China’s Growth Potential in Comparison with the US and the European Union: a Knowledge Production Perspective

Jinghai Zheng
Department of Economics
University of Gothenburg
Sweden

Preliminary, June 2010

Abstract
Although China’s recent pattern of extensive growth is not sustainable in the long run, we believe that China will be able to sustain a growth rate of 8 to 9 percent for an extended period if it moves from extensive to intensive growth. On the other hand, the growth trend in the US and EU has been in the range of 2-3% since the early 1970s and may prevail for the next decade or two. We discuss the differences in production structure and level of development across the three economies that may explain the countries’ varied intermediate-term growth prospects.

Since the mid-1990s, efforts by the US through ICT and financial innovation to achieve high growth have resulted in a worldwide credit crunch and global economic recession. The situation cast significant doubt on the theory of “New Economy” and presented challenges to the industrial nations for future economic growth. The knowledge production function framework we recommend here may open new avenues for considering the relationship between innovation and long run growth. We expect that the framework would help draw useful lessons from the financial crisis, provide analytical instruments for developing countries in strategic planning, and improve our understanding about the success of the Chinese economic development.

Key words: Chinese economy, the US economy, EU Economy, growth potential, neoclassical growth theory, endogenous growth, knowledge production function
1. Introduction

Having experienced the dramatic development in manufacturing and trade during the last three decades, China is expected to overtake Japan and become the second largest economy just behind the United States in 2010. If this trend continues, China’s GDP may catch up with that of the US in about two to three decades. Because China has been deeply integrated into the world economic system and its success and failure now affect the well-beings of the other nations in the world, whether China can sustain its rapid economic development has become an intriguing issue for both policy makers and academic researchers.

In this study we try to elaborate the findings in a previous study concerning the growth of “potential” in China in comparison with those in the US and EU. We noted that structural characteristics, rapid accumulation in capital stock, and improvement in labor quality were the major factors behind China’s phenomenal economic growth. China’s future TFP growth was also assumed to be faster than that of the US and EU, since it enjoys growth in the stock of world knowledge that can be accessed at affordable price and that enhances production possibilities (Prescott 2002).

We therefore suspected that differences in production structure and in the levels of development were the main reasons as to why growth miracles are recent phenomena as questioned in Prescott (1998). Examples include the East Asian NICs, to some extent the post WWII Japan and Germany, and the Soviet Union both in the between war period and the early years of the cold war. More recently due to the rapid industrialization process China has started to join the ranks of the high performers of East Asia nations. An understanding of the causes and the conditions about how economic miracles took place may provide useful experience to other developing countries. Differences in production structure and in the level of development may also provide explanations as to why productivity slowed in the US and EU from the early 1970s, and how US productivity has surged for ten years from the mid-1990s while productivity stagnated in Europe.

The recent financial crisis seems to demonstrate that concepts such as rationality, optimization, perfect competition, and most of all - "equilibrium" are important and should have been taken more seriously in policy analysis. Thus, the essential ingredients of the mainstream microeconomics were considered to be those related to the general equilibrium model. The discussions revolved around the concept of "equilibrium" are particularly useful. Interesting issues include whether the equilibrium exists, and if so whether it's unique, stable and how to get there. Especially, as the benchmark model in modern economics, the competitive equilibrium framework possesses attractive properties.

In terms of economic applications, the neoclassical growth model of Solow, which is still considered as the core of modern growth theory in the literature (McAdam and Allsopp 2007), has been rather successful in economic planning and in the analysis
about the sources of economic growth. Solow once said that he had Mao’s China on his mind when the model was conceived (Solow 2007). While imposing a basic structure on a macro economic model of general equilibrium nature, the Solow framework can be applicable to both well-developed market economies and planned economies (Solow 2001). The accuracy and precision of Solow’s approach to aggregate productivity analysis were rather striking. Examples in the literature include the growth accounting studies on the Soviet economy before the fall of Berlin Wall (Ofer 1987); Productivity studies on East Asia tigers were also well-known (Kim and Lau 1994, and Young 1995); and a more recent example is a series of studies by Robert Gordon on the US economy (e.g., Gordon 2000).

On the other hand, the 1980s were marked with the emergence of endogenous growth literature and increasing interests in the impact of R&D expenditure on productivity performance of firms and industries, while studies on productive efficiency experienced explosive expansion. Conceptually the three branches of literature are well connected. The endogenous growth model studies the effects of knowledge accumulation on the productivity growth of the aggregate economy in the long run; the studies on innovation-productivity link focus on the empirical estimation of the return to R&D; and the literature of productive efficiency specializes in models that measure technical efficiency and technical progress.

Although the size of the literature that estimates productive efficiency is huge, empirical investigations on the determinants of the productivity performance based on these estimates appear to be difficult and the relationships between explanatory and explained variables are not well structured. The endogenous growth models paid more attention to mathematical formulations, but the structures employed so far are often restrictive and simplistic, for example as in Romer (1990), Aghion and Howitt (1992), and Jones (1995). One of the purposes of the study is therefore to sort out the potential microstructures that can be formulated regarding the relationship between innovative inputs and total factor productivity.

The basic idea of this article is to propose in a formal manner an analytical framework that can be established when considering the relationship between innovation and productivity. The concept of interests might be naturally referred to as the “knowledge production function” as in Romer (1990). However, our formal presentation was inspired to a large extent by the “technology function” of Phelps (1966). When approaching the issues involved through the aspect of the standard production theory, the advantage is that the basic properties required of ordinary production function can be employed to infer the microstructure of the knowledge acquisition process. This approach is also attractive in an empirical sense that the techniques of applied productivity analysis might be effective in investigations on the determinants of innovation and economic growth.
2. Comparing China’s Growth Potential with the US and EU

The literature prior to current worldwide credit crunch has documented many examples that foresaw the looming economic crisis years before (e.g., Gordon 2005, Phelps 2004, Stiglitz 2002, and Brenner 2000 and 2004). Robert Gordon’s application of the growth accounting framework to the study of the US productivity revival and slowdown stands out as convincing evidence that economic theory can be a powerful instrument in empirical analysis for macroeconomic planning.

Since the publication of Solow’s seminal work on technical progress and the aggregate production function, the practice of growth accounting has been used to assess the economic performance of the former Soviet Union (Ofer 1987), in raising concerns about the sustainability of the economies of the East-Asian tigers just a couple years prior to the Asian financial crisis (Young 1995, Kim and Lau 1994, and Krugman 1994), and recently in alarming planners about the macroeconomic imbalances in China (Zheng and Hu 2006).

Although often used as a tool for economic planning, growth accounting and its related concept of TFP are not well understood (Zheng 2008). Some economists believe that the framework as an empirical tool lacks theoretical foundations, although it has been used frequently for economic forecasting. In many cases, applications have been rather mechanical and results were not interpreted properly. Some critics even regarded growth accounting as practically useless or irrelevant. However, if adequately implemented and understood, growth accounting can be a rather useful instrument in improving analysis of growth experience for many countries and regions. Plenty examples in the literature show that growth accounting methods are sensitive enough to pick up significant changes in productivity performance if parameters of production are carefully chosen.

Growth accounting decomposes growth in output into its components as follows.

\[
\frac{\dot{Y}}{Y} = \frac{\dot{A}}{A} + \alpha \frac{\dot{K}}{K} + (1 - \alpha) \frac{\dot{L}}{L}
\]  

(2.1)

where \(\dot{Y}\) stands for GDP and \(\dot{Y}\) for change in GDP over time, \(K\) represents capital stock and \(\dot{K}\) the change in capital stock, and labor is denoted as \(L\) and \(\dot{L}\) the change in labor input; TFP growth is represented similarly by \(\dot{A}/A\); and \(0 < \alpha < 1\) is the output elasticity with respect to capital, \((1 - \alpha)\) the output elasticity associated with labor. Knowledge of growth potentials with its components and output elasticity will provide estimates of potential growth in GDP with reference to a specific country. Differences in output elasticity reflect structural differences among different
economies. A typical growth accounting structure for China can be represented as follows (Chow and Li 2002, and Chow 2008).

\[
\frac{\dot{Y}}{Y} = \frac{A}{A} + 0.6 \frac{\dot{K}}{K} + 0.4 \frac{\dot{L}}{L}
\] (2.2)

for the US (CBO 2001)

\[
\frac{\dot{Y}}{Y} = \frac{A}{A} + 0.3 \frac{\dot{K}}{K} + 0.7 \frac{\dot{L}}{L}
\] (2.3)

and for EU (Mossu and Westermann 2005)\(^1\)

\[
\frac{\dot{Y}}{Y} = \frac{A}{A} + 0.4 \frac{\dot{K}}{K} + 0.6 \frac{\dot{L}}{L}
\] (2.4)

China has output elasticity for capital of 0.6 in comparison with 0.3 for the US. Differences of this magnitude are large enough to generate significant differences in growth potentials between the two economies. For example, a 10% growth in capital stock will enable China to grow by at least 6% per year while the growth rate will be only 3% for the US if everything else is kept constant. Growth difference can also be due to differences in investment in physical capital as well as in TFP growth. For developing economies such as China, investment opportunities are plenty due to its relatively low level of development as opposed to the level prevailed in the US and EU industrialized countries. The potential for absorbing useful information from the existing world knowledge pools is also the greatest in China, while developed economies especially the US will have to rely on new knowledge and innovations in order to shift its production frontier.

The growth accounting framework provides a compact formula for the study of potential output growth. We define potential output as the highest level of real GDP that can be sustained over the period of interests. Growth associated with potential output can therefore be termed sustainable growth. We divide sustainable growth into three categories in this study according to the different timeframes under discussion. In this section, given the structural differences we calibrate the US and EU economy in comparison with a typical scenario for the Chinese economy. It demonstrates that growth potentials across the three major economies vary because of differences in production structure, the level of development, and opportunities for absorbing foreign technologies. Growths in developed countries are mainly relying on technological innovations since investment opportunities are far less than in developing countries. Since technology development often presents patterns of cyclical fluctuations, attempts to counterbalance business cycles or alter the trajectory of growth potentials may result in short-term gains but long-term losses.

\(^1\) Proietti, Musso, and Westermann (2007) took 0.35 as capital elasticity and 0.65 for labor.
Understanding of this is crucial for central banks to carry out sound monetary policies and to prevent financial crisis from happening in the future.

By now it is clear that China may have benefitted from an extensive growth pattern in the intermediate term but is not sustainable in the long run. China may still enjoy a high growth rate of 8 to 9% even if it manages the transformation from extensive to intensive growth.

Taking a longer-term perspective, China will be able to keep its momentum as a rapidly developing economy well into the next two decades or so while the US and EU may only manage a growth rate of 2 to 3% as calibrated in Table 1. Structural differences are one of the main reasons explaining the large differences in growth potentials between China on the one hand and the US and EU on the other. It accounts for almost one half of the contribution from the capital if compared with the US case. The level of development provides even greater opportunities for the Chinese economy than for those of the US and EU. Investment opportunities in China can be twice as much as those in the US, and the potentials for China to absorb new technologies from the developed nations are double of those in the US and EU.

Table 1 Growth Projections (2009-2030)

<table>
<thead>
<tr>
<th></th>
<th>Capital stock</th>
<th>Labor</th>
<th>TFP</th>
<th>GDP</th>
</tr>
</thead>
<tbody>
<tr>
<td>China</td>
<td>8%</td>
<td>3%</td>
<td>2.5%</td>
<td>8%</td>
</tr>
<tr>
<td>US</td>
<td>4%</td>
<td>0.5%</td>
<td>1.2%</td>
<td>3.1%</td>
</tr>
<tr>
<td>EU</td>
<td>3%</td>
<td>0.0%</td>
<td>1.0%</td>
<td>2.2%</td>
</tr>
</tbody>
</table>

Note: output elasticity of both capital and labor is assumed as of 0.5 for China, 0.4 for EU, and 0.3 for the US.

Moreover, shortage of labor as another important input to the production process will prevent the developed economies from growing faster in the intermediate term.  

---

2 The growth rate in Table 1 is somewhat too optimistic according to what the economists in the US would like to agree. “[M]ainstream economists are exceptionally united right now around the proposition that the trend growth rate of real gross domestic product (GDP) in the United States — the rate at which the unemployment rate neither rises nor falls — is in the 2 percent to 2.5 percent range” (Blinder 2002).

3 Sterman (1985) presented a behavioral model of the economic long wave, which showed that “capital self-ordering” was sufficient to generate long waves. By capital self-ordering Sterman (1983) refers to the fact that the capital-producing sector must order capital equipment such as large machinery to build up productive capacity. Investment expansions in the 1950s and 1960s accumulated large excess capacity in the US and EU countries. “But while stimulating basic research and training the labor force for “new-wave” technologies are important, innovation alone will not be sufficient to lift the economy into a sustained recovery
China will face similar problem as encountered in developed countries today as its population ages in twenty years. Demographic change due to China’s baby boomers of 1960s and 1970s entering retirement age may bring significant impact on both labor supply and the country’s capacity to save and invest.

Table 2 Productivity slowdown in Soviet Union, US, and EU

<table>
<thead>
<tr>
<th></th>
<th>GDP</th>
<th>Capital</th>
<th>Labor</th>
<th>TFP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Soviet Union</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1950-1970</td>
<td>5.4</td>
<td>8.8</td>
<td>1.8</td>
<td>1.6</td>
</tr>
<tr>
<td>1970-1985</td>
<td>2.7</td>
<td>7.0</td>
<td>1.1</td>
<td>-0.4</td>
</tr>
<tr>
<td>US</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1950-1972</td>
<td>3.94</td>
<td>2.60</td>
<td>1.35</td>
<td>1.64</td>
</tr>
<tr>
<td>1972-1996</td>
<td>3.30</td>
<td>3.08</td>
<td>1.70</td>
<td>0.55</td>
</tr>
<tr>
<td>1996-2004</td>
<td>3.62</td>
<td>2.64</td>
<td>0.65</td>
<td>1.49</td>
</tr>
<tr>
<td>EU (Euro zone)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1960-1973</td>
<td>5.1</td>
<td></td>
<td>4.8</td>
<td>3.2</td>
</tr>
<tr>
<td>1973-2003</td>
<td>2.2</td>
<td>0.5</td>
<td>2.8</td>
<td>1.0</td>
</tr>
</tbody>
</table>


From a longer-term perspective, although economic prosperity ultimately depends on innovation-driven productivity growth, there are signs that even worldwide innovation activities might have been ineffective in recent decades. There is a literature on diminishing technological opportunities since the early 1960s, and recent studies on endogenous growth have also discussed related issues intensely (Jones 1999, Segerstrom 1998, and Kortum 1997). More recently, Gordon in a series of articles addresses the issue in terms of demand creation for new products and technological advances (e.g., Gordon 2004) and suggests that the US productivity revival since 1995 might not be sustainable (see Table 2). This would mean that the productivity slowdown in the developed world since the early 1970s may continue into the next decade or so. Given the input constraints on potential output growth in the US and EU, productivity is left as the only source of extra growth.

In this regard, historical lessons from the former Soviet Union need to be taken seriously. Soviet growth was spectacular to the extent that its industrial structure changed dramatically from an economy with 82% rural population and most GNP produced in agriculture to one that is 78 percent urban with 40-45% of GNP originating in manufacturing and related industries (Ofer 1987). The growth achievement was largely due to the pattern of extensive growth that lasted virtually as long as excess capacity in basic industries continues to depress investment. (Sterman 1983)"
for a period of 70 years from the late 1920s to the mid-1980s. By 1970, Soviet total factor productivity growth was zero and negative ever since (Table 2). Although the problem in western countries nowadays is different in nature in the sense that their patterns of growth have not been every extensive with the rate of growth in capital stock of 3 to 4%, the growth in TFP has been worrisome.

This situation has important implications for macroeconomic planning. A straightforward strategy to boost productivity growth is of course to increase spending on R&D activities. However, as many policy makers would like to believe that the R&D activities with regard to information and computer technologies (ICT) may benefit an economy in the long run, but one has to take into account the lag between the emergence of a new technology and the generation of sufficient demand when managing the macro economy. For example, the US economy has recorded impressive productivity growth since the mid-1990s thanks to innovations and massive investments in ICT. But the on-going financial crisis may dramatically alter the interpretation of the US productivity boom of the past decade. Some critics suggest that the problem lies in the desire to maintain growth above what is sustainable by encouraging excessive investment in technology and cutting loose regulations for risky innovation in the financial sector. This amounts to taking the concept of “potential output” more seriously as far as macroeconomic planning is concerned. ⁴

Long run growth depends on growth in labour force, capital accumulation, and TFP. Since labour force growth is hard to change and capital stock usually can only grow at a rate of about 4% per year, this leave TFP growth as the only way to increase long run growth. The contents of TFP usually consist of technical progress, spill-over, improved efficiency, scale economies, and so on, among which technical progress is the most important component in TFP growth for the US as the world technology leader. However, caution should be exercised with the understanding of the technologies. An example is the emergence of the so-called “new economy” in the late 1990s.

---

⁴ Krugman (1997) noted that standard economic analysis suggested that the US could not expect its economy to grow at much more than 2% over the next few years. The reason Krugman gave was that if the Federal Reserve Bank would try to force faster growth by keeping interest rates low, serious inflation could be a problem. Of course inflation didn’t rise until recently, but the US economy was overheating starting from the mid-1990s already. Jorgenson, Ho, and Stiroh (2006) report their base-case scenario projections for the US GDP growth at 2.97% per annual for 2005-2015, with the uncertainty ranging from 1.9% to 3.5%. McNamee and Magnusson (1996) gave a detailed discussion on why a long run growth rate 2% could be a problem for the US economy as a whole.
The “new economy” theory postulates that the Western economy is experiencing a fundamental transformation. Productivity slowdown, inflation, and the business cycle will all disappear. Gordon (2000) was rather suspicious of the notion of “New Economy.” For example, the “new economy” proponents believe that information technology is responsible for the surge in productivity growth in the US economy during the second half of the 1990s, and suspect that something fundamental has changed (Oliner et al. 2007). However, Gordon (2003) argues that during the boom in ICT in the late 1990s investment was unsustainable because computer hardware investment grew at more than 30 percent per year for five years in a row, and the rapid growth of computer power need not be matched with an equal expansion of the corresponding demand. In other words, “supply does not create its own demand.”

Even after the IT bubble burst in 2001 the productivity performance of the US economy continued to be better than what it had been during the more than two decades of sluggish productivity growth following the first oil crisis in the early 1970s. Many explanations focus on fundamental technological progress outside of IT production since total factor productivity in other sectors continues to grow impressively (Jorgenson et al., 2007 and Bosworth and Triplett, 2007). A recent study predicts that US trend productivity is expected to be around 2.25 percent a year (Oliver et al, 2007), which is a fairly high rate compared with our calibration in Table 3. However, Gordon (2006) is rather suspicious of the notion that ICT investment would break through all limits set by the “scarcity of economic resources.”

During the US productivity surge, a puzzle is why Europe failed to register a post-1995 productivity growth revival, since the same hardware and software are widely used across Europe and Nokia and Ericson are the leading companies in mobile phone technologies. Gordon’s (2004) explanation is that regulatory barriers and land-use regulations in Europe prevent the retail sector from many productivity gains. But a further question to ask for both the US and EU is that for how long these productivity gains would last?

---


6 Robert Gordon was recently quoted in Mandel (2008) as saying: “The continuing decline in the productivity growth trend provides further evidence that the productivity growth revival of 1995-2004 was a one-time event. In the late 1990s the primary cause of the productivity growth revival was the dot.com boom and invention of the WWW. During 2001-03 the further good news on productivity growth was due to a combination of the delayed impact of the 1990s technology surge (the “intangible capital” hypothesis) with unusually savage corporate cost cutting that caused the prolonged decline in payroll employment between 2001 and 2003.”
gains will last. The ICT boom only continued for half a decade and the successive housing boom did not sustain any longer. To break the limit to growth potentials in both the US and EU, future productivity growth will have to come from some long lasting sources.

A lesson learned from the US experience can be summarized as follows. While innovation increases productivity, it also involves risks. Very often new developments in technology can be delayed, or only bring “a one-time jump in productivity” (Solow 1994). Although the US as the world technology leader may have to rely mainly on product innovation to improve productivity, it is important for developing countries like China to maintain a balance between process and product innovations, and to absorb foreign technologies efficiently. As it is well-known that process innovations can improve efficiency in the short term while product innovation is highly uncertain – “New knowledge accumulates in a way that is neither predictable [and] steady, nor continuous.”(CBO 2005). This situation leads to a question on whether or not it was wise to involve large government interventions to support the development of new technologies such as computers, telecommunications, and the Internet. 

The economics profession for decades has recognized that R&D is important for long run economic growth. The productivity slowdown of the 1970s in the major industrialized countries prompted the US economists to call for public policy support of science and technology because R&D activities generate more benefit for the public (spill-over effects) rather than the entities that incur the major costs of the innovation (Nelson and Romer 1996). During the quarter century up to the mid-1990s some economists seem to have concluded that R&D not only contributed significantly to economic growth in the US in the past but also is a promising way to promote productivity in the future (Walsh 1981). However, empirical evidence didn’t seem to be conclusive (Griliches 1988).

As a result of significant support from the US government, the ICT boom during the second half of the 1990s encouraged the “New Economy” proponents to imagine that the economy could grow faster than before without renewed inflation (Stiroh 1999).

---

7 Schiff (1993) notes "Don't fight the long wave. Yes, the economic recovery has been disappointing, but that's all part of a cycle, the long cycle. Advocates of long-wave theory hold that there are fluctuations in economic growth much longer than normal business cycles, stretching for 40 to 60 years. If you're on the downswing of that cycle, as they say the U.S. is now, there's only so much you can do to pick up the pace. So don't bust the budget trying.” Sterman (1992) noted “We are currently in the trough of the long wave economic cycle during which there is an overcapacity to produce almost everything (steel, autos, semiconductors, laid optical fiber, …).”
However, by the end of the 1990s the debate on the sustainability of the US productivity revival cast doubt on whether the “New Economy” is really new and whether it has changed fundamentally the way our economies work. Following the IT stock market crash of 2000, the phenomenal housing boom and bust, and the ongoing financial turmoil may indicate that monetary policy may not be effective if used to intervene with the “real business” cycles. It probably brings short-term gains while the long run cost might be very high.

3. The Knowledge Production Function Defined

Empirical studies on the link between R&D and productivity performance have been mainly carried out with the approach introduced in Griliches (1979 and 1980a). Examples of applications include Baily and Chakrabarti (1985) on innovation and productivity in industries, for instance. Two issues appear need to be clarified under this framework. Conceptually it is not very clear to what extent the link between R&D expenditure and productivity growth can be studied with well-defined economic relationships. It is natural to expect that the endogenous growth literature should provide testable structure for the empirical studies. However, the restrictive and simplistic formulation specified in the R&D based endogenous growth models do not appear particularly useful. Moreover, in his critics to the endogenous growth model, Solow (1994) wrote:

The idea of endogenous growth so captures the imagination that growth theorists often just insert favorable assumptions in an unearned way; and then when they put in their thumb and pull out the very plum they have inserted, there is a tendency to think that something has been proved. Suppose that the production function is \( A f(K,L) \) where \( A \) carried (Hicks-neutral) technological progress. (The neutrality is just for clarity; it is inessential) Successful innovations make \( A \) larger. But how much larger?

For this purpose, take it for granted that there is something meaningful called “an innovation” and a stream of these innovations occurs as a result of decisions made by firms. It is easy to agree that the flow of innovations per unit time depends on the amount of resources devoted to creating them. If an innovation generates a proportionate increase in \( A \), then we have a theory of easy endogenous growth. Spend more resources on R&D, there will be more innovations per year and the growth rate of \( A \) will be higher. But suppose that an innovation generates only an absolute increase in \( A \): then greater allocation of resources to R&D buys a one-time jump in productivity, but not a faster rate of productivity growth. I do not know which is the better assumption, and these are
only two of many possibilities. But merely to adopt the more powerful assumption is no more than to assume the more powerful conclusion.

We now turn formally to define the knowledge production function as termed in Romer (1990) but the generalization of the concept is more in line with the concept of “technology function” of Phelps (1966). As we will see that Griliches (1980a) formulation can be generalized as a flow variable form using the concept of knowledge production function defined in Romer (1990) when we endogenize total factor productivity growth with respect to innovation and other possible explanatory variables such as trade. The knowledge production function is specified in terms of both stock and flow variables.

At the outset, we try to be as general as possible when defining the concept. Assuming that the production unit is the firm, its production function takes the standard form as follows:

\[ Y = A f (K, L_Y) \]  

(3.1)

where \( Y \) stands for the ordinary output, \( K \) and \( L_Y \) refer to capital and labor, and \( A \) is the level of technology or useful knowledge stock that can be employed in the production process. The knowledge production is measured as the change in knowledge stock \( \dot{A} \), and the basic knowledge production function can be expressed as follows.\(^8\)

\[ \dot{A} = f (K_A, L_A, L, A) \]  

(3.2)

where \( K_A \) and \( L_A \) stand for technology capital or research capital stock and research labor input, and \( L = L_A + L_Y \) is the labor force of the firm. Generally one can specify the knowledge production function with properties as follows.

**A1.** The knowledge production function is monotone in inputs

**A2.** The input set of the knowledge production is convex

**A3.** Researcher is the essential input, i.e., \( \dot{A} = f (K_A, 0, L, A) = 0 \)

**A4.** The input set is nonempty and closed

**A5.** The knowledge production function is finite, nonnegative, real valued, and single valued for all nonnegative and finite inputs

\(^8\) Phelps (1966) uses the term “technology function” in referring to the relationship between research inputs and research outcomes; Gumolka (1970) mentioned “the production function of innovations”; and Jones (1999, 2002, and 2006) prefers to phrase the relationship as “the idea production function.”
The general specification, especially the choice of variables is due to Phelps (1966) where he discussed a possible construction of the “technology function” for an economy as follows.\(^9\)

\[
A(t) = \int_{-\infty}^{t} f(E(L_A(v), L(v), K_A(v)), A(v - \omega))dv 
\]  

(3.3)

where \(A\) is the knowledge stock function, \(f(\cdot)\) is the knowledge production function in flow form, and \(E\) the effective research function. The formulation (3.3) can be used to discuss a whole host of structural specifications. For example, the knowledge production function in flow form is given as:

\[
\dot{A}(t) = f(E(L_A(v), L(v), K_A(v)), A(v - \omega)) 
\]  

(3.4)

and the effective research function is as follows:

\[
E = \varphi(L_A, L, K_A) 
\]  

(3.5)

includes \(L_A\) as the amount of employment in the research industry, \(K_A\) the stock of research capital, and \(L\) the labor force.\(^{10}\) Standard assumptions of the production function apply here. For example, Phelps (1966) assumes that “effective research” is homogenous of degree one in \(L_A\), \(L\), and \(K_A\). It is increasing and strictly concave in \(L_A\) and \(K_A\). Besides the standard assumptions, the presence of the total labor force in the function gives specific structure as follows. Assume that no worker has an absolute advantage in the production of commodities, but that all workers differ in their effectiveness in research. Phelps argues that if just one researcher is needed, the worker with the greatest comparative advantage in research will be assigned for the purpose. If an additional worker is needed, then the next effective worker will be given the task. This means that the increase of “effective research” will not be proportional because “the second worker is inferior to the first.” Therefore Phelps assumes that “with efficient allocation of labor, effective research increases as the number of researchers at a decreasing rate.”\(^{11}\)

An interesting structure implied in \(f(\cdot)\) is that knowledge stock of the economy \(A\) as a whole is separable from the “effective research function” of the industry. This is a

---

\(^9\) As to the controversial choice of variables for the knowledge production function, in one extreme case Beaudreau (2005) shown that “by including energy, the cornerstone of production processes as modeled in engineering, in simple growth accounting exercises, the Solow residual is nearly eliminated.”

\(^{10}\) The inclusion of capital in the knowledge production function could also be justified because “Technology capital is accumulated know-how from intangible investments in R&D, brands, and organizations” (McGrattan and Prescott 2007).

\(^{11}\) Solow (1994) also has a good reason to include ordinary workers into the knowledge production can as follows: “It is an important fact of life that many instances of product improvement and cost reduction have little to do with the R&D activity, but originate in some other way, for instance from the accumulation of small suggestions coming from production workers, process engineers, and even customers.”

13
very important formulation, which means under certain circumstances research efficiency can be studied independently prior to the study of knowledge production. Another specification introduced in $f(\cdot)$, that may have structural implications in applied work, is that knowledge production at time $t$ in producing technical progress at time $v$ is an increasing function of the level of technology at time $v$-$w$. Phelps interprets $w$ as “publication lag.” Note that homogeneous of degree one for $E$ with respect to $L_A$, $L$, and $K_A$ in combination with homogeneous of the same degree for $f$ in $L_A$ and $A$ lead to homogeneous of degree one for $f(\cdot)$ in $L_A$, $L$, $K_A$, and $A$. It appears that the property of homogeneous of degree one is very important in knowledge production. This is not so obvious in many commodity productions.

Phelps also mentioned the assumption that the marginal effective function, $E$ was stationary in $L_A$ so that researchers do not get more productive over time. This assumption is important in the sense that modeling knowledge production is equivalent to modeling time trend for productivity growth, i.e., there might be better explanations than simply considering time as an input to the production process.\footnote{Gomulka (1970) noted “innovation of all kinds and, in particular, the flow of technological innovations had been treated as something given from the outside, as a factor of considerable importance but of no cost.”}

$\dot{A}$ as the change in knowledge stock $A$ needs to be homogeneous of degree one to be a valid knowledge production function in the context of Phelps (1966). Definition (3.2) simplifies the structure of the flow function in (3.3) because the research effective function are not present. Thus, knowledge is treated as internal and the researchers and the total labor force engage in a joint production of knowledge that generate total factor productivity growth for the ordinary production process. The appearance of two types of labor sometimes might be interpreted as that the task of the researchers is to push forward the production frontier while ordinary workers can be involved in improving technical efficiency. This is consistent with the applied productivity models that TFP can be decomposed into technical progress and technical efficiency improvement components.

The homogenous property can be crucially important in understanding the knowledge production process. For example, growth accounting for knowledge production might be conducted in a similar fashion as in the ordinary accounting practice. The pattern of knowledge production can be characterized with the way research capital and labor
forces are utilized as in Zheng, Bigsten and Hu (2009). Whether or not capital-intensive research activities are sustainable should be a relevant question.\(^\text{13}\)

Economies of scale and technical progress are also intriguing issues in this context. One should not be surprised if the knowledge production function does not usually present a time trend. If research effectiveness declines as the next effective researcher is assigned to the sector there should be no particular reason to expect that constant returns to “the production of the new ideas” prevail\(^\text{14}\)

Another interesting question is that a secondary growth model with respect to knowledge production might be established. Then one can imagine that sustainable growth in labor productivity of knowledge production would require technical progress in the knowledge production process itself. In the ordinary growth model, investment is usually a fraction of the total output (GDP), while investment in research capital can be realized through borrowing. Very often research capital is ignored in endogenous growth models. Does this mean that labor’s share in total cost of research dominates that of research capital?

If the contribution of the ordinary workers can be ignored or is insignificant equation (3.5) is simplified as follows:

\[
\dot{A} = f(K_A, L_A, A) \quad (3.6)
\]

Note that in (3.6) \(\dot{A} = f(K_A, L_A, A)\) is the flow form. It is convenient for discussions on properties of the knowledge production function such as the curvature of the production function, returns to scale, output elasticity of inputs, and elasticity of substitutions. A typical flow form is of course the Romer (1990) specification, \(\dot{A} = \delta H_A A\). In Rivera-Batiz and Romer (1991b), the flow function was written as

\[
\dot{A} = f(H_A, K_A, L_A, A)
\]

where \(H_A\) stands for human capital employed in the research. We do not adopt this formulation here because we try to emphasize the essentiality of the research labor in the knowledge acquisition process. Human capital measured such as average years of schooling can be incorporated at a later stage.

If \(A\) is homogenous of degree one then we have

\[\text{\textsuperscript{13}It is not clear at this point that the growth accounting should emphasize } \dot{A}/A \text{ itself or the growth rate of } \dot{A}/A \text{ since the later concept may not be well-defined.}\]

\[\text{\textsuperscript{14}This phrase was due to Jones, for example Jones (1999).}\]
\[
\dot{A}/A = f(K_A/A, L_A/A, 1)
\]

which is quite useful for theoretical analysis since homogeneous of degree one implies that growth rate in total factor productivity declines as the stock of knowledge increases. This property comes from the monotone assumption for the technology in knowledge production.

It may not be always the case that knowledge stock matters that much, then a more concise form can be given by

\[
\dot{A} = f(K_A, L_A)
\]

(3.7)

Different specifications may fit into different industries depending on the nature of the production process. The reason (3.7) is motivated is that endogenous models of economic growth often concentrate on labor force in the research sector (Romer 1990, and Aghion and Howitt 1992) while the R&D-productivity link literature emphasizes R&D expenditure (Griliches 1979 and 1980a) or R&D capital at firm level. The situation seems to indicate that labor and capital might be considered as the main inputs to the knowledge production process along with knowledge stock. Then an immediate question is if missing either of the variables in empirical studies will result in biased estimations of structural parameters. Another related question is also whether capital and labor should always be included in the knowledge production function estimation.

The endogenous growth literature seem to have difficulties in incorporating research capital or capital used in research in the knowledge production function. The Romer (1990) specification \( \dot{A} = \delta H_A A \) includes only labor and knowledge stock as inputs to the production process. A strong assumption involved in this specification is that unskilled labor and physical capital are not effective in R&D. An alternative specification in Rivera-Batiz and Romer (1991) was attempted in the spirits of (3.7) as follows:

\[
\dot{A} = BH^\alpha L^\beta \int_0^A x(t)^{1-\alpha-\beta} dt
\]

(3.8)

where \( H, L, \) and \( x_i \) represent inputs used in R&D and \( B \) denotes a constant scale factor. This case refers to a technology for R&D that uses the same inputs as the manufacturing technology. Specifically, human capital \( H \), unskilled labor \( L \), and capital goods \( x \) such as personal computers or oscilloscopes are all productive in research. But the difference of this specification in comparison with those of Phelps (1966) and Romer (1990) is that “knowledge per se has no productive value.”
This specification does not include knowledge as input since it is assumed to have “no productive value.” Our general formulation of the knowledge production function in (3.5) seems to provide a framework that different specifications can be tested in empirical work. This latter specification is interesting in the sense that it has a vintage interpretation for physical capital used in the research process. But there are two issues that need to be discussed. The first is that we need to index capital for different vintage, which can be simply interpreted as an measurement issue. The second is that new vintages of capital will still involve research labour and research capital, then the question goes back to the original one, i.e., what are the inputs to be considered for the production of $A$?

The notion of knowledge production function in general do not rule out a two-sector model of endogenous growth in which there is a research sector that specialized for knowledge production. The usefulness of this framework is obvious. It can be used to evaluate the efficiency of the research sector on one hand and can also be applied to answer questions concerning spill over effects and learning.

To be able to further appreciate the notion of the knowledge production function we need to return to Phelps (1966) for the dynamic properties required of a valid “technology function”, i.e., the knowledge production function in stock form, $A(t)$. Phelps listed preferred properties as follows:

- **B1.** Diminishing returns
- **B2.** Diminishing marginal rate of substitution
- **B3.** The marginal effectiveness of current research is an increasing function of the level of technology recently attained (technical progress in research)
- **B4.** Exponential growth of researchers will produce an exponential increase of the level of technology

With the four properties above, Phelps demonstrated the problem with a simple time trend specification $A(t)=A_0 e^{\gamma t}$. Assume that

$$A(t) = A_0 e^{\sum_{v=0}^{\beta R(v)}}$$

(3.9)

then the marginal effectiveness in research at time $v$ $\frac{\partial A(t)}{\partial R(v)} = \beta A(t)$ is increasing with $R(v)$. Models with this property will be explosive in the long run. For diminishing returns Phelps gave an example as follows:
A(t) = \beta \left[ \int_{-\infty}^{t} R(v)dv \right]^\alpha, \quad 0 < \alpha < 1 \quad (3.10)

this formulation presents constant elasticity of substitution between research at two
points of time for technology level at time t. This is a common specification often
seen in the endogenous growth literature with Schumpetarian interpretations. The
next specification is more favorable in the context of Phelps (1966):

A(t) = \int_{-\infty}^{t} G(R(v))dv \quad (3.11)

where \( G(R) \) is increasing and strictly concave, which implies both diminishing
marginal returns and diminishing marginal rate of substitution. The marginal
effective function, \( G'(R) \), is stationary so that “researchers do not get more productive
over time.” Phelps considered an example including number of researchers and the
level of technology simultaneously

A(t) = \int_{-\infty}^{t} H(R(v), A(v-w))dv, w > 0 \quad (3.12)

If \( H(R, A) \) is increasing in \( R \) and \( A \), and strictly concave in \( R \), then \( A(t) \) will satisfy
B1 through B3. B4 will be satisfied if \( H(R, A) \) is homogeneous of degree one in \( R \)
and \( A \) and \( H(0, A) = 0 \). We are now ready for exploration of other properties and
specifications of the knowledge production function.

4. Consequences of Structural Misspecifications

The popularity of the neoclassical economics texts throughout the world is not
accidental. The main reason is perhaps that the neoclassical framework remains an
open system at the basic level. However, to be able to take advantage of its openness
it depends on how well model builders understand the system. To some extent the
neoclassical principles in setting up the general equilibrium model is much like
insisting on having a build-in firewall in an operating system for personal computers.
It’s not convenient but necessary to make sure that the model will be safe.

For example, drifting away from the standard neoclassical practice may lead to
models that give problematic solutions. In fact, there has been a debate since the early
1990s on the scale effects exhibited in some standard endogenous growth models
(Jones 1995, 1999, and 2002). With the assumption of the increasing returns, some
endogenous growth models predict that growth can be unlimited as long as one
wishes. For example, the R&D based endogenous growth model of Romer (1990)
relies on a knowledge production function that gives explosive solutions in the long

\[15\] Recall that convexity of the input set and concavity of the production function implies
linear homogeniety (Chambers 1988).
run (the sign of an unsafe model). It is a small wonder that the "New Economy" advocates found Romer's theory comfortable in the late 1990s. The model was even suspected to have landed support to the advocates of the “New Economy” (Schwartz and Leyden 1997). We have now realized that people were too optimistic with the so-called "New Economy" at the time.

Jones (1999) points out that Romer (1990) and Aghion and Howitt (1992) are in the same category as far as “scale effects” is concerned. We discuss the long run theoretical consequences and empirical implications of the formulations that have been attempted in the endogenous growth and R&D-productivity link literature. Nordhaus (1969) used a formulation as follows. The rate of technological change as a function of the number of inventions can be written as the simple log-linear form:

$$\frac{\dot{A}}{A} = N^\beta A^{-\gamma}$$  \hspace{1cm} (4.1)

Equation (4.1) means first that the rate of technological advance is an increasing function of the number of inventions produced at a given time, $N$. For a given initial stock of knowledge, $A$, inventions add less and less to the level of technology (Machlup 1962). In other words knowledge production presents diminishing returns to $N$. The other feature of equation (4.1) is that the rate of advance may be a function of the level of knowledge. If knowledge has advanced to a high level, for a given number of inventions the rate of return to a higher knowledge level might be diminishing. Nordhaus also pointed out that the opposite might also be the case. Dynamic analysis shows that $A$ must be a retarding force on technological advance for sake of stability. Rewrite (4.1) slightly,

$$\dot{A} = N^\beta A^{1-\gamma}$$  \hspace{1cm} (4.2)

Diminishing returns in both $N$ and $A$ imply $0 < \beta < 1$ and $0 < (1-\gamma) < 1$. If $\beta + (1-\gamma) = 1$ we end up with a standard Neoclassical production function in Cobb-Douglas form. Note that that the $N$th invention gives an increase in productivity as follows:

$$A'(N) = \beta N^{\beta-1} A^{1-\gamma}$$  \hspace{1cm} (4.3)

In Romer (1990) a production function for output $Y$ is specified as follows:

$$Y = A^\varphi L_{\varphi}$$

It was assumed that all researchers could take advantage of the entire stock of knowledge $A$ at the same time, so the output of researcher $j$ was $\delta L_j A$, where $L_j$ is the human capital devoted to research by each researcher. When aggregating over all researchers, TFP growth rate was assumed to be proportional to the number of researchers employed as follows:

16 In Jones (2002), number of patents was replaced by the number of researchers.
\begin{equation}
\frac{\dot{A}}{A} = \delta L_A
\end{equation}

(4.4)

where $\dot{A}$ stands for the absolute change in TFP, $\delta$ is a constant and $L_A$ the number of researchers devoted to R&D activities in a labour force $L = L_Y + L_A$. And $L_A = sL$, where $0 < s < 1$ is a constant. Note that the long run growth rate for per capita income $y$ is as follows:

$$g_y = \frac{\dot{Y}}{Y} - \frac{\dot{L}}{L} = \sigma \delta s L$$

(4.5)

Loosely speaking one can also interpret it as the labour productivity in the context of a firm. But this formulation implies that no steady state solutions exist and the rate of productivity growth differs as the proportion of researchers to total employed people diverges. Hence a more flexible formulation was considered in Jones (1995a) as follows:

$$\dot{A} = \delta L_A A^\phi$$

(4.6)

where TFP growth rate is allowed to slowdown as the level of TFP increases if $\phi$ is restricted to be less than unity. Equation (4.6) is only slightly more general than the specification in Nordhaus (1969) and it reduces to (4.4) if $\phi$ is set to unity.\(^{17}\) In both (4.4) and (4.6) absolute change in TFP level depends on the number of researchers involved. The long run growth rate of $A$ is given by

$$g_A = \frac{n}{1-\phi}$$

(4.7)

and the steady state growth rate in output per capita exists if $\phi$ is not equal to one:

$$g_y = \sigma g_A = \frac{\sigma n}{1-\phi}$$

(4.8)

But in this case the model ceases to be endogenous since the steady state growth rate is exogenously determined by $\sigma$, $n$, and $\phi$. On the other hand long run labour productivity is still endogenous with respect to the numbers of researchers employed:

$$y^*(t) = (1-s) \left( \frac{\delta (1-\phi)}{n} s L(t) \right)^{\sigma (1-\phi)}$$

(4.38)

This result might be used in empirical studies both at the macro and the micro level. A slightly different formulation was provided in Jones (2002) in which the long run solution for $A$ is given as follows:

$$A = A_0 \left( \frac{x_0}{x} e^{\lambda t} + 1 \frac{x_0}{x} \right)^{\gamma/L}$$

where $x = \dot{A}/A$.

\(^{17}\) Benhabib and Spiegel (1994) adopted a specification suggested in Nelson and Phelps (1966) in which $\rho$ was set to zero. A more detailed discussion based on the concept of “technology function” can be found in Phelps (1966b).
Rewrite (4.6) in terms of TFP growth, one gets
\[
\frac{\dot{A}}{A} = \delta L_A^{\phi} A^{\delta-1}
\] (4.10)
which might be used as a basic structure for regression analysis. The relation between (4.10) and (4.7) is interesting since one may use (4.10) to estimate $\phi$ for different, say, countries and then use (4.7) to test if there exists a steady state growth rate in TFP.

Note that (4.10) cannot be easily log linearized. The problem with (4.10) is that it does not allow non-positive TFP growth for log linearization. Again one option suggested in Zheng (2008) is to linearise (4.10) around (1,1) according to the Taylor expansion as first-order approximation as follows:
\[
\frac{\dot{A}}{A} \approx \delta + \delta(\phi - 1)(A - 1) + \delta\lambda(L_A - 1)
\] (4.41)

5. The Structural Tests

The knowledge production function approach can be attractive in an empirical sense that the standard techniques of applied productivity analysis can be effective in investigations on the determinants of productive efficiency and total factor productivity growth. We use the translog production function specification as an example to demonstrate the usefulness of the approach introduced so far.

Separability and linear homogeneity appear to be the two most important properties that a knowledge production function should possess. For matter of practicality as well as convenience we now discuss the structure of the knowledge production function through a specification of the translog production function (Christensen, Jorgenson, and Lau 1973). An important reason for choosing Translog is that the Cobb-Douglas specification of (3.19) is one of its special cases. We specify a knowledge production function with three inputs as follows:
\[
\ln \dot{A} = \ln f(K_A, L_A, A) = \beta_0 + \beta_A \ln A + \beta_K \ln K_A + \beta_L \ln L_A + \\
\frac{1}{2} \beta_{AA} \ln (A)^2 + \frac{1}{2} \beta_{KK} \ln (K_A)^2 + \frac{1}{2} \beta_{LL} \ln (L_A)^2 + \\
\beta_{AK} \ln A \ln K_A + \beta_{AL} \ln A \ln L_A + \beta_{LA} \ln K_A \ln L_A
\] (5.1)
Output elasticity of each input is given by:
\[
\varepsilon_A = \frac{\partial \ln \dot{A}}{\partial \ln A} = \beta_A + \beta_{AA} \ln A + \beta_{AK} \ln K_A + \beta_{LA} \ln L_A
\]
\[
\varepsilon_K = \frac{\partial \ln \hat{A}}{\partial \ln K_A} = \beta_K + \beta_{KK} \ln K_A + \beta_{AK} \ln A + \beta_{KL} \ln L_A \tag{5.2}
\]
\[
\varepsilon_L = \frac{\partial \ln \hat{A}}{\partial \ln L_A} = \beta_L + \beta_{LL} \ln L_A + \beta_{LK} \ln K_A + \beta_{LA} \ln A
\]

Elasticity of scale:
\[
\varepsilon = \varepsilon_A + \varepsilon_K + \varepsilon_L \tag{5.3}
\]
For linear homogeneity of \(\ln f\), one needs,
\[
\varepsilon = \varepsilon_A + \varepsilon_K + \varepsilon_L = 1
\]
which requires
\[
\beta_A + \beta_K + \beta_L = 1 \tag{5.4}
\]
\[
\beta_{AA} + \beta_{AK} + \beta_{LA} = 0
\]
\[
\beta_{KK} + \beta_{AK} + \beta_{KL} = 0 \tag{5.5}
\]
\[
\beta_{LL} + \beta_{KL} + \beta_{LA} = 0 \tag{5.6}
\]
For separability (or Hicks-neutrality) of \(A\) from \(K_A\) and \(L_A\), we use the definition as follows:
\[
\frac{\partial}{\partial A} \frac{\partial f}{\partial K_A} = 0 \tag{5.7}
\]
which requires both \(\beta_{AK} = 0\) and \(\beta_{LA} = 0\). If \(\hat{A} < 0\) then (5.1) is not operational but it also violates the basic assumption of the knowledge production function, A5. Generalized Leontief production function (Diewert 1971) will be more convenient in this case. However, our purpose for using translog is to be consistent with the Cobb-Douglas specification as the simplest and most used functional form in regression analysis with empirical work. A simpler test is to impose separability first and then check the degree of homogeneity of \(\varphi(K_A, L_A)\) in terms of \(K_A\) and \(L_A\). We take advantage of (3.1) and use a specification as follows:
\[
\ln \frac{\hat{A}}{A} = \ln \varphi(K_A, L_A) \tag{5.8}
\]
\[
= \beta_0 + \beta_K \ln K_A + \beta_L \ln L_A + \frac{1}{2} \beta_{KK} (\ln K_A)^2 + \frac{1}{2} \beta_{LL} (\ln L_A)^2 + \beta_{KL} \ln K_A \ln L_A
\]
We next demonstrate how the elasticity of substitution for the translog production function can be calculated since it can be useful to characterize the knowledge production function if knowledge stock is not separable from research capital and labour. Very often in policy analysis one needs to consider the balance between innovation and imitation (through trade, for example), this requires the knowledge production function to accommodate inputs other than research capital and labour in order to say something about substitution possibilities in the process of knowledge acquisition. Research is often considered as capital intensive as emphasized in the
formulation of Griliches (1980a) concerning the R&D-productivity relationship. On the other hand, endogenous growth theorists focus on the importance of human capital possessed by researchers in the R&D sector of the economy as in Romer (1990) and Aghion and Howitt (1992). Do these mean that complementarity rather than substitutability between labour and capital should prevail in the research sector? So to be more general we extend our specification to accommodate more inputs, 18

$$\ln A = \ln \beta_0 + \sum_{i=1}^{n} \beta_i \ln x_i + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \beta_{ij} \ln x_i \ln x_j$$  

(5.9)

Follow Berndt and Christensen (1973b), note that

$$f_i = \frac{\ln A}{x_i} \frac{A}{\ln x_i} = \left[ \beta_i + \sum_{j=1}^{n} \beta_{ij} \ln x_j \right] \frac{A}{x_i}$$  

(5.10)

The Allen elasticity of substitution can be expressed as follows:

$$\sigma_{ij} = \frac{x_if_i + x_jf_j}{x_ix_j} \frac{F_{ij}}{F}$$  

(5.11)

where $F$ is the determinant of the bordered Hessian matrix and $F_{ij}$ is the cofactor of $f_{ij}$ and $F$.

$$F = \begin{bmatrix} 0 & f_1 & f_2 & \ldots & f_n \\ f_1 & f_{11} & f_{12} & \ldots & f_{1n} \\ f_2 & f_{22} & f_{22} & \ldots & f_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ f_n & f_{n1} & f_{n2} & \ldots & f_{nn} \end{bmatrix}$$  

(5.12)

The concept of substitution can be meaningful concerning questions about whether there should be a balance between product innovation and imitation or between investment in R&D and participation in trade.

6. Potential for Empirical Applications

We expect that the knowledge production function approach should have good potentials in empirical studies on determinants of productivity. Since the publication of Farrell’s (1957) paper on the measurement of productive efficiency, the applied productivity analysis has become an area with a huge literature using econometric and programming techniques to measure technical progress, technical efficiency, and productivity growth. Especially applied studies using the non-parametric deterministic frontier method (DEA) experienced explosive growth in the 1980s and 1990s thanks to the seminal paper of Charne, Cooper, and Rhodes (1978).

---

18 Notations are due to Chambers (1988), and see del Valle and Astorkiza (2003) for a recent application to the Spanish fishing sector.
However, analyses of the productivity estimates have been simply relying on various regression models with little built in structure for accurate interpretations of the estimated parameters. Issues concerned for researchers in regression analyses are often revolving around how to choose statistical models for proper inference of parameter estimates (Grosskopf 1996). Using productivity index such as the DEA efficiency scores and Solow residuals to regress on a set of explanatory variables is termed as Two-Step procedure. Zheng (2008b) notes:

The two-step procedure has been well received in the literature on the determinants of technical efficiency and productivity growth with deterministic frontier approach. This procedure seems preferable for two reasons. First, it allows one to treat the non-production factors as fixed and indivisible non-discretionary inputs and avoids the free disposability assumption with respect to these inputs. Second, it permits considerable flexibility in the choices of the non-production inputs to be considered (Ray, 1991). Studies in line with this type of research can be found in Madden and Savage (2001), and Zheng, Liu, and Bigsten (1998, and 2003).

From the point of view of applied production analysis, the advantage of having the efficiency scores or Solow residuals separately estimated from the second step regression analysis is that economists often have more information about how the production structure looks like at the first step while it is not very clear what explanatory variables should be included in the second step. Imposing strongly a priori both theoretical and empirical restrictions at the first step can avoid the trouble of misspecifications for the basic structure of the production function when non-structural variables are involved at the very beginning.

The concept of the knowledge production function combined with the two-step procedure opens new avenues for considering the relationship between innovation and productivity growth. Promising areas for application should include productivity growth accounting (in comparison with ordinary growth accounting), tests of knowledge production structure, efficiency in research, and knowledge measurement.

Growth accounting for the knowledge production function may generate new insights about how the research sector of the aggregate economy should be analyzed. Industries with high productivity may have to be evaluated together with other industries in order to provide a whole picture as to how innovation activities promote
growth at the macroeconomic level. Research sector needs to be complimentary to the production process not a showcase for the economy.

More general functional forms instead of Cobb-Douglas production functions can be specified. Complicated functional forms can be employed such as translog and generalized-Lionief. Efficiency of knowledge production can be studied under this very general framework. The issue of technical progress in knowledge production is even more interesting since it will be very informative for growth accounting. If time trend in the knowledge production function is difficult to prevail than all growth in knowledge production can be attributed to the inputs of the function such as capital and labor. Then the question is what the factor shares are, whether or not the distribution of income follows the first order conditions, so on and so forth. Substitution and even technical progress of the research sector become relevant issues through the lenses of the knowledge production function. Questions such as whether innovation and trade are substitutable can be conveniently studied with this analytical tool.

The notion of the knowledge production function in this study links innovation inputs to total factor productivity as the output measure. Of course, there are jumps with the direct link. Knowledge production would remind people naturally of patents, new products, research papers, new technology and process innovation.19 The advantage of using total factor productivity as a knowledge measure lies in that numbers of patents and publications are hardly ideal measures of knowledge creation and only useful knowledge are accounted for since productivity can be regarded as an ultimate performance measure.20 Therefore, the jumps in linking innovation inputs to productivity outcomes might be seen as an advantage rather than disadvantage.

The now standard practice of decomposing total factor productivity (TFP) growth into technical progress and technical efficiency improvement components has inspired many applications in empirical work. However, it presented both challenges and opportunities for the investigation of determinants of TFP. Moreover, factors behind productivity dynamics may impact firm performance through different channels, innovation and purchase of new technologies may shift production frontiers, and trade and learning may enhance efficiency. We suspect that the now standard practice of

20 Griliches found no evidence of decline in the speed of invention, but there was obviously a slowdown in the rate that available technologies were adopted (Brenner 2002).
decomposing productivity growth into technical progress and technical efficiency change components (Nishimizu and Page, 1982; and Färe et al., 1994) shall have important implications for the microstructure of knowledge production functions.

Statistics on the research labor force now has become standardized such as the number of engineers and scientists (United Nations: *Yearbook of International Statistics*). Statistics on research capital are also largely available both due to the popularity of Griliches type of R&D-productivity link studies and recently available Community Innovations Survey in more than a dozen European countries (Zheng 2008). So the concept of knowledge production function introduced in this study should have great potentials in empirical as well as theoretical applications for future research.

To apply the knowledge production function for the purpose of Griliches (1980a) specification is rather straightforward (Zheng 2008b). The Nelson-Phelps specification involves catch-up effects with the frontier; the frontier approach in applied productivity analysis will be useful if firm data or cross-country/region data are involved. The effect of trade might be studied through the estimation of elasticity of substitution. We now turn to the issue of scale effects when different levels of data are used. This is important because the knowledge production function is in essence a dynamic concept. Now the question is whether scale effects of the endogenous growth model is also a valid issue at firm level.

One of the purposes in our study is to provide the testable hypothesis for empirical work based on firm data, taking advantages of various structure specifications in the literature. Those specifications are mostly found in endogenous growth models such as in the R&D based models of Romer (1990). Jones (1995) not only noticed the problem of the scale effect in the R&D based endogenous growth models, he also carried out empirical tests using US time series aggregate data. Abdinh and Joutz (2006) also use cointegration analysis to estimate knowledge production function specified in several well-known R&D based endogenous growth models. Then the question is whether similar tests can be carried out using industry or firm data. One recent empirical work is Zheng (2008) in which Norwegian firm data was used to test the structural specifications from Romer (1990) and Jones (1999). Theoretical model of endogenous growth with R&D investment and stochastic innovation also took advantage of the specifications from the quality ladder models in the macro growth

---

literature (Klette and Griliches 2000). To conclude this section, we take an example from the literature to demonstrate how “scale effects” of the macro growth models has been tested empirically using data from aggregate economy to industries and to firm level.

Dinopoulos and Thompson (1999) surveyed the literature concerning the empirical tests of models with scale effects using aggregate economy, industry, and firm data. They provide straight forward examples linking theoretical formulation to empirical estimation. The knowledge production function is specified as follows:

\[ \frac{\dot{A}}{A} = \gamma L_A(t) \]  
\[ \text{(6.1)} \]

If we rewrite it into a flow form as

\[ \dot{A} = f(L_A, A) = \gamma L_A A \]  
\[ \text{(6.2)} \]

we can see it more clearly as a knowledge production function with a unrestricted Cobb-Douglas form. This knowledge production function is not homogeneous of degree one in its arguments \( L_A \) and \( A \). The consequence of this violation is termed as scale effect and has been discussed in detail in the last section. We here only note that in addition, research capital is missing from this production function if we take the concept of knowledge production function seriously. This misspecification may result in estimation bias in empirical studies due to missing relevant variable problems. For example, Backus et al. (1992) investigated the effects of input scale in a cross-sectional setting. They conducted cross-country regressions on equations of the form:

\[ \frac{\dot{A}}{A} = \beta_0 + \beta_1 L_A + \beta_2 \frac{L_A}{L} \]  
\[ \text{(6.3)} \]

which allows for discriminating tests between the scale and intensity of inputs used in knowledge creation. Their regressions used the growth rates of GDP per capita and manufacturing per worker, while \( L_A \) was measured with the number of students, the number of scientists, and R&D expenditure. They found that there was no support for scale effects in GDP, while intensity effects exist only when inputs are measured with the number of students. For industrial data Backus et al. consider a model as follows:

\[ Y_i = \gamma_i A_i N_i^{1-\alpha_i} K_i^{\alpha_i} \]  
\[ \text{(6.4)} \]

where \( Y_i \) is value-added in industry \( i \), \( i=1, \ldots, I, N \) refers to labour and \( K \) is capital.

Taking \( A \) as a measure of learning by doing, they further specify:

\[ A_{t+1} = A_t (1 + \beta Y_t)^\gamma \]  
\[ \text{(6.5)} \]

There has been a literature on diminishing technological opportunities since the early 1960s, and recent studies on endogenous growth have also discussed related issues intensely (Jones 1999, Segerstrom 1998, and Kortum 1997).
If there are spillovers across industries inside a country, they assume that

\[ A_{t+1} = A_t \left(1 + \sum_{j=1}^{I} \beta_y Y_j \right)^\rho \]  

(6.6)

In comparison with the GDP data, “scale effects in the manufacturing are evident for all three input measures” (Dinopoulos and Thompson 1999). These findings are relevant to the issue we discussed in the first article of our trilogy. Empirical findings of positive relationship between innovation and productivity have been documented in the literature using firm level data. However, Jones (1995) shows that at the macro level time series evidence of the US and European data do not support this assumption. One problem might be that only positive findings got published when firm data were used.

Models of industry evolution are also interested in the issue of scale effects. Concerning the relationship between firm growth and size, Dinopoulos and Thompson (1999) note that extensive research has showed that the growth rate of firms is not for sure independent of their size (i.e., there may or may not be firm-level scale effects). Negative correlation between size and firm growth was found in Evans (1987a,b) and Hall (1987), but for large firms the correlation was not significant. Other related studies in this area include Jovanovic and MacDonald (1994), and Klepper (1996). In these investigations about the determinants and effects of industry shake outs, the conclusion appears to be that “firm-level evidence does not support the assumption of long-run growth scale effects” (Dinopoulos and Thompson 1999).

As to trade, Backus et al. (1992) suggest that specialization index can be constructed using the United Nations’ Yearbook of International Trade Statistics. They used the following specification to incorporate trade:

\[ g(y_t) = \beta Y_t \sum_{i=1}^{I} (Y_i / Y_t)^2 \]  

(6.7)

where \( g \) is growth rate in per capita output \( y \), the summation in (6.7) is a specialization index. This formulation can be related to the knowledge production function in a straightforward manner.

---

23 For more general discussion on R&D and spillovers, see Griliches (1991 and 1992), and Griliches (1984) re-examine the relationship between interindustry technology flows and productivity growth.

7. Conclusions
The economics profession for decades has recognized the importance of research and
development for long run economic growth (Solow 1956 and 1957). Since the early
1970s the productivity slowdown in the major industrialized countries has prompted
economists to call for public policy support of science and technology in the US. 25
During the quarter century up to the mid-1990s the mainstream thinking in economics
seem to have concluded that R&D contributed significantly to economic growth in the
US in the past (Griliches 1987)26 and R&D is “a promising way of promoting future
productivity” (Walsh 1981). Along with the discussion this period witnessed the
emergence of endogenous growth models of Romer (1990) and Aghion and Howitt
(1992). This new generation of growth models appears to have shaped theoretical
foundations for public policies rewarding innovation in science and technology
because R&D activities often bring large social returns and the US “underinvests in
research” (Nelson and Romer 1996). 27

The ICT boom during the second half of the 1990s encouraged the “New Economy”
proponents to imagine that the economy can now “grow faster than before without
renewed inflation” (Shepard 1997, cited in Stiroh 1999). However, by the end of the
1990s the R&D-based growth models were overshadowed with the presence of “scale
effects” (Jones 1995a, b, and 1999). Efforts to eliminate the scale effects often lead to
elimination of the endogenous features of productivity growth in this type of models.
Moreover, the debate on the sustainability of the US productivity revival cast doubt

25 Mansfield (1972) surveys the discussion on the contribution of R&D to economic growth in
the US. At the meantime, the publication of “The Limit to Growth” (Meadow et al. 1972)
stirred up discussions on the possibilities for future economic growth. Hueckel (1975)
sumarizes different views and policy perspectives. For a discussion on “productivity
slowdown” and R&D expenditure, see Griliches (1980b). Griliches (1988 and 1997) discusses
other aspects of productivity slowdown.
26 The 1987 official recognition of the Solow’s contribution to the theory of economic
growth was also seen as a mainstream approval to the notion that ”technology plays key role
in economic growth” (Marshall 1987).
27 The US R&D policy may have been interpreted as counterbalancing the ”real business
cycle” judging from what Romer (1998) wrote as follows: ”One of the reasons we are now
enjoying the benefits of the information technology revolution is that the federal government
created the academic discipline of computer science more than four decades ago.” However,
different views existed as far back as in the early 1980s: “There is no doubt that renewed
emphasis on basic research and R&D are important ingredients for economic revitalization.
However, vigorous stimulus of research is not a sufficient response to the current crisis”
(Sterman 1983). Dickson (1983) reviewed the long wave theory as argument for increased
support for the basic research together with policies to offset unemployment due to
technological change.
on whether the “New Economy” is really new and whether it has changed fundamentally the way our economies work.  

In connection with the current financial crisis, the neoclassical framework still provides the necessary methodology to think of, for example, policies for promoting economic growth in industrialized countries. The Federal Reserve was most likely working in the wrong direction when monetary policy was used to maintain growth rate well above the US output potentials. Some people say that Greenspan was probably looking at the inflation rate instead of potential output or he simply believed that the US output potential had been increased permanently from what the Solow model would usually predict due to the IT boom (the so-called “New Economy”). Ironically, this happened during a period that neo-liberalism and “Washington Consensus” had been the dominant doctrine in both industrialized countries and in economies in transition. When the US economy failed as the role model of free market in its recent financial collapse, fingers were pointed to the neoclassical economics and its relation to the neo-liberal emphasis on the importance of the free market and private ownership.

In comparison with the Cold War period, economic reforms in the former socialist countries including China have largely enriched economic practices in the global economy. We should retain an open mind if neoclassical economics is found useful in the studies of those economies (especially as a methodology), not to mention if careful applications of the neoclassical tool may lead to the emergence of new models for economic growth and development.

The knowledge production function framework presented in this study opens new avenues for considering the relationship between innovation and long run growth, and for analyses on the determinants of productivity performance. As we have shown, for productivity to grow one needs inputs, not just ideas, but also physical inputs such as capital and labor, which are scarce resources. It should be emphasized that productivity outcomes are the results of balancing supply of and demand for innovative activities and supply does not automatically create its own demand. The framework also brings in rich structures for endogenizing productivity performance,

---

making possible the applications of the techniques developed for applied productivity analysis during the last decades. More efforts should be made in this line of research.
References


Walsh, John, Is R&D the Key to the Productivity Problem? As new Administration seeks to spur economic growth economists can offer a diagnosis of what went wrong. Science, Vol. 211, 13 February 1981.


