# The Great Divergence: A Network Approach *

Ines Lindner† and Holger Strulik**

First Version: March 2014. This Version: November 2014.

**Abstract.** We present a multi-country theory of economic growth in which countries are connected by a network of mutual knowledge exchange. Knowledge in any country depends on the human capital of the countries it exchanges knowledge with. The diffusion of knowledge throughout the world explains a period of increasing world inequality after the take-off of the forerunners of the industrial revolution, followed by decreasing relative inequality. Knowledge diffusion through a Small World network explains the 'New Kaldor facts' and produces an extraordinary diversity of country growth performances, including the overtaking of individual countries in the course of world development.

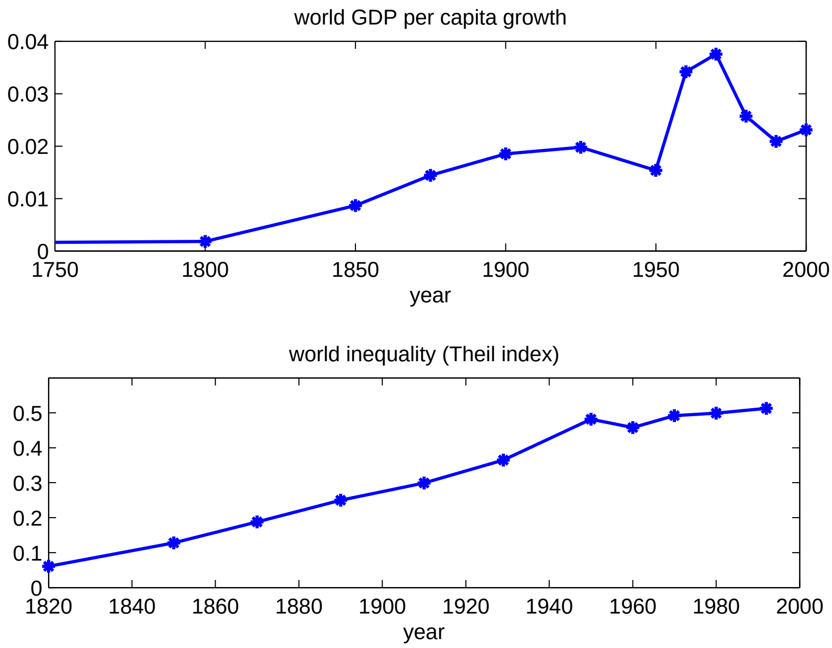
*Keywords:* networks, knowledge diffusion, economic growth, world income distribution.
*JEL:* O10; O40; D62; D85; F41.

*We would like to thank Jean-Louis Arcand, Francois Bourguignon, Clive Bell, Christiane Clemens, Guido Cozzi, Oded Galor, Jan Willem Gunning, Sophia Kan, Alessandra Pelloni, Robert Waldmann, and David Weil for helpful comments.
†VU University Amsterdam, Department of Econometrics and Operations Research, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands, email: ilindner@feweb.vu.nl
**University of Göttingen, Department of Economics, Platz der Göttinger Sieben 3, 37073 Göttingen, Germany; email: holger.strulik@wiwi.uni-goettingen.de.

## 1. Introduction

In pre-modern times, before the take-off to long-run growth of the countries that led the industrial revolution, national income differences were minuscule from today's perspective. Bairoch (1993, Ch. 9) reviews the literature and comes to the conclusion that even in the mid-18th century the income differential between future least developed countries and future developed countries was of the order of 1.1 to 1.3. With the beginning of the industrial revolution, the world witnessed the "great divergence". Income inequality between countries, measured by the Theil index, increased from 0.06 in 1820 to 0.25 in 1870 to 0.48 in 1950 to 0.5 in 1980 (according to Bourguignon and Morrisson, 2002). Since then, the increase of inequality has slowed down to a point such that researchers speculate whether it has settled at a steady state or started to decline (e.g. Jones, 1997; Acemoglu and Ventura, 2002; Sutcliffe, 2002). Figure 1 shows the gradual increase of world income growth and the evolution of world inequality since the onset of the industrial revolution.[1]

Figure 1: World Economic Growth and Inequality




Data from De Long (1998) and Bouguignon and Morrisson (2002).

The country-specific differences in the timing of the gradual take-off from stagnation to long-run growth are a major theme in unified growth theory (Galor, 2005, 2011). It is

[1]We refer to the inequality of GDP per capita between countries, which is the relevant measure for the theory developed below. For inequality of income between world citizens, the evidence is stronger that it actually declined since the 1980s due to the take-off of populous China and India (Sala-i-Martin, 2006). The term Great Divergence was initially coined by Pommeranz (2000) with respect to the divergent evolution of China and the West. It is now more broadly applied to the divergent evolution of income per capita across the world (Galor, 2005).

argued that the varying time of the take-off to growth contributed significantly to both increasing world inequality and the emergence of convergence clubs (i.e. clusters of countries that grow similarly with respect to each other but differently to other countries). Unified growth theory, however, largely focuses on countries conceptualized as closed economies, which implies, in particular, the notion that each country independently generated its own impulse for the take-off to growth.[2]

This paper proposes a different approach. It considers a world of many countries connected by a network of knowledge exchange. As knowledge diffuses gradually through the world, more and more countries are "infected", their population starts investing in education, and their economy takes off to long-run growth. With more and more countries jumping on the bandwagon of growth, world income per capita increases gradually towards a balanced growth path. The individual timing of the take-off is explained by the countries' closeness to the leaders of the industrial revolution. Countries which are connected directly, or with only a few links, adopt the knowledge created in the leader countries earlier compared to poorly connected or "remote" countries. Take-off to growth of the forerunners of the industrial revolution is, naturally, accompanied by increasing world inequality, as the income gap with respect to the backward countries gets larger. Eventually, however, knowledge diffuses through the whole world and the remote countries also take off. Because the available knowledge has increased tremendously since the take-off of the original leaders, the latecomers of the industrial revolution have more to learn from and thus, they take off faster, at rates that temporarily exceed the balanced growth rate. This feature, that the growth rates of latecomers temporarily overshoot the balanced growth rate implies that *relative* world inequality eventually declines.

Most of the related literature focuses exclusively on relative inequality measured e.g. by the conventional Gini index or Theil index. One exception is Atkinson and Brandolini (2010) who discuss alternative measures of absolute inequality based on Kolm (1976) and find that it accelerated since the 1950s, i.e. during the period when relative inequality leveled off. We show that our network theory of long-run growth captures this phenomenon as well. We compute the absolute Gini index, defined by the product of the Gini index and average income (Chakravarty 1988), and show that declining relative income inequality is predicted to be accompanied by increasing income gaps in absolute levels.

In order to focus on the knowledge diffusion process, the underlying economic model is a deliberately simple one. It is basically a multi-country version of Romer (1986). The main difference is that a country's factor productivity not only depends on the human

[2] See Galor and Weil (2000), Kögel and Prskawetz (2001), Jones (2001), Galor and Moav (2002, 2004), Boucekkine et al. (2002), Doepke (2004), Strulik and Weisdorf (2008), and many others. The study by Galor and Mountford (2008) is an exception in that it considers two interacting countries (or regions). It investigates trade – but not knowledge exchange – and argues that the fact that countries are connected delays the take-off to growth of the initially backward country.

capital of its own workforce but also on the human capital of the countries with which it is connected. Figuratively speaking, countries need not to re-invent the wheel or the steam engine. Backward countries benefit from knowledge exchange with the technological leaders. Together with a network for knowledge exchange, this set-up provides a *minimum setup* in order to explain how a diversity of individual growth experiences evolves out of initial similarity between countries. In contrast to conventional unified growth theory, the model is too simple to explain how leader countries transition out of stagnation. Here, we simply give the leader countries a small head start that supports the initial growth process. We think that this approach is justifiable because of our focus on the differentiated take-offs and growth experiences across countries that happened *after* the take-off of the leader countries.

Our study was inspired by the work of Lucas (2000, 2009). Similar to our study, Lucas investigates the initial divergence and subsequent convergence of income across countries in a multi-country version of a simple economic model. In the 2000 paper, growth is exogenous and in the 2009 paper, growth is, as in our study, driven by human capital accumulation and access to the human capital of others. An important difference is that in Lucas' studies, countries have either full access or no access to world knowledge. There is no network of knowledge exchange and no evolving diversity of individual growth experiences. A stochastic mechanism determines when countries gain access to world knowledge. According to our approach, in contrast, the economic take-off and subsequent growth of the leaders, followers and trailers of the industrial revolution is endogenously explained and understood by the increasing diffusion of knowledge throughout the world. Growth of the leading country (the knowledge frontier) is exogenous in Lucas' world, whereas in our model, all growth is endogenous. This allows us to explain a richer set of phenomena. For example, according to Lucas' approach, the US would never had outpaced England, the industrial leader. In our setup, countries do not only temporarily diverge and converge but they also (occasionally) overtake each other. We also demonstrate that the network architecture itself has an impact on the success of industrialization and that, in this sense, the industrial revolution was not inevitable.

There also exists a relatively large literature on R&D spillovers between countries (e.g. Eaton and Kortum, 1999; Howitt, 2000). This literature usually involves a rather sophisticated modeling of households and firms but the way knowledge is exchanged between countries is straightforward and (most of) the analysis concerns the steady state. In this paper, the economic model is straightforward but the process of knowledge diffusion is non-trivial and the analysis focuses on transitional dynamics.[3] Another strand of literature investigates multi-country models in which convergence is driven by capital accumulation and trade (e.g. Acemoglu and Ventura, 2002). Conceptually, the available multi-country growth literature focuses on the question of whether and how countries at initially different income levels converge while we also investigate how countries that were initially similar diverged. In other words, as with the available multi-country growth literature, we also share an interest in the question of where the steady-state cross-country income distribution lies. Additionally, as with unified growth theory, we share an interest in the question of how the presently observable diversity of growth experiences evolved out of an initial similarity between countries. In a unifying framework, our network theory of knowledge diffusion offers an explanation for both "the great divergence" as well as "the great convergence".

There is a relatively small body of literature on networks in the context of economic growth. The most closely related work is perhaps the study by Cavalcanti and Giannitsarou (2012) who investigate learning externalities between households (or schools) in simple networks and focus on convergence behavior. Fogli and Veldkamp (2012) provide a study on the role of network connectivity for the diffusion of knowledge and diseases. Lindner and Strulik (2014) investigate how economic development is affected by globalization conceptualized as an evolving network, i.e. how decreasing local connectivity affects occupational choice and investment behavior through eroding trust and trustworthiness.

According to Jones and Romer (2010) the "old" stylized facts of growth, set up by Kaldor (1961), are now largely explained by neoclassical theory. Meanwhile, they argue that results from contemporary growth empirics established "new Kaldor facts" as a guideline and challenge for future developments in the theory of economic growth. We think that our theory addresses five of the six new facts: 1) an increasing flow of ideas through globalization, 2) accelerating growth rates, 3) cross-country variation of growth rates that increases with distance from the technology frontier, 4) large income and TFP differences across countries and 5) increases in human capital per worker throughout the world.[4] Since the underlying Romer (1986) model is – as a stand-alone unit – too simple

[3]Klenow and Rogriguez-Clare (2005) survey the literature on knowledge externalities in economic growth and propose some extensions. In particular, they consider treating knowledge diffusion as being country-pair specific and depending on distance (but they do not pursue this approach very far, cf. pp. 852-3). Recently, Comin et al. (2012) developed a micro-founded theory of spacial knowledge diffusion based on the random interaction of individuals. Their paper is also indirectly supportive of our approach by empirically showing that knowledge diffuses slower to countries farther away from the technological leaders.

[4]We cannot address Jones and Romer's (2010) sixth fact concerning the stable skill premium because countries are, for simplicity, assumed to be populated by representative households.

to allow for an explanation of (most of) the new facts, we conclude a non-negligible value added from the consideration of networks in the theory of economic growth.

The paper is organized as follows. The next section introduces the basic model. Section 3 then provides analytical results for comparative statics and comparative dynamics (steady-state, S-shaped transitions, overshooting growth of latecomers, rising and eventually declining world inequality). Proofs of the propositions are delegated to the Appendix. In Section 4 we investigate the implied growth dynamics for some very simple networks in order to provide a better understanding of the impact of the network architecture on knowledge diffusion. In Section 5, we introduce the Small World network (Watts and Strogatz, 1998) and investigate the distribution and growth of world income when countries are connected by such a network. We argue that the Small World network is already sufficiently complex to generate growth trajectories consistent with the new Kaldor facts. We provide a sensitivity analysis with respect to network parameters and initial conditions and discuss the phenomenon of endogenously changing world economic leaders and overtaking in general in the course of global development. Section 6 concludes.

## 2. The Model

2.1. **Households and Firms.** Consider a world consisting of a number $n$ of countries indexed by $i$. Countries share the same economic structure but differ potentially in their endowments and their connections to other countries. Each country is inhabited by two overlapping generations of measure one. A household consists of one adult and one child. The representative adult born at time $t-1$ in country $i$ maximizes the utility experienced from consumption above subsistence needs $c_{it}$ and from the future income of his or her child $y_{i,t+1}$. Assuming, for simplicity, a log form of the utility function, the objective is to maximize

$$u_{it} = \log c_{it} + \beta \log y_{i,t+1}. \tag{1}$$

for $y_{it} > \bar{c}$. The parameter $\beta$ denotes the weight of future income in utility. The household budget constraint is given by

$$y_{it} = \bar{c} + c_{it} + e_{it}, \tag{2}$$

in which $\bar{c}$ are subsistence needs and $e_{it}$ is expenditure for child education. Moreover, we require $c_{it}, e_{it} \geq 0$ and set $c_{it} = y_{it}$ and $u_{it} = -\infty$ for $y_{it} \leq \bar{c}$. The existence of subsistence needs motivates the fact that investment in education increases gradually with increasing family income. Education translates into human capital via a simple iso-elastic production function,

$$h_{i,t+1} = \max\left\{1, \;\; Be_{it}^{\eta}\right\}. \tag{3}$$

The assumption that the next generation has at least one unit of human capital (innate skills) ensures the long-run survival of mankind. It implies that the equation of motion exhibits a kink at unity. Such a kink could be motivated by assuming that $h_{it} = 1$ is associated with the skills to perform subsistence farming, i.e. skills which are obtained without cost by observing the parents working on the fields and which become obsolete when the individual gets an education and an occupation outside subsistence farming. The assumption allows us to capture the fact that all countries experienced a history in which mass education was not worthwhile. We assume $0 < \eta \leq 1$ but focus mainly on the case where $\eta = 1$ because it allows for positive growth at the steady-state. Our description of human capital accumulation is admittedly very stylized compared with its sophisticated modeling in unified growth theory (Galor and Weil, 2000; Galor, 2005) but it captures the same stylized facts: mass education becomes only worthwhile when technology has advanced sufficiently far and it then gradually increases with economic development. We introduce this "shortcut" modeling of intergenerational decision making because we want to focus on the phenomenon of knowledge diffusion around the world.

Countries are populated by a large number of competitive firms. The representative firm of country $i$ at time $t$ produces per capita output (GDP)

$$y_{it} = A_{it} h_{it}^{\alpha}. \tag{4}$$

Factor productivity $A_{it}$ is given for a single firm but endogenously determined through knowledge externalities from other firms. Because there are no other private inputs in production aside from human capital, the representative household earns income $y_{it}$. Inserting (2)–(4) into (1) and solving the problem provides education effort

$$e_{it} = \begin{cases} \frac{\alpha\beta\eta}{1+\alpha\beta\eta} \cdot (y_{it} - \bar{c}) & \text{for } \frac{\alpha\beta\eta}{1+\alpha\beta\eta} \cdot (y_{it} - \bar{c}) > (1/B)^{1/\eta} \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

Education effort increases in household income except for poor households, i.e. those that are close to subsistence and cannot afford education. For perpetually growing income, the income share spent on education converges towards a constant.

Inserting income (4) and education (5) into (3), we obtain the (gross) growth rate of human capital.

$$g_{it}^{h} \equiv \frac{h_{i,t+1}}{h_{it}} = \max\left\{\frac{1}{h_{it}},\, L \cdot \frac{(A_{it} h_{it}^{\alpha} - \bar{c})^{\eta}}{h_{it}}\right\}, \qquad L \equiv B\left(\frac{\alpha\beta\eta}{1+\alpha\beta\eta}\right)^{\eta}. \tag{6}$$

2.2. **Knowledge Externalities.** Following Arrow (1962) and Romer (1986) we assume that firm productivity is given by the knowledge that can be absorbed from knowledge created via production elsewhere. In contrast to Arrow and Romer, we assume that firms have access to knowledge created in other countries (for example, through the flow of goods and people). A link between two countries $i$ and $j$ thus means that these countries are open with respect to each other and that they are in mutual knowledge exchange.

The world network of knowledge exchange is represented by a matrix $\mathbf{W}$ whose elements indicate whether countries are linked with each other. For simplicity, we assume that links are unweighted and undirected. This means that the entry $w_{ij} = w_{ji}$ is equal to one if countries $i$ and $j$ are linked, and zero otherwise. The nodes to which country $i$ is linked are called neighbors of $i$. Assume country $i$ has $d_i$ links. By definition, each country is not linked to itself such that $d_i$ can assume any value between 0 (isolation) and $n-1$ (connected to all other countries). Let $N_i$ denote the neighborhood of $i$, i.e. the set of countries to which country $i$ is linked to.

Let $\epsilon$ denote the share of international knowledge spillovers, $\epsilon \in [0,1]$. Knowledge spillovers from abroad are derived from the *externality matrix* $\bar{\mathbf{W}}$, which is obtained by normalizing $\mathbf{W}$ such that for every linked country, $d_i > 0$, the sum of weights to neighbors in $\bar{\mathbf{W}}$ is equal to $\epsilon$, that is $\bar{w}_{ij} = \epsilon/d_i$ for $j \in N_i$. In case of isolation, $d_i = 0$ and we set $\bar{w}_{ij} = 0$ for all $j \neq i$. Finally, we assign $\bar{w}_{ii} = 1-\epsilon$ for all $i$. Hence, all rows of $\bar{\mathbf{W}}$ have positive elements and sum up to one if every country has at least one link.

We define the knowledge externality for country $i$ as the average human capital input in production of all neighbors, including the country itself.

$$A_{it} = \left(\sum_{j=1}^{n} \bar{w}_{ij} h_{jt}\right)^{1-\alpha} = \left(\epsilon\, \bar{h}_{it} + (1-\epsilon) h_{it}\right)^{1-\alpha}, \qquad (7)$$

in which the average human capital of the neighbors of $i$ is denoted by $\bar{h}_{it} = \frac{1}{d_i}\sum_{j \in N_i} h_{jt}$. For $\epsilon = 0$, the model collapses to the Romer (1986) model, in which countries are treated as if in isolation.[5]

The intuition behind the use of country averages is that at any time increment any person in country $i$ can exchange knowledge either with a person in country $j$ or country $k$. The fact that aggregate time for knowledge exchange per country is normalized to unity then implies that the total knowledge acquired is given by the average of the human capital of its neighbors. Ceteris paribus, a link to a backward country (with $h_{jt} < h_{it}$) leads to a lower knowledge externality for country $i$ and a link to a forward country (with $h_{jt} > h_{it}$) implies a higher knowledge externality. This means that initially backward countries that are well connected to initially rich countries have an advantage in learning from abroad whereas initially rich countries that are mainly connected with poor countries experience a drawback in knowledge accumulation because there is relatively little to be learned from abroad.

[5]In the Appendix we also discuss the case of $A_{it} = \left(\sum_{j=1}^{n} \bar{w}_{ij} h_{jt}\right)^{\omega}$, $\omega < 1-\alpha$. In this case, knowledge externalities are too weak for positive growth to be sustainable at the steady-state. Nevertheless, the calibrated model can produce plausible adjustment dynamics vis-a-vis the historical trends. In this sense, the "knife-edge" assumption $\omega = 1-\alpha$ is not crucial for our main results.

## 3. Long Run Dynamics

From (6) and (7) we get a description of the world as a vector-valued difference equation,

$$h_{it+1} \quad = \quad \max\left\{1,\ L\cdot\left(\left[\epsilon\,\bar{h}_{it}+(1-\epsilon)h_{it}\right]^{1-\alpha}h_{it}^{\alpha}-\bar{c}\right)^{\eta}\right\}, \tag{8}$$

$i = 1, \ldots, n$, and a network $\bar{\mathbf{W}}$. Note that in case of isolation, $\bar{h}_{it} = 0$ which is always harmful to growth.

Let $H_t \equiv \sum_{i=1}^{n} h_{it}$ and $\bar{h}_t \equiv H_t/n$ denote the world-wide stock of human capital and the average human capital endowment per person, respectively. Furthermore, let $g_t^h \equiv H_{t+1}/H_t$ denote the growth rate of world human capital and $x_{it} = h_{it}/\bar{h}_t$ country $i$'s relative endowment, i.e. the ratio between country $i$'s human capital and average human capital. Whether economies are able to have positive growth in the long run is determined by $\eta$ and $L$.

Proposition 1. *There is no long run growth, i.e.* $\lim_{t\to\infty} g_{it}^h = 0$ *for all* $i = 1, \ldots n$*, for* $\eta < 1$ *and for* $\eta = 1$ *if* $L \leq 1$.

Definition 1. *A steady state is defined by each country growing at the same constant growth rate, i.e.* $g_{it} = g$ *such that the relative human capital stays the same for each country, i.e.* $x_{it+1} = x_i$.

Proposition 2. *When all countries start with equal initial human capital* $h_0$ *and there are no isolated countries, the network is irrelevant and the economy is always at the steady-state where* $x_{it+1} = x_i = 1$ *for all* $i \in N$ *and* $t = 1, 2, \ldots$.

This means that the Romer (1986) model is included as a special case when the world consists of identical countries. The next proposition rules out unequal equilibria for $\eta < 1$ if the network is connected.[6]

Proposition 3. *Consider the world as a connected network* $\bar{\mathbf{W}}$. *For* $\eta < 1$*, human capital and income of all countries eventually converge to the same level. There are at most two different steady states above subsistence level. If* $\eta = 1$ *and* $L \leq 1$*, the world eventually converges to subsistence level* $\bar{c}$.

So far, results have just confirmed for the multi-country world what we already know from the closed country Romer model. From now on, however, results deviate.

[6] A network is connected if there exists a path between any two nodes. Formally, this means that for any $i$ and $j$, $i \neq j$, there exists a $k \geq 1$ such that $\left(\bar{\mathbf{W}}^k\right)_{ij} > 0$.

Proposition 4 (Long-Run Growth). *For $\eta = 1$ and $L > 1$, the economy has the potential for positive long-run growth at a uniquely determined rate $g^* = L$. If the network $\bar{\mathbf{W}}$ is connected, the world economy converges towards a steady state (of growth or stagnation).*

Notice the difference to Romer (1986). There, $L > 1$ and linear factor accumulation, i.e. $\eta = 1$ (in our notation), inevitably produce long-run growth. Here, these conditions make long-run growth *feasible*. In fact, if the network architecture is unfavorable, long-run growth will not occur. We elaborate on this phenomenon in Section 5.2. If the network is connected, all knowledge is eventually shared by all countries. This feature implies that the world economy converges towards a steady state, i.e. a situation in which all countries grow at a common rate. A steady-state of positive growth means that all countries of the world irrespective of their backward initial situation are eventually "infected" by knowledge diffusion and will grow eventually at the same rate as the leaders of the industrial revolution.

The shape of the path of a country's economic development is crucially determined by the ratio between the externality and its own human capital. In order to see this, we drop the indices of (8) and define the externality ratio $\xi = \bar{h}/h$ such that the growth rate for $\eta = 1$ is given by

$$g = L \cdot [\epsilon\xi + (1-\epsilon)]^{1-\alpha} - L\frac{\bar{c}}{h}. \tag{9}$$

The growth rate is determined by two terms. The second term reduces the growth rate and the term is large for poor countries (i.e. when $h$ is low) and vanishes for rich countries. Intuitively, education expenditure in poor countries is relatively low because subsistence needs are relatively important. The first term depends on the externality ratio $\xi$. For an initially poor country, we have $\xi > 1$ because the country is poorer than the average of its neighborhood. The fact that $\xi > 1$ is favorable for growth does not necessarily imply that poor countries catch up quickly. Firstly, as stated above, the second term is large for the poor country, especially immediately after take-off. Secondly, the average neighborhood may actually grow at a higher rate because it is favorably linked to rich countries implying an increasing $\xi$. Altogether, this means that growth is increasing during the early phase after the take-off. As the country gets richer, the second term vanishes and $\xi$ will eventually decline to unity as the country converges towards the steady state. A typical S-shaped growth rate is therefore driven by an increasing $\xi$ in the beginning. For the moment, we assume that $\xi$ increases monotonically to a maximum and then decreases (thereby excluding oscillating behavior). For all networks investigated below, it turns out that $\xi$ behaves in this way.

Proposition 5 (S-shaped Transitions). *Suppose a steady-state of long-run growth $g^*$ exists. Let the externality ratio $\xi(h)$ of a country be twice continuously differentiable,* *concave and increasing in $h$ to a maximum. Then there exists a $T \in \{0, 1, \ldots, \infty\}$ such that the country's growth rate changes from accelerating to decelerating.*

Recall that both neoclassical growth theory and Romer's (1986) endogenous growth theory fail to predict the gradual take-off to growth (the new Kaldor fact number 2). The neoclassical model predicts that growth is highest for the poorest countries, i.e. for the countries farthest away from the steady-state. This counterfactual feature follows immediately from decreasing returns of factor accumulation. According to the linear Romer (1986) model there are no transitional dynamics. The original Romer (1986) article – in contrast to its textbook presentations – also discussed the case of accelerating growth rates by assuming increasing returns to scale. This mechanism, however, leads to an explosion of the economy without further assumption that stops growth rates from growing. Growth is not S-shaped. The present model, in line with the historical observation, generates S-shaped transitions. The growth rate of GDP per capita accelerates gradually during the first phase after take-off. During the second phase growth decelerates, a phenomenon which renders convergence towards the steady-state.

Proposition 6 (Overshooting Growth). *Suppose that $\bar{\mathbf{W}}$ is connected and a steady state of long-run growth $g^* > 0$ exists. A) Forerunners of the industrial revolution converge monotonously towards $g^*$. B) Followers of the industrial revolution converge non-monotonously at growth rates that are temporarily above $g^*$ if their initial lag of human capital with respect to the neighborhood average is sufficiently large.*

Intuitively, overshooting growth occurs when initially backward countries start investing in education and there is a lot to learn from the rest of the world or, more precisely, from the countries to which a link of mutual knowledge exchange exists (i.e. the neighbors). The opportunity to tap into a greater pool of knowledge creates an advantage of backwardness (Gerschenkron, 1962). In line with the historical evidence the initially backward countries manage to double their income per capita in a much shorter amount of time than the leaders of the industrial revolution (Parente and Prescott, 1994, 2005).[7]

In order to develop a comprehensive picture of the evolution of world income inequality, we distinguish between relative and absolute income inequality. To see the difference, consider a world of 2 countries with endowment $(y_1, y_2) = (10, 40)$. Assume the endowment changes to $(20, 80)$. This means that the absolute gap changes from 30 to 60, while the relative difference 30/50 stays constant. Relative income inequality can be expressed by the Gini index (or the Theil index), whereas absolute income inequality can be measured

[7]Parente and Prescott (1994, 2005) offer an alternative explanation of the phenomenon. According to their theory world knowledge grows at an exogenously given rate and adoption of world knowledge is easier for a country when it is farther away from world knowledge because barriers to adoption are smaller.

by the so-called *absolute* Gini index, defined as the product of the Gini index and average income (Chakravarty, 1988). In the present example of 2 countries the Gini index is 0.3 for both distributions but the absolute Gini index changes from 7.5 to 15. In order to operationalize these ideas, let $D_t = \max_{i,j}(y_{it} - y_{jt})$ denote the absolute income gap between the richest and poorest country and $Y = \sum_{i=1}^{n} y_i$. Let the relative gap be defined by $d_t = D_t/Y_t$. The following Lemma then summarizes the main properties of these measures of inequality.

LEMMA 1.
*(i) The Gini index stays constant if income grows at the same rate for all countries. It tends to zero for $t \to \infty$ if and only if $d_t \to 0$.*
*(ii) The absolute Gini index stays constant if income in every country increases by the same absolute amount. It tends to zero if and only if $D_t \to 0$.*
*(iii) If the absolute Gini index tends to zero, then the Gini index tends to zero as well, but not vice versa.*

Statement (iii) follows immediately from the fact that $d_t \to 0$ occurs if the gap between rich and poor countries $D_t$ increases at a lower rate than growth of $Y_t$.

PROPOSITION 7 (The World Kuznets Curve). *Suppose $\bar{\mathbf{W}}$ is connected and that a steady state of growth $g^* > 0$ exists. A) Relative income inequality between countries eventually vanishes. The Gini index converges to zero. B) If some countries initially grow and others stagnate, then relative income inequality increases initially and declines subsequently.*

PROPOSITION 8. *[No Convergence in Levels] Suppose $\bar{\mathbf{W}}$ is connected and that a steady-state of growth $g^* > 0$ exists. Despite eventually declining relative world inequality there is not necessarily convergence of income levels.*

The model produces not only a "great divergence" (Pommeranz, 2000), initiated by the take-off of the leaders of industrial revolution, but also a "great convergence" in terms of relative income levels. Convergence occurs after the take-off of the latecomers of the industrial revolution. The latecomers are identified as the countries with inferior initial endowments *and* missing links to the forerunners of the industrial revolution. Recall that a connected network ensures the existence of a steady-state. This means that eventually all knowledge is shared between all countries, which explains the phenomenon of vanishing relative income inequality. This result is in disagreement with some popular articles on the world income distribution (Jones, 1997; Acemoglu and Ventura, 2002) but it is in line with Lucas (2000, 2009) vision of the world's future development. Notice that the prediction that relative income inequality vanishes eventually does not imply that absolute income levels converge. In fact, as we show later, countries may even overtake each other (several times) and the absolute income gap may increase while income inequality measured by the (relative) Gini or Theil index disappears. In order to investigate this phenomenon and other interesting features, we next turn to a numerical presentation of the model.

## 4. Adjustment Dynamics: The Great Divergence And The Great Convergence

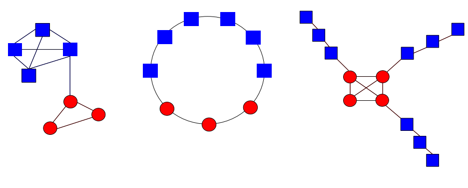
We assume that initially, there are at most two distinct groups of countries. A small group of countries with relatively high human capital endowment (the rich) and a large group with relatively low endowment (the poor). Initial endowments are such that rich countries are growing, albeit at a very low rate, while poor countries are stagnating because aggregate productivity is so low that investment in human capital is not worthwhile. This setup is the most interesting case because it allows for evolving country heterogeneity. As time proceeds and knowledge crosses borders, income and productivity of the countries grow differently according to their connections with other countries and countries become more dissimilar with respect to economic growth. Having two different groups of countries is the *minimum setup* to discuss evolving heterogeneity (cf. Proposition 2).

### 4.1. Stylized Networks.

We first investigate adjustment dynamics for some particularly simple examples of the externality matrix $\bar{\mathbf{W}}$. This allows us to provide an understanding of the main mechanism behind the international flow of knowledge and world income dynamics. Suppose the world network is given alternatively by a stylized network from the set of networks depicted in Figure 2. Rich countries, i.e. countries in which the population is better educated, are represented by red circles and poor countries are represented by blue squares.

A bridge network is partitioned into two components. The rich and the poor are each internally representing a complete network. The two components share exactly one link, the bridge. The bridge network could be understood as a metaphor for a world of different continents connected by a minimum of links.

Figure 2: Stylized Networks



Stylized networks: Bridge (left), Ring (middle), and Core-Periphery (right). Rich countries are represented by red circles and poor countries are represented by blue squares.

A ring network is obtained by positioning each country along a line, ordered by country-specific human capital. In order to establish symmetric architecture, the line is closed to form a circle. Each country is connected to its $k$ nearest neighbors (not counting itself as a neighbor). This means that there are $2k$ poor countries connected with rich countries. In the example we have $k = 1$. The ring network emphasizes the role of geographic proximity for knowledge exchange. The world is "round" and countries are directly connected only with their geographical neighbors.

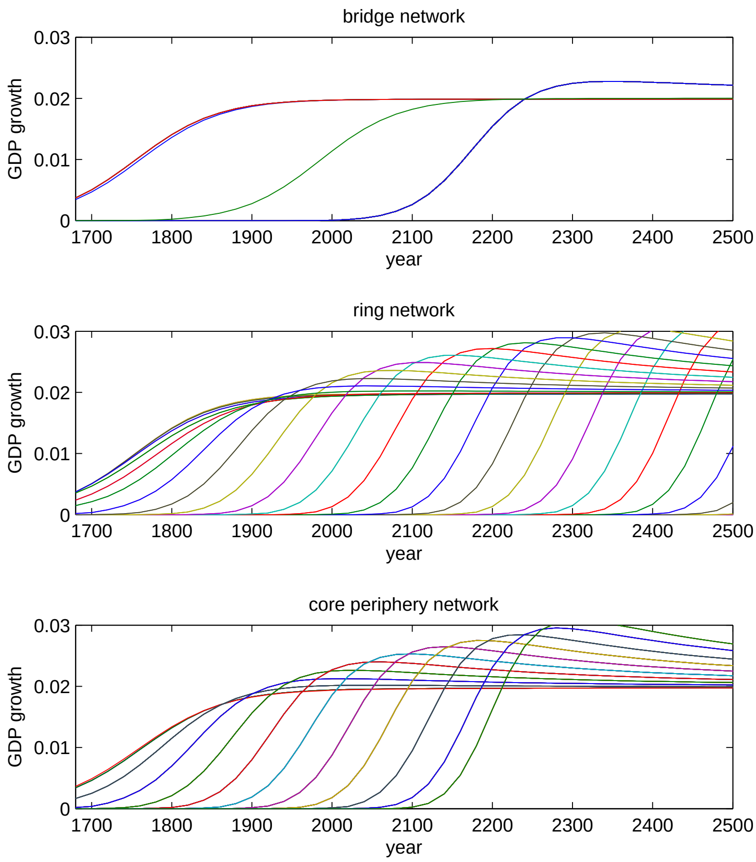
Finally we consider the core-periphery network. Here, the core consisting of initially rich countries forms a complete network to which a number of peripheries consisting of initially poor countries are connected. The poor countries are connected in series implying that there is one bridge per periphery, linking it with the core. The core-periphery network describes a situation in which a subset of rich countries is already fully integrated and another subset of poor countries (the colonies) is less well integrated.

4.2. **Numerical Specification.** We begin with a benchmark specification of the model. Later on, we discuss the sensitivity of results on parameter choice. Suppose the world consists of 100 countries of which 10 percent are initially rich, i.e. better endowed with human capital. We do not ask where the initial differences between countries come from but assume, in line with the historical evidence on economic conditions in pre-modern times, that the initial differences are small. The challenge is thus to explain how a great variety of growth performances evolves out of small initial differences.

All countries share the same parameters values. We set $\eta = 1$ in order to allow for long-run growth. We set $\alpha = 0.5$, such that at the steady state, productivity growth explains half of GDP per capita (approximating the findings by e.g. Barro and Sala-i-Martin, 2004 and Easterly and Levine, 2001). We set the utility weight for children's income to $\beta = 0.23$ such that at the steady state, 7 percent of income are spent on education, as observed approximately for most fully developed countries (Eurostat, 2012). We then adjust the value of $B$ such that the implied steady-state growth rate is 2 percent annually. The parameter values of $\epsilon$ and $\bar{c}$ are irrelevant for the steady-state but shaping adjustment dynamics. Eaton and Kortum (1999) estimate, for a sample of fully developed countries, that between one half and three-fourths of the knowledge adopted has been generated abroad. We take the benchmark value for our (temporarily) more heterogenous set of countries from the lower bound of their estimates and set $\epsilon = 0.5$. This means that one half of the knowledge available in a country has been generated by domestic firms and the other half stems from international knowledge diffusion. We set $\bar{c} = 0.33$ in order to get the best fit of economic growth of the leader country with the historical data.

After running the simulation we convert the generational data into annual data assuming that a generation takes 20 years. We convert generational growth rates into annual ones for better comparability with real data. We set initial time to the year 1700, i.e. shortly before the onset of the first industrial revolution. The poor countries are endowed with minimum human capital, $h_{poor,0} = 1$. The rich countries are assumed to be 20 percent better endowed $h_{rich,0} = 1.2$. This implies that – depending on the respective network – income in the rich countries is initially 1.23 to 1.3 times higher than in poor countries. These values correspond well with the estimates of the head start of Western European countries vis-a-vis the rest of the world at the dawn of the first industrial revolution (Bairoch, 1998, Ch. 9). Most importantly, this specification means that poor countries initially stagnate while rich countries initially grow at a low rate of around 0.3 percent.

Figure 3: Evolution of Economic Growth in 100 Countries when the World is a...




For all three " worlds": 10% of countries initially better endowed with human capital (in 1680). $h_0 = 1$ for poor countries, $h_0 = 1.2$ for rich countries. Parameters: $\alpha = 0.33$, $\beta = 0.4$, $\eta = 1$, $\bar{c} = 0.33$, $\epsilon = 0.5$. Middle panel: 2 neighbors per node. Bottom panel: 9 peripheries a 10 countries.

Figure 3 shows the evolution of growth predicted by the numerical experiments. The upper panel assumes that the world network is a bridge. Knowledge diffusion through the network generates four visibly distinct adjustment trajectories. Naturally, the rich countries take off first. The rich country linked directly to the poor world takes off a bit later because there is less to learn from the poor neighbor. In contrast, the poor country equipped with a direct link to the rich world experiences a huge advantage vis-a-vis its poor neighbors and takes off about two centuries earlier, fueled by knowledge diffusion from its rich neighbor. The remaining club of less developed countries takes off late but experiences an “advantage of backwardness” (Gerschenkron, 1962) in the sense that their income growth surpasses the income growth of the forerunners of the industrial revolution.

The fact that growth rates of latecomers overshoot the balanced growth rate means that relative income inequality declines eventually. The explanation is that latecomers, once growth is initiated, tap into a greater reservoir of world knowledge. This knowledge has been accumulated in the recent past and was not yet available when the forerunners took off. This phenomenon relates the model to the new Kaldor fact no. 1: the increasing flow of ideas via globalization. Globalization here means that an increasing share of countries gets out of stagnation with education becoming worthwhile and that the educated workforce benefits from and contributes to an increasing stock of knowledge diffusing through the network.

The bridge network already displays one important phenomenon of growth in networks, the overshooting growth of latecomers, but it generates insufficient variety of economic performance across countries. This is different for the ring network, as evidenced in the center panel of Figure 3. The initially rich countries are again experiencing a very similar take-off to growth, in which the countries surrounded by other rich countries perform only slightly better than those at the border to the poor world. The poor countries, on the other hand, experience a very varied take-off. The reason is that new knowledge is “handed over” along the circle. The two countries neighboring the rich take off first among the poor, then the countries next to these countries follow, etc. There is also more variety in growth rates. Generally, we observe that overshooting growth is higher, the later the take-off time is. This is the case because there is more to learn from the neighbors once education becomes worthwhile for the latecomers. Consequently, growth during the early take-off phase is much faster for latecomers. While it took about 200 years for the forerunners of the industrial revolution to reach a growth rate of 2 percent, the countries taking off in the 1950s needed only about two generations to achieve the same rate of growth.

Compared with other networks, the circle predicts a very long period of take-offs, implying a very long period of increasing world inequality. The reason is that it takes time until knowledge is passed on along the circle from neighbor to neighbor toward the most unfortunate country “at the other side of the world”. Moreover, the take-offs are “too predictable”. Their sequence follows the position of countries on the circle.

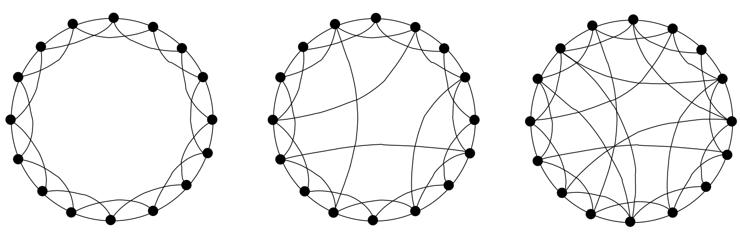
The core-periphery network, shown in the bottom panel of Figure 3, eliminates some of the flaws of the two previous networks. It produces a variety of growth experiences, largely overshooting growth rates, and a reasonable duration of the “era of take-offs to growth” from 1700 to the mid 21st century. Yet the growth experience of countries is still too easily predicted. The countries next to the bridges to the core take off just after the initially rich and then we observe departures from stagnation according to the order of countries along the peripheries. Altogether, we observe “only” 10 different growth paths, one for the core countries and one for each position on the periphery. There is still too little heterogeneity in the world. Moreover, the connectivity between the initially rich countries is “too high” in all three simple networks. This is evident from the result that the take-off of the forerunners of the industrial revolution happens too fast in all three panels of Figure 3. By the year 1800, the forerunners of the industrial revolution are counterfactually predicted to grow already at a rate of 1.5 percent annually.

## 5. Knowledge Diffusion and Income Evolution in A Small World

5.1. **Model Setup.** The small world model (Watts and Strogatz, 1998) is a device to investigate an irregular network that features both local connectivity and long-distance links between nodes in a simple way. Mathematically, it is easily understood but complex enough to allow for an application to a plethora of biological and social phenomena (see Newman, 2003, for an overview). The small world model appears to be particularly suited for our purpose because it retains the importance of local connectivity, capturing the fact that most knowledge diffuses from direct neighbors, but at the same time allows for the establishment of long-distance links between distant countries.

Here, we consider a modification of the Watts and Strogatz model, developed by Newman and Watts (1999), which appears to be more appropriate for our purpose. The idea of the Small World model can be illustrated best by considering a network on a one-dimensional lattice. It is constructed from a regular network in which any node (country) is connected with its direct neighbors that are $m$ or fewer lattice spaces away. In the example of Figure 4, $m = 2$. Each country is connected to 4 neighbors, 2 at each side. The regular network is then modified by randomly adding long-distance links. The probability for a long-distance link per link of the underlying lattice is denoted by $p$. The middle panel of Figure 4 shows an example for which $p$ is low and the panel on the right shows an example for larger $p$.

Figure 4: Small World Network



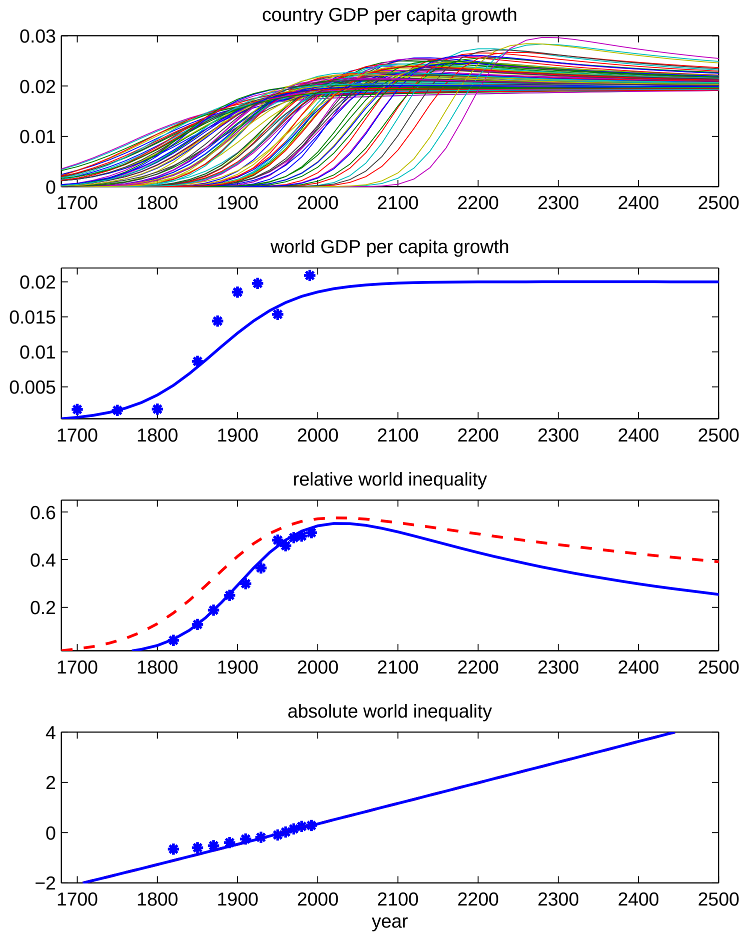
For international knowledge flows, the feature of local connectivity, created through positioning the countries on a ring, captures the empirical fact that knowledge spillovers, in principle, decline with geographic distance (e.g. Keller, 2002). The presence of long-distance links means that this generality is occasionally broken and that the effective distance is (much) shorter than geographic distance. Figuratively speaking we could imagine the US to be geographically only two neighbors away from Guatemala but exchanging much more knowledge with England because both countries are connected with a long-distance link. This may turn out to be crucial for comparative development because the US benefits directly from knowledge created in England while Guatemala benefits only indirectly via the US. Moreover, in order for the knowledge to arrive in Guatemala it has to cross Mexico, another initially less educated country, such that a part of the knowledge created in England gets "lost in transition".

5.2. **Results for the Benchmark Model.** As a benchmark we take the specification of the economic model from above and consider $p = 0.3$, i.e. the case in which 30 percent of the countries are equipped with a long-distance link. We subsequently provide sensitivity analysis with respect to $p$ and other important parameters. Figure 5 shows the implied adjustment dynamics for the benchmark case. In contrast to the simple networks discussed above, the small world generates a lot of heterogeneity. Basically, each of the 100 countries follows its own idiosyncratic growth trajectory. Recalling that initially, in the year 1700, there were only two different types of countries and that the initial difference between rich and poor countries was small (1:1.2), we conclude that diversity evolves out of similarity.

Comparing the model prediction with the historical facts (Bairoch, 1993; Galor, 2005), we would imagine the group of initially rich countries as Western Europe, which reaches on average a growth rate of 1 percent from 1820-1870 also coinciding with the phase in which some of the Latin American countries started to grow. In the 20th century, when the latecomers take off, the initially rich countries grow at an almost constant rate of 2 percent annually. It is also interesting to observe that growth of the leaders is already surpassed by growth of some followers in 19th century and that despite the presence of long-distance links, some countries are predicted to take off very late in the mid-21st century. The differentiated and relatively rapid take-offs of the latecomers of the industrial revolution in the 20th century produce the picture of a great variety of subsequent growth experiences of countries that were almost equally poor just a generation ago.

The second panel of Figure 5 shows the implied average economic growth in the world. Dots represent the data points from De Long (1998) shown in Figure 1. The model predicts the take-off of aggregate world growth reasonably well. World growth rises from almost zero to just below 1 percent in the mid-19th century and to about 1.5 percent in the mid-20th century. Compared to the data, the take-off is somewhat too slow, an

Figure 5: Economic Growth in a Small World




Dots in the second panel: world GDP growth according to De Long (1998); Third panel: solid line: Theil-index; dashed line: Gini index; dots: Theil index from Bourguignon and Morrison (2002); Fourth Panel: log of absolute absolute Gini index; dots: absolute Gini from Atkinson and Brandolini (2010). Sample size: 100 Countries. Parameters as for Figure 3; small world based on ring with 1 neighbor to left- and right-hand side; long-distance links added with probability $p = 0.3$.

outcome that could be corrected (by assuming a higher $p$ or $\epsilon$) at the expense of predicting a take-off that is "too early" for the latecomers. Altogether, however, the model generates plausible S-shaped transitions. On the individual level, as well as on the global level, the model provides an explanation for the new Kaldor fact no. 2, the gradual increase of the rate of economic growth.

The differentiated take-off of countries produces the Great Divergence: world inequality increases strongly from 1800 to 2000. This is shown in the third panel of Figure 5, in which dots represent the data points from Figure 1 (Bourguignon and Morrison, 2002). The solid and dashed line, respectively, show the model's prediction for the evolution of the Gini index and the Theil index, computed from the individual income trajectories of the 100 countries. According to the model, for its benchmark calibration, inequality stops growing in the late 20th century. From the early 21st century onwards, the model predicts a "great convergence". As more and more latecomers catch up with overshooting growth rates, *relative* world inequality declines. The inequality curve, however, is skewed. The great convergence is predicted to take several centuries longer than the great divergence. The intuition is straightforward. The fact that the original leaders of the industrial revolution keep growing makes the catch up harder than the quick departure of the leaders from the almost stagnant income of the followers and latecomers two centuries earlier.

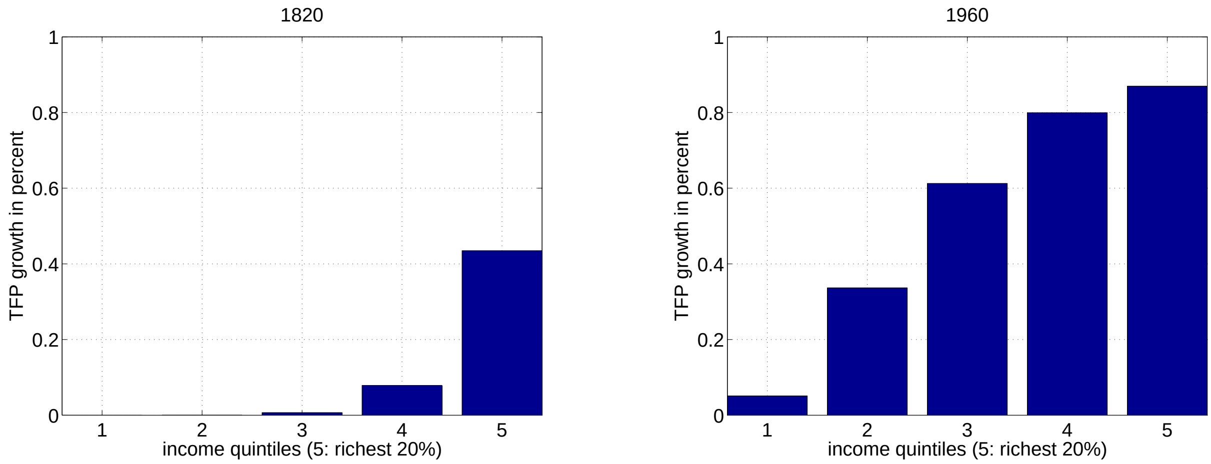
The focus on the conventional Gini index, however, conceals that absolute world inequality keeps on rising. The bottom panel of Figure 5 shows the absolute Gini, i.e. the relative index from the third panel multiplied by mean income. The log-scaling means that absolute inequality grows exponentially. These findings illustrate Proposition 8 and highlights the importance of distinguishing relative and absolute convergence. The relative income gap between rich and poor tends to zero because the absolute gap grows slower than the total level of income (cf. Lemma 1). Dots in the bottom panel show the absolute Gini index computed for the Bourguignon and Morrison (2002) data by Atkinson and Brandolini (2010). The network model somewhat underestimates absolute inequality at the dawn of industrialization but gets the exponential increase over the 20th century about right. It predicts this trend to continue in the future.

5.3. **The Evolution of TFP across Time and Space.** With rising income, households spend more on education, in line with the new Kaldor fact no. 5. The share of income spent on education converges gradually towards 7 percent of GDP. Increasing education at home as well as abroad means that aggregate productivity (TFP) also increases. TFP of country $i$ at time $t$ is given by $A_{it}$. In order to explore the evolution of TFP growth, we sorted the countries for any time $t$ into income quintiles, with the poorest 20 percent of countries in the first quintile and the richest 20 percent in the 5th quintile. We then computed the average growth rate of TFP for each quintile.

Figure 6 shows excerpts of the computation for two specific years, 1820 and 1960. In line with the historical evidence, TFP growth, compared to contemporary rates, was low in the 18th century, even for the leaders of the second industrial revolution (Crafts, 2003). According to the model, TFP growth was about 0.8 percent for the richest countries and about 0.2. percent in the second richest income group. That TFP growth is predicted to be lower in poorer countries comes as no surprise. It is perhaps more interesting to observe that the growth of TFP increases over time across all income groups. This is the case because at earlier times in history, more countries had neighbors who were not yet well educated and produced hardly new knowledge. This means that average knowledge was low, and that on average, comparatively little new knowledge diffused internationally.

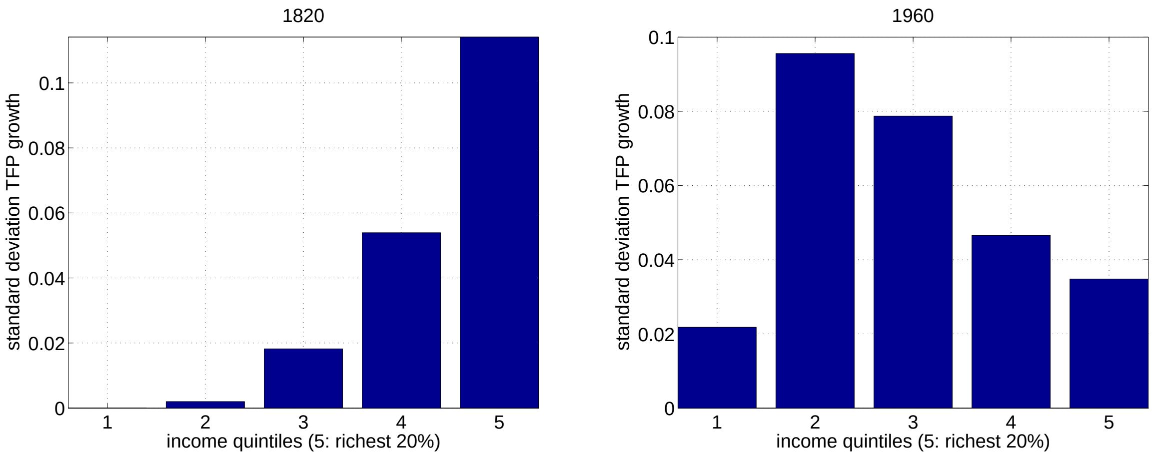
In general, as evidenced by Figure 5 and 6 taken together, the model captures the new Kaldor fact no. 4, the large income and TFP differences across countries.

Figure 6: TFP Growth Across Countries



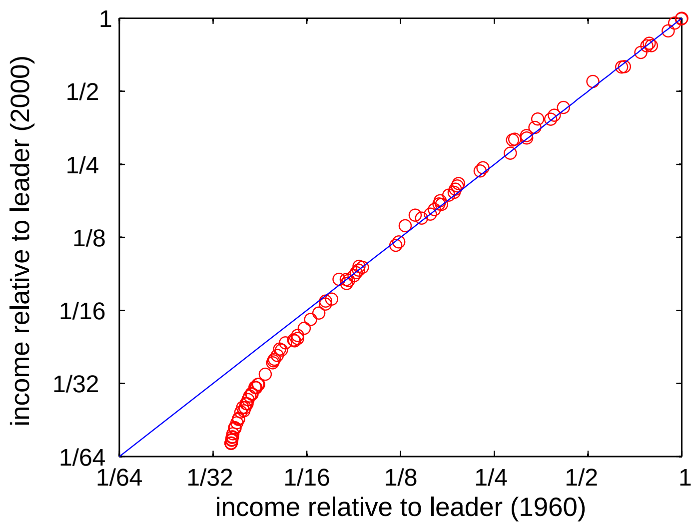

The variability of growth rates is shown in Figure 7, again exemplarily for 1820 and 1960. The figure shows the standard deviation (in percent) of TFP growth for each income quintile. The panel for the year 1960 corresponds with the new Kaldor fact no. 3, stating that the variance of growth rates across countries increases with distance to the technological frontier. Since there has been only few take-offs in the bottom quintile in the 1960s, the highest standard deviation is observed for the second poorest income group. But otherwise, the variance is subsequently lower among the richer groups. The figure also suggests that fact no. 3 is a phenomenon of the 20th century. In the 19th century, when the frontier countries themselves sequentially experienced their take-offs to growth, while the rest of the world was still close to subsistence, the variance of growth rates was highest among the richest countries.

Figure 7: Standard Deviation of TFP Growth Across Countries



5.4. **Is the Present World Income Distribution Close to Its Steady State?** As evidenced in Figure 5, the model predicts that relative income inequality across the world will eventually decline after the take-off of the latecomers of industrialization. In contrast to such an optimistic outlook, some related studies developed theories in order to explain a *constant* world income distribution at a state of high inequality, most notably perhaps the study of Acemoglu and Ventura (2002). Acemoglua and Ventura's work was inspired by the observation of "a relatively stable" world income distribution in the second half of the 20th century.

Figure 8: Income Relative to Leader: 1960 vs. 2000




Country GDP relative to leader country benchmark model; loglog scale.

A "relatively stable" distribution, however, could also be inferred from an actually slowly evolving distribution. This is particularly the case if the window of observation is relatively short and if the observation happens to be taken at a period of time when world inequality is flat because it is close to its maximum. In order to verify this claim by way of example, we compute for the outcome from the benchmark economy a relative income plot similar to the one displayed in Jones (1997, Figure 2) and Acemoglu and Ventura (2002, Figure 1). Specifically, we compute from the time series shown in Figure 5 for all countries, the relative income with respect to the leader country in the year 1960 and in the year 2000 and plot the result on a loglog scale, as shown in Figure 8. In accordance with the earlier studies, we observe little deviation from the 45 degree line. Relative income in 1960 is a good predictor of relative income in 2000. Confronted with this picture alone, one could indeed be tempted to conclude that the world income distribution is basically constant. In fact, however, we know from Proposition 7 that income relative to the leader country moves to unity for all countries as time goes to infinity. This convergence process, however, is very slow and presumably not discernable within a 40 year time window. The observation of an almost stable distribution of high inequality is consistent with a moving distribution toward equality.

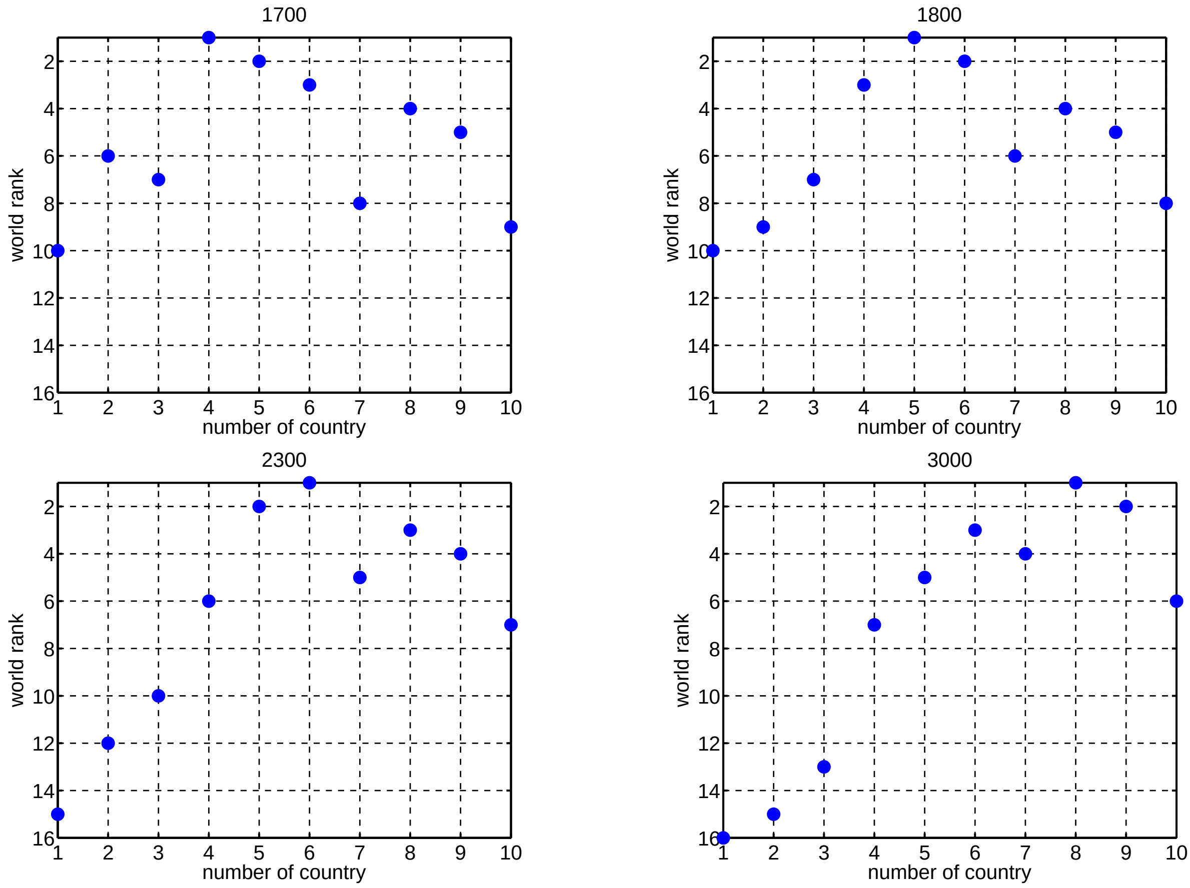
5.5. **Overtaking and Falling Behind.** The phenomenon of income convergence is at the center of modern growth economics. The phenomenon of overtaking, however, is rarely investigated in the context of an economic growth model.[8] Since overtaking and the falling behind of countries is observed in the real world, it is challenging to model this behavior while not relying on stochastic (i.e. unexplained) fluctuation and not giving up global stability, i.e. convergence to a global steady-state with vanishing income inequality. To demonstrate this extraordinary behavior, we perform the following numerical experiment. We follow the 10 initially richest countries, named $1, 2, \ldots, 10$, along the way towards the steady state and visualize their relative position in the world income ranking. Figure 9 shows the resulting "income ladders" for four different years. For example, a dot at the (1,10) position in the 1800 diagram means that country 1 was ranked 10th place in the year 1800.

In the numerical experiment, country 4 leads the world income ranking in the year 1700. Obviously, it was favorably connected with other rich countries. By the year 1800, country 4 gave up the lead to country 5, which presumably benefitted from favorable connections with quick followers of the industrial revolution. We also observe that country 2 and 3 fall behind whereas country 7 advances by two positions. In 2300, country 6 is at the top and the original number one dropped to the sixth place while country 1 and 2 fell out of the top 10 altogether. And in the year 3000, we observe yet another leader, country 8, as well as a new number 2, country 9, while country 3 also left the top 10. These changes in rank are explained by the changing advantage of links as knowledge is accumulated and diffused through the network. For example, an initially rich country connected only to another initially rich country, which in turn is connected only to latecomers of the industrial revolution, grows initially fast and then slows down. It is potentially overtaken by a country that is connected with initially poor countries, which are, however, well connected and "infected" by the growing knowledge of their neighbors at an early stage of the diffusion process.

In order to develop an intuition for these results, consider a "network" of two countries, one with initial human capital level $h$, the other with initial human capital level $h + \Delta$. Neglecting the corner solution, the equation of motion (8) for the first country is given by $f_1(h, h+\Delta) = L\left\{\left[\epsilon(h+\Delta) + (1-\epsilon)h\right]^{1-\alpha} x^{\alpha} - \bar{c}\right\}$. For the second country, it is given by

[8] Some researchers modeled overtaking in a purely stochastic context of Markov chains of income distributions, see e.g. Jones (1997).

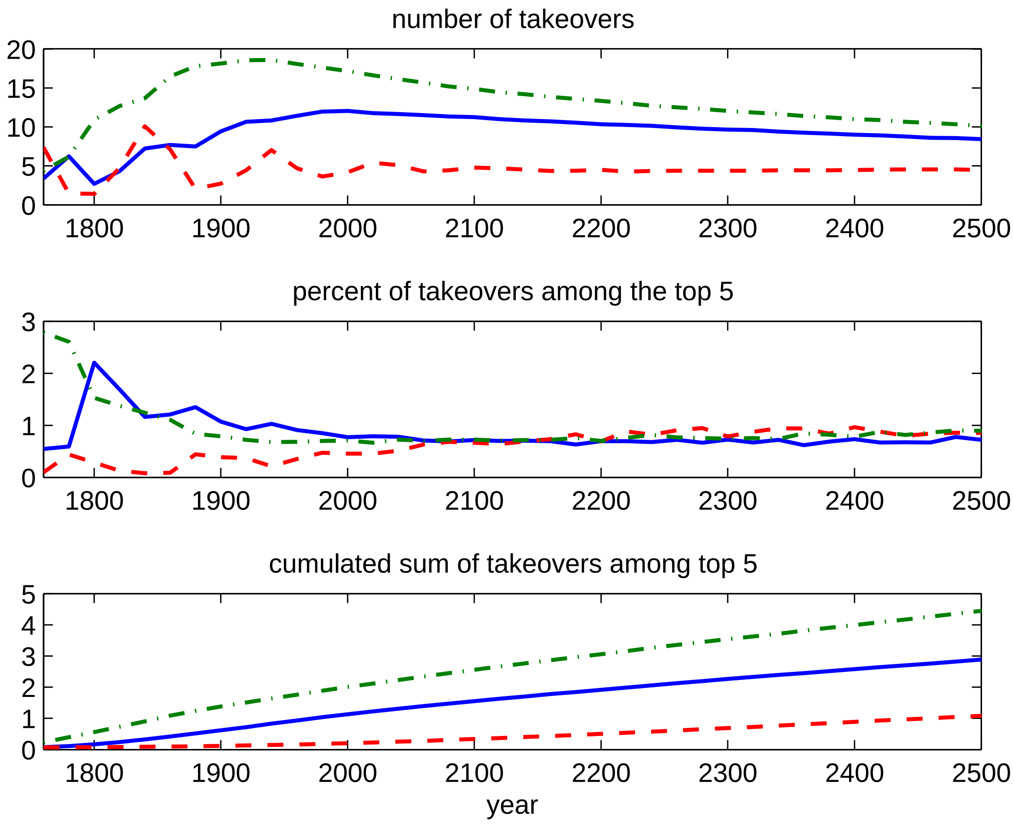
Figure 9: World Ranking Position for Country 1 - 10




$f_1(h+\Delta, h) = L\left\{[\epsilon h + (1-\epsilon)(h+\Delta)]^{1-\alpha}(h+\Delta)^{\alpha} - \bar{c}\right\}$. Consider the implausible yet illuminating case in which *all* knowledge comes from abroad, i.e. $\epsilon = 1$, and human capital is not a stand-alone production factor, i.e. $\alpha = 0$. In this case, the two economies have changed their roles in the next period. Now, the first country is the better endowed one but it keeps this status only for one period after which the advantage is again transferred to the second country. There is overtaking in every period.

Generally, overtaking seems to be more likely the greater $\epsilon$ and the lower $\alpha$. To verify this claim for the simple example, take a first order Taylor approximation around $\Delta = 0$ and compute $f_1 - f_2 = L\Delta(2\epsilon(1-\alpha) - 1)$. This expression becomes positive, indicating overtaking, for $\epsilon(1-\alpha) > 1/2$. For the actual model with a complex network of one hundred participating economies we cannot obtain a simple condition for overtaking. Instead, we investigate overtaking frequencies by way of numerical experiments. For that purpose, we run the model 5000 times (i.e. for 5000 alternative specifications of the Small World network) and count the average number of overtakings in each period. An overtaking is defined as the advancement by one step in the income-ranking of countries. Countries of the same income level are assigned the same rank. If, for example, a country advances from rank 5 to 4 in one period, it is recorded as 1 overtaking. However, if it advances from rank 5 to rank 3 we count 2 it as overtakings.

Figure 10: Overtaking Frequencies




Solid lines: benchmark run ($\epsilon = 0.5$). Dashed lines: low international knowledge diffusion ($\epsilon = 0.2$). Dash-dotted lines: high international knowledge diffusion ($\epsilon = 0.8$).

The results for the benchmark model are shown by solid lines in Figure 10. The top panel shows the total number of overtakings per period. On average, we observe about 8 overtakings. Overtakings are relatively rare during early global development, gradually increasing until they reach a maximum in the late 20th century and then gradually declining to a level of about 8. Then, they stay at this level even if the the simulation runs far way beyond the year 2500. This suggests that overtaking never stops. The world reaches a steady state only in terms of growth rates and relative income levels (see Section 2).

Although overtaking takes place surprisingly frequently at the world level, it is at the same time quite rare among the world leaders. But even the world leader cannot expect to maintain his position permanantly. This is shown in the middle panel of Figure 10 where we consider the top 5 countries in terms of GDP per capita. On average, only about 1 percent of overtaking takes place among the top 5. If overtakings were equally distributed among countries, we would have expected about 5 percent of them taking place in the top 5. When 1 percent of overtaking takes place among the world leaders, and there are on average 10 overtakings, this means that there are on average $10 \times 0.01 = 0.1$ overtakings among the leaders in that year (as for example in the year 1900). In order to better assess these results quantitatively, the bottom panel computes the cumulated sum of average overtakings among the top 5. For the benchmark case, there is about 1 overtaking happening before the year 2000 and 2 overtakings before the year 2500.

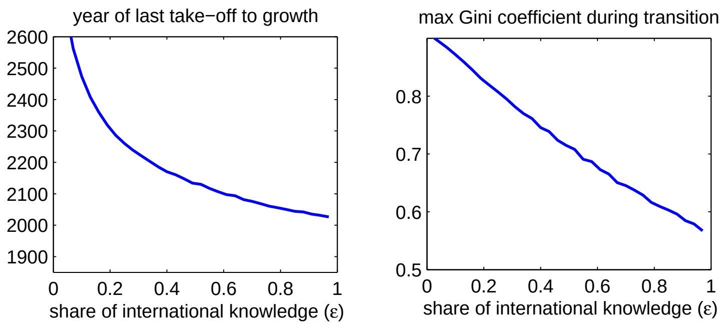
The incidence of overtaking, naturally, depends on the degree of openness ($\epsilon$). Dashed lines in Figure 10 show that there are fewer overtakings in total and among the top 5, when only 20 percent of productivity advancements are learned from abroad ($\epsilon = 0.2$). More overtakings can be expected when openness is large, as demonstrated by the dashed lines for $\epsilon = 0.8$. We thus find numerical evidence in large networks for the theoretical conclusions about the role of $\epsilon$ derived from small (two-country) networks. Results for similar numerical experiments on $\alpha$ (available upon request) confirm the conclusion derived from small country networks as well: more overtaking is taking place when $\alpha$ is low, i.e. when knowledge is relatively more important in production than human capital.

Interestingly, Figure 10 also reveals that the degree of openness affects the chronological nature of overtaking behavior. When $\epsilon$ is large, we observe just one historical peak in the number of overtakings (the roaring 20's). When $\epsilon$ is low, however, knowledge diffuses so slowly through the world network that we observe "waves of overtaking". Relatively turbulent times are followed by calm periods, concerning rank changes in the world income distribution.

5.6. **Network Effects on Global Inequality and Growth.** We next investigate how the specific make up of the network affects the evolution of the world income distribution. For that purpose we focus on two characteristic numbers, the calendar time when the last country takes off from stagnation and the maximum Gini index reached during the transition. Since long-distance links are set at random in the Small World model, we ran each specification of the model 1000 times and took averages afterwards.

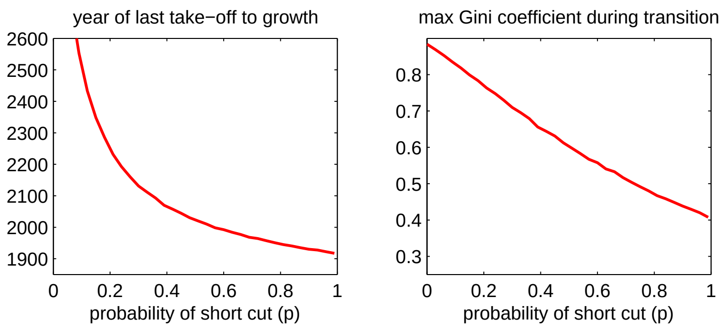
Figure 11 shows that a large contribution of international knowledge flows to productivity (that is large $\epsilon$) increases the pace of world development. Larger international knowledge spillovers are helpful to reduce world-wide inequality faster because a greater share of the initial knowledge advantage of the leaders is passed on through the network. The original Romer (1986) model is a knife-edge case since there will never be convergence for $\epsilon = 0$, that is in isolation. When no knowledge flows towards the initially poor countries, they do not escape from stagnation (without exogenous impulse) and the world remains at an asymmetric equilibrium in which only the club of rich countries is growing.

Figure 11: Inequality and Growth – Degree of International Knowledge Diffusion




Example with 100 countries; 10 percent initially rich, $p = 0.3$, varying $\epsilon$.

Figure 12: Inequality and Growth – Share of Long-Distance Links




Example with 100 countries; 10 percent initially rich, $\epsilon = 0.5$, varying $p$.

Next, in Figure 12, we investigate how the share of long-distance links affects the evolution of the world income distribution. The year of the last take-off decreases very quickly for low values of $p$ but remains rather insensitive for $p$'s larger than 0.5. The outcome reflects a well known feature of the Small World model, namely that average path length between nodes decreases sharply at low values of $p$ and not much at high values (Watts and Strogatz 1998). Maximum inequality also decreases sharply with increasing $p$, in an almost linear way. If every country had a long-distance link ($p = 1$), the last take-off would have been, according to the model, around the year 1900 with an associated maximum Gini index of 0.4.[9]

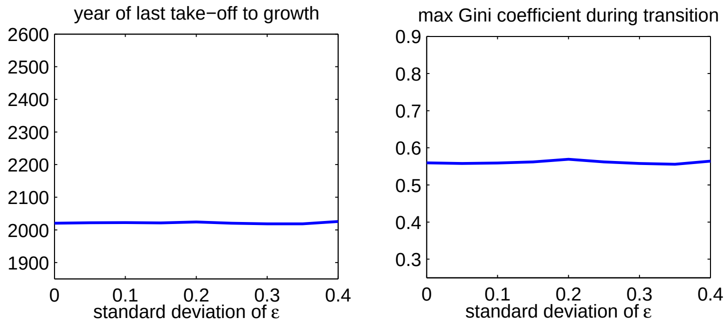
5.7. **Country-Specific Degrees of Openness.** It could be argued that the degree of openness to knowledge flows from abroad ($\epsilon$) varies across countries. We thus finally

[9]It can be shown that the year of the last take-off and the maximum Gini also depends quite strongly on the share of initially rich countries. The initial income ratio between rich and poor countries, in contrast, does not much affect the speed of transition and maximum inequality. The reason is that the negative impact on income inequality of an initially higher income gap is almost completely balanced by the fact that more can be learned from initially better endowed economies.

demonstrate that allowing for country-specific openness adds more realism but leaves our main results basically unaffected. The performance of individual countries, of course, depends crucially on their degree of openness. In particular, we expect initially backward countries with high degree of openness to catch up relatively quickly and relatively closed countries to be latecomers of industrialization. At the world level, however, we expect little change in performance. In order to verify this claim we assume that the degree of openness is a normally distributed random variable with mean $\epsilon$ and standard deviation $\sigma$. We then run the Small World model 1000 times for alternative values of $\sigma$ and record the year of the last take-off and the maximum Gini index during the transition.

Figure 13 shows the outcome for alternative $\sigma \in (0, 0.4)$ and $\epsilon$ drawn from a (truncated) normal distribution.[10] For better comparison, we kept the scaling of Figure 11. There is basically no change in the average maximum Gini along the transition and the year of the last take-off as the standard deviation of the degree of openness increases from 0 (our benchmark case) to 0.4.

Figure 13: Inequality and Growth – Varying Distribution of Openness




Example with 100 countries; 10 percent initially rich. Country-specific degree of openness, $\epsilon$ is normally distributed with mean 0.5 and standard deviation $\sigma \in (0, 0.4)$.

Summarizing the results from the last two subsections, we observe that the speed of industrialization and the evolution of world inequality crucially depends on the aggregate structure of the network, summarized in the (average) degree of openness ($\epsilon$) and the share of long-distance links ($p$). The distribution of comparative advantage of countries as measured, for example, by the standard deviation of country-specific $\epsilon$, has comparatively minor effects on aggregate behavior.

[10] In the rare event when the random draw provided a value above unity or below zero, we assign a value of 0.01 and 0.99, respectively.

## 6. Conclusion

In this paper, we laid out a network-based theory of knowledge diffusion as an explanation for the divergence of countries as well as for their subsequent global convergence. Besides the endogenous evolution of the world income distribution, the theory contributes also to the explanation of the new Kaldor facts (Jones and Romer, 2010). The theory generates S-shaped transition paths with gradual take-off from stagnation as well as overshooting growth rates at later stages of development. In the long run, it thus predicts (slow) convergence of relative income across the globe.

The model could be extended such that it predicts permanent relative income inequality by assuming that some countries use the available knowledge less efficiently than others. From the perspective of the very long run, however, convergence appears to be more intuitively appealing. However, even with all knowledge eventually diffusing through the whole world, inequality vanishes only in relative terms, measured, for example, by the conventional Gini index. Absolute inequality, measured, for example, by the absolute Gini index, is predicted to keep on rising with increasing global development.

Although the underlying economic model has been a deliberately simple one, the theory can already explain a great variety of individual growth performances, including the overtaking of countries in the course of global development. Naturally, further extensions are conceivable. For example, the reliance on learning-by-doing could be relaxed by introducing a two-state process according to which learning-by-doing eventually triggers market R&D activities as in Strulik et al. (2013). Another interesting application could be to investigate the spread of the fertility transition through a global network. In the present paper, we conceptualized globalization as the increasing flow of knowledge through the world. The network itself, however, may be subject to globalization as well. Integrating an increasing share of long-distance links over time, as in as in Lindner and Strulik (2014), could be an interesting extension of our network-based theory of global knowledge diffusion and growth.

## References


Acemoglu, D. and Ventura, J., 2002, The world income distribution, *Quarterly Journal of Economics* 117, 659-694.

Arrow, K.J., 1962, The economic implications of learning-by-doing, *Review of Economic Studies* 29, 155-173.

Atkinson, A.B. and Brandolini, A., 2010, On analyzing the world distribution of income, *The World Bank Economic Review* 24, 1-37.

Bairoch, P., 1993, *Economics and World History*, University of Chicago Press, Hemel Hempstead.

Barro, R.J. and Sala-i-Martin, X., 2004, *Economic Growth* (2nd ed.), MIT Press, Cambridge, MA.

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

Bourguignon, F. and Morrisson, C., 2002, Inequality among world citizens: 1820-1992, *American Economic Review* 92, 727-744.

Cavalcanti, T. and Giannitsarou, D., 2012, Network structure and human capital dynamics, Working Paper, University of Cambridge.

Chakravarty, S. R., 1988, Extended Gini indices of inequality, *International Economic Review* 29, 147-156.

Comin, D. A., Dmitriev, M., and Rossi-Hansberg, E., 2012. The spatial diffusion of technology, NBER Discussion Paper 18534.

Crafts, N.F.R., 2003, Quantifying the contribution of technological change to economic growth in different eras: a review of the evidence, Economic History, London School of Economics and Political Science.

De Long, B., 1998, Estimates of World GDP, One Million BC – Present. University of California, Berkeley, Department of Economics, `http://delong.typepad.com/print/20061012_LRWGDP.pdf`.

Diamond, J., 1997, *Guns, Germs, and Steel: The Fates of Human Societies* WW Norton, New York.

Doepke, M., 2004, Accounting for fertility decline during the transition to growth. *Journal of Economic Growth* 9, 347–383.

Easterly, W. and Levine, R., 2001, What have we learned from a decade of empirical research on growth? It's not Factor accumulation: stylized facts and growth models, *The World Bank Economic Review* 15, 177-219.

Eaton, J. and Kortum, S., 1999, International patenting and technology diffusion: Theory and measurement, *International Economic Review* 40, 537-70.

Eurostat (2012), Educational expenditure statistics (epp.eurostat.ec.europa.eu/statistics_explained/index.php/Educational_expenditure_statistics)

Fogli, A. and Veldkamp, L., 2011, Germs, social networks and growth, mimeo, University of Minnesota.

Galor, O., 2005, From stagnation to growth: unified growth theory, in: P. Aghion and S. Durlauf, *Handbook of Economic Growth* Vol 1A, Amsterdam, North-Holland.

Galor, O. and Weil, D.N., 2000, Population,technology and growth: From the Malthusian regime to the demographic transition and beyond, *American Economic Review* 110, 806–828.

Galor, O. and Moav, O., 2002, Natural selection and the origin of economic growth, *Quarterly Journal of Economics* 117, 1133-1192.

Galor, O. and Moav, O., 2004, From physical to human capital accumulation: Inequality and the process of development, *Review of Economic Studies* 71, 1001-1026.

Galor, O. and Mountford, A., 2008, Trading population for productivity: theory and evidence, *Review of Economic Studies* 75, 1143-1179.

Gerschenkron, A., 1962, Economic backwardness in historical perspective, in: Hoselitz, B. (ed.) *The Progress of Underdeveloped Areas*, Chicago University Press, Chicago.

Howitt, P., 2000, Endogenous growth and cross-country income differences, *American Economic Review* 90, 829-846.

Jackson, M.O. and Yariv, L., 2007, Diffusion of Behavior and Equilibrium Properties in Network Games, *American Economic Review* (Papers and Proceedings), 97(2), 92-98.

Jones, C. I., 1997, On the evolution of the world income distribution, *Journal of Economic Perspectives* 11(3), 19-36.

Jones, C.I., 2001, Was an industrial revolution inevitable? Economic growth over the very long run, *Advances in Macroeconomics* 1, 1-43.

Jones, C.I. and Romer, P.M., 2010, The new Kaldor facts: ideas, institutions, population, and human capital, *American Economic Journal: Macroeconomics* 2, 224-245.

Kaldor, N., 1961, Capital accumulation and economic growth, in: *The Theory of Capital*, ed. F.A. Lutz and D.C. Hague, St. Martins Press, New York.

Keller, W., 2002, Geographic localization of international technology diffusion, *American Economic Review* 92, 120-142.

Klenow, P.J. and Rodriguez-Clare, A., 2005, Externalities and growth, in: Aghion, P. and S.N. Durlauf (eds.) *Handbook of Economic Growth* Vol. 1, 817-861, Elsevier, Amsterdam.

Kolm, S.-C., 1976, Unequal inequalities I, *Journal of Economic Theory* 12, 416-442.

Kuznets, S., 1966, *Modern Economic Growth: Rate, Structure and Spread*, Yale University Press, New Haven.

Lindner, I. and Strulik, H., 2014, From Tradition to Modernity: Economic Growth in a Small World, *Journal of Development Economics*, forthcoming.

Lucas, R.E., 2000, Some Macroeconomics for the 21st century, *Journal of Economic Perspectives* 14, 159-168.

Lucas, R.E., 2009, Trade and the industrial revolution, *American Economic Journal: Macroeconomics* 1, 1-25.

Newman, M.E.J. and Watts, D.J., 1999, Renormalization group analysis of the small-world model, *Physics Letters A* 263, 341-346.

Newman, M.E.J., 2003, The structure and function of complex networks, *SIAM Review* 45, 167-256.

Parente, S.L. and Prescott, E.D., 1994, Barriers to technology adoption and development, *Journal of Political Economy* 102, 298–32.

Parente, S.L. and Prescott, E.D., 2005, A unified theory of the evolution of international income levels, in: Aghion, P. and S.N. Durlauf (eds.) *Handbook of Economic Growth*, Elsevier, Amsterdam, Chapter 21.

Pomeranz, K., 2000, *The Great Divergence: China, Europe, and the Making of the Modern World*, Princeton, Princeton University Press.

Romer, P.M., 1986, Increasing returns and long-run growth, *Journal of Political Economy* 94, 1002-1037.

Sala-i-Martin, X., 2006, The world distribution of income: falling poverty and convergence, period, *Quarterly Journal of Economics* 121, 351-397.
336-352.

Strulik, H. and Weisdorf, J., 2008, Population, food, and knowledge: A simple unified growth model, *Journal of Economic Growth* 13, 169-194.

Strulik, H., Prettner, K., and Prskwetz, A., 2013, The past and future of knowledge-based growth, *Journal of Economic Growth*, 18, 411-437.

Sutcliffe, B., 2004, World inequality and globalization, *Oxford Review of Economic Policy* 20, 15-37.

Ventura, J., 2005, A global view of economic growth, in: P. Aghion and S. Durlauf, *Handbook of Economic Growth* Vol 1B, Amsterdam, North-Holland.

Watts, D.J., and Strogatz, S.H., 1998, Collective dynamics of "small-world" networks, *Nature* 393, 440-442.

## Appendix

**Proof of Proposition 1.** Assume the society starts in $h_0 = (h_{10}, ..., h_{n0})$. Put $h_{max0} = \max\{h_{10}, ..., h_{n0}\}$. Let $i$ denote the country with $h_{i0} = h_{max0}$. Since $A_{it}$ takes weighted averages of human capital we get as an upper bound , $A_{i\max 0} \leq h_{i\max 0}$. Growth of country-specific human capital is therefore bounded from above by the scenario where $i$ is linked exclusively to a country $j \neq i$ that belongs to the richest group at the beginning and there are no other links between $i$ and $j$. In this case, we get $A_{it} = h_{jt} = h_{it}$ and (8) simplifies to $h_{i,t+1} = \max\{1, L \cdot (h_{it} - \bar{c})^\eta\}$. The scalar function $f(h) = L \cdot (h - \bar{c})^\eta$ is concave for $\eta < 1$ and depending on parameter values has two or less intersection points with the identity line.

*Case 1*: Two intersection points with the identity line. In this case, the larger fix point of the equation $h^* = f(h^*)$ is stable.

*Case 2:* Less than two intersection points. Here, $f(h)$ is either equal or below the identity line for all $h$. Hence $h_{i\max,t+1} \leq \max\{1, L \cdot (h_{i\max 0} - \bar{c})^\eta\}$ for all $t \geq 0$ such that growth eventually declines.

For $\eta = 1$, $L \leq 1$ ensures that $h_{i\max,t+1} = \max\{1, L \cdot (h_{i\max t} - \bar{c})\} \leq \max\{1, h_{i\max 0}\}$, which verifies the absence of long-run growth.

**Proof of Proposition 2.** With identical human capital endowments, the neighborhood weighted average of human capital is given by $A_{it} = h_{it}$. Equation (6) simplifies to

$$g_{it}^h \equiv \frac{h_{i,t+1}}{h_{it}} = \max\left\{\frac{1}{h_{it}},\ L \cdot \frac{(h_{it} - \bar{c})^\eta}{h_{it}}\right\},$$

which provides equal growth rates for equal levels of human capital.

**Proof of Proposition 3.** We first discuss the case $\eta < 1$. Proposition 1 states convergence to an equilibrium for all countries. Assume that there is an asymmetric stable equilibrium $h^* = (h_1^*, ..., h_n^*)$ and put $h_{max}^* = max\{h_1^*, ..., h_n^*\}$. Since the network is connected, there exists a country $r$ with $h_r^* = h_{max}^*$ but lower average human capital of its neighborhood $\bar{h}_r^* < h_{max}^*$. We will show below that the function

$$f(h) = L \cdot \left([\epsilon\bar{h} + (1 - \epsilon)h]^{1-\alpha} h^\alpha - \bar{c}\right)^\eta \tag{A.1}$$

derived from (8) is concave in $h$ for $\eta < 1$. Note that (A.1) is the same for all countries aside from the neighborhood average $\bar{h}$. Since $f(h)$ is concave, the stable equilibrium is the upper intersection of $f(h)$ with the identity line. The upper intersection increases with $\bar{h}$. The country with the highest intersection $h_{max}^*$ is therefore a country with maximal neighborhood average $\bar{h}_{max}$ which contradicts our construction of $r$. We conclude that there are no stable asymmetric equilibria which also rules out unstable asymmetric equilibria (lower intersection of (A.1)).

It remains to be shown that $f(h)$ in (A.1) is concave. Put $f(h) = z(g(h))$, with $g(h) = \left[\varepsilon\bar{h} + (1-\varepsilon)h\right]^{1-\alpha} h^{\alpha} - \bar{c}$ and $z(y) = Ly^{\eta}$. For the second derivative of $f$ follows

$$f''(h) = z''(g(h))(g(h))^2 + z'(g(h))g''(h). \tag{A.2}$$

For $\eta < 1$ we have $z''(g(h)) < 0$ such that the left term in the sum of (A.2) is negative. A sufficient condition for $f''(h) < 0$ is therefore that $g''(h) < 0$ since $z'(h) > 0$. Put

$$g(h) = u(h)v(h) \text{ with } u(h) = \left[\varepsilon\bar{h} + (1-\varepsilon)h\right]^{1-\alpha} \text{ and } v(h) = h^{\alpha}.$$

The second derivative of $g$ is $g'' = u''v + 2u'v' + uv''$ and after dividing $g''(h) < 0$ by $u'v'$ we get

$$\frac{u''}{u'}\frac{v}{v'} + 2 + \frac{u}{u'}\frac{v''}{v'} < 0. \tag{A.3}$$

The fractions in this inequality are given by

$$\begin{aligned} \frac{u''}{u'} &= -\frac{\alpha(1-\varepsilon)}{\varepsilon\bar{h} + (1-\varepsilon)h}, \quad \frac{v}{v'} = \frac{h}{\alpha}, \\ \frac{u}{u'} &= \frac{\varepsilon\bar{h} + (1-\varepsilon)h}{(1-\alpha)b}, \quad \frac{v''}{v'} = -\frac{(1-\alpha)}{h}. \end{aligned} \tag{A.4}$$

Inserting (A.4) into (A.3) gives

$$-(1-\varepsilon)\frac{h}{\varepsilon\bar{h} + (1-\varepsilon)h} + 2 - \frac{1}{(1-\varepsilon)}\frac{\varepsilon\bar{h} + (1-\varepsilon)h}{h} < 0. \tag{A.5}$$

Multiplying (A.5) by

$$-\frac{1}{(1-\varepsilon)}\frac{h}{\varepsilon\bar{h} + (1-\varepsilon)h}$$

provides

$$\left(\frac{h}{\varepsilon\bar{h} + (1-\varepsilon)h}\right)^2 - 2\frac{h}{\varepsilon\bar{h} + (1-\varepsilon)h}\frac{1}{(1-\varepsilon)} + \left(\frac{1}{(1-\varepsilon)}\right)^2 > 0$$

which simplifies to

$$\left(\frac{h}{\varepsilon\bar{h} + (1-\varepsilon)h} - \frac{1}{(1-\varepsilon)}\right)^2 > 0,$$

which always holds unless $\bar{h} = 0$.

Finally, consider the case $\eta = 1$. The term $z''(g(h)$ in (A.2) equals zero such that a sufficient condition for $f''(h) < 0$ is $g''(h) < 0$, which is analogous to the case $\eta < 1$. Analogously, we can rule out asymmetric equilibria for $\eta = 1$. For $\bar{h} = h$ and $\eta = 1$, the function $f(h)$ in (A.1) is linear and lies below the identity line $f(h) = h$ without intersection, a fact that implies convergence to subsistence.

**Proof of Proposition 4.** For $\eta = 1$ the growth rate of country $i$ after the take-off is given by

$$
\begin{aligned}
h_{it+1}/h_{it} &= L\left([\epsilon\,\bar{h}_{it} + (1-\epsilon)h_{it}]^{1-\alpha}h_{it}^{\alpha} - \bar{c}\right)/h_{it} = L\left([\epsilon\,\frac{\bar{h}_{it}}{h_{it}} + (1-\epsilon)]h_{it} - \bar{c}\right)/h_{it} \\
&= L\left[\epsilon\,\frac{\bar{h}_{it}}{h_{it}} + (1-\epsilon)\right] - L\frac{\bar{c}}{h_{it}}.
\end{aligned}
$$

As we will show in Proposition 7, the relative distance between countries in connected components tends to zero. This is equivalent to $\bar{h}_{it}/h_{it}$ tending to one, either because both $\bar{h}_{it}$ and $h_{it}$ stagnate, or because both grow forever. The implied growth rate in the latter case is $h_{it+1}/h_{it} = L[\epsilon \cdot 1 + (1-\epsilon)] - L \cdot 0 = L$.

**Proof of Proposition 5.** Jackson and Yariv (2007) use a similar argument to prove S-shaped rates of adoption (see their Proposition 7).

Put $u(\xi) = [\epsilon\,\xi + (1-\epsilon)]^{1-\alpha}$ such that $g(\xi(h), h) = u(\xi(h)) - \bar{c}/h$. Notice that the conditions of the propositions imply $\xi''(h) < 0$ and that $\xi'(h)$ changes from positive to negative. For the derivatives follows

$$
\begin{aligned}
\frac{d}{dh}g(\xi(h),h) &= u'(\xi)\xi'(h) + \frac{\bar{c}}{h^2} = \epsilon\,(1-\alpha)[\epsilon\,\xi + (1-\epsilon)]^{-\alpha}\xi'(h) + \frac{\bar{c}}{h^2} > 0 \\
\frac{d^2}{(dh)^2}g(h) &= \underbrace{\epsilon^{\,2}(1-\alpha)(-\alpha)[\epsilon\xi + (1-\epsilon)]^{-\alpha-1}\left[\xi'(h)\right]^2}_{<0} + \underbrace{\epsilon\,(1-\alpha)[\epsilon\,\xi + (1-\epsilon)]^{-\alpha}\xi''(h)}_{<0 \text{ since } \xi''<0} \underbrace{-2\frac{\bar{c}}{h^3}}_{<0}.
\end{aligned}
$$

Put $f(h) = [\epsilon\,\xi(h) + (1-\epsilon)]^{1-\alpha}h - \bar{c} = u(\xi(h))h - \bar{c}$. We get

$$
\begin{aligned}
u'(\xi) &= \underbrace{(1-\alpha)\epsilon[\epsilon\,\xi + (1-\epsilon)]^{-\alpha}}_{>0} = \frac{(1-\alpha)\epsilon}{\epsilon\,\xi + (1-\epsilon)}u(\xi) < u(\xi) \text{ if } \xi \geq 1. \\
f'(h) &= \underbrace{u'(\xi)\xi'(h)h}_{\text{positive first, then negative}} + u(\xi(h)) \\
&= \frac{(1-\alpha)\epsilon}{\epsilon\,\xi + (1-\epsilon)}u(\xi)\xi'(h)h + u(\xi) = \left[\frac{(1-\alpha)\epsilon}{\epsilon\,\xi + (1-\epsilon)}\xi'(h)h + 1\right]u(\xi) \\
f''(h) &= u''(\xi)\left(\xi'(h)\right)^2 h + u'(\xi)\left[\xi''(h)h + \xi'(h)\right] + u'(\xi)\xi'(h) \\
&= \underbrace{u''(\xi)\left(\xi'(h)\right)^2 h}_{<0} + u'(\xi) \underbrace{\left[\xi''(h)h + 2\xi'(h)\right]}_{<0 \text{ if } \xi''(h) <0 \text{ and } \xi'(h) \text{ approaches 0 or gets negative.}}
\end{aligned}
$$

Now put $y = f(h)$. Since $h$ is increasing, we have $y > h$. We know that for $\xi' > 0$ it follows that $g'(y)$ and $g'(h)$ are positive. Since $g''(h) < 0$ it holds that $g'(y) > g'(h)$. We get $(g(y) - g(h))' = g'(y)f'(h) - g'(h)$ and $(g(y) - g(h))' > 0$ is equivalent to

$$
f'(h) > \frac{g'(h)}{g'(y)}.
$$

Note that eventually $f''(h) < 0$.

We distinguish 3 cases:

(1) $f'(h)$ is larger than $g'(h)/g'(y)$ initially and equal to $g'(h)/g'(y)$ for finite $\hat{h} > 1$, i.e. finite time $T > 0$. This implies that the growth rate reaches a maximum $T$ and decreases afterwards. There is overshooting growth.

(2) $f'(h)$ is always smaller than $g'(h)/g'(y)$. In that case $T = 0$ and the growth rate keeps decelerating during the whole process.

(3) $f'(h)$ is larger than $g'(h)/g'(y)$ but does not pass it in finite time such that $T = \infty$. This means that there is long run growth but no overshooting growth.

**Proof of Proposition 6.** We define overshooting as the temporary surpassing of the long run growth rate (global overshooting). In case of positive long-run growth, the long-run growth rate is given by $L$. Thus, country $i$ overshoots at time $t$ if $g^h_{it} > L$. We next show that overshooting occurs due to a sufficiently high lag of human capital with respect to the neighborhood average. For $\eta = 1$ we have $g^h_{it} > L$ *iff*

$$\left[\epsilon \, \frac{\bar{h}_{it}}{h_{it}} + (1-\epsilon)\right]^{1-\alpha} = \left[1 + \epsilon \left(\frac{\bar{h}_{it}}{h_{it}} - 1\right)\right]^{1-\alpha} \quad > \quad 1 + \frac{\bar{c}}{h_{it}}.$$

We first prove that the richest country can never overshoot. The second step provides a sufficient condition for overshooting.

(i) Put $m_t = \arg\max_i h_{it}$. It follows that growth at time $t$ is bound from above by a growth scenario in which country $m$ is exclusively linked with an identical twin and vice versa. For $g^h_{mt}$ we get

$$g^h_{mt} \quad \leq \quad \frac{L \cdot ([\epsilon \, h_{mt} + (1-\epsilon) h_{mt}]^{1-\alpha} h^{\alpha}_{mt} - \bar{c})}{h_{mt}} = \frac{L \cdot (h_{mt} - \bar{c})}{h_{mt}} \leq L.$$

(ii) The growth rate is given by $g(h) = L \cdot \big([\epsilon \, \bar{h} + (1-\epsilon)h]^{1-\alpha} h^{\alpha} - \bar{c}\big)/h$.

From Proposition 3, we know that the long-run growth rate is $L$. Overshooting means that at some point $g(h) > L$, i.e.

$$\left[\epsilon \, \frac{\bar{h}}{h} + (1-\epsilon)\right]^{1-\alpha} - \frac{\bar{c}}{h} > 1 \quad \Leftrightarrow \quad \left[\epsilon \, \frac{\bar{h}}{h} + (1-\epsilon)\right]^{1-\alpha} > 1 + \frac{\bar{c}}{h}.$$

Note that for $\epsilon = 0$, the left-hand side simplifies to 1 and hence, the inequality is never fulfilled which confirms that global overshooting is triggered by sufficiently large $\bar{h}/h$. In particular, overshooting occurs if

$$\epsilon\frac{\bar{h}}{h}+(1-\epsilon)>\left(1+\frac{\bar{c}}{h}\right)^{\frac{1}{1-\alpha}} \quad\Leftrightarrow\quad \frac{\bar{h}}{h}>\frac{1}{\epsilon}\left(1+\frac{\bar{c}}{h}\right)^{\frac{1}{1-\alpha}}-\frac{1-\epsilon}{\epsilon}>1$$

$$\Leftrightarrow\ \frac{\bar{h}}{h}>1+\frac{1}{\epsilon}\left[\left(1+\frac{\bar{c}}{h}\right)^{\frac{1}{1-\alpha}}-1\right].$$

**Proof of Lemma 1.** For notational convenience we omit the dependence on time $t$.
(i) The Gini index is defined by

$$G=\frac{1/2-B}{1/2},$$

where $B$ is the area under the Lorenz curve. Without loss of generality assume the countries are labeled such that $y_1 \leq y_2 \leq ... \leq y_n$. Put $Y_k=\sum_{i=1}^{k} y_i$. The Lorenz curve is a polygonal line defined by the set of points

$$\left\{(0,0),(\frac{1}{n},\frac{Y_1}{Y}),(\frac{2}{n},\frac{Y_2}{Y}),...,(\frac{n-1}{n},\frac{Y_{n-1}}{Y}),(1,1)\right\}.$$

If all countries grow by the same rate, the fractions $Y_i/Y$ stay the same for all $i=1,...,n$. The Gini index is zero if and only if the Lorenz curve is the identity line which means, in particular, that the first slope of the polygonal line $ny_1/Y$ and the slope of the last polygonal line $ny_n/Y$ are identical. We conclude that the relative Gini index converges to zero if and only if

$$n\hat{d}=n\left(\frac{y_n}{Y}-\frac{y_1}{Y}\right)\to 0.$$

(ii) The term $\tilde{B}=nHB$ measures the area under the re-scaled Lorenz curve where the horizontal axis ranges from 0 to $n$ and the vertical axis from 0 to $Y$. The re-scaled Lorenz curve is a polygonal line defined by the set of points

$$\{(0,0),(1,Y_1),(2,Y_2),...,(n-1,Y_{n-1}),(n,Y)\}.$$

We get

$$\begin{aligned}\tilde{B} &= \frac{y_1}{2}+(y_1+\frac{y_2}{2})+(y_1+y_2+\frac{y_3}{2})+(y_1+y_2+y_3+\frac{y_4}{2})+...\\ &\quad +(y_1+y_2+...+y_{k-1}+\frac{y_k}{2})+...+(y_1+y_2+...+y_{n-1}+\frac{y_n}{2})\\ &= (n-1)y_1+\frac{y_1}{2}+(n-2)y_2+\frac{y_2}{2}+...+(n-k)y_k+\frac{y_k}{2}+...+\frac{y_n}{2}\\ &= (n-1)y_1+(n-2)y_2+...+(n-k)y_k+...+y_{n-1}+\frac{Y}{2}.\end{aligned}$$

For the absolute Gini index, we get

$$G\cdot\bar{Y} = G\cdot(1/n)Y=\frac{(nY)/2-(nH)B}{(nH)/2}\cdot(1/n)Y=\frac{(nY)/2-\tilde{B}}{(nH)/2}\cdot(1/n)Y=\frac{(nY)/2-\tilde{B}}{n^2/2}$$

$$= \frac{2}{n^2}\left[(nY)/2 - \tilde{B}\right] = \frac{2}{n^2}\left[(nY)/2 - \tilde{B}\right]. \tag{A.6}$$

We are now ready to prove the first claim of (ii) by induction. Note that the first term in (A.6) is just a scaling factor such that it suffices to prove the statement for the second term in brackets

$$T_n = \frac{n}{2}Y_n - \tilde{B}_n,$$

where the index indicates the amount of countries.

For $n = 2$ follows

$$\tilde{B}_2 = \frac{y_1}{2} + (y_1 + \frac{y_2}{2}) = y_1 + \frac{y_1 + y_2}{2}.$$

$$T_2 = (2 * Y_2)/2 - \tilde{B}_2 = Y_2 - y_1 - \frac{y_1 + y_2}{2} = \frac{y_2 - y_1}{2},$$

which proves the claim for $n = 2$.

Suppose Lemma is true for $n$ countries. We get

$$T_{n+1} = \frac{(n+1)}{2} * Y_{n+1} - \tilde{B}_{n+1} = \frac{(n+1)}{2} * (Y_n + y_{n+1}) - \tilde{B}_{n+1}.$$

$$\tilde{B}_{n+1} = \tilde{B}_n + y_1 + ... + y_n + \frac{y_{n+1}}{2} = \tilde{B}_n + Y_n + \frac{y_{n+1}}{2}.$$

$$\begin{aligned}
T_{n+1} &= \frac{(n+1)}{2} * (Y_n + y_{n+1}) - \tilde{B}_{n+1} = \frac{n}{2} * (Y_n + y_{n+1}) + \frac{1}{2} * (Y_n + y_{n+1}) - \tilde{B}_{n+1} \\
&= \frac{n}{2} * Y_n + \frac{n}{2} * y_{n+1} + \frac{1}{2} * Y_n + \frac{1}{2} * y_{n+1} - \tilde{B}_n - Y_n - \frac{y_{n+1}}{2} \\
&= \frac{n}{2} * Y_n - \tilde{B}_n + \frac{n}{2} * y_{n+1} - \frac{1}{2} * Y_n = T_n + \frac{1}{2}(n * y_{n+1} - Y_n).
\end{aligned} \tag{A.7}$$

From (A.7) we conclude that the term in brackets does not change if all incom levels increase by the same amount.

Finally, note that the relative Gini index is given by the area between the identity line and the Lorenz curve divided by the total area under the identity line from 0 to 1 (which is 1/2). Multiplying the relative Gini index by $Y/n$ is equivalent to studying this ratio with a rescaling of the vertical axis ranging from 0 to $Y/n$. Here, the rescaled Lorenz curve is a polygonal line defined by the set of points

$$\left\{(0,0), (\frac{1}{n}, \frac{Y_1}{n}), (\frac{2}{n}, \frac{Y_2}{n}), ..., (\frac{n-1}{n}, \frac{Y_{n-1}}{n}), (1, \frac{Y}{n})\right\}.$$

The area between the identity line and the rescaled Lorenz curve is 0 if and only if the slope if the first polygonal line of the Lorenz curve has the same slope as the last one. This is equivalent to $y_1 = y_n$.

(iii) Average income $\bar{Y}$ is bounded from below by 1 by (3). This implies that the product $G\bar{Y}$ can only tend to zero if the relative Gini index tends to zero. However, the latter is not a sufficient condition since $G\bar{Y}$ does not decrease if $\bar{Y}$ grows with a higher rate than the declining rate of $G$.

**Proof of Proposition 7.** From Proposition 1, we know that for $\eta < 1$, long run growth is not possible and all countries eventually converge to the same income level if the network connected. This implies that for $\eta < 1$, the Gini coefficient always tends to zero if there is a path between any two countries.

For $\eta = 1$, we first show that the dynamics

$$h_{it+1} \quad = \quad L \cdot \max\left\{1, u(h_{it}, \bar{h}_{it})\right\} \tag{A.8}$$

with

$$u(h_{it}, \bar{h}_{it}) = [\epsilon\, \bar{h}_{it} + (1-\epsilon)h_{it}]^{1-\alpha} h_{it}^{\alpha} - \bar{c} \tag{A.9}$$

consists of 2 subsequent steps: (i) $u$ is a contraction mapping which decreases distance between human capital levels and (ii) scaling, i.e. multiplication by $L$.

Put $h_{rt} = \max_i h_{it}$, $h_{pt} = \min_i h_{it}$ such that $D_t = h_{rt} - h_{pt}$.

Lemma 2. *Let the network be connected. For $\eta = 1$ and $L \leq 1$ the absolute income gap $D_t$ decreases monotonically to zero after the last take-off. For $L \geq 1$ distance $D_t$ grows slower than $L^t$, in short,*

$$D_t = o\left(L^t\right). \tag{A.10}$$

*Proof.* From the definition of $D_t$ follows

$$D_{t+1} \;\leq\; L[\epsilon\, h_{rt} + (1-\epsilon)h_{rt}]^{1-\alpha} h_{rt}^{\alpha} - L[\epsilon\, h_{pt} + (1-\epsilon)h_{pt}]^{1-\alpha} h_{pt}^{\alpha} \leq Lh_{rt} - Lh_{pt} \leq D_t \text{ for } L \leq 1.$$

Let $S_{rt} = \{i \in N | h_{it} = \max_{j\in N} h_{jt}\}$ denote the set of countries with the highest income level at time $t$. From time $t$ to $t+1$ it is possible that countries leave or join this set due to externality effects. However, since the network is connected, it is not possible that countries in $S_{rt} \neq N$ develop forever with an externality level $\bar{h}_{rt} = h_{rt}$. At some point in time, the richest countries face an externality level $\bar{h}_{rt} < h_{rt}$ such that $h_{rt+1} < h_{rt}$. Analogously, since the poorest countries are not isolated, we have $h_{p,t+1} > h_{pt}$ at some point in time such that $D_{t+1} < D_t$.

Finally, we conclude that mapping $u$ from (A.9) never increases distance and that there exists a $t > 0$ such that $u$ is a contraction. This contraction is followed by multiplication with $L$ in (A.8). We conclude that $D_t$ has the shape of $L^t$ multiplied by a zero sequence, in short, $D_t$ grows slower than $L^t$, as displayed in (A.10). □

Lemma 3. *The Gini coefficient tends to zero in connected networks.*

*Proof.* From Lemma 2, we know that all $y_{it}$ values lie in an interval $I_t$ with measure $\mu(I_t) = o(L^t)$. In other words, there exists a $y^*$ such that $y_{it}/L^t \to y^*$ for any $i \in N$. For $Y_t = \sum_{i=1}^n y_{i,t}$ follows $Y_t/L^t \to ny^*$ and hence,

$$\hat{d}_t = \frac{D_t}{Y_t} = \frac{D_t/L^t}{Y_t/L^t} \to \frac{0}{ny^*} = 0.$$

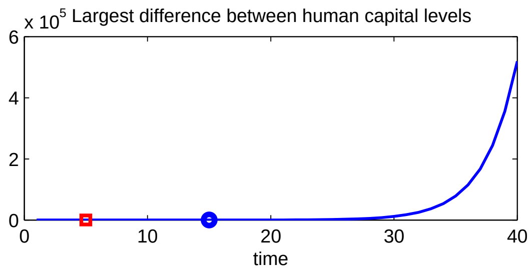
From Lemma 1 we conclude that the Gini coefficient tends to zero. □

**Proof of Proposition 8.** The proposition is proven by way of example. Consider the bridge network from Figure 3 with $n = 10$ and where half of the population starts rich. All other settings are as in the benchmark case, in particular, $\eta = 1$ and $L = 1.48 > 1$ ensure long-run growth. Figure A.1 illustrates the time path of the largest human capital difference $d_t$ with an example. The largest difference grows exponentially, which can be explained by inspecting

$$\begin{aligned} f(h) &= L \cdot \left([\epsilon \bar{h} + (1-\epsilon)h]^{1-\alpha} h^\alpha - \bar{c}\right) = L \cdot \left([\epsilon \frac{\bar{h}}{h} + (1-\epsilon)]h - \bar{c}\right) \\ &= L \cdot \left([\epsilon \frac{\bar{h}}{h} + (1-\epsilon)] - \bar{c}/h\right) h \approx L \cdot [\epsilon \frac{\bar{h}}{h} + (1-\epsilon)]h \end{aligned}$$

for large $h$. The dense connectedness of neighbors with similar endowments of human capital implies that $h$ essentially grows each time by a factor $L$ which explains why the shape of $d_t$ behaves similar to $L^t d_1$. Given the explosive growth of largest distance it might seem puzzling that the Gini coefficient declines. For that, recall that the Gini coefficient works with a normalization. It uses relative instead of absolute numbers.

Figure A.1: Largest Difference of Human Capital Levels. Bridge Network




Time measured in generations. The red square indicates the time of the last take-off to growth; the blue circle indicates the time when the Gini coefficient assumes its maximum.

**Decreasing and Increasing Returns.** Consider a generalization of the model in which $A_{it} = \left(\sum_{j=1}^{n} \bar{w}_{ij} h_{jt}\right)^{\omega}$, $\omega \neq 1-\alpha$, replaces (7). Aggregate production displays decreasing returns to scale for $\omega < (1-\alpha)$ and increasing returns to scale for $\omega > (1-\alpha)$. The latter case is interesting because it allows for positive growth at the steady-state without the need of linear returns to education (without the need of $\eta = 1$). Figure A.2 shows such a case in which a high value for the importance of international knowledge flows $\omega$ compensates for decreasing returns in education ($\epsilon = 0.75$). Nevertheless, there is long-run growth. The value of $B$ has been adjusted in order to arrive at a steady-state growth rate of 1.5 percent. To the naked eye, the figure looks very similar to Figure 6 in the paper.

Figure A.2: Decreasing Returns to Education and Steady Growth

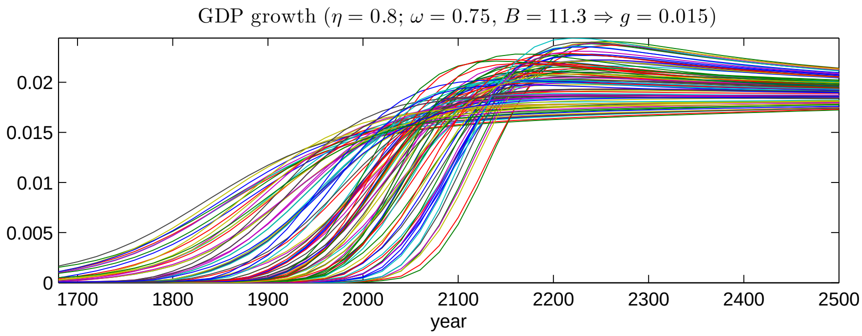
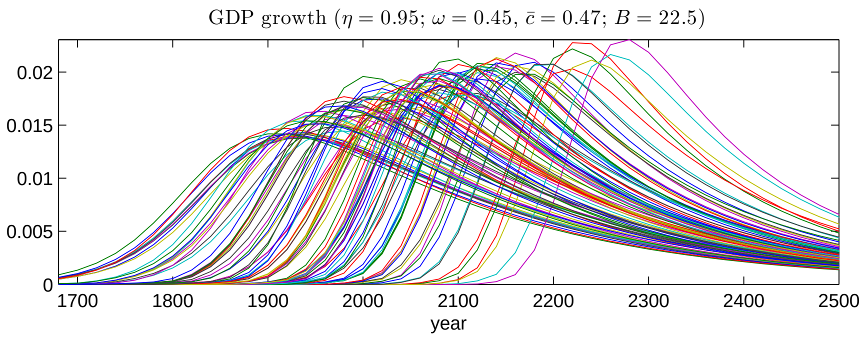



Figure A.3: Decreasing Returns to Education and Decreasing Returns to Scale




The case of decreasing returns to scale is interesting to investigate dynamics when there is no positive growth at the steady state. Figure A.3 shows the result for an example in which there are mildly decreasing returns to scale ($\omega = 0.45$) as well as mildly decreasing returns to education ($\eta = 0.95$). It demonstrates that historically plausible growth trajectories can be obtained without relying on positive exponential growth at the steady-state.

**Just England.** It could be argued that initially, there was just one forerunner of the industrial revolution, England. From there, the take-off to growth spread across the world. To cover this case, we assume that initially, there was only 1 rich country and 99 poor countries. All other parameters are kept from the benchmark model. Recall that in the benchmark case $\bar{c} = 0.33$ such that industrialization is possible (but not inevitable) when there is just one initially growing country. Figure A.4 shows the growth trajectories for such an example. Compared to the benchmark case it takes much longer until the industrial revolution diffuses around the world but otherwise, the evolution of world income is structurally similar.

Figure A.4: Just England