The Knowledge-Diffusion Bottleneck in Economic Growth and Development

Staley, Mark

University of Ontario Institute of Technology

9 June 2018

The Knowledge-Diffusion Bottleneck in Economic Growth and Development

Mark Staley*

June 9, 2018

Abstract

If learning and teaching were easy, the whole world would be developed and frontier knowledge would spread quickly. But learning is difficult and there is a finite capacity for teaching. The concept of a knowledge-diffusion "bottleneck" can be used to explain why there is a negative correlation between per-capita income growth rates and fertility rates, especially amongst countries that are in the early stages of industrialization. It can also explain why income distributions in rich countries have power-law tails (both upper and lower) and why income growth rates are independent of scale. A simple model of knowledge diffusion is calibrated to the observed paths of structural transformation in newly-industrializing countries. The resulting parameters are found to be consistent with the tail exponents seen in U.S. income data, suggesting a common mechanism of knowledge diffusion operating in developing countries and fully-developed countries.

Keywords: Growth, Diffusion, Development, Bottleneck.

*University of Ontario Institute of Technology, Oshawa, Canada. Email: mstaley@uoit.ca. I am grateful to P. Berg and M. Li for comments on an earlier version of the manuscript.

1 Introduction

It is commonly agreed that economic growth is driven by knowledge accumulation. Increases in knowledge lead to increases in productivity. Knowledge accumulation occurs in two steps, the first being the creation of new ideas, and the second being the diffusion of those ideas to other people. Intuition would suggest that the first step is the key to growth, but this paper argues that knowledge diffusion is actually more important. It is important because it leads to the spread of productivity improvements. It is also important because it enables the creation of new ideas, as innovators stand on the shoulders of past innovators. It follows that the speed of diffusion drives the growth rate of an economy.

In the case of developing countries, new ideas likely originate from outside the country and are then spread internally, so it is plausible that growth is driven by knowledge diffusion.[1] In the case of fully-developed countries, the argument for why growth is mainly driven by knowledge diffusion is as follows.

Innovation is difficult. But if many people are making the attempt simultaneously, some of them succeed, even if by luck. Suppose there is a mechanism of knowledge diffusion that favors good ideas. In that case innovations spread, and once they do, other people inevitably discover new ideas that build upon the previous discoveries. This mechanism of growth does not require the existence of geniuses performing heroic acts of innovation. Nor does it require large investments in R&D. Innovation can be small-scale and random. It can even be *noise* (including noise having zero mean impact on productivity). If noise is passed through a filter of rational imitation, the economy grows (Staley, 2011, Luttmer, 2012).[2]

The central message of this paper is that there is a knowledge-diffusion "bottleneck" that determines the growth rates of both industrializing countries and developed countries. To motivate that message, let us look at two empirical facts that are suggestive of a bottleneck. The first is the observed

[1] Jones & Romer (2010) state that "For some reason - perhaps because the diminishing returns to capital in a production function are much easier to measure than the process of technology adoption - the idea explanation for catch-up growth is not as well-established as it probably should be."

[2] The Luttmer and Staley models are based on a model of diffusion developed by Alvarez *et al.* (2008) and Lucas (2009a).

Figure 1: *The number of workers in non-agricultural sectors in China. The best-fit line uses data from 1978 onwards. Source: Penn World Table Version 9.0, ILO, and author's calculations.*

historical pattern of structural change in China (that country being a prominent example of a newly-industrializing country). The second is the shape of the income distribution in the United States (that country being the leading developed country). It will be argued that these two facts are linked by a common bottleneck parameter.

Figure 1 shows the number of workers in non-agricultural sectors in China by year from 1962 to 2014. The numbers have been compiled using employment data from the Penn World Table (Feenstra *et al*, 2015) along with sector composition data from the U.N. International Labour Organization (ILO).[3] The line is slightly convex, and an exponential function fits the data quite well after 1978 (the start of reforms). The estimated growth rate is 3.8% per year. The population growth rate in China during the reform years was on average only 1% per year, so more than two thirds of the growth seen in Figure 1 must have been due to migration of workers from the agricultural sector. That pattern is consistent with the idea of development in a dual-sector economy. Workers migrate from a Traditional Sector (mostly subsistence

[3] www.ilo.org/ilostat.

agriculture) to a Modern Sector (mostly manufacturing and services).[4]

The shape of the curve in Figure 1 is consistent with many convex functions, but the exponential function lends itself to easy interpretation in terms of knowledge diffusion. It implies that the rate of absorption of new workers in the Modern Sector is proportional to the number of workers already there. To see how such a proportionality might arise, consider a Modern Sector consisting of workers and managers. It is plausible that some constant percentage of the workforce is skilled enough to be a manager, and managers have a finite span of control. Hence there is a limit to the number of migrants from the traditional sector that can be hired and trained during a given unit of time. Alternatively, imagine that knowledge transmission occurs in a formal education sector. Some percentage of the modern workforce has the aptitude to be a teacher, and there is a limit to the number of students each teacher can handle. Both of these mechanisms of teaching imply that the rate of entry of migrants should be proportional to the number of workers already in the Modern Sector.[5]

We can formalize the above by writing

$$\frac{\mathrm{d}N_M}{\mathrm{d}t} = (\gamma - \delta)N_M, \tag{1}$$

where N_M is the number of workers in the Modern Sector, γ is the absorption rate, and δ is the mortality rate. The absorption parameter γ captures the rate of education of newborns and migrants together. The average adult mortality rate in China from 1978 to 2014 was $\delta = 0.6\%$ per year.[6] Using the observed growth rate of 3.8% per year, we can rearrange Equation (1) to obtain

$$\gamma = 3.8\% + 0.6\% = 4.4\% \text{ per annum.}$$

[4]Lewis (1954) introduced the dual-sector theory of development. Kuznets (1955) supported that model based on observed income inequality in developing countries. More recently, Young has emphasized the role of migration in explaining the economic data for the Asian Tigers (Hong Kong, Singapore, South Korea and Taiwan) and for China (Young, 1995, 2003; see also Hsieh, 2002 and Krugman, 1994).

[5]This picture is consistent with Lucas (1990) who suggested that knowledge accumulation may be the limiting ingredient in economic development. It is also consistent with the idea that cities act like ports for the importation of knowledge, and there is an urban agglomeration bottleneck that limits the spread of that knowledge to the rest of the country (Lucas, 2009b).

[6]Mortality data is available from data.worldbank.org.

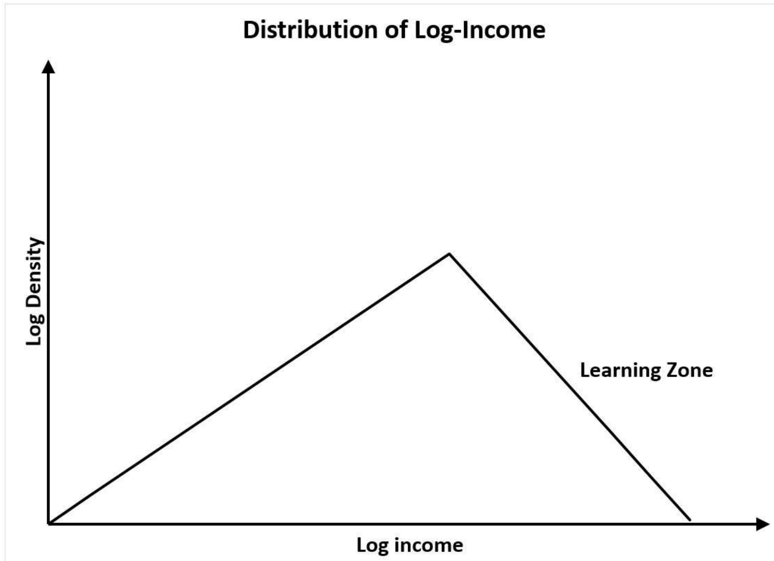

Figure 2: *A stylized distribution of income. Both axes are log scale. The density is the number of workers per unit of log income.*

Of course Equation (1) cannot be true forever because the supply of traditional workers eventually dries up. The parameter γ must eventually decline until it is equal to the birth rate. A more complete model of development must incorporate a process of deceleration as a country becomes fully modern. In other words, diminishing returns to the accumulation of human capital must eventually set in. But in the early stages of industrialization there are, in the words of Lewis (1954), "unlimited supplies of labour", hence no diminishing returns.

Let us now consider a second empirical fact, the shape of the income distribution in the United States. At first glance, the shape of the U.S. income distribution would seem to have nothing to do with the pattern of structural transformation in China. But the two facts can be linked conceptually and numerically using the idea of a knowledge-diffusion bottleneck.

It has been known since the time of Pareto that the right tail of the U.S. income distribution follows a power law with negative exponent (Pareto, 1896). In 1953 Champernowne proposed a model that generates a power law also in the left tail (Champernowne, 1953). Reed (2001, 2003) and Toda (2011, 2012) have recently confirmed Champernowne's prediction. The entire distribution has an asymmetrical "double-Pareto" shape (the left tail is thicker than the right tail). Figure 2 shows a stylized double-Pareto distribution, similar to what is seen empirically.[7]

One way to generate a double-Pareto distribution for income is to combine knowledge diffusion along the lines discussed above for China with random shocks to income. The mechanism works as follows.

Imagine that some percentage of people per unit time wish to upgrade their knowledge. Assuming rationality, they seek the teacher with the most productive knowledge. That teacher has a finite capacity for teaching. Once that teacher fills up his or her "classroom", students seek the teacher with the next most productive knowledge, and so on. A long as the parameters describing this process satisfy some reasonable conditions, the end result is that some upper portion of the income distribution contains enough teachers to accommodate students from across the entire income distribution. All learning then takes place in the upper portion of the income distribution, which we can call the "learning zone".

When random income shocks are added to the above model of diffusion, two things happen. First, the distribution takes on the form seen in Figure 2. The downward-sloping part of the distribution on the right is the learning zone, similar to the Modern Sector in a developing country. The upward-sloping part of the distribution on the left is similar to the Traditional Sector in a developing country.

The second thing that happens is that the distribution moves to the right at a constant speed. Motion is seen even when the mean of the stochastic productivity shocks is zero. That behavior is reminiscent of the behavior described in Staley (2011) and Luttmer (2012, 2015). The distribution of income forms a so-called traveling wave in which the shape is maintained and moves to the right, mimicking economic growth.

One of the parameters of the above-described model is the rate at which workers can be absorbed into the learning zone. It is similar to the parameter γ described earlier. When the model of traveling-wave growth is calibrated to the observed power-law exponents reported by Toda (2014) and to the

[7] See also Toda & Walsh (2015), Toda (2017). König *et al* (2016) show that the distribution of *firm productivity* is also double-Pareto, suggesting a link between income and firm productivity. Alfaranoa *et al* (2012) find that both tails of the distribution of firm profit rates also follow power laws.

observed growth rate of income in the U.S., the absorption parameter takes on the value 4.2% per year. That number is close to the value of γ for China (4.4% per year).

The near coincidence between the absorption rate in the traveling-wave model for the U.S. and the absorption rate in the dual-sector model for China suggests a common mechanism of knowledge diffusion. This paper presents a unified model of growth and development based on a simple mechanism of learning. The model has three parameters: γ, which has already been discussed, a parameter β describing the rate at which people seek new teachers, and a parameter σ describing the volatility of productivity shocks.

A striped-down version of the model using only γ and β is constructed for the purpose of describing knowledge-transmission in a dual economy. Structural transformation occurs because workers migrate from the Traditional Sector to the Modern Sector, picking up knowledge from workers in the Modern Sector. It turns out that over 90% of the historical variance of structural transformation rates in countries that have industrialized since World War II can be explained by this model.

The model predicts that during the early stage of industrialization there is a strong negative correlation between the structural transformation rate (and by extension the growth rate of per-capita income) and the fertility rate. That prediction is supported by the data. China has been helped by the one-child policy. The Philippines has been hindered by high fertility. The intuition is that during the early stage of development there is a severe learning bottleneck in the Modern Sector. If a country's fertility rate is high, many children do not gain access to productive knowledge when they grow up because there are not enough "teachers" available for them in the Modern Sector, so they are stuck in the Traditional Sector. The higher the birth rate, the more young people are stuck, which results in a low rate of income growth at the aggregate level. The model also predicts that once the size of the Modern Sector reaches a certain threshold, the bottleneck is broken and the process of structural transformation is no longer held back by high fertility. That prediction is also supported by the data.

When stochastic productivity shocks are added to the model via the parameter σ, a solution describing steady-state growth is admitted. The income distribution takes on the shape of a double-Pareto distribution and the distribution moves as a solid mass, i.e. as a traveling wave. When the three parameters are calibrated to the shape and speed of the log-income distribution in the U.S., the resulting values of γ and β are close to the values in the dual-economy model. That is the main result of this paper. It suggests that there is a common mechanism of knowledge diffusion driving growth in industrializing countries and fully-developed countries.

A prediction of the steady-state growth model is that the rate of growth of per-capita income is given by

$$g = \sigma\sqrt{2(\gamma - b)}, \tag{2}$$

where b is the fertility rate. Equation (2) predicts a negative correlation between the income growth rate and the fertility rate. The intuition behind that result is that even in fully-developed countries there is a learning bottleneck. The more new people there are seeking teachers, the larger is the learning zone. Hence the vertex of the inverted 'V' in Figure 2 is located further to the left, which in turn means that the average gain in productivity per unit time is lower.

The U.S. has likely been on a steady-state growth path for some time, making it ideal for testing Equation (2) using historical data. It so happens that the growth rate of real per-capita GDP has been gradually increasing since the early 1800s, while the fertility rate has been slowly declining over that same period of time. It is shown in this paper that those two shifts are consistent with Equation (2).[8]

The final prediction of the steady-state growth model is a limited scale effect. One might expect to see a positive correlation between the number of workers and the growth rate of income. The more people there are, the more discoveries are being made. A simulation exercise shows that this is indeed the case for the steady-state growth model, but only for small populations. As the population grows, the scale effect becomes weaker, until it eventually disappears. For modern developed countries, the prediction is that there is no discernible scale effect, in agreement with observation. The intuition for this result is that the finite speed of diffusion puts a brake on the rate of

[8] In a somewhat related vein, Fwyrer (2011) suggests that the entry of the baby boomers into the workforce in the 1970s and 1980s contributed to the slowdown in productivity growth during those decades. He postulates that the slowdown occurred partly because the average quality of management declined as lower-quality managers were required to handle the influx of young workers. That decline is analogous to a leftward shift of the vertex in Figure 2.

absorption of new ideas. As a result, there is some duplication of discoveries, or "stepping on toes", as Bloom *et al.* (2017) put it.

Much of the recent literature on diffusion-based models of growth is inspired by Lucas (2009a), who models heterogeneous agents with differing levels of productivity. Lucas shows that knowledge diffusion between agents can be an engine of growth. His paper has triggered several modifications and elaborations. Two representative examples are Lucas & Moll (2014) and Perla & Tonetti (2014), who expand Lucas's model by incorporating optimal choosing between producing and imitating. In those models, and in Lucas (2009a), growth eventually stops unless the distribution of productivities has an infinitely long right tail. Steady-state growth requires that all productivity-enhancing discoveries have already been made.

Working independently, Staley (2011), Luttmer (2012, 2015) and König *et al* (2016) have introduced stochastic productivity shocks into models of knowledge diffusion, showing that steady-state growth can persist even when the initial distribution of productivity is bounded. In the case of Staley and Luttmer the productivity shocks are Brownian, while in König *et al* the shocks are jumps on a Schumpeterian quality ladder. Hongler *et al* (2016) expand Staley and Luttmer's model by introducing an imitation neighborhood. They show that there is a critical imitation strength above which steady-state growth is possible.

Luttmer (2015) models stochastic innovation and knowledge transmission between students and teachers. Students with different capacities for learning are optimally matched to teachers having different levels of productivity. A simplification of the model leads to a growth formula similar to (2). The present model is similar to Luttmer's model but abstracts from the details of student-teacher assignments and pursues a more phenomenological modeling approach.[9] The present model also incorporates demographics, makes a link with structural transformation in developing countries, and can capture the thick left tail seen in U.S. data. In summary, the main contributions of the present paper are:

- providing an explanation of why high fertility rates hinder the rates of

[9] A *phenomenological* model is based on parsimony and goodness of fit to empirical data rather than on fundamental principles. It is hoped that such a model is consistent with fundamental principles.

structural transformation in developing countries in early years but not in later years,

- providing a explanation of the asymmetric double-Pareto shape of income distribution seen in the U.S.,
- suggesting how the increase in the income growth rate in the U.S. over the last 200 years may have been due to the decline in fertility,
- discussing the bounded scale effect first pointed out by Staley (2011),
- and finally, showing quantitatively that knowledge diffusion is a promising basis for a unified theory of growth and development.

The plan of the paper is as follows. In Section 2, a simple two-parameter model of learning applicable to structural transformation in industrializing countries is presented. The two parameters γ and β are calibrated to cross-country employment-share data, and the demographic implications are explored. Section 3 develops a diffusion-based model of steady-state growth consistent with the structural transformation model, using the same two parameters γ and β, and incorporating stochastic productivity shocks through a third parameter σ. Consistency with the dual-economy model is demonstrated. The demographic and scale implications of the steady-state growth model are explored. Finally, Section 4 provides some concluding remarks.

2 A Dual-Economy Model of Development

This section presents a simple two-parameter model describing the transmission of knowledge between two groups of people. The model is useful for two reasons. First, it builds intuition that will prove useful in constructing a more complex model of income dynamics. Second, it has application to the process of development in a dual-sector economy. The model can explain much of the pattern of structural transformation in countries that have successfully industrialized since World War II.

2.1 The Model

Consider an economy consisting of two sectors. One of the sectors, labeled "Traditional", contains workers that have low productivity and low income. The other sector, labeled "Modern", contains workers that have high productivity and high income. Workers migrate from the Traditional Sector to the Modern Sector, acquiring knowledge from those already in the Modern Sector and thereby increasing their productivity and income.

Consider the situation where the Modern Sector is small compared with the Traditional Sector. Repeating Equation (1) from the Introduction,

$$\frac{\mathrm{d}N_M}{\mathrm{d}t} = (\gamma - \delta)N_M, \tag{3}$$

where N_M is the number of workers in the Modern Sector, γ is the absorption rate, and δ is the mortality rate. N_M increases over time because new workers are born into the Modern Sector, and because migrants arrive there from the Traditional Sector. It is assumed that γ is higher than the fertility rate, so there is excess capacity for the Modern Sector to absorb migrants. The parameter γ represents the *maximum* rate at which new workers can be absorbed. In other words, γ represents a teaching bottleneck. If the supply of new workers is lower than γM_M, then the lower number is absorbed. When the Modern Sector is small, there are more potential migrants than can possibly be absorbed into the Modern Sector, so the teaching constraint is binding and Equation (3) holds. We will assume that γ is the same for all countries, but δ is country-specific.

Let L be the total number of workers in the country. L grows as

$$\frac{\mathrm{d}L}{\mathrm{d}t} = nL = (b - \delta)L,$$

where n is the population growth rate and b is the fertility rate. We are assuming that there is no childhood or retirement in this economy. Alternatively, we are assuming that the number of workers is a fixed fraction of the total population. This simplification serves the purpose of model building, but we will need to be careful when calibrating the model to data.

Define $p = N_M/L$ to be the proportion of workers in the Modern Sector. Combining the above two equations, we have

$$\frac{\mathrm{d}p}{\mathrm{d}t} = (\gamma - b)p, \tag{4}$$

where γ is the same for all countries but b is country-specific. An immediate implication of Equation (4) is that high fertility is detrimental to development, at least when the Modern Sector is small. The reason is basic arithmetic. The more young people there are, the more are turned away from the Modern Sector for want of teachers. At the aggregate level, that translates into a lower rate of growth of p.

Equation (4) cannot be true forever because at some point p will become greater than 1, which is nonsensical. Intuition suggests that as p approaches 1 the supply of workers from the Traditional Sector shrinks, causing a decrease in the rate of migration. We need a model of the Traditional Sector.

Let us assume that the Traditional Sector can absorb new workers, just like the Modern Sector, but the only people being educated there are those born there. Simply stated, nobody in the Modern Sector wants to move to the Traditional Sector so that they can absorb traditional knowledge and suffer a drop in income. Assume that the mortality rate in the Traditional Sector is the same as that in the Modern Sector. Finally, assume that a fixed fraction β of workers in the Traditional Sector are capable of migrating to the Modern Sector during each unit of time.

The last assumption above can be justified in a manner similar to the justification of a teaching bottleneck in the Modern Sector (see Introduction). Let us imagine that in the Traditional Sector there is a system of education. It might be formal, informal or even parental. Before a person can contemplate migrating to the Modern Sector he or she must obtain some level of knowledge from teachers in the Traditional Sector. If the number of teachers in the Traditional Sector is a fixed fraction of the working population there, and each teacher has a finite capacity for teaching, then the number of potential migrants to the Modern Sector per unit time is a fixed faction of the number of workers in the Traditional Sector.

Putting this all together, we have for the Traditional Sector

$$\frac{\mathrm{d}N_T}{\mathrm{d}t} = (n - \beta)N_T, \tag{5}$$

where N_T is the number of workers in the Traditional Sector. The parameter β represents the *maximum* rate at which people can leave the Traditional Sector. If the Modern Sector is small, there may be insufficient capacity there to absorb all potential migrants (as discussed earlier), in which case the rate of migration is lower than above. When the Modern Sector is large enough to absorb all migrants, the above equation applies. Given that $p = 1 - N_T/L$, we then have

$$\frac{\mathrm{d}p}{\mathrm{d}t} = \beta(1-p). \tag{6}$$

Equations (4) and (6) cannot both be true at the same time, except at one particular value of p, which we will call p_I. When p is small, the ability of the Modern Sector to absorb new workers is the binding constraint on the rate of migration. When p is large, the ability of the Traditional Sector to supply new workers is the binding constraint. Putting (4) and (6) together, we have

$$\begin{aligned}\frac{\mathrm{d}p}{\mathrm{d}t} &= \min\left[(\gamma - b)p,\, \beta(1-p)\right] \\ &= \begin{cases} (\gamma - b)p & \text{when } p \le p_I, \\ \beta(1-p) & \text{otherwise,} \end{cases}\end{aligned} \tag{7}$$

where p_I is an inflection point given by $p_I = \beta/(\beta + \gamma - b)$. Equation (7) describes an S-shaped path of economic development. The first stage is one of accelerating migration, and the second stage is one of decelerating migration, or convergence. Note that a condition for structural transformation to get started is $\gamma > b$.

2.2 A Calibration

Figure 1 in the introduction showed a pattern of accelerating structural transformation in China, consistent with Equation (3) and with the upper expression in Equation (7). Figure 3 presents the share of non-agricultural employment in Japan. It shows a pattern of deceleration that looks consistent with the lower expression in Equation (7).

Given the examples of China and Japan, it is tempting to map the traditional and Modern Sectors of the dual-economy model to the agricultural and non-agricultural sectors respectively. Anecdotally, such a mapping seams reasonable given the oft-observed inequality between rural incomes and urban incomes in developing countries. There is a strong negative correlation between overall per-capita income and the percentage of workers in agriculture.[10] But

[10] See for example Figure 11 in Lucas (2009b).

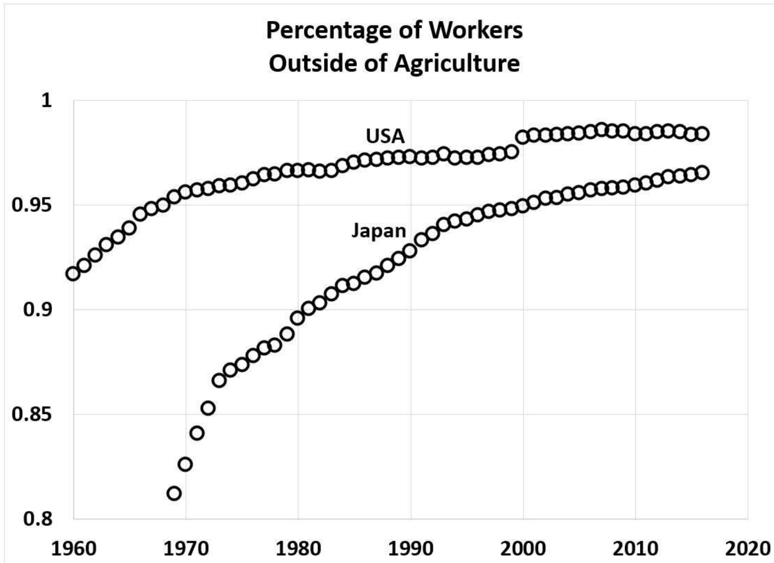

Figure 3: *Percentage of workers outside of agriculture in Japan. Also shown is the curve for the U.S., which is used in Equation (8) to normalize employment shares in the dual-economy model. Source: ILO.*

the mapping cannot be exact. Countries that are fully developed have small agricultural sectors that are far from "Traditional". Agricultural sectors within developing countries contain a mix of subsistence farms and modern farms.[11] In order to calibrate our dual-economy model to employment-share data we need to find a way to translate employment shares into p, taking into account that there is a mix of traditional and modern agriculture.

A simple approach, consistent with an extreme vision of a dual economy, comes to mind. Assume that the Traditional Sector is entirely agricultural. Assume that the Modern Sector is a mini version of the United States, complete with manufacturing, services, and modern agriculture. In particular, assume that the share of employment in agriculture within the Modern Sector is the same as in the United States. These assumptions can be formalized as follows. Define p to be the share of employment in the Modern Sector within a developing country. Define q to be the share of employment in non-agricultural sectors. Finally, define q_{USA} to be the share of employment in

[11] Lewis (1954) offers the image of "highly capitalized plantations, surrounded by a sea of peasants".

non-agricultural sectors in the U.S. Then $q = p\, q_{\mathrm{USA}}$, or rearranging,

$$p = \frac{q}{q_{\mathrm{USA}}}. \tag{8}$$

For a given set of countries, we can use (8) to derive a set of curves $p(t)$. The curve of $p(t)$ for Japan is concave and approaches the value of 1 as the two curves in Figure 3 converge towards each other, reflecting the notion of "catch-up growth" or "convergence growth".[12]

The parameters γ and β in Equation (7) were calibrated using cross-country regressions. The strategy was as follows. Countries were classified as being either in the first stage of development or in the second stage. The method by which this was accomplished is described below. For countries in the first stage, the upper expression in (7) was tested by regressing growth rates of p against fertility rates and checking whether the slope was near -1, as predicted. The parameter γ was then equated with the y-intercept. Growth rates were defined to be the change in log-p over the historical interval, divided by that interval. Fertility rates (net of infant mortality rates) were averaged over the same interval, but lagged by 18 years in order to approximate worker entry rates.

For countries in the second stage of development, the following definition of *Migration Rate* was used:

$$R = \frac{p(t_f) - p(t_i)}{t_f - t_i}, \tag{9}$$

where t_i was the initial year in the historical period and t_f was the final year. Integrating the lower expression in Equation (7) we have

$$R = \frac{1 - e^{-\beta(t_f - t_i)}}{t_f - t_0}[1 - p(t_i)], \tag{10}$$

which says that the migration rate should be a declining linear function of the starting value of p, assuming a common interval of time. To test that prediction, t_i and t_f were fixed for all stage-two countries. Migration rates

[12] The time series of q_{USA} used in Equation (8) is an exponential fit to the share of U.S. workers in agriculture. The exponent is -0.03. It reflects increases in total factor productivity in U.S. agriculture, assuming inelastic demand for food (see Wang & Ball, 2014).

Country	**Class**	**Stage**	**Start p**	**End p**	b	**Interval**
China	NIC	1	0.31	0.56	0.026	1978 - 2005
Thailand	NIC	1	0.32	0.69	0.028	1978 - 2015
India	NIC	1	0.38	0.55	0.031	1991 - 2015
Indonesia	NIC	1	0.41	0.65	0.032	1978 - 2011
Philippines	NIC	1	0.50	0.72	0.035	1978 - 2015
Turkey	NIC	2	0.55	0.81	0.029	1991 - 2015
Brazil	NIC	2	0.75	0.86	0.027	1991 - 2015
Poland	OECD	2	0.77	0.90	0.017	1991 - 2015
Malaysia	NIC	2	0.77	0.89	0.028	1991 - 2015
Mexico	NIC	2	0.77	0.88	0.031	1991 - 2015
Greece	OECD	2	0.80	0.88	0.013	1991 - 2015
Chile	OECD	2	0.83	0.92	0.022	1991 - 2015
Portugal	OECD	2	0.85	0.94	0.014	1991 - 2015
S. Korea	OECD	2	0.86	0.96	0.019	1991 - 2015
S. Africa	NIC	2	0.87	0.95	0.031	1991 - 2015
Latvia	OECD	2	0.88	0.94	0.014	1991 - 2015
Ireland	OECD	2	0.88	0.96	0.018	1991 - 2015

Table 1: *Countries used to calibrate γ and β, in order of starting p. China and Indonesia transitioned to stage two in* 2005 *and* 2011 *respectively. Data for India is only available starting in* 1991*. Source: World Bank, ILO and author's calculations.*

were then regressed against $p(t_i)$, and the slope measured. The parameter β was then backed out of Equation (10). Finally, the fertility rate b was added to the regression equation in order to test the prediction that migration rates are independent of b for stage-two countries.

There are 195 countries in the world, but most countries are undeveloped so are not good candidates for calibrating γ and β. "Western" countries such as those in Western Europe or offshoots such as Canada and Australia are also not good candidates because they underwent industrialization well before employment-share statistics were collected in a systematic fashion. Of the 195 countries, only 17 can be definitively identified as having industrialized during the last few decades. The list of 17 countries includes 10 that are commonly classified as *Newly Industrialized Countries* (NICs), as well as 7 from the list of OECD countries. The countries are listed in Table 1, along with the data used to calibrate γ and β.[13]

Figure 4 shows growth rates of p plotted against fertility rates for stage-one countries. The slope of the solid line in Figure 4 has been set to -1, and the R^2 of the resulting fit is 0.92. The data strongly supports the demographic prediction. The y-intercept of the solid line is 0.47, which we can equate with γ.

Figure 5 shows migration rates for second-stage countries plotted against starting values of p. As expected, the pattern is one of convergence. The lower the starting value of p, the faster the rate of migration, as per Equation (10). The slope of the best fit line is 0.22, which from (10) translates into $\beta = 0.031$. When fertility rates are added to the regression, the resulting coefficient is insignificant ($t = 0.55$, p value 0.59), confirming the prediction that migration rates for second-stage countries are independent of fertility.

There is still the question of how countries were classified as stage one or stage two. Some iteration was required. The criteria for a country to be in stage one is that $p \leq p_I$ for all years where $p_I = \beta/(\beta + \gamma - b)$ (see Equation 7). Similarly, the criteria for a country to be in stage two is that $p > p_I$ for all years. There is some circularity because calibrated values of γ and β depend on the choice of countries, hence the need for some trial and error. For the above set of countries, both criteria are satisfied. China and Indonesia transitioned from stage one to stage two towards the end of the interval. The post-transition data for those two countries is quite short, so was simply discarded.

[13] The list of newly-industrialized countries was taken from en.wikipedia.org/wiki/Newly_industrialized_country, which is based on an intersection of countries listed in Bozyk (2006), Guillen (2003), Waugh (2000) and Mankiw (2007). For OECD countries see www.oecd.org. The years 1991 - 2015 were used for stage-two countries to ensure full data coverage from ILO (www.ilo.org/ilostat) and to have a common interval of time for calibrating Equation 10. The seven stage-two OECD countries used for calibration had values of p less than 0.9 in 1991. The excluded OECD countries ($p > 0.9$) were all advanced "Western" countries, including Australia, Canada, Israel, New Zealand, Japan, and most of the Western European countries apart from Greece, Portugal and Ireland. Also excluded were Taiwan, Hong Kong and Singapore due to a lack of agricultural employment data in ILO. The latter two are city-states so they get a pass. Fertility data was available back to 1960 from the World Bank (data.worldbank.org). The lag of 18 years precluded employment shares earlier than 1978 from being used for stage-one countries.

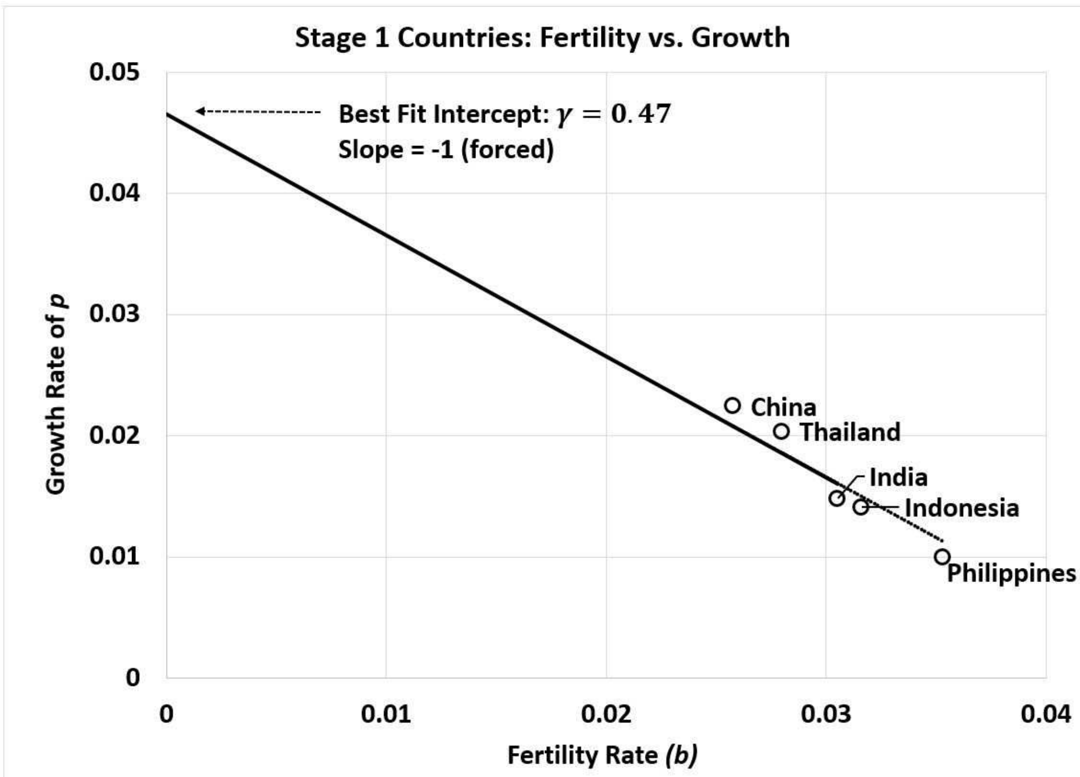

Figure 4: *First-stage countries. The pattern conforms to the upper expression in Equation (7). The R^2 of the fit is 0.92. Source: data in Table 1 and author's calculations.*

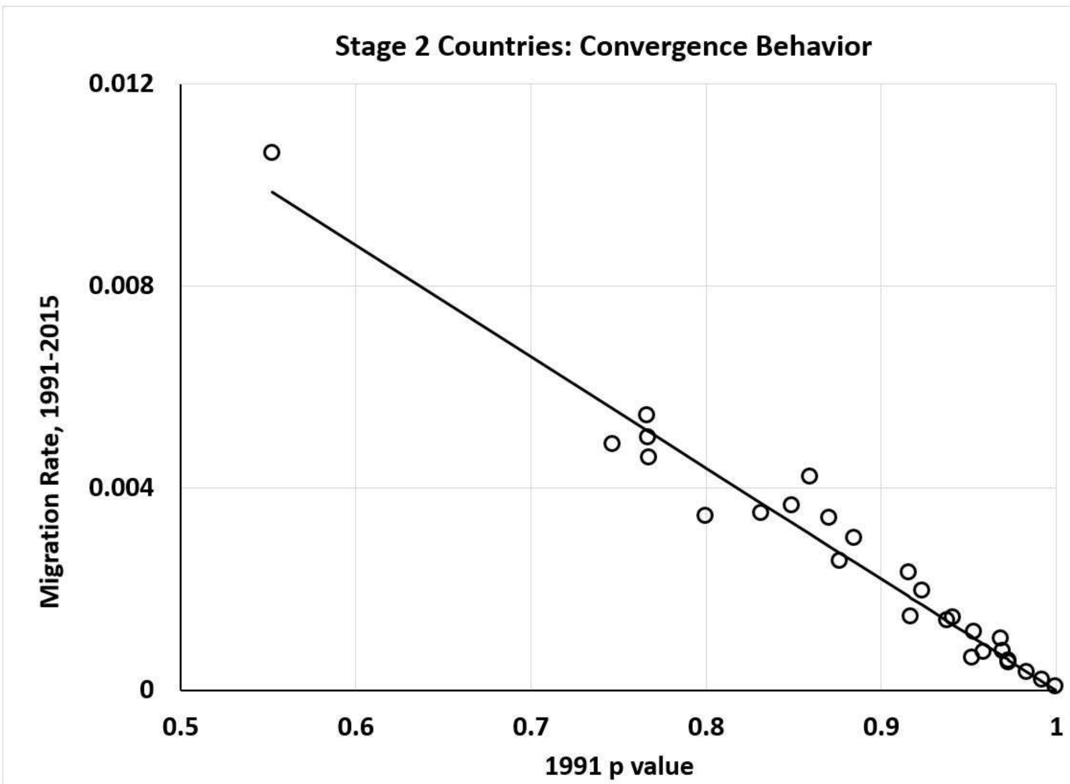

Figure 5: *Migration rates between 1991 and 2015 for second-stage countries. The solid line is a best fit to all points except those with* 1991 p *value greater than* 0.9 *(*$R^2 = 0.91$*). The latter represent "Western" countries and are shown only for information purposes, but nevertheless are surprisingly close to the line. Source: data in Table 1 and author's calculations.*

In summary, the calibrated parameter values are as follows:

$$\begin{aligned} \gamma &= 0.047, \\ \beta &= 0.031. \end{aligned} \tag{11}$$

3 A Model of Income Dynamics

> I was observing the motion of a boat which was rapidly drawn along a narrow channel by a pair of horses, when the boat suddenly stopped - not so the mass of water in the channel which it had put in motion; it accumulated round the prow of the vessel in a state of violent agitation, then suddenly leaving it behind, rolled forward with great velocity, assuming the form of a large solitary elevation, a rounded, smooth and well-defined heap of water, which continued its course along the channel apparently without change of form or diminution of speed. I followed it on horseback, and overtook it still rolling on at a rate of some eight or nine miles an hour, preserving its original figure some thirty feet long and a foot to a foot and a half in height.
>
> John Scott Russell (1845), discoverer of traveling waves.

This section presents a three-parameter model describing the transmission of knowledge between workers having a variety of productivities. It builds on the dual-sector model presented previously, expanding the number of productivity categories from two to infinity (a continuum), and adding stochastic productivity shocks. The latter are required to sustain long-run steady-state growth.

Two of the parameters, γ and β, have the same interpretation as in the dual-sector model. They determine the speed at which knowledge is transmitted between workers. The third parameter σ represents the volatility of productivity shocks. One of the main goals of this section is to test whether the calibrated values of γ and β from Section 2.2 are compatible with the data on U.S. income dynamics and growth. A successful test would support the thesis that there is a common mechanism of knowledge diffusion operating in developing countries and fully-developed countries.

3.1 Empirical Preliminaries

Before presenting the model it is worth examining three qualitative features of the empirical data on U.S. income dynamics that have bearing on the endeavor. The first is that American workers experience a lot of income-growth volatility, some of it transient and some of it persistent. But there is an overall positive tendency, especially for those at the lower-end of the income spectrum (Guvenen *et al*, 2015). This can be interpreted as being due to job changes and learning. Low-income workers have more scope for positive gains associated with learning from others than do high-income workers.

The second feature of the data is that income is positively correlated with age across the entire spectrum of income, suggesting that workers become more productive as they gain experience (Guvenen *et al*, 2015, Karahan & Ozkan, 2013, Toda, 2012). Toda regresses detrended log-income on age as well as on years of education using a Mincer equation and finds that the shape of the residual is independent of age. In particular, young people enter the workforce with the same distribution of log-income as older workers, albeit shifted left due to a lack of experience. We can interpret this finding to mean that young entrants are randomly matched to older workers (e.g. managers) whom they learn from. They earn the same income as their "teachers" less an experience penalty. Their income then evolves deterministically due to experience, and also randomly. The random shocks are partly due to jumps associated with job-change and, presumably, learning from others. Given that the age-dependent income drift does not survive death, it does not contribute to long-run economic growth. So we can ignore the age-dependent component of income dynamics and concentrate on the residual.

The third feature of the data is that the residual has a double-Pareto shape, as shown in Figure 2 in the Introduction (Toda, 2012). A model of income dynamics should be able to explain that shape.

3.2 The Model

Consider an economy in which there is only one factor of production, labor, and income is equated with productivity. There is no land or capital. Let x stand for log-income or log-productivity, and let $\Phi(x, t)$, $x \in (-\infty, +\infty)$ be the density of workers near x at time t. That is, the number of workers having log-productivity between between x and $x + \mathrm{d}x$ at time t is $\Phi(x,t)\mathrm{d}x$. The integral of Φ over x is equal to the total number of workers L.

Let us postulate the following differential equation for Φ:

$$\frac{\partial \Phi}{\partial t} = (\alpha - \beta + i - \delta)\Phi, \tag{12}$$

where α and β are knowledge-diffusion parameters, i is the immigration rate, and δ is the mortality rate. Equation (12) describes the "learning zone", as termed in the Introduction. The terms on the right-hand-side of this equation are now described in turn:

- $\alpha\Phi$ This term represents a teaching bottleneck. Workers arrive through two means. Either they are born and are randomly matched to workers having productivity near x, whom they learn from. Or they arrive from regions of lower productivity and learn from workers with productivity near x. There is a maximum number of new workers that can be absorbed per number of existing workers, regardless of the means by which they arrive, and that maximum is α.

- $-\beta\Phi$ There is flow $\beta\Phi$ of workers exiting the neighborhood of x in order to move to higher productivity regions in x-space. The parameter β is directly analogous to the parameter β used in section 2.1 to quantify the flow of workers out of the Traditional Sector in a dual-sector economy.

- $+i\Phi$ Immigration was not considered in the dual-economy model of development, but is important for countries such as the United States. It is assumed that the distribution of productivity of immigrants is the same as that of natives, so they do not need to be taught upon arrival. This strong assumption can be relaxed, but as we shall see later it seems to be consistent with historical data for the U.S.

- $-\delta\Phi$ There is a constant rate of mortality δ.

Define $\rho(x,t) = \Phi(x,t)/L$ to be the density of workers, normalized to a total mass of 1. The number of workers grows as

$$\frac{\mathrm{d}L}{\mathrm{d}t} = (b + i - \delta)L = nL,$$

where b is the fertility rate and n is the growth rate of workers. Combining the above with (12) we have

$$\frac{\partial \rho}{\partial t} = (\gamma - b)\rho, \tag{13}$$

where $\gamma = \alpha - \beta$. The parameter γ is assumed to be positive. It is directly analogous to the absorption rate of the Modern Sector in the dual-economy model presented in Section 2.1.

Consider the case where $\gamma - b > 0$, as in the Modern Sector of the dual-sector economy. The density is then growing exponentially, which cannot be true for all regions of x, else the condition of normalization will be broken. Equation (13) can only be valid in some limited region of x-space.

It is natural to assume that Equation (13) applies to regions where x is high, because in those regions rational workers will crowd opportunities to improve their productivity. Without specifying the precise process of matching between students and teachers, we can simply postulate that there is some level of log-income, call it m, which divides x-space into two halves. Above m, Equation (13) applies, whereas below m workers are leaving.

For $x < m$ let us postulate the following

$$\frac{\partial \Phi}{\partial t} = (n - \beta)\Phi \quad \Longrightarrow \quad \frac{\partial \rho}{\partial t} = -\beta\rho. \tag{14}$$

The rationale for this equation is as follows. As in the upper region, new workers are randomly matched to existing workers. As well, there is constant mortality, and immigrants arrive with productivity similar to that of natives. These worker flows account for the term $n\Phi$. There is also exit of workers to the region $x \geq m$, which is captured by the term $-\beta\Phi$.

Combining Equations (13) and (14) we have

$$\frac{\partial \rho}{\partial t} = \begin{cases} (\gamma - b)\rho & \text{when}\, x \geq m. \\ -\beta\rho & \text{otherwise,} \end{cases} \tag{15}$$

The dividing point m is determined by the requirement that the flow of workers out of the region $x < m$ must balance the flow into the region $x \geq m$. In other words, the total mass must be conserved:

$$\frac{\partial}{\partial t}\int_{-\infty}^{\infty} \rho(x,t)\,\mathrm{d}x = \int_{-\infty}^{\infty} \frac{\partial \rho}{\partial t}\,\mathrm{d}x = 0. \tag{16}$$

If Equation (15) was all there was to economic dynamics, all workers would eventually attain the maximum productivity, and economic growth would stop. Let us now introduce some noise to keep the economy growing. Assume that workers experience continuous Brownian shocks to their log-productivity or log-income x over a time interval $\mathrm{d}t$ as follows

$$\mathrm{d}x = \sigma\,\mathrm{d}z, \tag{17}$$

where σ is a constant and $\mathrm{d}z \sim N(0, \mathrm{d}t)$. Workers experiences shocks that are independent of other worker's shocks. One might also include a drift parameter in the above but that would just add an exogenous influence that has a trivial influence on the dynamics. The intuition around this equation is that people experience fluctuations in their own productivity, and they also experiences changes in income due to external influences such as fluctuating costs of production.[14] The choice of a Gaussian implies that very large shocks are possible, which seems unrealistic. However simulation exercises show that the exact form of the shocks does not matter to the overall dynamics, and the Gaussian formulation provides analytic tractability.

Stochastic shocks described by (17) can be captured as follows:[15]

$$\frac{\partial \rho}{\partial t} = \frac{\sigma^2}{2} D^2 \rho, \tag{18}$$

where D stands for the partial derivative with respect to x, $D = \frac{\partial}{\partial x}$. Combining (18) with (15) we have the final form of the differential equation

$$\frac{\partial \rho}{\partial t} = \begin{cases} (\gamma - b)\rho + \dfrac{\sigma^2}{2} D^2 \rho, & \text{when } x \geq m(t), \\ -\beta\rho + \dfrac{\sigma^2}{2} D^2 \rho, & \text{otherwise,} \end{cases} \tag{19}$$

where (16) determines the point $m(t)$. Here we have included an explicit time-dependence in $m(t)$ to emphasize that the dividing point changes over time. If one repeats the above derivation, this time assuming that immigrants arrive with no knowledge and need to be taught in the same manner as young entrants, the only change is that b in Equation (19) is replaced with $b + i$.

[14] Harberger (1998) models growth as a process of real cost reduction. He shows that cost increases are just as common as cost decreases.

[15] This is an application of the Fokker-Planck-Kolmogorov equation. See for example Cox & Miller (1996).

3.3 The Dual-Economy Model Revisited

In is now possible to recast the dual-economy model in terms of the income-dynamics model described above. Define the cumulative fraction of workers between X and ∞ to be

$$p_X(t) = \int_X^\infty \rho(x,t)\,\mathrm{d}x.$$

Dropping the subscript X and integrating Equation (19) we have[16]

$$\frac{\mathrm{d}p}{\mathrm{d}t} = \begin{cases} (\gamma - b)p - \dfrac{\sigma^2}{2} D\rho|_{x=X} & \text{when } X \geq m, \\ \beta(1-p) - \dfrac{\sigma^2}{2} D\rho|_{x=X} & \text{otherwise,} \end{cases} \tag{20}$$

where $D\rho|_{x=X}$ is the first derivative of ρ evaluated at X. It has been assumed that $D\rho$ vanishes as $x \to \pm\infty$.

Let us imagine that X divides the log-income space into two sectors. Above $x = X$ is the Modern Sector, and below $x = X$ is the Traditional Sector. If the two sectors have disjoint income distributions, then $\rho(X) = 0$ and $D\rho(x)|_{x=X} = 0$ as well. Then Equation (20) is the same as the equation of dynamics in the dual-economy model (Equation 7 in Section 2.1). Equality of the two expressions in (20) occurs when $X = m$, which happens when $p = \beta/(\beta + \gamma - b)$. That is the transition point between the first stage and second stage of development in the dual-economy model.

The dual-economy model is thus cast in a new light. The Modern Sector is high-income, and the Traditional Sector is low income. When the seeds of a Modern Sector are first planted in a country, m must lie somewhere to the left of X. In that case the upper expression in (20) applies. But when the Modern Sector grows to such an extent that m crosses X, the lower expression in (20) starts to apply and the country enters the second phase of transformation.

In reality, the distributions of income in the Traditional and Modern sectors probably overlap, in which case $D\rho(x)|_{x=X} \neq 0$. The term $\sigma^2/2\, D\rho|_{x=X}$ describes the flow of workers across the boundary due to stochastic effects.

[16] For $X \geq m$ integrate the upper expression in Equation (19). For $X < m$, define $q = 1 - p$ and integrate the lower expression.

Intuition would suggest that this quantity should be small compared to the overall flow of workers from the Traditional sector to the Modern sector.[17]

Given that the dual-sector model can be derived from the model of income dynamics, the calibrated values of γ and β listed in Equation (11) in Section 2.2 should be relevant to other applications of the income dynamics model, including steady-state growth.

3.4 A Steady-State Growth Solution

If one simulates an economy described by Equation (19), starting with an arbitrary initial shape for ρ, one sees an interesting phenomena.[18] After an initial period of time, the distribution rearranges itself into shape that is similar to that seen in Figure 2. Both tails appear to follow power laws, as seen in U.S. income data, and the mode of the distribution appears as a sharp point similar to the residual of the Mincer Equation seen by Toda (2012). The entire mass moves as a solid block, or traveling wave. The mean of the distribution increases at a constant rate while the variance remains fixed. It is worth formulating an analytic description of this traveling wave because it may be a candidate for describing steady-state growth in developed countries.

In Equation (19) there are two regions. Within each region the differential equation is second-order linear. There is a "birth/death" process in each region and a discontinuity separating the two regions. These characteristics remind one of Gabaix's (2009) analysis of power laws, suggesting that there are two power-law solutions, one for each region. We are looking for a traveling wave solution, so let us try the following

$$\rho(x,t) = \begin{cases} A\exp\{-\lambda_{II}(x-gt)\}, & \text{when } x \geq gt, \\ A\exp\{\lambda_I(x-gt)\}, & \text{otherwise,} \end{cases} \tag{21}$$

[17] Numerical simulation using parameters calibrated later in this section confirm that the stochastic boundary term is indeed small.

[18] Create a large number L of workers, say $100,000$. During each unit of time, do the following. Add noise to the log-productivity x of each worker. Randomly select βL students. Find the dividing point m such that α times the mass of workers having $x > m$ (the "learning zone") equals βL. Randomly match teachers in the learning zone with students from across the income spectrum (with the proviso that no student should see a drop in income) setting the values of x of the students to be equal to the values of x of their teachers.

where λ_I and λ_{II} are the two power-law exponents (here written in terms of exponential functions because we are working in log-income space), g is the speed of the wave (the growth rate of income), and A is a normalizing constant. The boundary point is gt, which is the value of x where the two solutions match. Both expressions have the same coefficient A to ensure that the expressions are equal at $x = gt$. A can be determined by the requirement that the integral of ρ over x must be equal to 1, which gives us

$$A = \frac{\lambda_I \lambda_{II}}{\lambda_I + \lambda_{II}}.$$

It can be verified that Equation (16) is then automatically satisfied. Substituting (21) into (19) we get

$$\begin{aligned} 0 &= \gamma - b - g\lambda_{II} + \frac{\sigma^2}{2}\lambda_{II}^2, \\ 0 &= -\beta + g\lambda_I + \frac{\sigma^2}{2}\lambda_I^2. \end{aligned} \tag{22}$$

Assuming that the parameters γ, β, σ and b are known, there are three unknowns, g, λ_I and λ_{II}. But there are only two equations above. For any value of g there is a solution for λ_I and λ_{II}. So any value of g is permitted. However when one simulates the model, a single definite speed is observed!

This type of situation, in which the speed of a traveling wave is mathematically ambiguous, has been the subject of investigation by mathematical biologists and biophysicists working on problems of gene propagation and microbial evolution. It turns out that partial differential equations are not the whole story. Differential equations assume continuity, but in the far reaches of the tails the number of elements (e.g. workers) is small and there are stochastic effects that require a different treatment. The interaction between those stochastic effects and continuous dynamics in the body give rise to stability issues, and there is only one viable speed.[19] The upper tail has received more attention in the literature than the lower tail, so in order to make links with that literature our focus is also on the upper tail.

[19] Fisher, a biophysicist, discusses traveling waves of fitness seen in microbial populations and offers the amusing analogy of the random snuffing of an exploring dog's nose. "The balance between the irregular snuffling and the inertial motion of the body determine the overall speed; yet, as the owner of a large, headstrong dog knows, predicting its speed is very hard!" (Fisher, 2011).

A mathematical trick borrowed from Tsimring *et al* (1996) can be used to express g in terms of the other parameters. Define the following transformed variables

$$\begin{aligned}\tilde{x} &= \frac{\sqrt{2(\gamma - b)}}{\sigma}\, x \\ \tilde{t} &= (\gamma - b)\, t\end{aligned} \tag{23}$$

Substituting these into the upper expression in (19) ($x \geq m(t)$), all the parameters cancel and we end up with

$$\frac{\partial \rho}{\partial \tilde{t}} = \rho + \frac{\partial^2 \rho}{\partial \tilde{x}^2}. \tag{24}$$

In the frame of reference defined by (23) the wave travels at some speed, call it c, which is unrelated to the parameters γ, β and σ. Transforming back to original coordinates (x, t), the speed g must be

$$g = \sigma \sqrt{\frac{\gamma - b}{2}}\, c.$$

It remains to find c.

Equation (24) shows up as a limiting expression for the upper tail of distributions covering disparate nonlinear phenomena having traveling-wave solutions. The most famous example is gene propagation, originally described by Fisher (1937) using what is now known as the Fisher-Kolomgorov equation. Fisher analyzed stochastic behavior in the far right tail of his system, which happened to be described by Equation (24), and found that $c = 2$ was the only speed that leads to stable wave propagation, in which the stochastic fluctuations at the leading edge move at the same speed as the body of the distribution.[20] Fisher presented his argument in a verbally terse manner. It is reproduced in the Appendix in a more explicit mathematical form.

Substituting $c = 2$ into the above formula for g we have

$$g = \sigma \sqrt{2(\gamma - b)}. \tag{25}$$

[20] Kolmogorov *et al.* (1937) were investigating various biological and chemical systems and obtained the same result. Van Saarloos (1987) analyses the stability properties of these systems using modern mathematical techniques. Grindrod (1996) provides a good overview of these types of systems.

A condition for growth is that $\gamma > b$, which was also the condition for structural transformation to occur in the dual economy model. Note that if we had included a drift term in (17) the above formula for the growth rate would have simply included that extra term, but as we will see below, that does not seem to be necessary.

Equation (25) implies that the growth rate of income in developed countries is an increasing function of σ and γ and a decreasing function of b. The positive dependence on σ makes intuitive sense. The higher the standard deviation of shocks, the greater the magnitude of positive shocks, which are the focus of attention of knowledge seekers. The positive dependence on γ also makes intuitive sense because the higher the rate of absorption the faster rate of knowledge transmission. The negative dependence of g on b is a bottleneck effect. When the fertility rate increases, the additional new workers make new demands on teachers, causing the vertex in Figure 2 to be shifted to left (m is lower). The average quality of knowledge transmitted to students is then lower.

3.5 Testing the Steady-State Growth Solution

In this section, the model of steady-state growth is calibrated to U.S. growth and income data, with the goal of testing for consistency with the dual-economy model. The steady-state growth model has four parameters, γ, β, b and σ. The fertility rate b can be directly extracted from U.S. demographic data. The other three parameters can be calibrated to the observed growth rate of income and the Pareto exponents of the U.S. income distribution:[21]

$$\begin{aligned} b &= 0.0185, \\ g &= 0.0206, \\ \lambda_I &= 1.25, \\ \lambda_{II} &= 2.25. \end{aligned} \tag{26}$$

[21]The fertility rate b is the average over the decades 1950s to 1990s from Haines (2008). The growth rate g is the average of growth rates of real income per capita between 1968 and 2014 (adding eighteen years to account for childhood) from the Penn World Table (Feenstra *et al*, 2015). The values of λ_I and λ_{II} are the midpoints of the ranges 1 - 1.5 and 2 - 2.5 respectively from Table 1 in Toda (2014).

Calibrated Parameter	γ	β	σ
Steady-State Growth Model	0.042	0.033	0.096
Dual-Economy Model	0.047	0.031	

Table 2: *Consistency of knowledge-diffusion parameters.*

These values can be used in Equations (22) and (25) to back out σ, γ and β, which are shown in Table 2 (top row).

Table 2 shows a marked degree of consistency in the values of γ and β between the steady-state growth model and the dual-economy model, suggesting a common mechanism of knowledge diffusion driving growth in industrializing countries and the United States. The calibrated value of σ is close to the value of 0.1 estimated by Karahan & Ozkan (2013) to be the minimum volatility of persistent income shocks over the life cycle of U.S. workers.

The steady-state growth model has been assumed to apply to the United States since World War II. Going back further in time, one sees that the growth rate was somewhat lower. The average growth rate of real income during the nineteenth century was about 1.5% per year, while for the twentieth century it was about 2% per year (Maddison, 2009). According to the growth-rate formula (25), an increase in the growth rate would be expected if the fertility rate had been dropping.

Figure 6 shows that the fertility rate has indeed been dropping in the U.S. since at least 1850 (Haines, 1994, 2008). The dashed line shows the total "entry rate", which includes immigration. A back-of-the-envelope calculation shows that the drop in the fertility rate can explain the rise in the growth rate. Differentiating Equation (25) with respect to b, the elasticity of growth with respect to fertility is $-\sigma^2/g$. Using $\sigma = 0.096$ from Table 2, $g = 0.02$ and a slope of $\Delta b = -0.133\%$ per decade (fit to the solid line in Figure 6), the prediction is that the growth rate should have been increasing by 0.059% per decade, or just over half a percent over one hundred years. That estimate is close to the observed increase in g between the nineteenth and twentieth centuries according to Maddison's data.[22]

[22]According to Maddison's data, there is also a negative correlation between growth rates and fertility rates post-WWII across other Western countries. However some of that may be due to capital investment after the war. Even a simple Solow-Swan model predicts a negative correlation between growth rates and fertility rates during the transition to

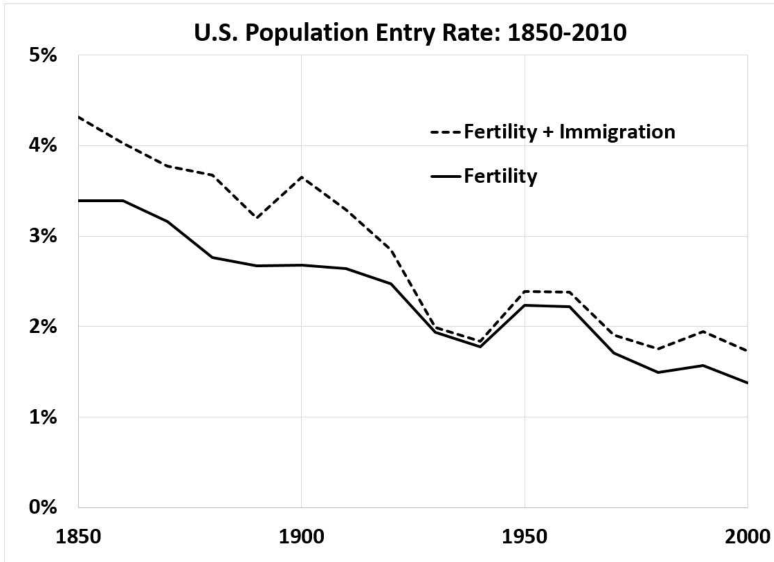

Figure 6: *The decline in U.S. fertility and immigration by decade, 1850-2010. Each point represents the decade forward from that year. Source: Haines (*1994, 2008*).*

Recall that an assumption of the income-dynamics model in Section 3.2 was that immigrants do not have to be taught upon first arrival. An alternative version of the model assumes that immigrants are like young entrants to the job market, in which case the parameter b must include the immigration rate. Repeating the back-of-the-envelope calculation using this alternative specification, one obtains a predicted rate of increase in g of 0.079% per decade, which is too high. Figure 7 shows a more exact comparison of the two alternative model predictions (back-casts). The version of the model that includes immigration seriously underestimates growth rates in the mid-nineteenth century. It seems that immigrants to the United States did not take up space as students in the "classrooms" of their adopted country upon arrival. They were teachers.

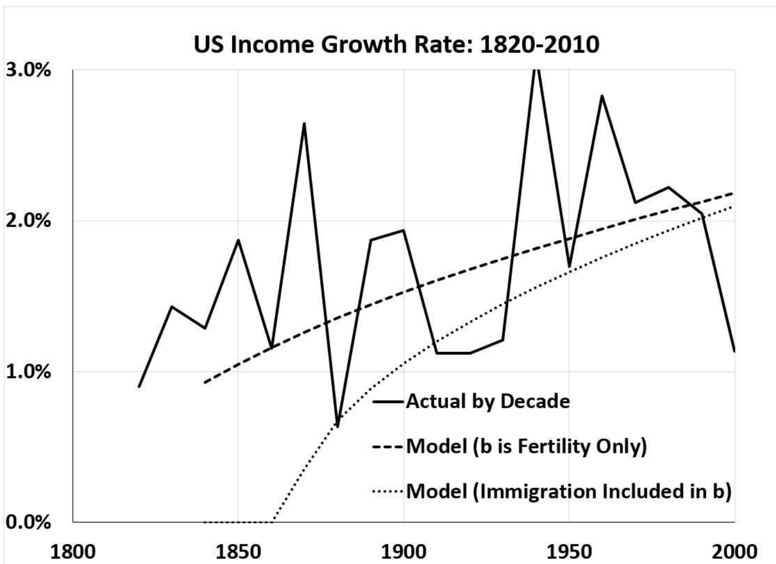

Figure 7: *Testing the growth formula* $g = \sigma\sqrt{2(\gamma - b)}$. *Each point represents the decade forward from that year. The version of model where* b *reflects just fertility is more consistent with the historical data than the version where* b *also captures immigration.*

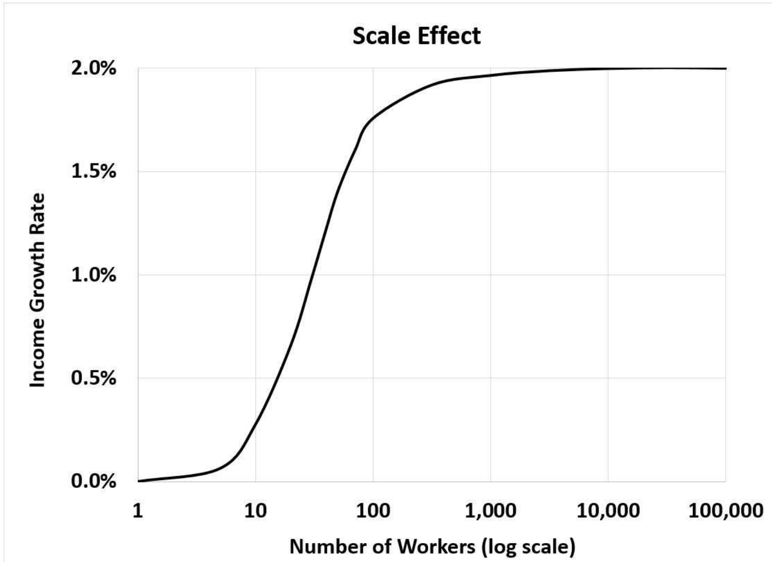

Figure 8: *Growth rate g vs number of workers L. Steady-state parameters are from Table 2. For each choice of L the growth rate is averaged over* 100 *seeds of the random number generator.*

3.6 The Scale Effect

The growth rate formula $g = \sigma\sqrt{2(\gamma - b)}$ is only valid in the continuum limit. When the number of workers is finite the growth rate is lower, so there is a scale effect but it is bounded from above. There are no analytic expressions for these types of scale effects available in the literature, but the effect can be measured using simulation.[23]

Figure 8 shows the result of the simulation exercise. When the number of workers L is small, the growth rate is approximately linear in L. This shows up as a convex curve in Figure 8 because of the log scale. The linearity of g with L is intuitive: the more people there are making discoveries, the higher the growth rate. But when L becomes large, the finite rate of diffusion overwhelms the effect of having more innovators. There is much reinventing of wheels going on, or "stepping on toes" as Bloom *et al.* (2017) put it. The teaching bottleneck puts a brake on the scale effect.

steady-state growth.

[23] Staley (2011) explores some analytic approximations available in the literature.

Scale effects have been a topic of lively discussion in the endogenous growth literature. The earliest versions of endogenous growth models exhibited strong scale effects in which the growth rate is proportional to the level of population. Jones' (1995) critique of those models prompted the development of second-generation models that eliminated that prediction. According to Jones (1999), there are two broad approaches to eliminating the strong scale effect. The first posits a growing number of sectors such that the number of people in each sector is constant. The second assumes that it is getting harder to find new ideas (Bloom *et al.*, 2017, update this argument). The current model suggests another approach to eliminating the strong scale effect, which does not assume an expanding number of sectors, nor an increasing level of difficulty in finding new ideas.

4 Conclusion

In this paper, a unified model of economic growth and development based on knowledge diffusion has been presented. Two of the parameters γ and β describe knowledge diffusion, while a third, σ, describes stochastic productivity shocks. A simplified version of the model containing only γ and β describes the process of structural transformation in developing countries. The two parameters have been calibrated to historical employment-share data covering countries that have industrialized since World War II. A second version of the model containing all three parameters describes steady-state growth. That model has been calibrated to the shape and speed of the U.S. income distribution.

The main result of the paper is that the calibrated values of the knowledge-diffusion parameters γ and β are similar in the two versions of the model, implying a common mechanism of knowledge diffusion operating in developing countries and fully-developed countries. Or turning this around, the model of economic growth presented in this paper is consistent with the expectation that learning behavior is similar across countries.

The model of structural transformation predicts that the rate of transformation is hindered by high fertility during the first (accelerating) stage of migration, reflecting a teaching bottleneck, but not during the second (decelerating) stage. That prediction is supported by the data. The model of steady-state growth predicts a negative correlation between fertility and income growth, which is supported by U.S. data going back to the mid-nineteenth century.

Some possible avenues for future work include the following. On the empirical side, the model could be tested against historical income-distribution data in developing countries, rather than relying on employment share data. Income distribution data for developing countries is sparse in comparison with U.S. data, but is becoming more available (see for example Wu & Perloff, 2004, Sala-i-Martin, 2006, Chotikapanish *et al*, 2007). On the theoretical side, work could be done to endogenize the parameters γ, β and σ in terms of rational behavior. Lastly, there remains the question of why, in some countries, knowledge becomes available to the masses, while in other countries it is hoarded, or does not exist.

A Fisher's Explanation for c=2

Fisher's (1937) explanation for $c = 2$ is rather terse. Given the central importance of his result it is worth spelling out his explanation in some detail. Equation (24) without the tildes reads

$$\frac{\partial \rho}{\partial t} = \rho + \frac{\partial^2 \rho}{\partial x^2}. \tag{27}$$

The power-law solution can be written as

$$\rho = A \exp\left[-\lambda(x - ct)\right],$$

where A is a constant, λ is the power-law exponent, and c is the speed of the traveling wave. Substituting this solution into (27) we have

$$c = \frac{1}{\lambda} + \lambda,$$

which is a U-shaped function with minimum at $\lambda = 1$ corresponding to $c = 2$. Fisher showed that this minimum speed is the only speed consistent with stochastic behavior in the far-right tail.

Let us imagine that at the leading edge some worker leaves the pack due to a stochastic shock to productivity and ends up at $x = x_0$. The density function describing this person is initially a delta-function centered on x_0, but spreads out due to subsequent stochastic shocks. The amplitude of the density also grows at the rate e^t due to the teaching of others. The variance of this localized group of workers grows as $2t$ (recall Equation 18), hence the density evolves as

$$\rho(x,t) = \frac{1}{\sqrt{4\pi t}} \exp\left\{-\frac{(x-x_0)^2}{4t}\right\} e^t. \tag{28}$$

It can be verified that this expression satisfies Equation (27).

Imagine a point x_r located somewhere to the right of x_0 such that the count of workers to the right of x_r is q. Note that x_r is a function of q, $x_r(q)$. The subscript r stands for "right tail". Let us further imagine that as the wave expands and grows in amplitude, x_r moves to the right in such a manner that q stays constant. If the resulting speed of x_r matches the speed of the power-law distribution then the traveling wave will keep its shape. Hence in order to find the stable traveling wave speed c it suffices to find the speed of x_r such that q is constant. The path followed by x_r as a function of t is implicitly defined by

$$q = \int_{x_r}^{\infty} \rho(y,t)\,\mathrm{d}y \quad \text{(for any } q \text{ with } 0 < q < 0.5),$$

where ρ is defined in (28). Substituting $z = (y - x_0)/\sqrt{2t}$ (the distance from x_0 to y in units of standard deviation) we have

$$q = e^t \int\limits_{z_r = \frac{x_r - x_0}{\sqrt{2t}}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\,\mathrm{d}z,$$

which is now expressed in terms of the cumulative standard normal distribution. There is no analytic expression for this integral, but an expansion in $1/z_r$ can be obtained via integration by parts. Rewrite the integral as

$$\int_{z_r}^{\infty} \frac{1}{z} \frac{1}{\sqrt{2\pi}} z e^{-z^2/2}\,\mathrm{d}z.$$

Integrating by parts this becomes

$$\frac{1}{z_r\sqrt{2\pi}} e^{-z_r^2/2} - \int_{z_r}^{\infty} \frac{1}{z^2} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\,\mathrm{d}z.$$

The second term above can be further integrated by parts to obtain a term in $1/z_r^2$, and so on. If we restrict attention to the far-right tail then to first order in $1/z_r$ we have

$$q = e^t \frac{1}{z_r\sqrt{2\pi}} e^{-z_r^2/2}.$$

Substituting $z_r = (x_r - x_0)/\sqrt{2t}$ and taking logarithms,

$$0 = 4t^2 - (x_r - x_0)^2 - 4t\ln(\sqrt{2\pi}q) + 2t\ln(2t) - 4t\ln(x_r - x_0).$$

As x_r and t become large only the first two terms on the RHS of the above expression survive, hence

$$x_r \to x_0 + 2t,$$

so the speed approaches 2.

This finding is not dependent on the choice of q ($0 < q < 0.5$) and, hence, the entire profile for $x_r > x_0$ moves at this speed. A similar argument can be made for $x_r < x_0$. Therefore, an initial profile $\rho(x, 0)$ will approach a traveling wave as time grows, moving with speed $c = 2$.

References

[1] Alfaranoa, S., Milakovic, M., Irlex, A., Kauschkec, J. (2012) A Statistical Equilibrium Model of Competitive Firms. *Journal of Economic Dynamics and Control* **36**: 136-149.

[2] Alvarez, F., Buera, F., Lucas, R. (2008) Models of Ideas Flows. NBER Working Paper no. 14135.

[3] Bloom, N., Jones, C., Van Reenen, J., Webb, M. (2017) Are Ideas Getting Harder to Find? NBER Working Paper No. 23782.

[4] Bozyk, P. (2006). *Newly Industrialized Countries.* Globalization and the Transformation of Foreign Economic Policy. Ashgate Publishing, Burlington, Vt.

[5] Champernowne, D.G. (1953) A Model of Income Distribution. *The Economic Journal* **63**: 318-351.

[6] Chotikapanish, D., Rao, D., Tang, K. (2007) Estimating Income Inequality in China using Grouped Data and the Generalized Beta Distribution. *The Review of Income and Wealth* **53**: 127-147.

[7] Cox, D., Miller, H. (1995) *The theory of stochastic processes.* Chapman and Hall, London.

[8] Feenstra, R. C., Inklaar, R., and Timmer, M. (2015), The Next Generation of the Penn World Table. *American Economic Review*, **105(10)**: 3150-3182. Available for download at www.ggdc.net/pwt.

[9] Feyrer, J. (2011) The US productivity slowdown, the baby boom, and management quality. *Journal of Population Economics* **24**: 267-284.

[10] Fisher, D.S. (2011) Leading the Dog of Selection by its Mutational Nose. *Proceedings of the National Academy of Sciences* **108**: 2633-2634.

[11] Fisher, R.A. (1937) The Wave of Advance of Advantageous Genes. *Annals of Eugenics* **7**: 335-369.

[12] Gabaix, X. (2009) Power Laws in Economics and Finance. *Annual Reviews of Economics* **1**: 255-293.

[13] Grindrod, P. (1996). *The Theory and Applications of Reaction Diffusion Equations: Patterns and Waves.* Oxford University Press, Oxford.

[14] Guillen, M (2003). *Multinationals, Ideology, and Organized Labor.* The Limits of Convergence. Princeton University Press, Princeton.

[15] Guvenen, F., Karahan, F., Ozkan, S., Song, J. (2015) What Do Data on Millions of U.S. Workers Reveal about Life-Cycle Earnings Risk? NBER Working Paper, No. 20913.

[16] Haines, M. (1994) The Population of the United States, 1790-1920. NBER Historical Paper #56.

[17] Haines, M. (2008) Fertility and Mortality in the United States. *EH.Net Encyclopedia*, edited by Robert Whaples, March 19. URL http://eh.net/encyclopedia/fertility-and-mortality-in-the-united-states/

[18] Harberger, A. (1998) A Vision of the Growth Process. *The American Economic Review* **88**: 1-32.

[19] Hsieh, C. (2002) What Explains the Industrial Revolution in East Asia? Evidence from the Factor Markets. *The American Economic Review* **92**: 502-526.

[20] Hongler, M., Gallay, O., Hashemi, F. (2016) Impact of Imitation on the Dynamics of Long Wave Growth. Working Paper.

[21] Jones, C. (1995). Time Series tests of Endogenous Growth Models. *Quarterly Journal of Economics* **110**: 495-525.

[22] Jones, C. (1999) Growth: With or Without Scale Effects? *American Economic Review* **89**: 139-144.

[23] Jones, C., Romer, P. (2010). The New Kaldor Facts: Ideas, Institutions, Population, and Human Capital. *American Economic Journal: Macroeconomics* **2**: 224-245.

[24] Karahan, F., Ozkan, S. (2013) On the persistence of income shocks over the life cycle: Evidence, theory, and implications. *Review of Economic Dynamics* **16**: 452-476.

[25] Kolmogorov, A., Petrovsky, I., Piscounov (1937) Etudes de l'equations de l diffusion avec croissance de la quantite de matiere et son application a un problem biologique. *Moscou Univ. Bull. Math. A.* **1**: 1-25.

[26] König, M.D., Lorenz, J., Zilibotti, F. (2016) Innovation vs Imitation and the Evolution of Productivity Distributions. *Theoretical Economics* **11**: 1053-1102.

[27] Krugman, P. (1994) The Myth of Asia's Miracle. *Foreign Affairs* **73**: 62-78.

[28] Kuznets, S. (1955) Economic Growth and Income Inequality. *The American Economic Review* **45**: 1-28.

[29] Lewis, W.A. (1954) Economic Development with Unlimited Supplies of Labour. *The Manchester School* **22**: 139-191.

[30] Lucas, R. (1990) Why Doesn't Capital Flow from Rich to Poor Countries? *The American Economic Review* **80**: 92-96.

[31] Lucas, R. (2009a) Ideas and Growth. *Economica* **76**: 1-19.

[32] Lucas, R. (2009b) Trade and the Diffusion of the Industrial Revolution. *American Economic Journal: Macroeconomics* **1**: 1-25.

[33] Lucas, R. Moll, B. (2014) Knowledge Growth and the Allocation of Time. *Journal of Political Economy* **122**: 1-51.

[34] Luttmer, E. (2012) Eventually, Noise and Imitation Implies Balanced Growth. Working Paper 699, Federal Reserve Bank of Minneapolis.

[35] Luttmer, E. (2015) An Assignment Model of Knowledge Diffusion and Inequality. Staff Report 509, Federal Reserve Bank of Minneapolis.

[36] Maddison, A. (2009) *The Contours of the World Economy 1-2030 AD.* Oxford University Press, Oxford. Data available from http://www.ggdc.net/maddison.

[37] Mankiw, G. (2007) *Principles of Economics, 4th Edition.* Thomson Southwestern, Mason, OH.

[38] Pareto, V. (1896) Cours d'Économie Politique. F. Rouge, Lausanne.

[39] Perla, J., Tonetti, C. (2014) Equilibrium Imitation and Growth. *Journal of Political Economy* **122**: 52-76.

[40] Reed, W.J. (2001) The Pareto, Zipf and Other Power Laws. *Economics Letters* **74**: 15-19.

[41] Reed, W.J. (2003) The Pareto Law of Incomes - an Explanation and an Extension. *Physica A* **319**: 469-486.

[42] Russell, J.S. (1845) Report of the fourteenth meeting of the British Association for the Advancement of Science; York. September 1844, 311-390, John Murray, London.

[43] Sala-i-Martin, X. (2006) The World Distribution of Income: Falling Poverty and Convergence, Period. *Quarterly Journal of Economics* **121**: 351-397.

[44] Staley, M. (2011) Growth and the Diffusion of Ideas. *Journal of Mathematical Economics* **47**: 470-478.

[45] Toda, A.A. (2011) Income Dynamics with a Stationary Double Pareto Distribution. *Physical Review E* **83** (4): 046122.

[46] Toda, A.A. (2012) The Double Power Law in Income Distribution: Explanations and Evidence. *Journal of Economic Behavior and Organization* **84**: 364-381.

[47] Toda, A.A. (2014) Incomplete market dynamics and cross-sectional distributions. *Journal of Economic Theory* **154**: 310-348.

[48] Toda, A.A. (2017) A Note on the Size Distribution of Consumption: More Double Pareto than Lognormal. *Macroeconomic Dynamics* **21**: 1508-1518.

[49] Toda, A.A. Walsh, K. (2015) The Double Power Law in Consumption and Implications for Testing Euler Equations. *Journal of Political Economy* **123**: 1177-1200.

[50] Tsimring, L., Levine, H., Kessler, D. (1996) RNA Virus Evolution via a Fitness-Space Model. *Physical Review Letters* **76**: 4440-4443.

[51] Wu, X., Perloff, J. (2004) China's Income Distribution Over Time: Reasons for Rising Inequality. UC Berkeley Working paper.

[52] van Saarloos, W. (1987) Dynamical Velocity Selection: Marginal Stability. *Physical Review Letters* **58**: 2571-2574.

[53] Wang, S., Ball, E. (2014) Agricultural Productivity Growth in the United States: 1948-2011. *Amber Waves*, United States Department of Agriculture. February 3.

[54] Waugh, D. (2000) *Manufacturing industries (chapter 19), World development (chapter 22).* Geography, An Integrated Approach (3rd ed.). Nelson Thornes, Oxford.

[55] Young, A. (1995) The Tyranny of Numbers: Confronting the Statistical Realities of the East Asian Growth Experience. *The Quarterly Journal of Economics* **110**: 641-680.

[56] Young, A. (2003) Gold into Base Metals: Productivity Growth in the People's Republic of China during the Reform Period. *Journal of Political Economy* **111**: 1220-1261.