}, \epsilon^{HL})$ is

$$\mathcal{J} = \begin{pmatrix} \frac{\alpha^{HH}\psi}{\bar{\tau}\gamma\beta_H\theta_H(1+(\alpha^{HH}\epsilon^{HH}+\alpha^{HL}\epsilon^{HL}+\sigma)\psi)^2} & \frac{\alpha^{HL}\psi}{\bar{\tau}\gamma\beta_H\theta_H(1+(\alpha^{HH}\epsilon^{HH}+\alpha^{HL}\epsilon^{HL}+\sigma)\psi)^2} \\ \frac{\alpha^{HH}\psi}{\bar{\tau}\gamma\beta_H\theta_L(1+(\alpha^{HH}\epsilon^{HH}+\alpha^{HL}\epsilon^{HL}+\sigma)\psi)^2} & \frac{\alpha^{HL}\psi}{\bar{\tau}\gamma\beta_H\theta_L(1+(\alpha^{HH}\epsilon^{HH}+\alpha^{HL}\epsilon^{HL}+\sigma)\psi)^2} \end{pmatrix}. \quad (34)$$

The Eigensystem of this matrix is

$$\lambda_1 = 0, \quad EV_1 = \left(-\frac{\alpha^{HL}}{\alpha^{HH}}, 1 \right)^\top \quad (35)$$

$$\lambda_2 = \frac{(\alpha^{HL}\theta_H + \alpha^{HH}\theta_L)\psi}{\bar{\tau}\gamma\beta_H\theta_H\theta_L(1+(\epsilon^{HH}\alpha^{HH} + \epsilon^{HL}\alpha^{HL} + \sigma)\psi)^2}, \quad EV_2 = \left(\frac{\theta_L}{\theta_H}, 1 \right)^\top. \quad (36)$$

Inserting the steady state values $(\epsilon_1^{HH}, \epsilon_1^{HL})$ we find that $0 < \lambda_2 < 1$. For the second steady state $(\epsilon_2^{HH}, \epsilon_2^{HL})$ we have $\lambda_2 \geq 1$. □

Proof of Proposition 4. It is obvious that the steady state values for education increase in the family background β_H and in the cognitive ability parameters θ_H and θ_L . The partial derivatives of the steady state values for education wrt. compulsory schooling σ are given by

$$\begin{aligned} \frac{\partial \epsilon_{1,2}^{HH}}{\partial \sigma} &= \frac{1}{2(\alpha^{HL}\theta_H + \alpha^{HH}\theta_L)} \left[-2\alpha^{HL}\theta_H - (1 + \alpha^{HH} - \alpha^{HL})\theta_L \pm \right. \\ &\quad \left. \pm \frac{(-1 + \alpha^{HH} + \alpha^{HL})\beta_H\gamma\theta_H\theta_L^2\bar{\tau}(-1 + (-\alpha^{HH} - \alpha^{HL} + (-1 + \alpha^{HH} + \alpha^{HL})\sigma)\psi)}{\sqrt{\Delta}} \right], \\ \frac{\partial \epsilon_{1,2}^{HL}}{\partial \sigma} &= \frac{1}{2(\alpha^{HL}\theta_H + \alpha^{HH}\theta_L)} \left[-2\alpha^{HH}\theta_L - (1 - \alpha^{HH} + \alpha^{HL})\theta_H \pm \right. \\ &\quad \left. \pm \frac{(-1 + \alpha^{HH} + \alpha^{HL})\beta_H\gamma\theta_H^2\theta_L\bar{\tau}(-1 + (-\alpha^{HH} - \alpha^{HL} + (-1 + \alpha^{HH} + \alpha^{HL})\sigma)\psi)}{\sqrt{\Delta}} \right] \end{aligned}$$

with

$$\Delta := \beta_H\theta_H\theta_L\gamma\bar{\tau}(-4\psi(\alpha^L H\theta_H + \alpha^{HH}\theta_L) + \beta_H\gamma\theta_H\theta_L\bar{\tau}(-1 + (-\alpha^{HH} - \alpha^{HL} + \sigma(-1 + \alpha^{HH} + \alpha^{HL}))\psi)^2).$$

In conclusion,

$$\frac{\partial \epsilon_{1,2}^{HH}}{\partial \sigma} > 0 \Leftrightarrow \beta_H < \frac{(2\alpha^{HL}\theta_H + \alpha^{HH}\theta_L - \alpha^{HL}\theta_L)^2\psi}{\gamma\theta_H\theta_L(\alpha^{HL}(\theta_H - \theta_L) + \theta_L)\bar{\tau}(-1 + \alpha^{HH}(-1 + \sigma)\psi + \alpha^{HL}(-1 + \sigma)\psi - \sigma\psi)^2},$$

$$\frac{\partial \epsilon_{1,2}^{HL}}{\partial \sigma} > 0 \Leftrightarrow \beta_H < \frac{((-1 + \alpha^{HH} - \alpha^{HL})\theta_H - 2\alpha^{HH}\theta_L)^2\psi}{\gamma\theta_H\theta_L((1 - \alpha^{HH})\theta_H + \alpha^{HH}\theta_L)\bar{\tau}(-1 + \alpha^{HH}(-1 + \sigma)\psi + \alpha^{HL}(-1 + \sigma)\psi - \sigma\psi)^2}.$$

□

Proof of Propositions 8 and 9. We assume that compulsory schooling increases in period t moderately from σ to $\sigma + \tilde{\sigma}$. Thus, in this period we have

$$\epsilon_t^{HH} = \epsilon_{t-1}^{HH} - \tilde{\sigma} > 0, \quad \epsilon_t^{HL} = \epsilon_{t-1}^{HL} - \tilde{\sigma} > 0 \quad (37)$$

due to the time lag in life expectancy. Applying equation (13) in period t , provides

$$\beta_L < \frac{1}{(1 - \sigma - \tilde{\sigma})\gamma\theta_H\tau_{t,1}}. \quad (38)$$

Thus, $\epsilon_t^{LH} = \epsilon_t^{LL} = 0$. Life expectancy in the following period increases due to compulsory schooling, i.e.

$$\tau_{t,t+1} = \bar{\tau} (1 + \psi (\alpha_t^{HH}(\epsilon_{t-1}^{HH} - \tilde{\sigma}) + \alpha_t^{HL}(\epsilon_{t-1}^{HL} - \tilde{\sigma}) + \sigma + \tilde{\sigma})) = \tau_{t,1} + \bar{\tau}\psi\tilde{\sigma}(1 - \alpha_t^{HH} - \alpha_t^{HL}).$$

An increase of compulsory schooling does not necessary lead to a transition towards mass education, it depends on the choice of the parameters. Furthermore, there are two possibilities for the takeoff, namely that the $\{\beta_L, \theta_H\}$ -individuals prefer a higher education directly in the next period, or that these individuals remain for now at the corner solution without higher education. In that case the takeoff may take place some periods later due to the successive increases in life expectancy. This results from the immediate increase in life expectancy, $\bar{\tau}\psi\tilde{\sigma}(1 - \alpha_t^{HH} - \alpha_t^{HL})$, which in turn raises education of the HH - and HL -individuals. This in turn increases life expectancy further and so on.

In the first case we have $\epsilon_{t+1}^{LH} > 0$. From equation (13) it follows that

$$\beta_L > \frac{1}{(1 - \sigma - \tilde{\sigma})\gamma\theta_H\tau_{1,t+1}} = \frac{1}{(1 - \sigma - \tilde{\sigma})\gamma\theta_H (\tau_{t,1} + \bar{\tau}\psi\tilde{\sigma}(1 - \alpha_t^{HH} - \alpha_t^{HL}))}. \quad (39)$$

Defining $\zeta_t := 1 - \alpha_t^{HH} - \alpha_t^{HL}$, the condition can be reformulated as

$$\beta_L\gamma\theta_H [-\tilde{\sigma}^2\bar{\tau}\psi\zeta_t - \tilde{\sigma}(-\bar{\tau}\psi\zeta_t + \bar{\tau}\sigma\psi\zeta_t + \tau_{t,1}) + \tau_{t,1}(1 - \sigma)] - 1 > 0. \quad (40)$$

The expression on the left hand side is strictly concave in $\tilde{\sigma}$. The zeros of the this expression are given by

$$\tilde{\sigma}_{1,2} = \frac{1}{2\beta_L\gamma\zeta_t\theta_H\bar{\tau}\psi} \left(-\beta_L\gamma\theta_H(\tau_{1,t} + \zeta_t(\sigma - 1)\bar{\tau}\psi) \pm \sqrt{\beta_L\gamma\theta_H(-4\zeta_t\bar{\tau}\psi + \beta_L\gamma\theta_H(\tau_{1,t} - \zeta_t(\sigma - 1)\bar{\tau}\psi)^2)} \right)$$

Thus, for ϵ_{t+1}^{LH} to be positive we need $\tilde{\sigma} \in (\tilde{\sigma}_1, \tilde{\sigma}_2) \cap (0, 1 - \sigma)$.

The second possible solution is that individuals with high cognitive ability and low educated parents remain at the corner solution in period $t + 1$, i.e. $\epsilon_{t+1}^{LH} = 0$. In order to compute under which conditions the shock of compulsory schooling leads to a delayed take-off, assume that

$\epsilon_t^{LH} = \epsilon_t^{LL} = 0 \forall t$. Due to the moderate increase and because the second steady state is unstable the economy is in transition to the first, stable steady state. In that particular steady state, steady state life expectancy is

$$\tau_1(\sigma + \tilde{\sigma}) = \frac{\beta_H \gamma \theta_H \theta_L \bar{\tau} (1 + (\sigma + \tilde{\sigma} - 1)(\alpha^{HH} + \alpha^{HL}) - (\sigma + \tilde{\sigma})) + \sqrt{\Delta}}{2\beta_H \gamma \theta_H \theta_L} \quad (41)$$

with

$$\Delta := \beta_H \gamma \theta_H \theta_L \bar{\tau} \left[-4(\alpha^{HL} \theta_H + \alpha^{HH} \theta_L) \psi + \beta_H \gamma \theta_H \theta_L \bar{\tau} (-1 + ((\sigma + \tilde{\sigma} - 1)(\alpha^{HH} + \alpha^{HL}) - (\sigma + \tilde{\sigma})) \psi)^2 \right].$$

Furthermore, life expectancy at the newly attained steady state is larger than before the compulsory schooling shock. Given that the education class is not left during the transition phase, it must hold that

$$\beta_L < \frac{1}{(1 - \sigma - \tilde{\sigma}) \gamma \theta_H \tau_1(\sigma + \tilde{\sigma})}, \quad (42)$$

providing

$$\frac{\beta_L(1 - \sigma - \tilde{\sigma})}{2\beta_H \theta_L} \left[\beta_H \gamma \theta_H \theta_L \bar{\tau} (1 + ((\alpha_t^{HH} + \alpha_t^{HL})(1 - \sigma - \tilde{\sigma}) + \sigma + \tilde{\sigma}) \psi) + \sqrt{\Delta} \right] - 1 < 0. \quad (43)$$

The expression on the left hand side is strictly concave in $\tilde{\sigma}$ with zeros

$$\begin{aligned} \tilde{\sigma}_{1,2} = & \frac{2(1 - \sigma) \beta_L^{3/2} \sqrt{\gamma \theta_H \bar{\tau}} \psi (\alpha^{HL} \theta_H + \alpha^{HH} \theta_L) + \beta_H \theta_L \sqrt{\beta_L \gamma \theta_H \bar{\tau}} (-1 + (1 - 2\alpha^{HL} + 2(-1 + \alpha^{HH} + \alpha^{HL}) \sigma) \psi)}{2\psi \sqrt{\beta_L \gamma \theta_H \bar{\tau}} [(\beta_H(1 - \alpha^{HH}) + \alpha^{HH} \beta_L) \theta_L + \alpha^{HL} (\beta_L \theta_H - \beta_H \theta_L)]} \\ & \pm \frac{\sqrt{\beta_H \theta_L} \sqrt{-4\beta_L (\alpha^{HL} \theta_H + \alpha^{HH} \theta_L) \psi + \beta_H \theta_L (4(-1 + \alpha^{HH} + \alpha^{HL}) \psi + \beta_L \gamma \theta_H \bar{\tau} (1 + \psi)^2)}}{2\psi \sqrt{\beta_L \gamma \theta_H \bar{\tau}} [(\beta_H(1 - \alpha^{HH}) + \alpha^{HH} \beta_L) \theta_L + \alpha^{HL} (\beta_L \theta_H - \beta_H \theta_L)]}. \end{aligned}$$

For $\tilde{\sigma} \in (\tilde{\sigma}_1, \tilde{\sigma}_2) \cap (0, 1 - \sigma)$ condition (42) is not fulfilled and there will be a transition to mass education. \square

Proof of Proposition 10.

(1) The steady state growth factor of GDP per capita is given by

$$\frac{y_{t+1}}{y_t} = \frac{Y_{t+1}}{Y_t} \frac{L_t}{L_{t+1}} = \frac{Y_{t+1}}{Y_t},$$

since the population is constant at the steady state. The growth factor of aggregate GDP is

$$\begin{aligned} \frac{Y_{t+1}}{Y_t} &= \frac{B_{t+1}}{B_t} \left(\frac{H_{t+1}}{H_t} \right)^\alpha \left(\frac{S_{t+1,Y}}{S_{t,Y}} \right)^\eta \frac{A_{t+1}}{A_t} \left(\frac{x_{t+1}}{x_t} \right)^{1-\alpha-\eta} \\ &= \frac{B_{t+1}}{B_t} \left(\frac{H_{t+1}}{H_t} \right)^\alpha \left(\frac{S_{t+1,Y}}{S_{t,Y}} \right)^\eta \frac{A_{t+1}}{A_t} \left(\frac{B_{t+1}}{B_t} \left(\frac{H_{t+1}}{H_t} \right)^\alpha \left(\frac{S_{t+1,Y}}{S_{t,Y}} \right)^\eta \right)^{1-\alpha-\eta} \\ &= \frac{A_{t+1}}{A_t}, \end{aligned}$$

because B , H and S_Y are constant at the steady state.

(2) Aggregate human capital of scientists is given by

$$S_t = S_{t,A} + S_{t,Y} = S_{t,A} + \frac{\eta(1 + S_{t,A})^\phi}{\delta(\alpha + \eta)(1 - \alpha - \eta)}. \quad (44)$$

Hence, the steady state value of $S_{t,A}$ is determined by the steady state value of S_t .

- (3) The partial derivative of the growth rate (19) with respect to human capital of scientists $S_{t,A}$ is given by

$$\frac{\partial g}{\partial S_{t,A}} = \bar{\delta} (1 + S_{t,A})^{-1-\phi} (1 + (1 - \phi)S_{t,A}) > 0,$$

since $\phi \in [0, 1]$. Equation (44) defines an implicit function

$$G(x, S_{t,A}) := S_t - S_{t,A} - \frac{\eta(1 + S_{t,A})^\phi}{\bar{\delta}(\alpha + \eta)(1 - \alpha - \eta)} \quad \text{for } x = S_t, \eta, \alpha, \phi, \bar{\delta}.$$

Applying the implicit function theorem we have

$$\begin{aligned} \frac{\partial S_{t,A}}{\partial S_t} &= -\frac{\frac{\partial G}{\partial S_t}}{\frac{\partial G}{\partial S_{t,A}}} > 0, \\ \frac{\partial S_{t,A}}{\partial \eta} &= -\frac{\frac{\partial G}{\partial \eta}}{\frac{\partial G}{\partial S_{t,A}}} = \frac{\frac{\alpha(1-\alpha)+\eta^2}{\bar{\delta}(-1+\alpha+\eta)^2(\alpha+\eta^2)}}{-1 - \phi \frac{\eta}{\bar{\delta}(\alpha+\eta)(1-\alpha-\eta)} (1 + S_{t,A})^{-1-\phi}} < 0, \\ \frac{\partial S_{t,A}}{\partial \bar{\delta}} &= -\frac{\frac{\partial G}{\partial \bar{\delta}}}{\frac{\partial G}{\partial S_{t,A}}} = -\frac{\frac{\eta}{\bar{\delta}^2(\alpha+\eta)(1-\alpha-\eta)} (1 + S_{t,A})^\phi}{-1 - \phi \frac{\eta}{\bar{\delta}(\alpha+\eta)(1-\alpha-\eta)} (1 + S_{t,A})^{-1-\phi}} > 0, \\ \frac{\partial S_{t,A}}{\partial \phi} &= -\frac{\frac{\partial G}{\partial \phi}}{\frac{\partial G}{\partial S_{t,A}}} = -\frac{(1 + S_{t,A})^\phi \log(1 + S_{t,A})}{-1 - \phi \frac{\eta}{\bar{\delta}(\alpha+\eta)(1-\alpha-\eta)} (1 + S_{t,A})^{-1-\phi}} > 0 \end{aligned}$$

and

$$\frac{\partial S_{t,A}}{\partial \alpha} = -\frac{\frac{\partial G}{\partial \alpha}}{\frac{\partial G}{\partial S_{t,A}}} = \frac{\frac{\eta(1-2(\alpha+\eta))}{\bar{\delta}(-1+\alpha+\eta)^2(\alpha+\eta)^2}}{-1 - \phi \frac{\eta}{\bar{\delta}(\alpha+\eta)(1-\alpha-\eta)} (1 + S_{t,A})^{-1-\phi}} > 0 \Leftrightarrow \eta(1-2(\alpha+\eta)) > 0 \Leftrightarrow \frac{1}{2} > \alpha + \eta.$$

Therefore, the comparative statics of the steady state growth rate are

$$\begin{aligned} \frac{\partial g}{\partial \eta} &= \frac{\partial g}{\partial S_{t,A}} \frac{\partial S_{t,A}}{\partial \eta} < 0, \\ \frac{\partial g}{\partial \eta} &= \frac{\partial g}{\partial S_{t,A}} \frac{\partial S_{t,A}}{\partial \alpha} > 0 \Leftrightarrow \frac{1}{2} > \alpha + \eta, \\ \frac{\partial g}{\partial \bar{\delta}} &= \frac{\partial g}{\partial S_{t,A}} \frac{\partial S_{t,A}}{\partial \bar{\delta}} > 0 \end{aligned}$$

and

$$\frac{\partial g}{\partial \phi} = \bar{\delta} \left[\frac{\partial S_{t,A}}{\partial \phi} (1 + S_{t,A})^{-\phi} + S_{t,A} \left(-(1 + S_{t,A})^{-\phi} \log(1 + S_{t,A}) \right) \frac{\partial S_{t,A}}{\partial \phi} \right] > 0 \Leftrightarrow 1 > S_{t,A} \log(1 + S_{t,A}).$$

□

REFERENCES

- Acemoglu, D., Johnson, S., 2007, Disease and development: The effect of life expectancy on economic growth, *Journal of Political Economy* 115, 925-985.
- Aghion, P. and Howitt, P., 1992, A model of growth through creative destruction. *Econometrica*, 60, 323-702.
- Aghion, P. and Howitt, P., 2009, *The economics of growth*, MIT Press.
- Becker, S.O., Hornung, E., and Woessmann, L., 2011, Education and catch-up in the Industrial Revolution, *American Economic Journal: Macroeconomics* 3, 92-126.
- Ben-Porath, Y., 1967, The production of human capital and the life cycle of earnings, *Journal of Political Economy* 75, 352-365.
- Bjoerklund, A. and Salvanes, K.G., 2010, Education and family background: Mechanisms and policies, in: *Handbook in Economics of Education*, 201-247.
- Boucekinne, R., de la Croix, D., and Licandro, O., 2002, Vintage human capital, demographic trends, and endogenous growth, *Journal of Economic Theory* 104, 340-375.
- Boucekinne, R., de la Croix, D., and Licandro, O., 2003, Early mortality declines at the dawn of modern growth, *Scandinavian Journal of Economics* 105, 401-418.
- Card, D., 1999, The causal effect of education on earnings, *Handbook of Labor Economics* Vol. 3, 1801-1863, Elsevier, Amsterdam.
- Cervellati, M. and Sunde, U., 2005, Human capital formation, life expectancy, and the process of development, *American Economic Review* 95, 1653-1672.
- Cervellati, M. and Sunde, U., 2011, Life expectancy and economic growth: the role of the demographic transition, *Journal of Economic Growth* 16, 99-133.
- Cervellati, M. and Sunde, U., 2013, Life Expectancy, Schooling, and Lifetime Labor Supply: Theory and Evidence Revisited, *Econometrica* 81, 2055-2086.
- Cutler, D.M., Lleras-Muney, A. and Vogl, T., 2011, Socioeconomic status and health: dimensions and mechanisms, in: Glied, S. and Smith, P.C. (eds), *The Oxford Handbook of Health Economics*, Oxford University Press, Oxford, pp. 124-163.
- Cipolla, C.M., 1969, *Literacy and Development in the West*, Penguin Books, Harmondsworth.
- De la Croix D., 2010, Adult longevity and economic take-off: from Malthus to Ben-Porath, in: *Institutional and Social Dynamics of Growth and Distribution*, Neri Salvadori ed., Edward Elgard, chap 8, 172-190.
- Doepke, M., 2004, Accounting for fertility decline during the transition to growth. *Journal of Economic Growth* 9: 347-383.
- Eckstein, Z. and Zilcha, I., 1994, The effects of compulsory schooling on growth, income distribution and welfare, *Journal of Public Economics* 54, 339-359.

- Feinstein, L., Duckworth, K., Sabates, R., 2008, Education and the family: Passing success across the generations. Routledge.
- Fischer, M., Karlsson, M., and Nilsson, T., 2013, Effects of compulsory schooling on mortality: Evidence from Sweden, *International Journal of Environmental Research and Public Health* 10, 3596-3618.
- Galor, O., 2005, From stagnation to growth: unified growth theory, in: P. Aghion and S. Durlauf, *Handbook of Economic Growth* Vol 1A, Amsterdam, North-Holland.
- Galor, O., 2007, *Discrete dynamical systems*, Springer.
- Galor, O., 2011, *Unified Growth Theory*, Princeton University Press.
- Galor, O. and D.N. Weil, 2000, Population, technology and growth: From the Malthusian regime to the demographic transition and beyond, *American Economic Review* 110, 806–828.
- Galor, O. and O. Moav, 2002, Natural selection and the origin of economic growth, *Quarterly Journal of Economics* 117, 1133-1192.
- Galor, O., and Moav, O., 2006, Das human-kapital: A theory of the demise of the class structure, *Review of Economic Studies* 73, 85-117.
- Galor, O. and A. Mountford, 2008, Trading population for productivity: theory and evidence, *Review of Economic Studies* 75, 1143-1179.
- Gancia, G., and Zilibotti, F., 2005, Horizontal innovation in the theory of growth and development, in: P. Aghion and S. Durlauf, *Handbook of Economic Growth* Vol 1A, Amsterdam, North-Holland., 111-170.
- Gillard, D., 2011, *Education in England: a Brief History*, www.educationengland.org.uk/history.
- Green, A., 1990, *Education and State Formation*, St. Martins Press, New York.
- Hansen, G. D., and Prescott, E. C., 2002, Malthus to Solow, *American Economic Review* 92(4), 1205-1217.
- Hansen, C.W. and Lønstrup, L., 2012, Can higher life expectancy induce more schooling and earlier retirement?, *Journal of Population Economics*, 25(4), 1249-1264.
- Hansen, C.W., 2013, Life expectancy and human capital: evidence from the international epidemiological transition, *Journal of Health Economics* 32, 1142-1152.
- Hazan, M., 2009, Longevity and lifetime labor supply: Evidence and implications, *Econometrica* 77, 1829-1863.
- Jayachandran, S., and Lleras-Muney, A., 2009, Life expectancy and human capital investments: Evidence from maternal mortality declines, *Quarterly Journal of Economics* 124, 349-397.
- Jones, C.I., 2001, Was an industrial revolution inevitable? Economic growth over the very long run. *Advances in Macroeconomics*, 1, 1–43.

- Jones, C.I., 2005, R&D-based Models of Economic Growth. *Journal of Political Economy* 103, 759-783.
- Koblik, S., 1975, Sweden's Development from Poverty to Affluence 1750-1970.
- Kögel, T. and Prskawetz, A., 2001, Agricultural productivity growth and escape from the Malthusian trap, *Journal of Economic Growth* 6, 337-357.
- Kremer, M. (1993). Population growth and technological change: one million B.C. to 1990. *Quarterly Journal of Economics*, 108, 681-716.
- Kuznets, S., 1955, Economic growth and income inequality, *American Economic Review* 45, 1-28.
- Maddison, A., 2001, *The World Economy: A Millennial Perspective*. Development Center Studies, OECD, Paris.
- Madsen, J.B., Ang, J.B. and Banerjee, R., 2010, Four centuries of British economic growth: the roles of technology and population. *Journal of Economic Growth*, 15, 1-28.
- Meghir, C., and Palme, M., 2005, Educational Reform, Ability, and Family Background. *American Economic Review*, 95(1), 414-424.
- Mitch, D., 1992, *The Rise of Popular Literacy in Victorian England: The Influence of Private Choice and Public Policy*, University of Pennsylvania Press, Philadelphia.
- Mincer, J., 1974, Schooling and Earnings. *Schooling, Experience, and Earnings*. Columbia University Press
- Morrison, C. and Murtin, F., 2009, The century of education, *Journal of Human Capital* 3, 1-42.
- Mokyr, J., 2002, *The Gifts of Athena*. Princeton University Press, Princeton.
- Mokyr, J., 2005, Long-term economic growth and the history of technology. *Handbook of economic growth* 1, 1113-1180.
- MPI, 2007, The Human Life Table Database, Max Planck Institute for Demographic Research, Rostock (<http://www.lifetable.de/>)
- Peretto, P.F., 1998, Technological change and population growth. *Journal of Economic Growth*, 3, 283-311.
- Romer, P.M., 1990, Endogenous technological change. *Journal of Political Economy*, 98, S71-S102.
- Segerstrom, P., 1998, Endogenous growth without scale effects. *American Economic Review*, 88, 1290-1310.
- Sokoloff, K. L., and Engerman, S. L., 2000, History lessons: institutions, factor endowments, and paths of development in the new world, *Journal of Economic Perspectives* 14(3), 217-232.

- Strulik, H. and Weisdorf, J., 2008, Population, food, and knowledge: A simple unified growth model. *Journal of Economic Growth*, 13, 169-194.
- Strulik, H., Prettnner, K., and Prskawetz, A., 2010, The past and future of knowledge-based growth, *Journal of Economic Growth* 18, 411-437.
- Strulik, H. and Werner, L., 2013, 50 is the New 30 – Long-run Trends of Schooling and Retirement Explained by Human Aging, Discussion Paper, University of Goettigen.