DEPOCEN

Working Paper Series No. 2013/17

Geographical factors, Growth and Divergence

Nguyen Thang DAO[1]
Julio DÁVILA[2]

[1] Corresponding author, CORE, Université catholique de Louvain, B-1348 Louvain-la-Neuve, Belgium and Vietnam Centre for Economic and Policy Research (VEPR), Hanoi, Vietnam. Email: nguyen.dao@uclouvain.be
[2] CORE, Université catholique de Louvain, B-1348 Louvain-la-Neuve, Belgium and Paris School of Economics, Paris, France. Email: julio.davila@uclouvain.be.

Geographical factors, Growth and Divergence

Nguyen Thang DAO[1] and Julio DÁVILA[2]

Abstract

This paper develops a unified growth model capturing issues of endogenous economic growth, fertility, and technological progress considering the effects of geographical conditions to interpret the long transition from Malthusian stagnation, through demographic transition to modern sustained growth, and the great divergence in GDP per capita across societies. The paper shows how the interplay of size of "land" and its "accessibility" and technological progress play a very important role for an economy to escape Malthusian stagnation and to take off. Thus differences in these geographical factors lead to differences in take-off timings, generating great divergence across societies.

Keywords: Geographical land, land accessibility, level of technology, human capital, fertility.

JEL Classification: J11, O11, O33

"The factors we have listed (innovation, economies of scale, education, capital accumulation, etc.) are not causes of growth; they are growth." (North and Thomas, 1973, p.2 in *The Rise of the Western World: A New Economic History*)

1. Introduction

The long transition from stagnation to sustained growth along with great divergence across societies is an interesting topic in development economics and has been the subject of intensive research in recent years. This paper aims at contributing to the literature a mechanism to interpret the long transition from stagnation, through demographic transition to modern sustained growth, as well as the great divergence in GDP per capita across societies. The paper highlights the role of "size of land" (i.e. the amount of resource suitable for living and production) and its "accessibility" in supporting an economy in early stage of development to escape stagnation and to take off. There exists a set of these geographical factors under which an economy starting from low initial conditions never escapes stagnation, as well as an other set guaranteeing a take-off. The paper shows that differences in these geographical factors

[1]Corresponding author, CORE, Université catholique de Louvain, B-1348 Louvain-la-Neuve, Belgium and Vietnam Centre for Economic and Policy Research (VEPR), Hanoi, Vietnam. Email: nguyen.dao@uclouvain.be

[2]CORE, Université catholique de Louvain, B-1348 Louvain-la-Neuve, Belgium and Paris School of Economics, Paris, France. Email: julio.davila@uclouvain.be.

The authors thank funding from a research grant "Mobilité Ulysse F.R.S.-FNRS" from the Belgian FNRS. The first author also thanks CORE Foundation for financial support for his PhD program.

This paper presents research results of the Belgian Program on Inter-university Poles of Attraction initiated by the Belgian State, Prime Minister's Office, Science Policy Programming. The scientific responsibility is assumed by the authors.

lead to differences in take-off timings, generating great divergence across societies. The paper also points out a threshold of technological level, which depends positively on the size of productive land and negatively on population size, that households invest in education for their children when the technological level exceeds this threshold.

The major portion of human history is marked by the so-called Malthusian stagnation in which technological progress and the population growth rate are very small. Thomas R. Malthus proposed the most basic description of the relationship between population growth and income in his famous work "*An Essay on the Principles of Population*" first published in 1798. He is credited with the view that population has a tendency to outrun the food supply and to be held in check by war, disease, and starvation. He may have been the first to significantly modify his view that population growth inevitably presses against the food supply (as stated in Gale Johnson, 2000, p.1). The first of two key components in the Malthusian model is the existence of land which is fixed in supply, implying decreasing returns to scale of all other factors of production. The second one is a positive effect of the living standard on population growth. According to the Malthusian model, in the absence of technological progress, the size of population will be self-equilibrating. The living standard will be high when the size of population is small, and population will grow as a natural result of passion between the sexes. When size of population is large, the living standard is low and population will be reduced by the either the intentional reduction of fertility or by malnutrition, disease, and famine, etc. Furthermore, exogenous increases in resources will be offset by increase in population size in the long run. The predictions of the Malthusian model are consistent with historical facts. For thousands of years, the living standard was nearly constant and did not differ considerably across societies. Estimations from Angus Maddison (1982) show that the growth rate of GDP per capita in Europe between 500 and 1500 was zero. Lee (1980) shows that the real wage in England in 1800 was roughly the same to it had been in 1300. Chao (1986) shows that real wages in China at the end of eighteenth century were even slightly lower they had been at the beginning of the first century. Lucas (1999) argues that even in the richest countries, the sustained growth phenomenon in living standards is only a few centuries old. The evolution of population before industrial revolution is also consistent with the predictions of the Malthusian model. For thousand of years, the growth rate of population was nearly zero. Massimo Livi-Bacci (1997) estimates the growth rate of world population from the year 1 to 1750 at around 0.064 percent per year.

Although the Malthusian model explains well the stagnation of nearly all of human history, it fails to explain the demographic transition to the regime of modern sustained growth. The two last centuries are marked by the significant technological progress associated with the industrial revolution and the generalization of basic education. As a consequence, many societies got out of Malthusian stagnation and experienced a considerable increase in the income per capita and population growth rate, as well as increases in human capital. These two last centuries are also marked by the demographic transition leading eventually to a decline in the growth rate of population. However, many other societies seem still to be trapped in the Malthusian stagnation. The transition from Malthusian stagnation through the demographic transition to the modern sustained economic growth and the phenomenon of the great divergence, as depicted in Figure 1 and Figure 2, have shaped considerably the contemporary world economy and needs to be explained.

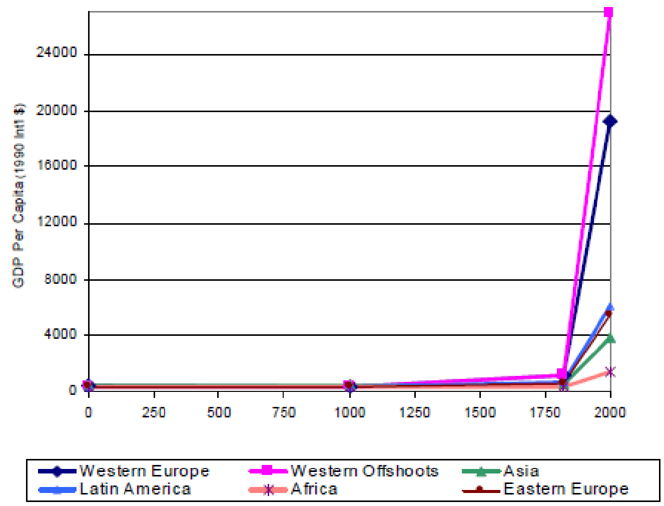

Figure 1. The evolution of regional income per capita: years 0 - 2000. Source: Maddison (2003)[3]

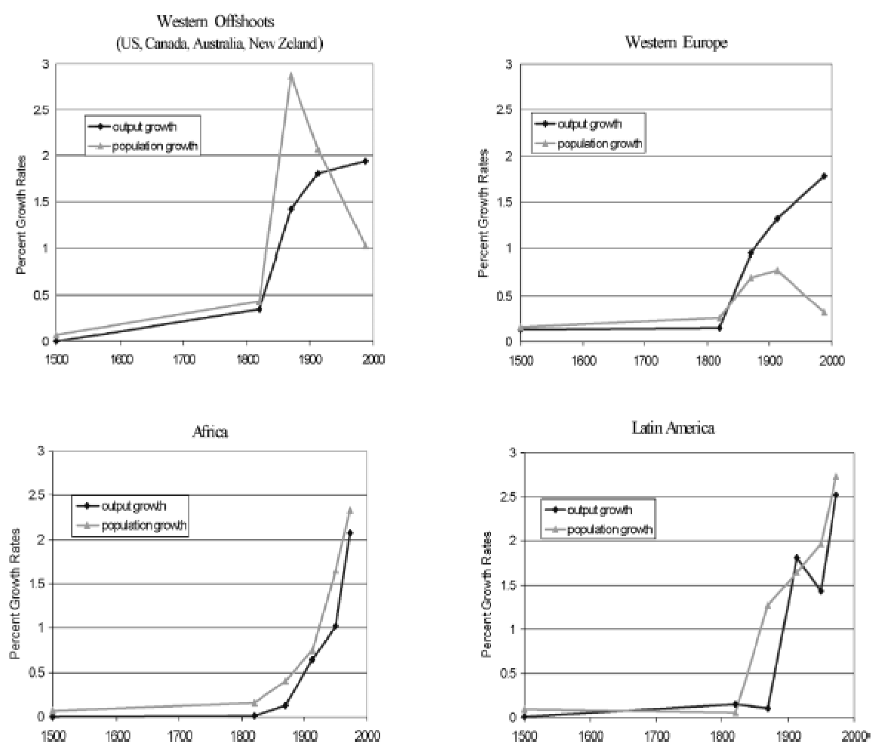

Figure 2. Growth of GDP per capita and population: 1500 - 2000. Source: Maddison (2001)[4]

[3]Quoted in Galor (2005). According to classification from Maddison, "Western Offshoots" consist of the United States, Canada, Australia and New Zealand.
[4]Quoted in Galor (2005)

2. Related Literature

The modeling of the very long transition from thousands of years of stagnation with small population size and growth rate, as well as low economic growth, to the modern regime of large populations and low fertility but sustained economic growth, along with the subsequent great divergence across countries, are among the most significant research challenges facing researchers interested in growth and development. This modeling aims to answer questions like: *(i) what does account for the Malthusian stagnation?; (ii) what is the origin of the sudden spurt in growth rates of output per capita and population?; (iii) what was the cause of the reversal in the positive relationship between income per capita and fertility that existed throughout most of human history?; (iv) what triggered the demographic transition?* (see Galor (2005), p.177). Galor and Weil (2000) first advanced a unified growth model interpreting the historical evolution and interaction mechanisms between population, technology, and output. It encompasses the endogenous transition between three distinct regimes (from Malthusian stagnation through demographic transition to modern growth) that have characterized economic development. The authors focus on the two most important differences between these regimes from a macroeconomic viewpoint: (i) the behavior of income per capita; and (ii) the relationship between the level of income per capita and the growth rate of population. Galor and Weil (2000) show that during the Malthusian period, the dynamical system of an economy would be characterized by a conditional Malthusian equilibrium. Galor and Weil assume that the driving forces for technological progress are the average level of education and the size of population. Technological progress appears nonetheless even when education is zero and population is small. So, eventually, the dynamical system would make the Malthusian equilibrium vanish endogenously, leaving the arena to the gravitation forces of the modern growth regime and permitting the economy to take off and to converge to a modern steady state growth. This model explains well the evolutions of population, technology, and output for societies in Western Europe and many other societies in the world during last two thousand years. Building on this seminal work of Galor and Weil, several papers have tried to merge within a single framework the change of regimes through demographic transition to answer the questions above. An incomplete list of papers includes Boucekkine et al. (2002, 2007), Doepke (2004), Galor and Moav (2002), Hansen and Prescott (2002), Jone (2001), Kogel and Prskawetz (2001), Lagerlof (2007), Lucas (2002), Strulik (2003), Strulik and Weisdorf (2009), Tamura (2002), and Weisdorf (2004). Since technological progress still appears even when education is zero and the size of population is small, all societies eventually escape stagnation to converge to a modern steady state growth. However, in reality, some societies show no sign of escaping the stagnation on their own, in particular small and isolated societies. Also, Galor and Weil (2000) does not explain the great divergence in income per capita across countries in the last two centuries.

A unified growth modeling also needs to answer the fundamental question: *what are fundamental causes of economic growth*? According to Acemoglu (2009, p.109), in explaining cross-countries income differences, any explanation relying simply on technology, physical capital, as well as human capital differences across countries is, at some level, incomplete. Acemoglu argues that if technology, physical capital, and human capital are so important in understanding differences in the wealth of nations and if they can account for many-fold differences in per capita income across countries, then *why some countries do not improve their technology, accumulate physical capital, and invest in human capital as much as others do*? So there should be other and deeper reasons that are referred as fundamental causes of economic growth.

So what could these fundamental causes be? Innumerable causes of economic growth have been proposed in the literature by economists, historians, and social scientists. Acemoglu (2009) classifies the major candidates into four categories of hypotheses: (i) the luck hypothesis, (ii) the geography hypothesis, (iii) the culture hypothesis, and (iv) the institutions hypothesis. In this paper, we focus on the geography hypothesis. Geography refers to all factors that are imposed on individuals as part of the physical, geographic, and ecological environment in which they live. The geography hypothesis, first and foremost, is the fact that not all regions of the world are equally opt for living and production. Nature, that is, the ecological and geographical environment of nations may play a major role in their economic experiences. There are at least three main branches of geography hypothesis, each emphasizing different mechanism for how geography affects prosperity. The first one and also the earliest one is proposed by Montesquieu in 1748. He believed that climate, in particular heat, shaped human attitudes and effort, and through this channel, it affects both economic and social outcome. The second one, which is developed by Gunnar Myrdal, emphasizes the impact of geography on the technologies available to a society, especially in agriculture. Myrdal (1968, p.2121) wrote: "*Serious study of the problems of underdevelopment should take into account the climate and its impacts on soil, vegetation, animals, humans and physical assets - in short, on living conditions in economic development*". The third variant of the geography hypothesis, which is proposed by Jefferey Sachs, links poverty in many areas of the world to their disease burden, emphasizing that "the burden of infectious disease is higher in the tropics than in the temperate zones" (Sachs, 2000, p.32). In our paper, the geographical factor is likely closer to the second version of geography hypothesis.

In the widely popular book "*Guns, Germs, and Steel: The Fates of Human Societies*", Jared Diamond (1997) provides a historical account along with research results from other sciences such as biology, geography, archeology, epidemiology, etc., to explain why the world becomes so unequal across people communities. And why some regions, peoples, and cultures developed more quickly than the others. Diamond pushes the series of causes and consequences back to 13,000 years ago to reach a conclusion that the origin of great divergence is due to initial differences in geographical and biological conditions. Like espousing the second view of geography hypothesis, he argues that geographical differences between Americas and Eurasia determined the timing and nature of settled agriculture and, by means of this channel, shaped whether societies have been able to develop complex organizations and advanced civilian and military technologies. Although whether historical studies of human societies can be pursued as scientifically is still a controversy, the work of Diamond depicts the most general picture of human history during the last 13,000 years and shows a way for other sciences to develop theories of development.

Gallup et al. (1999) address the complex relationship between geography and economic development. They show the ways in which geography may matter growth directly, controlling for economic policies and institutions, as well as the effects of geography on institutions and policy choices. They find that climate and location of a country have large effects on income levels and income growth, through their effects on disease burdens, intrinsic agricultural productivity, and transport cost. Furthermore, they show that geography seems to be a factor for the choice of economic policy itself.

Ashraf and Galor (2011) provide a mechanism linking exogenous effect of geographical isolation on cultural diversity affecting creation and accumulation of knowledge to explain the asymmetric evolutions across societies. They show that societies characterized by less geographical vulnerability to cultural diffusion benefited from enhanced assimilation, lower cultural diversity, and more intense accumulation of society-specific human capital. So, these societies were more efficient with respect to their production-possibility frontiers and flourished in the agricultural stage of development. However, the lack of cultural diffusion diminishing abilities of these societies in adapting to a new technological paradigm, which delayed their industrialization and, hence, their take-off to sustained growth regime. Their empirical analysis shows that (i) geographical isolation prevalent in pre-industrial period has had a persistent negative impact on the extent of contemporary cultural diversity; (ii) pre-industrial geographical isolation had a positive impact on economic development in the agricultural stage but has had a negative impact on income per capita in the course of industrialization; and (iii) cultural diversity has had a positive impact on economic develpopment in the industrialization process.

Of course, geography is not everything in explaning the divergence across countries. But from previous research in the literature, geography seems to play a crucial role in the early stages of development for an economy to escape from stagnation and to take off. Building on this rich literature, this paper reconsiders the role of geographical factors to interpret the long transition from stagnation to modern sustaine growth, as well as interpret the divergence across societies. This paper differs from the literature by highlighling the role of geographical factors in guaranteeing a sufficiently large size of population to make technology grow, so that differences in geographical factors lead to differences in take-off timings, generating divergence.

3. The model

We consider an overlapping generations economy in which each agent lives for two periods, say childhood and adulthood. In each period the economy produces a single homogenous final good by using human capital a factor of production. The production process is harmful for the health of workers making them pay for health costs. In each period of time, the adult agents in the economy maximize their utility by choosing quantity and quality of children as well as their consumptions under the budget constraint.

3.1. Land

We refer by land here to the entire geographical environment supporting the economy. It is a fact that not all regions in the economy are equal in the sense that some regions have better natural conditions for living and producing than the others. Obviously, our lives depend on how suitable the ecosystem around us is. Due to technological constraints, people may not make the most of the available land (i.e. resources and enviroment), e.g. they may just occupy the part of their geographical territory that is the most suitable for their lives given the technology available. This part of land is called productive land whose size depends on the level of technology and land accessibility.[5] The parameter accessibility of land captures its intrinsic suitability for people to live in the ecosystem as a whole, such as temperature, humidity, rainfall, river density, and bio-diversity, etc. So we assume that the size of the productive land of the economy in period t, X_t, is characterized by the following equation

$$X_t = \chi(\theta, A_t)X \tag{1}$$

where $\chi(\theta, A_t) \in [0,1)$, θ is an accessibility parameter of land; A_t is the level of technology at time t; and $X > 0$ is total land (i.e. the entire resources and environment of the economy). Moreover, $\chi(0,0) = 0$, $\chi_\theta(\theta, A_t) > 0$, $\chi_A(\theta, A_t) > 0$, $\chi_{AA}(\theta, A_t) < 0$, and for all $\theta \geq 0$, $\lim_{\theta \to +\infty} \chi(\theta, A_t) = 1$, $\lim_{A_t \to +\infty} \chi(\theta, A_t) = 1$, $\lim_{A_t \to +\infty} \chi_\theta(\theta, A_t) = \lim_{A_t \to +\infty} \chi_A(\theta, A_t) = \lim_{A_t \to +\infty} \chi_{AA}(\theta, A_t) = 0$.

3.2. Production and health cost

The productivity of each unit of time of a household in period t is given by its human capital, h_t, and the level of technology, A_t. We assume that the output produced by households is linear in the amount of time devoted to production. Each period is normalized to be one unit of time. The output produced per unit of time in period t and per household is

$$y_t = f(A_t)h_t$$

where $f(A_t) > 0 \; \forall A_t \geq 0$, $f'(A_t) > 0$, and $f''(A_t) < 0$.[6]

We assume that in each period t, each household is endowed a unit of time. It allocates its time between supplying labor and raising children. Let us denote $(1-\xi_t) \in [0,1]$ to be the fraction of time devoted by a household to raise its children in period t, and ξ_t to be the fraction of time working, determined by the optimal choice discussed in Section 3.4. So the real output or income per household in period t therefore is $\xi_t y_t$.

[5] The size of productive land might also depend positively on the size of population. Here, for simplicity we don't take into account the effect of the population size on the productive land. Introducing the role of population size in productive land formation does not change the qualitative results crucially as long as its marginal effect is not too strong.

[6] In this paper, the production function is simplified in two aspects: (i) We ignore the role of physical capital because we would focus on and highlight the role of human capital and the mechanism for human capital accummulation. (ii) Land does not appear because, in fact, introducing land in the production function does not change the qualitative analysis.

The production process harms the health of workers (e.g. because of associated polution) requiring each of them to pay for a health cost. We assume that for a given level of technology the effect of production on the health of workers depends on the amount of produced output. Technology itself has two opposite impacts on the health of workers. For a given human capital of workers, higher levels of technology lead to higher amounts of produced output. Through this channel, technology harms the health. On the other hand, for a given amount of output, a higher level of technology is less harmful for the health of workers because technology enhances productivity of workers. We define health costs that each household has to pay as a proportional to its output

$$m_t = \varphi(A_t)\xi_t y_t \tag{2}$$

where $\varphi(A_t) \in (0,1)$, $\varphi'(A_t) < 0$ and $\lim\limits_{A_t \to +\infty} \varphi(A_t) = 0$.[7]

3.3. Technological progress and obsolescence

The dynamics of the technological level of the economy depends on the obsolescence rate and on technological progress. So the level of technology in period $t+1$ is defined by

$$A_{t+1} = (1-\lambda)[1+g_t]A_t \tag{3}$$

where $\lambda \in (0,1)$ is the obsolescence rate of technology, g_t is technological progress in t, and $(1-\lambda)[1+g_t]$ is technological growth rate between periods $t+1$ and t. As in Galor and Weil (2000), we assume that g_t depends on the average education, e_t, and the size L_t of the working generation in period t, i.e.

$$g_t = g(e_t, L_t) \tag{4}$$

in which, for any period t, for all $e_t \geq 0$, $L_t > 0$, we have $g(0, L_t) > 0$, $\lim\limits_{L_t \to 0} g(0, L_t) = 0$, and $g_e(e_t, L_t) > g_L(e_t, L_t) > 0$, $g_{LL}(e_t, L_t) < 0$ $\forall L > 0$ $\forall e \geq 0$, and $g(e_t, L_t)$ is not bounded from above.[8]

From equations (3) and (4) we know that if the education of the working generation t is zero, then in period $t+1$ the economy has positive technological growth if and only if the size of population is large enough, i.e.

$$(1-\lambda)[1+g(0, L_t)] > 1 \tag{5}$$

$$\Leftrightarrow \quad g(0, L_t) > \frac{\lambda}{1-\lambda} \tag{6}$$

[7] The qualitative analysis still remains the same for the more general case $\lim\limits_{A_t \to +\infty} \varphi(A_t) = \underline{\varphi} \geq 0$

[8] The assumption $g_e(e_t, L_t) > g_L(e_t, L_t)$ $\forall L > 0$ $\forall e \geq 0$, which implies that the marginal effect of education on technological progress is always stronger than that of population size, is rather reasonable. One can justify that education equips people with knowledge and skills intensively, making people generate new ideas to enhance technological progress, while population size just enhances technological progress by interaction between people.

which implies that for positive technological growth to exist $L_t > \underline{L}$ where $\underline{L}$ satisfies

$$g(0, \underline{L}) = \frac{\lambda}{1-\lambda}$$

3.4. Households

3.4.1. Preferences and constraints

We follow the standard model of household fertility behavior introduced first in Becker (1960) extended to consider the impact of health cost and population density. Households are homogenous. A household chooses the number of children and their quality under the constraint of the unit of time they can use to childrearing and production. The only input required to produce both child quantity and child quality is time.

In each period t, a generation consists of L_t identical working households. Each household lives for two periods. In the first period (say childhood), $t-1$, it uses up a fraction of his parent household's time. In the second period (say parental), t, it is endowed with one unit of time, which it allocates between child-rearing and production. It chooses the optimal mix of quantity and quality of children and it devotes its remaining time to production to consume its income and pay for the health cost. The preferences of the household born in period $t-1$ are defined over the number and quality of its children, n_t and h_{t+1} respectively, as well as from its consumption in period t, c_t, as follows

$$u_t = \gamma \left[\ln n_t + \ln h_{t+1}\right] + (1-\gamma) \ln c_t \tag{7}$$

We assume that the time to raise children physically, regardless education investment, is decreasing in per household space X_t/L_t, i.e. productive land per household. This idea is introduced in Goodsell (1937) and Thompson (1938), and recently cited by de la Croix and Gosseries (2011) to take into account that, when households have small dwellings, child production is more costly and households have fewer children. For simplicity, we assume that the cost in time for raising n_t children physically is $(\frac{L_t}{X_t})^\beta n_t$, where $\beta \in (0,1)$ captures the importance of space in raising children. We define $\frac{L_t}{X_t}$ as the effective population density.

So the fraction of time that households devote to raise n_t children with education e_{t+1} for each child is $1 - \xi_t = n_t[(\frac{L_t}{X_t})^\beta + e_{t+1}]$, and the opportunity cost for doing so is $y_t n_t[(\frac{L_t}{X_t})^\beta + e_{t+1}]$. Households have to pay a health cost m_t, as defined in (2), to compensate for the negative impact of production on their health. Hence, the agent born at date $t-1$ maximizes at date t its utility (7) under the following budget constraint

$$y_t n_t[(\frac{L_t}{X_t})^\beta + e_{t+1}] + c_t + m_t \leq y_t \tag{8}$$

3.4.2. Human capital formation

Galor and Weil (2000) assume that human capital formation of children born at date t,

h_{t+1}, depends positively on education investment, e_{t+1}, and negatively on the growth rate of technological progress from period t to period $t+1$, g_t. They argue that education lessens the obsolescence of human capital in a changing technology. And, hence, households have incentives to invest in education when technological progress appears regardless the level of technology. From our viewpoint, however, the incentives in investing in education for their offspring depend on the level of technology, A_{t+1}, rather than the growth rate of technological progress, g_t. In effect, for an economy with a high enough level of technology, even if there is no technological progress, agents have incentives to educate their offspring in order to able them to make use of the technology. Hence, we assume that

$$h_{t+1} = h(e_{t+1}, A_{t+1}) \tag{9}$$

where $h(e,A) > 0$, $h_A(e,A) < 0$, $h_{AA}(e,A) > 0$, $h_{eA}(e,A) > 0$ $\forall (e,A) \geq (0,0)$; $h_e(e,A) > 0$, $h_{ee}(e,A) < 0$, $\forall (e,A) \neq (0,0)$, $h_e(0,0) = 0$, $\lim_{A \to +\infty} h(e,A) > 0$ $\forall e > 0$ and $\lim_{A \to +\infty} h(0,A) = 0$, and $\lim_{A \to +\infty} h_e(e,A) > 0$ $\forall e \geq 0$.

3.4.3. Household's optimization

In each period t, each household of the working generation t chooses the quantity n_t and quality h_{t+1} of its offspring, and its own consumption c_t after paying for health cost m_t so as to maximize its utility. From (7), (8), and (9), the optimization problem is

$$\max_{\substack{n_t, c_t > 0 \\ e_{t+1} \geq 0}} \gamma \left[\ln n_t + \ln h(e_{t+1}, A_{t+1})\right] + (1-\gamma) \ln c_t$$

subject to

$$y_t n_t [(\frac{L_t}{X_t})^\beta + e_{t+1}]) + c_t + m_t \leq y_t$$

(The first-order conditions are shown to be not only necessary but also sufficient in Appendix A1). Solving the household's optimization problem we have, first,

$$c_t = (1-\gamma)(y_t - m_t) \tag{10}$$

$$n_t = \frac{\gamma[1 - (m_t/y_t)]}{(\frac{L_t}{X_t})^\beta + e_{t+1}} \tag{11}$$

Regarding the households' choice, Becker (1960) advanced the argument that a rise in income and an associated rise in the opportunity cost of raising children makes fertility decline. He suggests that the rise in income induced a decline in fertility because the positive income effect on fertility is dominated by a negative substitution effect brought by the rising opportunity cost of raising children. Similarly, Becker and Lewis (1973) argued that the income elasticity with respect to investment in children's education is greater than the one with respect to quantity of children, and as a consequence the increase in income leads to a decrease in fertility along with an increase in the educational investment for each child. In contrast to the prediction of Beckerian theory, Galor and Weil (2000) showed that when regardless of whether potential income is high or low, increase in income will not change the division of child-rearing time between quality and quantity. However, when income is so low that the household's consumption is binding at the subsistence level, then an increase in potential income will increase the time spent raising children. The household can generate the subsistence consumption with smaller labor force participation and the fraction of time devoted to childrearing increases. Nevertheless, when income is sufficiently high, so as to ensure the consumption is not binding at the subsistence level, then an increase in income will not change the amount of time spent by a household raising children. Hence, according to Galor and Weil (2000), the increase in income may have no effect of the number of children. Interestingly, in taking into account the negative effect of production on health, our model shows that in contrast to predictions of both Beckerian theory and theory of Galor and Weil, given other factors constant, the fertility depends positively on income. Moreover, the fertility rate depends negatively on the health costs and the effective population density. Equation (11) also provides a trade-off between quantity and quality of children which is stated from the literature.

Finally, the first-order condition with respect to e_{t+1} requires the following relationship between e_{t+1} and A_{t+1}, $\frac{L_t}{X_t}$ to hold:

$$G\left(e_{t+1}, A_{t+1}, \frac{L_t}{X_t}\right) = h_e(e_{t+1}, A_{t+1})[(\frac{L_t}{X_t})^{\beta} + e_{t+1}] - h(e_{t+1}, A_{t+1}) \begin{cases} = 0 \text{ if } e_{t+1} > 0 \\ \leq 0 \text{ if } e_{t+1} = 0 \end{cases} \tag{12}$$

4. Sensitivity analysis of household's choices

In this section, we shall study in each period the responses of endogenous variables with respect the changes in exogenous variables.

Proposition 1: *In a competitive overlapping generations economy as set up above, in any period t there exists a threshold of technological level, $\hat{A}_{t+1} = \hat{A}\left(\frac{L_t}{X_t}\right) > 0$, depending on the efective population density, such that households invest in the education of their offspring if, and only if, the level of technology exceeds this threshold, i.e.,*

$$e_{t+1} = e\left(A_{t+1}, \frac{L_t}{X_t}\right) \begin{cases} = 0 \; if \; A_{t+1} \leq \hat{A}\left(\frac{L_t}{X_t}\right) \\ > 0 \; if \; A_{t+1} > \hat{A}\left(\frac{L_t}{X_t}\right) \end{cases} \qquad \forall X_t, L_t > 0$$

moreover

$$\hat{A}'\left(\frac{L_t}{X_t}\right) < 0$$

and $\forall A_{t+1} > \hat{A}\left(\frac{L_t}{X_t}\right)$

$$\frac{\partial e\left(A_{t+1}, \frac{L_t}{X_t}\right)}{\partial A_{t+1}} > 0, \qquad \frac{\partial e\left(A_{t+1}, \frac{L_t}{X_t}\right)}{\partial (\frac{L_t}{X_t})} > 0$$

Proof: See Appendix A2.

We assumed that the size of productive land in period t, X_t, depends on the land accessibility θ, the level of technology in period t, A_t, and geographic size of land, X; and the level of technology in period $t+1$, A_{t+1}, depends average education level of working generation t, e_t, the size of population L_t, and the level of technology in period t, A_t. Therefore, from Proposition 1, Corollary 1 follows

Corollary 1:

$$\frac{\mathrm{d}\hat{A}_{t+1}}{\mathrm{d}\theta} > 0, \ \frac{\mathrm{d}\hat{A}_{t+1}}{\mathrm{d}A_t} > 0, \ \frac{\mathrm{d}\hat{A}_{t+1}}{\mathrm{d}X} \geq 0$$

and $\forall A_{t+1} > \hat{A}_{t+1}$ *then*

$$\frac{\mathrm{d}e_{t+1}}{\mathrm{d}e_t} > 0, \ \frac{\mathrm{d}e_{t+1}}{\mathrm{d}\theta} < 0, \ \frac{\mathrm{d}e_{t+1}}{\mathrm{d}X} < 0, \ \frac{\mathrm{d}e_{t+1}}{\mathrm{d}L_t} > 0$$

Proof: See Appendix A3.

From the Proposition 1 and the Corollary 1, the threshold of technological level $\hat{A}\left(\frac{L_t}{\chi(\theta,A_t)X}\right)$ depends positively on the accessibility of land θ, the level of technology A_t, as well as geographic size of land, but depends negatively on the population size L_t. This happens because a higher land accessibility, as well as a higher level of technology or geographic size of land, results in a larger productive land X_t. The larger productive land, i.e. lower population density, makes the cost of raising children physically less expensive, enhancing the preferences of households toward quantity of children rather than quality of children. So in order to have positive education investment for children, the level of technology should be higher to attract interests of households in educating their children. The larger size of population L_t makes population density increases, raising the cost of childrearing so that households tend to be more interested in the quality rather than the quantity of children. As a result, the threshold of technology decreases in the size of population.

If the level of technology in period $t+1$ exceeds the threshold, education investment for children in period t, e_{t+1}, increases with respect to education of their parents, e_t, and the size of population, L_t. This is because parent's education and the size of population in period t enhance the level of technology in the period $t+1$ then, through this channel, households have more incentive to educate their offsprings. In addition, as we argue above, a larger population makes the cost of raising children increase so that households will be more interested in the quality of children. Education investment depends negatively on the accessibility of land. That is because a higher accessibility of land makes raising children less costly due to the decreasing in population density.

So far we have not considered the general equilibrium yet. However, it is interesting to note how households respond optimally to exogenous changes. Proposition 2 summarizes these responses.

Proposition 2: *In an overlapping generations economy as set up above, the household's optimal choice, in any period t, is such that:*

(i) An increase in level of technology in period $t+1$, A_{t+1}, results in a decline in the fertility rate and an increase in education of children, i.e.,

$$\frac{\partial n_t}{\partial A_{t+1}} \leq 0 \quad \text{and} \quad \frac{\partial e_{t+1}}{\partial A_{t+1}} \geq 0$$

(ii) An increase in potential income of household results in an increase in fertility rate but does not affect the investment in education for children, i.e.,

$$\frac{\partial n_t}{\partial y_t} > 0 \quad \text{and} \quad \frac{\partial e_{t+1}}{\partial y_t} = 0$$

(iii) An increase in the health costs results in a decrease in fertility rate but does not affect the investment in education for children, i.e.,

$$\frac{\partial n_t}{\partial m_t} < 0 \quad \text{and} \quad \frac{\partial e_{t+1}}{\partial m_t} = 0$$

(iv) The effective population density negative effect on fertility rate but has positive effect on education for children, i.e.,

$$\frac{\partial n_t}{\partial (\frac{L_t}{X_t})} < 0 \quad \text{and} \quad \frac{\partial e_{t+1}}{\partial (\frac{L_t}{X_t})} \geq 0$$

Proof:
These statements follow straightforwardly from the FOCs (11) and (12). Q.E.D.

Statement (i) in Proposition 2 implies that under an increase in the level of technology, households have more incentives to invest in the quality of their children and educate them more. The trade-off between quality and quantity of children results in a decline in fertility rate.

Statement (ii) may look surprising as the effect of income on per child education is zero. That is because of the composition of two effects here. The first one is the cost effect in the sense that the higher the income the higher the opportunity cost of educating children. The second one is an income effect: a higher income makes households dispose for a bigger budget for education. In the model these two effects completely offset each other. Notwithstanding, the higher income allows to devote more budget to raise children, resulting in an increase in fertility. So, in fact, the total education expenditures increase along with fertility thanks to any increase in income. One could raise the question, given the actual observations across countries, of *why high-income countries have higher education than low-income countries?* Our model shows that both income and education may have a strong correlation via the levels of technology. Indeed, income (or more precisely labor productivity) depends positively on the level of technology, while high education is driven by a high level of technology via optimal choices of households. The levels of technology that we mention here are two levels of technology in two consecutive periods. However, from the dynamics of technology, these two levels of technology are strongly correlated.

The statement (iii) is also interesting as the health cost negatively affects fertility but does not affect education per child. A higher health cost results in a decline in fertility. In fact, the decline in fertility reduces the total expenditure for education, but education per child remains unchanged because this reduction is offset by the quality quantity trade-off effect.

Statement (iv) makes sense as it refers the effect of population density on fertility and education through the channel of affecting the costs of raising children.

5. Competitive equilibria

Competitive equilibria of this economy are characterized by (i) the agents' utility maximization under constraints, (ii) the households' potential income, (iii) the population dynamics, (iv) the technological progress dynamics, (v) the determinant of productive land, and (vi) the determinant of the health costs. Therefore, a competitive equilibrium is fully determined by the following system of equations (13)-(19), given θ,α, β, γ, and X.

FOCs

$$n_t = \frac{\gamma[1-(m_t/y_t)]}{(\frac{L_t}{X_t})^\beta + e(A_{t+1}, \frac{L_t}{X_t})} \tag{13}$$

$$h_e(e_{t+1}, A_{t+1})[(\frac{L_t}{X_t})^\beta + e_{t+1}] - h(e_{t+1}, A_{t+1}) \begin{cases} = 0 \; if \; e_{t+1} > 0 \\ \leq 0 \; if \; e_{t+1} = 0 \end{cases} \tag{14}$$

Production

$$y_t = A_t h_t \tag{15}$$

Population

$$L_t = n_{t-1} L_{t-1} \tag{16}$$

Technology

$$A_{t+1} = (1-\lambda)[1 + g(e_t, L_t)]A_t \tag{17}$$

Land

$$X_t = \chi(\theta, A_t)X \tag{18}$$

Health costs

$$m_t = \varphi(A_t)\xi_t y_t \tag{19}$$

The competitive equilibrium can be fully characterized by the following reduced system describing the equilibrium dynamics of the economy's population L_{t+1}, technology A_{t+1}, and education e_{t+1}:

$$L_{t+1} = \frac{1 - \frac{1-\gamma}{1-\gamma\varphi(A_t)}}{(\frac{L_t}{\chi(\theta,A_t)X})^{\beta} + e\left((1-\lambda)[1+g(e_t, L_t)]A_t, \frac{L_t}{\chi(\theta,A_t)X}\right)} L_t \tag{20}$$

$$A_{t+1} = (1-\lambda)[1+g(e_t, L_t)]A_t \tag{21}$$

$$e_{t+1} = e\left((1-\lambda)[1+g(e_t, L_t)]A_t, \frac{L_t}{\chi(\theta, A_t)X}\right) \begin{cases} = 0 \; if \; A_{t+1} \leq \hat{A}\left(\frac{L_t}{\chi(\theta,A_t)X}\right) \\ > 0 \; if \; A_{t+1} > \hat{A}\left(\frac{L_t}{\chi(\theta,A_t)X}\right) \end{cases} \tag{22}$$

for a given initial condition L_0, A_0, and e_0.

6. Steady states

6.1. Stagnation Trap

In this section, we study the conditions on geographical size of land X and its accessibility θ under which an economy starting from very low initial conditions never escapes stagnation. We assume an initial education $e_0 = 0$, an intial level of technology $A_0 \leq \underline{A}$ where $\underline{A}$ solves $\underline{A} = \hat{A}\left(\frac{\underline{L}}{\chi(\theta,\underline{A})X}\right)$, and an initial population size $L_0 < \underline{L}$. We characterize below a set of geographical factors that does not support an economy to reach the population size exceeding the critical level $\underline{L}$ guaranteeing a positive net rate of technological progress. As a consequence, the technological level will be lower than the take-off threshold, locking the economy at zero education. Zero education associated with small population size can not guarantee a technological progress able to offset obsolescence, so that the economy can not expand its productive land to enhance fertility and reach a bigger population size. This negative feedback loop prevents the economy from escaping stagnation.

Let S be

$$S = \left\{(X, \theta) \in \mathbb{R}_+^2 : \left(1 - \frac{1-\gamma}{1-\gamma\varphi(\hat{A}(\underline{L}/X))}\right)\left(\frac{\chi(\theta, \hat{A}(\underline{L}/X))X}{\underline{L}}\right)^{\beta} \leq 1\right\} \tag{23}$$

and $\underline{X}$ be such that[9]

$$\left(1 - \frac{1-\gamma}{1-\gamma\varphi(\hat{A}(\underline{L}/\underline{X}))}\right)\left(\frac{\underline{X}}{\underline{L}}\right)^{\beta} = 1 \tag{24}$$

Proposition 3: *If an economy with $(X, \theta) \in S$, and has initial conditions $L_0 < \underline{L}$, $e_0 = 0$, and $A_0 \leq \underline{A}$, where $\underline{A}$ solves $\underline{A} = \hat{A}\left(\frac{L}{\chi(\theta,\underline{A})X}\right)$,[10] then it will be locked in a low stable steady state with small population $\tilde{L}$, zero level of technology $\tilde{A}$, and zero education $\tilde{e}$, specifically*

$$\tilde{L} = \left(\frac{\gamma[1-\varphi(0)]}{1-\gamma\varphi(0)}\right)^{1/\beta}\chi(\theta,0)X < \underline{L} \qquad \tilde{A} = 0 \qquad \tilde{e} = 0$$

Proof: See Appendix A4.

The Proposition 3 characterizes sufficient conditions under which an economy starting from very low initial state will be locked in a low and stable steady state. This proposition applies to economies with an initial state that rather matches that of early stages of development. Nevertheless, the condition in Proposition 3 is not necessary. The proof of Proposition 3 shows that the geographical factors $(X, \theta) \in S$ do not allow an economy starting from low initial conditions to reach a sufficiently large population size ($L > \underline{L}$) to guarantee a positive net technological progress. Therefore, the technology converges to a basic level normalized to zero and there is no educational investment, making the economy converge to a low stable steady state.

5.2. Conditions for escaping stagnation

The previous subsection studies the conditions under which an economy is locked in a low stable steady state. In this subsection we study the conditions under which an economy can escape stagnation and take off. First of all, we study the dynamical system when the economy really takes off in the sense that there is always positive net technological progress, i.e. and $(1-\lambda)[1+g(e_t, L_t)] > 1$ $\forall t \geq \tau$, where τ is a point of time that the economy starts having positive net technological progress, and in the

[9] Note that $\underline{X}$ is well defined since the left-hand side of (24) is a monotonically increasing and continuous function of $\underline{X}$. We assumed that $\varphi(A) \in (0,1)$ $\forall A > 0$. Hence, for given $\underline{L}$, γ, and β, we have

$$\lim_{\underline{X}\to 0}\left(1 - \frac{1-\gamma}{1-\gamma\varphi(\hat{A}(\underline{L}/\underline{X}))}\right)\left(\frac{\underline{X}}{\underline{L}}\right)^{\beta} = 0$$

and

$$\lim_{\underline{X}\to \infty}\left(1 - \frac{1-\gamma}{1-\gamma\varphi(\hat{A}(\underline{L}/\underline{X}))}\right)\left(\frac{\underline{X}}{\underline{L}}\right)^{\beta} = \infty$$

So there exists a unique $\underline{X}$ satisfying (24). And we have $(\underline{X}, \theta) \in S$ $\forall\theta$.

[10] From the proof of Lemma 2 in Appendix A4, we prove that the function $\hat{A}\left(\frac{L}{\chi(\theta,A)X}\right)$ is increasing and strictly concave in A. Moreover, this function is bounded from above by $\hat{A}\left(\frac{L}{X}\right)$, and $\hat{A}\left(\frac{L}{\chi(\theta,0)X}\right) > 0$. Then there exists a unique $\underline{A}$ solving $\underline{A} = \hat{A}\left(\frac{L}{\chi(\theta,\underline{A})X}\right)$.

long run the level of technology grows unboundedly. We assumed that $\lim_{A\to+\infty}\varphi(A)=0$ and $\lim_{A\to+\infty}\chi(\theta,A)=1\ \forall\theta$, so the regime that the economy converges is characterized by

Technology grows at a constant rate $(1-\lambda)[1+g(\bar{e},\bar{L})]$

$$\frac{\gamma}{(\bar{L}/X)^{\beta}+\bar{e}}=1 \tag{25}$$

$$\lim_{A\to+\infty}\left(h_e(\bar{e},A)[(\frac{\bar{L}}{X})^{\beta}+\bar{e}]-h(\bar{e},A)\right)=0 \tag{26}$$

Now, the following assumption on the human capital formation function and child preference parameter suffices to guarantee the existence of a solution to (25) and (26)

$$\lim_{A\to+\infty}\frac{h\left(\frac{\gamma[1-\varphi(0)]}{1-\gamma\varphi(0)},A\right)}{h_e\left(\frac{\gamma[1-\varphi(0)]}{1-\gamma\varphi(0)},A\right)}>\gamma \qquad (A1)$$

In effect, from (25) and (26) we have

$$\lim_{A\to+\infty}\frac{h(\bar{e},A)}{h_e(\bar{e},A)}=\gamma \tag{27}$$

So the existence of a solution to (25) and (26) is equivalent to the existence of a solution to equation (27). In effect, the left hand side of (27) is continuous and monotonically increasing function of $\bar{e}$. We have

$$\lim_{\bar{e}\to0^+}\frac{h(\bar{e},\infty)}{h_e(\bar{e},\infty)}=0<\gamma \tag{28}$$

So from (28) and the assumption $(A1)$ there always exists a unique $\bar{e}\in(0,\frac{\gamma[1-\varphi(0)]}{1-\gamma\varphi(0)})$ solving equation (27). And it is straightforward to compute the steady state size of population

$$\bar{L}=(\gamma-\bar{e})^{1/\beta}X \tag{29}$$

Now we characterize conditions for an economy to escape stagnation and take off. We defined the set ES as follows

$$ES=\left\{(X,\theta)\in\mathbb{R}^2_+:\ \frac{\frac{\gamma[1-\varphi(0)]}{1-\gamma\varphi(0)}}{(\frac{L}{\chi(\theta,0)X})^{\beta}+\bar{e}}\geq1\right\} \tag{30}$$

Applying the implicit function theorem to the FOC (12) with respect to e_{t+1} and A_{t+1}, as well as with respect to e_{t+1} and L_t to $G(e_{t+1},A_{t+1},\frac{L_t}{X_t})=0$, we find that

$\frac{\partial^2 e(A_{t+1},\frac{L_t}{X_t})}{\partial A_{t+1}^2}$ and $\frac{\partial^2 e(A_{t+1},\frac{L_t}{X_t})}{\partial (\frac{L_t}{X_t})^2}$ depend on the third derivatives of the human capital formation function $h(e_{t+1}, A_{t+1})$, so there is enough room to assume that the optimal education investment e_{t+1} is a concave function in the level of technology A_{t+1} and in the population size L_t, i.e.,

$$\frac{\partial^2 e(A_{t+1},\frac{L_t}{X_t})}{\partial A_{t+1}^2} < 0 \quad \forall A_{t+1} > \hat{A}\left(\frac{L_t}{X_t}\right) \qquad (A2)$$

$$\frac{\partial^2 e(A_{t+1},\frac{L_t}{X_t})}{\partial (\frac{L_t}{X_t})^2} < 0 \quad \forall A_{t+1} > \hat{A}\left(\frac{L_t}{X_t}\right) \qquad (A3)$$

Concave responses of the level of education to the level of technology and to the population density are necessary conditions to guarantee that the education investment is bounded from above.[11]

From (12) and the proof of Proposition 1 (see Appendix A2), it is straightforward that

$$G\left(e(A_{t+1},\frac{L_t}{X_t}), A_{t+1}, \frac{L_t}{X_t}\right) = 0 \quad \forall A_{t+1} > \hat{A}\left(\frac{L_t}{X_t}\right)$$

so we have

$$\lim_{A_{t+1}\to\hat{A}(\frac{L_t}{X_t})^+} G\left(e(A_{t+1},\frac{L_t}{X_t}), A_{t+1}, \frac{L_t}{X_t}\right) = 0$$

Since the function $G\left(e_{t+1}, A_{t+1}, \frac{L_t}{X_t}\right)$ is continuous in e_{t+1} and A_{t+1} for all $e_{t+1} \geq 0$ and $A_{t+1} \geq 0$, then

$$G\left(\lim_{A_{t+1}\to\hat{A}(\frac{L_t}{X_t})^+} e(A_{t+1},\frac{L_t}{X_t}), \hat{A}(\frac{L_t}{X_t}), \frac{L_t}{X_t}\right) = 0 \qquad (31)$$

From the proof of of Proposition 1 we know that

$$G\left(0, \hat{A}(\frac{L_t}{X_t}), \frac{L_t}{X_t}\right) = 0 \qquad (32)$$

The function $G(e_{t+1}, A_{t+1}, \frac{L_t}{X_t})$ is decreasing in e_{t+1} for all $e_{t+1} \geq 0$, in effect

$$G_e\left(e_{t+1}, A_{t+1}, \frac{L_t}{X_t}\right) = h_{ee}(e_{t+1}, A_{t+1})(\frac{L_t}{X_t})^{\beta} < 0$$

[11] Alternatively, if the optimal education investment is strictly convex in the level of technology or in the size of population, we may assume, for physiological or some other reasons, that the maximum amount of education investment that a child can receive is bounded from above. As in the model of Galor and Weil (2000), our model also ignores integer constraints on the quantity of children. So households can choose an infinitesimally small quantity of children with infinitely high quality. Thus, introducing integer constraints may be taken as one justification for an upper bound on level of education.

then

$$G\left(e_{t+1}, \hat{A}(\frac{L_t}{X_t}), \frac{L_t}{X_t}\right) < 0 \quad \forall e_{t+1} > 0 \tag{33}$$

Therefore, from (31), (32), and (33) we have

$$\lim_{A_{t+1}\to\hat{A}(\frac{L_t}{X_t})^+} e(A_{t+1}, \frac{L_t}{X_t}) = 0$$

So along with the assumption $(A1)$, we have $e(A_{t+1}, \frac{L_t}{X_t})$ is continuous, nondecreasing in A_{t+1}, and it is increasing concave in A_{t+1} for all $A_{t+1} > \hat{A}(\frac{L_t}{X_t})$.

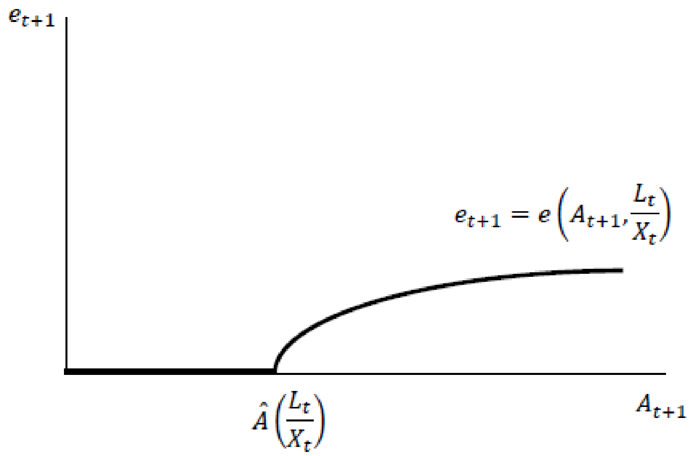

Figure 3.

As from equations (1), (3) and (12), by applying the implicit function theorem with respect to e_{t+1} and A_t for $G\left(e_{t+1}, (1-\lambda)[1+g(e_t, L_t)]A_t, \frac{L_t}{\chi(\theta,A_t)X}\right) = 0$, we find that $\mathrm{d}e_{t+1}/\mathrm{d}A_t$ depends on the slope of $\chi(\theta, A_t)$ with respect to A_t and the derivatives of the function $h(e_{t+1}, A_{t+1})$. In effect, $\forall A_{t+1} > \hat{A}\left(\frac{L_t}{\chi(\theta,A_t)X}\right)$ we have

$$\frac{\mathrm{d}e_{t+1}}{\mathrm{d}A_t} = \frac{h_{eA}(1-\lambda)[1+g(e_t, L_t)][(\frac{L_t}{\chi(\theta,A_t)X})^{\beta} + e_{t+1}] - h_e(\frac{L_t}{X})^{\beta}\frac{\chi_A(\theta,A_t)}{\chi(\theta,A_t)^2} - h_A}{-h_{ee}[(\frac{L_t}{\chi(\theta,A_t)X})^{\beta} + e_{t+1}]}$$

where to simplify notation, we denote h_e, h_A, h_{ee}, h_{eA} as the first derivatives and second derivatives of function $h(e_{t+1}, A_{t+1})$. Hence, there are also enough room to assume that optimal education investment e_{t+1} is increasing in current level of technology A_t, i.e.

$$\frac{\mathrm{d}e_{t+1}}{\mathrm{d}A_t} > 0 \quad \forall A_{t+1} > \hat{A}\left(\frac{L_t}{\chi(\theta, A_t)X}\right) \qquad (A4)$$

In fact, the level of technology A_t has two impacts on education investment e_{t+1}. The first impact is through increase in the level of technology in the next period $t+1$ which affects positively on education, e_{t+1}. The second one is through expansion of productive land which reduces population density lowering costs of raising children, and hence increasing fertility and reducing education investment. The assumption $(A4)$ implies that the first effect always dominates the second effect.

Proposition 4: *Under assumptions* $(A1)$, $(A2)$, $(A3)$, *and* $(A4)$, *any economy satisfying* $(X, \theta) \in ES$ *and starting from arbitrarily low initial conditions* $L_0 > 0$, $A_0 > 0$ *and* $e_0 \geq 0$ *will escape stagnation and converge to a regime of large population, high education and unbounded technological progress.*

Proof: See Appendix A5.

The geographical factors $(X, \theta) \in ES$ guarantee an economy with low initial conditions to reach a sufficiently large population $L > \underline{L}$ which makes positive net technological progress to appear. When technology accummulates enough, households invest education for their offsprings enhancing the take-off process and the economy will converge to a regime of large population, high education and unbounded technological progress.

It is straightforward that $ES \cap S = \oslash$. For intuition, from (23) and (30) we can represent the sets S and ES in the (X, θ) plane.

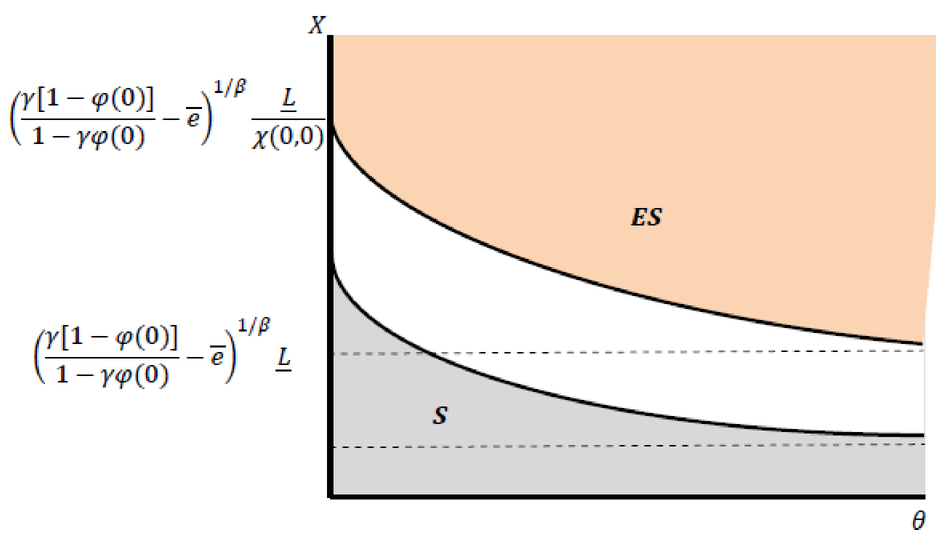

Figure 4. The stagnation set S and take-off set ES

7. Long run Growth and Divergence

In this section, we analyze the long transition of an economy from the Malthusian regime through the demographic transition to modern growth and analyze the impact of geographical factors on the differences in the timing of take-off generating the divergence between countries.

7.1. Long run Growth

In Proposition 3, we study conditions of geographical factors under which an economy starting from very low initial states will be locked in a low and stable steady state. Proposition 4 provides conditions allowing an economy to escape stagnation, then take off and converge to a high and stable steady state. We consider an economy eligible to escape from stagnation, i.e. $(X, \theta) \in ES$, to analyze the long transition from Malthusian stagnation through demographic transition to modern growth. Consider an economy in an early stage of development with a very small population size $L_0 < \underline{L}$, a low level of technology $A_0 \leq \underline{A}$ and zero education $e_0 = 0$. In the early stages of development, population is so small that its effect on technological progress may be dominated by the obsolescence. The low level of technology, associated with negative net technological progress, makes the level of technology in the next period to be below the threshold that incentives households to invest in the education for their offsprings. The low level of technology, on the other hand, has two effects making fertility to be low. First, it makes the size of productive land to be small so that the cost of raising children is high. Second, it makes the cleanness factor to be high and households have to pay higher fraction of income for the health cost, leaving less resources for raising children and consumption. So low fertility makes small size of population in the next period. And along with low level of technology and zero education, the economy continues to experience low stage of development. As long as the size of population is sufficiently small, $L_t < \underline{L}$, no crucial quantitative changes occur in the dynamical system.

In our model, in an early stage of development with the absence of positive net technological progress, the size of geographical land and its land accessibility play crucial roles in enhancing population growth. Over time, with the support of these geographical factors, the slow growth in population taking place in the Malthusian regime will generate positive net technological progress. The technological progress, in turn, reinforces population growth through the channel of impact of technological level on the size of productive land and then on fertility, making population reach a large size. And so on, when the level of technology is high enough, exceeding the threshold, education investment appears reinforcing the technological progress. When education appears due to a high level of technology, there will be a trade-off between quantity and quality of children in the decision of households. And the economy enters in the demographic transition regime characterized by the increasing education investment and decreasing in fertility along with sharp increases in technology.

Over time, the level of technology increases unboundedly thanks to education investment and the large size of population, the economy will enter a modern growth era with large size of population and low fertility but sustained growth. Our model predicts that the level of technology will approach to infinity while education investment converges to a high level $\bar{e}$ and the size of population converges to a constant level, $\bar{L}$, which depends linearly on the size of geographical land (Proposition 4).

7.2. Divergence

In this section we study the impact of geographical factors on timing of escaping from stagnation and then take-off which may generates the Great Divergence across countries. We consider two countries N and M having geographical lands and their land accessibilities as (X^N, θ^N) and (X^M, θ^M), respectively. Assume that these two countries start with the similar and very low initial conditions as characterized in Proposition 3, i.e. $L_0^i < \underline{L}$, $e_0^i = 0$, and $A_0^i \leq \underline{A}^i = \hat{A}\left(\frac{L}{\chi(\theta^i, \underline{A}^i)X^i}\right)$, $i \in \{N, M\}$.

It is obvious that if $(X^N, \theta^N) \in ES$ and $(X^M, \theta^M) \in S$ then there will be a divergence between two countries in the long run in which country N will escape from stagnation era and converge to a high and stable steady state as stated in Proposition 5 while country M will be locked in a low and stable steady state as characterized in Proposition 3.

Now we consider the divergence between two countries in the case both countries are eligible to take off in the long run. We assume that

$$\frac{1 - \frac{1-\gamma}{1-\gamma\varphi(0)}}{\left(\frac{L}{\chi(\theta^N, 0)X^N}\right)^\beta + \bar{e}} > \frac{1 - \frac{1-\gamma}{1-\gamma\varphi(0)}}{\left(\frac{L}{\chi(\theta^M, 0)X^M}\right)^\beta + \bar{e}} \geq 1 \tag{34}$$

implying (X^N, θ^N), $(X^M, \theta^M) \in ES$, and countries N has more initial advantage to escape from stagnation than country M.

For a simplification and an intuition, we will study two simple cases, satisfying the inequalities above, which may help to explain the differences in timing of take off generating the divergence between two countries.[12]

Case 1: $X^N = X^M$, $\theta^N > \theta^M$

In this case, in the early stage of development, the size of productive land in country N is larger than one in country M. Hence, in the early stage, population size of country N is greater than the one of country M because with the same initial sizes of population, the cost of raising children in country N is cheaper than in country M. With the larger size of population, country N will have higher level of technology compared to country M, enlarging the gap in productive lands between two countries. The higher level of technology in country N, on the other hand, reduces the fraction of health costs over the households' potential incomes then leave more resources for households to raise children. In the early stage of development, education is zero, hence households will have more children. By this mechanism, due to the difference in land accessibility, the size of population and level of technology, as well as productive land in country N always greater than those in country M during the early stage of development. Hence, the timing of appearing education arrives in country N before country M. We know from (4) that education has stronger marginal effects on technological progress than population has. And from our model, education investment is a very important preparation for an economy to take off.

[12]For the other cases, as long as satisfying condition (34), the analysis of mechanism for divergence between two countries are analogous to the analyses above.

Therefore, difference in land accessibilities between two countries leads to a difference in timing of take off, generating a divergence between to countries.

Case 2: $X^N > X^M$, $\theta^N = \theta^M$

In this case, the geographical land of country N is greater than the one of country M while their land accessibilities are the same, then with other factors are the same, the size productive land of country N is initially greater than one of country M, making the cost of raising children in country N less than in country M given the same population sizes between two countries. Hence, during the development process, the size of population in country N is always larger than in country M, leading to the timing of take off in country N is before in country M, generating a divergence between these two countries. The mechanism making this divergence in this case is rather similar to that in case 1.

8. Conclusion

This paper develops a unified endogenous growth model to explain the long transition from Malthusian stagnation to modern growth along the thousands of years of the human history and explain the divergence across countries qualitatively. The model captures the geographical factors in explaining the differences in timing of take-off which may generating the great divergence. The model shows the existence of a threshold for level of technology that households will provide education investment for their children if the level of technology exceeds the threshold. This threshold depends positively on the size of productive land and negatively on the size of population (Proposition 1). This finding suggests that for thousand of year there is no investment in education because the levels of technology were not high enough, i.e. below the threshold. This paper, in taking into account the health costs and the effective population density, shows that in contrast to predictions of both Beckerian theory and theory of Galor and Weil, given other factors constant the fertility depends positively on potential income. Moreover, fertility rate depends negatively on the health costs and effective population density (Proposition 2). The model shows the conditions on geographical factors under which an economy starting from very low initial state cannot escape from stagnation trap (Proposition 3), as well as conditions under which an economy will be eligible to escape from stagnation and take-off in the long-run (Proposition 4). Hence, the model suggests that geographical factors play important role, particularly in the very early stage of development.

The model abstracts from several factors such as institutions and cultures, etc. which are relevant for economic growth. These factors would be reflected in their ability to escape from stagnation trap and in the speed of their take-off. Similarly, differences in policies, for example education reforms, would change the dynamics of the model crucially. The construction of a unified growth model to capture other important factors above in explaining the long transition of human history and the great divergence across countries, as well as predicting the world in the future would be significant development from this paper. And this task should be left for our future research.

References

Acemoglu, D. *Introduction to Modern Economic Growth.* Princeton University Press, (2009).

Acemoglu, D., Johnson, S. and Robinson, J. "Reversal of Fortune: Geography and Institutions in the making of the Modern World Income Distribution". The Quarterly Journal of Economics, pp. 1231 - 1294, (2002).

Ashraf, Q. and Galor, O. "Cultural Diversity, Geographical Isolation, and the Origin of the Wealth of Nations", NBER Working Paper (2011).

Becker, Gary S. "An Economic Analysis of Fertility" in Ansley J. Coale, ed., *Demographic and economic change in developed countries.* Princeton, NJ: Princeton University Press, pp. 209 - 240, (1960).

Becker, Gary S. *A Treatise on the Family.* Harvard University Press, (1981).

Boucekkine, R., D. de la Croix and O. Licandro. "Vintage human capital, demographic trends, and endogenous growth". Journal of Economic Theory 104, pp. 340-375, (2002).

Boucekkine, R., D. de la Croix and D. Peeters. "Early Literacy Achievements, Population Density, and the Transition to Modern Growth". Journal of the European Economic Association 5, pp 183-226, (2007).

Chao, Kang. *Man and land in Chinese history: An economic analysis.* Standford, CA: Stanford University Press, (1986).

De la Croix D. and A. Gosseries. The Natalist Bias of Pollution Control. Journal of Environmental Economics and Management, 63, pp. 271 - 287, (2012).

Diamond, Jared. "*Guns, Germs, and Steel: The Fates of Human Societies*". New York: W. W. Norton. (1997)

Gallup, J. L., Sachs, J. D. and Mellinger, A. D. Geography and Economic Development. International Regional Science Review 22 (2), pp. 179 - 232. (1999)

Galor, Oded. "From Stagnation to Growth: Unified Growth Theory". Chapter 4, in Handbook of Economic Growth, Vol 1A. Edited by Philippe Aghion and Steven N. Durlauf, (2005).

Galor, Oded. "The Demographic Transition: Causes and Consequences". NBER Working Paper 17057. Cambridge, MA, (2011)

Galor, Oded and Moav, Omer. "Natural Selection and the Origin of Economic Growth". Quarterly Journal of Economics, Vol. CXVII, Issue 4, (2002).

Galor, Oded and Weil, David N. "From Malthusian Stagnation to Modern Growth". American Economic Review, Vol. 89, pp. 150 - 154, (1999).

Galor, Oded and Weil, David N. "Population, Technology, and Growth: From Malthusian Stagnation to the Demographic Transition and Beyond". American Economic Review, Vol. 90, pp. 806 - 828, (2000).

Galor, Oded., Moav, Omer., and Vollrath, Dietrich. "Inequality in Land Ownership, the Emergence of Human Capital Promoting Institutions and the Great Divergence". Review of Economic Studies, Vol. 76, pp 143 - 179, (2009).

Goodsell, W. "Housing and birth rate in Sweden". American Sociological Review 2 (6), pp 850 - 859, (1937).

Hansen, Gary D. and Prescott, Edward C. "Malthus to Solow". American Economic Review, Vol. 92, pp 1205 - 1217, (2002).

Johnson, D. Gale. "Population, Food, and Knowledge". American Economic Review, Vol. 90, pp 1 - 14, (2000).

Kogel, T. and Prskawetz, A. "Agricultural Productivity Growth and Escape from the Malthusian Trap". Journal of Economic Growth, Vol. 6, pp. 337 - 357, (2001).

Lagerlof, Nils-Peter. "From Malthus to Modern Growth: Can Epidemic explain the three Regimes". International Economic Review, Vol. 44, No. 2, (2003).

Lagerlof, Nils-Peter. "The Galor-Weil model revisited: A quantitative exercise". Review of Economic Dynamics, Vol. 9, pp. 116 - 142, (2007).

Lee, Roland D. "A History Perspective on Economic Aspects of the Population Explosion: The case of Preindustrial England" in Richard A. Easterlin, eds. *Population and Economic change in developing countries.* Chicago: University of Chicago Press, pp. 517 - 566, (1980).

Livi-Bacci, Massimo. *A concise history of world population.* Oxford, Blackwell, (1997).

Lucas, Robert E., Jr. "The industrial Revolution: Past and Future". Mimeo, Department of Economics, University of Chicago, (1999).

Malthus, Thomas R. *An essay on the principle of population.* Cambridge: Cambridge University Press, (1826).

Tamura, R. "Human capital and the switch from agriculture to industry". Journal of Economic Dynamics and Control 27, pp. 207-242, (2002).

Thompson, W.S. "The effect of housing upon population growth". The Milbank Memorial Fund Quarterly, 16 (4), pp. 359 - 368, (1938).

Wrigley, Edward A. and Schofield, Roger S. *The population history of England 1541 - 1871: A reconstruction.* Cambridge, MA: Harvard University Press, (1981).

Appendix

A1. Solving optimization problem of household

The Lagrangian for optimization problem of households in period t

$$\mathcal{L}_t = \gamma\left[\ln n_t + \ln h(e_{t+1}, A_{t+1})\right] + (1-\gamma)\ln c_t$$

$$-\rho_t\left(y_t n_t[(\frac{L_t}{X_t})^\beta + e_{t+1}] + c_t + m_t - y_t\right) - \eta_t e_{t+1}$$

where $\rho_t, \eta_t \geq 0$ is Lagrangian multipliers associated with the constraints.

The Kuhn-Tucker conditions are

$$\frac{\partial \mathcal{L}_t}{\partial n_t} = \frac{\gamma}{n_t} - \rho_t y_t[(\frac{L_t}{X_t})^\beta + e_{t+1}] = 0 \tag{35}$$

$$\frac{\partial \mathcal{L}_t}{\partial e_{t+1}} = \frac{\gamma h_e(e_{t+1}, A_{t+1})}{h(e_{t+1}, A_{t+1})} - \rho_t y_t n_t - \eta_t = 0 \tag{36}$$

$$\frac{\partial \mathcal{L}_t}{\partial c_t} = \frac{1-\gamma}{c_t} - \rho_t = 0 \tag{37}$$

$$\rho_t\left(y_t n_t[(\frac{L_t}{X_t})^\beta + e_{t+1}] + c_t + m_t - y_t\right) = 0 \tag{38}$$

$$\eta_t e_{t+1} = 0 \tag{39}$$

$$\rho_t, \eta_t \geq 0 \tag{40}$$

From equation (37) we have

$$\rho_t = \frac{1-\gamma}{c_t} > 0 \tag{41}$$

then from equation (47) we have

$$c_t = y_t(1 - n_t[(\frac{L_t}{X_t})^\beta + e_{t+1}]) - m_t \tag{42}$$

So from equations (35) and (42) we have

$$n_t = \frac{\gamma\,[1 - (m_t/y_t)]}{(\frac{L_t}{X_t})^\beta + e_{t+1}} \tag{43}$$

and

$$c_t = (1-\gamma)(y_t - m_t) \tag{44}$$

From equations (41), (42) and (44), equation (36) becomes

$$\frac{h_e(e_{t+1}, A_{t+1})}{h(e_{t+1}, A_{t+1})} - \frac{1}{(\frac{L_t}{X_t})^\beta + e_{t+1}} - \eta_t = 0 \tag{45}$$

(i) If $\eta_t = 0$ then (45) becomes

$$\frac{h_e(e_{t+1}, A_{t+1})}{h(e_{t+1}, A_{t+1})} - \frac{1}{(\frac{L_t}{X_t})^\beta + e_{t+1}} = 0 \tag{46}$$

$$\Rightarrow\ h_e(e_{t+1}, A_{t+1})[(\frac{L_t}{X_t})^\beta + e_{t+1}] - h(e_{t+1}, A_{t+1}) = 0 \tag{47}$$

(ii) If $\eta_t > 0$ then $e_{t+1} = 0$ and (45) gives us

$$\frac{h_e(0, A_{t+1})}{h(0, A_{t+1})} - \frac{1}{(\frac{L_t}{X_t})^\beta} = -\eta_t < 0 \tag{48}$$

$$\Rightarrow\ h_e(0, A_{t+1})(\frac{L_t}{X_t})^\beta - h(0, A_{t+1}) < 0 \tag{49}$$

Therefore, from (47) and (49) we have the following relationship between e_{t+1} and A_{t+1}, X_t, L_t:

$$G(e_{t+1}, A_{t+1}, \frac{L_t}{X_t}) = h_e(e_{t+1}, A_{t+1})[(\frac{L_t}{X_t})^\beta + e_{t+1}] - h(e_{t+1}, A_{t+1}) \begin{cases} = 0\; if\; e_{t+1} > 0 \\ \leq 0\; if\; e_{t+1} = 0 \end{cases} \tag{50}$$

Since the optimization problem is not convex, then we have to check the second order conditions, the bordered Hessian matrix is

$$\bar{H} = \begin{pmatrix} 0 & 0 & y_t[(\frac{L_t}{X_t})^\beta + e_{t+1}] & y_t n_t & 1 \\ 0 & 0 & 0 & -1 & 0 \\ y_t[(\frac{L_t}{X_t})^\beta + e_{t+1}] & 0 & -\frac{\gamma}{n_t^2} & 0 & 0 \\ y_t n_t & -1 & 0 & \zeta & 0 \\ 1 & 0 & 0 & 0 & \frac{\gamma-1}{c_t^2} \end{pmatrix}$$

where $\zeta = \gamma \frac{h_{ee}(e_{t+1},A_{t+1})h(e_{t+1},A_{t+1}) - h_e^2(e_{t+1},A_{t+1})}{h^2(e_{t+1},A_{t+1})}$.

We have

$$(-1)^3 \left|\bar{H}_3\right| = -\begin{vmatrix} 0 & 0 & y_t[(\frac{L_t}{X_t})^\beta + e_{t+1}] \\ 0 & 0 & 0 \\ y_t[(\frac{L_t}{X_t})^\beta + e_{t+1}] & 0 & -\frac{\gamma}{n_t^2} \end{vmatrix} = 0$$

$$(-1)^4 \left|\bar{H}_4\right| = \begin{vmatrix} 0 & 0 & y_t[(\frac{L_t}{X_t})^\beta + e_{t+1}] & y_t n_t \\ 0 & 0 & 0 & -1 \\ y_t[(\frac{L_t}{X_t})^\beta + e_{t+1}] & 0 & -\frac{\gamma}{n_t^2} & 0 \\ y_t n_t & -1 & 0 & \zeta \end{vmatrix}$$

$$= -\begin{vmatrix} 0 & 0 & y_t[(\frac{L_t}{X_t})^\beta + e_{t+1}] \\ y_t[(\frac{L_t}{X_t})^\beta + e_{t+1}] & 0 & -\frac{\gamma}{n_t^2} \\ y_t n_t & -1 & 0 \end{vmatrix} = y_t^2[(\frac{L_t}{X_t})^\beta + e_{t+1}]^2 > 0$$

$$(-1)^5 \left|\bar{H}_5\right| = -\left|\bar{H}\right| = \frac{1-\gamma}{c_t^2}\left|\bar{H}_4\right| - \begin{vmatrix} 0 & 0 & 0 & -1 \\ y_t[(\frac{L_t}{X_t})^\beta + e_{t+1}] & 0 & -\frac{\gamma}{n_t^2} & 0 \\ y_t n_t & -1 & 0 & \zeta \\ 1 & 0 & 0 & 0 \end{vmatrix}$$

$$= \frac{1-\gamma}{c_t^2}\left|\bar{H}_4\right| + \begin{vmatrix} 0 & 0 & -1 \\ 0 & -\frac{\gamma}{n_t^2} & 0 \\ -1 & 0 & \zeta \end{vmatrix} = \frac{1-\gamma}{c_t^2}\left|\bar{H}_4\right| + \frac{\gamma}{n_t^2} > 0$$

So, the solution to the household's problem is a maximum indeed.

A2. Proof of Proposition 1

Indeed, we prove that for any X_t and L_t, there always exists unique $\hat{A}_{t+1}$ such that $G(0, \hat{A}_{t+1}, \frac{L_t}{X_t}) = 0$. From the assumptions of the function $h(e_{t+1}, A_{t+1})$ and equation (12), we find that $G(0, A_{t+1}, \frac{L_t}{X_t})$ is monotonically increasing in A_{t+1}. In effect,

$$G_{A_{t+1}}(0, A_{t+1}, \frac{L_t}{X_t}) = h_{eA}(0, A_{t+1})(\frac{L_t}{X_t})^\beta - h_A(0, A_{t+1}) > 0$$

Furthermore, $\forall X_t, L_t > 0$, $\lim\limits_{A_{t+1}\to\infty} G(0, A_{t+1}, \frac{L_t}{X_t}) > 0$, whereas from (9) $h_e(0,0) = 0$ and $h(0,0) > 0$ indicate that $G(0,0,\frac{L_t}{X_t}) < 0$. So there always exists a unique $\hat{A}_{t+1} > 0$ such that $G(0, \hat{A}_{t+1}, \frac{L_t}{X_t}) = 0$, and therefore, as follows from (12), $e_{t+1} = 0$ for $A_{t+1} \leq \hat{A}_{t+1}$.

We have

$$\frac{\partial G(0, \hat{A}_{t+1}, \frac{L_t}{X_t})}{\partial \hat{A}_{t+1}} = h_{eA}(0, \hat{A}_{t+1})(\frac{L_t}{X_t})^\beta - h_A(0, \hat{A}_{t+1}) > 0$$

$$\frac{\partial G(0, \hat{A}_{t+1}, \frac{L_t}{X_t})}{\partial(\frac{L_t}{X_t})} = \beta h_e(0, A_{t+1})(\frac{L_t}{X_t})^{\beta-1} > 0$$

So by applying the implicit function theorem to $G(0, \hat{A}_{t+1}, \frac{L_t}{X_t}) = 0$, we get

$$\hat{A}_{t+1} = \hat{A}(\frac{L_t}{X_t})$$

in which

$$\hat{A}'(\frac{L_t}{X_t}) = -\frac{\beta h_e(0, A_{t+1})(\frac{L_t}{X_t})^{\beta-1}}{h_{eA}(0, \hat{A}_{t+1})(\frac{L_t}{X_t})^{\beta} - h_A(0, \hat{A}_{t+1})} < 0$$

Furthermore, for all $A_{t+1} > \hat{A}_{t+1}$ we have

$$G(e_{t+1}, A_{t+1}, \frac{L_t}{X_t}) = h_e(e_{t+1}, A_{t+1})[(\frac{L_t}{X_t})^{\beta} + e_{t+1}] - h(e_{t+1}, A_{t+1}) = 0$$

and

$$\frac{\partial G(e_{t+1}, A_{t+1}, \frac{L_t}{X_t})}{\partial e_{t+1}} = h_{ee}(e_{t+1}, A_{t+1})[(\frac{L_t}{X_t})^{\beta} + e_{t+1}] < 0$$

$$\frac{\partial G(e_{t+1}, A_{t+1}, \frac{L_t}{X_t})}{\partial A_{t+1}} = h_{eA}(e_{t+1}, A_{t+1})[(\frac{L_t}{X_t})^{\beta} + e_{t+1}] - h_A(e_{t+1}, A_{t+1}) > 0$$

$$\frac{\partial G(e_{t+1}, A_{t+1}, \frac{L_t}{X_t})}{\partial(\frac{L_t}{X_t})} = \beta h_e(e_{t+1}, A_{t+1})(\frac{L_t}{X_t})^{\beta-1} < 0$$

By applying the implicit function theorem, it follows that

$$e_{t+1} = e(A_{t+1}, \frac{L_t}{X_t})$$

where

$$\frac{\partial e_{t+1}}{\partial A_{t+1}} = -\frac{h_{eA}(e_{t+1}, A_{t+1})[(\frac{L_t}{X_t})^{\beta} + e_{t+1}] - h_A(e_{t+1}, A_{t+1})}{h_{ee}(e_{t+1}, A_{t+1})[(\frac{L_t}{X_t})^{\beta} + e_{t+1}]} > 0$$

$$\frac{\partial e_{t+1}}{\partial(\frac{L_t}{X_t})} = -\frac{\beta h_e(e_{t+1}, A_{t+1})(\frac{L_t}{X_t})^{\beta-1}}{h_{ee}(e_{t+1}, A_{t+1})[(\frac{L_t}{X_t})^{\beta} + e_{t+1}]} > 0$$

Q.E.D.

A3. Proof of Corollary 1

We have

$$\hat{A}_{t+1} = \hat{A}(\frac{L_t}{X_t}) = \hat{A}(\frac{L_t}{\chi(\theta, A_t)X})$$

hence,

$$\frac{\mathrm{d}\hat{A}_{t+1}}{\mathrm{d}\theta} = \frac{\partial \hat{A}_{t+1}}{\partial X_t}\frac{\mathrm{d}X_t}{\mathrm{d}\theta} > 0$$

$$\frac{\mathrm{d}\hat{A}_{t+1}}{\mathrm{d}A_t} = \frac{\partial \hat{A}_{t+1}}{\partial X_t}\frac{\mathrm{d}X_t}{\mathrm{d}A_t} > 0$$

We have $\forall A_{t+1} > \hat{A}_{t+1}$

$$e_{t+1} = e(A_{t+1}, \frac{L_t}{X_t}) = e((1-\lambda)[1+g(e_t, L_t)]A_t, \frac{L_t}{\chi(\theta, A_t)X}) > 0$$

hence,

$$\frac{\mathrm{d}e_{t+1}}{\mathrm{d}e_t} = \frac{\partial e_{t+1}}{\partial A_{t+1}}\frac{\mathrm{d}A_{t+1}}{\mathrm{d}e_t} = \frac{\partial e_{t+1}}{\partial A_{t+1}}(1-\lambda)g_e(e_t, L_t)A_t > 0$$

$$\frac{\mathrm{d}e_{t+1}}{\mathrm{d}\theta} = \frac{\partial e_{t+1}}{\partial X_t}\frac{\mathrm{d}X_t}{\mathrm{d}\theta} \leq 0$$

$$\frac{\mathrm{d}e_{t+1}}{\mathrm{d}L_t} = \frac{\partial e_{t+1}}{\partial A_{t+1}}\frac{\mathrm{d}A_{t+1}}{\mathrm{d}L_t} + \frac{\partial e_{t+1}}{\partial L_t} = \frac{\partial e_{t+1}}{\partial A_{t+1}}(1-\lambda)g_L(e_t, L_t)A_t + \frac{\partial e_{t+1}}{\partial L_t} > 0$$

A4. Proof of Proposition 3

In order to prove the Proposition 3, we first prove the Lemma 1 and Lemma 2 belows

Lemma 1: *The following inequality always holds*

$$\chi(\theta, \hat{A}(\underline{L}/X))X \leq \underline{X} \qquad \forall (X, \theta) \in S$$

Proof:

In effect, suppose that there were $(X, \theta) \in S$ such that $\chi(\theta, \hat{A}(\underline{L}/X))X > \underline{X}$, then $X > \underline{X}$ and

$$\left(1 - \frac{1-\gamma}{1-\gamma\varphi(\hat{A}(\underline{L}/X))}\right)\left(\frac{\chi(\theta, \hat{A}(\underline{L}/X))X}{\underline{L}}\right)^{\beta} > \left(1 - \frac{1-\gamma}{1-\gamma\varphi(\hat{A}(\underline{L}/\underline{X}))}\right)\left(\frac{\underline{X}}{\underline{L}}\right)^{\beta} = 1$$

which implies a contradiction that $(X, \theta) \notin S$.

Q.E.D.

Lemma 2:

$$\rho\hat{A}\left(\frac{L_t}{\chi(\theta,A_t)X}\right) < \hat{A}\left(\frac{L_t}{\chi(\theta,\rho A_t)X}\right) \quad \forall\rho\in(0,1) \quad \forall A_t>0$$

Proof: We know from Corollary 1 that $\hat{A}\left(\frac{L_t}{X_t}\right)$ is a increasing in A_t. Now, we prove that $\hat{A}\left(\frac{L_t}{X_t}\right)$ is concave in A_t. In effect,

$$\frac{\mathrm{d}^2\hat{A}\left(\frac{L_t}{X_t}\right)}{\mathrm{d}A_t^2} = \frac{\partial^2\hat{A}\left(\frac{L_t}{X_t}\right)}{\partial X_t^2}\left(\frac{\mathrm{d}X_t}{\mathrm{d}A_t}\right)^2 + \frac{\partial\hat{A}\left(\frac{L_t}{X_t}\right)}{\partial X_t}\frac{\mathrm{d}^2X_t}{\mathrm{d}A_t^2}$$

We have

$$\frac{\partial^2\hat{A}\left(\frac{L_t}{X_t}\right)}{\partial X_t^2} = \frac{-h_{eA}(0,\hat{A}_{t+1})(\frac{L_t}{X_t})^\beta + (1+\beta)h_A(0,\hat{A}_{t+1})}{\left[h_{eA}(0,\hat{A}_{t+1})(\frac{L_t}{X_t})^\beta - h_A(0,\hat{A}_{t+1})\right]^2} < 0$$

hence,

$$\frac{\mathrm{d}^2\hat{A}\left(\frac{L_t}{\chi(\theta,A_t)X}\right)}{\mathrm{d}A_t^2} < 0$$

So $\hat{A}\left(\frac{L_t}{\chi(\theta,A_t)X}\right)$ is increasing concave in A_t. For any given θ and L_t, we consider the following function

$$\Lambda(A_t) = \rho\hat{A}\left(\frac{L_t}{\chi(\theta,A_t)X}\right) - \hat{A}\left(\frac{L_t}{\chi(\theta,\rho A_t)X}\right)$$

We have

$$\Lambda'(A_t) = \rho\frac{\mathrm{d}\hat{A}\left(\frac{L_t}{\chi(\theta,A_t)X}\right)}{\mathrm{d}A_t} - \rho\frac{\mathrm{d}\hat{A}\left(\frac{L_t}{\chi(\theta,\rho A_t)X}\right)}{\mathrm{d}(\rho A_t)} = \rho\left[\frac{\mathrm{d}\hat{A}\left(\frac{L_t}{\chi(\theta,A_t)X}\right)}{\mathrm{d}A_t} - \frac{\mathrm{d}\hat{A}\left(\frac{L_t}{\chi(\theta,\rho A_t)X}\right)}{\mathrm{d}(\rho A_t)}\right] < 0$$

Hence,

$$\Lambda(A_t) < \Lambda(0) = (\rho-1)\hat{A}\left(\frac{L_t}{\chi(\theta,0)X}\right) < 0 \quad \text{i.e.} \quad \rho\hat{A}\left(\frac{L_t}{\chi(\theta,A_t)X}\right) < \hat{A}\left(\frac{L_t}{\chi(\theta,\rho A_t)X}\right)$$

Q.E.D.

Now we prove the Proposition 3. We argue that for this economy, education is always zero and level of technology will converge monotonically to zero. Indeed, we have

$$A_0 \leq \hat{A}\left(\frac{\underline{L}}{\chi(\theta, \underline{A})X}\right) < \hat{A}\left(\frac{\underline{L}}{X}\right) \quad (\text{since}\, \chi(\theta, \underline{A}) \in (0,1))$$

$$\Rightarrow A_0 < \hat{A}\left(\frac{\underline{L}}{\chi(\theta, \hat{A}(\underline{L}/X))X}\right) \leq \hat{A}\left(\frac{\underline{L}}{\underline{X}}\right) \quad (\text{under Lemma 1})$$

and we also have $L_0 < \underline{L}$, hence

$$A_1 = (1-\lambda)[1+g(0,L_0)]A_0 < A_0 < \hat{A}\left(\frac{\underline{L}}{\chi(\theta, A_0)X}\right)$$

$$\Rightarrow A_1 < \min\left\{\hat{A}\left(\frac{L_0}{\chi(\theta, A_0)X}\right), \hat{A}\left(\frac{\underline{L}}{\underline{X}}\right)\right\}$$

$$\Rightarrow \quad e_1 = 0$$

$$\Rightarrow L_1 = \left(1 - \frac{1-\gamma}{1-\gamma\varphi(A_0)}\right)\left(\frac{\chi(\theta, A_0)X}{L_0}\right)^{\beta} L_0 < \left(1 - \frac{1-\gamma}{1-\gamma\varphi(\hat{A}(\underline{L}/\underline{X}))}\right)\left(\frac{\underline{X}}{\underline{L}}\right)^{\beta}\underline{L} = \underline{L}$$

$$\Rightarrow A_2 = (1-\lambda)[1+g(0,L_1)]A_1 < A_1 < (1-\lambda)[1+g(0,L_0)]\hat{A}\left(\frac{\underline{L}}{\chi(\theta, A_0)X}\right)$$

$$< \hat{A}\left(\frac{\underline{L}}{\chi(\theta, (1-\lambda)[1+g(0,L_0)]A_0)X}\right) = \hat{A}\left(\frac{\underline{L}}{\chi(\theta, A_1)X}\right) \quad (\text{under Lemma 2})$$

$$\Rightarrow A_2 < \min\left\{\hat{A}\left(\frac{L_1}{\chi(\theta, A_1)X}\right), \hat{A}\left(\frac{\underline{L}}{\underline{X}}\right)\right\}$$

$$\Rightarrow \quad e_2 = 0$$

$$\Rightarrow L_2 = \left(1 - \frac{1-\gamma}{1-\gamma\varphi(A_1)}\right)\left(\frac{\chi(\theta, A_1)X}{L_1}\right)^{\beta} L_1 < \left(1 - \frac{1-\gamma}{1-\gamma\varphi(\hat{A}(\underline{L}/\underline{X}))}\right)\left(\frac{\underline{X}}{\underline{L}}\right)^{\beta}\underline{L} = \underline{L}$$

$$\Rightarrow A_3 = (1-\lambda)[1+g(0,L_2)]A_2 < A_2 < (1-\lambda)[1+g(0,L_1)]\hat{A}\left(\frac{\underline{L}}{\chi(\theta, A_1)X}\right)$$

$$< \hat{A}\left(\frac{\underline{L}}{\chi(\theta, (1-\lambda)[1+g(0,L_1)]A_1)X}\right) = \hat{A}\left(\frac{\underline{L}}{\chi(\theta, A_2)X, \underline{L}}\right) \quad (\text{under Lemma 2})$$

$$\Rightarrow A_3 < \min\left\{\hat{A}\left(\frac{L_2}{\chi(\theta, A_2)X}\right), \hat{A}\left(\frac{\underline{L}}{\underline{X}}\right)\right\}$$

$$\Rightarrow \quad e_3 = 0$$

$$\vdots$$

and so on, we have for all t

$$A_{t+1} < \min\left\{\hat{A}\left(\frac{L_t}{\chi(\theta, A_t)X}\right), \hat{A}\left(\frac{\underline{L}}{\underline{X}}\right)\right\}$$

$$e_{t+1} = 0$$

$$L_{t+1} = \left(1 - \frac{1-\gamma}{1-\gamma\varphi(A_t)}\right)\left(\frac{\chi(\theta, A_t)X}{L_t}\right)^{\beta} L_t < \underline{L}$$

and

$$A_{t+1} = A_0(1-\lambda)^{t+1}\prod_{i=0}^{t}[1 + g(0, L_i)]$$

Since $L_t < \underline{L}$ $\forall t$ then $(1-\lambda)[1+g(0,L_t)] < 1$ $\forall t$, so the level of technology converges monotonically to

$$A_\infty = 0$$

So for this economy, the existence of a steady state for the system (20)-(22) is the existence of a solution $\tilde{L}$ to

$$\frac{\gamma[1-\varphi(0)]}{1-\gamma\varphi(0)}\left(\frac{\chi(\theta, 0)X}{L}\right)^{\beta} = 1$$

$$\Leftrightarrow \quad \tilde{L} = \left(\frac{\gamma[1-\varphi(0)]}{1-\gamma\varphi(0)}\right)^{1/\beta}\chi(\theta, 0)X$$

And the steady state of the system (20)-(22) is

$$(\tilde{L}, \tilde{A}, \tilde{e}) = \left(\left(\frac{\gamma[1-\varphi(0)]}{1-\gamma\varphi(0)}\right)^{1/\beta}\chi(\theta, 0)X, 0, 0\right)$$

Since for this economy education is always zero during the transition then the dynamical system can be simply characterized by the following system of two equations.

$$L_{t+1} = \frac{\gamma[1-\varphi(A_t)]}{1-\gamma\varphi(A_t)}\left(\frac{\chi(\theta, A_t)X}{L_t}\right)^{\beta} L_t = \psi(L_t, A_t, 0)$$

$$A_{t+1} = (1-\lambda)[1+g(0, L_t)]A_t = \Gamma(L_t, A_t, 0)$$

Let linearize this system around its steady state, we have

$$\begin{pmatrix} L_{t+1} \\ A_{t+1} \end{pmatrix} = \begin{pmatrix} 1-\beta & \frac{\partial\psi(\tilde{L},0,0)}{\partial A_t} \\ 0 & (1-\lambda)[1+g(0,\tilde{L})] \end{pmatrix} \begin{pmatrix} L_t \\ A_t \end{pmatrix} + \begin{pmatrix} \epsilon_L \\ \epsilon_A \end{pmatrix}$$

So the associate Jacobian matrix has two real eigenvalues

$$\lambda_1 = 1-\beta \qquad \lambda_2 = (1-\lambda)[1+g(0,\tilde{L})]$$

where $0 < \lambda_1, \lambda_2 < 1$; hence, this steady state is stable.

Q.E.D.

A5. Proof of Proposition 4

Firstly, we prove the Lemma 3 below

Lemma 3: *Under* $(A2)$ *and* $(A3)$, $e_{t+1} = e\left(A_{t+1}, \frac{L_t}{X_t}\right)$ *is increasing concave in* L_t *for all* $A_{t+1} > \hat{A}\left(\frac{L_t}{\chi(\theta,A_t)X}\right)$.

Proof:

In effect, $\forall A_{t+1} > \hat{A}\left(\frac{L_t}{\chi(\theta,A_t)X}\right)$, we have

$$\frac{\mathrm{d}e_{t+1}}{\mathrm{d}L_t} = \frac{\partial e_{t+1}}{\partial A_{t+1}}\frac{\mathrm{d}A_{t+1}}{\mathrm{d}L_t} + \frac{\partial e_{t+1}}{\partial L_t}$$

hence,

$$\frac{\mathrm{d}^2 e_{t+1}}{\mathrm{d}L_t^2} = \frac{\partial^2 e_{t+1}}{\partial A_{t+1}^2}\left(\frac{\mathrm{d}A_{t+1}}{\mathrm{d}L_t}\right)^2 + \frac{\partial e_{t+1}}{\partial A_{t+1}}\frac{\mathrm{d}^2 A_{t+1}}{\mathrm{d}L_t^2} + \frac{\partial^2 e_{t+1}}{\partial L_t^2}$$

We have

$$\frac{\mathrm{d}^2 A_{t+1}}{\mathrm{d}L_t^2} = (1-\lambda)g_{LL}(e_t, L_t)A_t < 0$$

$(A2)$ and $(A3)$ guarantee that $\frac{\partial^2 e_{t+1}}{\partial A_{t+1}^2} < 0$ and $\frac{\partial^2 e_{t+1}}{\partial L_t^2} < 0$. Therefore,

$$\frac{\mathrm{d}^2 e_{t+1}}{\mathrm{d}L_t^2} < 0 \qquad \forall A_{t+1} > \hat{A}\left(\frac{L_t}{\chi(\theta,A_t)X}\right)$$

Q.E.D.

Now we prove the Proposition 4. Indeed, for such economy there does not exist such a low steady state as characterized in Proposition 3 because the condition

$$\frac{\frac{\gamma[1-\varphi(0)]}{1-\gamma\varphi(0)}}{\left(\frac{L}{\chi(\theta,0)X}\right)^{\beta} + \bar{e}} \geq 1$$

guarantees that when the economy starts from initial conditions $L_0 \in (0, \underline{L})$, $A_0 > 0$ and $e_0 = 0$,[13] it is still able to reach a population size exceeding the threshold $\underline{L}$. When the population size exceeds $\underline{L}$, and with an arbitrary positive level of technology, however small, the net positive technological progress will appear. We now prove that when $L_t > \underline{L}$ then $L_{t+1} > \underline{L}$, and hence $L_{t+i} > \underline{L}$ $\forall i \geq 0$. We rewrite the dynamical system between period t and $t+1$ defined over $L_t > \underline{L}$.

$$L_{t+1} = \frac{\frac{\gamma[1-\varphi(A_t)]}{1-\gamma\varphi(A_t)} L_t}{\left(\frac{L_t}{\chi(\theta,A_t)X}\right)^{\beta} + e\left((1-\lambda)[1+g(e_t,L_t)]A_t, \frac{L_t}{\chi(\theta,A_t)X}\right)} = \psi(L_t, A_t, e_t)$$

$$A_{t+1} = (1-\lambda)[1+g(e_t,L_t)]A_t$$

$$e_{t+1} = e\left((1-\lambda)[1+g(e_t,L_t)]A_t, \frac{L_t}{\chi(\theta,A_t)X}\right) \begin{cases} = 0 \; if \; A_{t+1} \leq \hat{A}\left(\frac{L_t}{\chi(\theta,A_t)X}\right) \\ > 0 \; if \; A_{t+1} > \hat{A}\left(\frac{L_t}{\chi(\theta,A_t)X}\right) \end{cases}$$

For any $A_t > 0$ and $e_t > 0$, there always exists a unique $\hat{L}_t > 0$ such that

$$(1-\lambda)[1+g(e_t,\hat{L}_t)]A_t = \hat{A}\left(\frac{\hat{L}_t}{\chi(\theta,A_t)X}\right)$$

since the left-hand side is unboundedly increasing in $\hat{L}_t$ while the right-hand side is monotonically decreasing in $\hat{L}_t$, and they are both defined over $\hat{L}_t > 0$. For the dynamics of population above, we examine for each case of $\hat{L}_t$ as follows

(i) If $\hat{L}_t \leq \underline{L}$ then $\psi(L_t, A_t, e_t)$ is differential in L_t for all $L_t \geq \underline{L}$. And it is straightforward that $\psi(L_t, A_t, e_t)$ is an increasing function in L_t defined over $L_t \geq \underline{L}$. Therefore, $\forall L_t > \underline{L}$ we have

$$\begin{aligned} L_{t+1} &= \frac{\frac{\gamma[1-\varphi(A_t)]}{1-\gamma\varphi(A_t)} L_t}{\left(\frac{L_t}{\chi(\theta,A_t)X}\right)^{\beta} + e\left((1-\lambda)[1+g(e_t,L_t)]A_t, \frac{L_t}{\chi(\theta,A_t)X}\right)} \\ &> \frac{\frac{\gamma[1-\varphi(A_t)]}{1-\gamma\varphi(A_t)} \underline{L}}{\left(\frac{\underline{L}}{\chi(\theta,A_t)X}\right)^{\beta} + e\left((1-\lambda)[1+g(e_t,\underline{L})]A_t, \frac{\underline{L}}{\chi(\theta,A_t)X}\right)} \\ &> \frac{\frac{\gamma[1-\varphi(A_t)]}{1-\gamma\varphi(A_t)} \underline{L}}{\left(\frac{\underline{L}}{\chi(\theta,A_t)X}\right)^{\beta} + \bar{e}} > \frac{\frac{\gamma[1-\varphi(0)]}{1-\gamma\varphi(0)} \underline{L}}{\left(\frac{\underline{L}}{\chi(\theta,0)X}\right)^{\beta} + \bar{e}} \geq \underline{L} \end{aligned}$$

(ii) If $\hat{L}_t > \underline{L}$ then $\psi(L_t, A_t, e_t)$ is differential and increasing in L_t over $L_t \in (\underline{L}, \hat{L}_t]$ and $L_t \in (\hat{L}_t, \infty)$.

[13] Here we consider the initial education e_0 is zero, the analysis of course does not change quantitatively when $e_0 > 0$.

(iia) If $L_t \in (\underline{L}, \hat{L}_t]$ then

$$L_{t+1} = \frac{\gamma[1-\varphi(A_t)]}{1-\gamma\varphi(A_t)}\left(\frac{\chi(\theta, A_t)X}{L_t}\right)^{\beta} L_t > \frac{\gamma[1-\varphi(A_t)]}{1-\gamma\varphi(A_t)}\left(\frac{\chi(\theta, A_t)X}{\underline{L}}\right)^{\beta} \underline{L} > \frac{\frac{\gamma[1-\varphi(0)]}{1-\gamma\varphi(0)}\underline{L}}{\left(\frac{\underline{L}}{\chi(\theta,0)X}\right)^{\beta} + \bar{e}} \geq \underline{L}$$

(iib) If $L_t \in (\hat{L}_t, \infty)$ we have

$$L_{t+1} > \lim_{L_t \to \hat{L}_t^+} \frac{\gamma[1-\varphi(A_t)]}{1-\gamma\varphi(A_t)}\left(\frac{\chi(\theta, A_t)X}{L_t}\right)^{\beta} L_t = \frac{\gamma[1-\varphi(A_t)]}{1-\gamma\varphi(A_t)}\left(\frac{\chi(\theta, A_t)X}{\hat{L}}\right)^{\beta} \hat{L}$$

$$> \frac{\gamma[1-\varphi(A_t)]}{1-\gamma\varphi(A_t)}\left(\frac{\chi(\theta, A_t)X}{\underline{L}}\right)^{\beta} \underline{L} > \frac{\frac{\gamma[1-\varphi(0)]}{1-\gamma\varphi(0)}\underline{L}}{\left(\frac{\underline{L}}{\chi(\theta,0)X}\right)^{\beta} + \bar{e}} \geq \underline{L} \qquad \forall L_t > \hat{L}$$

So for this economy, in the long run $L_t > \underline{L}$, guaranteeing positivity of net technological progress, and the level of technology approaches monotonically to infinity. So in the long run, the whole dynamical system of the economy will be

$$L_{t+1} = \frac{\gamma L_t}{(\frac{L_t}{X})^{\beta} + e(A_{t+1}, \frac{L_t}{X})}$$

$$A_{t+1} \to \infty$$

$$e_{t+1} = e(A_{t+1}, \frac{L_t}{X})$$

and the sustained growth is characterized by

$$\bar{L} = (\gamma - \bar{e})^{1/\beta} X, \qquad \bar{A} = \infty, \qquad \frac{h(\bar{e}, \infty)}{h_e(\bar{e}, \infty)} = \gamma$$

Indeed, this dynamical system can be reduced in a single dynamical equation of population as follows

$$L_{t+1} = \frac{\gamma L_t}{(\frac{L_t}{X})^{\beta} + e(\infty, \frac{L_t}{X})} = \mathcal{F}(L_t)$$

And the stability of the steady state of this dynamical equation guarantees the stability of the steady-state size of population and the steady-state education investment. We differentiate this equation with respect to L_t, evaluated at $L_t = \bar{L}$, we have

$$\mathcal{F}'(\bar{L}) = \frac{\gamma\left[(\frac{\bar{L}}{X})^{\beta} + e(\infty, \frac{\bar{L}}{X})\right] - \gamma\bar{L}\left[\beta X^{\beta}\bar{L}^{\beta-1} + \frac{1}{X}\frac{\partial e(\infty, \frac{\bar{L}}{X})}{\partial \frac{\bar{L}}{X}}\right]}{\left[(\frac{\bar{L}}{X})^{\beta} + e(\infty, \frac{\bar{L}}{X})\right]^2}$$

$$= 1 - \frac{\beta(\frac{\bar{L}}{X})^{\beta} + \frac{\bar{L}}{X}\frac{\partial e(\infty,\frac{\bar{L}}{X})}{\partial(\frac{\bar{L}}{X})}}{\gamma} < 1$$

Since the function $e(\infty, \frac{\bar{L}}{X})$ is nondecreasing concave in $\frac{\bar{L}}{X}$ and $e(\infty, 0) = 0$ then $\frac{\bar{L}}{X}\frac{\partial e(\infty,\frac{\bar{L}}{X})}{\partial(\frac{\bar{L}}{X})} < e(\infty, \frac{\bar{L}}{X}) = \bar{e}$. Hence,

$$\mathcal{F}'(\bar{L}) > 1 - \frac{\beta(\frac{\bar{L}}{X})^{\beta} + \bar{e}}{\gamma} > 1 - \frac{(\frac{\bar{L}}{X})^{\beta} + \bar{e}}{\gamma} \qquad \text{(since } 0 < \beta < 1\text{)}$$

$$\Rightarrow \mathcal{F}'(\bar{L}) > 0$$

Therefore $0 < \mathcal{F}'(\bar{L}) < 1$ implying that the steady-state size of poupualtion and steady-state education investment are stable.

Q.E.D.