Università degli Studi di Salerno

Scuola Dottorale "Antonio Genovesi"
Dottorato di Ricerca in Economia del Settore Pubblico - XI Ciclo

Essays on Human Capital, Long Run Growth and Comparative Development

Candidato:

Mario F. Carillo

Coordinatore:
Chiar.mo Prof.
Sergio Destefanis

Supervisor:
Chiar.mo Prof.
Marcello D'Amato

Anno Accademico 2013-2014

Contents

List of Figures

List of Tables

Acknowledgments

I'm grateful to Sergio Destefanis e Marcello D'Amato for advice and support over the years of the program. I also thank Oded Galor for advice and suggestions and to seminar participants at Brown University and ACDD conference (Strasbourg) for helpful comments.

A special thank of mine goes to Maria Rosaria Carillo for her special support and encouragement in following my own way.

Introduction

This research argues that differences in the distribution of human capital across countries and their impact on the advancement and the adoption of technology contributed to the differential timing of the transition from the Malthusian stagnation to modern growth and the persistent differences in income per capita across the globe. Polarization in the distribution of human capital within an economy implied a trade-off between innovation and adoption of technologies that, in turn, influenced the transition from stagnation to growth. Despite the contribution of the upper tail of the human capital distribution to technological innovation, the absence of wide group of educated individuals among the working population delayed technology adoption and the transition from stagnation to growth.

The work is developed in three chapters. Chapter 1 is an introduction to the theories of long run growth and adoption of technology on which this research stands. First, it introduces the Malthusian theory with its theoretical and empirical predictions. Second, it reviews some of the theories of the demographic transition, which is a phenomenon that the most advanced economies of the globe experienced starting from the beginning of the 19th century and that will be extensively analyzed in the following chapters. The emphasis of Chapter 1 is on the part of the literature that deals with the determinants of the demographic transition that focus on human capital as fundamental engine of technology advancement and accumulation of knowledge, in turn determining the timing of the transition from stagnation to growth. Finally, Chapter 1 highlights both empirical and theoretical literature that deals with adoption of technology as determinant of economic development via growth in technology. The overview highlights, on one hand, the relevance of the emergence of human capital sufficiently diffused in the population as mechanism that can explain cross country differences in the timing of the transition. On the other hand, the focus is on the importance in considering how technology adoption and technological innovation interact in order to explain technological progress. Which has been widely considered in the economic growth literature with focus on short run proximate causes of growth, but not (to my knowledge) in the long run economic growth literature which analyses the fundamental determinants of economic development, which is the contribution of this work.

Chapter 2 presents the theoretical model, which is the main contribution of this work. Such model advances the above mentioned hypothesis that the distribution of human capital had an effect on the demographic transition and its timing, affecting the long run performance of the economy. First, the chapter introduces historical examples in which innovations were not adopted in the production process and thus did not contribute to economic development. Second, the theoretical model is introduced. In particular, the implication of the model can be summarized in three main predictions. The first prediction advances that economies characterized by significantly low, or extremely high, polarization in the distribution of human capital will be disadvantaged in terms of the transition from stagnation to growth. In other words, highly polarized societies were such that knowledge was in the hands of a small fraction of the population (such as small elites or noble families). Such economies experienced a late transition because, despite the contribution to innovation given by the upper tail of the human capital distribution, new innovations were not understood and thus adopted in the production process by the uneducated population. On the opposite side, economies characterized by extremely low polarization of the human capital distribution were also disadvantaged from the point of view of the timing of the transition. The reason is that, in these societies, despite the fact that the population was sufficiently educated to understand and adopt innovations, the scarcity of highly educated individuals was detrimental for the process of *innovation* of new technologies.

The second prediction of the model describes the effects of the distribution of human capital after the demographic transition. In particular, societies that are characterized by low polarized distribution of human capital, despite the late transition timing, have an *advantage* in the after-transition growth rate of the economy. The reason is that such economies experience a longer Malthusian era with a larger accumulation of population. Therefore, once the transition to industrialization is completed (as enshrined by the demographic transition) the large population will result in an advantage for the performance of the economy. These predictions are consistent with the growth miracle of the "Asian Tigers". Namely, economies that were characterized by low polarization in the distribution of human capital (with respect to Europe) experienced a late demographic transition yet with a significant advantage in growth rates.

The last prediction of the model is not directly related to human capital, yet it describes the effect of a change in parameters of the model on fertility choices. Namely societies that, due to environmental conditions, are characterized by higher cost of raising children (which is parametrized in the model) relative to the cost of human capital, will experience the demographic transition in advance but with a lower after-transition growth rate. Here the mechanism works as follows. Larger cost of raising children implies a lower relative cost of human capital, in turn affecting technology accumulation and thus implying early demographic transition. However, the high cost of raising children is detrimental in the long run because the economy will be characterized by smaller population.

The latter prediction of the model is empirically investigated in Chapter 3. In particular, the empirical approach considers calories intake as a proxy for cost of raising children in early stages of development. Since calories intake is strongly correlated with body size, the empirical specification exploits the cross-country variations in body size as explanatory variable for the differences in the timing of the take-off from stagnation to growth that we observe across economies today. The findings are consistent with the theoretical prediction. Namely, countries that are characterized by environmental conditions that imply high cost of raising children (as proxied by average body size) are on average characterized by early demographic transition.

Chapter 1

Literature Review

1.1 Introduction

For most of human existence the world has been in an epoch of stagnation during which living standards were at level of subsistence. Only in the last two centuries, we observe an improvement in living standards that is indicated by an increase in GDP per capita, that was nearly zero in early stages of development. Figure 1.1 illustrates this point[1].

In particular, we observe that the all world was in an epoch of stagnation before the second half of the eighteenth century, where GDP per capita is low in all economies and cross country disparity in income and in living standards is nearly nonexistent. After the Industrial Revolution, we observe the abandon of the subsistence toward an era of sustained growth in GDP per capita, and consequently living standards, with significant effects on the differences in income per capita distribution across countries, that is the reason why is also called 'Great Divergence'. Namely, The ratio of GDP per capita between the richest region and the poorest region in the world was only 1.1:1 in the year 1000, $2:1$ in the year 1500, and $3:1$ in the year 1820. In the course of the 'Great Divergence' the ratio of GDP per capita between the richest region and the poorest region has widened considerably from a

[1] According to Maddison's classification, "Western Offshoots" consist of the United States, Canada, Australia and New Zealand.

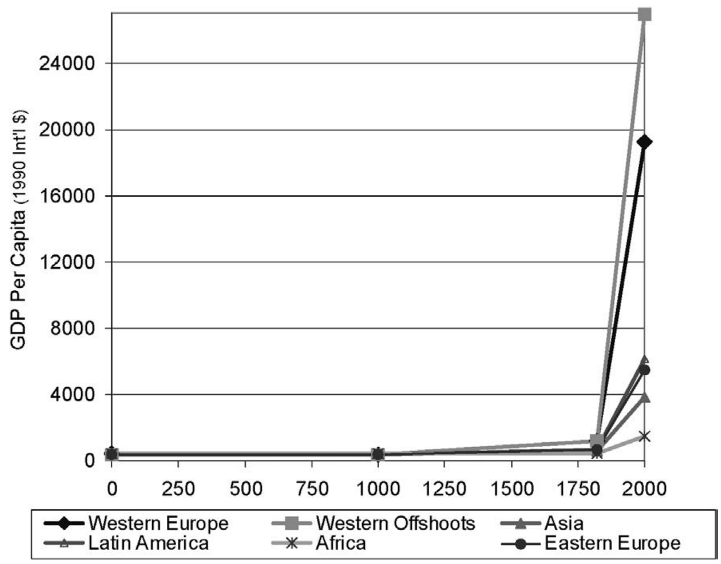

Figure 1.1: The evolution of regional income per capita over the years 1-2000. Source: Maddison (2003).

modest 3 : 1 ratio in 1820, to a 5 : 1 ratio in 1870, a 9 : 1 ratio in 1913, a 15 : 1 ratio in 1950, and a 18 : 1 ratio in 2001 (Galor, 2005).

The period of stagnation is an era in which income per capita is constant over time. As it will be analyzed in detail in section 1.2, this is a period of great dynamism in which population growth offsets any increase in total GDP, in turn, keeping per capita income constant over time. Such effect was first analyzed by Malthus (1798), therefore this long period of time is oftentimes called Malthusian era. Interestingly, the Malthusian mechanism of population growth that keeps living standards at level of subsistence was coming to an end in those days in which Malthus was publishing his famous essay.

As will be extensively analyzed in the following, the mechanisms that determine the transition from the Malthusian era to an era of sustained economic growth, and in particular its timing, plays a crucial role in understanding fundamental questions of comparative development. Such transition is a demographic phenomenon, indeed it determines the end of the pressure that fertility operates keeping low living standards, ultimately increasing growth in output per capita. Moreover, understanding the mechanisms that trigger

the timing of the transition is crucial because we observe that countries that experience the demographic earlier are more developed nowadays. In other words there is a strong and significant correlation between the timing of the transition and GDP per capita today, therefore the understanding of this specific aspect is crucial. Indeed, the analysis of the transition timing will be the fundamental object of the analysis in chapter 2 as well as the core of this work.
Clearly, there is huge evidence that highlights the presence of large disparities in income across countries today, and such inequality doesn't seem to shrink over time. In particular, adopting contemporaneous data to analyze the variations in the disparities in cross-country GDP per capita today, it is evident that the gap is not shrinking over time. Figure 1.2 illustrates the distribution of income per capita across countries. In particular, it is evident that the gap widened from 1960 to 2000. We can analyze in more detail

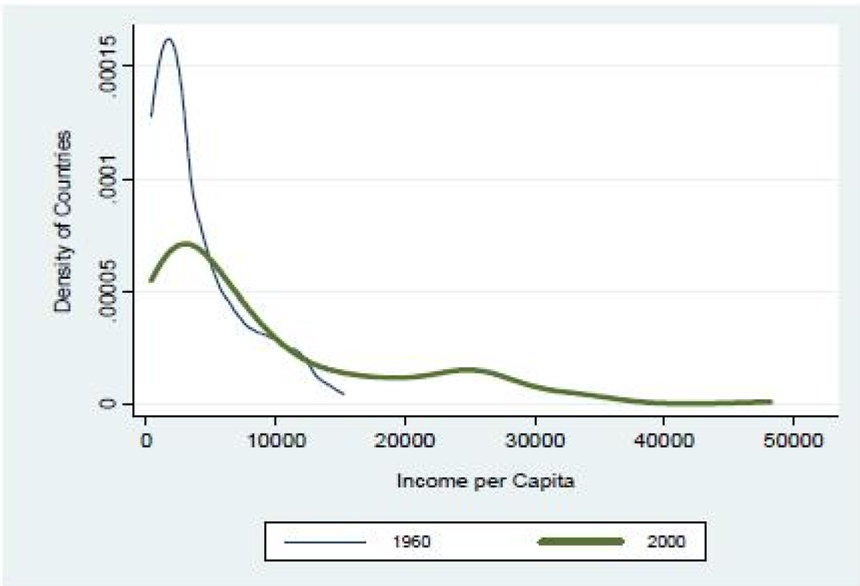

Figure 1.2: Evolution of disparity in Income per capita over the period 1960-2000.

the distribution of income across countries in figure 1.3, where it is depicted the distribution of the income per capita, normalized using the natural logarithm, in three years: 1960, 1980 and 2000. The figure depicts that there has been no absolute convergence over the period considered and that we actually observe divergence, which is represented by the fact that the distribution of income across countries becomes more and more disperse over the period considered.

Figure 1.4 highlights that economies over the period from 1960 to 2000 are

Figure 1.3: Persistent inequality in world income distribution over the years 1960-2000.

not characterized by any source of convergence. Namely, countries that were poor relative to the United States in 1960 are still poor 40 years after as graphically confirmed by the proximity of the data points to the 45 degree line of figure 1.4.

The data presented above display two important features. First, in the data we do not observe any source of absolute convergence of income per capita across countries. Second, all economies were in an era of stagnation prior to the Industrial Revolution after which the disparity in income widened. Therefore, the fundamental challenge consists in developing a theory that takes into account the evolution of economies from an epoch of stagnation to an era of economic growth in order to understand the disparity in income

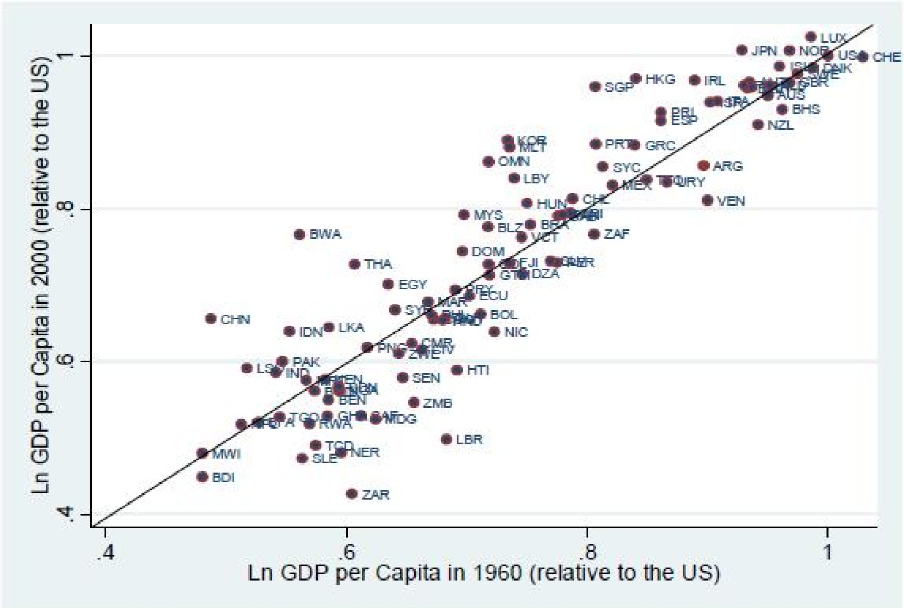

Figure 1.4: Lack of Convergence across the Globe: 1960-2000.

that we observe across countries today.

The chapter is organized as follows. Section 1.2 reviews the fundamental structure of the Malthusian theory, highlighting the features of the model which is the fundamental theory on which the model presented in chapter 2 is based. Section 1.3 describes part of the literature related to the demographic transition. In particular, the focus is on the literature that is related to human capital as main mechanism for the transition. Where human capital, in this framework, is both considered in the form of education and in the form of health. Also some of the literature related to institutions that are conducive to the accumulation of human capital is summarized. Section 1.4 deals with the analysis of technology adoption in the literature, both from the theoretical point of view and from an empirical point of view. The aim is to present such part of the literature of technology adoption because the hypothesis advanced in this research aims to apply a more comprehensive view of technology, as developed in the technology adoption literature, in the context of long run growth and demographic transition. Last section concludes.

1.2 The Malthusian Epoch

1.2.1 Main Features of the Malthusian Theory

During the Malthusian Epoch societies are characterized by living standards at the level of subsistence. Eventual improvement in technology or land discoveries are translated in a transitory improvement in living standards, which in turn is translated in a larger number of children that, in turn, reduces again the living standards. This dynamism of population and income would be reflected in fluctuations of GDP per capita over time (see figure 1.5) around a constant long-run average as observed in the data depicted in figure 1.1. Due to the Malthusian mechanism, fluctuations in income per

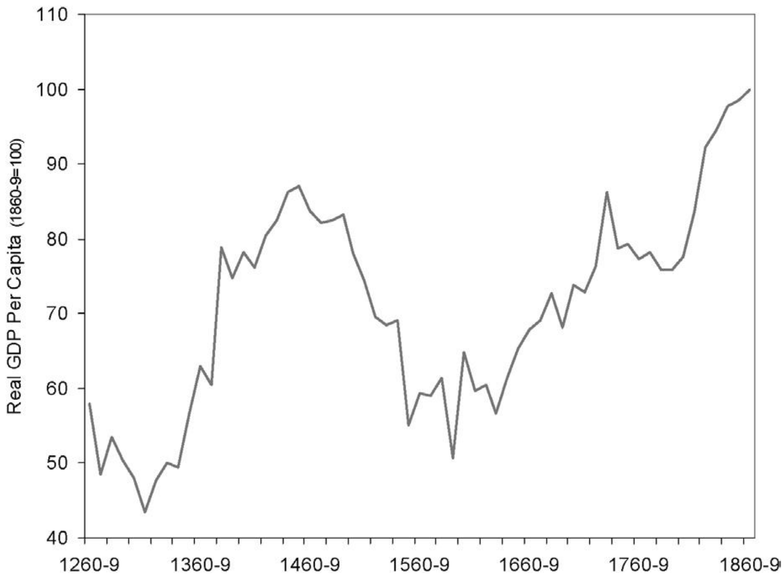

Figure 1.5: Real GDP per capita: England, 1250-1750. Sources: Clark (2001).

capita are counterbalanced by variation in population. This aspect explains the negative correlation between income and population that characterizes this period of time.

Such negative correlation is observed also in cases of exogenous shocks to population that, causing larger income per capita and wages which, in turn, have a feedback on fertility implying larger population. This is what we

Figure 1.6: Population and real wages: England, 1250-1750. Sources: Clark (2001, 2002).

observe in England with the catastrophic decline in population due to the Black Death (1348-1349) in which England population is decimated from 6 million to 3 million, implying a huge increase in real wages that can cause, after a century, an increase of the population almost at the pre-plague level.[2].
The natural consequence of the Malthusian mechanism is that areas of the globe that are more prosperous are characterized by larger population.[3]
The elements of the Malthusian theory can be described by a formal, although simple, model, which is developed in subsection 1.2.2.

1.2.2 A Simple Model

Based on Ashraf and Galor (2011), in the following it is reported a simple model that incorporates the features of the Malthusian theory. The model that will be developed in Chapter 2 will stand on this Malthusian structure,

[2]Reliable population data is not available for the period 1405-1525 and figure 1.6 is depicted under the assumption maintained by Clark (2001) that population was rather stable over this period of time.

[3]As reported in Galor (2011), Adam Smith in 1776 writes: "The most decisive mark of the prosperity of any country [was] the increase in the number of its inhabitants".

therefore illustrating the insights of the simple Malthusian model can give a better understanding of the more involved model that is the main contribution of this work.
Consider an overlapping generation model, time evolves discretely. One single homogeneous good is produced using land and labor. Land is fixed over time, whereas labor is determined by parents' decision regarding the number of children that they want to raise in the previous period.

Production

One homogeneous good in period t, Y_t, is produced according to a constant returns to scale technology,

$$Y_t = (AX)^{\alpha} L_t^{1-\alpha},\ \alpha \in (0,1) \tag{1.1}$$

where L_t is the amount of labor employed at time t, whereas X is the amount of land employed in the production process. A measures the technological level and, for the sake of simplicity, is assumed to be constant over time.[4] The level of technology can be seen as the level of knowledge in the economy, that can improve land used (think about new discoveries of lands suitable for agriculture or new technologies that increase the productivity of land, such as the plough). Therefore the term (AX) can be seen as effective land resources employed in production.
Output per worker in period t, $y_t \equiv Y_t/L_t$, is given by

$$y_t = (AX/L_t)^{\alpha}. \tag{1.2}$$

Preferences and Budget Constraints

In each period t there are two generations: parents and children. Children are passive economic agents, they consume parental resources and in the second period of live they will become parents. Parents work and earn wage that equals y_t, they devote their resources to raise children and to consume.

[4] In the model developed in Chapter 2 this assumption will be relaxed.

Individual utility is determined by the amount of the good they consume and the number of surviving children they raise:

$$u^t = (1-\gamma)ln(c_t) + \gamma ln(n_t) \tag{1.3}$$

where c_t is consumption and n_t is the number of surviving children of an individual of generation t.
Parental resources at time t are allocated to consumption, c_t, and to raise children, τn_t, where τ is the cost of raising children. Therefore the budget constraint is given by

$$c_t + \tau n_t \leq y_t. \tag{1.4}$$

Optimization

Parents allocate their resources optimally in order to maximize their utility given by (1.3) subject to the budget constraint, which is given by (1.4). The result of the optimization is given by,

$$c_t = (1-\gamma)y_t \tag{1.5}$$

$$n_t = \gamma y_t/\tau. \tag{1.6}$$

In the latter we can already see the characteristic of the Malthusian theory. An increase in income per capita is associated with a larger number of surviving children and, therefore, larger population growth. In other words, in the Malthusian theory, the correlation between population growth and income per capita is positive.

Population Dynamics

The size of the working population is determined by the following difference equation

$$L_{t+1} = n_t L_t \tag{1.7}$$

where L_t is the size of working population at time t and $L_0 > 0$. Substituting into the latter equation (1.6) combined with (1.2) we get the first order difference equation that governs the evolution of population dynamics

$$L_{t+1} = (\gamma/\tau)(AX)^{\alpha} L_t^{1-\alpha} \equiv \phi(L_t; A), \tag{1.8}$$

where the $\phi(\cdot)$ function is increasing and concave in L_t and satisfies the Inada conditions.
From the latter it is possible to determine the unique stable steady state level of the adult population, $\bar{L}$, such that

$$\bar{L} = (\gamma/\tau)^{1/\alpha}(AX) \tag{1.9}$$

and population density, $\bar{P}_d$, given by

$$\bar{P}_d \equiv \bar{L}/X = (\gamma/\tau)^{1/\alpha} A. \tag{1.10}$$

Note that technology has a first order positive effect on population and labor force, namely

$$\frac{\partial \bar{L}}{\partial A} > 0 \text{ and } \frac{\bar{P}_d}{\partial A} > 0 \tag{1.11}$$

In other words, positive technology shocks increase the *steady state level* of population.

The Evolution of Income per Worker

From equations (1.2) and (1.7), income per worker in period $t+1$ is given by

$$y_{t+1} = \Big[(AX)/L_{t+1}\Big]^{\alpha} = \Big[(AX)/n_t L_t\Big]^{\alpha} = y_t/n_t^{\alpha} \tag{1.12}$$

Substituting (1.6) into (1.12), then we obtain the equation for the evolution of income per worker in the Malthusian economy, which is given by

$$y_{t+1} = (\tau/\gamma)^{\alpha} y_t^{1-\alpha} \equiv \psi(y_t) \tag{1.13}$$

where the $\psi(\cdot)$ function is increasing and concave and satisfies Inada conditions.
Noting that $y_0 > 0$, there exists a unique non stable steady state of the above

first order difference equation, $\bar{y}$, which is given by

$$\bar{y} = (\tau/\gamma). \tag{1.14}$$

Note that in this case technology has an effect on the level of output per worker at time t, y_t, but it does not have any effect on the steady-state level of output per worker, $bary$. Namely,

$$\frac{\partial y_t}{\partial A} > 0 \text{ and } \frac{\bar{y}}{\partial A} = 0. \tag{1.15}$$

Predictions of the Model

The Malthusian theory is formally represented in the model and its main predictions are represented in (1.11) and (1.15). Namely, a positive shock to technology, A, would temporarily affect income per worker but it would not have any effect on the steady state level of income per worker. The reason is that, the temporary effects on income per worker would be compensated by an increase in the number of surviving children that parents can afford, increasing working population in the following periods and, in turn, implying convergence of income per worker to its previous steady state level. This has an important empirical implication: during early stages of development, economies that are more developed from a technological point of view (A), or from the point of view of endowments of resources (that can be represented in the model by X) are characterized by larger population. In other words, before the demographic transition, the correlation between population growth and economic development is positive. On the opposite, as widely analyzed in the growth literature since Solow (1956), it is observed that in late stages of development (i.e. after the demographic transition), economies that are more developed are characterized by lower population growth with respect to developing countries. Therefore the correlation between economic development and population growth becomes negative in later stages of development. The demographic transition plays a key role in explaining this fundamental structural change and has been extensively investigated in the literature. A subset of this literature described in the next section.

1.3 The Demographic Transition

After centuries of Malthusian stagnation, at the beginning of 1800 some regions of the world start experiencing an increase in living standards that has never been observed before. Triggered by structural changes in production techniques associated with great improvement in technology, we observe a huge increase in income per capita at world level, which is depicted in figure 1.7.

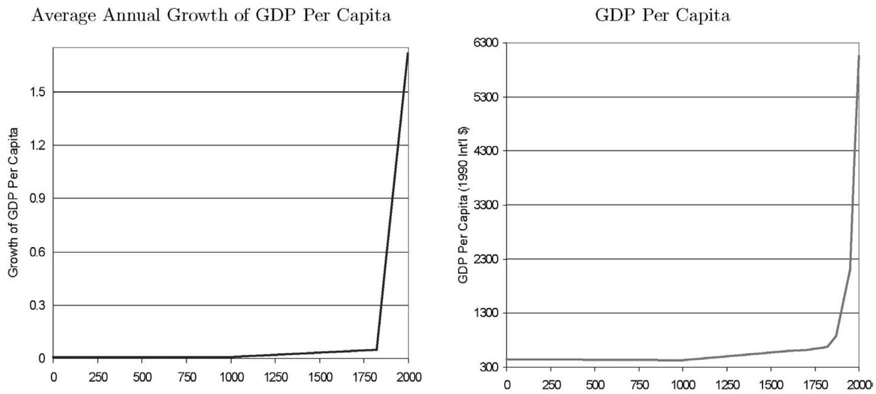

Figure 1.7: The evolution of the world income per capita over the years 1-2000. Source: Maddison (2003)

As depicted in figure 1.8, there are regions of the world that experienced this transition much in advance with respect to others. For instance Western Europe and Western Offshoots (USA, Canada, New Zealand and Australia) experienced a spike in growth of income per capita that has been experienced by part of Asian societies much more recently and has not been experienced yet by some African countries.

Understanding the difference in the timing of the transition is a fundamental challenge for scholars of any kind with the aim of understanding differences in income per capita across countries today. Figure 1.9 illustrates that after the take-off we observe some source of divergence, which can be partly explained by differences in the timing of the transition.

Figure 1.10 reinforces the relevance of understanding differences in the timing of the transition. In particular it illustrates that countries that experience

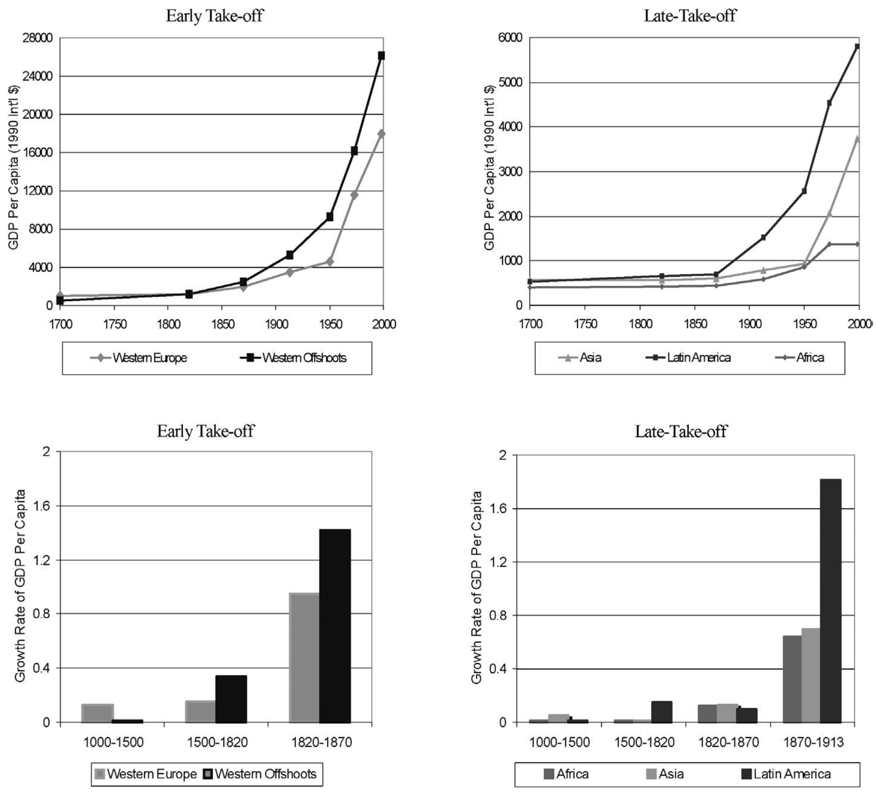

Figure 1.8: Differences in Timing and Magnitude of the Take-off from Malthusian Stagnation to Growth. Source : Galor (2005).

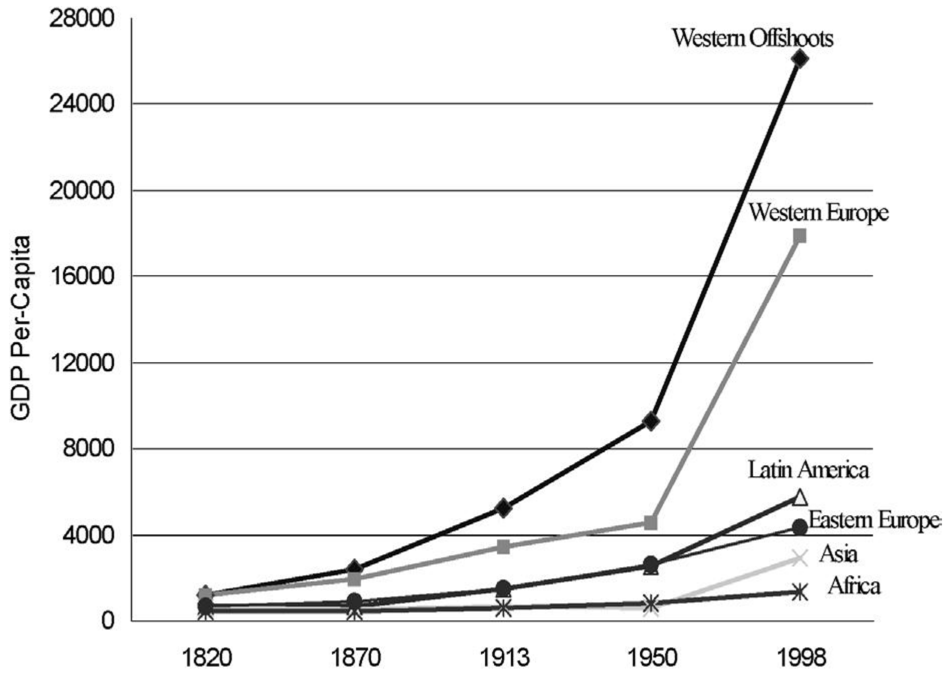

Figure 1.9: Divergence across regions: 1820-2000. Source: (Maddison, 2001, 2003).

the transition in advance are characterized by higher income today. Therefore, in order to answer fundamental questions of comparative development, it is necessary to understand what is the role played by initial conditions in explaining the transition from Malthusian stagnation and the differences in living standards that we observe today.

Naturally there is great interest in answering such fundamental questions, and clearly many theories that explain the demographic transition have been developed. in the following the emphasis will be on the branch of the literature that focuses on human capital and its interaction with technology, which is the segment of the literature on which the theoretical contribution developed in Chapter 2 stands on.

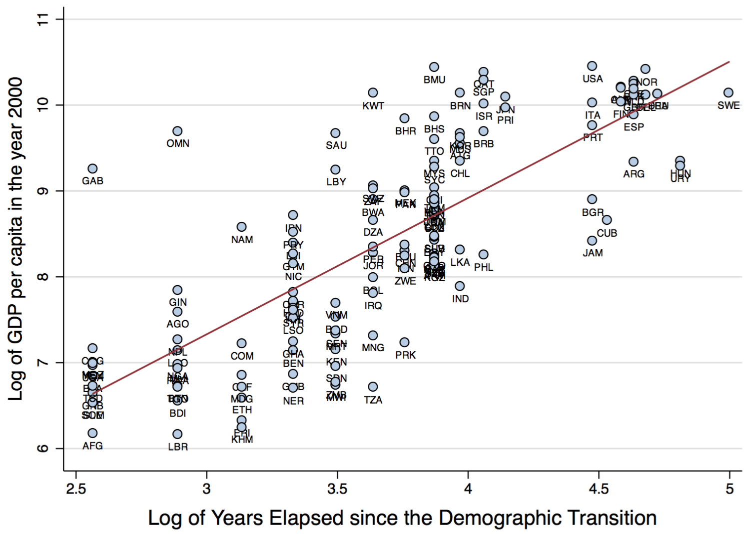

Figure 1.10: Differences in Transition Timing and Economic Performance Today. Data Source: Reher (2004)

1.3.1 The Rise in Human Capital Demand

Galor and Weil (1999, 2000) and Galor and Moav (2002) argue that the acceleration in the rate of technological progress during the second phase of industrialization induced an increase in the demand for human capital, implying larger investment at family level toward the quality (i.e. human capital) of the children rather than their quantity. This mechanism, standing on the quantity-quality trade-off by Becker et al. (1960), implies that the increase in the rate of technological progress is associated with a reduction in fertility. Thus the correlation between GDP and population becomes negative with the onset of the demographic transition.

The rise in human capital demand entails an increase in human capital supplied in the economy, that, in turn, has a feedback on knowledge and technology advancements implying sustained growth that we observe after the demographic transition. This is one of the fundamental mechanisms in the literature, which is complementary to others related to institutions and health, which are developed in the following.

1.3.2 The Role of Institutions

The Decline in Child Labor

Institutions play a key role in explaining differences in the timing of the take-off. One of the mechanisms that affected the timing of the transition is related to the reduction in child labor imposed by the adoption of child labor regulation in those countries that experienced the demographic transition in advance.

Doepke and Zilibotti (2005) analyze this mechanism from a theoretical point of view, supporting their theory with the observed pattern in England. In their paper, individuals might not have incentives to ban child labor when they have numerous families and child labor is a significant source of household income. However, skilled biased technical change reduces the return from unskilled labor (and thus from child labor) and then increses the return from human capital of the child. The implication is that families become less numerous and the incentives from preventing child labor regulation shrink, implying the adoption of institutions that ban child labor.

1.3.3 Technology and Health

Another channel through which an increase in the rate of technology growth can affect the demographic transition is through improvement in living standards given by health. In particular, the raise in life expectancy can be a channel through which the return from human capital of children. In particular Hazan (2009) finds sufficient conditions from a theoretical point of view such that the Ben-Porath (1967) theory is satisfied. That is longevity, increasing length of working life, increases the return to human capital of the children, reduces fertility as a consequence of the quantity-quality trade-off, ultimately triggering the demographic transition.

From an empirical point of view the health channel has been established by Bleakley (2007). Namely, he argues that the eradication of a particular disease from the South America had a positive effect on the schooling rate and human capital accumulation, with respect to those areas where the disease was still operating. This is additional evidence that technology is the fundamental engine of growth that can operate also via health. Therefore, in order to understand what are the determinants of he demographic transition that ultimately affected the cross-country disparity in income per capita that we observe today, it is crucial to identify the fundamental determinants of technology growth.

1.4 The Role of Technology Adoption in the Growth Literature

1.4.1 Selected Theories of Technology Adoption

The theory related to adoption of technology was generalized by the prominent paper by Nelson and Phelps (1966). Their theory of the phenomenon is based on the assumption that, while growth of the technology frontier reflects the rate at which new discoveries are made (innovation), growth of total factor productivity depends on the implementation of these discoveries (adoption), and varies positively with the distance between the technology frontier and the level of current productivity. Applied to the diffusion of technology between countries, with the country leading in total factor productivity representing the technology frontier, this is a formalization of the catch-up hypothesis that was originally proposed by Gerschenkron et al. (1962). There is another component of Nelson-Phelps hypothesis, which suggested that the rate at which the gap between the technology frontier and the current level of productivity is closed (the catch-up rate) depends on the level of human capital. This was a break with respect to the standard human capital theories, because according to the Nelson-Phelps theory human capital should not be considered as a factor of production *per se*, but it should only affect the level of adoption of existing technology. Benhabib and Spiegel (2005) generalize the model by Nelson and Phelps allowing for a more flexible specification and test it empirically as we will see in detail in the empirical section of this project.

Nelson and Phelps (1966) breaking path paper stimulated the literature about the potential friction to adoption of technology and the consequences on how technology diffuses. An important paper in this literature is by Basu and Weil (1998), in which the authors focus their analysis on the concept of *appropriate technology.* That is, technological advances can benefit certain types of technologies and not others. Namely, each technological advancement can benefit only those technologies that have the same capital labor ratio. As an example, consider an advancement in transportation technology in Japan, which might be an improvement in the latest maglev train. Even if such an advancement is free to move across economies, it can have very little effect on transportation technology in Bangladesh, which relies in large part on bicycles and bullock carts.[5] In other words they model a friction which acts as a barrier to technological adoption, the friction is due to the fact that countries take time to achieve a level of development that can take advantage of the progress being made by the technology leaders. The friction is defined as “appropriateness”, that is technology are appropriate for a specific level of capital intensity. Innovation is modeled as a process that expands the production possibilities frontier for a given capital-labor ratio; they justify this approach based on they idea that, in such way, it is possible to model learning by doing. Therefore, technology transfer, in the model, are not immediate because countries take time to achieve a level of development that can take advantage of the progress being made by the leaders. This friction justifies the persistence in the disparity of adoption of technologies that we observe across countries today and that will be documented in the next sections.

However there are other mechanisms that affect technology adoption that have been analyzed in the literature, such as mechanisms that are based on institutions. For instance certain countries might follow strategies based on imitation rather than based on innovation. This is the argument in Acemoglu et al. (2006), namely they argue that relatively backward economies can grow rapidly by investing in, and adopting, already existing technologies, or by pursuing what we call an *investment-based growth strategy.* This is their

[5]This example is extracted by Basu and Weil (1998).

explanation of the experience of a number of European countries during the nineteenth century discussed by Gerschenkron et al. (1962). At the other extreme, there is the process of *innovation-based growth*, where the selection of successful managers and firms, as well as a variety of activities aimed to innovate, are more important. The latter type of strategy would be undertaken by countries at the frontier of technology, while the countries that are lagging behind in terms of technological advancement would instead follow the investment- based strategy. In Acemoglu et al. (2006) paper, the difference between these strategies is determined by the distance to the technological frontier toward institutions, that is countries would develop different institutions that can promote one of the other strategy in turn affecting their level pf technological, and economic, development

If countries that are lagging behind in terms of development could adopt immediately the innovations produced by technological leaders, then there would not be differences in technologies across countries. In other words if such differences exist it means that there are frictions to technology adoptions. Such frictions can be related to institutions and talent allocation as in Acemoglu et al. (2006), but there are many other source of frictions that can cause the time lag in adoption of technology. However, an empirical approach in the analysis of technology adoption phenomena is essential to understand the magnitude of the disparities that we observe across countries and eventually the possible frictions by which they are generated.

1.4.2 Selected Empirical Works on Technology Adoption

Several papers modeled the process of technology diffusion with important cross-country implications. In particular, as discussed by Benhabib and Spiegel (2005), we can summarize two main approaches: the exponential diffusion process and the logistic model. There is fundamental difference in the empirical predictions of the two processes: the first implies that countries that are less developed can catch-up the technology leaders, while the logistic model predicts that the gap between followers and leaders can keep growing.

It is very interesting to notice that there functional forms are very similar. Based on Nelson and Phelps (1966), the exponential process of diffusion of technology in country i can be modeled as

$$\frac{\dot{A}_i(t)}{A_i(t)} = g(H_i(t)) + c(H_i(t))\Big(\frac{A_m(t)}{A_i(t)} - 1\Big) \tag{1.16}$$

where $A_i(t)$ is the total factor productivity, $g(H_i(t))$ is the component of total factor productivity growth that depends on the level of education $H_i(t)$ in country i and the term $c(H_i(t))\Big(\frac{\dot{A}_m(t)}{A_i(t)} - 1\Big)$ represents the rate of technology diffusion from the leader country m to country i.
The logistic specification can be represented with a slight difference with respect to the (1.16), that is

$$\frac{\dot{A}_i(t)}{A_i(t)} = g(H_i(t)) + c(H_i(t))\Big(\frac{A_i(t)}{A_m(t)}\Big)\Big(\frac{A_m(t)}{A_i(t)} - 1\Big) \tag{1.17}$$

Therefore both in equation (1.16) and (1.17), education has a direct effect on total factor productivity growth via the $g_i(\cdot)$ function which basically represents the effect of innovation. The difference is in the adoption component of the two processes. In both specifications education affects also the rate at which the gap between country i and the leader m is closed. However in equation (1.17) there is an extra term $\Big(\frac{A_i(t)}{A_m(t)}\Big)$. This term acts to dampen the rate of diffusion as the distance to the leader increases, reflecting perhaps the difficulty of adopting distant technologies. This is consistent with the theoretical literature seen above in which the frontier of technology might not be immediately "appropriate" (Basu and Weil, 1998), and is consistent with the empirical evidence of convergence clubs, as documented by Durlauf and Johnson (1995). Catch-up, according to the logistic specification, might be slower when the leader is either too distant or too close, and is fastest at intermediate distances.
Using a sample of 85 countries over the period 1960-1995, Benhabib and Spiegel (2005) display results that are in favor of a logistic specification. They derive a point estimate from their estimation results for the minimum initial

human capital level necessary to exhibit catch-up in total factor productivity relative to the leader nation, which is the United States in their sample. The point estimate in their favored specification indicates that an average of 1.78 years of schooling was required in 1960 to achieve convergence in total factor productivity growth with the United States. Using this approach, they identify 27 countries in the sample that their point estimates predict to exhibit slower total factor productivity growth than the United States. Their data show that over a period of 35 years, 22 of these 27 countries did indeed fall farther behind the United States in total factor productivity, while the remaining nations exhibited positive catch-up in total factor productivity.
They repeat the exercise to identify the set of countries that are still falling behind in total factor productivity growth in 1995. Because the United States had higher education levels in 1995, they estimates a higher threshold level for total factor productivity growth convergence with the United States. The point estimate is that 1.95 average years of schooling in the population over the age of 25 was necessary to faster total factor productivity growth at the level of the leader nations. They identify four countries that, at the end of the 35 years period considered, are still below the threshold in 1995: Mali, Mozambique, Nepal, and Niger. Therefore, their results indicate that, with the exception of these four countries, most of the world is not in a permanent development trap, at least in terms of total factor productivity growth.
While Benhabib and Spiegel (2005) analyze the process of technology adoption in a 35 years window of time, economists and economic historians have also long debated the importance of the long -run approach. More specifically, the idea is that past technology adoption affects subsequent technological development and especially what triggered the Industrial Revolution in Europe. While the scholars disagree on an initial technology advantage as a sufficient cause of the Industrial Revolution, their description of technology history reaches a consensus on many mechanisms that cause past technology to have an effect on future technology. The long run persistence of technology adoption is empirically documented in Comin et al. (2010). The authors assemble a dataset on technology adoption in 1000 BC, 0 AD, and 1500 AD for the predecessors of today's nation states in order to test the hypothesis

of technology persistence. In other words, the aim is to test whether it is possible that economies that were technologically developed long ago, have still today an advantage in technology adoption that ultimately affects the disparity in cross country economic development that we observe today. The main finding of the paper consists in documenting a strong and statistically significant association between the level of technology adoption in 1500 AD with per capita income and technology adoption today, where adoption is measured as the number of technologies available at a specific point in time. They also find robust and significant technological persistence from 1000 BC to 0 AD, and from 0 AD to 1500 AD.

Naturally, the question that arises is why we have such persistence and what are the possible sources. The authors documents the channels that have been illustrated in the economic history literature. They document 9 possibly related mechanisms. Fist, there are strong *complementarities* between existing technologies and new technologies (think about cement masonry and roads or aqueducts under Romans). If this complementarity is present, then the cost of adopting the new technology is lower, in turn implying that the number of technologies at a specific point in time is positively related to the number of technologies in the future. Second, *recombination* of old technologies to make new technologies is another possible channel which, again, implies persistence in the number of technologies over time. Third, the *feedback from technology to science*. That is, when a technique "works", this gives new evidence to scientists to test theories about why it works; thus science implies more innovations that, in turn will imply more adoption. Feedback from technology to *lower access costs for knowledge*. *Spillover* of technology between sectors. Some technologies are affected by *economies of scale*, implying positive correlation between past and future technology adoption. The introduction of new *General Purpose Technologies* can clearly increase the possibility to adopt new innovations in the economy (for instance the invention of Gutenberg printing process, information technology and so on). Feedback from technology to the *improvement of lab equipment*. Finally *learning by doing* and tacit knowledge. All these 9 mechanisms can explain the persistence of disparities in adoption of technology that we still observe today.

Naturally, the dataset built by Comin et al. (2010) is yet a remarkable contribution. It is constructed based on technologies available at cross-country level in the three period considered: 1000 BC, 0 AD, and the pre-colonial period 1500 AD. Technology adoption is measured on the extensive margin. That is, they document whether a technology is present or not in an economy and in a specific point in time, however they do not document how intensively a particular technology is used (the intensive margin). Economies are defined using modern day nation states, which creates a methodological issue. In particular, in some countries multiple cultures with heterogeneous level of development were present. The authors then decide to assign the country to the culture with the highest prevalence of population. I think this is a big limitation of their approach, because the largest cultures are also those who are more advanced technologically, while the number of less prevalent cultures as well as their effect on technology is not taken into account. In other words they do not take into account diversity, either genetic (Ashraf and Galor, 2013) or ethnic (Michalopoulos, 2012). Another important limitation is that the datset is extracted by the "Atlas of Cultural Evolution" (Peregrine, 2003) which does not tale into account crucial technologies like the plough, mathematics, astronomy, or medicine. Moreover, there is an important issue, that applies generally to the literature of technology adoption and more importantly in a long run perspective, which concerns technologies that depend on local circumstances. A famous example is that the Aztecs used wheels in toys for children but not in production activities, probably because of geographical conditions that did not allow them to use such technology in the agricultural sector, or maybe due to the complementarity to other innovations that were not present, such as roads and other infrastructures.

Another interesting paper that documents the disparity in technology adoption that we observe across the globe is the work by Comin and Hobijn (2010). In particular, they find that the adoption lags across economies are very large. In particular, the average adoption lag across countries and several technologies is 45 years. Moreover they find that there is substantial variation in these lags, both across countries and across technologies. The overall standard deviation is 39 years. In their analysis they find that 53 percent of such variation is explained across technologies and 18 percent is explained by cross-country variation, and 11 percent by the covariance between the two. They also find that over the last two centuries there is an acceleration in technology adoption, so later technologies have been adopted more rapidly than older ones.

They also analyze the phenomenon from a theoretical point of view. In the model, they assume that there is a fixed cost of adopting a new production method and the size of such cost determines the length of time between the invention and the eventual adoption of a production method: the adoption lag. Once the production method has been introduced in an economy, its productivity determines how many units of the good that is associated with it are produced. Therefore, in their model they have both extensive and intensive margin of technology. Where, by extensive margin is meant the variety of available production methods, while the intensive margin is the amount of goods produced using a specific production method.

Their model is crucially based on the assumption that cost of adoption takes the form of a fixed cost of production. They do not provide evidence for this assumption and they don't consider any source of firms' heterogeneity. The following section hypothesizes what are the potential implication of considering heterogeneity in firm size under the assumption of fixed cost of adoption.

1.5 Conclusions

This Chapter focused on selected literature on which the theoretical contribution developed in Chapter 2 stands on. In particular, the emphasis is, first, related to the main elements of the Malthusian theory that incorporates those latent structural changes that will bring the economy to the transition to a regime of sustained growth. Second, it is reviewed some of the main contributions on the theory of the demographic transition related to the fundamental mechanism of technology growth and rising in the demand for human capital that ultimately trigger knowledge accumulation and economic development.

Finally, the last part of the Chapter focuses on technology adoption in the growth literature both from a theoretical point of view and from an empirical point of view. These elements of the literature briefly presented in this chapter will be combined in the theoretical model presented in the next chapter. Namely the theoretical contribution developed in chapter 2 introduces technology adoption in the context of a long run growth model to explain differences in the timing of the demographic transition and the persistent differences in income that we observe across the globe today.

Chapter 2

The Theoretical Contribution

2.1 Introduction

Since the days of ancient Greece and before, philosophers, mathematicians and scholars of any sort have been the main source of technological innovations, involving creation and accumulation of knowledge over time. Despite the simplicity of great ideas, at all point in time significant inventions required refined skills and superior knowledge to be created. On the other hand, when innovations where not understood by masses they were not adopted in the production process and thus languished without having the possibility to contribute to accumulation of knowledge and output formation. While the upper tail of the human capital distribution, being composed by highly educated individuals more prone to innovate, implied a higher rate of innovation, the presence of a wide portion of the population enough educated to adopt innovations in the production process is essential to enhance technological advancement, ultimately affecting the timing of the transition from the Malthusian stagnation to modern growth and the long run growth rate of output per capita.

The proposed theory suggests that innovation is a necessary but not sufficient condition for technological progress: only if innovations are understood and adopted in the production process can contribute to output formation and induce technological progress and growth, therefore the rate of adoption as well as its genesis, is a crucial mechanism in understanding the huge disparities in income today. Historically there are many examples in which innovation and adoption did not go hand in hand causing many ideas to languish. For instance, one of the precursors of the steam engine was the Italian scholar Giambttista Della Porta (1535? - 1615) who used steam power to pump water in 1606, while "the first commercially successful engine did not appear until around 1712.[1] It was invented by Thomas Newcomen and paved the way for the Industrial Revolution" (Stuart, 1829). However, his first application was to pump water as well (Rolt et al., 1963). This evidences that one of the most powerful innovations was not sufficient to entail a 'major technological breakthrough' (Galor and Tsiddon, 1997) in the South of Italy, and then only after a century in England there was enough knowledge among the population that gave the possibility to understand the potential of the Newcomen's pump, so that, once materially produced, it was improved by James Watt to create the final version of the steam engine. A further example is given by the 12th century European Inquisition, during which many scientists were branded as heretics and forced to recant their proposed innovations (including Galileo Galilei and the above mentioned Giambattista Della Porta) and, if Inquisition is considered as a consequence of lack of diffusion of human capital in the population, it results in an additional source of evidence that human capital distribution had an impact on technological development.

These historical examples support the hypothesis that innovations are not sufficient to entail technological progress, the absence of a wide portion of the population enough educated to understand innovations and adopt them in the production process can imply the abandon of ideas that potentially might have been major technological breakthroughs, delaying the transition from stagnation to growth. If an economy is characterized by highly polarized human capital, that is much of the mass is in the tails of the human capital distribution, knowledge is concentrated in a few highly educated individuals in opposition to a mass of uneducated population. As a consequence, despite the contribution of highly educated individuals to the rate of inno-

[1] Brown, Richard (1991).*Society and economy in modern Britain*, 1700-1850 (Repr. ed.). London: Routledge. pp. 60.

vation, technological progress is harnessed by lack of education among the working population that implies a low rate of adoption of such innovation in the production process. In other terms, high polarization in the distribution of human capital involves high rate of innovation due to the presence of highly educated individuals but a low rate of adoption because of shortage of educated workers. Nevertheless if polarization in the distribution of human capital is significantly low, then a wide fraction of the labor force is enough educated to understand and adopt new innovations, entailing a high rate of adoption. However, the lack of highly educated individuals, who compose the fraction of the population inclined to move forward the frontier of innovation, implies an exiguous rate of innovation that, in turn, adversely affect technological progress in the pre-transition era and thus delaying the take-off to a regime of sustained economic growth.

Polarization in the distribution of human capital can have effects not only on the timing of the transition to a modern growth regime, but also it can affect the economic outcome in the long run. That is, consider that, due to the effect of polarization on the transition timing, low polarized economies took-off later, thus facing the Malthusian mechanism for longer. During such period, population growth offset any production improvement, entailing a more numerous population for those economies that were trapped in the Malthusian stagnation for longer with respect to economies that take-off earlier. Such advantage in terms of population gives an advantage in terms of human capital resources once the transition occurs. A larger population means a large number of highly educated agents and this could compensate for the adverse effect caused by the low *fraction* of such individuals by which low polarized economies are characterized, implying the possibility of higher rate of growth in the long run.

Polarization in the distribution of human capital implied two sources of trade-off. First, in early stages of development, polarization in the distribution of human capital entailed a trade off between innovation and adoption that affected timing of the transition from stagnation to growth. Second, polarization implied a trade off between the timing of the transition and the long run growth of output per capita.

This research highlights the role of adoption of new innovations in the production as a key element to explain the timing of the take-off and the long run growth rate in output per capita. More specifically, adoption rate had a twofold effect. First, only if an innovation is adopted in the production process it can have a direct effect on the production performance. Second, adoption has an indirect effect on future innovations because it acts as an alternative channel through which is possible to store, make accessible, transmit and then develop knowledge. In early stages of development, the level of knowledge was relatively low and the available techniques to store knowledge were primitive and imperfect relatively to modern era. Moreover, in a period in which a common scientific language has not been developed yet, the difficulties in accessing stored information were significant. For instance, consider that only in 1662 the first scientific journal, the *Royal Society of London*, was developed (Bekar and Lipsey, 2004) and thus innovations were stored and made accessible to the scientific community only in late stages of development. Before that time, innovations can be intergenerationally transmitted only when they were adopted, and thus transformed from theoretical projects in material "gadgets" (Ashton, 1955) that can be stored, transmitted and understood despite the lack of a common scientific language. This is a crucial channel because once innovations were adopted and transformed in "gadgets", they can be improved over generations (Bekar and Lipsey, 2004) based on a process of "tinkering" (Mokyr, 1990) that paved the way for a "basic mechanization experimentation based" (Musson, 1963). This indirect effect of adoption on technological advancement is primarily originated by workers that are skilled in tinkering and producing gadgets, namely they supply a combination of intellectual and manual labor, in other terms they compose the mass in the middle of the human capital distribution.

Therefore, an economy characterized by low polarization in the distribution of human capital will experience a high rate of adoption of new innovations because of the wide fraction of the labor force that can understand and employ new ideas improving the production process and consequently, through the creation of gadgets that materially represent such ideas, have a feedback effect on the accumulation of knowledge and technological progress. The interplay of this feedback effect on the accumulation of knowledge together with the direct effect on production implies that polarization in the distribution of human capital is a crucial element in the analysis of the transition from the Malthusian era to sustained growth and in understanding the disparity in the economic performance across the globe today.

The chapter is organized as follows. Next section present a model that formalizes how polarization in the distribution of human capital, via the mechanism of creation and adoption of new technology, affects the timing of the transition from Malthusian stagnation to growth and the long run implications for growth rate of output per capita. Last section concludes.

2.2 The Model

Consider an overlapping generation model that evolves over infinite discrete time. Every period t, a finite homogeneous good, Y_t, can be produced according to two alternatives regimes of production, defined as *old regime* and *new regime* of production. Factors of production are three sources of labor force reflecting three levels of human capital of workers[2]. The three sources of labor force are: manual labor, human capital intensive labor and a combination between the two.[3]

Manual labor, L_t, reflects the amount of labor supplied by those individuals that are characterized by the lowest level of human capital. Human capital-intensive labor, H_t, reflects the amount of labor supplied by highly educated labor force, that, being highly educated, is more prone to innovate and thus they are defined as *innovators*. A third source of labor force, which repre-

[2]In addition, I could consider Land as a constant factor of production, assuming the absence of property rights (Galor and Moav, 2002), the results would not change qualitatively. Considering capital as a factor of production, could complicate the model to the point of intractability.

[3]The aim is to represent the distribution of human capital in the population as a discrete distribution with three mass accumulation points that correspond to three levels of education in the population. In other terms this approach takes into account the fact that investment in human capital it's not continuous: people can decide to go to the college or not, to learn a particular job or not, but taking the 75% of a college degree or of a skill necessary to be employed in a particular sector is not rewarded on the market and then can be excluded from the analysis.

sents the middle of the human capital distribution, is given by a combination between manual labor and intellectual labor, M_t, which is supplied by individuals that are enough educated to understand and adopt new innovations, thus they are defined as *adopters*.

2.2.1 Production

Production can take place according to two alternative regimes, defined as *old regime* and *new regime*.

The old regime The *old regime* of production is such that only manual labor and highly educated labor are employed[4].

$$Y_t^o = A_t^o H_t^\alpha L_t^{1-\alpha} = A_t^o L_t h_t^\alpha \tag{2.1}$$

where h_t is the proportion of highly educated over manual labor force. That is

$$h_t \equiv \frac{H_t}{L_t} \tag{2.2}$$

That is, in early stages of technological development, new inventions, that were principally produced by highly skilled labor force, can be directly adopted by manual labor force. This represents the idea that, when the level of knowledge is low, new innovations are not too complex to adopt and thus the understanding of the practical functioning is affordable even by the fraction of the low skilled population. Namely, even though the theoretical foundation could be unknown to the masses, the simplicity of early technology is reflected in simplicity of its adoption, implying that there is no necessity of specific forms of education in order to have the adoption of such innovations in output production. For instance, although high chemical skills were needed in order to invent *ley farming*[5], that was introduced in one of the most advanced forms

[4] Alternatively, innovators may be considered as monopolist of an intermediate sector of innovations (following Aghion and Howitt (1992)). Despite the complexity of such alternative approach, the result would be identical.

[5] Ley farming is an agricultural system where the field is alternately seeded for grain and left fallow. During the fallow period the soil is filled with roots of grasses and other plants.

at the beginning of the seventeenth century (Stapledon et al., 1948), can be easily understood and adopted in the agricultural sector: the farmer needs simply to know what are the grasses to fill the soil during the fallow period in order to increase production of the cultivation. In other words, in early stages of technological development, production of innovations required a early form of human capital while the minimum level of knowledge needed in order to adopt new inventions is low. However, despite the fact that manual labor is sufficiently skilled to adopt innovations, it is not enough educated to fully understand the theoretical principle of such innovations, implying a scarce tendency to improve them over time. Going back to the above example, the farmer can adopt ley farming without a comprehensive view of the chemical reasons of the improvement in land productivity, however this entails a slow inter generational improvement of such technique.[6] According to such view, production in early stages of development is represented by the contribute of manual labor force, corroborated by the presence of educated individuals that directly contributed to production through the creation of innovations. Moreover, as analyzed in the following, the population in the upper tail of the distribution also contributes indirectly to economic development through accumulation of knowledge.

The new regime Production can take place according to the *new regime* which, in addition to manual labor and innovators' labor, employs adopters' labor force.

$$Y_t^n = A_t^n H_t^\beta M_t^\phi L_t^{1-\beta-\phi} = A_t^n L_t h_t^\beta m_t^\phi \tag{2.3}$$

where m_t is the proportion of adopters over manual labor force,

$$m_t \equiv \frac{M_t}{L_t} \; ; \; h_t \equiv \frac{H_t}{L_t} \tag{2.4}$$

The intuition is the following. Considering an environment in which tech-

[6]This additional mechanism of accumulation of knowledge will be analyzed in detail in the following.

nological frontier is sufficiently developed, then innovations can be principally adopted by individuals that are enough educated to understand and then employ them in the production process. This is the case of most advanced technical innovations, for which some form of education is needed in order to acquire those technical skills that are necessary to thoroughly understand and adopt them in the production process. For example, during the process of development, the urbanization phenomenon was associated to the creation, or consolidation, of certain classes of workers broadly defined as artisans. Such workers were educated and specialized in order to compute specific tasks and employ gadgets, that needed specific training to be performed. Such kind of labor force, supplying a combination of manual and intellectual labor, composed the mass in the middle of the human capital distribution.

Therefore, an environment sufficiently advanced from a technological point of view is such that productivity of manual labor force in food production is sufficiently high to make the subsistence constraint no longer binding, entailing the possibility for a portion of the population to invest in learning technical skills that are necessary to adopt new technologies. Such portion of the labor force will compose the mass in the middle of the human capital distribution.

Factor prices Markets are perfectly competitive, the inverse demands for factors of production depend on the regime employed.

The inverse demand for highly skilled labor, given 2.1 and 2.3,

$$w_t^h = \begin{cases} \alpha A_t^o h_t^{\alpha-1} & \text{if } Y_t^o > 0 \\ \beta A_t^n h_t^{\beta-1} m_t^{\phi} & \text{if } Y_t^n > 0 \end{cases} \tag{2.5}$$

where w_t^h is the wage of innovators.

The inverse demand for manual labor, given 2.1 and 2.3, is

$$w_t^l = \begin{cases} (1-\alpha) A_t^o h_t^{\alpha} & \text{if } Y_t^o > 0 \\ (1-\beta-\phi) A_t^n h_t^{\beta} m_t^{\phi} & \text{if } Y_t^n > 0 \end{cases} \tag{2.6}$$

where w_t^l is the wage of unskilled labor.
The inverse demand for adopters' labor, given 2.3, is

$$w_t^m = \phi A_t^n h_t^\beta m_t^{\phi-1} \quad \text{if } Y_t^n > 0 \tag{2.7}$$

where w_t^m is the wage of adopters, that will be employed only in new regime Moreover, given 2.5 and 2.6, the innovators over manual labor wage ratio is

$$\frac{w_t^h}{w_t^l} = \begin{cases} \frac{\alpha}{1-\alpha}\frac{1}{h_t} & \equiv \omega\big(h_t^o\big) \quad \text{if } Y_t^o > 0 \\ \Big(\frac{\beta}{1-\beta-\phi}\Big)\frac{1}{h_t} & \equiv \omega\big(h_t^n\big) \quad \text{if } Y_t^n > 0 \end{cases} \tag{2.8}$$

Given 2.7 and 2.6 the adopters over manual labor wage ratio is given by

$$\frac{w_t^m}{w_t^l} = \Big(\frac{\phi}{1-\beta-\phi}\Big)\frac{1}{m_t} \equiv \omega^m\big(m_t\big) \text{ if } Y_t^n > 0 \tag{2.9}$$

From the properties of the production functions, it follows that wage ratios are characterized by the following properties:
$\omega^{'}(j_t) < 0$, $\lim_{j\to 0}\omega^j(j_t) \to \infty$, $\lim_{j\to\infty}\omega^j(j_t) \to 0$ with $j = m, h$ and $\forall j \in \big[0, \infty\big)$.

The individual choice Consider an economy in which individuals live for two periods of time: childhood and parenthood. During the first period of life they consume a fraction of parental endowment that consists in one unit of time. All decisions are made in the adult period of life. Parents are endowed with one unit of time as manual labor, l, adopters labor, m, or innovators labor, h, depending on the level of education they received during childhood. Such endowment is allocated between children rearing and consumption.

2.2.2 Preferences and budget constraints

Preferences are defined over parental consumption and the potential aggregate income of their children (Galor and Mountford, 2008). Parents i, where $i = l, m, h$, choose the number of children $n^{i,j}$ for each level of education j, with $j = l, m, h$, and parental utility from each child depends on the wage she gets on the market. In other terms, parents get their own utility according to the utility function

$$u_t^i = (1-\gamma)lnc_t^i + \gamma ln\big(w_{t+1}^l n_t^{i,l} + w_{t+1}^m n_t^{i,m} + w_{t+1}^h n_t^{i,h}\big) \tag{2.10}$$

where c_t^i is parental consumption at time t, $n_t^{i,j}$ is the number of children of type j reared by parent i at time t.[7]

The budget constraint is given by

$$c_t^i + w_t^i\big(n_t^{i,l}\tau^l + n_t^{i,m}\tau^m + n_t^{i,h}\tau^h\big) \leq w_t^i \tag{2.11}$$

Optimization

$$\{c_t^i, n_t^{i,l}, n_t^{i,m}, n_t^{i,h}\} = argmax\Big[(1-\gamma)lnc_t^i + \gamma ln\big(w_{t+1}^l n_t^{i,l} + w_{t+1}^m n_t^{i,m} + w_{t+1}^h n_t^{i,h}\big)\Big] \tag{2.12}$$

subject to

$$c_t^i + w_t^i\big(n_t^{i,l}\tau^l + n_t^{i,m}\tau^m + n_t^{i,h}\tau^h\big) \leq w_t^i \tag{2.13}$$

$$c_t^i \geq \tilde{c} \tag{2.14}$$

where $\tau^l < \tau^m < \tau^h$. In particular τ^j is the cost of having a child of type j with $j = l, m, h$, therefore the higher the level of human capital of the offspring the higher the cost of producing a child with that particular level of education[8].

The optimal level of consumption is given by,

[7]Notice that, since mortality is not explicitly modelled, n_t can be interpreted as the number of surviving children.

[8]Alternatively, one may argue that, despite the lower level of human capital intrinsic in artisans skills with respect to philosophers or mathematicians, the scarcity of certain skills, such as the carpenter or armourer ones, entails difficulties in acquiring them with the consequence of higher costs. However, most artisans skills were acquired through job training or, in more advanced stages of urbanization, through apprenticeship under the supervision of masters. Both these approaches of acquiring this source of human capital are characterized by a higher degree of economies of scale with respect to the acquisition of high level human capital, in turn implying a lower cost relative to other forms of human capital.

$$c_t^i = \begin{cases} \tilde{c} & \text{if } (1-\gamma)w_t^i < \tilde{c} \\ (1-\gamma) & \text{if } (1-\gamma)w_t^i \geq \tilde{c} \end{cases} \tag{2.15}$$

The amount of time invested in child rearing is given by,

$$n_t^{i,l}\tau^l + n_t^{i,m}\tau^m + n_t^{i,h}\tau^h = \begin{cases} \frac{w_t^i - \tilde{c}}{w_t^i} & \text{if } (1-\gamma)w_t^i < \tilde{c} \\ \gamma & \text{if } (1-\gamma)w_t^i \geq \tilde{c} \end{cases} \tag{2.16}$$

During the old regime of production,

$$\begin{array}{ll} n_t^{i,h} = 0 & \text{if } w_t^h/w_t^l < \tau^h/\tau^l \\ n_t^{i,h} > 0 \text{ and } n^{i,l} > 0 & \text{only if } w_t^h/w_t^l = \tau^h/\tau^l \\ n_t^{i,l} = 0 & \text{if } w_t^h/w_t^l > \tau^h/\tau^l \end{array} \tag{2.17}$$

which means that if the wages ratio equals the cost ratio all the types available in the old regime will exist.

Equivalently, during the new regime of production,

$$\begin{array}{rll} n_t^{i,h} = 0 & \textit{if } w_t^h/w_t^l < \tau^h/\tau^l & \textit{or } w_t^h/w_t^m < \tau^h/\tau^m \\ n_t^{i,m} = 0 & \textit{if } w_t^m/w_t^l < \tau^m/\tau^l & \textit{or } w_t^m/w_t^h < \tau^m/\tau^h \\ n_t^{i,l} = 0 & \textit{if } w_t^h/w_t^l < \tau^h/\tau^l & \textit{or } w_t^h/w_t^m < \tau^h/\tau^m \\ (n_t^{i,h}, n_t^{i,m}, n_t^{i,l}) \gg 0 & \textit{only if } w_t^h/w_t^l = \tau^h/\tau^l & \textit{and } w_t^m/w_t^l = \tau^m/\tau^l \end{array} \tag{2.18}$$

Lemma 1 *Consider the old regime of production. There exists a unique ratio of innovators to manual labor ratio, $(h^o)^*$ such that*

$$\frac{w_t^{o,h}}{w_t^{o,l}} = \omega((h^o)^*) = \frac{\tau^h}{\tau^l} \tag{2.19}$$

where,

$$\begin{array}{ll} n_t^{i,l} = 0 & \text{if } h_t < (h_t^o)^* \\ n_t^{i,h} = 0 & \text{if } h_t > (h_t^o)^* \end{array} \tag{2.20}$$

Proof. The uniqueness of $(h_t^o)^*$ follows from the properties of $\omega((h_t^o)^*)$. The remaining part is a corollary of 2.17. □

Hence, during the old regime, if $h_t < (h_t^o)^*$ the relative reward for having uneducated offspring is low with respect to the relative cost and thus there are no incentives to produce them, implying an increase in h_t. Whereas, if $h_t > (h_t^o)^*$ there are no incentives to raise high human capital offspring, implying a decrease in h_t up to the equilibrium proportion,$(h_t^o)^*$.

Corollary 1 *If the old regime of production is employed then $h_t = (h_t^o)^*$, that is,*

$$h_t = (h_t^o)^* \quad \text{if } Y_t^o > 0 \tag{2.21}$$

and therefore wages for innovators are

$$w_t^h = \alpha A_t^o \big[(h_t^o)^*\big]^{\alpha-1} \quad \text{if } Y_t^o > 0 \tag{2.22}$$

wages for manual labor are

$$w_t^l = (1-\alpha) A_t^o \big[(h_t^o)^*\big]^{\alpha} \quad \text{if } Y_t^o > 0 \tag{2.23}$$

and thus

$$(h_t^o)^* = \Big(\frac{\alpha}{1-\alpha}\Big)\frac{\tau^l}{\tau^h} \tag{2.24}$$

where the latter comes from 2.8, given Lemma 1 .

Importantly, notice from 2.24 that during the old regime the optimal proportion of innovators over manual labor force is constant over time, that is,

$$(h_t^o)^* = (h^o)^* \; \forall t \tag{2.25}$$

Lemma 2 *Consider the new regime. There exists a unique innovators to manual labor ratio, $(h^n)^*$, and a unique adopters to manual labor ratio, m^*, such that*

$$\frac{w_t^{n,h}}{w_t^{n,l}} = \omega((h_t^n)^*) = \frac{\tau^h}{\tau^l} \tag{2.26}$$

$$\frac{w_t^m}{w_t^{n,l}} = \omega(m_t^*) = \frac{\tau^m}{\tau^l} \tag{2.27}$$

$$\begin{array}{lll} n_t^{i,h} = 0 & \text{if } h_t > (h^n)^* & \text{or } m_t < m_t^* \\ n_t^{i,m} = 0 & \text{if } m_t > m_t^* & \text{or } h_t < (h^n)^* \\ n_t^{i,l} = 0 & \text{if } h_t < (h^n)^* & \text{or } m_t < m_t^* \end{array} \tag{2.28}$$

Proof. The uniqueness of $(h_t^n)^*$ and m_t^* follows from the properties of $\omega((h_t^n)^*)$ and $\omega(m_t^*)$ respectively. The remaining part is a corollary of (2.18). □

Hence, during the new regime, if $h_t > (h_t^n)^*$ there are no incentives to raise highly educated offspring, entailing a reduction in h_t. However, also in the case in which $m_t < m_t^*$ there are no incentives to raise neither highly educated children nor uneducated children because the relative reward of raising offspring with an intermediate level of education is higher, therefore resources will move in this direction, increasing m_t. In other words, if the proportion of one of the three source of labor force is lower than the optimal one, the potential relative wage of that child is higher than the relative cost, therefore parents will invest their resources in rearing offspring with that particular level of education, increasing their proportion with respect to the other two with a consequent reduction in the relative wage until (2.26) and (2.27) are satisfied.

Corollary 2 *If the new regime of production is employed then* $h_t = (h_t^n)^*$ *and* $m_t = m_t^*$*, that is,*

$$h_t = (h_t^n)^* \text{ and } m_t = m_t^* \quad \text{if } Y_t^n > 0 \tag{2.29}$$

and therefore wages for innovators are

$$w_t^h = \beta A_t^n \left[(h^n)^*\right]^{\beta-1} \left[m^*\right]^{\phi} \quad \text{if } Y_t^n > 0 \tag{2.30}$$

wages for adopters are

$$w_t^m = \phi A_t^n \left[(h^n)^*\right]^{\beta} \left[m^*\right]^{\phi-1} \quad \text{if } Y_t^n > 0 \tag{2.31}$$

wages for manual labor are

$$w_t^l = (1-\beta-\phi)A_t^n\big[(h^n)^*\big]^\beta\big[m^*\big]^\phi \quad \text{if } Y_t^n > 0 \tag{2.32}$$

and thus

$$(h_t^n)^* = \Big(\frac{\beta}{1-\beta-\phi}\Big)\frac{\tau^l}{\tau^h} \tag{2.33}$$

$$m_t^* = \Big(\frac{\phi}{1-\beta-\phi}\Big)\frac{\tau^l}{\tau^m} \tag{2.34}$$

where 2.33 and 2.34 come from 2.8 and 2.9 , given Lemma 2.2.2 . Importantly, notice from 2.33 and 2.34 that during the new regime the optimal proportions of innovators over manual labor force and adopters over manual labor force are constant over time, that is,

$$(h_t^n)^* = (h^n)^* \text{ and } m_t^* = m^*;\ \forall t \tag{2.35}$$

Furthermore from 2.19 and 2.26,

$$\omega((h^o)^*) = \omega((h^n)^*) \tag{2.36}$$

that implies

$$\Big(\frac{\alpha}{1-\alpha}\Big)(h^n)^* = \Big(\frac{\beta}{1-\beta-\phi}\Big)(h^o)^* \tag{2.37}$$

Notice that under reasonable parametrization[9] the prevalence of innovators with respect to manual labor force is higher in the new regime, that is,

$$(h^n)^* > (h^o)^* \tag{2.38}$$

[9]The condition on parameters is that $(1-\alpha) < (1-\beta-\phi)$, such condition ensures that production during the old regime is manual labor intensive.

2.2.3 Technological Progress

Suppose that, during the old regime, technology formation between time t and $t+1$, depends on the number of innovators in the economy at time t:

$$\frac{A^o_{t+1} - A^o_t}{A^o_t} = \Omega(H_t) \tag{2.39}$$

where A^o_0 is historically given and the innovation function $\Omega(H_t)$ is an increasing and concave function $\Omega' > 0$ and $\Omega'' < 0$ and $\Omega \in (0, \infty)$.
That is, technological progress during early stages of development depends on the population of innovators in the economy. Note that $\Omega(0) > 0$, that is, in the absence of innovators in the economy, manual labor can give a contribution to technological advancement.
During the new regime of production, in addition to the effect of innovators, the presence of a fraction of the labor force that can adopt, transmit and improve innovations is an additional source of technological accumulation

$$\frac{A^n_{t+1} - A^n_t}{A^n_t} = \Omega(H_t)\ \big(1 + \lambda(M_t/N_t)\big) \tag{2.40}$$

where A^n_0 is historically given and the adoption rate $\lambda' > 0$, $\lambda'' < 0$ and $\lambda \in (0, \infty)$ with $\lambda(0) > 0$[10]. The adoption rate λ depends on the fraction of adopters in the economy, $\frac{M_t}{N_t}$, the higher the fraction of adopters the closer the economy is to perfect adoption. Note that technological progress is faster in the new regime.
During the old regime, despite the fact that the new regime of production is not operative, knowledge advancement permits the *potential* productivity of the new regime to grow over time. That is, when the new regime is not efficient, adopters are not employed in the production process, and thus not rewarded on the market,however there is a latent technology advancement due to those workers that, throughout a process of tinkering and learning

[10] Consider that, if the fraction of adopters is null the adoption rate is assumed to be positive (i.e. $\lambda(0) > 0$). This assumption ensures that, during stages of development in which the degree of specialization is higher (i.e. the new regime), technological accumulation advances faster with respect to earlier stages of development.

by doing in laboratories rather then on the job, acquire those skills that are necessary to the process of adoption and make the new regime more efficient.

2.2.4 Viability of production regimes

The two regimes are available at each point in time, thus each agent chooses the preferred regime depending on the reward he can get. In other terms,

$$\text{Every agent } i \text{ chooses the regime } j \text{ if } w_t^{j,i} \geq w_t^{-j,i} \forall i = l, m, h; \forall t \tag{2.41}$$

Lemma 3 *At each point in time, only one regime of production is operative.*

Proof. It comes from condition 2.41 given Lemma 1 and Lemma 2.2.2 . □

Considering that each agent chooses the preferred regime depending on the reward he can get, Lemma 1 and 2.2.2 imply that, during the old and the new regime respectively, the proportions of factors of production in the economy are constant over time (see 2.24, 2.33 and 2.34), therefore factors of production are not free to adjust up to an equilibrium wage that permits the coexistence of the two regimes. Conversely, in each regime wages are given (but not constant: they depend on the level of technology A_t^j, *for* $j = old, new$) and at each point in time agents compare such wages determining the operative regime.

Lemma 4 *The new regime is economically viable if*[11]

$$w_{t+1}^{n,l} \geq w_{t+1}^{o,l} \tag{2.42}$$

where $w_{t+1}^{n,l}$ *is the wage that uneducated children at time* t *will get at time* $t+1$ *in regime* $j = old, new$

[11] It is assumed that, in the case in which wages for a specific source of labor force are equal in both regimes, the new regime is preferred. However, as will be clear in the following, such equality can persist only for one period of time.

Proof. It comes from condition 2.41 given 2.19, 2.26 and 2.27. □

Since the wage ratios are constant over time, if the new regime is economically viable for manual labor force it will be economically viable for all agents in the economy. Notice that during the old regime adopters are not rewarded, therefore parents will choose to invest in that particular level of education only when the new regime will be operative. That is, only when children that are educated at high level (innovators) or at low level (manual labor) will get a higher wage in that regime.
So that, the difference between the wage that manual labor can get in the new regime and the one available in the old regime represents a threshold rule for the escape from the old to the new regime of production.

2.2.5 The Time Path of the Economy

The productivity parameters are restricted so that the new regime is not economically viable in period 0, that is,

$$\frac{A_0^o}{A_0^n} > \frac{1-\beta-\phi}{1-\alpha}\,\frac{[(h^n)^*]^\beta[m^*]^\phi}{[(h^o)^*]^\alpha} \tag{2.43}$$

Lemma 5 It exists a time t^* such that the new regime is viable, that is,

$$\exists t^* | \forall t \geq t^*, w_t^{n,l} \geq w_t^{o,l} \tag{2.44}$$

Proof. It comes from lemma 4, 2.32 and 2.23 given 2.39 and 2.40 □

Since the productivity of the new regime grows faster[12] and given that the unique source of time variation of wages is due to total productivity growth, necessarily exists a point in time, t^*, in which wages of the new regime are equal or higher, implying the transition to the new regime. This also means

[12] Despite the fact that the new regime of production is not employed for a certain period of time, knowledge advancements implies an increase in the potential technology. Innovations stimulates productivity of the old regime of production, as well as the advancement of knowledge of those labour force that acquire adopters' skills through a learning by doing process although such skills are not rewarded on the market. Such Assumption is consistent with the literature, among others Galor and Mountford (2008).

that wherever an economy is optimally positioned in the h_t, m_t plane, see figure 2.1, sooner or later it will experience the transition from stagnation to growth.

Lemma 6 It exists a time t^c such that the Malthusian constraint is no longer binding, that is,

$$\exists t^c | \forall t \geq t^c, w_t^{o,l} \geq \tilde{c}/(1-\gamma) \tag{2.45}$$

Proof. It comes from equation 2.23 , given 2.39 □

Equilibrium wages increase of time due to technology advancement, therefore necessarily exists a point in time at which the subsistence constraint is no longer binding.
Consistently with the historical pattern, it is assumed that $t^* = t^c$. This assumption is equivalent to assume that the wage level w_{t^*} such that $w_{t^*} \equiv w_t^{n,l} = w_t^{o,l}$ is such that $w_{t^*} = \tilde{c}/(1-\gamma)$. Since the new regime of production implies that the level of knowledge is sufficiently advanced to entail the consolidation of a new source of labor force based on the understanding and adoption of innovation, it is plausible to consider that such improvement in labor force productivity implies a wage that is at least sufficient for the subsistence.

The Timing of the Transition Given the equilibrium quantities h^*; m^* and the threshold G_t, it possible to solve for the time at which an economy will experience the transition to the new regime, t^*. Where G_t is such that 2.42 is satisfied with equality, that is,

$$G_t = \left\{ G(h_t, m_t) | w_t^{n,l} - w_t^{o,l} = 0 \right\} \forall t \leq t^* \tag{2.46}$$

where $h_t \equiv h_t^n$.

Lemma 7 *The threshold G_t, before the transition, is a function of the proportions of factors of production in the new regime and parameters of the* *model, that is,*

$$G_t = w_t^{n,l} - w_t^{o,l} = G\big(h_t^n, m_t; A_t^o, A_t^n, \zeta\big) = 0 \ \forall t \leq t^* \tag{2.47}$$

where $\zeta = \zeta(\tau^l, \tau^h; \alpha, \beta, \phi)$

Proof. It comes from 2.23, 2.24 and 82 given Lemma 5 noting that only the old regime is operative. □

In other words, before the transition takes place, only the old regime is operative (see condition 2.43), entailing labor forces to be rewarded at the equilibrium wages. The threshold G_t, $\forall t < t^*$, represents the sets of points (h_t, m_t), where $h_t^n \equiv h_t$, such that the new regime is viable, therefore, it can be represented on a h_t, m_t plane (see figure2.1). The new regime is viable when the equilibrium proportions (h^*, m^*) satisfies the threshold rule G_t. Knowing that, from condition 2.42, in period 0 the new regime is not viable, the timing of the transition can be measured considering the time elapsed between period 0 and the period in which the optimal proportions of factors of production belong to the threshold, $t^*|(h^*, m^*) \in G_t^*$. Figure 2.1 depicts the movement of the threshold, G_t, until the transition is experienced (i.e. $t \leq t^*$). Therefore, the timing of the transition, t^*, is a function of the distance from the equilibrium point (h^*, m^*) and the curve $G_{t=0}$. More specifically, from the basic rule of physics, time is given by the ratio between distance and speed, thus the timing of the transition from stagnation to growth is given by,

$$t^* = \frac{d^*}{s_{t^*}} \tag{2.48}$$

where, d^* is the minimum distance between (h^*, m^*) and $G_{t=0}$, s_{t^*} is the speed of convergence to the new regime[13]. Therefore, noticing 2.24, 2.33 and 2.34 the time of the transition can be expressed as follows,

$$t^* = t(\tau^l, \tau^m, \tau^h; \xi) \tag{2.49}$$

where $\xi \equiv \xi(\alpha, \beta, \phi, A_0^o, A_0^n, \lambda(0))$. The costs of raising offspring, τ^j, with

[13]See Appendix for the specification of d^* and s_{t^*}

The Timing of the Transition

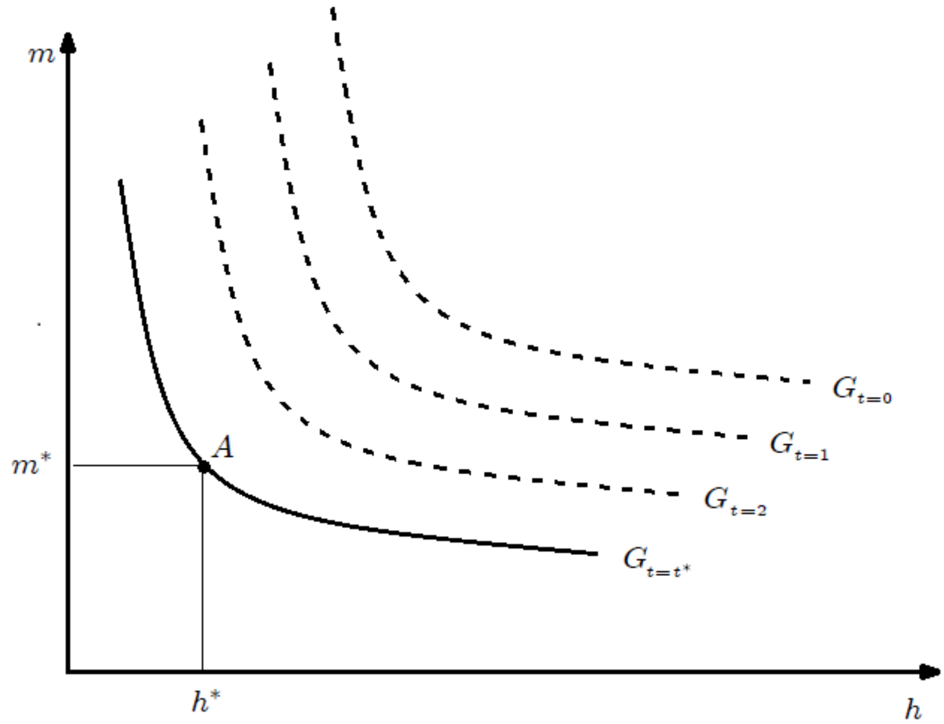

Figure 2.1: At t^* Economy A will experience the transition.

$j = l, m, h$, determine polarization in distribution of human capital in the economy. Therefore, analysing the effect of variations in such parameters on t^*, it is possible to investigate the effect of polarization in the human capital distribution on the timing of the transition from stagnation to growth.

The Effect of Polarization on the Timing of the Transition Variations in the costs of raising offspring, depicted by parameters τ^l, τ^m and τ^h, determine the prevalence of each source of labor force in the economy, in other words they determine the human capital distribution. For instance, the higher the cost of raising a child such that once adult, she will be able to adopt technology, τ^m, the lower the prevalence of adopters in the economy entailing higher polarization in the distribution of human capital. Graphi-

cally, this situation is depicted as a reduction of m^* for a given h^*. Looking at figure 2.2 an increase of τ^m can be seen as a movement from point A to point C . The same exercise can be done with the other costs parameters, for instance, raising the cost of highly educated offspring, τ^h, would decrease the presence of innovators implying a low degree of polarization in human capital distribution, which corresponds to a change of the position of the economy in figure 2.2 from point A to point B. Alternatively, increasing the cost of uneducated offspring, τ^l, reduces the lower tail of the distribution of human capital due to a relative increase in the prevalence of innovators and adopters, in other terms, the level of human capital in the economy increases. This is illustrated in figure 2.3 with a shift from point A to point D.
In other words, each point in the h, m plane depicted in figure 2.3 represents a potential distribution of human capital once the transition occurs. An economy, depending on the costs of raising offspring, will be located on one point that represents the optimal distribution of human capital. Therefore, the distance from such point to the threshold $G_{t=0}$ is proportional to the timing of the transition from stagnation to growth. Since variations in the cost parameters are associated with changes in the distribution of human capita and its degree of polarization, the analysis of the effects the cost parameters' variations on the transition timing, t^*, sheds light on how polarization in the distribution of human capital affected the transition from stagnation to growth.

Lemma 8

$$\frac{\partial t^*}{\partial \tau^l} < 0; \; \frac{\partial t^*}{\partial \tau^m} > 0; \; \frac{\partial t^*}{\partial \tau^h} > 0 \tag{2.50}$$

Proof. *See Appendix.* □

As depicted in figure 2.2, the higher the cost of raising highly educated children, τ^h, the lower the prevalence of innovators (i.e. lower polarization in the human capital distribution) implying a shift from point A to point B which, in turn, corresponds to a larger distance AA' with respect to BB' and thus economy B would need more time in order to experience the transition with respect to economy A.[14] Namely, low polarization in the distribution of human capital implied a relatively low level of innovation in the economy, ultimately delaying the transition from stagnation to growth. Similarly, higher cost of raising offspring that once adult will be enough educated to adopt innovations, τ^m, can delay the transition because it is associated with higher polarization due to the low fraction of adopters (a shift from A to C), entailing a delay in the transition from stagnation to growth (i.e. $CC' > AA'$) due to the low rate of adoption in economy C. Finally, high cost in raising uneducated offspring, τ^l, entails a larger prevalence of innovators and adopters, consisting in higher *level* of human capital, which accelerate the process of technological progress implying an early take-off.

Such exercise compares economies with different environmental conditions which, affecting the costs of raising children as represented by the parameters of the model, thus determine polarization in the human capital distribution and ultimately the timing of the transition from the Malthusian stagnation to the modern growth regime. This cross country comparison has to take into account the underlying assumption that economies considered in the analysis are isolated from a technological point of view, in other words they do not exchange either innovations or adoptions techniques. While it is understood that this is a strong condition, it is also important to consider that, in early stages of development, the transmission of technology was imperfect. In particular, the techniques to store knowledge were relatively primitive. Therefore, consistently with the proposed theory, technology were spatially diffused mainly when innovations were adopted in the production process entailing the creation of gadgets that can be spatially and inter-temporally transmitted. Implying not only that the adoption mechanism was a fundamental phenomenon of technological growth as advanced in this research, yet also that the diffusion of innovation was severely harnessed by the imperfection of transportation technology.

The model developed in this section predicts that, in early stages of devel-

[14]The effect of the speed function due to variations in τ^j, with $j = l, m, h$, exacerbates the effect of the variations of such parameters on the distance function. Therefore, it is sufficient to analyse the effect on the distance function to understand the direction of the overall effect on the timing on the transition.

opment, polarization in the distribution of human capital implied a trade-off between innovation and adoption of innovation, determining the timing of the transition from stagnation to growth. Economies characterized by excessively pronounced polarization in the distribution of human capital experienced a delayed take-off due to the disadvantage caused by an insufficient adoption rate. However, also those economies in which human capital distribution was characterized by extremely low polarization were disadvantaged in the transition due to lack of innovations. In other terms the prediction of the model is that there is a trade-off in early stages of development between excess and lack of polarization. Nevertheless, as next section analyses, despite the disadvantage in terms of the transition' timing, economies who experienced a low level of polarization are advantaged in therms of long run economic performance.

Long Run Effect of Polarization During the Malthusian stagnation, any improvement in output is eroded by a proportional increase in population that keeps constant output per capita. Namely, any increase in parental income is reflected in higher fertility rate, entailing that technological advancement are not reflected in improvement of living standards. Consistently with such mechanism, the model predicts that output per capita growth rate is null during the old regime when the Malthusian constraint is binding. Nevertheless, once the escape from the Malthusian mechanism is experienced, the transition to a modern growth regime entails a significant growth rate of output per capita thanks to human capital accumulation and its effect on technology advancements. Growth rate of output in the two regimes are given by[15],

Lemma 9

$$a)\ g_y^o = 0\ \forall\ t < t^* \tag{2.51}$$

$$b)\ g_y^n = \Omega(H_t)\ \big(1 + \lambda(M_t/N_t)\big)\ \forall\ t \geq t^* \tag{2.52}$$

Proof. See Appendix □

[15] Where $g_y^j = \frac{Y_{t+1}^j/N_{t+1}^j - Y_t^j/N_t^j}{Y_t^j/N_t^j}$ and $j = old, new$.

Comparative Statics

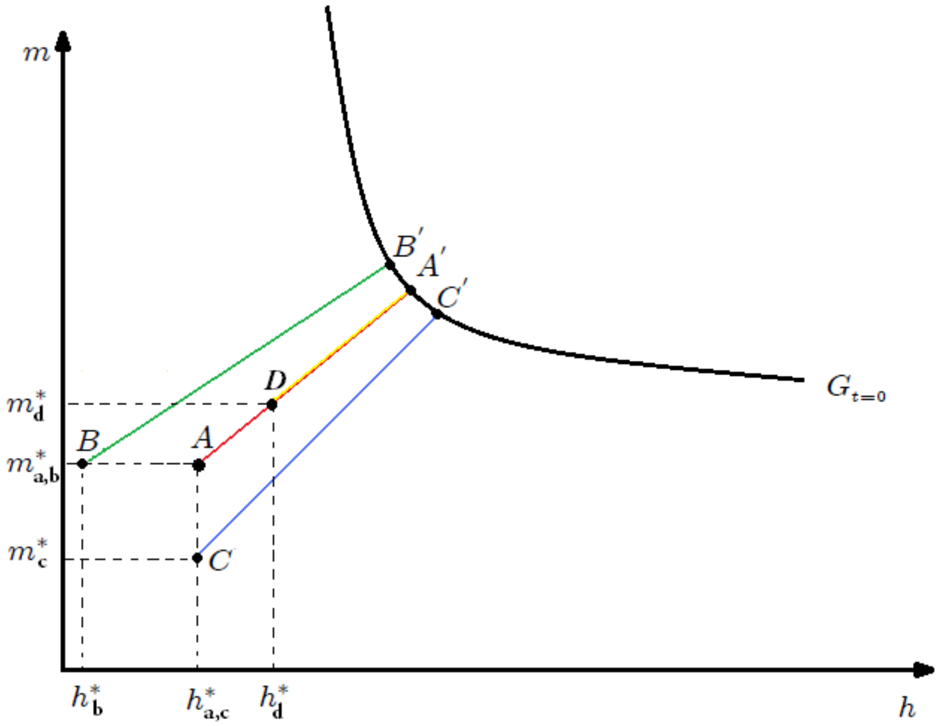

Figure 2.2: *Case 1* $[\tau^l_d > \tau^l_a]$ - In country D the level of human capital is higher with respect to A; *Case 2* $[\tau^h_b > \tau^h_a]$ - In country B polarization is lower with respect to A ; *Case 3* $[\tau^m_c > \tau^m_a]$ - In country C polarization is higher with respect to A.

Lemma 10

$$g_y^n = \Omega(n^{t-t^*})\chi \ (1 + \lambda(M_t/N_t)) \quad \forall \ t \geq t^* \tag{2.53}$$

Proof. *See Appendix* □

Given equation 2.53 it is possible to investigate the effect of polarization on the growth rate of output per capita. Such results are illustrated in figure 2.3. Namely, in the first case economy D has a higher cost of raising uneducated offspring with respect to A, $\tau_d^l > \tau_a^l$, therefore in D more educated children will be raised since their relative cost is lower. In turn, entailing an higher level of human capital in the economy andthus an earlier transition. Therefore, an economy positioned in D will be advantaged in terms of timing of the transition to a growth regime. However, increasing τ^l the benefits of the early transition are eroded by the long run disadvantage. Namely, economy in point D is characterized by higher level of human capital in early stages, however the early is not corroborated by a numerous population, inducing a disadvantage in terms of growth in output per capita in the long run. The negative effect of the parameter τ^l on the growth rate of the economy decreases marginally in magnitude, in other words for a given t the cost of having a non educated child has a U-shaped relationship with the long run growth rate of the economy. The second case is the effect of an increase in the cost raising highly educated children, τ^h, that implies lower polarization in B with respect to economy A. As depicted in figure 2.2 an 2.3 an economy with a low level of polarization will have a delayed take off and this will imply that the Malthusian mechanism will act longer in this economy, in turn entailing a higher level of population once the transition will be experienced. A larger population implies a larger number of innovators, which, being supported by a larger fraction of the population able to adopt such innovation, ultimately implies a higher growth after the take off from the Malthusian stagnation. Therefore the model predicts a sort of reversal of fortune in growth rates. Namely, countries that are disadvantaged in terms of the timing of the transition because of low polarization in the human capital distribution, experienced an advantage in terms of long run growth rate of

output per capita.
The last case is the case in which the cost of raising offspring enough educated to contribute to technology formation through the adoption of innovations in the production process, τ^m, is higher in economy C with respect to A. In this framework, economy in C is disadvantaged both in terms of a lower prevalence of labor force that can adopt new technologies and in terms of population accumulated through the Malthusian mechanism, therefore it will experience a late transition and a lower growth rate of output per capita in the long run.

2.3 Conclusions

This chapter examines the effect of polarization in the distribution of human capital on the timing of the transition from stagnation to growth and on the long run growth rate of output per capita. Polarization has an inverted-U relationship with the timing of the transition. High polarization induces a high prevalence of innovators in the economy which implies high innovation rate. However, innovation is harnessed by the absence in the population of a fraction of the labor force that is not enough educated to adopt such innovations, implying that such innovations do not translate into growth in technology, delaying the timing of the transition and involving a disadvantage in the growth rate of output per capita in the long run. In the case in which polarization in the distribution of human capital is extremely low, the fraction of labor force that is enough educated to understand and adopt new innovation is significant, entailing a high rate of adoption yet the rate of innovations will be lower due to the lower level of innovators, implying a disadvantage in terms of timing of the transition from stagnation to growth. Nevertheless, despite the disadvantage of a low polarized economy in terms of the timing of the transition, in the long run such economy will experience a higher growth rate of output per capita. The reason is that, due to the longer Malthusian era and the low cost of raising children sufficiently educated, population will be large even after the transition, implying higher growth rate of the economy. Finally, the model predicts that economies in

Long Run Effect of Polarization

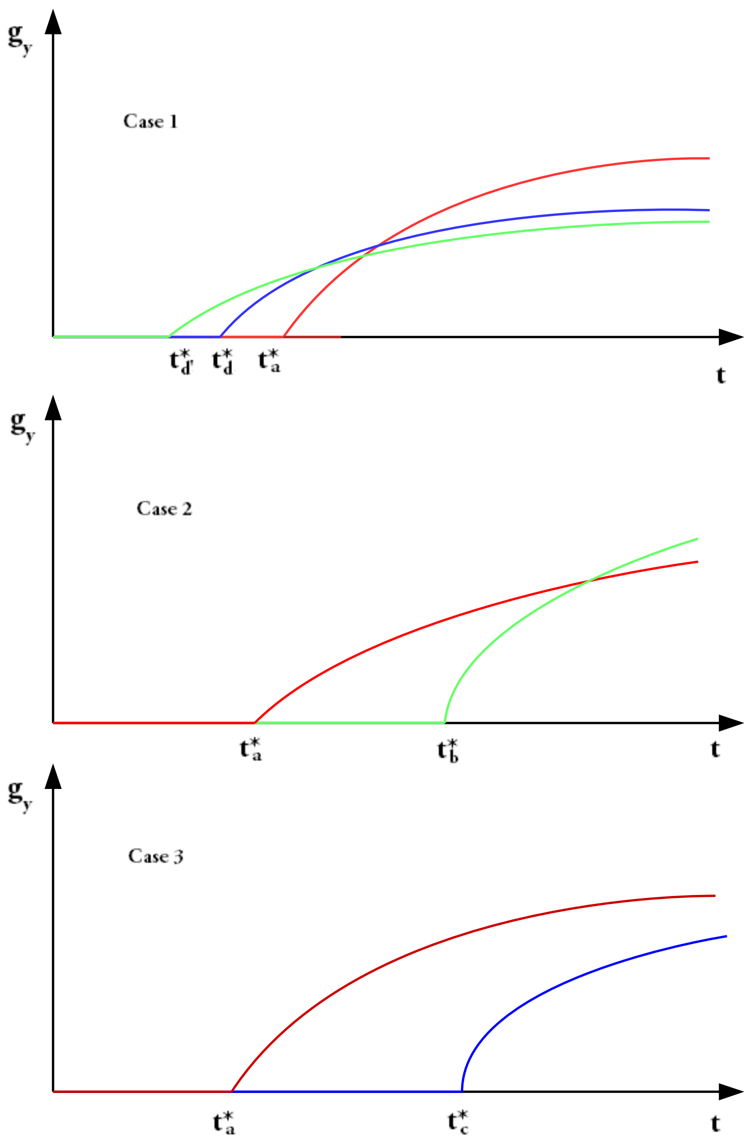

Figure 2.3: *Case 1* $[\tau^l_{d'} > \tau^l_d > \tau^l_a]$ - In country D the level of human capital is higher with respect to A; *Case 2* $[\tau^h_b > \tau^h_a]$ - In country B polarization is lower with respect to A ; *Case 3* $[\tau^m_c > \tau^m_a]$ - In country C polarization is higher with respect to A.

which the cost of producing a child independently of the level of education (such as calories intake or cost of food) is high will experience an early take off. At the same time while such source of cost of raising children has a U-shaped relationship with the long run growth rate of output per capita.
The implication of the effect of cost of raising children, given the level of human capital, on the timing of the demograohic transition will be empirically investigated in the last chapter.

Appendices

.1 Mathematical Appendix

Speed function From equation 2.47, the threshold can be written, without loss of generality, as,

$$G_t \equiv \big[h_t\big]^{\beta}\big[m_t\big]^{\phi}\frac{(1-\beta-\phi)}{(1-\alpha)}\big((h^o)^*\big)^{-\alpha}-\Big(\frac{A_t^o}{A_t^n}\Big)=0 \tag{54}$$

Therefore the speed of convergence to the new regime is calculated as the movement the points $(h_t, m_t) \in G_t$ have to do each period in order to compensate the time variation of the productivity ratio, $\Big(\frac{A_t^o}{A_t^n}\Big)$. Therefore, speed at $t+1$ is given by

$$\begin{aligned}
s_{t+1} &= \Big\{\big[h_{t+1}\big]^{\beta}\big[m_{t+1}\big]^{\phi}-\big[h_t\big]^{\beta}\big[m_t\big]^{\phi}\Big\}\frac{(1-\beta-\phi)}{(1-\alpha)} && (55)\\
&= \Big\{\frac{\big[h_{t+1}\big]^{\beta}\big[m_{t+1}\big]^{\phi}-\big[h_t\big]^{\beta}\big[m_t\big]^{\phi}}{\big[h_{t+1}\big]^{\beta}\big[m_{t+1}\big]^{\phi}}\Big\}\big[h_{t+1}\big]^{\beta}\big[m_{t+1}\big]^{\phi}\frac{(1-\beta-\phi)}{(1-\alpha)} && (56)\\
&\approx ln\Big(\frac{\big[h_{t+1}\big]^{\beta}\big[m_{t+1}\big]^{\phi}}{\big[h_t\big]^{\beta}\big[m_t\big]^{\phi}}\Big)\big[h_{t+1}\big]^{\beta}\big[m_{t+1}\big]^{\phi}\frac{(1-\beta-\phi)}{(1-\alpha)} && (57)\\
&= ln\Big(\frac{A_{t+1}^o}{A_t^o}\frac{A_t^n}{A_{t+1}^n}\Big)\frac{A_t^o}{A_t^n}\big((h^o)^*\big)^{\alpha} && (58)\\
&= \Big[ln\Big(\frac{A_{t+1}^o}{A_t^o}\Big)-ln\Big(\frac{A_t^n}{A_{t+1}^n}\Big)\Big]\frac{A_t^o}{A_t^n}\big((h^o)^*\big)^{\alpha} && (59)\\
&\approx \|\Big[-\Omega(H_t)\lambda(0)\Big)\Big]\frac{A_t^o}{A_t^n}\big((h^o)^*\big)^{\alpha}\| && (60)\\
& && (61)
\end{aligned}$$

where it is understood that speed cannot be negative, thus the absolute value is considered.

Distance function The distance d^* is the square of the minimum distance between the point (h^*, m^*) and the curve $G_{t=0}$[16]. The (h^*, m^*) point is given by equations 2.33 and 2.34; the function $G_{t=0}$ is given by the threshold

[16] It is considered the distance squared for simplicity of calculations. The results are not affected by this simplification.

function in period 0, that is

$$G_0 \equiv \big[h_t\big]^{\beta}\big[m_t\big]^{\phi}\frac{(1-\beta-\phi)}{(1-\alpha)}\big((h^o)^*\big)^{-\alpha}-\Big(\frac{A_0^o}{A_0^n}\Big)=0 \tag{62}$$

In order to find the minimum distance between a point and a curve, first it is necessary to find a point $(\tilde{h},\tilde{m}) \in G_{t=0}$ that minimize the distance function, that is,

$$\{\tilde{h},\tilde{m}\} = argmin\Big\{\big(\tilde{h}-h^*\big)^2+\big(\tilde{m}-m^*\big)^2\Big\} \tag{63}$$

taking into account that, since $(\tilde{h},\tilde{m}) \in G_{t=0} = 0$, then $\tilde{m} = \tilde{m}(\tilde{h})$. Thus, substituting $(\tilde{h},\tilde{m})$ into the generic distance function[17] squared the distance function, d^* is obtained, that is,

$$d^* = \Big\{\Big[\tilde{h}(h^*,m^*,(h^o)^*)-h^*\Big]^2+\Big[\tilde{m}\big(\tilde{h}(h^*,m^*,(h^o)^*),(h^o)^*\big)-m^*\Big]^2\Big\}$$

Notice that from the distance minimization it is sufficient to find $\tilde{h}$ to uniquely determine $\tilde{m}$. Whereas it is not possible to find an explicit solution for $\tilde{h}$, the sign of variations in the cost parameters are derived through the implicit function theorem.

Comparative Statics on the Transition Timing Given that

$$t^* = \frac{d^*}{s_{t^*}}$$

the comparative statics exercise is made on the speed function, s_{t^*}, and on the distance function, d^*.

Given equation for the speed derived above, it is straightforward to derive that

$$\partial s_{t^*}/\partial\tau^h < 0 \forall t \leq t^* \tag{64}$$

[17]The generic distance function, d, is given by

$$d = \sqrt{\Big\{\Big[\tilde{h}-h^*\Big]^2+\Big[\tilde{m}-m^*\Big]^2\Big\}}$$

and

$$\partial s_{t^*}/\partial \tau^l > 0 \forall t \leq t^* \tag{65}$$

The comparative statics exercise on d^* implies,

$$\frac{\partial d^*}{\partial \tau^m} > 0; \; \frac{\partial d^*}{\partial \tau^h} > 0; \; \frac{\partial d^*}{\partial \tau^l} < 0 \tag{66}$$

In order to make the comparative statics exercises on the distance function, d^*, it can be useful to consider the quantities $\tilde{m}$ and $\tilde{h}$. Where, from 62

$$\tilde{m} = \Big[\frac{\big((h^o)^*\big)^{\alpha}\frac{1-\alpha}{1-\beta-\phi}A_0^o}{\tilde{h}A_0^n}\Big]^{1/\phi} \tag{67}$$

from the first order conditions of the minimization problem given by 63, taking into account 67, thus $\tilde{h}$ is implicitly defined by function $\tilde{K}$, where,

$$\tilde{K} \equiv (h-h^*)-\frac{1}{\phi}\Big[\Big(\frac{\big((h^o)^*\big)^{\alpha}\frac{1-\alpha}{1-\beta-\phi}A_0^o/A_0^n}{\tilde{h}^{\beta}}\Big)-m^*\Big]\Big(\frac{\big((h^o)^*\big)^{\alpha}\frac{1-\alpha}{1-\beta-\phi}A_0^o/A_0^n}{\tilde{h}^{\beta}}\Big)\frac{\beta}{\tilde{h}} = 0 \tag{68}$$

The comparative statics is done for each of the parameters τ^j with $j = l, m, h$, where derivatives of $\tilde{h}$ are implicitly derived from $\tilde{K}$. The effect on d^* of a variation in τ^m is given by,

$$\frac{\partial d^*}{\partial \tau^m} = 2\big[\tilde{h} - h^*\big]\Big[\frac{\partial \tilde{h}}{\partial m^*}\frac{\partial m^*}{\partial \tau^m}\Big]+2[\tilde{m} - m^*]\Big[\Big(\frac{\partial \tilde{m}}{\partial \tilde{h}}\frac{\partial \tilde{h}}{\partial m^*} - 1\Big)\frac{\partial m^*}{\partial \tau^m}\Big] > 0 \tag{69}$$

Population The number of individuals in the manual labor force

$$N_t^l = \begin{cases} (1-\alpha)N_t & \forall t < t^c \\ \frac{(1-\beta-\phi)/\tau^l}{(1-\beta-\phi)/\tau^l+\beta/\tau^h+\phi/\tau^m}N_t & \forall t \geq t^* \end{cases} \tag{70}$$

The number of Adopters

$$N_t^m = \frac{(\phi)/\tau^m}{(1-\beta-\phi)/\tau^l+\beta/\tau^h+\phi/\tau^m}N_t \quad \forall t \geq t^* \tag{71}$$

The number of Innovators

$$N_t^h = \begin{cases} \alpha N_t & \forall t < t^c \\ \frac{(\beta)/\tau^h}{(1-\beta-\phi)/\tau^l+\beta/\tau^h+\phi/\tau^m} N_t & \forall t \geq t^* \end{cases} \tag{72}$$

Proof. It comes form the optimization, given the equilibrium quantities $(L_t^*, M_t^*, H_t^*) = (N_t^l l_t^{l,*}, N_t^m l_t^{m,*}, N_t^h l_t^{h,*})$, where l_t^j is the amount of working time of individual $j = l, m, h$ □

Population Dynamics

$$N_{t+1} = \begin{cases} N_t\Big(\frac{1}{(1-\alpha)\tau^l+\alpha\tau^h}\Big)\Big[1 - \frac{\tilde{c}}{A_t^o}\Big(\frac{1+h^{o,*}}{((h^o)^*)^\alpha}\Big)\Big] & \forall t < t^* \\ \\ N_t\big((1-\beta-\phi)/\tau^l + \beta/\tau^h + \phi/\tau^m\big)\gamma & \forall t \geq t^* \end{cases} \tag{73}$$

Proof. It comes form the optimization, given the equilibrium quantities $(N_t^j)$with $j = l, m, h$, given $\sum_i n_t^{i,j} = N_{t+1}^j \; \forall i \in (1, N_t)$ where i is the number of parents at time t □

Lemma 9

$$a)\ g_y^o = 0\ \forall\ t < t^* \tag{74}$$

Proof. Lemma 9 a) comes from feasibility condition that is such that

$$Y_t^o = \tilde{c} N_t^o\ \forall t^* < t \tag{75}$$

Where N_t given derived by equations 70, 71 and 72 considering that $t < t^*$ □

$$b)\ g_y^n = \Omega(H_t)\ \big(1 + \lambda(M_t/N_t)\big)\ \forall\ t \geq t^* \tag{76}$$

Proof. Lemma 9 b) comes from the fact that from equation 77, L_t is a constant fraction of N_t. Where N_t is derived by (the sum of) equations 70, 71 and 72 □

Aggregate Labor Allocation

$$L_t = \begin{cases} \frac{\tilde{c}N_t}{A_t^o}\left[\frac{\alpha}{1-\alpha}\frac{\tau^l}{\tau^h}\right]^{-\alpha} & \forall t < t^* \\ \frac{(\ 1-\gamma)(1-\beta-\phi)/\tau^l}{(1-\beta-\phi)/\tau^l+\beta/\tau^h+\phi/\tau^m}N_t & \forall t \geq t^* \end{cases} \tag{77}$$

$$H_t = \begin{cases} \frac{\tilde{c}N_t}{A_t^o}\left[\frac{\alpha}{1-\alpha}\frac{\tau^l}{\tau^h}\right]^{1-\alpha} & \forall t < t^* \\ \frac{(\ 1-\gamma)(\beta)/\tau^h}{(1-\beta-\phi)/\tau^l+\beta/\tau^h+\phi/\tau^m}N_t & \forall t \geq t^* \end{cases} \tag{78}$$

$$M_t = \begin{cases} 0 & \forall t < t^* \\ \frac{(\ 1-\gamma)(\phi)/\tau^m}{(1-\beta-\phi)/\tau^l+\beta/\tau^h+\phi/\tau^m}N_t & \forall t \geq t^* \end{cases} \tag{79}$$

Equilibrium wages

$$w_t^h = \begin{cases} \alpha A_t^o\left[(h^o)^*\right]^{\alpha-1} & \text{if } Y_t^o > 0 \\ \beta A_t^n\left[(h^n)^*\right]^{\beta-1}\left[m^*\right]^{\phi} & \text{if } Y_t^n > 0 \end{cases} \tag{80}$$

$$w_t^m = \phi A_t^n\left[(h^n)^*\right]^{\beta}\left[m^*\right]^{\phi-1} \quad \text{if } Y_t^n > 0 \tag{81}$$

$$w_t^l = \begin{cases} (1-\alpha)A_t^o\left[(h^o)^*\right]^{\alpha} & \text{if } Y_t^o > 0 \\ (1-\beta-\phi)A_t^n\left[(h^n)^*\right]^{\beta}\left[m^*\right]^{\phi} & \text{if } Y_t^n > 0 \end{cases} \tag{82}$$

The Dynamical System In order to find H_{t^*} it is necessary to solve the following dynamical system.
From 73 and 78

$$H_{t+1} = H_t\frac{A_t^o}{A_{t+1}^o}\Big(\frac{1}{(1-\alpha)\tau^l+\alpha\tau^h}\Big)\Big[1-\frac{\tilde{c}}{A_t^o}\Big(\frac{1+(h^o)^*)}{((h^o)^*)^{\alpha}}\Big)\Big] \ \forall t < t^* \tag{83}$$

where $(h^o)^*)$ is given by 2.24.
From 2.39

$$A_{t+1}^o = A_t^o\left[1+\Omega(H_t)\right] \tag{84}$$

In other terms the dynamical system is given by

$$H_{t+1} = f(A_{t+1}^o, A_t^o)H_t \tag{85}$$

The Dynamical System

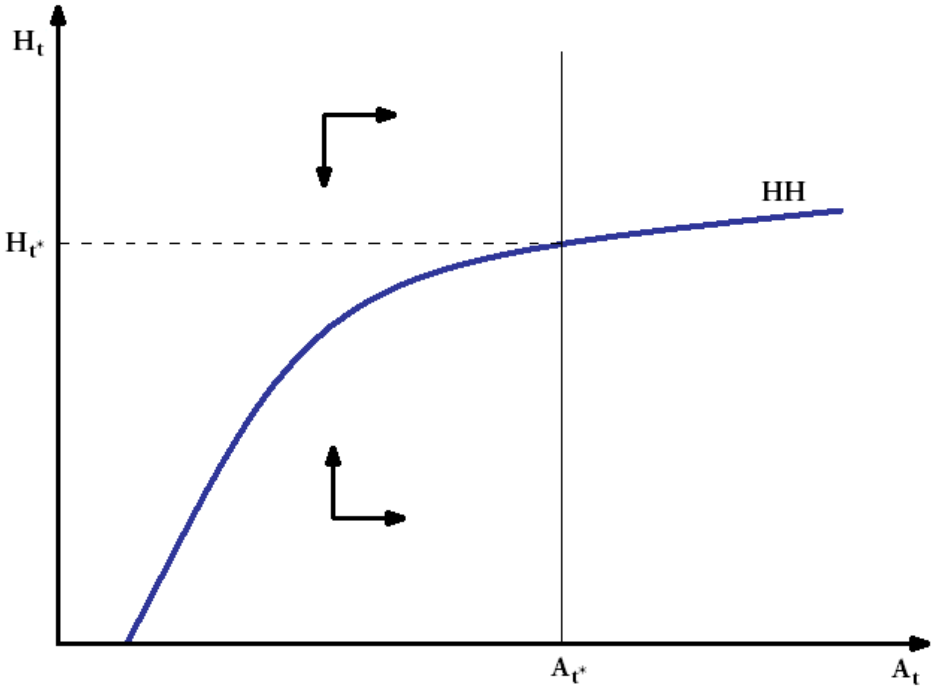

Figure 4: H_t^* is the solution to the dynamical systemwith $A_t^o = A_{t^*}$.

and

$$A^o_{t+1} = f(A^o_t)H_t \tag{86}$$

Therefore, the AA locus is defined as

$$AA = \{(H_t, A^o_t)|\Delta A^o_t = 0\} \tag{87}$$

where $\Delta A^o_t \equiv A^o_{t+1} - A^o_t = A^o_t\big[1 + \Omega(H_t)\big] - A^o_t = 0$ if $\Omega(H_t) = 0$ which is not feasible since $\Omega(0) > 0$ by construction. The intuition is that in this economy technology does not have a steady state in levels because there will always be technological advancements.
The HH locus is given by

$$HH = \{(H_t, A^o_t)|\Delta H_t = 0\} \tag{88}$$

where $\Delta H_t = H_{t+1} - H_t = 0$. Such condition, taking into account 86, is satisfied by

$$H_t = \Omega^{-1}\Big(\big(\frac{1}{(1-\alpha)\tau^l + \alpha\tau^h}\big)\big[1 - \frac{\tilde{c}}{A^o_t}\big(\frac{1+(h^o)^*)}{((h^o)^*)^\alpha}\big)\big] - 1\Big) \;\forall t \leq t^* \tag{89}$$

Let define, H_t as,

$$H_t \equiv \Omega^{-1}\Big(\chi(A^o_t)\Big) \;\forall t \leq t^* \tag{90}$$

The 85 implies that $\frac{\partial A_t}{\partial H} > 0$ and $\frac{\partial^2 A_t}{\partial H_t^2} < 0$ as represented by the blue line in figure 4.
Finally, $H_{t^*} = \Big\{H_t|A^o_t = A_{t^*}\Big\}$ where $A_{t^*} = \Big\{A^o_t|w^{o,l}_t = \tilde{c}/(1-\gamma)\Big\}$ where $w^{o,l}_t$ is the equilibrium wage in the old regime, given by 2.23. That is

$$A_{t^*} = \frac{\tilde{c}}{(1-\gamma)(1-\alpha)\big((h^o)^*\big)^\alpha} \tag{91}$$

Lemma 10

$$g^n_y = \Omega(n^{t-t^*})\chi \;\; \lambda(M_t/N_t) \quad \forall \; t \geq t^* \tag{92}$$

Proof. It comes from equation 76, given that from the dynamics of population, $\lambda(M_t/N_t)$ is constant over, considering that H_t is given by

$$H_t = n^{t-t^*} H_{t^*} \ \forall t > t^*$$

. Where n is given by

$$n = \Big(\frac{1-\beta-\phi}{\tau^l} + \frac{\beta}{\tau^h} + \frac{\phi}{\tau^m}\Big)\gamma \tag{93}$$

Given that H_{t^*} is the solution of the dynamical system, which is given by

$$H_{t^*} = \Omega^{-1}(\chi(A_{t^*}))$$

, where $\chi(A_{t^*})$ is given by

$$\chi(A_{t^*}) = \Big(\frac{1}{(1-\alpha)\tau^l + \alpha\tau^h}\Big)(1 - (1 + \big((h^o)^*\big))(1-\alpha)(1-\gamma)) \tag{94}$$

Where in the latter I make use of equation 91. $\square$

Chapter 3

The Empirical Contribution

3.1 Introduction

This Chapter performs an empirical test of the hypothesis that economies characterized by high cost of raising children experienced an early demographic transition. In particular, as predicted by the theoretical model presented in chapter 2, an increase in the cost of raising children, for a given level of human capital, decreases the cost of education in relative terms, implying accumulation of knowledge and thus technological advancement in early stages of development that, in turn, affects the timing of the transition from stagnation to growth.

One of the main challenges undertaken in the empirical exercise consists in finding a variable that can be a reasonable measure for the cost of raising children. In principle, there are both geographical and institutions related variables that affect the cost of raising children. However, due to the difficulty in finding an exogenous measure of institutions, the focus here will be on geographic variation that can affect the cost of raising children, conditioning on institutions.

In particular, since in early stages of development the main cost parents had to sustain for their children was (and in some countries still is) food, the focus of the empirical exercise is to consider calories intake as a measure of cost of raising children. One question might be about the source of cross-country variation in calories intake. A possible answer can be found in two fundamental laws of Biology. In particular, calories intake depends on temperature and body size. Namely, for a given body size, the colder the environment the larger the amount of calories needed in order to keep the body warm. At the same time, a bigger body also implies larger calories consumption due to metabolic needs (Kleiber, 1932). Finally, as will be explained in the following, in colder climates individuals attain larger body size as established by Bermann's rule (Newman, 1953). In other words, since the three variables - calories intake, temperature and body size- are strictly correlated, it is possible to exploit the cross country variations of temperature and body size, as proxy for calories intake.

While we can be relatively ensured that average yearly temperature in a country remained relatively the same over the last centuries, body size changed significantly from the onset of the industrial revolution (Dalgaard and Strulik, 2010). However, under the assumption that the cross country variations in body size remained relatively stable over time, the empirical exercise support the prediction of the model that countries characterized by higher cost of raising children (as proxied by bigger body size) experienced the demographic transition earlier. In addition, consistently with the Biology literature, such countries are on average colder.

The hypothesis is empirically investigated based on the analysis of cross-country data. The aim is to contribute to the empirical growth literature in two dimensions. First, in the literature typically there are controls for absolute latitude or temperature (see for instance Ashraf and Galor (2013) and Michalopoulos (2012)). This work aims to advance that, controlling for temperature or equivalently absolute latitude, consists in controlling for cost of raising children. Another contribution of this exercise is that it gives additional cross-country evidence on the empirical literature that establishes the quantity-quality trade-off first introduced by Becker (Becker et al., 1960) and widely accepted in the literature.

The chapter is organized as follows. Next Section introduces the fundamental mechanism that is behind the measurement methodology to empirically analyze the cost of raising children independently of human capital. Namely, it presents an established law in biology and zoology according to which individuals who live in colder climates are also characterized by larger body size, in addition it illustrates a well known fact that body size is strongly correlated with calories intake, thus establishing that calories intake (that acts as a proxy for cost of raising children) can be measured by body size. Such rule will be exploited in the empirical approach which is presented in Section 3.3. Last Section concludes.

3.2 Measuring Cost of Raising Children

3.2.1 Calories Intake as Measure of Cost of Raising Children

Parental choice to raise children is naturally associated to the costs related to such investment in offspring. The assumption introduced in this empirical analysis is that calories intake of children is a fundamental measure cost of raising children. Under this assumption, it is possible to exploit variation in environmental conditions that affect calories intake and ultimately have a consistent measure of cost of raising children.

Despite the simplifying assumption of measuring cost or raising children based on calories intake, the latter is far from being easy to measure. However, it is possible to make use of a well known principle in Biology: the Kleiber's law. According t this law, for the vast majority of animals (including humans), the metabolic rate is positively associated with body mass (Kleiber, 1932). Figure 3.1 depicts the graphical representation of the data presented in the original work by Kleiber (1947).

Moreover Kleiber's law has been statistically proved to hold within species and in particular for humans, taking into account gender and age. Namely, "each centimeter per kg$^{1/3}$ increase in specific stature produces, on the average, an increase of 1 per cent of the metabolic rate for men, and 1.8 per cent of the metabolic rate of women" (Kleiber, 1947).[1] The fundamental conse-

[1]The metabolic rate is the amount of calories consumed by an individual in a given period of time. The study takes into account deviation from the age of 30 years.

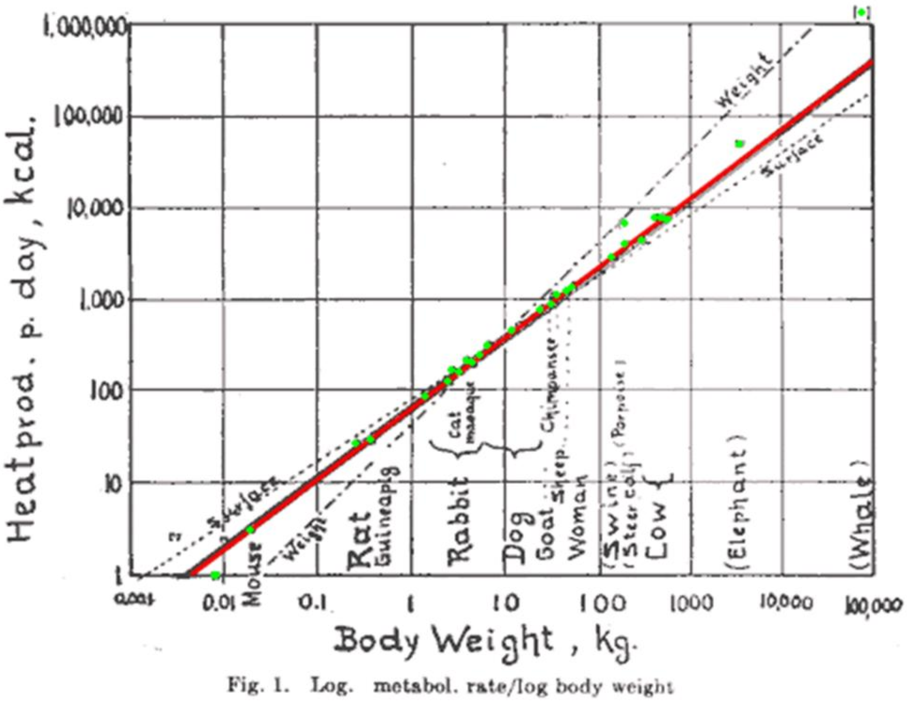

Figure 3.1: Graphical Representation of Kleiber's Law. Source: Kleiber (1947)

quence of this principle is that body size acts as a measure of the calories intake in particular for women, given that the correlation is much stronger for females. In particular, in the empirical specification the measure of average body size chosen is country average height. Such variable is the proposed measure of cost of raising children.
At this point, a potential question could be related to the environmental conditions that make calories intake, and thus the cost of raising children, vary across countries. The answer to this question will be developed in the following.

3.2.2 Calories Intake and Climatic Conditions

Correlations between the physical characters of warm-blooded vertebrates and their environment have been established by zoologists that formulated several rules. In particular, they established the so called Bermann's rule. That is "[...] within a single wide-ranging species of warmblooded animal, the subspecies or races in colder climates attain greater body size than those in warmer climates" (Newman, 1953).
The correlation between body size and temperature established by the Bergmann's rule, has important implications for what concerns the cross country variations of the cost of raising children. Namely, countries that are characterized by colder average temperature will be characterized by larger body size that, in turn, implies larger calories intake and thus higher cost of raising children. As noted above, it is necessary to adopt a measure of body size that has to take into account gender differences. The reason is that, since gender differences are not necessarily constant across countries, this would be translated in additional source of noise that can imply attenuation bias. Therefore, to deal with this problem the measure of body size that will be used in the empirical specification is country average height of women.
There are several reasons why women body size is a better indicator than body size of men. First, as mentioned above, the correlation between women body size and calories consumption is stronger, as established in Kleiber (1947). Second, children size is strongly correlated with mother body size

(Charnov, 1991, 1993), which, in turn affects calories consumption of the child both in utero and during breast feeding (Dalgaard and Strulik, 2010). Therefore country average height of females is simply a superior measure of children body size and therefore will be used in the empirical analysis. As already explained, the measure of body size is used as proxy for calories intake with the aim of capturing the cross country variability in the cost of raising children.
Using this approach it is therefore possible to test for the advanced hypothesis that countries characterized by higher calories intake, and thus higher cost of raising children, are prone to experience early take-off from Malthusian stagnation to the modern growth regime. The test of this hypothesis will be developed in the following section.

3.3 Empirical Exercise

3.3.1 Data

The data are from several sources and the unit of observation is a country. Data on the demographic transition timing are from Reher (2004); the control variables are from **?**Ashraf and Galor (2013) and Comin et al. (2010) that are based on Peregrine (2003) and many other sources.The variable Average female height is from World Bank Development indicator and refers to average height of adult women. The variable Temperature is yearly average temperature in a country.

3.3.2 Empirical Specification and Results

Figure 3.2 depicts the cross country correlation between the indicator of body size (acting as a proxy for cost of raising children) and number of years elapsed since the demographic transition. The graphical analysis of the data supports the hypothesis that countries that are characterized by larger body size for females, and then are characterized by higher costs of raising children, experienced the demographic transition earlier with respect to those

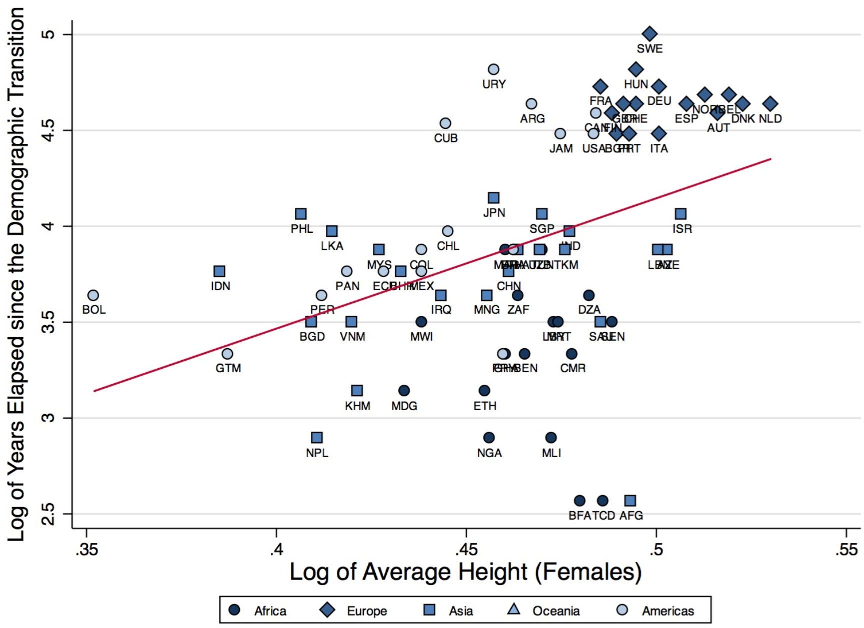

Figure 3.2: Correlation between Calories Intake Indicator (Average Height of Females) and Years Elapsed since the Demographic Transition

countries that are characterized by low cost of raising children. In particular, without additional controls, the effect is estimated (as represented by the slope of the line in figure 3.2) to be 6.796. This is also illustrated in the first column of the reduced form specification of table 3.1. This is consistent with the hypothesis that higher cost of raising children, as measured by females body size, is associated with early demographic transition.

Table 3.1 illustrates the reduced form specificationand investigates whether the expected positive correlation survives to the introduction of controls generally used in the long-run growth literature. Namely, controlling for genetic diversity and genetic diversity squared it is possible to control for economic development (Ashraf and Galor, 2013) that would imply reverse causality. Similarly, controlling for average elevation and standard deviation of elevation introduces a control for ethno-linguistic diversity, consistently with Michalopoulos (2012), which could affect both the body size indicator and economic development in the early stage of development (timing from the transition), in turn implying bias in the estimated coefficients. Thus, to min-

imize the severity of omitted variable bias, a number of covariates are added to control also for land productivity. Such specification is depicted in column 2, which report that the coefficient of interest is positive and precisely estimated. Column 3 performs a similar specification, adding continental fixed effects. Again the coefficient is statistically significant at 1 per cent. Column 4 introduces controls for urbanization level in year 0 AD (Peregrine, 2003) and a dummy variable that is 1 if the country has been a European colony (Comin et al., 2010). The control for urbanization is introduced to possibly condition for the cost of education. In order to further control for institutions that can affect cost of raising children and timing of the demographic transition it is introduced tha control related to European colonial institutions. Column 5 illustrates a similar specification with the introduction of continental fixed effects. Again the coefficient of interest is precisely estimated.
The reduced form analysis illustrated in table 3.1 is consistent with the hypothesis that an increase in cost of raising children, on average, associated with earlier transition.

Table 3.1: Reduced Form Estimates of the effect of Calories Intake, as measured by the Body Size indicator, on the Timing Elapsed since the Demographic Transition. Ordinary Least Squares.

	Log of Years Elapsed Since the Demographic Transition				
VARIABLES	(1)	(2)	(3)	(4)	(5)
Log of Avg Height of Females	6.796***	10.74***	4.793***	8.063***	5.071**
	(1.440)	(1.754)	(1.589)	(1.939)	(1.903)
Log of Years Since the Neolithic		0.154	-0.225	0.0150	-0.260*
		(0.139)	(0.139)	(0.137)	(0.153)
Genetic Diversity (predicted)		65.02**	-4.081	41.14	-7.972
		(30.74)	(35.51)	(32.23)	(35.98)
Genetic Diversity Squared (predicted)		-52.00**	1.060	-34.37	4.225
		(22.64)	(25.41)	(23.55)	(25.98)
Average Elevation		-0.550	-0.516*	-0.604	-0.475*
		(0.358)	(0.266)	(0.380)	(0.264)
Standard Deviation of Elevation		0.464	0.497	0.437	0.513
		(0.427)	(0.324)	(0.429)	(0.325)
Log of Arable Land		0.0705	0.0362	0.0358	0.0358
		(0.113)	(0.104)	(0.121)	(0.111)
Log of Land Suitability		0.0200	-0.0307	0.0251	-0.0408
		(0.0892)	(0.0771)	(0.0966)	(0.0873)
Urbaniz. 0 AD				-0.0193	-0.0509
				(0.108)	(0.0923)
Dummy for Colony				-0.496***	0.119
				(0.131)	(0.154)
Continental FE	N	N	Y	N	Y
Observations	84	69	69	65	65
R-squared	0.167	0.498	0.730	0.575	0.743

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 3.2 represent an empirical assessment of the Bermann's rule in the cross country data. The aim is to test whether such law is satisfied in this sample. The specifications in the columns 1-5 of table 3.2 are very similar to the ones in table 3.1, with the crucial difference that here the aim is to investigate whether temperature affects body size and thus calories intake. The table illustrates that the Bergmann's rule is satisfied in this data and thus average height of women is a good indicator for body size.
Naturally, according to the Bergmann's rule, a natural candidate for an instrument is exactly temperature. Despite the fact that such variable is exogenous, it might affect the dependent variable through dimensions that are unobservable, thus violating the exclusion restriction assumption. However, if we believe that such violation is not severe we can perform the exercise based on the IV specification under such important caveat. In this case the estimates of the coefficient of interest are statistically significant and larger in magnitude in all specifications, suggesting that attenuation bias from measurement error is playing a role.[2]

[2]The IV estimates are not reported. However, they are available from the author on request.

Table 3.2: The Bergmann's Rule in Cross-Country Data. Ordinary Least Squares.

	Log of Avg Height of Females				
VARIABLES	(1)	(2)	(3)	(4)	(5)
Log of Temperature	-0.0693***	-0.0479***	-0.0521***	-0.0419***	-0.0422***
	(0.0146)	(0.0150)	(0.0143)	(0.0130)	(0.0125)
GDP pc 2000		1.05e-06***	7.86e-07**	4.39e-07*	6.93e-07**
		(2.93e-07)	(3.21e-07)	(2.55e-07)	(3.14e-07)
Genetic Diversity (predicted)			0.220***	0.388**	-1.772
			(0.0630)	(0.170)	(1.896)
Genetic Diversity Squared (predicted)					1.429
					(1.429)
Average Elevation			0.00255	-0.000506	0.00376
			(0.0143)	(0.0143)	(0.0126)
Standard Deviation of Elevation			-0.0244	-0.0139	-0.0278
			(0.0185)	(0.0197)	(0.0171)
Log of Years Since the Neolithic					0.0255**
					(0.0111)
Log of Arable Land					0.00946*
					(0.00483)
Log of Land Suitability					-0.00608
					(0.00413)
Urbaniz. 0 AD					-0.00752
					(0.00605)
Dummy for Colony					-0.00946
					(0.00959)
Continental FE	N	N	N	Y	Y
Observations	91	91	88	88	79
R-squared	0.271	0.347	0.573	0.646	0.742

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

3.4 Conclusions

This Chapter empirically investigates one of the prediction of the model presented in Chapter 2. In particular, economies that are characterized by environmental and specifically climatic conditions that imply high cost of raising children (such as low temperature) on average experienced the demographic transition in advance with respect to others. The theory proposed suggests that high cost of raising children implies a low relative cost of human capital, affecting knowledge and then technology accumulation, ultimately implying an early transition from stagnation to growth.
Naturally, it is understood that testing for a subset of the predictions of the model is not a test for the all model itself. Nevertheless, the empirical exercise provides additional evidence that the theory advanced in this research is consistent with what observed in the world.

Conclusions

Over the centuries, scholars from many disciplines developed several theories to answer questions related to the fundamental factors that can explain disparities in living standards across societies. Naturally, such big questions do not have a unique answer. However, after the seminal book by Diamond (1998), a significant portion of the literature aimed to highlight that fundamental determinants of the disparities we observe across countries are rooted in the far past. In particular, environmental conditions determined the structure of incentives for individuals, in turn, affecting their decisions about fertility as well as investment in human capital of their offspring. This literature is, though only partially, reviewed in Chapter 1.

This research, somehow ambitiously, advanced the hypothesis that, being complementary to others in this literature, aims to shed light on possible channels through which initial conditions in the far past have long lasting effects on the distribution of income across countries today. Namely, the distribution of human capital had an effect on the timing of the transition from stagnation to growth, in turn explaining part of the disparities in income per capita that we observe today. In particular, the interplay between innovation and adoption of technologies as determined by fertility and education choices, affecting growth in technology, deeply affected the long run performance of the economy. Chapter 2 analyses in details the theoretical model that describes the advanced hypothesis and its implication.

At this point, the natural question is related to what are the possible determinants of the distribution of human capital. In the theory developed in chapter 2 the parameters of the model play a key role in determining such distribution. These parameters represent the cost of raising children independently on their level of education, as well as the cost of their human capital. Such costs can be determined both by environmental conditions related to geographical characteristics as well as institutions. In particular, geography can determine both the calories intake needed as well as the cost of food production. Institutions instead can affect the cost of acquiring human capital depending on the presence of investment in public goods and in particular schooling. Chapter 3 is a first attempt to exploit the geographic variation in explaining the distribution of human capital, controlling for the institutions channel. However more research has to be done in order to better understand what are the environmental conditions that, affecting institutions, would imply variation in the cost of acquiring human capital and, importantly, to find exogenous variation in order to identify empirically such effects.

Bibliography

Acemoglu, D., Aghion, P., and Zilibotti, F. (2006). Distance to frontier, selection, and economic growth. *Journal of the European Economic association*, 4(1):37–74.

Aghion, P. and Howitt, P. (1992). A model of growth through creative destruction. *Econometrica: Journal of the Econometric Society*, pages 323–351.

Ashraf, Q. and Galor, O. (2011). Dynamics and stagnation in the malthusian epoch. *The American Economic Review*, 101(5):2003–41.

Ashraf, Q. and Galor, O. (2013). The out of africa hypothesis, human genetic diversity, and comparative economic development. *The American Economic Review*, 103(1):1–46.

Ashton, T. S. (1955). *An economic history of England: the 18th century*, volume 3. Taylor & Francis.

Basu, S. and Weil, D. N. (1998). Appropriate technology and growth. *The Quarterly Journal of Economics*, 113(4):1025–1054.

Becker, G. S., Duesenberry, J. S., and Okun, B. (1960). An economic analysis of fertility. In *Demographic and economic change in developed countries*, pages 209–240. Columbia University Press.

Bekar, C. and Lipsey, R. (2004). Science, institutions and the industrial revolution. *Journal of european economic history*, 33(3):709–753.

Ben-Porath, Y. (1967). The production of human capital and the life cycle of earnings. *Journal of Political Economy*, 75(4):pp. 352–365.

Benhabib, J. and Spiegel, M. M. (2005). Human capital and technology diffusion. *Handbook of economic growth*, 1:935–966.

Bleakley, H. (2007). Disease and development: evidence from hookworm eradication in the american south. *The Quarterly Journal of Economics*, 122(1):73–117.

Charnov, E. L. (1991). Evolution of life history variation among female mammals. *Proceedings of the National Academy of Sciences*, 88(4):1134–1137.

Charnov, E. L. (1993). *Life history invariants: some explorations of symmetry in evolutionary ecology*. Oxford University Press Oxford.

Clark, G. (2001). The secret history of the industrial revolution. *Manuscript, University of California, Davis.*

Clark, G. (2002). Land rental values and the agrarian economy: England and wales, 1500–1914. *European Review of Economic History*, 6(3):281–308.

Comin, D., Easterly, W., and Gong, E. (2010). Was the wealth of nations determined in 1000 bc? *American economic journal. Macroeconomics*, 2(3):65–97.

Comin, D. and Hobijn, B. (2010). An exploration of technology diffusion. *American Economic Review*, 100(5):2031–59.

Dalgaard, C.-J. and Strulik, H. (2010). The physiological foundations of the wealth of nations. *Univ. of Copenhagen Dept. of Economics Discussion Paper*, (10-05).

Diamond, J. M. (1998). *Guns, germs and steel: a short history of everybody for the last 13,000 years*. Random House.

Doepke, M. and Zilibotti, F. (2005). The macroeconomics of child labor regulation. *American Economic Review*, pages 1492–1524.

Durlauf, S. N. and Johnson, P. A. (1995). Multiple regimes and cross-country growth behaviour. *Journal of Applied Econometrics*, 10(4):365–384.

Galor, O. (2005). From stagnation to growth: unified growth theory. *Handbook of economic growth*, 1:171–293.

Galor, O. (2011). *Unified growth theory*. Princeton University Press.

Galor, O. and Moav, O. (2002). Natural selection and the origin of economic growth. *The Quarterly Journal of Economics*, 117(4):1133–1191.

Galor, O. and Mountford, A. (2008). Trading population for productivity: theory and evidence. *The Review of Economic Studies*, 75(4):1143–1179.

Galor, O. and Tsiddon, D. (1997). Technological progress, mobility, and economic growth. *The American Economic Review*, pages 363–382.

Galor, O. and Weil, D. N. (1999). From malthusian stagnation to modern growth. *The American Economic Review*, 89(2):150–154.

Galor, O. and Weil, D. N. (2000). Population, technology, and growth: From malthusian stagnation to the demographic transition and beyond. *American economic review*, pages 806–828.

Gerschenkron, A. et al. (1962). Economic backwardness in historical perspective. *Economic backwardness in historical perspective.*

Hazan, M. (2009). Longevity and lifetime labor supply: Evidence and implications. *Econometrica*, 77(6):1829–1863.

Kleiber, M. (1932). Body size and metabolism. *ENE*, 1:E9.

Kleiber, M. (1947). Body size and metabolic rate. *Physiol. Rev*, 27(4):511–541.

Maddison, A. (2001). *The World Economy: A Millennial Perspective.* OECD Publishing.

Maddison, A. (2003). *Development Centre Studies The World Economy Historical Statistics: Historical Statistics.* OECD Publishing.

Malthus, T. (1798). An essay on the principle of population (printed for j. johnson, in st. paul's church-yard, london).

Michalopoulos, S. (2012). The origins of ethnolinguistic diversity. *The American Economic Review*, 102(4):1508–1539.

Mokyr, J. (1990). *The lever of riches: Technological creativity and economic progress.* Oxford University Press.

Nelson, R. R. and Phelps, E. S. (1966). Investment in humans, technological diffusion, and economic growth. *The American Economic Review*, 56(1/2):pp. 69–75.

Newman, M. T. (1953). The application of ecological rules to the racial anthropology of the aboriginal new world*. *American Anthropologist*, 55(3):311–327.

Peregrine, P. N. (2003). *Atlas of cultural evolution.*

Reher, D. S. (2004). The demographic transition revisited as a global process. *Population, space and place*, 10(1):19–41.

Rolt, L. T. C., Rolt, L. T. C., and Rolt, L. T. C. (1963). *Thomas Newcomen: the prehistory of the steam engine.* David and Charles.

Solow, R. M. (1956). A contribution to the theory of economic growth. *The quarterly journal of economics*, 70(1):65–94.

Stapledon, R. G., Davies, W., et al. (1948). Ley farming. *Ley Farming*, (Edn 2 (revised)).

Stuart, R. (1829). *A descriptive history of the steam engine.* Number 24502. printed for Knight and Lacey.