Appendix A

A.1. The two agricultural revolutions

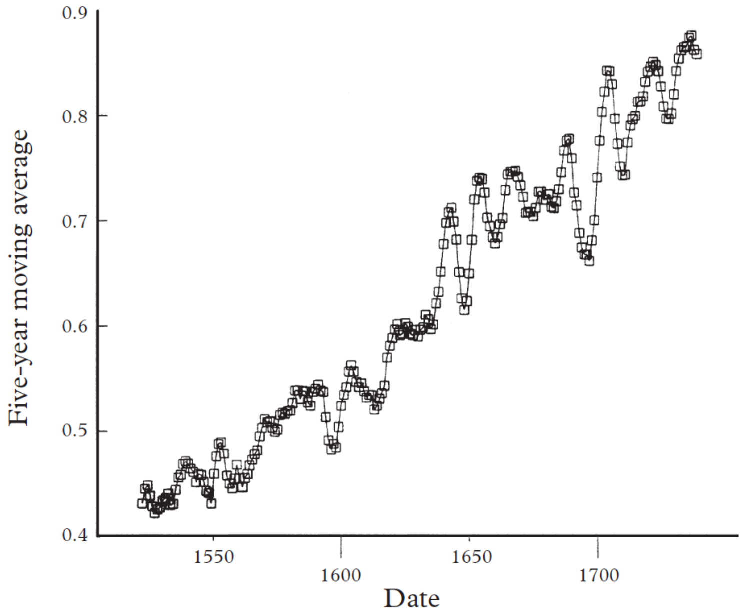

Figure 1A. The first agricultural revolution: index of farm output, 1520-1739. Source, Allen (1999)

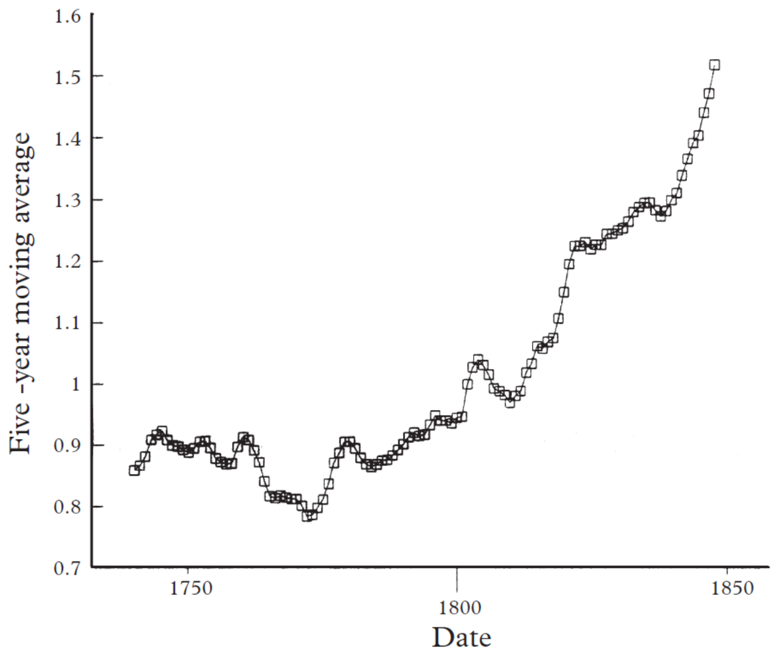

Figure 2A. The second agricultural revolution: index of farm output, 1740-1850. Source, Allen (1999)

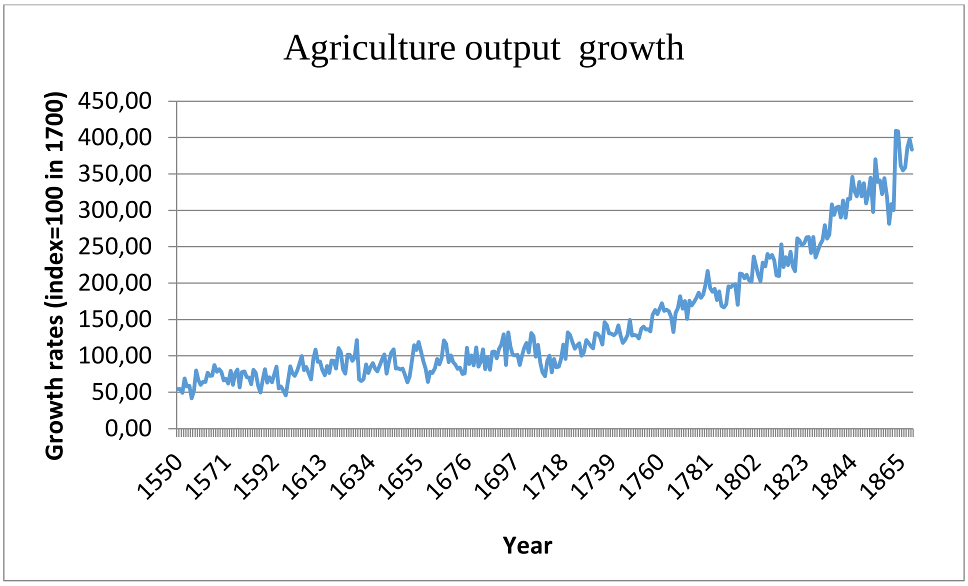

Figure 3A: Total Agricultural Output, normalized to 100 in the year 1700. Source: Thomas and Dimsdale (2017), authors' own calculations.

A.2. Assumptions concerning the production function in the industrial sector

The model explains the macro-dynamics during the industrial revolution through several aspects. One of these aspects is the onset of the industrial sector, which is closed during the Malthusian period and opens afterwards giving rise to the Industrial Revolution. This can be ensured by guaranteeing that when employment in the agricultural sector decreases, the demand for labor increases without bound while productivity in the industrial sector remains finite. This fact ensures that the industrial sector will open if and only if labor productivity in this sector exceeds the marginal productivity of labor in the agricultural sector, under the assumption that all labor is employed in the agricultural sector (see also Lemma 1). Thus, some assumptions have to be made regarding the production function of the industrial sector. The assumption taken to ensure that this sector is closed during the Malthusian period, as historically happened, is the latent production function $Y_t^I = A_t^I H_t$ because in this equation marginal labor productivity does not rely on the labor level. Since it is only a latent function, it does not directly affect the dynamics of the economy regarding output or labor decisions.

After the onset of the industrial revolution, a traditional Cobb-Douglas production function in the industrial sector $Y_t^I = (A_t^I)^{1-\theta}(H_t)^{\theta}$ is assumed. This will be the effective production function of the industrial sector in the economy.

There are two points that make a strong case for this approach. The first is the fact that in the pre-industrial regime, the production function of the industrial sector is only latent. It is merely a benchmark used to quantify the gains of productivity necessary to make the industrial sector economically viable in the economy. In the end, the only valid production function is the one employed after the industrialization takes place and is also the one affecting the dynamics of output and employment in each sector. The second point is the presence in the literature that supports the existence of these structural breaks in the ways of production and of technology advancements for different economies at different periods of time in our modern history (see, for example, Hansen (2001) and Miyagiwa and Papageorgiou (2007)). As these papers show, aggregate elasticity of substitution evolves with the process of economic development, which gives further support to this assumption as it appears to be a valid way of modeling the enormous change to the economy that took place at this point in time in some European countries.

A.3. Sectoral distribution of labor force

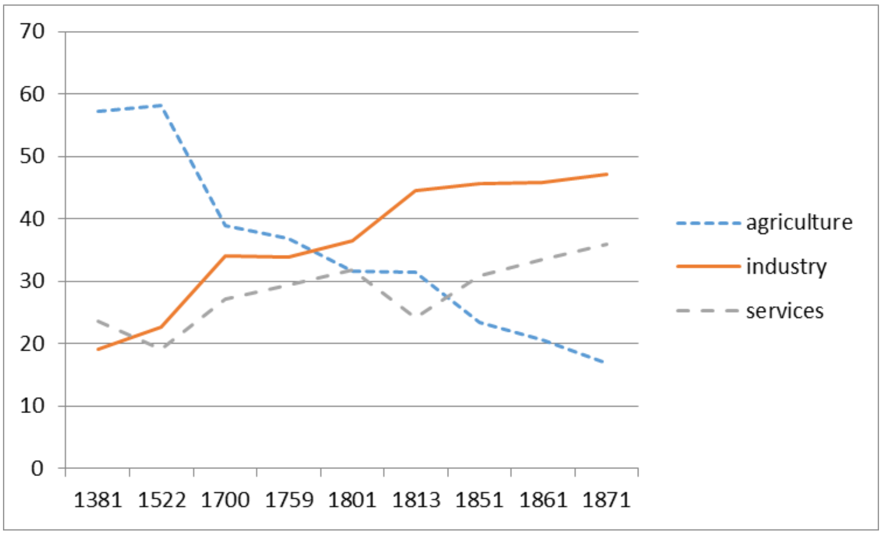

Figure 4A: Sectoral distribution of labor force between 3 sectors, from 1381-1871. Source: Broadberry et al. (2013)

Table 1A. Sectoral distribution of labor force

Sector / **Year**	**1381**	**1522**	**1700**	**1759**	**1801**	**1813**	**1851**	**1861**	**1871**
Agriculture	57.2	58.1	38.9	36.8	31.7	31.4	23.5	20.6	16.9
Industry	19.2	22.7	34	33.9	36.4	44.5	45.7	45.9	47.1
Services	23.6	19.2	27.2	29.3	31.9	24.1	30.9	33.5	36
Share in Industry / Share in Agriculture	25%	28%	47%	48%	53%	59%	66%	69%	74%

Source: Broadberry et al. (2013)

Appendix B

B.1 Proof of Lemma 1

If we equalize (6) and (7), using the assumption of perfect labor mobility, we know that workers are also employed in the industrial sector if the marginal productivity in the industrial sector A_t^I is equal to or exceeds the marginal productivity in the agricultural sector.

B.2 Proof of Lemma 2

First we need to derive the properties of e_{t+1}. Using (14) and (16), the optimization with respect to e_{t+1} shows how the implicit function $E(.)$ only depends on e_{t+1} and T_t:

$$E(e_{t+1}, T_t) = h'_{t+1}(\tau^r + g(\tau^e, T_t)e_{t+1}) - h_{t+1}g(\tau^e, T_t) \ , \tag{B.1}$$

where $E_e(e_{t+1}, T_t) < 0$. $E_T(e_{t+1}, T_t) = g'(\tau^e, T_t)[h'_{t+1}e_{t+1} - h_{t+1}] > 0$ according to the properties of $g(.)$, $n_t(.)$ and $h_t(.)$. To guarantee that for a positive level of T_t, the chosen level of education is higher than zero, it is assumed that:

$$E(0,0) = h'_{t+1}(0)\tau^r - h_{t+1}(0)g(\tau^e, 0) = 0 \ , \tag{B.2}$$

We know that $E(0, T_t)$ is increasing in T_t. Also, the $lim_{T\to\infty} E(0, T_t)$ is higher than $(0, 0)$, so from (A1) it is positive. In the concrete case of this paper, where the set of equations of $g(.)$, $n_t(.)$ and $h_t(.)$ is assumed, e_{t+1} will only depend on T_t. Using the first order conditions obtained from the optimization of the household problem for n_t and e_{t+1} yields:

$$\frac{h_{t+1}(e_{t+1})}{h'_{t+1}(e_{t+1})} = \frac{(\tau^r + g(\tau^e, T_t)e_{t+1})}{g(\tau^e, T_t)}$$

Applying the definition of $h_t(.)$, and after some manipulations, we arrive at the expression:

$$e_{t+1} = \frac{\beta\tau^r - g(\tau^e, T_t)}{(1-\beta)g(\tau^e, T_t)} \qquad \text{(B.3)}$$

From here we can discern what is meant by condition (B.2). This condition is a strong assumption in the model. Since the objective of the transferences from landowners to the households is to promote education, $e_{t+1}(T_t) < 0$ is not realistic. Therefore, we choose parameters in the model that make sure that a sufficient condition $(\beta\tau^r - g(\tau^e, 0) \geq 0)$ will be valid and $e_{t+1} \geq 0$ will always apply, as was shown in (B.2). This assumption can actually be relaxed if we want to assume that some residual education prevails in the economy. But, as explained in the introduction, costs of education were quite large by the time of the Industrial Revolution, therefore we enforce condition (B.2) - $\beta\tau^r = g(\tau^e, 0)$ - such that $e_{t+1}|_{T_t=0} = 0$.

As we can observe, e_{t+1} depends only on the variable T_t: $e_{t+1} = e_{t+1}(T_t)$. From here, and using the definition of $g(.)$, we can also derive the derivative of e_{t+1} with respect to T_t:

$$\begin{aligned} e'_{t+1}(T_t) &= \frac{-1}{(1-\beta)}\left[\frac{\beta\tau^r - g(\tau^e, T_t)}{[g(\tau^e, T_t)]^2}\frac{dg}{dT_t} + \frac{1}{g(\tau^e, T_t)}\frac{dg}{dT_t}\right] \\ &= \frac{-1}{(1-\beta)g(\tau^e, T_t)}\frac{dg}{dT_t}\left[1 + \frac{\beta\tau^r - g(\tau^e, T_t)}{g(\tau^e, T_t)}\right] \qquad \text{(B.4)} \\ &= -\frac{1}{(1-\beta)g(\tau^e, T_t)}\frac{dg}{dT_t}\left[\frac{\beta\tau^r}{g(\tau^e, T_t)}\right] \end{aligned}$$

From above, and since $\frac{dg}{dT_t} < 0$, it is definitely the case that $e'_{t+1}(T_t) > 0$, when applying the definitions of $g(.)$.

B.3 Proof of Proposition 1 – part (a)

This proposition claims that the number of offspring is decreasing in T_t. Using (17):

$$\frac{dn_t}{dT_t} = \frac{-(1-\gamma)\left[\frac{dg}{dT_t}e_{t+1} + \frac{de_{t+1}}{dT_t}g(.)\right]}{(\tau^r + g(\tau^e, T_t)e_{t+1})^2} \gtrless 0$$

By the formulations of $e'_{t+1}(.)$above,

$$\frac{dn_t}{dT_t} = \frac{-(1-\gamma)}{(\tau^r + g(\tau^e, T_t)e_{t+1})^2}\left[\frac{dg}{dT_t}\left(e_{t+1} - \frac{1}{(1-\beta)}\left[1 + \frac{\beta\tau^r - g(\tau^e, T_t)}{g(\tau^e, T_t)}\right]\right)\right]$$

Applying (B.3),

$$\frac{dn_t}{dT_t} = \frac{-(1-\gamma)}{(\tau^r + g(\tau^e, T_t)e_{t+1})^2}\left[\frac{dg}{dT_t}\left(\frac{-g(\tau^e, T_t)}{(1-\beta)g(\tau^e, T_t)}\right)\right]$$

Since $\frac{dg}{dT_t} < 0$, then it is straightforward that $\frac{dn_t}{dT_t} < 0$.

B.4 Proof of Lemma 3

Since we are maximizing the utility, we want the values of t_t for the interval $[0,1]$ to yield that maximum. So,

When $\frac{du}{dt_t} \neq 0$ for the interval of $t_t \,\epsilon\, [0,1]$;

If $\frac{du}{dt_t} > 0 \;\Rightarrow G(.) = \frac{d\rho_{t+1}}{dt_t} - b_t > 0 \;\Rightarrow t_t = 1$

If $\frac{du}{dt_t} < 0 \;\Rightarrow G(.) = \frac{d\rho_{t+1}}{dt_t} - b_t < 0 \;\Rightarrow t_t = 0$

Since $\frac{du}{dt_t}$ is a decreasing function, from numerical simulations, these are the only valid cases, and $\frac{du}{dt_t}|t_t = 0 < 0$ and $\frac{du}{dt_t}|t_t = 1 > 0$ do not apply.

Take the definition of ρ_{t+1} from (5):

$$\rho_{t+1} = \alpha(A^A_{t+1})^\alpha (X_{t+1})^{\alpha-1}(L^A_{t+1})^{1-\alpha} =$$

$$= \alpha(A^A_{t+1})^\alpha (X_{t+1})^{\alpha-1}\big((1-\lambda_{t+1})L_{t+1}\big)^{1-\alpha}$$

Depending on if the industrial sector is already economically viable or not, the definition of λ_{t+1} changes. Before the take-off of the industrial sector, landowners assume prospective values of λ_{t+1}, given (6), (7) and Lemma 1, such that:

$$\rho_{t+1} = \alpha (A_{t+1}^{A})^{\alpha} (X_{t+1})^{\alpha-1} \left(\frac{(1-\alpha)^{\frac{1}{\alpha}} A_{t+1}^{A} X_{t+1}}{(h_{t+1} A_{t+1}^{I})^{\frac{1}{\alpha}} L_{t+1}} L_{t+1} \right)^{1-\alpha}$$

Using the dynamic conditions from Subsection 4.1:

$$= \alpha L_t^{\varepsilon} (A_t^{A})^{\delta} \left[\frac{1-\alpha}{(1+h_t L_t^{\Delta})^{\zeta} A_t^{I}} \right]^{\frac{1-\alpha}{\alpha}} \frac{(1+e_{t+1}(A_t^{I})^{b})}{(1+e_{t+1})^{\frac{\beta(1-\alpha)}{\alpha}}}$$

As we can observe from above, on the one hand, rents are positively dependent on agricultural technology, which depends on industrial technology and education itself. On the other hand, they depend negatively on the industrial technology since more technology will increase the share of workers in the industrial sector. However, the industrial sector is still closed; therefore, both incentives are not strong enough to make landowners' take action and set positive taxes. Taking now the implicit function $G(.)$, we show this conclusion below:

$$G(.) = \frac{d\rho_{t+1}}{dt_t} - b_t =$$

$$= \alpha (L_t)^{\varepsilon} \left(A_t^{A}\right)^{\delta} \left[\frac{(1-\alpha)^{\frac{1}{\alpha}}}{\left(\left(1+h_t L_t^{\Delta}\right)^{\zeta} A_t^{I}\right)^{\frac{1}{\alpha}}} \right]^{1-\alpha} \left(\frac{\left(A_t^{I}\right)^{b}}{(1+e_{t+1})^{\frac{\beta(1-\alpha)}{\alpha}}} \frac{de_{t+1}}{dt_t} - \frac{\beta(1-\alpha)}{\alpha} \frac{\left(1+e_{t+1}\left(A_t^{I}\right)^{b}\right)}{(1+e_{t+1})^{\frac{\beta(1-\alpha)}{\alpha}+1}} \frac{de_{t+1}}{dt_t} \right) - b_t$$

$$\Leftrightarrow G(.) = \frac{\alpha (L_t)^{\varepsilon} \left(A_t^{A}\right)^{\delta}}{(1+e_{t+1})^{\frac{\beta(1-\alpha)}{\alpha}}} \left[\frac{(1-\alpha)^{\frac{1}{\alpha}}}{\left(\left(1+h_t L_t^{\Delta}\right)^{\zeta} A_t^{I}\right)^{\frac{1}{\alpha}}} \right]^{1-\alpha} \frac{de_{t+1}}{dt_t} \left(\left(A_t^{I}\right)^{b} - \frac{\beta(1-\alpha)}{\alpha} \frac{\left(1+e_{t+1}\left(A_t^{I}\right)^{b}\right)}{(1+e_{t+1})} \right) - b_t$$

Since b_t is always positive, if Lemma 4's condition (see Appendix B.6) is negative, then $G(.) < 0$. So only when Lemma 4's condition is positive and sufficiently high $G(.) \geq 0$. More explicitly, as $\frac{\beta(1-\alpha)}{\alpha}$ is constant, and A_t^{I} grows over time, then:

$$(A_t^{I})^{b} - \frac{\beta(1-\alpha)}{\alpha} \frac{(1+e_{t+1}(A_t^{I})^{b})}{(1+e_{t+1})} =$$

$$= (A_t^{I})^{b} - \frac{\beta(1-\alpha)}{\alpha} \frac{\left(\frac{1}{(A_t^{I})^{b}} + e_{t+1}\right)}{(1+e_{t+1})} (A_t^{I})^{b}$$

Given some parameterization that guarantees that $\frac{\beta(1-\alpha)}{\alpha} < 1$ and b sufficiently high, as A_t^I grows, when A_t^I is sufficiently large such that $(A_t^I)^b > 1$ and for any $\mathrm{e}_{t+1} \geq 0$, we have a positive condition. Thus, at some point in time, $G(.) \geq 0$. We observe in the numerical simulations that $G(.)$ will not be higher than zero before the onset of the industrial sector.

After the industrialization, landowners' decisions depend on a different implicit function that then makes landowners aware that allowing for education fosters not only agricultural production, but also industrial production. Therefore,

$$\rho_{t+1} = \alpha(A_{t+1}^A)^\alpha (X_{t+1})^{\alpha-1} \left(\left(1 - \frac{A_{t+1}^I (h_{t+1})^{\frac{\theta}{\alpha}}}{A_{t+1}^A X_{t+1} + A_{t+1}^I (h_{t+1})^{\frac{\theta}{\alpha}}} \right) L_{t+1} \right)^{1-\alpha}$$

And the given implicit function:

$$G(.) = \frac{d\rho_{t+1}}{dt_t} - b_t = \frac{\alpha(1-\alpha)^{\frac{1-\alpha}{\alpha}} (L_t)^{1+\varepsilon-\alpha} (A_t^A)^\delta}{\left[(1 + e_{t+1}(A_t^I)^b)(L_t)^\varepsilon (A_t^A)^\delta X_t (1-\alpha)^{\frac{1}{\alpha}} + \theta^{\frac{1}{\alpha}} (1 + h_t L_t^A)^\zeta A_t^I (1 + e_{t+1})^{\beta\frac{\theta}{\alpha}} \right]^{1-\alpha}}$$

$$\left(\frac{de_{t+1}}{dt_t} n_t^{1-\alpha} (A_t^I)^b \left(1 - (1-\alpha) \frac{(L_t)^\varepsilon (A_t^A)^\delta X_t (1-\alpha)^{\frac{1}{\alpha}} + \theta^{\frac{1}{\alpha}} (1 + h_t L_t^A)^\zeta A_t^I (1 + e_{t+1})^{\beta\frac{\theta}{\alpha}} \left(\beta \frac{\theta}{\alpha} \frac{1}{(1 + e_{t+1})(A_t^I)^b} \right)}{(L_t)^\varepsilon (A_t^A)^\delta X_t (1-\alpha)^{\frac{1}{\alpha}} + \theta^{\frac{1}{\alpha}} (1 + h_t L_t^A)^\zeta A_t^I (1 + e_{t+1})^{\beta\frac{\theta}{\alpha}} \left(\frac{1}{(1 + e_{t+1}(A_t^I)^b)} \right)} \right) \right.$$

$$\left. - (1-\alpha)(1 + e_{t+1}(A_t^I)^b) n_t^{-\alpha} (1-\gamma)(\tau^r + g(.)e_{t+1})^{-2} \left(\frac{dg(.)}{dt_t} e_{t+1} + \frac{de_{t+1}}{dt_t} g(.) \right) \right) - b_t$$

Since b_t is always positive, and from the proof of Proposition 1: $\left(\frac{dg(.)}{dt_t} e_{t+1} + \frac{de_{t+1}}{dt_t} g(.) \right) > 0$ then, if Lemma 5's condition (see Appendix B.7) is negative, $G(.) < 0$. Only when Lemma 5's condition is positive does $G(.) \geq 0$ at some point.

Lemma 1's condition depends on initial conditions of A_t^A and L_t. Since we assume the steady state values under the Malthusian regime, both A_t^A and L_t depend on the parameters of the model. We know from Lemma 1 that industrial technology guarantees that Lemma 1's condition holds at the beginning of the Malthusian and Pre-Industrial period. Under both the parameterization of the model and the assumption of the initial value of A_t^I at the Malthusian period condition, $1 > (A_t^I)^b$ originally holds.

Then, when the economy starts to industrialize and $(A_t^I)^b > 1$, we can certainly guarantee that:

$$0 < (1-\alpha)\frac{(L_t)^{\varepsilon}(A_t^A)^{\delta}X_t(1-\alpha)^{\frac{1}{\alpha}} + \theta^{\frac{1}{\alpha}}(1+h_tL_t^A)^{\zeta}A_t^I(1+e_{t+1})^{\beta\frac{\theta}{\alpha}}\left(\beta\frac{\theta}{\alpha}\frac{1}{(1+e_{t+1})(A_t^I)^b}\right)}{(L_t)^{\varepsilon}(A_t^A)^{\delta}X_t(1-\alpha)^{\frac{1}{\alpha}} + \theta^{\frac{1}{\alpha}}(1+h_tL_t^A)^{\zeta}A_t^I(1+e_{t+1})^{\beta\frac{\theta}{\alpha}}\left(\frac{1}{(1+e_{t+1}(A_t^I)^b)}\right)} < 1$$

And hence,

$$\left(1-(1-\alpha)\frac{(L_t)^{\varepsilon}(A_t^A)^{\delta}X_t(1-\alpha)^{\frac{1}{\alpha}} + \theta^{\frac{1}{\alpha}}(1+h_tL_t^A)^{\zeta}A_t^I(1+e_{t+1})^{\beta\frac{\theta}{\alpha}}\left(\beta\frac{\theta}{\alpha}\frac{1}{(1+e_{t+1})(A_t^I)^b}\right)}{(L_t)^{\varepsilon}(A_t^A)^{\delta}X_t(1-\alpha)^{\frac{1}{\alpha}} + \theta^{\frac{1}{\alpha}}(1+h_tL_t^A)^{\zeta}A_t^I(1+e_{t+1})^{\beta\frac{\theta}{\alpha}}\left(\frac{1}{(1+e_{t+1}(A_t^I)^b)}\right)}\right) > 0$$

Note that, as $\beta\frac{\theta}{\alpha}$ is constant, $\frac{1}{(1+e_{t+1})(A_t^I)^b} > \frac{1}{\left(1+e_{t+1}(A_t^I)^b\right)}$ only if $1 > (A_t^I)^b$

Therefore, again, when A_t^I is sufficiently large, we have a positive condition. So, at some point in time $G(.) \geq 0$ – we observe this also in the numerical simulations. Moreover, as we show in the sensitivity analysis below, $\frac{\partial G}{\partial A_t^I}$ is always positive for the range of parameters chosen for the main analysis in the paper, and additionally for a wide range of values for each parameter, giving support to the claim above.

This dual effect will make landowners delay the onset of taxes and only later will they slowly raise them and, hence, increase education.

B.5 Sensitivity Analysis on $G(.)$

Because the analytical solution of the model is intractable, it is too complex to determine the impact of several variables, such as t_t, A_t^A, and A_t^I on landowners' decisions, $G(.)$. Therefore, for the sake of clarity, a numerical sensitivity analysis was done so that how the dynamics of landowners' decisions depend on the evolution of the economy and its parameters could be understood. This sensitivity analysis is only possible when using numerical simulations of the analytical solutions of the derivatives. To calculate the derivatives, we use the values from the real model at three specific moments in time of the simulation – before the industrialization, just after the industrialization and after education becomes prevalent in the economy. These three moments of time allow us to observe how each derivative behaves relying on its analytical solution, which changes between before and after industrialization (see above), and also relying on the emergence or not of education. Additionally, we can also test if variables that change over time, such as A_t^A, A_t^I and L_{t} affect the behavior of each derivative. We verify that they do not affect $\frac{\partial G}{\partial t_t}$, but affect the behavior of $\frac{\partial G}{\partial A_t^A}$ and slightly alter the behavior of $\frac{\partial G}{\partial A_t^I}$.

Concerning the parameters, we present below the results for different ranges of parameters in each table. The ranges are not shown but we test for the whole range $(0,1]$ in almost all parameters except for ϕ, which ranges from $(-1,-0.5]$. These ranges are reasonable for most parameters, but can be too large for parameters such as γ or α; in these cases, we stopped at the value 0.9.

We start by showing that $\frac{\partial G}{\partial t_t}$ is always negative for a different range of parameters and also in different stages of the economy. We consider the necessary restrictions on the parameters to guarantee a globally stable steady state. Therefore, for example, we only consider $\tau^r > \bar{\tau}^r$ due to condition 1 below, (see Proof of Proposition 2).

Table 1B. Sensitivity analysis for $\frac{\partial G}{\partial t_t}$

	ϕ	τ^r	γ	b
Before industrialization	< 0	< 0	< 0	< 0
After industrialization	< 0	< 0	< 0	< 0
After industrialization and after rise of education	< 0	< 0	< 0	< 0

Source: own computations

As we can see from Table 1B above, $\frac{\partial G}{\partial t_t}$ is negative for almost the whole range of parameters - here only the most relevant ones to determine t_t are shown. Note that for initially very low values of $A_t^I \leq 0.6$, $\frac{\partial G}{\partial t_t}$ is positive. Since we assume an initial value of $A_0^I = 0.6$, this does not pose problems to the sensitivity analysis here. In the period after the emergence of the industrial sector, it is generally valid for all simulations made. The only exception ($\gamma > 0.9$) would not be in the chosen range of parameterization since it implies that households would dedicate more than 90% of their time to raising children, which is not empirically observed. Therefore, we can be confident in assuming that under the standard range of parameterization, $\frac{\partial \mathrm{G}}{\partial \mathrm{t_t}} < 0$ will always hold.

The same procedure was implemented to determine the relationship between $G(.)$ and A_t^A. From Table 2B below, the sign of $\frac{\partial G}{\partial A_t^A}$ is ambiguous.

Table 2B. Sensitivity analysis for $\frac{\partial G}{\partial A_t^A}$

	ϕ	τ^r	γ	b	α	δ	Δ	ε
Before industrialization	< 0	< 0	< 0	< 0	< 0	< 0	< 0	< 0
After industrialization	$\lesseqgtr 0$	$\lesseqgtr 0$	$\lesseqgtr 0$	$\lesseqgtr 0$	$\lesseqgtr 0$	$\lesseqgtr 0$	$\lesseqgtr 0$	$\lesseqgtr 0$
After industrialization and after rise of education	$\lesseqgtr 0$	$\lesseqgtr 0$	$\lesseqgtr 0$	$\lesseqgtr 0$	$\lesseqgtr 0$	$\lesseqgtr 0$	$\lesseqgtr 0$	$\lesseqgtr 0$

Source: own computations

Before industrialization, it is clear that $\frac{\partial G}{\partial A_t^A}$ is always negative. However, in the time period after the industrialization, the sign of the derivative depends on the range of values of A_t^A. For small values of A_t^A, i.e. values of A_t^A at the three specific moments of the model shown in the sensitivity analysis, $\frac{\partial G}{\partial A_t^A}$ is positive. As A_t^A increases, $\frac{\partial G}{\partial A_t^A}$ becomes negative, *ceteris paribus*. For the same range of values of A_t^A, the higher A_t^I is, the higher the threshold of A_t^A such that $\frac{\partial G}{\partial A_t^A}$ becomes negative. We also know that the technology of agriculture is lower than the industrial sector's technology after industrialization, and the former's growth dynamics follow the growth dynamics of the latter. Since both technologies are increasing, it is thus always the case that $\frac{\partial G}{\partial A_t^A} > 0$. All parameters do not affect these results and the same patterns referred to previously hold. The derivative will always behave as described above.

Regarding $\frac{\partial G}{\partial A_t^I}$, shown in Table B3 below, we observe that it is almost always positive, the exception being only after industrialization for the range of parameters α and b. However, this only happens for values of b lower than 0.66 and for values of α around 0.1, which is too low and not an option for calibration. This analysis gives strength to the claims made in part B.4 in this Appendix, where $G(.)$ will at some point be positive due to increasing A_t^I.

Table 3B. Sensitivity analysis for $\frac{\partial G}{\partial A_t^I}$

	ϕ	τ^r	β	γ	b	α	δ	Δ	ε
Before industrialization	> 0	> 0	> 0	> 0	> 0	> 0	> 0	> 0	> 0
After industrialization	> 0	> 0	> 0	> 0	$\lesseqgtr 0$	$\lesseqgtr 0$	> 0	> 0	> 0
After industrialization and after rise of education	> 0	> 0	> 0	> 0	$\lesseqgtr 0$	$\lesseqgtr 0$	> 0	> 0	> 0

Source: own computations

B.6 Lemma 4:

Lemma 4: Before industrialization, for $G(.) \geq 0$ the condition $(A_t^I)^b - \frac{\beta(1-\alpha)}{\alpha}\frac{\left(1+e_{t+1}(A_t^I)^b\right)}{(1+e_{t+1})}$ must be positive.

Proof: From (5), (6), (7), (20) and Lemma 3 we can determine the implicit function $G(.)$ before industrialization:

$$G(.) = \frac{\alpha(L_t)^{\varepsilon}(A_t^A)^{\delta}}{(1+e_{t+1})^{\frac{\beta(1-\alpha)}{\alpha}}}\left[\frac{(1-\alpha)^{\frac{1}{\alpha}}}{((1+h_t L_t^A)^{\zeta} A_t^I)^{\frac{1}{\alpha}}}\right]^{1-\alpha}\frac{de_{t+1}}{dt_t}\left((A_t^I)^b - \frac{\beta(1-\alpha)}{\alpha}\frac{(1+e_{t+1}(A_t^I)^b)}{(1+e_{t+1})}\right) - b_t\ ,$$

where only e_{t+1} depends on t_t.

B.7 Lemma 5:

Lemma 5: After industrialization, for $G(.) \geq 0$ it must be true that:

$$\left(1-(1-\alpha)\frac{(L_t)^{\varepsilon}\left(A_t^A\right)^{\delta}X_t(1-\alpha)^{\frac{1}{\alpha}}+\theta^{\frac{1}{\alpha}}\left(1+h_tL_t^A\right)^{\zeta}A_t^I(1+e_{t+1})^{\beta\frac{\theta}{\alpha}}\left(\beta\frac{\theta}{\alpha}\frac{1}{(1+e_{t+1})\left(A_t^I\right)^b}\right)}{(L_t)^{\varepsilon}\left(A_t^A\right)^{\delta}X_t(1-\alpha)^{\frac{1}{\alpha}}+\theta^{\frac{1}{\alpha}}\left(1+h_tL_t^A\right)^{\zeta}A_t^I(1+e_{t+1})^{\beta\frac{\theta}{\alpha}}\left(\frac{1}{\left(1+e_{t+1}\left(A_t^I\right)^b\right)}\right)}\right)>0$$

Proof: It follows from the implicit function condition.

B.8 Globally Stable Steady State

Using Lemmas 7 and 8, the pre-industrial steady-state values of productivity in the agricultural sector A_{ss}, and the size of the working population L_{ss}, are given by:

$$A_{ss}^A=\left[\frac{(1-\tau^r)(1-\alpha)}{\tilde{c}}\right]^{\frac{\varepsilon}{\alpha(1-\delta-\varepsilon)}}X^{\frac{\varepsilon}{(1-\delta-\varepsilon)}},$$

$$L_{ss}=\left[\frac{(1-\tau^r)(1-\alpha)}{\tilde{c}}\right]^{\frac{1-\delta}{\alpha(1-\delta-\varepsilon)}}X^{\frac{1-\delta}{(1-\delta-\varepsilon)}},$$

Proof of Proposition 2:

Under the Pre-Industrial, Pre-Malthusian, the following dynamic system applies:

$$\begin{cases}A_{t+1}^A=(1+e_{t+1}(A_t^I)^b)(L_t)^{\varepsilon}(A_t^A)^{\delta} \\ L_{t+1}=\dfrac{1-\dfrac{\tilde{c}}{w_th_t}}{(\tau^r+g(\tau^e,\iota_t)e_{t+1})}L_t\end{cases}\quad \text{for } z_t<\tilde{z}$$

In this regime we assume that there is still no education ($e_{t+1}=0$) because at the beginning of the model the conditions of Lemma 4 do not hold yet. The Jacobian matrix will be represented by:

$$J(A^A,L)=\begin{bmatrix}\dfrac{dA^A}{dA_t^A} & \dfrac{dA^A}{dL_t} \\ \dfrac{dL}{dA_t^A} & \dfrac{dL}{dL_t}\end{bmatrix}=$$

$$= \begin{bmatrix} \delta L_t^{\varepsilon}(A_t^A)^{\delta-1} & \varepsilon L_t^{\varepsilon-1}(A_t^A)^{\delta} \\ \frac{1}{\tau^r}\frac{\tilde{c}\alpha}{(1-\alpha)}L_t^{1+\alpha}X_t^{-\alpha}(A_t^A)^{-\alpha-1} & \frac{1}{\tau^r}\left[1-\frac{\tilde{c}(1+\alpha)}{(1-\alpha)}L_t^{\alpha}X_t^{-\alpha}(A_t^A)^{-\alpha}\right] \end{bmatrix}$$

Under the steady-state values, the Jacobian matrix is given by:

$$J(A_{ss}^A, L_{ss}) = \begin{bmatrix} \frac{dA^A(A_{ss}^A, L_{ss})}{dA_t^A} & \frac{dA^A(A_{ss}^A, L_{ss})}{dL_t} \\ \frac{dL(A_{ss}^A, L_{ss})}{dA_t^A} & \frac{dL(A_{ss}^A, L_{ss})}{dL_t} \end{bmatrix} =$$

$$= \begin{bmatrix} \delta & \varepsilon\left[\frac{(1-\tau^r)(1-\alpha)X_{ss}^{\alpha}}{\tilde{c}}\right]^{-\frac{1}{\alpha}} \\ \frac{1}{\tau^r}\frac{\tilde{c}\alpha}{(1-\alpha)}\left[\frac{(1-\tau^r)(1-\alpha)}{\tilde{c}}\right]^{\frac{1+\alpha}{\alpha}}X_{ss} & \frac{1-(1+\alpha)(1-\tau^r)}{\tau^r} \end{bmatrix}$$

The eigenvalues are given by $\{\lambda_1, \lambda_2\}$. We know that: $det(A_{ss}^A, L_{ss}) = \lambda_1\lambda_2$ and $tr(A_{ss}^A, L_{ss}) = \lambda_1 + \lambda_2$

$$tr(A_{ss}^A, L_{ss}) = \delta + \frac{1-(1+\alpha)(1-\tau^r)}{\tau^r}$$

$$det(A_{ss}^A, L_{ss}) = \delta\frac{1-(1+\alpha)(1-\tau^r)}{\tau^r} - \varepsilon\frac{1}{\tau^r}\frac{\tilde{c}\alpha}{(1-\alpha)}\left[\frac{(1-\tau^r)(1-\alpha)}{\tilde{c}}\right]^{\frac{1+\alpha}{\alpha}-\frac{1}{\alpha}}X_{ss}X_{ss}^{-\frac{\alpha}{\alpha}}$$

$$\Leftrightarrow det(A_{ss}^A, L_{ss}) = \frac{1}{\tau^r}(\delta - (1-\tau^r)[\delta(1+\alpha)+\varepsilon\alpha])$$

so the equilibrium is globally stable if: $\lambda_1, \lambda_2 \in (-1,1)$.

(1) To guarantee that the convergence to the steady state is monotonically stable:

i. $Det(A_{ss}^A, L_{ss}) > 0$;

ii. and $Tr(A_{ss}^A, L_{ss}) > 0$.

For (i):

$$\frac{1}{\tau^r}(\delta - (1-\tau^r)[\delta(1+\alpha)+\varepsilon\alpha]) > 0$$

$$\Leftrightarrow \delta > \delta(1-\tau^r)(1+\alpha) + \varepsilon\alpha(1-\tau^r)$$

Condition 1: $\delta > \delta(1-\tau^r)(1+\alpha) + \varepsilon\alpha(1-\tau^r)$ is a necessary condition. We guarantee this condition through the parameterization of the model.

For (ii): $Tr(A_{ss}^A, L_{ss})$ is always higher than zero from the inequality below:

$$(1+\alpha)(1-\tau^r) < 1 \Leftrightarrow \tau^r > \frac{\alpha}{1+\alpha}$$

Condition 2: $\tau^r > \frac{\alpha}{1+\alpha}$ is a necessary condition. We will assume it in the parameterization of the model.

Given the parameterization in Subsection 6.1, we guarantee that the Jacobian matrix $J(A_{ss}^A, L_{ss})$ has real eigenvalues with modulus less than 1, meaning that the convergence to the steady state is monotonically stable.

(2) To guarantee that the equilibrium is globally stable:

i. $-2 < Tr(A_{ss}^A, L_{ss}) < 2$;

ii. $-1 < Det(A_{ss}^A, L_{ss}) < 1$;

iii. $Det(A_{ss}^A, L_{ss}) - Tr(A_{ss}^A, L_{ss}) \geq -1$;

iv. and $Det(A_{ss}^A, L_{ss}) + Tr(A_{ss}^A, L_{ss}) \geq -1$.

For (i): from before we know that $Tr(A_{ss}^A, L_{ss}) > 0 > -2$

$Tr(A_{ss}^A, L_{ss}) < 2$

$\Rightarrow \delta + \frac{1-(1+\alpha)(1-\tau^r)}{\tau^r} < 2 \Leftrightarrow 1-(1+\alpha)(1-\tau^r) < (2-\delta)\tau^r \Leftrightarrow \tau^r(1-\delta-\alpha) + 1 + \alpha > 1$

$\Rightarrow$ **Condition 3:** $\delta + \alpha < 1$, from the parameterization of the model, this condition holds:

$\Rightarrow Tr(A^A_{ss}, L_{ss}) \epsilon (-2,2)$

For (ii):

$Det(A^A_{ss}, L_{ss}) > -1$:

From Condition 1 we know this inequality holds.

$Det(A^A_{ss}, L_{ss}) < 1$:

$\frac{1}{\tau^r}(\delta - (1-\tau^r)[\delta(1+\alpha) + \varepsilon\alpha]) < 1$

$\Leftrightarrow \delta < \tau^r + (1-\tau^r)[\delta(1+\alpha) + \varepsilon\alpha]$

$\Rightarrow$ We guarantee that this condition holds under the parameterization of the model in Subsection 6.1 $\Rightarrow$ $Det(A^A_{ss}, L_{ss}) < 1$

For (iii):

$$\frac{1}{\tau^r}(\delta - (1-\tau^r)[\delta(1+\alpha) + \varepsilon\alpha]) - \delta - \frac{1-(1+\alpha)(1-\tau^r)}{\tau^r} \geq -1$$

$$\Leftrightarrow \frac{1}{\tau^r}[(1-\delta)[(1-\tau^r)(1+\alpha) - 1] - (1-\tau^r)\varepsilon\alpha] - \delta \geq -1$$

$\Rightarrow$ **Condition 4:** $1 \geq \varepsilon + \delta$, under the parameterization of the model this condition holds.

For (iv):

$$\frac{1}{\tau^r}(\delta - (1-\tau^r)[\delta(1+\alpha) + \varepsilon\alpha]) + \delta + \frac{1-(1+\alpha)(1-\tau^r)}{\tau^r} \geq -1$$

$$\Leftrightarrow \frac{1}{\tau^r}\big[1 + \delta + (1-\tau^r)[(1+\alpha)(\delta-1) + \varepsilon\alpha]\big] + \delta \geq -1$$

$$\Leftrightarrow (1+\tau^r)(1+\delta) \geq (1-\tau^r)[1 - \delta - \alpha(\delta + \varepsilon - 1)]$$

If Conditions 3 and 4 hold, then this inequality will also hold.

Under the parameterization in Subsection 6.1, the conditions above will hold. Hence, under the Pre-Industrial, Pre-Malthusian regime, and under this parameterization, we guarantee that the equilibrium is globally stable.

B.9 Industrial Revolution Regimes

In this first regime the economy is governed by a three-dimensional non-linear first-order autonomous system:

$$\begin{cases} A_{t+1}^{A} = (1 + e_{t+1}(A_t^I)^b)(L_t)^{\varepsilon}(A_t^A)^{\delta} \\ A_{t+1}^{I} = (1 + h_t L_t^A)^{\zeta} A_t^I \\ L_{t+1} = \dfrac{1 - \dfrac{\tilde{c}}{w_t h_t}}{(\tau^r + g(\tau^e, \iota_t) e_{t+1})} L_t \end{cases} \quad \text{for } z_t < \tilde{z}$$

The second regime is governed by the same three-dimensional system, although population growth does not depend on income of workers:

$$\begin{cases} A_{t+1}^{A} = (1 + e_{t+1}(A_t^I)^b)(L_t)^{\varepsilon}(A_t^A)^{\delta} \\ A_{t+1}^{I} = (1 + h_t L_t^A)^{\zeta} A_t^I \\ L_{t+1} = \dfrac{1 - \gamma}{(\tau^r + g(\tau^e, \iota_t) e_{t+1})} L_t \end{cases} \quad \text{for } z_t \geq \tilde{z}$$

B.10 Long-run outcomes and sensitivity analysis on the parameters of the model

Before we start with the sensitivity analysis *per se*, here we also show the long run outcomes of the baseline scenario that appear in Subsection 6.2. As we observe in Figures 1B and 2B below, the paths of education, fertility, and growth rates of technology and income per capita, under the parameterization in Tables 1 and 2, tend to stabilize after more than 1000 periods.

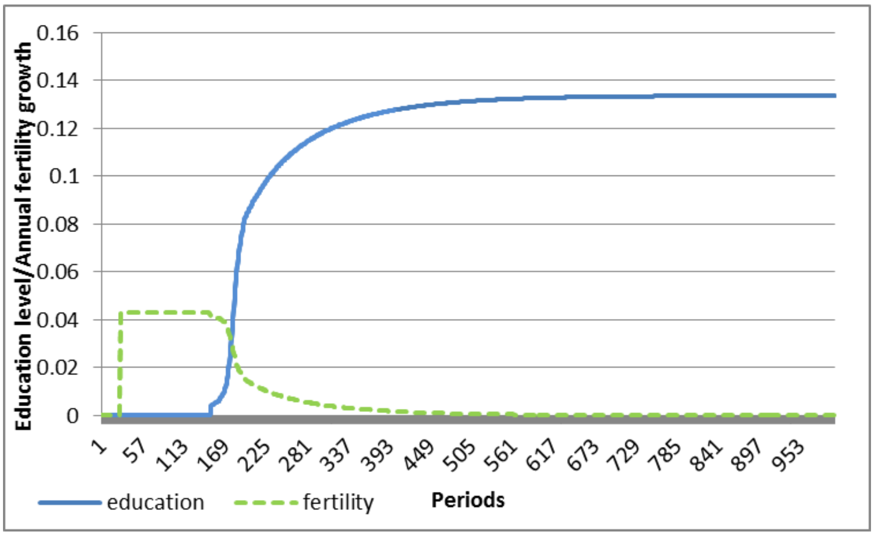

Figure 1B. Paths of education level and annual fertility growth in the long run

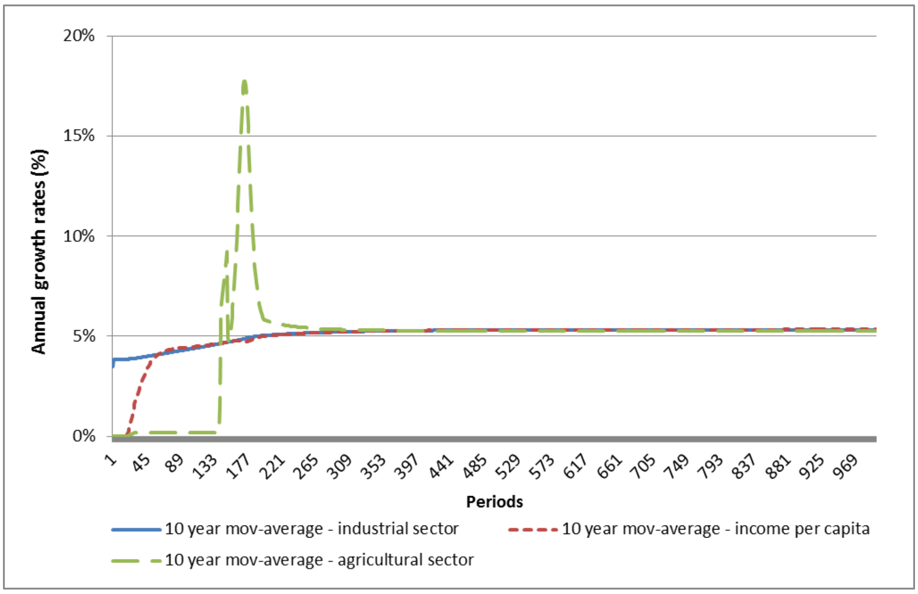

Figure 2B. Technology and income per capita paths in the long run

All conditions necessary to assume global stability hold under the chosen parameterization of the model. None of these conditions are strong and are assumed under a typical calibration used in other papers of the literature, as the ones cited in Subsection 6.1. Only condition (B.2) poses a stronger assumption on the parameters. We make some sensitivity analysis on some of the parameters for a fair range of possible values of the parameters to test how sensitive the results of the numerical simulation are. The parameters chosen are the fixed costs with children, which depart more from the values assumed in the literature, and some parameter values which are not taken from the literature. We assume a variation in each parameter such that, within the vicinity of the benchmark values, we can test if the model becomes unstable easily, and whether the results continue to hold or not. In table 4B, we summarize the parameters analyzed in the sensitivity analysis, and which ones we recalibrate in each new simulation.

Table 4B. Recalibrated parameters for new scenarios

	Baseline	$\tau^r = 0.2$ - only τ^e recalibrated	$\tau^r = 0.2$ - τ^e and γ recalibrated	$\phi = -0.6$	$\varepsilon = 0.12$	$b = 0.7$	$\zeta = 0.15$
γ	0.6349	0.6349	0.7950	0.6349	0.6349	0.6349	0.6349
τ^e	0.097	0.052	0.052	0.097	0.097	0.097	0.097
L_0	0.0761	0.1415	0.1415	0.0761	0.0584	0.0761	0.0761
A_0^A	0.8918	0.9168	0.9168	0.8918	0.6847	0.8918	0.8918

Source: Own computations

A lower level of τ^r

Since the literature assumes a much lower fixed cost of raising children, we lower τ^r to 0.20 and analyze the results derived. We recalibrate the values of τ^e, due to condition (B.2), as well as the initial conditions L_0 and A_0^A. Note that under $\tau^r = 0.2$, conditions 1 and 2 of Appendix B.9 do not hold. As we can observe in Figures 3B and 4B, the patterns of the simulation change substantially with lower costs of raising children because now the costs of raising children are too low and fertility rates increase too much after the industrialization period; however, fertility decisions are still made under the subsistence constraint (see equation 17). Even the rise of education does not lead to a decrease of fertility due to lower fixed child-raising costs. Thus, with increasing gains in technology and increasing income, population explodes.

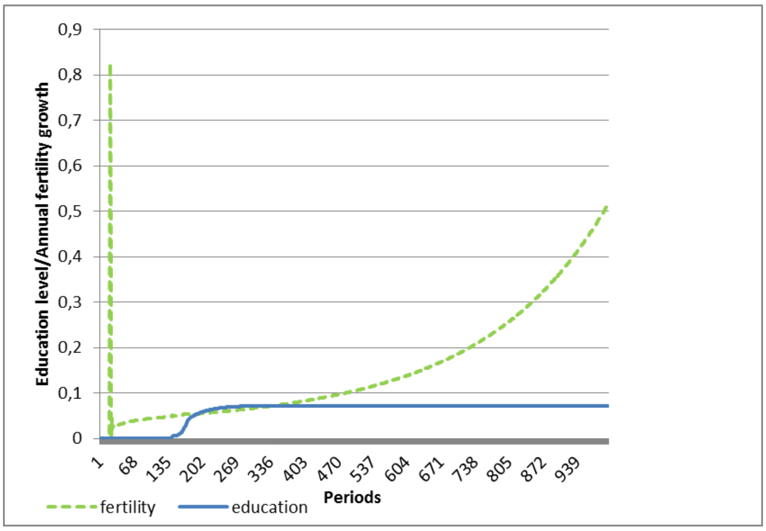

Figure 3B. Paths of education level and annual fertility growth with a lower level of τ^r

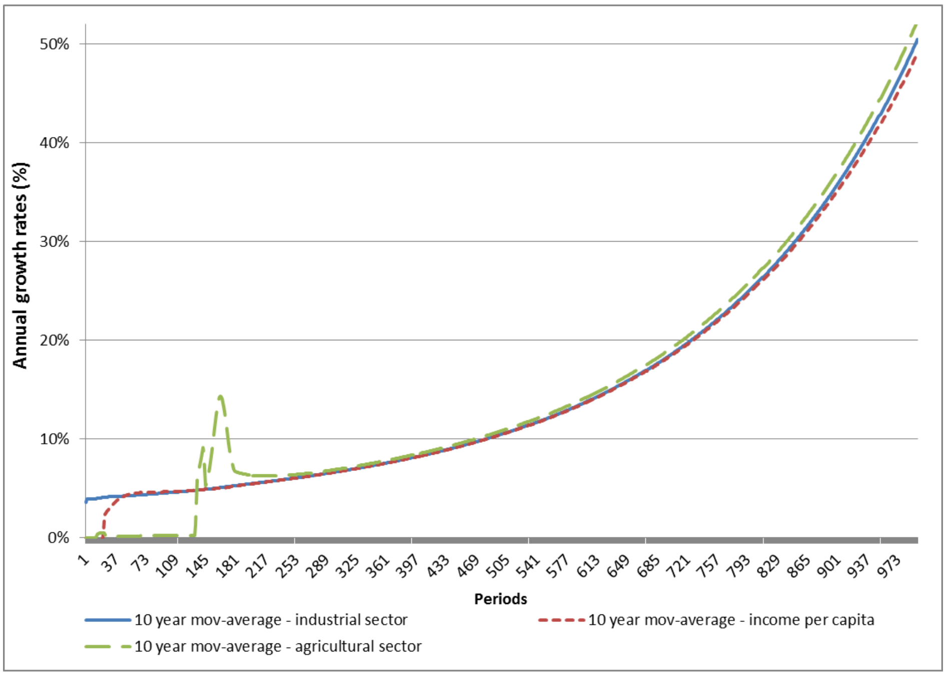

Figure 4B. Technology and income per capita paths with a lower level of τ^r

Only with τ^r close to 0.35, are the dynamics of the model out of the Malthusian trap and lead to the demographic change dynamics.

One of the main reasons for these patterns lies in the fact that low child-raising costs and high preferences for children lead to a boom in fertility. If, instead, we also recalibrate preferences for children by setting $\gamma = 0.7950$, then with low τ^r we can reach the same dynamics as in the baseline model. Figures 5B and 6B show exactly this.

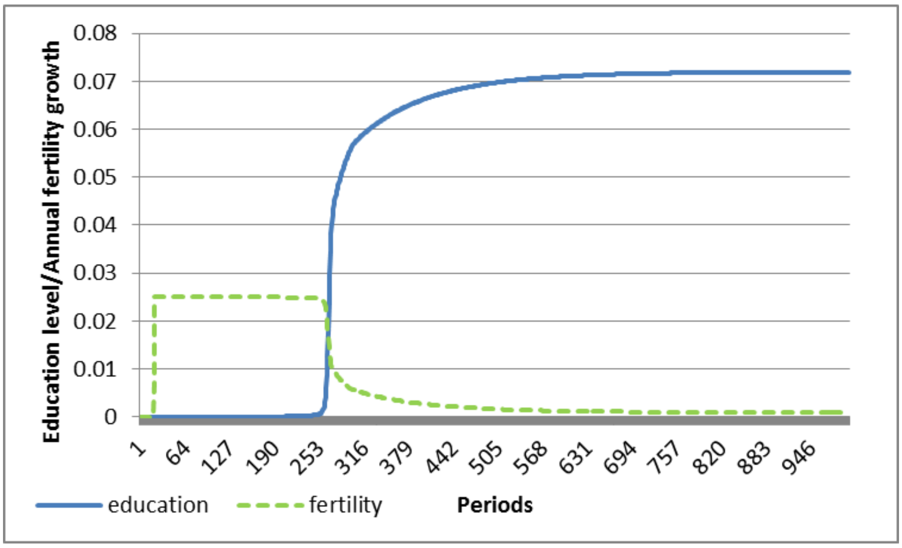

Figure 5B. Paths of education level and annual fertility growth with a lower level of τ^r, recalibrated

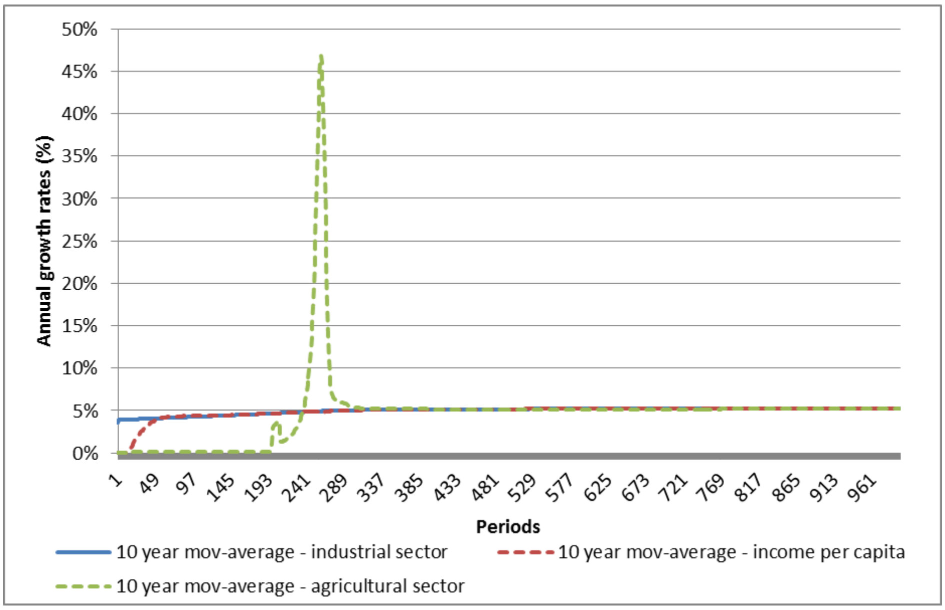

Figure 6B. Technology and income per capita paths with a lower level of τ^r, recalibrated

However, the fertility rates are now much lower and education takes longer to emerge. This occurs because the growth on technology depends on population growth and if population grows at a slower pace, technology gains will also take longer. And, as explained above, landowners will wait until their rents can better benefit from industrial technology advancements before they begin transferring resources to support education.

Alternatively, if we decide to break condition (B.2) by not recalibrating τ^e, the model will simply not converge. With lower τ^r, education decisions lead to negative education (see equation (B.3) in Appendix B.2), leading to complete instability in the model (figures not shown). If, instead, τ^r is higher, then workers will immediately prefer to educate their offspring and turn fertility rates negative. Both options of relaxing condition (B.2) lead to model instability and to outcomes far away from the patterns observed during the pre-industrial and industrial periods, as mentioned in Appendix B.8. Thus, condition B.2 is maintained throughout this paper.

A less negative value of ϕ

The parameters that are more sensitive and create more volatility in the model is the time endowment cost concavity, ϕ. Namely, under the purely calibrated parameters, changing ϕ creates instability in the model in the sense that it directly influences total costs on education and leads to a lower effect of landowners' transfers on workers' education decisions. As depicted in Figure 7B, education levels are now significantly lower compared to the baseline scenario; while fertility is comparatively higher than before. Nevertheless, the patterns of demographic transition are still untouched since fertility increases after the Malthusian regime and then declines as education rises. Regarding technology growth, it is no longer a sustained growth picture as average growth rates keep increasing over time due to the level effects of population on the dynamics of both technologies (see Figure 8B). The same pattern applies to income per capita, which increases over time. This contradicts our expectations that in the long run these variables should stabilize. All in all, the main trends observed in the baseline model are conserved, giving confidence that, despite some instability, the fall in fertility and the emergence of education decisions are in accordance with the hypotheses in this paper.

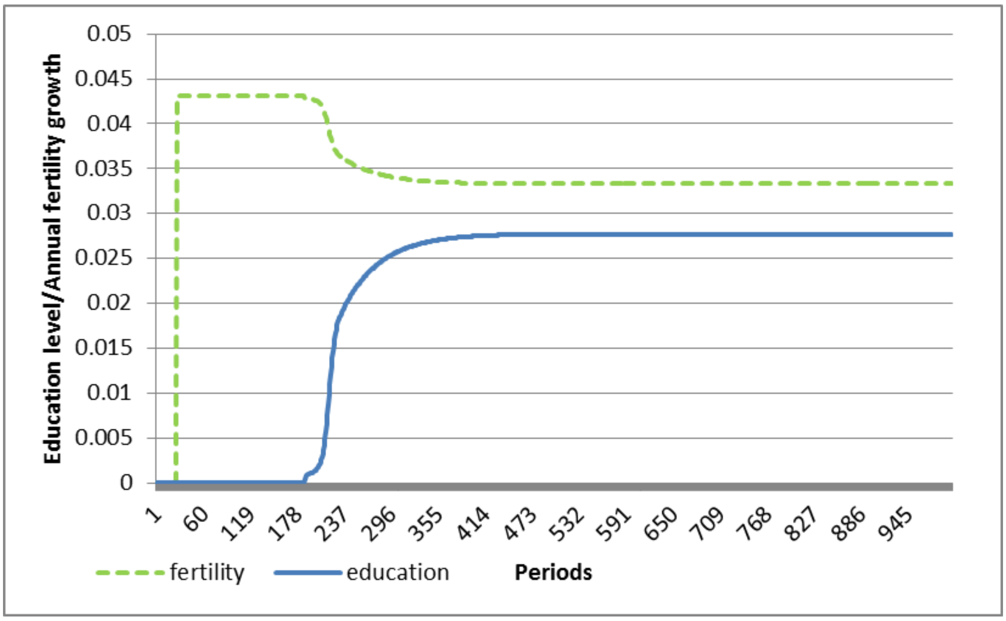

Figure 7B. Paths of education level and annual fertility growth with a less negative value of ϕ

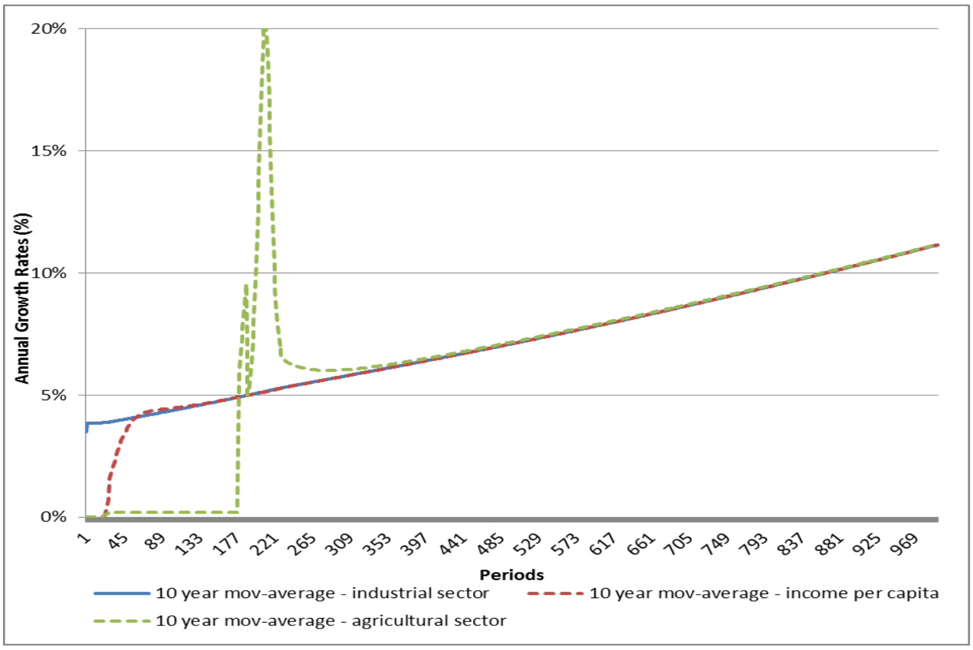

Figure 8B. Technology and income per capita paths with a less negative value of ϕ

Population effect on the agricultural technology (ε)

One of the parameters assumed in the model is the level of the learning by doing effect in equation (11). We assumed in the baseline parameterization a quite low weight (ε) of the population level that composes the total "learning by doing" effect. A lower level of ε means that population and all ideas that could emerge due to a quantity of people able to think and invent does not have much influence on agricultural technology. This assumption allows the dynamics of technology to increase at a slower pace and is in line with the same parameter for industrial technology Δ which is also low. However, this assumption could be debated because since the Industrial Revolution a huge amount of new ideas has emerged and reflected in agricultural technology also. Therefore, we relax the assumption of low ε and set a value of 0.12. What we observe is that the main results remain intact (see Figures 9B and 10B). The timing of the escape from the Malthusian trap is almost the same as in the baseline scenario as well as the timing of education. Technology and income per capita grow at slightly higher rates and tend to stabilization in the long run.

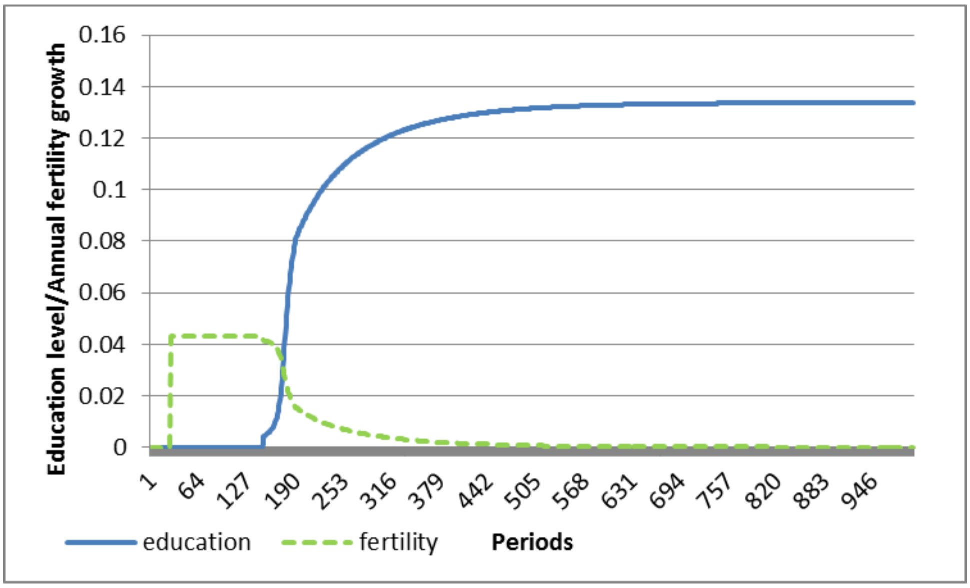

Figure 9B. Paths of education level and annual fertility growth with a lower weight of *ε*

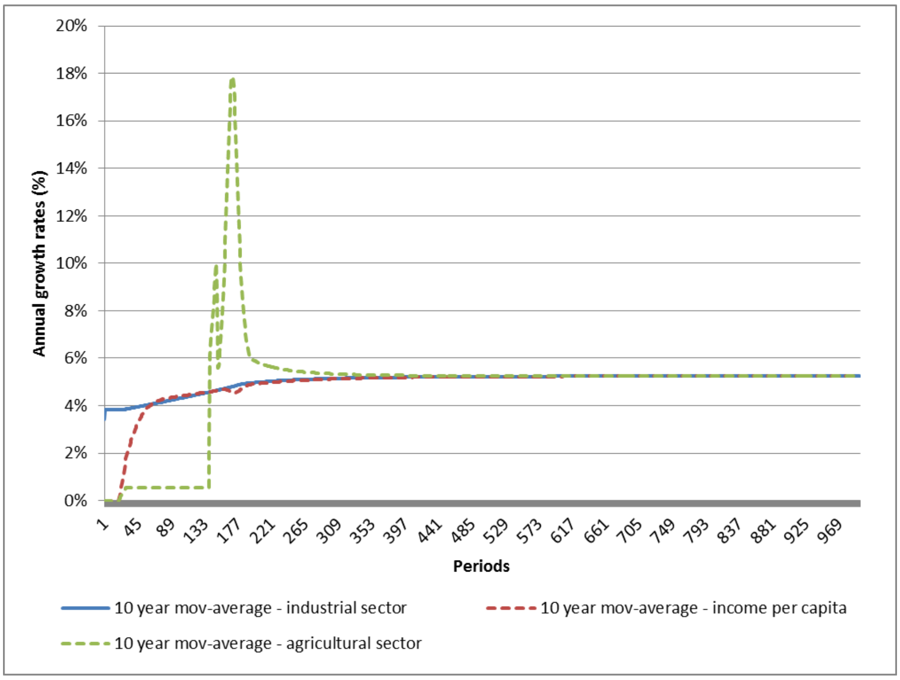

Figure 10B. Technology and income per capita paths with a lower weight of *ε*

***A lower externality effect of the industrial technology* (b)**

We assume in our model a very high level of externalities between the industrial sector and the agricultural sector $b = 0.89$. This affects technology growth rates which become very similar between sectors, as is visible in the baseline case in Section 6.2. One can argue that such a high externality did not occur by the time of the Industrial Revolution and the spillover effects might have been smaller than the ones captured in the baseline scenario. Therefore, we simulate the model using a lower spillover effect. As shown in Table 4B, now $b = 0.7$ and all the other parameters remain constant. This will decrease the growth rate of agricultural technology and lead to a stronger migration of workers to the industrial sector, as explained in subsection 6.2 and depicted in equation (9). As the simulation results show, this will have an impact on the growth rate of the agricultural technology, which is now lower than in the baseline case. Additionally, the fraction of workers in the agricultural sector decreases over time and is reduced to almost zero in the very long run. Therefore, contrary to the baseline case, the dynamics of the model will end up here with a very small agricultural sector, which will affect rents and technology growth, and, in the end, education. This last variable decreases because landowners no longer have a big incentive to promote education as they will not benefit as much from industrial technology and human capital in their sector. Nevertheless, for the medium- run, as shown in Figure 11B, the results remain stable as in the baseline case. These patterns tend to be more extreme the lower the parameter b. As for education and fertility, the results remain quite stable in comparison to the baseline scenario although the emergence of education occurs later than in the baseline scenario. This stems from the fact that industrial technology gains have less externalities in the agricultural sector, which creates disincentives on landowners to support education because their gains will be smaller for lower levels of technology. Only for a sufficiently high A_t^I (remember Lemma 5) do landowners have an interest in supporting education, therefore the spreading of education takes longer for a lower b.

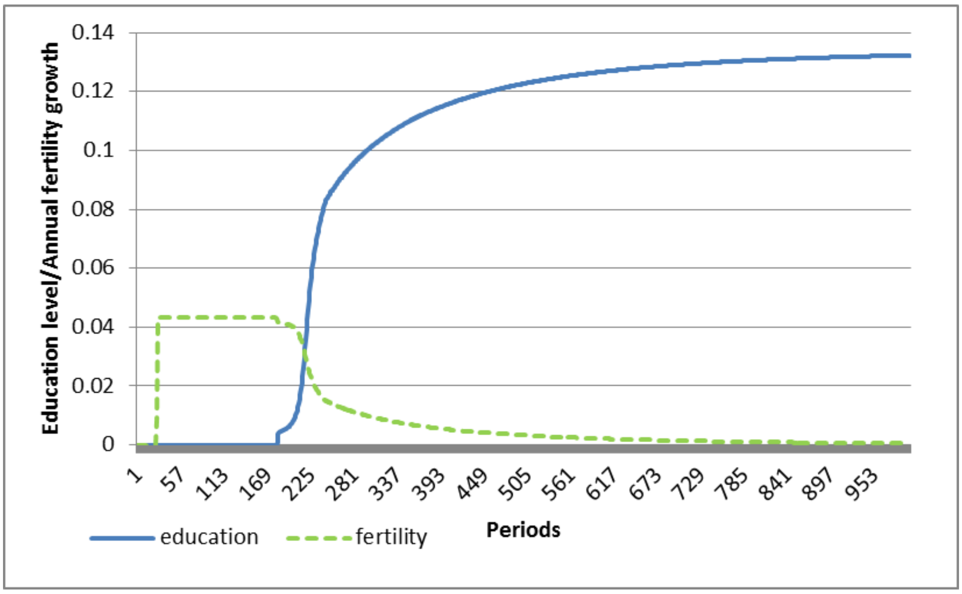

Figure 11B. Paths of education level and annual fertility growth with a higher level of externality, *b*

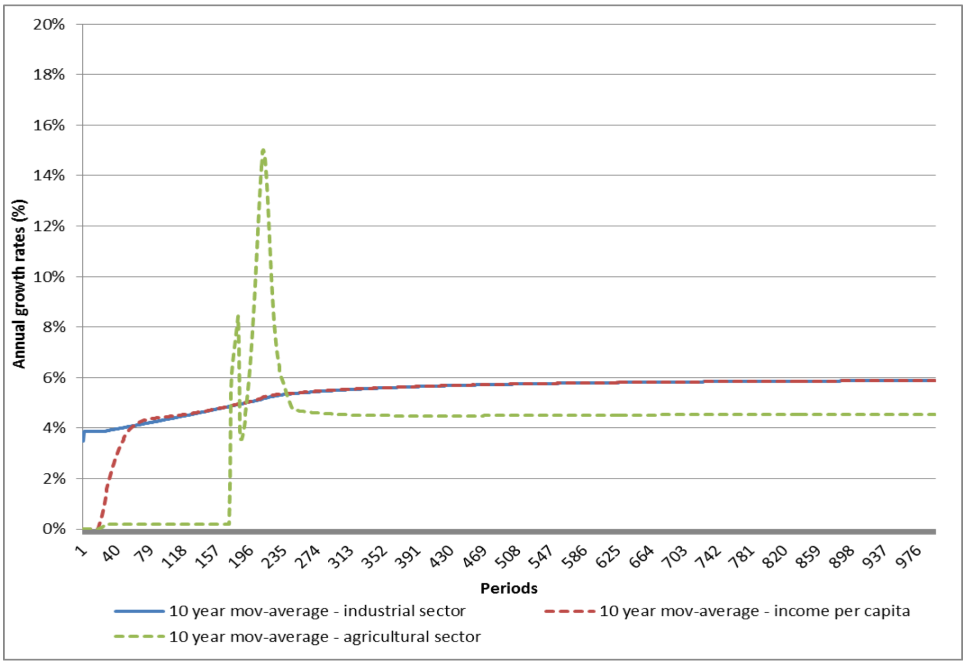

Figure 12B. Technology and income per capita paths with a higher level of externality, *b*

Ideas effect on the industrial technology (ζ)

Finally, regarding the effect of the economy's human capital level on the industrial technology, the value of ζ is higher than the same effect on the agricultural sector but still quite reduced. If we assume a higher value of $\zeta = 0.15$, such that ideas and inventions deriving from human capital and population levels expand the industrial technology even more than we would expect, there will be, on the one hand, a stronger boom of the industrial sector and, on the other hand, higher spillovers on the agricultural sector. The final outcome shows that demographics and education are not significantly affected but technology growth and output per capita grow at a much higher rate. The main dynamics are again not affected – see Figures 13B and 14B.

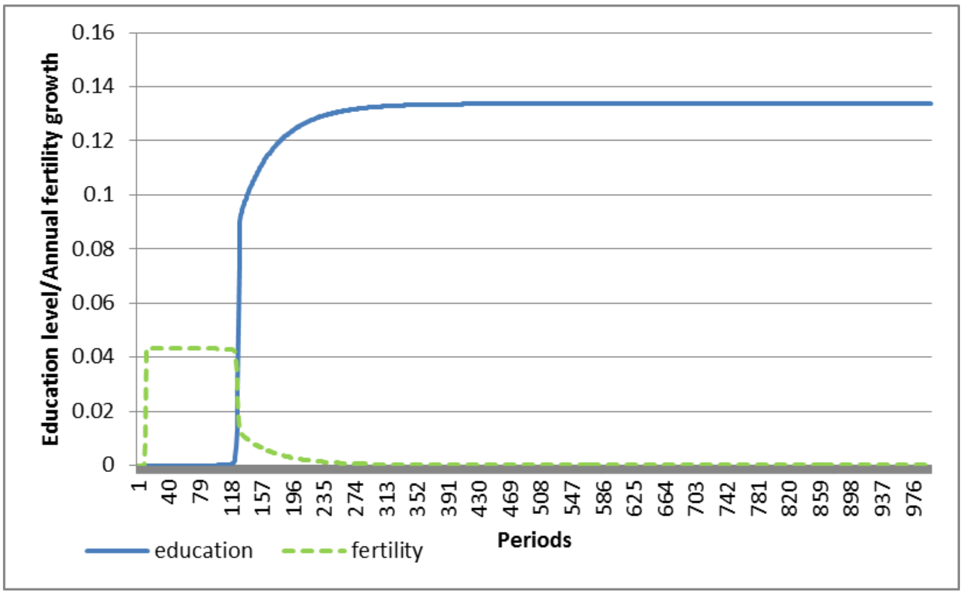

Figure 13B. Paths of education level and annual fertility growth with a higher value of ζ

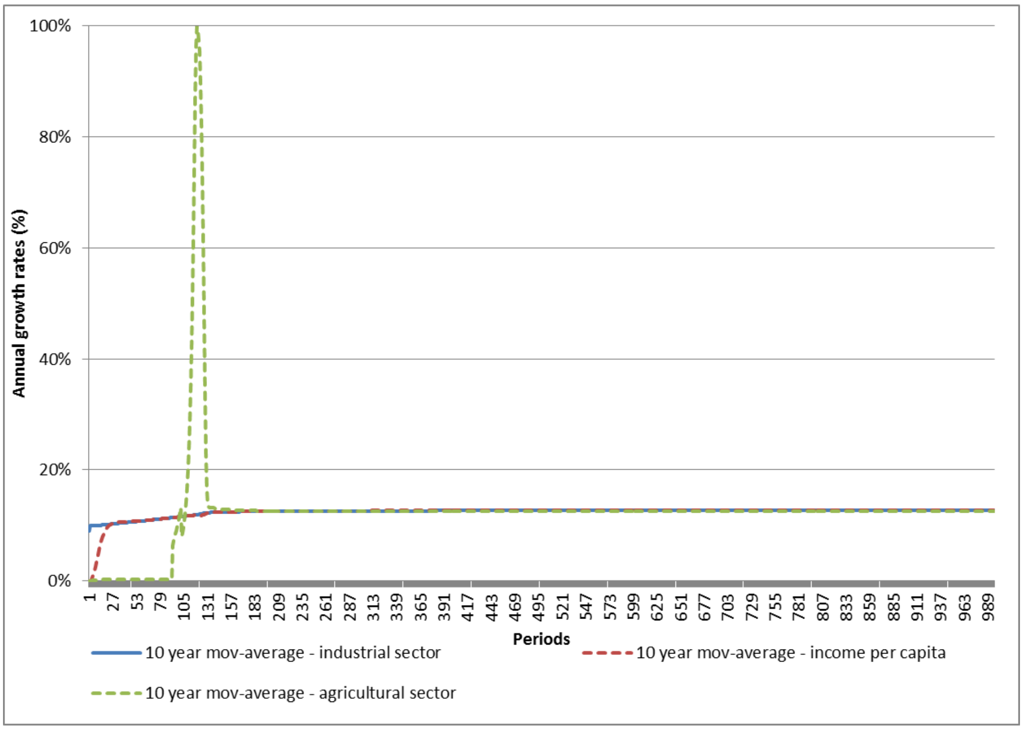

Figure 14B. Technology and income per capita paths with a higher value of ζ

B.11 Proof of Proposition 4

To ascertain the impact of agricultural technology on the decisions of the elite we need to apply the Implicit Function Theorem in equation (23) so that we can derive the impact of A_t^A on t_t. From the sensitivity analysis from Lemma 3 we already know that $\frac{dG}{dt_t} < 0$ for the entire time period, while $\frac{dG}{dA_t^A}$ is oscillatory. However, for the period previous to the onset of the Industrial Revolution $\frac{dG}{dA_t^A} < 0$, while afterwards $\frac{dG}{dA_t^A} > 0$, which in general holds according to our sensitivity analysis. Therefore, for our numerical example and under the range of values of each parameter in the sensitivity analysis:

$$\frac{dt_t}{dA_t^A} = -\frac{\frac{dG}{dA_t^A}}{\frac{dG}{dt_t}} \begin{cases} < 0 & before\ onset\ of\ Industrial\ Revolution \\ > 0 & after\ onset\ of\ Industrial\ Revolution \end{cases}$$

From this we can conclude that the higher the value of A_t^A on the onset of industrialization, the more likely it is that this onset follows sooner. The boost will also be stronger following the education spillovers on both sectors' technologies. Hence, the previous improvements in agriculture (during the 18th century) exert a positive influence on the early rise in education.