What Policy Makers Should Know About Total Factor Productivity

Jesus FELIPE*
Asian Development Bank
University of Cambridge
Australian National University

Abstract: Since the mid-1990s, development plans in many countries across developing Asia have emphasised the need to increase Total Factor Productivity (TFP), or its growth rate. The debates about the sources of growth in the region that plagued the literature during a good part of the decade made policy makers believe that TFP was an important objective of policymaking. This paper explains the theory behind the concept of TFP. Underlying this theory is a series of problems that, when taken together, make TFP a dubious, misleading and useless concept for policy making. The relevant variable for policy making should be labour productivity.

Keywords: Total factor productivity
JEL classification: O47, O49

"Despite all this work, there is still no general agreement on what the computed productivity measures actually measure, how they are to be interpreted and what are the major sources of their fluctuations and growth."

Griliches (1988: 363)

1. Introduction

Seldom does a paper written by economists cause immediate impact among policy makers. Most economics papers are, let us face it, written in a language that is alien to most of the latter. It takes some rounds of discussions until the main message (if any) is distilled and understood, and then can be put to practice by those who guide the economy everyday. This was not, however, the case of Krugman’s (1994) paper on the lack of total factor productivity growth (TFPG) in East Asia during the period of high growth, and its alleged consequences. Fortunately, the debate about whether Krugman was right or not faded. What has not disappeared, however, is the impact that Krugman’s paper had on many policy makers: TFPG is a very important concept that policy makers ought to understand. Total Factor Productivity (TFP), it is argued, is a measure of the combined efficiency with

* Central and West Asia Department, Asian Development Bank, 6 ADB Avenue Mandaluyong, 1550 MM, Philippines. The author is also affiliated to the Cambridge Centre for Economic & Public Policy, University of Cambridge and, the College of Business & Economics and Centre for Applied Macroeconomic Analysis, Australian National University.

Email: jfelipe@adb.org

I am grateful to a referee of this journal for his comments. This paper represents the views of the author and does not represent those of the Asian Development Bank, its Executive Directors, or the countries they represent.
which all inputs are used. If it is high, it is good; and if it is low, then there is something wrong with the economy.

Based on work by fellow economist Young (1992), Krugman claimed that Singapore’s TFPG accounted for virtually zero of overall output growth between the mid-1960s and 1990. Although many policy makers in East Asia neither liked nor agreed with Krugman’s story, the fear of being compared with the Soviet Union led the Singaporean Government to create the Singapore Productivity and Standards Board (SPSB) and set the target of achieving a TFP rate of 2 per cent a year. In the words of the country’s then Prime Minister, “We have reached a stage where TFP becomes more than just a theoretical concept” (Huff 1999: 238). I will briefly review Singapore’s case below. Today, many medium and long-term development plans across developing Asia make reference to the need to increase total factor productivity (TFP);¹ policy makers in meetings discuss it; and organisations like the Asian Productivity Organisation devote energy and resources to estimate it (APO 2004).² There is indeed a whole industry around the notion of TFP. In Felipe (1999), I indicated that this was problematic and warned against what was termed the Solow residualization of the East Asian economies. I still stand by this view.

The objective of this paper is to explain in layman’s terms what TFP is, and what is not, and the pitfalls that surround its use as an objective of policy making. As a consequence, I will argue that it is useless and even misleading for policy purposes. The goal of policy making should be, among other things, to increase labour productivity, as this is the main determinant of output per capita. This is not to say that the notion of labour productivity is trouble-free. However, the problems that underlie this concept are nowhere near those that affect TFP. I refer the reader interested in the technical arguments to the bibliography.

The discussion proceeds as follows. In Section 2, I review the theory that underlies the concept of TFP. In Section 3, I discuss some problems underlying this theory and, as a consequence, I question the interpretation of TFP as a measure of technical progress. Section 4 provides a reinterpretation of TFP. I argue that it is not a measure of technical progress, at least the way it is commonly understood. The last section reinforces the main messages of the paper by emphasising that policy makers should care about labour productivity, not TFP, and sketches the main points of a policy agenda.

2. The Theory of Total Factor Productivity Growth: a Refresher
Policy makers no doubt understand what labour productivity is, namely, some ratio of output, ideally expressed in physical terms (for example, tons of steel) to labour, naturally, also expressed in physical terms (for example, number of hours worked or number of workers). Its interpretation is straightforward. The units are tons of steel per, let us say, worker. At any aggregate level, however, labour productivity cannot be expressed as the ratio of two physical magnitudes. While one may stretch the argument and ‘sum’ numbers of workers of different types and who perform different activities (although see next section where I argue that this is not so simple), and still argue that aggregate labour is a physical magnitude; one can certainly not do the same with output: tons of apples and tons of steel cannot be

¹ At the expense of labouring the obvious, TFP denotes the level of total factor productivity, while TFPG is the growth rate.
² A referee has indicated that the APO has stopped its annual publication.
summed up. And if one tries to add the outputs of the different agricultural products and of the services provided by the banking sector, this becomes an insurmountable difficulty.

This problem occurs, taking the argument to the extreme, at all levels of aggregation (that is, even thinking of different type of apples). There is only one solution, namely, to express aggregate and heterogeneous outputs in value or monetary terms (for example, dollars) (van Ark 2004: 33). In this case, the values of the different outputs can be summed up. It will be noted that deflation of a monetary series does not get rid of the problem. Indeed, using a price deflator does not convert nominal dollars into physical quantities (and real value added is not a physical quantity). For this reason, the standard units of labour productivity estimates are dollars (or whatever other currency) per worker (or per hour) of a base year. As one can see, this may not be a perfect measure of what intrinsically productivity is, but certainly is a useful one.

At least in principle, one should also be able to speak of the productivity of capital, the ratio of output to the capital stock, where the latter is measured in terms of numbers of homogeneous machines. At the aggregate level, however, one would have to divide total output in value terms by a measure of capital also expressed in value terms. This faces huge complications, as it is the result of 'summing' the values of different types of capital. Nevertheless, this is a concept not widely used, although its inverse, the capital-output ratio, is widely used.

At some point in time (the 1930s according to Griliches (1995)), it was argued that labour productivity was an incomplete measure of productivity because it does not take into account other factors of production, mostly capital. It seemed therefore natural to extend the idea of a ratio of output to input by including capital in the denominator, giving rise to the notion of multifactor or total factor productivity (an output-over-total-input index). But in doing this, economists opened a can of worms. To start with, capital is not really a factor of production, the same as labour. Without getting into technical details, machines are also an output that somebody produced. This means that using it in the denominator of the ratio raises some serious conceptual problems. Second, there is the problem of how to calculate this combined measured of inputs, for example, should they be added or multiplied? Moreover, how should they be weighted and on what basis? Third, there is the problem of interpretation: once one divides output in value terms by, say, the product of the number of workers times the value of the capital stock (with some appropriate weights), the result is a number with no units. The latter two are mostly index number problems.

---

3 See the analysis in Rampa (2002) who discusses some problems associated with equating real value added with physical output.

4 At the aggregate level (that is, with output and capital expressed in value terms) 'capital productivity' is a percentage. Suppose the ratio is 0.66, estimated with annual data. This implies a rate of 66 per cent.

5 This led to the so-called Cambridge Capital debates. See Cohen and Harcourt (2003).

6 An example may help clarify the issue: suppose that total output is USD30,000,000; the number of workers is 100; the value of the stock of capital is USD60,000,000; the weights of labour and capital are 0.4 and 0.6, respectively. One can calculate TFP as the following ratio:

\[
\frac{30,000,000}{(100)0.4 \times (60,000,000)0.6}\]

=102.38. The value of TFP alters as the units of measurement change. For example, TFP changes if one measures labour input as number of workers or total hours worked (while not changing value added and the value of the capital stock).
During the 1940s and early 1950s, economists were struggling with these questions and provided different partial answers. It was not until 1957 that Solow (two decades later Nobel Laureate) came up with an economic rationale that became the basis for most modern analyses and estimations of TFP. This solution was most ingenious at first sight. Solow (1957) related the notion of TFP, that is, the ratio of output to inputs, to the neoclassical aggregate production function. The latter is a representation of the aggregate technology, that is, of how inputs are transformed into outputs in the economy. Economists write this (in the most simplified form) as \[ Y_t = A_t F(K_t, L_t), \] where \( Y \) denotes output, \( A \) is an index of technology (or TFP), \( K \) denotes the stock of capital and \( L \) is labour. \( F \) simply means that the relationship between \( Y, K, \text{and} L \) can take different forms. The most simple such form is the so-called Cobb-Douglas production function, namely, \[ Y_t = A_t K_t^\alpha L_t^\beta, \] where \( \alpha \) and \( \beta \) are the elasticities of output with respect to each input (that is, the percentage by which output changes given one percentage point change in the corresponding input). In this particular form they are constant, but in general they need not be.

What did Solow do? Very simple: TFP, the level of technology, equals \[ A_t = \frac{Y_t}{F(K_t, L_t)}. \]

This expression can now be written in growth rates. Its interpretation is as follows: the rate of TFP, that is, \( \text{TFPG} \), equals the growth rate of output \((\dot{Y})\) minus the growth rate of capital \((\dot{K})\), multiplied by the elasticity of capital \((\varepsilon_K)\), and minus the growth rate of labour \((\dot{L})\) multiplied by the elasticity of labour \((\varepsilon_L)\): \[ \text{TFPG} = \dot{Y} - \varepsilon_K \dot{K} - \varepsilon_L \dot{L} \] (1)

This result, however, is of not much use for empirical purposes since empirical estimates of the elasticities tend to be unreliable. There is a way around this problem: it can be shown that under the assumptions of profit maximisation (the firm’s objective), competitive factor markets (that is, each factor input is paid its marginal productivity) and constant returns to scale (that is, the factor elasticities add up to unity), the factor elasticities equal the factor shares in the national accounts (that is, the shares of capital and labour), that is, \( \varepsilon_K = 1-a \) and \( \varepsilon_L = a \).\footnote{Both factor elasticities and factor shares change in time. I have not written the subscript denoting time to keep the notation as simple as possible.} Since the latter are available, now the formula to estimate TFPG becomes very easy to implement (the other three series, that is, the growth rates of output of labour and of capital, exist or can be constructed): \[ \text{TFPG} = \dot{Y} - (1-a) \dot{K} - a \dot{L} \] (2)

To obtain the level of technology, TFP, Solow came up with yet another ingenious solution, he constructed an index such that the base year was equal to 1 and then constructed the rest of the series by adding (or subtracting) TFPG (that is, the growth rate).

It is very important to understand what this procedure does: it provides an estimate (obtained residually) of the rate of technical progress (referred to as TFPG). The estimated measure is interpreted as that part of overall output growth that is not due to factor accumulation and it is assumed that it is, broadly speaking, a measure of the rate of technical
progress. Stated in different terms, the method splits or apportions the overall growth rate into three components, namely, the contribution of capital (calculated as the product of the share of capital in the national accounts times the growth rate of capital, \((1-a) \dot{K}\)), the contribution of labour (calculated as the product of the share of labour in the national accounts times the growth rate of labour, \(aL\)), and the contribution of technical progress (that is, TFPG), that is, \(\dot{Y} - (1-a) \dot{K} - aL\). This adds up, by construction, to overall growth (\(\dot{Y}\)), and the exercise is referred to as growth accounting. It is also important not to forget that undertaking the exercise entails, obviously, accepting the theory and assumptions behind it.

This now seemingly simple idea revolutionised empirical growth at that time and gave rise to countless papers during the 1960s. The methodology has been greatly refined (for example, there are techniques that do not impose the assumption of profit maximisation and competitive markets) during the last decades, but its essence that output growth can be split into different components, remains. Today, economists and statistical agencies apply the methodology mechanically, produce hundreds of estimates and derive policy implications (on this, see Felipe 1999). But what lies behind the numbers obtained?

3. Some Problems with the Theory and Notion of TFP

In this section, I shall discuss the implications of the derivation of TFPG and its interpretation. I shall discuss six issues (problems) that greatly undermine the rationale that underlies the notion of TFP and growth accounting exercises (Felipe 2006).

(i) It is imperative to call the reader’s attention to the assumptions involved in the method: profit maximisation, competitive markets and constant returns to scale. Hypotheses and assumptions matter because they allow the researcher to build a theory. Estimates of TFPG are not the result of a trivial derivation, but the result of a whole theory underneath them. The problem in this case is that one has only to look at the empirical basis from which the weights (that is, the factor shares) are derived to see that the neoclassical assumptions are very crude. These weights, remember, are the shares of labour and capital in total output. This framework presupposes competitive markets, although sometimes researchers omit this crucial point. However, even casual observation leads to the conclusion that market imperfections are prevalent in developing economies. In the height of the TFPG discussions following Krugman’s (1994) paper, Stiglitz claimed:

"Does anyone who has studied wage setting in Singapore, for example, really believe that wages are set in a competitive process, so that the real wage equals the marginal product of labour, as most of the studies assume?"

(Stiglitz 1997: 16).

As I said above, there are other methods that do not depend on these extremely restrictive assumptions. These other methods, however, do not solve all the problems, as the issues discussed below still apply (Felipe and McCombie 2003).

---

8 Certainly economists acknowledge that it captures more than just technical progress. Some even argue that it is no more than “a measure of our ignorance” (see Griliches 1995: 5).

9 For example, while explaining the growth accounting methodology, APO (2004: 24) states: “Replacing the marginal productivities by factor prices...”
(ii) There is a very simple, but at the same time very deep problem about what this methodology does: the so-called ‘cake problem’. To see what the algebraic splitting of growth means, consider what growth accounting does, according to Nelson (1981). Imagine one bakes a cake. One combines flour, yeast, water, sugar, etc. Then after the cake is baked, suppose one makes the following claim: 30 per cent of the size (or of the taste) of the cake is due to the flour; another 5 per cent is due to the water, and so on, and there is a residual 10 per cent that is due to the baker’s cooking skills. This may seem silly. However, this is what growth accounting does. One thing is to ask: what would happen to the cake (economy) if one added a given amount of extra flour (capital)? Or one may speculate about what it would have happened to the cake (economy) if it had been baked (managed) by a more competent baker (Governor of the Central Bank). But this is different from apportioning the overall result to the individual components. Growth cannot be split the way it is done in growth accounting exercises because it does not make sense. Growth is the result of the interaction of a myriad of factors. Moreover, one has to be careful in interpreting these decompositions, as factor accumulation and productivity growth are both ‘endogenous.’ What this means is that finding that factor accumulation accounts for 75 per cent of growth, for example, does not imply that growth would have been 75 per cent as high in the absence of technical change. Indeed, in the absence of productivity change, the incentive to accumulate would have been much lower, and the resulting capital accumulation would have also been significantly lower. Or, stated in different terms: how is it possible to split the contributions of physical capital, labour and technology in the case of IT services? Aren’t capital and technical progress the two sides of the coin? What is the meaning of separating this from the contribution of labour? Who runs the computer?

The relevance of Nelson’s critique of standard growth accounting exercises is more obvious if one considers growth accounting exercises that include human capital as another, separate, factor of production. The role of human capital is to recognise that labour in different economies, or at different points in time in the same economy, may possess different levels of education and different skills, and therefore it has different capacity to perform work. However, the inexplicable aspect of this exercise is the separation of labour and human capital as ‘distinct factors of production.’ Indeed, it is very difficult to understand and comprehend what labour and human capital are and mean as distinct entities (see Braverman 1998: chap. 1).

My view is that one can list the possible sources of growth of an economy the way, for example, Olson (1996) does, that is, as an organisational device, or as a tool to think about growth in a systematic way. However, another quite different thing is to try to apportion these sources to account for overall growth the way growth accounting exercises do.

(iii) There is also a problem of interpretation of the results, neatly summarised in the following two quotations:

“If the Koreans do not have the TFP of the USA in the fifties despite having copied them, what can we say about this method? If Japan shows significant TFP during the fifties and Korea is the country that most closely followed the Japanese path to development, how is it that Korea does not show the same TFP? Since Singapore grew through heavy direct foreign investment, does the low TFP indicate a failure of foreign firms to use modern technology?”

(Rashid 2000: 152).
What Policy Makers Should Know About Total Factor Productivity

and:

"Alwyn Young’s (1992) often-cited study arguing that the freedom of markets in Hong Kong, China can explain the relatively rapid increase in its total factor productivity illustrates how the Solow technique can yield erroneous results. Not only is it the case that the measurement of total factor productivity increases can be unreliable [...] but the interpretation of the residual, what is left over after measuring inputs, is highly ambiguous. Assume that one could feel confident that Hong Kong’s residual was greater than that of Singapore. Is it because of better economic policies? Or is it because Hong Kong was the entrepôt for the mainland of China, and as the mainland’s economy grew, so did the demand for Hong Kong’s services? In this interpretation, Young’s explanation of Hong Kong’s higher TFP relative to Singapore is turned on its head: Hong Kong’s success actually was a result of the growth of perhaps the least free-market regime of the region”

(Stiglitz 2001: 512).

Therefore, as the following quotation indicates, the use of TFP (or TFPG) for purposes of analysis and policymaking is not a straightforward issue:

"The question then to be answered is whether the residual effect of ‘technical progress’ corresponds to anything interesting. I rather doubt it. There is no reason to suppose, for example, that technical progress, so defined, measures the effect of research and development expenditures. Indeed, I cannot think what it measures, except (tautologically) the difference between an actual increase in output and a purely hypothetical increase, which is based on a set of definitions that I can see no reason for using”

(Scott 1989: 88).

However, and despite these grim evaluations by well known economists, authors still interpret TFP estimates as measures of technical progress, and try to derive conclusions about the validity of different policies and growth strategies. Some authors have even run regressions with TFPG as the dependent variable on variables such as openness, R&D expenditures, inflation, and government expenditures. It is very difficult to interpret these regressions (see, for example, some of the papers in APO (2004) and the policy recommendations suggested). It is like playing the ‘piñata’ game to see if one is lucky and a couple of variables turn out to be statistically significant so that the researcher can ‘draw policy implications.’ This is bad economics.

(iv) Most estimates of TFPG are incorrect, even in the context of the neoclassical framework, in the sense of being ‘biased downward’ (Felipe and McCombie 2001). The reason is that the purpose of the growth accounting decomposition is to calculate the contributions to total output growth of technical progress and factor accumulation. There is a potential problem here. Recall that factor accumulation is computed as the sum of the growth rates of capital and labour, each multiplied by its factor share \( (1 - a) \dot{K} + a \dot{L} \). However, the factor shares that one uses are the ‘actual’ ones, taken from the National Accounts. These factor shares have been affected by technical progress and, therefore, the estimates of TFPG are incorrect. There is only one case when they would be correct. This is when technical progress had been Hicks-neutral. This means that technical progress did not change the relationship between the marginal products of capital and labour, and therefore it did not
affect the factor shares.\textsuperscript{10} Unless one can prove that this was indeed the case, it does not seem to be a realistic assumption. If on the other hand, technical progress was biased, in the sense that the relationship between the marginal products of capital and labour changed in such a way that the capital share decreased in time, then this has implications. As Nelson and Pack (1999) argued, this is very likely what happened in East Asia during the period of high growth. Why does this matter? In most countries capital grows at a faster rate than labour (that is, $\dot{K} > \dot{L}$). If one then multiplies the growth rate of the capital stock by the “capital share after eliminating the effect of technical progress” (a successively smaller capital share), then TFPG will increase. This revised TFPG would be

$$TFPG^* = \ddot{Y} - (1 - a^*) \dot{K} - a^* \dot{L},$$

where the asterisk denotes the factor shares after eliminating the effect of technical progress. This is higher than the conventional measure of TFPG. Empirical evidence showing that this is the case is provided in Felipe and McCombie (2001). None of the papers in the recent volume by APO (2004) considered this possibility.

This argument has another important implication. This is that if TFPG were calculated for a sufficiently long period, when the capital share should have decreased to zero, then

$$TFPG^* = \ddot{Y} - a^* \dot{L},$$

(where $a^*$ is the labour share after eliminating the effect of technical progress), that is, TFPG would coincide with the growth rate of labour productivity.

(v) Often it is argued that the relevant concept for long-term growth is productivity, in particular total factor productivity. Hence, this should be the relevant variable for policy purposes. Well, this is a half-baked truth, in the sense that it is the result of a model, and thus it is a theory-dependent result. This model is Solow’s (1956) original neoclassical growth model. That model also depends on a number of extremely restrictive assumptions. Under such assumptions, and in steady state,\textsuperscript{11} the model predicts that countries with high saving/investment rates will tend to be richer (per capita levels). These countries accumulate more capital per worker, and consequently have more output per worker; and countries that have high population growth rates will tend to be poorer. But savings rates and population growth do not affect the steady-state growth rates of per capita output. The latter is given by the rate of technical progress. In this model, this is TFPG. Therefore, yes it is true that TFPG determines long-run growth... but in the context of the neoclassical growth model. It is not a self-evident fact.

My view is that the relevant concept for long-run growth is Harrod’s ‘natural rate’ (Harrod 1939; see also Pasinetti 1974: essay IV) that is, the sum of the growth rates of labour-augmenting technical progress and population. This represents the maximum sustainable rate of growth that technical conditions make available to the economic system. This is a concept that policy makers can track.

\textsuperscript{10} This argument runs into the problem that Hicks-neutral is often impossible. Indeed, Steadman (1985) derived a number of sufficient conditions to render the concept of Hicks-neutral technical progress impossible for being, at the level of the economy, theoretically inconsistent.

\textsuperscript{11} Steady state refers to the situation where all variables grow at the same rate. This is supposed to happen in the long run.
vi) The final issue is so important that it cannot go unnoticed for, once understood, it should send the notion of TFP and the idea of decomposing overall growth, where it belongs to, namely the annals of economic history. This is the so-called aggregation problem. In theory, the production function is a representation of how a commodity (for example, a chair) is made. Output, labour and capital are measured in physical terms. The production function corresponding to the chair is $q_t = f(k_t, \ell_t)$, where output ($q$), labour ($\ell$) and capital ($k$) are measured, I insist, in physical terms. Each commodity or service that an economy produces (e.g., carrots, chairs, ships, banking services, oil, etc...) is made using a different production process, that is, with a different technology, and with different machines and workers (and possibly with other additional inputs, different in each case). As far back as the 1940s, economists started inquiring about the conditions under which so-called micro-production functions like the one above could be summed up so as to yield the aggregate production function $Y_t = A_t F(K_t, L_t)$ representing the technology to make something called ‘aggregate output’ ($Y$), the sum of the individual outputs of carrots, chairs, ships, banking services, oil and the millions of other products and services that an economy churns out every year, with some aggregates of capital ($K$) and labour ($L$). In this aggregation process, it is important to emphasise that the function $F$ has to be what economists refer to as ‘well behaved’, for example, with positive and diminishing marginal products.12

Franklin Fisher, among others, spent years in the 1960s, 1970s and 1980s working out different aspects of the problem. The results, for the interested reader, are summarised and discussed in Felipe and Fisher (2003; 2006). The conclusion is that the conditions under which aggregate production functions can be derived by aggregating micro-production functions are so stringent that it is difficult to believe that actual economies satisfy them. For example, a labour aggregate (yes, labour of different types also have to be aggregated) will exist if and only if a given set of relative wages induces all firms to employ different types of labour in the same proportion. Note that this condition requires the absence of specialisation in employment! Similarly, where there are many outputs, an output aggregate will exist if and only if a given set of relative output prices induces all firms to produce all outputs in the same proportion. This condition requires the absence of specialisation in production; indeed all firms must produce the same market-basket of outputs differing only in their scale. And do not forget that these conditions are derived under the assumption of constant returns to scale. Outside this case, forget about deriving aggregation conditions. Are these conditions true in the real world? Certainly not. They mean that, for all practical purposes, aggregate production functions do not exist, as they are constructs without sound theoretical foundations (Felipe and McCombie 2005).

It is important to emphasise that I am not saying that output and inputs are not linked. Certainly, a chair is made with labour, capital and intermediate materials. And I am not saying either that aggregate output (Y) does not exist. Aggregate output is the sum of private consumption, investment, government expenditures, and net exports (from the demand

---

12 In algebraic terms, and at the expense of simplifying a bit of the problem, the question is: when is it correct to aggregate the individual production functions of two firms, such that $q_1 + q_2 = f(k_1^1, \ell_1^1) + f(k_2^2, \ell_2^2) = F(K, L) = Y$?
side of the economy). Rather, the implication of the above paragraph is that the representation of an economy's aggregate technology as \( Y = F(K, L) \) is fictitious. There is no such a thing as the output of the economy (\( Y \)) from the point of view of neoclassical production theory, that is, a function that transforms aggregate capital (\( K \)) and aggregate labour (\( L \)) into aggregate (\( Y \)). It is worth quoting Felipe and Fisher on this issue:

"The use of aggregate production functions in theoretical and applied work reminds us of the long process it took mankind to understand the motions of the planets, delayed for over two thousand years. It resulted in the perpetuation of the Ptolemaic geocentric model of the movement of the planets long after Aristarchus, circa 250 B.C., first questioned this system and put forward the alternative heliocentric theory. This was partly due to the fact that by the use of epicycle after epicycle to describe the motions of the planets (although the epicycle is a totally artificial construct), the Ptolemaic system produced very good predictions. But prediction is not the same as explanation."

(Felipe and Fisher 2006: 157)

Why does all this matter? In the words of Nadiri, the aggregation problem matters because:

"without proper aggregation we cannot interpret the properties of an aggregate production function, which rules the behaviour of total factor productivity"

(Nadiri 1970: 1144)

It is important to ask if economists are aware of these six issues. Most are certainly aware of the first point, the assumptions underlying the method. However, for strange reasons, many economists do not seem to mind them and apply the method mercilessly, even to economies that greatly depart from the assumption of competitive markets, as reflected in Stiglitz's quotation above. Even worse are the analyses of China based on this methodology. The following justification reflects the poor state of affairs:

"The estimates of productivity growth for China may be biased in either direction if there are deviations from the assumptions imposed by the adopted methodology [...] However, since this methodology is widely used in studying sources of economic growth for members of the Organization for Economic Cooperation and Development, the newly industrialized economies of East Asia, and many developing countries with divergent income levels and economic structures, it is of interest certainly as a first step, to apply the same analysis to the Chinese economy to obtain what could be viewed as 'benchmark' estimates."

(Hu and Khan 1997: 108)

Three issues are worth mentioning regarding the above quotation. First, the authors admit that the assumptions needed to apply the method (see above) probably do not apply to the Chinese economy; second, applying a methodology simply because others use it is a very poor argument; and third, it is difficult to understand why anyone would like to calculate "as a first step" a number that is known to be inaccurate and whose interpretation is not clear, to say the least. How can this number be used as a "benchmark"? The interested reader can see Felipe and McCombie (2002a).

The cake and interpretation problems are trickier. Some economists dismiss the cake problem as an interesting and funny story. They argue that it is sensible to split total growth in order to get an idea of the contributions of the different inputs to total output. And they are convinced that TFGP is a proxy for the rate of technical progress. Unfortunately, the problem is true and "splitting total growth into its alleged sources" is a meaningless exercise.
Most economists do not seem to know how to be aware either of the fourth issue, namely, the implications of the fact technical progress is, most likely, not Hicks neutral. And since most economists today (certainly the younger generation) have been only exposed to the neoclassical growth model, they are convinced that TFPG determines long-run growth. Finally, most economists today are not aware of the damaging implications of the aggregation problem. It simply invalidates the whole rationale underlying the procedure. It cuts its oxygen supply. Economists like Kaldor (1957), Pasinetti (1959), Nadiri (1970), Nelson (1973; 1981), or Scott (1989) have aired on different occasions and from different points of view, the many problems underlying the concept of an aggregate production function and, as a consequence, of constructs such as TFP. Certainly Samuelson and Solow (and surely many others) are very well aware of these issues. Samuelson distanced himself from these concepts long ago. Solow, on the other hand, has never abandoned the idea of an aggregate production function, despite knowing very well the aggregation problem (he published a paper with Fisher and Kearl on the topic in 1977). But a large majority of economists is not even aware of it.\footnote{Incidentally, years ago, giving a technical seminar on the aggregation problem somebody asked me what was wrong with assuming that the aggregate production function exists. My reply was along the following lines: “You can assume whatever you want, but it will not be so just because you assume it.” I stand today by this same reply.}

The conclusion, in my view, is as follows: why would any policy maker be concerned with a number that requires very stringent assumptions to be derived; that is difficult to square with what growth is about; whose interpretation is not a clear cut issue; that is possibly biased downward because it does not consider that technical progress may be biased; that is of not much use for policy purposes; and that has no theoretical (economic) foundations? See also Felipe and McCombie (2002b).

4. So, What is TFP?

If the aggregate production function, the construct from which TFP and TFPG are derived does not exist, in the sense that it is a concept without sound economic foundations, what are economists, planning and statistical agencies calculating?

To provide an explanation of what TFPG is, I will proceed in two steps. First, I shall show what TFPG equals to if one looks at the National Income and Product Accounts (NIPA). Then, I will ask if this figure is something policy makers should care about. The answer is that perhaps yes, but if and only if one could show that such numbers represent, unambiguously, the rate of technical progress. I will argue that this is not the case.

As seen above, the standard derivation of TFPG consists in subtracting the growth rates of capital and labour, each weighted by its share in total output, from the growth rate of output. As all national accountants know, total output \( Y \) (disregarding conceptual differences between factor cost and market prices and gross and net) equals the sum of the total wage bill \( W \) plus the surplus of the economy (or total profits) \( P \). This is true by definition. Algebraically, \( Y \equiv W + P \). It is also true that the country’s wage bill can be written, as a matter of definition, as the product of the country’s average wage rate \( \bar{r} \) times total employment \( L \), that is, \( W = \bar{r} x L \); and that total profits can be written as the product of an average profit rate \( \bar{r} \) times the stock of capital \( K \), that is, \( P = \bar{r} x K \). This is implies that \( Y \equiv \)}
$W + P \equiv (w \times L) + (r \times K)$. It is now straightforward to express this relationship in growth rates (to simplify I do not write the subscript representing time, but, in general, all variables vary): the growth rate of output ($\hat{Y}$) equals sum of the growth rates of the wage rate ($\hat{w}$), profit rate ($\hat{p}$), employment ($\hat{L}$) and capital stock ($\hat{K}$), each weighted by its share in total output (these shares are the labour share ($a = (w \times L) / Y$) for the growth rates of the wage rate and of labour; and the capital share ($1-a = (r \times K) / Y$) for the growth rates of the profit rate and of the capital stock). Algebraically:

$$\hat{Y} \equiv a\hat{w} + (1-a)\hat{p} + a\hat{L} + (1-a)\hat{K} \quad (3)$$

It thus now follows that the difference between the growth rate of output and the growth rates of capital and labour, each weighted by its share in total output, that is, $\hat{Y} - a\hat{L} - (1-a)\hat{K}$ must be equal to the sum of the growth rates of the wage and profit rates, weighted by the labour and capital shares, respectively, that is, $a\hat{w} + (1-a)\hat{p}$. This is what TFPG equals to:

$$TFPG \equiv \hat{Y} - a\hat{L} - (1-a)\hat{K} = a\hat{w} + (1-a)\hat{p} \quad (4)$$

The reader will appreciate that I have not talked about aggregate production functions or made any assumption regarding the state of competition. The important point to note is that $\hat{Y} - a\hat{L} - (1-a)\hat{K}$ is exactly what we obtained above by assuming (i) that an aggregate production function exists, (ii) competitive markets, (iii) profit maximisation, and (iv) constant returns to scale.

Some economists will argue that I am simply rediscovering the wheel, that is, the above is well known. For example, Barro (1999) and Hsieh (2002) have argued that growth accounting can be performed using the income accounting identity. So, what am I saying? The issue is subtle but very important. When these authors also use the accounting identity to derive TFPG, they interpret the sum of the growth rates of the wage and profit rates (each weighted by the labour and capital shares), as a measure of technical progress. But you have to ask why? In other words, one should then ask if growth accounting can be performed directly from the accounting identity. The answer is obviously yes. The key issues, however, are about ‘theory’ and ‘interpretation’. Recall that toward the end of Section 2, I argued that performing a growth accounting exercise entails the acceptance of the theory and assumptions behind it. However, there is neither theory nor assumptions behind the accounting identity. Therefore, while one can also artificially decompose output growth through the accounting identity, its meaning is what is at stake; for, on what basis is the term in the NIPA identity (in growth rates) a measure of the rate of technical progress? Although these authors do not say it explicitly, they have in mind an aggregate production function, and that is how they associate this weighted average with TFPG. 14 Recall that from the production function one can derive $\hat{Y} = TFPG + (1-a)\hat{K} + a\hat{L}$; while from the National Accounts we know that $\hat{Y} \equiv a\hat{w} + (1-a)\hat{p} + a\hat{L} + (1-a)\hat{K}$. Then it is obvious,

---

14 They do this through the so-called Euler’s Theorem.
the argument goes, that $TFP_G = \alpha \hat{w} + (1-a)\hat{f}$. However, if the aggregate production function does not exist, why are we to interpret this weighted average in the accounting identity the way it is done in neoclassical economics, that is, as a measure of the rate of technical progress?

Is there any way to show that this weighted average is truly the rate of technical progress? It is true, and this is obvious, that the higher this weighted average 'the better.' But does it mean that this weighted average has an intrinsic and a clearly defined economic relevance? If it had it, then policy makers should know how to affect it, that is, increase it. The reality, however, is far from this. Are improvements in TFPG the result of more R&D, education, etc.? If, as Griliches (1988) argued (quoted at the start of the paper), it is not clear what TFPG measures are, then we cannot know what their determinants are. Therefore: "Public exhortations for deliberate efforts to 'improve' the rate of growth in aggregate productivity suffer from an underlying contradiction in logic. We simply cannot hope to affect consciously something that is defined to measure our lack of knowledge" (Sudit and Finger 1981: 7).

Note that since TFPG involves the factor shares and the growth rates of the rewards to labour and capital, it could well be referred to as a measure of income distribution (not in a zero-sum sense); and since I have derived it from the income accounting identity, it brings to mind issues relating to the functional distribution of income between labour and capital. Thus, the weighted average might go up and down for reasons unrelated to the technical conditions of production. Indeed, factor shares could merely reflect the bargaining power of labour and capital, and as such, they could change without there necessarily being any change in the technical conditions of production. Hence, in this case, factor shares cannot be taken as reflecting the aggregate output elasticities, even if these exist.

The original paper that provided the results upon which Krugman (1994) based his analysis, namely Young (1992), argued that Singapore's lack of TFPG was related to the country's extremely fast rate of structural transformation, and the latter was the result of the government's aggressive targeting policies. These prevented the city-state from realising all the learning-by-doing at each stage of development and implied a premature movement up the development ladder, which resulted in a fall in measured productivity and increasing costs of production. In simple terms: the government's policy of encouraging structural transformation was a mistake. While this thesis could be true (that is, Singapore pushed itself into technologies too far ahead of itself to benefit from learning by doing), the evidence is not clear. In fact, Singapore has done well and is one of the most technologically advanced countries in Asia. And this outcome is consistent with recent literature that argues that countries that push themselves and learn into areas where they do not have a comparative advantage, do well (Hausmann et al. 2005; Rodrik 2006; Hausmann and Rodrik 2006).

It is worth noting that indeed it is true that $TFP_G = \hat{y} - aL - (1-a)\hat{K} \equiv 0$ for Singapore, as Young (1992) documented. And from the accounting identity, we know that this implies that $TFP_G = a\hat{w} + (1-a)\hat{f} \equiv 0$, which itself implies that $a\hat{w} = -(1-a)\hat{f}$. Given that in Singapore $a \approx (1-a) \approx 0.5$ according to Young's (1992) data, all this implies is that

---

15 Young (1992) implicitly praised Hong Kong, whose TFPG was substantially higher than that of Singapore. Hong Kong's laissez faire economy had resulted in higher TFPG.
\( \hat{\omega} \equiv -\hat{r} \). What happened in Singapore between 1965 and 1990 was that the positive average growth rate of the wage rate was matched by an approximately equal but negative growth rate of the profit rate; and given that labour and capital took approximately the same share in total output, this implied a zero ‘residual.’ This is an interesting result, but it does not follow from any theory; and, it is not clear that it has to be attributed to the supposedly negative effects of Singapore’s industrial policies (Felipe 2000).\(^{16}\)

Finally, it could be argued that $TFPG \equiv \hat{Y} - a\hat{L} - (1 - a)\hat{K}$ is a weighted average of the growth rates of (aggregate) labour and capital productivity as it can be written as $TFPG \equiv a(\hat{Y} - \hat{L}) + (1 - a)(\hat{Y} - \hat{K})$. Is this interpretation meaningful? This derivation is obviously logically correct, as it is derived from the identity. It is important to emphasise, though, that here labour and capital productivity are constant-price value terms and not physical quantities. And (once again) there is no reference to a production function that determines the ‘contributions’ of capital and labour accumulation to output growth. To address this question, Felipe and McCombie (2006) asked if there is any relationship between the true rate of technical progress, calculated with purely micro and firm level data in physical terms and TFP estimated with aggregate data. To answer this question, they ran a simulation exercise. The exercise consisted of 10 firms whose output, labour and capital were given in physical units. The setup was such that all firms had the same rate of technical progress, 0.5 per cent per annum. As the rate of technical progress is the same for each firm, one could think of the rate of technical progress of the economy being 0.5 per cent per annum. But what happened at the aggregate level? To estimate the aggregate TFP, prices were used to generate the aggregates. These were set as a markup on unit labour costs. This markup gives the factor shares (these are unrelated to a production function).\(^{17}\) TFP was calculated through the standard method. The figure obtained was completely different, 1.48 per cent per annum, very different from the true rate of technical progress (0.5 per cent per annum).\(^{18}\) This implies that the only possible and meaningful way of talking about

\(^{16}\) It is also easy to explain why the Philippines’ TFP is often negative (APO 2004: 255-279). Felipe and Sipin (2004) showed that the Philippines’ profit rate is approximately constant. This means that \( \hat{r} \equiv 0 \). On the other hand, the real wage rate has declined for a long time (APO 2004: 258). This implies that \( \hat{\omega} < 0 \). Putting these two pieces together implies that $TFPG = \hat{Y} - a\hat{L} - (1 - a)\hat{K} = a\hat{\omega} + (1 - a)\hat{p} < 0$. Once again, there is no need to mystify this finding. And it is difficult to believe that the determinants of TFP in the Philippines (essentially \( a\hat{p} \)) are (APO 2004: 270-271) the ratio of exports to GDP, the ratio of imports to GDP lagged one year, the ratio of FDI to GDP lagged one year, the ratio of the share of manufacturing value added to GDP, the ratio of R&D expenditures to GDP lagged to years and the annual change in the GDP deflator.

\(^{17}\) Suppose prices are set as a markup on unit labor costs, that is, $P = \mu \frac{w}{(Y/L)}$ where $\mu$ is the markup. Assume $\mu = 1.33$. This implies that the labour share is $\frac{wL}{PY} = \frac{1}{\mu} = 0.75$. Hence the capital share is 0.25.

\(^{18}\) One implication of this result is that the arguments in this section do not apply if one had measures of output and inputs in pure physical quantities. In this case, one could estimate TFP (there is no aggregation problem at least).
multifactor or total factor productivity is at the microeconomic level, with TFPG computed with physical quantities.

The conclusion is that it is very difficult to argue convincingly that the residual measure TFPG is, unambiguously, a measure of the ‘aggregate’ rate of technical progress, or of anything meaningful that policy makers should care about because it can be affected by policy making. From the point of view of the aggregate production function, the interpretation of TFPG as a measure of the rate of technical progress is questionable, as the aggregate production function does not exist; and from the point of view of the income accounting identity, TFPG is a weighted average of the growth rates of the wage and profit rates. This weighted average can increase or decrease due to changes in the factor shares for reasons unrelated to the technical conditions of production.

5. By Way of Conclusion: Dear Policy Maker...
I hope that by now you start having serious doubts about the meaning and usefulness of the notion of total factor productivity. While the idea of splitting the rate of output growth seems attractive and powerful, the truth is that it has misled researchers and policy makers for decades. It is time to abandon it. I stress four points:

First, policy making is about understanding one’s economy and using the tools of economics to attain clearly defined objectives. TFP is neither a tool nor an objective. Some authors have even claimed that economics needs a theory of TFP. This is very dubious (see Felipe and McCombie 2007). It will only give you a headache because you will never understand the real meaning of the number that you calculated; why it moves up or down; or the policies to ‘improve’ it. Simply disregard it and do not waste time calculating it.

Second, algebraically, output per capita \( (Y/L) \) is the product of labour productivity \( (Y/L) \) – the value of the goods and services a country can provide with a given amount of workers, times the employment ratio \( (L/P) \) – the fraction of people who work. The latter varies somewhat around the world, but not that much. Labour productivity, however, varies enormously around the world. Differences in labour productivity explain most of the differential in GDP per capita around the world. This can be seen by plotting GDP per capita \( \text{vis-à-vis} \) labour productivity. Figure 1 shows the large differential between most developing Asian countries and the OECD average. Other than Hong Kong, Singapore, Taiwan and Korea, the level of productivity in the rest of developing Asia is still today very low (below the OECD average during the early 1970s). Policy makers in developing countries should devote their efforts to understanding the reasons for these differences in labour productivity with respect to the developed countries. The reasons underlying these differences are very complex and not easy to deal in terms of policy (Raes et al. 2004). These causes lie, probably, at the sector level (Lewis 2004). Indeed, one can think of aggregate labour productivity as the weighted sum of the sector level productivities, that is \( Y/L = \sum \lambda_i q_i \), where \( \lambda_i = L_i / L \) is the share of each sector’s employment in total employment, and \( q_i \) denotes each sector’s level of labour productivity.\(^{19} \)

\(^{19} \) A word of caution: aggregate labour productivity understood as the weighted average of the various sectors’ labour productivities in value terms also suffers from the aggregation problem. One implication of Felipe and McCombie’s (2006) simulations discussed above is that firm-level labour productivity (in physical terms) will differ from the aggregate (in value terms). Nevertheless, the notion of aggregate labour productivity, as a value measure, is certainly less cumbersome than total factor productivity.
look into at this level, and in particular at the firm level, in order to understand the determinants of labour productivity growth? Raes et al. (2004) indicate that the determinants are physical capital, human capital, innovation, competition, macro-fundamentals and firm dynamics. To this, one would have to add for the developing countries, “the invaluable contribution of ‘mere imitation’” (Baumol 2004:196-197).

Third, growth and development policies should promote structural change, higher savings, better education, a more skilled labour force, technological development and innovation – reflected in labour productivity growth—are the main drivers for the wealth of nations. If the economy is doing well, improvements in these variables should be reflected in increases in wages, labour productivity and in output per capita, the measure par excellence of well being. Productivity is definitely a key concept (most likely both as a tool and as an objective), but the measure to follow is labour productivity. Despite some measurement problems, it provides a much clearer gauge of how well the economy is doing. And Harrod’s (1939) model provides a rationale for why the growth of labour-augmenting technical progress is the relevant concept for long-run growth.

Finally, in this context, what is the role of government? If higher labour productivity growth is indeed a desirable objective, for it is the key long run growth, “a policy agenda towards higher

---

20 On this, see Felipe and McCombie (2005). The neoclassical model claims that steady-state labour productivity is a function of the savings rate and of the growth rate of population. There are empirical problems with this relationship.
productivity growth should take market failure as its starting point, and should be carefully designed so as to avoid government failure and reduce institutional failure" (Raes et al. 2004: 313).

References
Jesus Felipe


What Policy Makers Should Know About Total Factor Productivity

