Chapter 14 of *Palgrave Handbook of Economic Performance Analysis*, Thijs ten Raa and Bill Greene (eds.), New York: Palgrave McMillian, forthcoming, 2019.: 1-65.

14 An Overview of Issues in Measuring the Performance of National Economies

Anthony Glass,

Karligash Kenjegalieva,

Robin C Sickles,

Thomas Weyman-Jones

Contribution to *Handbook of Economic Performance Analysis*,

edited by William Greene and Thijs ten Raa

This draft January 15 2019

14 Issues in Measuring The Performance of National Economies

1 Introduction

In this chapter, we will review the ways that economists measure the aggregate economic performance of national economies. This is the lead-in to a number of separate chapters that develop particular themes so that this chapter is intended to give an overview and anticipation of general issues that may be met in more detail subsequently. Efficiency and productivity analysis using the methodologies of data envelopment analysis and stochastic frontier analysis has made a significant contribution to this challenge after the initial research which arose in the context of the analysis of economic growth. That initial research led to the idea of measuring total factor productivity change, TFP[1], and its identification with an unobserved data residual representing technological progress. The contribution of efficiency and productivity analysis has been to expand our understanding of what TFP could consist of and what could drive it and how we can extend our understanding of it beyond the idea of an unexplained data residual. Amongst the critical questions in this search is the exact definition of what measure of economic performance economists should use. The conventional answer is to measure economic performance by real gross domestic product, GDP, i.e. the gross value-added measure of GDP. However, it has been frequently suggested that a broader measure of economic welfare should be used and research in this area is particularly lively now in the early part of the 21st century.

It is important also to be clear about what this chapter cannot do and does not do. We cannot properly survey the existing literature on the performance of national economies since there

[1] We adopt the usage of representing *total factor productivity change* by the symbols TFP since that is the convention adopted elsewhere in the Handbook. Much of the macroeconomic literature simply calls this total factor productivity even though it is measured by the difference between two weighted rates of change over time. Where the reference is to the level of total factor productivity, we point this out.

are already tens of thousands of papers on this topic and no selection could possibly give a balanced or even-handed guide to this vast literature. Nor do we plan to survey the methodologies involved in measuring the performance of national economies using efficiency and productivity analysis: by these we mean growth accounting and the construction of index numbers, data envelopment analysis DEA, including free disposal hull methods FDH, stochastic frontier analysis SFA, stochastic non-parametric envelopment of data methods StoNED and other non-parametric regression methods. The reason is clear: the remainder of this book treats these methods in detail and it would be foolish to offer any duplication of this material. The purpose of this chapter, once these two impractical directions are excluded, therefore arises from the title and the idea of introducing the comparative ideas and themes that a researcher into the performance of national economies might want to consider when evaluating this massive research challenge. We have deliberately organized our range of topics very widely in order to meet the challenge set for us and while we are not so naïve as to imagine that our selection will meet with wide approval, we hope that we will stimulate researchers to think broadly about the sort of issues that a non-specialist might ask about when considering the wide-ranging topic of the performance of national economies.

The most important theme that we wish to emphasise is that we have interpreted the concept of the relative *performance* of national economies very widely. Anyone familiar with the upsurge in questions by economics students about the relevance of their studies will know that neoclassical economics and the overriding dominance of GDP as the only measure of performance is under serious debate. A good example of this is the CORE project[2] which is an innovative approach to widening the economics curriculum in response to student-centred requests and which is now being widely adopted in Europe and the USA. Consequently, we devote space to examining a wide range of different concepts of economic performance

[2] See https://www.core-econ.org

including but certainly not limited to the value-added definition of GDP. In part this reflects the expanding interest in behavioural economics, see Thaler (2018) who has emphasised how individual decision makers in reality appear to use concepts of altruism, fairness, subjective adjustment of objective frequencies, and heuristics that lead to behaviour that is at odds with the idea of maximising productivity growth.

This chapter is structured to fall broadly into two parts. The first part, which consists of sections 2 and 3 begins with the conventional neoclassical[3] definition of TFP using the change in the gross value-added measure of GDP in order to show what needs to be assumed to arrive at the identification of TFP with the unobserved residual that represents technological progress. There is a wide debate on why this measure of technological progress appears to have slowed down considerably in developed economies in recent years. The phrases 'productivity slowdown' or 'productivity gap' have become common in public discussion. Theories range from the idea that humanity has run out of new technological ideas all the way to secular stagnation meaning there is nothing worthwhile left in which to invest. We do not survey all of these ideas but we do discuss the methodological context in which they are considered.

It may of course simply be that the TFP measure used for this debate is severely at fault. This raises two types of question. First, do the conventional measures of GDP, i.e. aggregate expenditure on final goods and services or aggregate gross value-added, exclude important components of GDP? We briefly review some of the most important recent contributions to this question. Second, is GDP the appropriate indicator of national economic performance? Therefore, we follow this by a discussion of other measuring metrics of economic performance, for example economic welfare which can present a totally different picture.

[3] The proponents of the conventional treatment of TFP as a residual use the term *neoclassical* to describe it.

However, since the majority of the empirical work until now has focused on the measurement of economic performance in terms of GDP or GDP per capita, we treat this approach in the second part of the chapter, which consists of sections 4 to 8. We begin this part with the approach of efficiency and productivity analysis which explicitly relaxes the strong assumptions made to achieve the conventional residual TFP measure. By efficiency and productivity analysis we mean the whole range of methodologies which flowed from the pioneering work of Farrell (1957), Charnes Cooper and Rhodes (1978) and Aigner Lovell and Schmidt (1977), i.e. data envelopment analysis, stochastic frontier analysis and all the subsequent developments that measure the distance of economies from their technological frontier. We show how stochastic frontier analysis and data envelopment analysis modelling has been able through the idea of TFP decomposition and the measurement of inefficiency to tell us much more about TFP than the conventional approach. In particular, we show how these methodologies are able to relax the assumptions needed for the conventional neoclassical approach and we discuss how they attempt to model the components of TFP. We can conveniently classify these methodologies into regression-based approaches[4] and programming-based approaches depending on the importance attached to errors of measurement, sampling and specification. We discover that in the regression-based approach the critical ingredient in the TFP decomposition is the computation of the elasticities of the production, input distance, output distance or cost functions. In the programming-based approach, the critical ingredients are the estimated efficiency scores under different constraints. Naturally we leave to the other chapters in the handbook the technical details of implementing these methodologies.

We review a number of critical issues such as whether the size of national economies matters, whether there is an important role for exogenous variables in explaining the unobserved TFP

[4] In the regression-based methodology we concentrate on the frequentist approach and do not include discussion of Bayesian methods

residual, and the role of incentives to be efficient arising from the market structure of the economy. We will explain how programming-based approaches and regression-based approaches can model these issues and the difficulties and problems in doing so. There is an important distinction between the two broad approaches. In a regression-based methodology, the key to developing a deeper understanding of the components of TFP is by addition of further explanatory variables. In the mathematical programming methodology, the key is to develop additional constraints on the optimisation problem which is at the heart of computing the efficiency scores. Since the programming-based approach can be expressed in either primal envelopment form or dual multiplier form, adding (row) constraints to the primal involves adding (column) variables to the dual.

Following these methodological sections, the chapter turns to empirical issues, and for this we deliberately use the context of the regression-based approach, because the discussion of critical issues is, in our view, more transparent in this context than if we were to use the programming-based approach. We emphasise however that the concept of output or outputs used can include any of the performance metrics raised earlier, not only GDP. The first empirical problem we consider arises from the fact that interest in the performance of national economies is inseparable from the comparison of different national performances and this requires us to address the problem of latent heterogeneity in cross-country samples, i.e. *differences across countries.* In discussing these we review the issue of whether the performance of national economies converges over time, or whether, as suggested by endogenous growth models, the individual performance of different countries is endogenous to the country itself.

We identify a second empirical problem in the way that technological change is modelled in efficiency and productivity analysis. The majority of studies in both the regression-based approach and the programming-based approach treat technological change as a shift over time in the complete technology frontier. This shift may be Hicks-neutral or input increasing/decreasing but it assumes that all production techniques benefit simultaneously from technological change. However, there is an important literature which has a long history emphasising the idea of localised technical change in which innovation and progress applies to one or two production techniques but does not shift the whole frontier. We show that there are modelling problems for efficiency and productivity analysis in this idea but that programming-based methodology or other non-parametric approaches may offer a more fruitful starting point than conventional regression-based analysis. Finally, a third estimation issue that we identify refers to similarities amongst neighbouring countries rather than the differences between them that were discussed previously. This compels us to incorporate developments in spatial analysis into our review of the performance of national economies, and we do this in a particular example of the technological spillovers amongst neighbouring countries at the level of the aggregate production function. The issue here is how to meld together the spatial models with the standard error term specifications, and we review some very recent contributions to this problem.

This topic of the performance of national economies is very broad indeed. We can see that there is an implicit dilemma in the topic: is it macroeconomics or is it microeconomics? Certainly, in terms of plain numbers, the volume of macroeconomics treatments of national economic performance outstrips the volume of microeconomics treatments, but efficiency and productivity analysis is essentially embedded in microeconomics. There are different ways of addressing this dilemma, but we should explain ours clearly since it will not be in agreement with some approaches that other researchers may favour. We emphasise microeconomic developments particularly in regression-based and programming-based methodologies. However, we cannot pretend that the vast macroeconomics literature on the performance of national economies does not exist or is not relevant. Therefore, we set the scene by first reviewing the key ideas from the macroeconomic literature on national economic performance so that we bring out the four critical assumptions that underlie the conventional neoclassical measures of performance: *allocative efficiency, constant returns to scale, no exogenous variable shifts and no inefficiency of performance*. This enables us to motivate the microeconomic approach embedded in efficiency and productivity analysis because each of these assumptions is relaxed by the microeconomic approach to measuring national economic performance.

Our purpose in this chapter is not to present a detailed literature survey of the vast amount of research papers on the performance of national economies[5] – that would be an impossible task today. Instead we wish to present an analytical overview of how efficiency and productivity analysis can provide the appropriate tools for assessing national economic performance. This will therefore be an introduction to the more detailed range of topics developing this issue in the following chapters.

[5] One of the co-authors has already written a detailed survey of different stochastic frontier analysis models and specifications, Sickles et al (2017).

2 TFP: unobserved data residual representing technological progress

To most macroeconomists, TFP simply means the unobserved residual in aggregate data on the gross value-added measure of GDP when account is taken of the payments to inputs or factors of production. This is identified with technological progress, the key factor in raising per capita living standards over time. This measure which is known as growth accounting has been standard since the classic papers of Solow (1957) which stated that most of the growth in per capita GDP in the USA over the first half of the 20th century was not due to factor accumulation but was due instead to the unobserved residual which he named as technological progress, and Jorgenson and Griliches (1967) which demurred from this conclusion. In the macroeconomics literature, this measure is arrived at by the following calculation[6], see for example Goodridge, Haskel and Wallis (2016). Suppose that for sector or industry $j = 1 \ldots J$, aggregate labour used and aggregate capital used, L_j and K_j, produce gross value added, V_j. Then the relative change in aggregate real value added is

$$\Delta \ln V \equiv \sum_{j=1}^{J} w_j \Delta \ln V_j = \sum_{j=1}^{J} w_j v_j \Delta \ln L_j + \sum_{j=1}^{J} w_j v_j \Delta \ln K_j + \sum_{j=1}^{J} w_j \Delta \ln TFP_j \quad [1]$$

The weights v_j and w_j are respectively nominal value added in industry $j = 1 \ldots J$ as a share of aggregate value added and shares of factor cost in nominal industry value added. TFP is $\sum_{j=1}^{J} w_j \Delta \ln TFP_j$, i.e. the data residual required to ensure that the right-hand side aggregates sum to the left-hand side. This is what macroeconomists identify as technological progress. In ten Raa and Mohnen (2002) there is suggested a neat way of overcoming the problem in this growth accounting literature of using input prices as exogenous components in the weights to

[6] The chapter by Fox and Diewert elsewhere in this volume addresses this issue in much more detail

measure TFP when the input prices are themselves endogenous to the performance of the economy itself. ten Raa and Mohnen (2002) use shadow prices measured from an optimizing model of national economic performance in which the objective is to maximize the level of final demand given the endowments and technology of the economy represented by its input-output social accounting matrix.

From these data, national economic performance is often defined in terms of output per worker or output per hour worked: $\Delta \ln(V/N)$ or $\Delta \ln(V/H)$ using measures of the workforce, N, or hours worked, H. The apparent downwards trend in the major developed economies, USA, Japan and the European economies including the UK, in recent years is what constitutes the productivity slowdown. There are two types of explanation of national economic performance using this approach. The first is a careful deconstruction and refinement of the labour and capital data to ensure the minimum role for TFP. The second is a range of speculations on the variability of TFP measured in this way.

However, it is important to understand that very strong implicit assumptions about the structure of the aggregate economy are needed in order to use the growth accounting approach outlined above. These include the assumptions:

- that inputs are paid the value of their marginal products and output is priced at the marginal benefit of consumption, i.e. that there is *allocative efficiency* in all markets
- that there are *constant returns to scale* in every industry
- that no producers display inefficiency of performance due for example to agency problems or behavioural patterns different from those of rational economic agents, i.e. *every producing unit is on its production frontier*
- that *ceteris paribus* prevails i.e. there are *no important exogenous variable changes* or changes in the market structure or regulations of the economy under study.

Goodridge, Haskel and Wallis (2016) are careful to comment that they estimate TFP as a residual, but, they ask, what drives TFP? In theory, they say, it is technical progress, but it could also be: '*increasing returns to scale, omitted inputs, factor utilisation and cyclical effects, measurement error and a host of other factors*'. In other words, all of the factors which the neoclassical approach by necessity assumes to be absent. As we shall see, the approach of efficiency and productivity analysis is to focus on these other factors. In this way, the efficiency and productivity analysis methods reviewed here offer a much more flexible and open way of testing large theories of the nature of economic performance. There is no shortage of such theories, e.g. the encyclopaedic summary of growth under good and bad capitalism outlined by Baumol et al (2007). These authors offer, like others, a wide range of suggestions for enhancing economic growth which are testable using the methods of efficiency and productivity analysis but which are difficult to assess when the neo-classical assumptions of the growth accounting approach are used. Therefore, relaxing these assumptions becomes the key to understanding how national economic performance can be compared.

Before we do this, we briefly examine the two other explanations: deconstruction of the input data and speculation about the socio-economic determinants of TFP treated as technological progress alone.

A widely cited example of the input data deconstruction approach is Gordon (2003). Suppose that we examine the measure of the rate of change of real gross value added, GDP, $\Delta \ln(V)$. We might believe that a useful decomposition is[7]

$$V \equiv (V/H) \times (H/E) \times (E/N) \times (N/POP) \times POP$$

[7] Using an expanded identity as an analytical starting point is a popular technique for developing a new direction in research, but sooner or later it has to be supported by empirical evidence for testing theories about human behaviour.

[2]

Here:

(V/H) is gross value-added per hour worked in the sector under study

(H/E) is aggregate hours worked per employee

(E/N) is the employment rate – current employees as a share of the labour force

(N/POP) is the labour force participation rate – those in the labour force as a proportion of the relevant population

Only the population is regarded as a non-cyclical variable, the other ratios may all be cyclical. In Gordon (2003) the underlying trends in these ratios are identified using Hodrick-Prescott and Kalman filter time-series methods which then permit the development of socio-economic analyses of why the trends may be pointing in a particular direction. There are multiple versions of these speculative analyses in the literature. For example, with reference to the USA and other advanced economies, Baker, DeLong and Krugman (2005) highlight demographic and population issues suggesting that populations are aging and there are limited further reservoirs of female participation in employment because feminism is in a mature stage. To the extent that the productivity slowdown or weaker national performance is technical progress, Gordon (2016) is amongst the most prominent advocates of the argument that it has slowed because the modern age has run out of ideas. There has been a temporary boost to economic performance from the ICT based digital revolution including the smart-phone but this is ending and Gordon makes the bold claim that these innovations of the 21st century are as nothing compared with the great inventions of the previous 150 years: steam power, railways, natural gas pipelines, the internal combustion engine, electrical power generation and the jet engine. Why is technological progress slowing down? Gordon's explanation is that advanced economies are running into what he terms 'headwinds. These include demographics associated with the retirement of aging baby-boomers leading to lower labour force participation rates. Additionally, there is an education headwind because, he argues, there is no further room for greater high-school, i.e. secondary education, completion rates. Gordon adds that inequality is worsening as the top 1 percent stretch away from the rest and that this reduces incentives to raise productivity generally. There are echoes of these arguments in the revival of early Keynesian ideas about secular stagnation, and related research on the long-term trend towards a falling real return on capital and consequent disincentives to invest, e.g. Lukasz and Smith (2015) who characterise the global economy as experiencing higher saving rates due to aging populations and growing inequality, and lower returns due to falling public investment

We might expect that there should be an important role for the shift to the digital-knowledge economy in this type of analysis, and one approach focused largely on this is the shift towards "capitalism without capital" suggested by Haskel and Westlake (2017). The starting point is their observation that investment in tangible fixed assets is becoming much less important in developed economies than what they refer to as "intangible investment" which comprises investment in design, branding, R&D, data and software. They quote the example of Microsoft which in 2006 had recorded assets that amounted to about 30 percent of its then market value, but 85 percent of these assets consisted of cash while conventional plant and equipment accounted for only 3 percent of the assets and 1 percent of the market value. They cite Microsoft as one of the first examples of capitalism without capital. Haskel and Westlake (2017) use as a critical indicator the ratio of the value of the tangible assets on a firm's balance sheet to the market value of a firm. They show that for the world's five most valuable companies, this ratio is currently (2017) below five percent; they comment that that although these include the global 'tech' companies this phenomenon is spreading to every sector.

They argue that this makes the modern intangible-rich economy fundamentally different from one based on tangibles. Several problems arise for the measurement of performance as a result of this development. Investment in intangibles is difficult to measure in national statistics and often R&D is simply recorded as a cost rather than a form of investment. Haskel and Westlake argue that intangible investment such as a brand, or an algorithm can be scaled up much more easily than tangible investment through the transfer of software. In addition, intangible investment has spillovers making it more difficult to stay ahead of the competition but also driving a wedge between the private and social rates of return on this form of investment. Issues such as these suggest that the way that national economic performance can be measured is likely to change radically in the future compared with the way it has been carried out up to now.

There are therefore numerous analyses in the literature that allow economists to speculate in general socio-economic terms about perceived facts of modern society, but all of them suffer from a departure from well-formulated empirical analysis and the imposition of strong assumptions about markets and behaviour and that is a gap that efficiency and productivity analysis research tries to fill. Before considering the efficiency and productivity analysis in more detail however, we must first ask whether the gross-value added measure of GDP is adequate for addressing national economic performance.

3 Is GDP the right way to measure national performance?

There are two questions to ask in this context.

Is GDP measured properly? GDP in the national accounts is gross value-added and it equals not only spending on final goods and services but also factor incomes.

Is GDP the appropriate output variable to measure?

There is considerable literature on each of these issues for which we provide a brief introduction.

There is a widely perceived idea that measured GDP excludes many important areas of economic activity, particularly in relation to mispriced goods and services. The proper definition of GDP has been a subject of debate since the development of national accounts, which were an outcome of the problems that Keynes and his followers in the USA and the UK encountered in trying to measure the level of economic activity before and during the second world war – for a lively account of the early Keynesian efforts to define and understand GDP see Skidelsky (2003). A key issue is the definition of the production boundary, Coyle (2014, 2017). The production boundary[8] separates 'paid-for activities in the market economy from unpaid activities' so that firms and government are considered productive but households are not, (Coyle (2017). As a consequence, much of the work done largely by women in the home is not generally included in GDP – see Folbre and Nelson (2000), On the other hand, it is plain that the options for female participation in the labour force differ widely because child care provisions vary so markedly across even the developed economies in the OECD and EU, as shown by Bettio and Plantenga (2004). Most leisure

[8] The idea of the production boundary separating productive and non-productive services goes back to Adam Smith (1776) where Smith famously distinguished the output of productive labour from that of non-productive labour whose 'services generally perish in the very instant of their performance, and seldom leave any trace or value behind'. Many professions fell into this category, according to Smith, including the menial servant, the Sovereign, men of letters of all kinds, buffoons and opera-singers.

activity is excluded as well, and the digital economy is said to be having a massive but unmeasured effect on economic activity, Varian (2016). Once we start to unravel the definition of GDP the problems of using it to measure the economic performance of national economies seem to multiply exponentially. For example, there are massive policy changes under debate and in progress to combat climate change. The achievement of a viable carbon-neutral economy is the objective of many in the environmental movement, but without a clear consensus on the social cost of carbon, we have no way of measuring the benefits in GDP terms of the success or otherwise of environmental policy.

Consequently, economists have for decades argued that GDP is an inadequate measure of the economic performance of nations and have sought to develop alternative measures of national economic welfare. There have been many suggestions for a welfare or even a 'happiness'-based index instead, see Helliwell, Layard and Sachs (2012). Particularly important have been suggestions by international bodies like the UN which has developed its own human development index, HDI that includes measures of education and health. Many of the suggested substitute measures such as 'happiness' are based on survey responses, and Helliwell, Layard and Sachs (2012) is the most widely cited of these. The initial observation that commenced this line of research is the Easterlin (1974) paradox that states that at any point in time richer people appear to be happier than poorer people but over time society does not appear to become happier as it becomes richer. Easterlin's explanation is that individuals use relative income levels to evaluate their well-being but if these stay constant over time happiness is unchanged. The contention of Helliwell, Layard and Sachs is that happiness differs over time and across societies for identifiable reasons and it may be alterable by public policy. Their 2012 report used the Gallup World Poll, the World Values Survey, the European Values Survey and the European Social Survey from 2005 to 2011 to compile a broad happiness index. For example, the Gallup World Poll asked 1000 people aged 15 or over in 150 countries to evaluate the quality of their lives on an ascending score from 0 to 10. For the world as a whole weighted by population, the modal category, i.e. the category with the largest number of people reporting (26.2%), was 5, exactly the mid-point. In categories 6 to 10, there was a further 42.9% of respondents so that 69.1% of the total reported that they were not below the mid-point of the happiness scale. For North America, Australia and New Zealand, the modal life-satisfaction category was 8 with 92.9% of respondents reporting that their life satisfaction was not below the mid-point (5). By contrast in Sub-Saharan Africa only 47.4% of respondents were not below mid-point category 5. Clearly level of development with all of its associated implications plays a major role in the relative evaluation of happiness. In their analysis of responses, Helliwell Layard and Sachs identified key categories affecting life evaluations as: *work* (employment and quality); *social capital* (trust, freedom, equality); *values* (altruism, materialism, environment); *health* (mental, physical); *family*; *education level*; *gender.* For example, improvements in the nature of work or the support for social capital and health were evaluated as being worth several multiples of a 30 % increase in income. Based on results like these, Helliwell Layard and Sachs noted that changes in these factors can be brought about by policy reform, offering considerable scope for rich and deep analysis on the relative performance of national economies.

However, there has also been a consistent strand of economic research that attempts to measure economic welfare amongst nations empirically rather than subjectively. In Jones and Klenow (2016) for example, there is a detailed empirical study that compares the performance of a wide range of countries on a measure of economic welfare determined by an equivalent consumption metric. Their aim is to determine how an easily computable measure of economic welfare correlates with GDP as a measure of economic performance. It is interesting to examine this example of much of the recent work on the usefulness of GDP as a measure of national economic performance. Jones and Klenow imagine an individual living in an arbitrarily chosen country and drawing his/her life experiences from that country's distributions of *consumption, work-leisure trade-off opportunities, inequality and life expectancy*. Using simple logarithmic assumptions about preferences, they construct from observed macro and micro data a measure of utility for that individual in that country. They then construct a variable: the 'consumption equivalent measure of standard of living' which is the factor *lambda* which if applied to the random draws of consumption, leisure and life-expectancy from the distributions applying in the USA would make that individual indifferent between living in the USA and his/her original country. The factor *lambda* is the number which multiplicatively reduces the level of consumption of a US citizen sufficiently to provide a level of utility equivalent to the citizen of another country when utility depends on consumption, inequality and leisure, and when consumption in each country is a randomly distributed variable with a mean and variance particular to the country in question. In other words, the proportion of USA consumption – given the leisure, mortality and inequality in the USA – which would provide the same expected utility as the values elsewhere.

Jones and Klenow provide a simple example[9]. They postulate an intercept level of utility, $\bar{u}$, e.g. the lifetime subsistence level of consumption or value of life, and concentrate on two key variables: the first is consumption per capita, C, which is the individual's random draw from the consumption distribution for the country in which he/she lives, and the second is the utility of leisure time, $v(l)$, drawn from the leisure distribution in the country. The flow of utility is

$$u(C, l) = \bar{u} + \log C + v(l) \qquad [3]$$

[9] Jones and Klenow present a complex analysis of which this is the simplest example assuming a zero discount rate for utility of consumption and a zero growth rate for consumption.

They assume that consumption is log-normally distributed, a result often found to describe all but the top percentile of the income distribution in many countries:

$$\log C_i \sim N(\mu_i, \sigma_i^2)$$

[4]

where $\mu = E(\log C)$ and $\sigma^2 = var(\log C)$.

Jones and Klenow parameterise the mean of consumption in country i as $E(C_i) = c_i$, then, using the properties of the log-normal distribution, they are able to write[10]:

$$E(C) = exp\left(\mu + \frac{1}{2}\sigma^2\right) = c \Rightarrow \log E(C) = \mu + \frac{1}{2}\sigma^2 = \log c$$

[5]

i.e. after rearranging equation [5] and using [4]:

$$E(\log C) = \mu = \log E(C) - \frac{1}{2}\sigma^2 = \log c - \frac{1}{2}\sigma^2$$

[6]

Assuming that typical life expectancy in any year for a citizen in this country i is e_i, Jones and Klenow then write, in their simplest case, that a citizen's expected lifetime utility is the product of the flow of utility multiplied by life expectancy:

$$U_i = e_i\left(\bar{u} + \log c_i + v(l_i) - \frac{1}{2}\sigma_i^2\right)$$

[7]

This tells us that welfare of the typical citizen in this country is increasing in life expectancy, increasing in consumption per person, increasing in the utility of leisure available per person

[10] The derivation of these results is compressed in Jones and Klenow (2016) so we have expanded the explanation.

but decreasing in the variance of consumption per person, which is a measure of the inequality of the distribution of consumption per person.

Now for the case $i = USA$ multiply c_{USA} by λ_i, the multiplier by which U_{USA} must be reduced to yield the level of welfare that is equivalent to that of a citizen living in country i.

$$U_{USA}(\lambda_i) = U_i(1)$$

[8]

In other words, find λ_i that satisfies

$$e_{USA}\left(\bar{u} + \log(\lambda_i c_{USA}) + v(l_{USA}) - \tfrac{1}{2}\sigma_{USA}^2\right) = e_i\left(\bar{u} + \log c_i + v(l_i) - \tfrac{1}{2}\sigma_i^2\right)$$

[9]

The result is

$$\log \lambda_i = (e_i - e_{USA}/e_{USA})\left(\bar{u} + \log c_i + v(l_i) - \tfrac{1}{2}\sigma_i^2\right) + [\log c_i - \log c_{USA}]$$
$$+ [v(l_i) - v(l_{USA})] - \tfrac{1}{2}(\sigma_i^2 - \sigma_{USA}^2)$$

[10]

This 'consumption equivalent measure of standard of living' therefore consists of four additive terms for each country:

- Relative life expectancy in country i compared with USA weighted by the mean flow of utility of consumption and leisure in country i
- Relative mean consumption compared with USA
- Relative utility of leisure time compared with the USA
- Relative variance of consumption compared with the USA, which is a measure of consumption inequality.

In general, $\log \lambda_i$ will be negative so that $\lambda_i < 1$ due to the dominance of the second term, the gap between the county's per capita consumption and that of the USA. However, as the other terms have an impact, for some countries $\log \lambda_i$ will approach zero so that $\lambda_i \cong 1$. In other words, while many countries will have consumption per capita much lower than in the USA some may have higher leisure time, higher life expectancy and a more equitable distribution of income all of which contribute positively to the citizen's welfare in the Jones-Klenow social welfare function. Therefore, we may expect that compared to the ranking by GDP or consumption per capita, the consumption equivalent welfare measure may show that some countries rank equally highly with the USA in terms of national welfare performance but that others may be much more worse off than the raw GDP data indicate.

Jones and Klenow draw on the research literature to parameterise these components in particular using a value for the Frisch elasticity of labour supply of 1 which implies disutility from working rises with the square of the number of hours worked so that $v(l)$ then depends on the real wage and the marginal tax rate of labour income. Constructing λ_i for a wide range of countries provides a set of important results for the evaluation of GDP as a measure of comparative national performance compared with the consumption equivalent measure of welfare.

- The correlation between GDP per capita and consumption equivalent welfare is very high, of the order of 0.95-0.98.
- In western Europe, living standards are much closer to USA than income per capita suggests due to longer lives with more leisure
- In most developing countries welfare is much lower than income per capita due mainly to shorter lives with more inequality
- Economic growth in consumption equivalent welfare (except in sub-Saharan Africa) is 50 per cent higher than growth in GDP per capita due to declining mortality.

Jones and Klenow are conscious of leaving out other aspects of welfare in which they include morbidity, environmental quality, crime, political freedom and intergenerational altruism, nevertheless this example of a growing literature indicates how the measurement of the performance of national economies opens up a massive range of modelling developments. It is possible for example to consider the consumption equivalent measure of standard of living as providing an alternative conception of the frontier of national economic performance to which further efficiency and productivity analysis could then be applied. In general, efficiency and productivity analysis of the 'happiness frontier' or welfare frontier is a largely unexplored area.

4 National economic performance: programming analysis

From this point on we take up the second part of the chapter and focus on using real value-added GDP as the key measure of performance of national economies so that the estimation of TFP is the central preoccupation of the analysis. In this section we concentrate on the programming approach to measuring TFP, generally known by the generic name of data envelopment analysis.

The paper by Farrell (1957) and the comments by Winsten (1957) contributed hugely to the development of efficiency and productivity analysis and it is interesting that the initial example related to efficiency in aggregate agricultural production of the USA. However, it can be said that the pioneering paper in the application of efficiency and productivity analysis to measuring the performance of national economies is Färe, Grosskopf, Norris and Zhang (1994). This paper used data envelopment analysis to evaluate productivity change across different countries and introduced two major changes to the assumptions required by the neoclassical growth accounting method. Constant returns to scale was replaced by the possibility of variable returns to scale and the assumption that every country was on the international production frontier was replaced by the possibility that countries could display inefficiency of performance. In this way efficiency and productivity analysis moved on from the conventional neoclassical macroeconomic approach to measuring TFP. Färe, Grosskopf, Norris and Zhang used data envelopment analysis with variable returns to scale to develop Malmquist indices of TFP. Subsequently Ray and Desli (1997) refined the analysis on how the effect of variable returns to scale should be measured. In ten Raa and Shestalova (2011) the Solow residual concept is neatly reconciled with the data envelopment analysis approach to productivity measurement by embedding it in an input-output analysis. This approach which makes uses of duality and shadow prices offers a potentially interesting way to conceptualise the theoretical measurement of TFP.

The distance function contains the same information about the technology as the production function. Consequently Caves, Christensen and Diewert (1982) by adopting a normative approach rather than the neoclassical data residual approach show that productivity growth can be represented by a Malmquist index defined to be the ratio of the values of an output distance function after an event to the value of an output distance function before the event. The output distance function directly measures Farrell radial efficiency. The resulting index $M > 1$ if there is positive productivity growth. Färe et al (1994) developed this in several ways. First, by assuming that the producing unit need not be on the transformation surface either before or after the event. In other words, the producing unit could be technically inefficient despite the existence of productivity growth. The possibility of building in a measure of inefficiency allows the researcher to decompose the shift in the producing unit's position into two separate components: the efficiency change effect and the technical change effect. Secondly, Färe et al derived the equivalent Malmquist indices for the input distance function which measures the inverse of the Farrell radial efficiency, so that to maintain the convention that the resulting index $M > 1$ if there is positive productivity growth the inverses of the input distance functions, i.e. the Farrell radial technical efficiency scores were used. Finally, Färe et al addressed the issue of returns to scale by defining the Malmquist index for distance function values assuming constant returns and a separate Malmquist index for distance functions assuming variable returns to scale. The difference between the two technologies is defined by the description of the technology as a convex cone in the case of constant returns and a convex hull in the case of variable returns.

We isolate two periods for comparison: t and $t + 1$, representing the before and after positions relative to a productivity change. We need to compare the value of the distance function at $t + 1$ to its value at t, but there is the option of choosing the period t or the period $t + 1$ output possibility set as the reference technology. For example, Fare et al (1994) use the geometric mean of these two reference sets as the reference technology. We observe the inputs and outputs at each of these periods and set up the corresponding programming analysis with input-orientated radial efficiency measures (θ). Non-radial measures of efficiency can also be incorporated into developments of the distance function approach.

The output distance function is defined for a given technology such that a vector of inputs $\mathbf{x}' = (x_1 \dots x_K)$ can make a vector of outputs $\mathbf{y}' = (y_1 \dots y_R)$. The technology set is

$$T = \{\mathbf{x}, \mathbf{y}: \mathbf{x}\ can\ make\ \mathbf{y}\}$$

[11]

The output distance function is the smallest positive scalar divisor δ of a bundle of the production unit's outputs $\mathbf{y}$ such that $(\boldsymbol{y}/\delta)$ is in the technology set, T.

$$D_O(\mathbf{x}, \mathbf{y}) = min\{\delta: (\mathbf{x}, \boldsymbol{y}/\delta) \in T\} \leq 1$$

[12]

A piecewise linear representation of the technology of production with constant returns to scale is.

$$T^r = \{\mathbf{x}, \mathbf{y}: \mathbf{X\lambda} \leq \mathbf{x}, \mathbf{Y\lambda} \geq \mathbf{y}\}$$

[13]

The matrices $\mathbf{X}, \mathbf{Y}$ represent all of the observed data in the efficiency measurement exercise and the vectors $\mathbf{x}, \mathbf{y}$ represent one particular country. The output distance function can be measured by the Farrell radial efficiency, δ, of each country's outputs:

$$D_O(\mathbf{x}, \mathbf{y}) = min\{\delta: (\mathbf{x}, \boldsymbol{y}/\delta) \in T^r\} = min\{\delta: \mathbf{X\lambda} \leq \mathbf{x}, \mathbf{Y\lambda} \geq \boldsymbol{y}/\delta\}$$

[14]

There is an equivalent approach to the input distance function defined as the largest scalar divisor ρ of a bundle of inputs $\mathbf{x}$ such that $(\mathbf{x}/\rho)$ is still in the technology set which leads to a Farrell radial efficiency measure, θ, of the inverse input distance function:

$$D_I(\mathbf{x}, \mathbf{y}) = max\{\rho: (\mathbf{x}/\rho\,, \mathbf{y}) \in T^r\} = \; min\{\theta: \mathbf{X\lambda} \leq \theta\mathbf{x}, \mathbf{Y\lambda} \geq \boldsymbol{y}\}$$

[15]

In the both cases of the output and input distance function the assumption of variable rather than constant returns to scale is implemented by adding the constraint $\mathbf{e}'\boldsymbol{\lambda} = 1$ to the piecewise linear representation of the technology, where **e** is a vector of ones.

We use the notation $\theta_C^{t+1,t}$ to represent the input orientated measure of radial efficiency with constant returns to scale for a country observed in period t relative to the technology prevailing in period $t+1$ while $\theta_V^{t+1,t}$ refers to the variable returns to scale version.

Scale efficiencies are given by the ratio of the Farrell radial efficiency under CRS to the Farrell radial efficiency under VRS. There are four measures: respectively scale efficiency for the observation in period t with reference to the period t technology, scale efficiency for the observation in period t with reference to the period $t+1$ technology, scale efficiency for the observation in period $t+1$ with reference to the period $t+1$ technology, and scale efficiency for the observation in period $t+1$ with reference to the period t technology:

$$\sigma^t\left(\mathbf{x}_0^t, \mathbf{y}_0^t\right) = \frac{\theta_C^{\;t,t}}{\theta_V^{\;t,t}} \qquad \sigma^{t+1}\left(\mathbf{x}_0^t, \mathbf{y}_0^t\right) = \frac{\theta_C^{\;t+1,t}}{\theta_V^{\;t+1,t}} \quad \sigma^{t+1}\left(\mathbf{x}_0^{t+1}, \mathbf{y}_0^{t+1}\right) = \frac{\theta_C^{\;t+1,t+1}}{\theta_V^{\;t+1,t+1}} \quad \sigma^t\left(\mathbf{x}_0^{t+1}, \mathbf{y}_0^{t+1}\right) = \frac{\theta_C^{\;t,t+1}}{\theta_V^{\;t,t+1}}$$

[16]

This produces a scale decomposition:

$$SEC = \left[\frac{\sigma^{t}\left(\mathbf{x}_0^{t+1}, \mathbf{y}_0^{t+1}\right)}{\sigma^{t}\left(\mathbf{x}_0^{t}, \mathbf{y}_0^{t}\right)} \frac{\sigma^{t+1}\left(\mathbf{x}_0^{t+1}, \mathbf{y}_0^{t+1}\right)}{\sigma^{t+1}\left(\mathbf{x}_0^{t}, \mathbf{y}_0^{t}\right)} \right]^{0.5}$$

[17]

Malmquist indices can then be defined for CRS or VRS technology

$$M_{IC}\left(x_0^t, y_0^t, x_0^{t+1}, y_0^{t+1}\right) = \left[\frac{\theta_C^{\ t+1,t+1}}{\theta_C^{\ t,t}} \right] \left[\frac{\theta_C^{\ t,t}}{\theta_C^{\ t+1,t}} \frac{\theta_C^{\ t,t+1}}{\theta_C^{\ t+1,t+1}} \right]^{0.5}$$

[18]

$$M_{IV}(x_0^t, y_0^t, x_0^{t+1}, y_0^{t+1}) = \left[\frac{\theta_V^{\ t+1,t+1}}{\theta_V^{\ t,t}} \right] \left[\frac{\theta_V^{\ t,t}}{\theta_V^{\ t+1,t}} \frac{\theta_V^{\ t,t+1}}{\theta_V^{\ t+1,t+1}} \right]^{0.5}$$

[19]

Each index can be decomposed into a measure of efficiency change, EFC, the first ratio in square brackets, and technical change, TEC, the second ratio in square brackets.

Finally, we have a relationship between the indices:

$$M_{IC}\left(x_0^t, y_0^t, x_0^{t+1}, y_0^{t+1}\right) = M_{IV}\left(x_0^t, y_0^t, x_0^{t+1}, y_0^{t+1}\right) \times SEC$$

[20]

This provides us with a complete decomposition into efficiency change EFC, technical change TC and scale efficiency change, SEC:

$$M = EFC \times TC \times SEC$$

[21]

Fare et al (1994) and Ray and Desli (1997) applied the analysis to an international sample using data on real GDP, labour and capital inputs from the Penn World Tables, Summers and

Heston (1991) and this procedure has become commonplace in international productivity comparisons using the efficiency and productivity analysis approach. In summary, this early literature was able to introduce two ways in which the neoclassical assumptions could be relaxed by employing data envelopment analysis. The data envelopment analysis assumes that countries can be below the international production frontier and that they can operate with technologies that display variable returns to scale. The procedure for doing this is straightforward once we abandon the neoclassical approach and specify the ideas associated with the distance function and Farrell radial efficiency.

There are many ways in which this initial work on international TFP comparisons used efficiency and productivity analysis. For example Milner and Weyman-Jones (2003) in a study confined to developing nations also drew on the Penn World Tables to measure the radial efficiency component of different countries and to relate that to different measures of country heterogeneity, thereby combining the neoclassical approach that focused on determinants of differences in national GDP performance and the efficiency and productivity analysis approach which modelled the technology as a performance measure that varied with returns to scale. In their study of developing nations, Milner and Weyman-Jones (2003) looked for possible determinants of the measured Farrell radial efficiency scores of different countries that used inputs of labour, capital and agricultural land to generate real GDP. The explanatory factors they used in a 2-stage analysis included relative country size, per capita income, education level, health level, industrialization, degree of democracy, trade openness. There have been many further advances in the data envelopment analysis approach to international productivity comparisons. Giraleas et al (2012) demonstrated that the data envelopment analysis approach using Malmquist indices performed particularly well in simulation studies when compared against the neoclassical growth accounting and deterministic regression-based frontiers. Since the development of the programming approach by Färe, Grosskopf, Norris and Zhang there has been a massive expansion in the number of data envelopment analysis studies of the Malmquist estimation of TFP for a multiplicity of economies.

5 National economic performance: regression-based analysis

In assessing the performance of national economies from the point of view of efficiency and productivity analysis we have the choice of measuring either efficiency levels or productivity changes across space and time. The major part of the literature concentrates on productivity comparisons and changes because simply focusing on the measured distance to a frontier does not bring out the major factors that could be important in decomposing the changes in productivity over time. In section 4, we showed that by using data envelopment analysis to construct normative Malmquist indices of TFP, we are able to relax two critical assumptions of the conventional neoclassical approach: the assumption of constant returns to scale and the assumption that every country is on the international production frontier. The other two assumptions of the neoclassical approach can also be investigated. These are that allocative efficiency prevails and that *ceteris paribus* is invoked – i.e. there are no exogenous variable shifts to take into consideration. The neoclassical approach gets around the second of these two requirements by *ad hoc* qualitative speculation about long term socio-economic trends including demographics supported by detailed deconstruction of the data on inputs and GDP used in the traditional index number approach. We have now seen how efficiency and productivity analysis in the form of data envelopment analysis can address issues that the neoclassical approach cannot. It is natural to ask whether data envelopment analysis could also contribute to the relaxation of the allocative efficiency assumption and the *ceteris paribus* assumption. It is certainly the case that a vast amount of useful data envelopment analysis has addressed these issues as well. Allocative efficiency has been researched in the data envelopment analysis approach since the original Farrell contribution and data envelopment analysis models can be developed using input price data to capture allocative inefficiency, see e.g. Bogetoft and Otto (2011). It is also true that data envelopment analysis can be redesigned to accommodate additional shift factors representing the role of exogenous variables – this is done by adding constraint rows to the primal envelopment DEA programmes or equivalently adding variable columns to the dual DEA multiplier problems. In addition, some progress towards the inclusion of idiosyncratic error in the form of sampling error can also be made using bootstrapping approaches, Simar and Wilson (2007). However, these issues can also be addressed using stochastic frontier analysis and arguably in the context of comparing the performance of national economies rather than individual decision-making organizations, the stochastic frontier analysis approach offers clearer and more direct methods of analysis. Consequently, in continuing our discussion of the relaxation of the conventional neoclassical approach to TFP we turn to stochastic frontier analysis.

There are different ways of deriving the full TFP formulae for different representations of the technology. One procedure as we saw is to start from the generalised Malmquist index form shown in equations [20] and [21] in which the Malmquist index of distance function values, which can be decomposed into a technological shift and a frontier catch-up, is adjusted by a scale factor to take account of non-constant returns to scale. Orea (2002) showed how this can be developed in a stochastic frontier analysis framework to generate a Tornqvist index of total factor productivity change, TFP, by using the empirically estimated elasticities and the key idea of the quadratic identity lemma due to Diewert (1976). Coelli et al (2003) apply this to the production function, input distance function and multi-product cost function representations of the technology.

Another derivation starts from the basic properties of a TFP index. An index of TFP is the weighted growth rates of outputs minus the weighted growth rates of inputs. Two of the most important properties of the weights are monotonicity and proportionality. Monotonicity requires that the weighted output growth rates and input growth rates are chosen so that higher output and lower input unambiguously improve TFP. This requires that in an empirical application based on regression analysis the elasticities must all be adjusted to have non-negative signs. Proportionality requires that the weights in the output and input growth indices add to unity. We can apply these requirements to the empirical estimation of the single output production function, the multi-product input distance function, the multi-product output distance function and the multi-product cost function to derive TFP indices from each representation of the technology. We first differentiate our functional representation with respect to time to obtain the proportional rates of change of the outputs and the inputs then we choose a functional form to estimate by regression procedures so as to generate the required elasticity weights.

We apply this as follows in Table 1. The analysis here is based on the approach of Lovell (2003) who illustrated the output distance function. For each function: production, input distance, output distance and cost, we derive the elasticity weights from the logarithmic form of the function by total differentiation with respect to time. We use these elasticities as the output and input weights ensuring that they satisfy monotonicity and proportionality. In the case of the input distance function for example we must take care that the output elasticities which are negative when estimated are changed in sign to ensure monotonicity and in the case of the output distance function where the input elasticities are negative the same adjustment applies. The proportionality property is ensured by adjusting the elasticities by the elasticity of scale value for the function in question, respectively E, E^I, E^O, E^C for the production, input distance, output distance and cost functions. When $E, E^I, E^O, E^C > 1$ there are increasing returns or economies of scale, when $E, E^I, E^O, E^C = 1$ there are constant returns and when $E, E^I, E^O, E^C < 1$ there are decreasing returns or diseconomies of scale. These elasticity of scale formulae are derived for the distance functions in Färe and Primont (2012) and for the cost function in Panzar and Willig (1977). The resulting TFP index will therefore be decomposable into four components: scale efficiency change, SEC, the change due to exogenous variables, EXC, the technological change or frontier shift effect, TC, and the efficiency change component, EFC, derived from the inefficiency component of the error term in the empirical estimation form.

To construct the TFP measures shown in Table 1, we have identified for each function an elasticity weighted average of rates of output growth minus an elasticity weighted average of rates of input growth. Monotonicity requires that the TFP measure increases if outputs increase and decreases if inputs increase. This requirement determines the sign properties of the components. Proportionality requires that the different elasticity weights applying to outputs and to inputs each sum to one. This requirement is partly satisfied when the functional form representing the underlying technology has some form of linear homogeneity property, otherwise it must be satisfied by making a scale adjustment to the elasticity weighted rates of change of outputs or inputs. In stochastic frontier analysis, we find that different decompositions are available depending on the way in which we choose to model the technology and the behaviour of producers.

The simplest place to begin is the aggregate production function relating our preferred output measure, y, e.g. the real gross value added estimate of GDP to the economy's aggregated inputs: $x_1, \dots, x_K$, the exogenous variables held constant under the *ceteris paribus* assumption: $z_1, \dots, z_L$ and time representing the possibility of technological progress, t. Write the stochastic aggregate production function as

$$y = f(x_1, \dots, x_K, z_1, \dots, z_L, t)exp(v - u)$$

[22]

The error term has as usual two components: v is a two-sided symmetrically distributed zero-mean random variable representing idiosyncratic error which is usually assumed to encompass all the measurement error, sampling error and specification error in the data generating process; u is an asymmetrically distributed non-negative random variable with its distribution truncated at zero so that it has a positive expected value which is assumed to encompass the inefficiency of producer performance. In this way, we arrive at a measure of stochastic efficiency of performance

$$0 < TE \equiv e^{-u} = y/f(x_1, \dots, x_K, z_1, \dots, z_L, t)exp(v) \leq 1$$

[23]

To arrive at a productivity change measure, we must take the logarithmic derivative of the initial aggregate production function with respect to time. Write

$$\dot{y} \equiv \partial \ln y/\partial t = (\partial y/\partial t)(1/y)$$

[24]

We use the same convention for all of the other variables.

In table 1 we show in the first row the TFP decomposition for the aggregate production function representation of the technology. The term $E = \sum_k \varepsilon_{x_k}$ is the elasticity of scale (<1, = 1, >1, according as there are decreasing, constant and increasing returns to scale). The left-hand side of the TFP expression (in square brackets) is the growth rate of output minus the weighted growth rates of the inputs with the weights designed to sum to 1. Therefore, it is by definition a measure of total factor productivity change, TFP. The terms on the right-hand side represent first the growth rate of inputs adjusted by the elasticity of scale, which is scale efficiency change, *SEC*, second the weighted growth rates of the exogenous variables, *EXC*, third the growth of output over time when all other variables are held constant, i.e. technological progress or technical change, *TC*, and fourth the rate of decline of inefficiency over time, i.e. efficiency change, *EFC*. Each of these terms depends on knowledge of the relevant production function elasticities.

Consequently, the stochastic frontier analysis has permitted a generalisation of the previous expressions for TFP to give us:

$$TFP = SEC + EXC + TC + EFC$$

[25]

In this way, the stochastic frontier analysis has relaxed three of the four key assumptions of the conventional neoclassical growth accounting approach – we no longer assume constant returns to scale or that exogenous factors must be held constant, or that all producers are on their respective production frontiers. These factors can be added to technological progress as components of TFP so long as we are able to estimate them from the available data. We do this by econometric estimation of the functional form shown in the second column of Table 1, and we choose a functional form from which it is possible to extract the relevant elasticity information.

We can derive a productivity decomposition TFP for each of our functional forms using the log-derivatives as we did above, and the second, third and fourth rows of table 1 show the corresponding TFP decompositions[11]. Note that in each case, the left-hand side of the TFP expression contains the definition of TFP and it is required to be the difference between a weighted sum of the log-output changes minus a weighted sum of the log-input changes with the weights summing to 1 in each case. In the case of the input distance, or output distance this requires the use of a homogeneity property and a rescaling by the corresponding measure

[11] Key references on these ideas are Bauer (1990), Orea (2002) and Coelli, Estache, Trujillo and Perelman (2003)

of the elasticity of scale. For the input distance, the input elasticities must sum to 1 by the homogeneity property and it is the output elasticities which are rescaled on the left-hand-side. In the output distance, the output elasticities sum to 1 by homogeneity and the input elasticities are rescaled on the left-hand side.

None of these decompositions contains a measure of allocative inefficiency. To achieve this, we need to introduce input prices: $w_1, \dots, w_K$ and therefore move to a dual expenditure function or cost function for each industry or sector of gross value added separately. In the case of the dual expenditure function the cost-output elasticities on the left-hand side of the TFP decomposition are rescaled and the cost-input elasticities are weighted by their actual cost shares, s_k which must sum to 1.

Table 1 is completed by showing the TFP decomposition in this cost function case for industry j and this may be aggregated into an overall TFP decomposition using the industry weights as indicated in the conventional growth accounting approach. It is from this last row of Table 1 that we are able to incorporate an allowance for allocative efficiency change, AEC, which is measured by the log-input-price changes weighted by the difference between the actual cost shares and the optimal cost minimising cost shares, see Bauer (1990), and Orea (2002).

TABLE 1 HERE

The procedure for estimating the TFP decomposition then proceeds as follows. Since this form of research is often used to compare different national economic performances across countries and across time, we will illustrate with panel data.

- Select a function to represent the technology and the chosen TFP decomposition

- Select a functional form for the function
- Estimate the functional form using stochastic frontier analysis so that the efficiency change component, EFC is included. In the case of the input and output distance functions, the homogeneity property must be imposed to make estimation feasible by identifying the dependent variable. In the production function case, homogeneity is not assumed unless constant returns to scale is imposed. In the cost function homogeneity in input prices can be imposed or it may simply be tested on the estimated equation without homogeneity. In each case the relevant estimated elasticities from the homogeneous and non-homogeneous form are the same.
- The left-hand side of the TFP decomposition defines the measure of TFP and is not to be calculated, since it is by definition equal to the right-hand side. Instead, the elasticity estimates are used to calculate each of the right-hand side components.
- The elasticity estimates are used to multiply the log-variable changes to arrive at the corresponding TFP decomposition. However, all of the analysis so far has assumed continuous functions and the measures must be adapted to discrete data, e.g. annual changes.

Table 1 Functional forms and decompositions of TFP for different representations of the technology

Function for estimation	Form for estimation	Form for TFP components satisfying monotonicity and proportionality
Production: Elasticity of scale: $E = \sum_k \varepsilon_{x_k}$	$\ln y = \ln f(\mathbf{x}', \mathbf{z}', t) - u + v$ $\varepsilon_{x_k} = \partial \ln f / \partial \ln x_k \geq 0$	$\ln y = \ln f(\mathbf{x}', \mathbf{z}', t) - u, u \geq 0$ $[\dot{y} - (1/E)\boldsymbol{\varepsilon}_{\mathbf{x}}'\dot{\mathbf{x}}] = TFP = (E - 1/E)\boldsymbol{\varepsilon}_{\mathbf{x}}'\dot{\mathbf{x}} + \boldsymbol{\varepsilon}_{\mathbf{z}}'\dot{\mathbf{z}} + \varepsilon_t - (du/dt)$ Decomposition: $TFP = SEC + EXC + TC + EFC$
Input distance, $\sum_{\forall k} \varepsilon_{x_k} = 1$ Elasticity of scale: $E^I = \left(-\sum_r \varepsilon_{y_r}^I\right)^{-1}$	$-\ln x_K = \ln D_I\left((1/x_K)\mathbf{x}', \mathbf{y}', \mathbf{z}', t\right) - u + v$ $\varepsilon_{x_k}^I = \partial \ln D_I / \partial \ln(x_k / x_K) \geq 0$ $\varepsilon_{y_r}^I = \partial \ln D_I / \partial \ln y_r \leq 0$	$\ln D_I(\mathbf{x}, \mathbf{y}, \mathbf{z}, t) - u = 0, u \geq 0$ $[-E^I \boldsymbol{\varepsilon}_{\mathbf{y}}^{I\prime}\dot{\mathbf{y}} - \boldsymbol{\varepsilon}_{\mathbf{x}}^{I\prime}\dot{\mathbf{x}}] = TFP = (1 - E^I)\boldsymbol{\varepsilon}_{\mathbf{y}}^{I\prime}\dot{\mathbf{y}} + \boldsymbol{\varepsilon}_{\mathbf{z}}^{I\prime}\dot{\mathbf{z}} + \varepsilon_{\mathrm{t}}^I - du/dt$ Decomposition: $TFP = SEC + EXC + TC + EFC$
Output distance, $\sum_{\forall r} \varepsilon_{y_r} = 1$ Elasticity of scale: $E^O = \left(-\sum_k \varepsilon_{x_k}^O\right)$	$-\ln y_R = \ln D_O(\mathbf{x}', (1/y_R)\mathbf{y}', \mathbf{z}', t) + u + v$ $\varepsilon_{y_r}^O = \partial \ln D_O / \partial \ln(y_r / y_R) \geq 0$ $\varepsilon_{x_k}^O = \partial \ln D_I / \partial \ln x_k \leq 0$	$\ln D_O(\mathbf{x}, \mathbf{y}, \mathbf{z}, t) + u = 0, u \geq 0$ $[\boldsymbol{\varepsilon}_{\mathbf{y}}^{O\prime}\dot{\mathbf{y}} + (1/E^O)\boldsymbol{\varepsilon}_{\mathbf{x}}^{O\prime}\dot{\mathbf{x}}] = TFP = (1 - E^O/E^O)\boldsymbol{\varepsilon}_{\mathbf{x}}^{O\prime}\dot{\mathbf{x}} - \boldsymbol{\varepsilon}_{\mathbf{z}}^{O\prime}\dot{\mathbf{z}} - \varepsilon_t^O - du/dt$ Decomposition: $TFP = SEC + EXC + TC + EFC$
Dual expenditure (cost) for industry (value added sector) $j = 1, \ldots, J$, $\sum_{\forall k} \eta_{w_k} = 1$. Elasticity of scale: $E^C = \left(\sum_r \eta_{y_r}\right)^{-1}$	$\ln\left(C_j / w_K\right) = \ln c^j\left((1/w_K)\mathbf{w}', \mathbf{y}', \mathbf{z}', t\right) + u + v$ $\eta_{w_k} = \partial \ln c^j / \partial \ln(w_k / w_K) \geq 0$ $\eta_{y_r} = \partial \ln c^j / \partial \ln y_r \geq 0$	$\ln(\mathbf{w}'\mathbf{x})_j = \ln c^j(\mathbf{w}', \mathbf{y}', \mathbf{z}', t) + u, u \geq 0$ $[E^C \boldsymbol{\eta}_{\mathbf{y}}'\dot{\mathbf{y}} - \mathbf{s}'\dot{\mathbf{x}}] = TFP = (E^C - 1)\boldsymbol{\eta}_{\mathbf{y}}'\dot{\mathbf{y}} + (\mathbf{s} - \boldsymbol{\eta}_{\mathbf{w}})'\dot{\mathbf{w}} - \boldsymbol{\eta}_{\mathbf{z}}'\dot{\mathbf{z}} - \eta_t - (du/dt)$ s_k: observed cost share, $\sum_k s_k = 1$ Decomposition: $TFP^j = SEC + AEC + EXC + TC + EFC$ Aggregate: $TFP = \sum_j v_j TFP^j$

These derivations are adapted to discrete, e.g. annual, data as shown in Table 2. The analytical derivations of TFP used equation [24] which defines a Divisia index as the starting point. The formulation in Table 2 approximates the Divisia index by the Tornqvist index but this is not the only possibility. The paper by ten-Raa and Shestalova (2011) presents and explains the different possible approximations to the Divisia index for discrete data and outlines their properties. We illustrate in the example of the dual expenditure or cost function but the same ideas are applied to each of the other forms. Table 2 shows the TFP from the dual expenditure cost function for each component based on the estimated elasticity values for a panel data sample.

TABLE 2 HERE

Table 2 Index number calculations for TFP in the dual expenditure (cost) function case with panel data: $i = 1, \dots, I$ and $t = 1, \dots, T$

component	Tornqvist Expression	comment
SEC is $(E^C - 1)\boldsymbol{\eta}_{\mathbf{y}}'\dot{\mathbf{y}}$	$\frac{1}{2}\sum_{r=1}^{R}\left[(E_{it}^{C} - 1)\eta_{y_r it} + (E_{it-1}^{C} - 1)\eta_{y_r it-1}\right][\ln y_{\boldsymbol{rit}} - \ln y_{rit-1}]$	For one sector of the gross value added measure of GDP; multiple outputs are assumed
AEC is: $(\mathbf{s} - \boldsymbol{\eta}_{\mathbf{w}})'\dot{\mathbf{w}}$	$\frac{1}{2}\sum_{k=1}^{K}\left[(s_{kit} - \eta_{w_k it}) + (s_{kit-1} - \eta_{w_k it-1})\right][\ln w_{kit} - \ln w_{kit-1}]$	Weights are the divergence between actual and estimated optimal cost shares
EXC is: $-\boldsymbol{\eta}_{\mathbf{z}}'\dot{\mathbf{z}}$	$\frac{1}{2}\sum_{l=1}^{L}\left[(-\eta_{z_l it}) + (-\eta_{z_l it-1})\right][\ln z_{lit} - \ln z_{lit-1}]$	Exogenous variables could include quasi-fixed inputs in short run
TC is: $-\eta_t$	$-\frac{1}{2}\left((\partial \ln c_{it}^{j}/\partial t) + (\partial \ln c_{it-1}^{j}/\partial t)\right)$	Technological progress shifts the cost function for sector j downwards
EFC is $-(du/dt)$	$\ln(CE_{it}^{j}/CE_{it-1}^{j})$	Negative Inefficiency change is measured by the increase in cost efficiency

6 Estimation issues 1: between country differences: do they converge?

From the literature on macroeconomic growth models has evolved the idea of testing for convergence in the performance of countries over time.

Although economic historians have a long tradition of investigating national economic convergence, Baumol (1986) was one of the first papers by an economist to bring the topic to the forefront of economists' attention. Baumol's key empirical finding was relatively simple: he investigated the total productivity growth in GDP per labour-hour recorded in 16 major economies over the period 1870-1979 and regressed this against the productivity level measured in each country in 1870:

$$Growth\ Rate\ 1870-1979 = 5.25 - 0.75(\ln GDP\ per\ work - hour\ 1870)$$

[26]

With an $R^2 = 0.88$ Baumol concluded that the lower the starting level of productivity in a given country the higher was its subsequent rate of growth. In other words, unproductive economies caught up with the productivity leaders over a long period of time. However, Baumol also demonstrated that the catch-up effect was more pronounced in a cluster of market-orientated economies than in a cluster of centrally-planned economies, and the catch-up or convergence factor was absent in a cluster of less developed economies. Subsequently this empirical regularity was addressed theoretically and empirically by many other economists notably Barro and Sala-i-Martin (2004), who regress a model which states that the average growth rate of per capita real output y_{it} in country $i, i = 1 \ldots N$ over a fixed period depends negatively on the starting value $y_{i,t-T}$ and also depends on other variables, $\mathbf{z}'_{it}$:

$$(1/T)\ln\left(y_{it}/y_{i,t-T}\right) = a - \left(\left(1 - e^{\beta t}\right)/T\right)\left(\ln y_{i,t-T}\right) + \mathbf{z}'_{it}\boldsymbol{\mu} + u_{it}$$

[27]

They demonstrate that with a constant saving rate, the Solow-Swan theoretical one-sector growth model gives:

$$\beta = (1 - \alpha)(n + x + \delta)$$

[28]

α : elasticity of output with respect to capital in the Cobb-Douglas production function

n : rate of population growth

x : rate of labour augmenting technical progress, i.e. the steady state growth rate of output per capita, which we met earlier as Solow's residual measure of productivity

δ : rate of depreciation of the capital stock

Therefore, the log of income per effective worker is a weighted average of the initial value and the steady state value of income per effective worker, with the weight on the initial value declining exponentially at the rate β . Barro and Sala-i-Martin referred to this finding as beta-convergence. Subsequently, a different form of convergence was also identified: e.g. as stated in Young, Higgins and Levy (2008): 'when the dispersion of real per capita income across a group of countries falls over time there is sigma convergence; when the partial correlation between the growth in income over time and its initial level is negative there is beta-convergence'. Barro and Sala-i-Martin sum up the debate by stating:

Two concepts of convergence are:

i) a poorer country tends to grow faster than a rich one, (beta-convergence) (i.e. the transition growth rate to the steady state is higher the lower the initial value of output per capita)

and

ii) the dispersion of income per capita across countries diminishes over time (sigma-convergence). They suggest: beta-convergence can lead to sigma-convergence but new disturbances appear which offset this effect.

These ideas have been carried over to the literature on efficiency and productivity analysis of national economies, by incorporating into the analysis the measured data envelopment analysis or stochastic frontier analysis efficiency scores. Beta-convergence is measured by regressing the change in the log of country wide mean efficiency against the previous level of the log mean efficiency and the lagged log change. Beta convergence occurs if the coefficient on the lagged level is negative. Sigma-convergence is measured by regressing the change in the deviation in the log of country wide mean efficiency from the log of the whole sample mean efficiency against the lagged value and the lagged change of this deviation. Sigma convergence is said to occur if the coefficient on the lagged value of the deviation is negative. Panel least squares and GMM estimation are usually used. This type of analysis has typically been found in studies of national banking and financial systems, e.g. Casu and Girardone (2010).

However, a very different approach to the idea of convergence of TFP emerges from the development of endogenous growth theory in the 1990s. To understand how this relates to the neoclassical theory that we have discussed so far it is useful to go back to the simplest aggregate production function.

As we saw in section 2, TFP in much of the literature is measured as a residual between an index of outputs with weights summing to one and an index of inputs with weights summing to one as well. Such measures are defined by the standard neoclassical production function relating the aggregate output Y_{it} of country $i, i = 1 \ldots N$ at time $t, t = 1 \ldots T$ to its inputs of capital, K_{it} labour L_{it} and time. The impact of technological progress is contained in the role of the time variable which smoothly improves the production function as time passes

$$Y_{it} = f(K_{it}, L_{it}, t)$$

[29]

Usually an explicit assumption about the impact of technological progress would be made, and in the standard neoclassical growth model developed by Solow (1956) and Swan (1956) this took the form of labour enhancing technical change:

$$Y_{it} = f(K_{it}, L_{i0}e^{\rho_i t})$$

[30]

This form of the production function, when assumptions of positive but diminishing marginal products of the inputs and constant returns to scale are imposed in conjunction with the standard aggregate demand constraint, the definition of net investment as gross capital formation less depreciation of the capital stock and the Keynesian investment savings equilibrium condition, leads to an equilibrium in which income per capita grows at the constant rate ρ_i , and the capital income ratio and the consumption income ratios are constant. Eventually poorer countries would catch up with richer countries, and if there were international differences in ρ_i these would be unexplained since ρ_i is assumed to arise exogenously from a black box. In particular, only two explanations existed for the improvement in performance by different countries: either some had higher rates of input accumulation, especially capital, or some had faster trends in the productivity of labour. Initial research suggested that the second factor accounted for most US growth in the first half of the twentieth century, while considerable evidence (Krugman (1994), Young (1995)) favoured the first factor in the growth of the tiger-economies of south-east Asia in the second half of the twentieth century.

There emerged from this literature a set of 'stylized facts' about productivity growth, as described in Jones and Romer (2010).

1) Labour productivity has grown at a sustained rate.

2) Capital per worker has also grown at a sustained rate.

3) The real interest rate, or return on capital, has been stable.

4) The ratio of capital to output has also been stable.

5) Capital and labour have captured stable shares of national income

6) Among the fast-growing countries of the world, there is an appreciable variation in the rate of growth "of the order of 2–5 percent.''

In terms of the neoclassical growth model, the first five facts are predicted and fact 6) is left unexplained, it is simply the Solow residual which implies that growth arises in a country exogenously, who knows from where?

The importance of this from our point of view in comparing international economic performance is that equation [29] is the standard starting point for a very large part of the efficiency and productivity analysis literature. Technical change is identified with the passage of time and is usually assumed to be an exogenous factor in the estimation model, which seems to indicate no policy direction which could improve a country's prospects. However, this is very much at odds with subsequent developments in the macroeconomic productivity growth literature, leading Jones and Romer (2010) to define a new set of stylized facts appropriate to modern developments. In brief these are:

1) Increases in the extent of the market. Increased flows of goods, ideas, finance, and people—via globalization, as well as urbanization—have increased the extent of the market for all workers and consumers.

2) Accelerating growth. For thousands of years, growth in both population and per capita GDP has accelerated, rising from virtually zero to the relatively rapid rates observed in the last century.

3) Variation in modern growth rates. The variation in the rate of growth of per capita GDP increases with the distance from the technology frontier.

4) Large income and total factor productivity (TFP) differences. Differences in measured inputs explain less than half of the enormous cross-country differences in per capita GDP.

5) Increases in human capital per worker. Human capital per worker is rising dramatically throughout the world.

6) Long-run stability of relative wages. The rising quantity of human capital, relative to unskilled labour, has not been matched by a sustained decline in its relative price

It is fact 3) which started the trend towards endogenous growth models, and was originally noted by Romer (1986) who plotted the annual average growth rate of GDP per capita over the period 1960-85 for a large number of developing economies[12] against the income per capita in 1960 relative to the USA. The USA defined the technology frontier when the countries started growing and those with the lowest GDP per capita relative to the USA in

[12] Recall that the Baumol (1986) did not find convergence for countries outside a small sample of the most developed economies. Romer's finding of large variation in TFP rates for different countries used a much larger sample of chiefly developing countries.

1960 subsequently showed a much larger variation in annual growth rates than countries that started from a position closer to the USA. This led to the suggestion that it must be the behaviour of producers, consumers and policy makers in different countries that had the largest impact on the variations in national growth rates. Globalization, urbanization and human capital provision are now key factors in determining different rates of productivity growth and economic performance. This is both an incentive and an obstacle to efficiency and productivity analysis. It provides an incentive because the ability of efficiency and productivity analysis to incorporate different approaches and variables for modelling productivity growth is its main strength but the obstacle is that formulating a theory of production on these lines that can be summarised in an aggregate production function is very difficult. In particular, it is necessary to ensure that the return to capital including human capital does not diminish as capital is accumulated.

One way of thinking about this is shown by Romer (1994) and Stiroh (2001). Compare equation [29] above with the production function represented in equation [31] below:

$$Y_{it} = A(R)f(K_{it}, L_{it}, R_{it})$$

[31]

In this equation the new variable R is the stock of knowledge and ideas which may be partially embodied in human capital. Each country's output depends on its own stock of knowledge and ideas, R_{it}, but the production function shifts up over time because of the global stock of knowledge and ideas, R. It is the stock of knowledge that permits the non-diminishing returns to investment that mean that growth is not exogenously limited but can be endogenously determined. The return to investment in knowledge broadly defined is given by the marginal product of knowledge:

$$\partial Y_{it}/\partial R_{it} = A(R)f_{R_{it}} + f(K_{it}, L_{it}, R_{it})A'(R)(dR/dR_{it})$$

[32]

This can remain high even when $f_{R_{it}}$, the rate of return on the country's own knowledge stock for a constant state of global knowledge, tends to zero and it also incorporates a spill-over term in the last expression. Spillovers and more generally the concept that ideas and knowledge are non-rival goods which are only partially or perhaps not at all excludable means that productivity measurement incorporating endogenous growth theory offers a very wide range of modelling design possibilities but these, for example in the format of equation [31] may be difficult to incorporate into a standard efficiency and productivity analysis framework.

Nevertheless, there is a wide-ranging literature on spillover estimation particularly in the context of Leontief input-output analysis (I-O) which allows that commodities can be both intermediate inputs and final goods. ten Raa and Wolff (2000) offer an interesting suggestion for spillover measurement in the context of this input-output approach. Commenting that usually spillover effects in each sector are measured by a weighted average of R&D in the sectors supplying intermediate inputs, ten Raa and Wolff instead suggest that spillover effects in an industry can be measured by TFP growth in its supplying sectors and they build up an analysis of interindustry spillovers that distinguishes four factors: autonomous growth, R&D in the sector in question, direct productivity spillovers using the direct input output linkages between sectors to weight the supplying sectors' TFP growth rates and capital embodied spillovers using the investment coefficient of the supplying sector's capital to weight its productivity growth. ten Raa and Wolff (2000) then argue that productivity growth in a sector is counted in the sectors that trigger it. They find that for the I-O tables for the USA for 1958-

87, it is computers and office equipment and electronic components which are the sectors to which most productivity growth is imputed[13].

It seems essential therefore to allow for the widest possible range of variables in explaining international differences in productivity and performance and Jones and Romer suggest that differences in institutions must be the fundamental source of the wide differences in growth rates; by institutions they mean a very wide range of different factors in each society and economy that should be incorporated into TFP models at the international level, in particular that weak and strong institutions affect the adoption and utilization of ideas from leading nations on the frontier and that the potential for ideas to diffuse across nations amplifies the key role of institutions.

The role of institutions in TFP has been particularly strong in the work of North (1991) and Acemoglu et al (2005) whose definition of good economic institutions means those that provide security of property rights and relatively equal access to economic resources to a broad cross-section of society. The key argument here is the difference between the proximate causes of long run TFP, i.e. factors like innovation and the spill-over of ideas, which to North are not the causes of growth but are growth itself, and the fundamental causes of long run TFP which are embedded in the evolution of society and the emergence of good economic institutions. This poses a problem for researchers: efficiency and productivity analysis by necessity focuses only on the proximate causes of TFP and, even then, the issues such as the form of the production technology are difficult to model. Much more difficult is the problem of applying efficiency and productivity analysis to the understanding of the historical evolution of the fundamental causes of long run TFPC.

[13] The number of studies confirming the role of information technology in driving innovation and productivity growth throughput the world is expanding rapidly, see for example the long term study in Chen and Fu (2018).

One solution to the problem of measuring what we must now call the proximate causes of TFP, i.e. that incorporate the notions of innovation and ideas in the production technology is Sickles et al (2017). In this treatment, equation [29] is the starting point, i.e. the essential neoclassical formulation of the production function. When this is extended to incorporate ideas of endogenous growth theory, Sickles et al (2017) argue that the explanation for the spill-over that endogenously determines technology change is the loosening of constraints on the utilization of that technology, and that this this is just another way of saying that TFP is primarily determined by the efficiency with which the existing technology (inclusive of innovations) is utilized.

Transformed into an empirical equation: write y_{it} as the log of GDP per capita in country $i. i = 1 \dots N$ at time $t, t = 1 \dots T$, write $\mathbf{X}_{it}$ as the vector of logged inputs and other technology factors including innovations some of which may be endogenous, and write $\eta_i(t)$ to represent the country-specific fixed effect, which may be time varying, so that with the error term v_{it}

$$y_{it} = \mathbf{X}_{it}'\beta + \eta_i(t) + v_{it} \tag{33}$$

$$v_{it} \sim Nid(0, \sigma_v^2) \tag{34}$$

This is the generic stochastic frontier analysis model of the production function. This is the basic model for estimating efficiency change using panel data frontier methods. If we assume that innovations are available to all countries and that idiosyncratic errors are due to relative inefficiencies, then the country specific fixed effects can be used to capture the behavioural differences among countries that correspond to the key insight of the endogenous growth theory approach. Modern stochastic frontier analysis models offer a wide range of panel data methods for estimating the role of the country wide time varying fixed effects. The overall level of innovation change (innovation is assumed to be equally appropriable by all countries) can be measured directly by such factors as a distributed lag of R&D expenditures, or patent activity, or some such direct measure of innovation. In this way Sickles et al (2017) argue that the panel data methods incorporating endogeneity in the stochastic frontier analysis literature allow the researcher to address the issues raised by the endogenous growth models.

7 Estimation Issues 2: Technical change

The technical change component of the decomposition of TFP is written as we saw as the log-derivative of the technology representation with respect to time, e.g. in the case of the production function in Table 1 using the Tornqvist form:

$$-\tfrac{1}{2}\big((\partial \ln y_{it}/\partial t) + (\partial \ln y_{it-1}/\partial t)\big) \qquad [35]$$

The convention is to construct a very general role for the technical change component so that it shifts the whole production function (cost function) upwards (downwards) as illustrated in figure 1(a). The nature of the technical change may be classified as Hicks-neutral if the ratio of marginal products of two inputs remains unchanged when the ratio of the inputs is unchanged, or the technical change may be labour-augmenting or capital augmenting (Harrod neutral or Solow neutral). In the case of a Cobb-Douglas production function, all three forms of technical change have the same parametric form but the measured rate of technical change in the labour-augmenting (capital augmenting) case is the Hicks-neutral rate scaled down by the output elasticity of labour (capital). Alternatively, the technical change may be non-neutral in which case it will depend on the levels of the inputs and possibly other variables as well.

However, in an important but to some extent empirically neglected paper, Atkinson and Stiglitz (1969) discussed the idea of localized technical change, as illustrated in Figure 1 (b). Here, the smooth production function of Figure 1(a) is a limiting case of the piecewise linear production function arising in the activity analysis approach to representing technology. This is the approach that also underlies the concept of the efficient frontier in data envelopment analysis. Technical change may then apply to a subset only of the portfolio of blueprint techniques available to producers. Atkinson and Stiglitz give the appealing example of a technical change in textile production that applies to a single technique rather than to every technique from a fully automated loom to the crudest hand loom. The type of technical change which lifts the whole production function implies that technical progress spills over to every technique in the portfolio of technology. Localized technical change on the other hand limits the potential for spillovers from gains in knowledge from one form of production to another. The nature of the technical change is important here – the digital and information revolution may have much greater spill over potential for all techniques than for example the types of specific technological innovations which Gordon (2016) argues constituted the industrial revolution.

Two empirical issues arise from this concept of localized technical change, in addition to requiring a re-evaluation of the potential for spillovers. The first is that a technical improvement in a specific technique may have an effect on the piecewise linear production function that means other techniques are dropped from the portfolio of technologies. In Figure 1(b) we can see that a localized technical change in the marginal product of technique 2 only, shifts the corresponding segment of the piecewise linear production frontier so that technique 3 will no longer be relevant to production as the scale expands; production moves directly from technique 2 to technique 4. This has the added effect that improvements in a localized technology that result in potential techniques dropping out of the portfolio may reduce the elasticity of substitution between inputs.

The second issue is empirical. Modelling of the possibility of technical change and its part in the decomposition of total factor productivity change may require non-parametric or semi-parametric estimation techniques if the existence of localized technical change makes the usual parametric functions (in which technical change shifts the whole function) inappropriate for the data-generation process.

Figure 1(a) conventionally representing the role of technical change as a shift in the whole production function

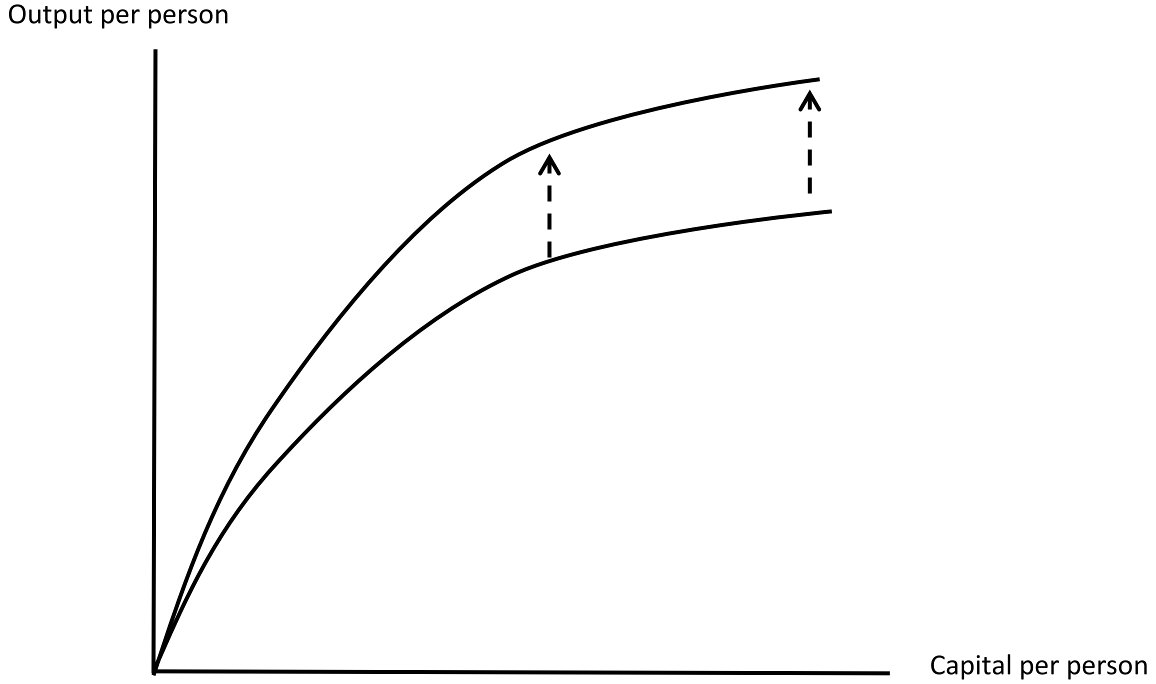
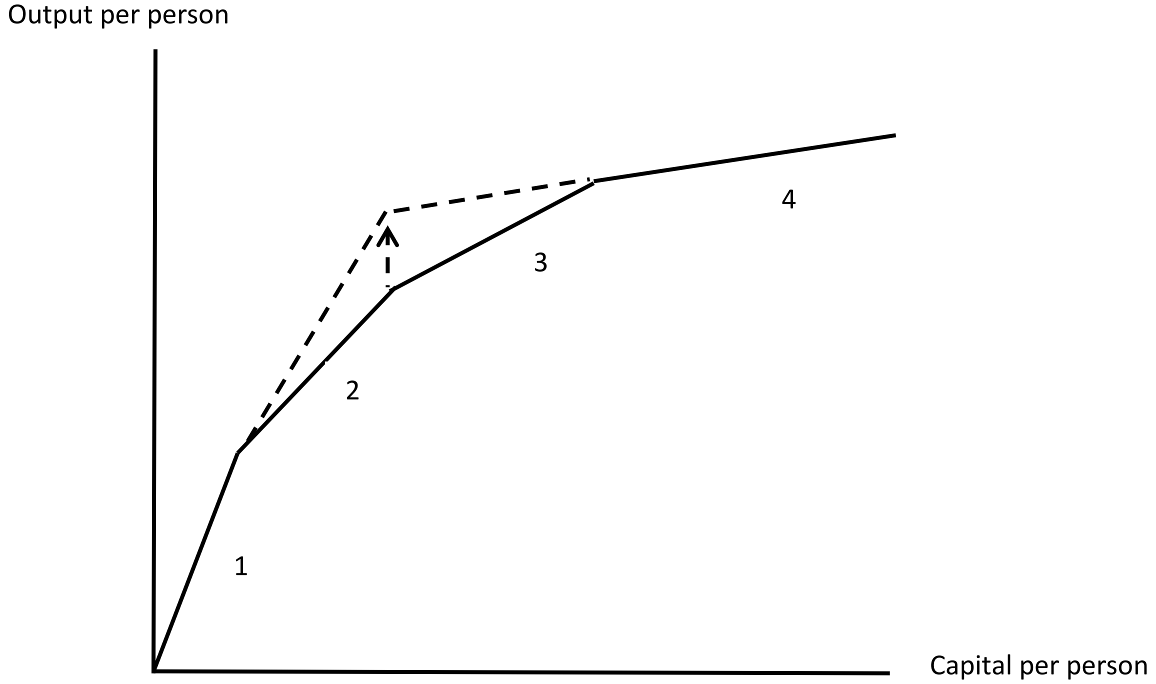


Figure 1(b) representing the role of technical change as a shift in a localised technique of production, Atkinson & Stiglitz (1969)



This idea has had a consistent following in the empirical literature since it first appeared and among the most recent treatments is Acemoglu (2015) where the link is made with induced technical change and developments in *localised, biased and directed* technological change. For example, Acemoglu points out that frontier technologies developed in rich, capital-intensive countries may be inappropriate to a capital-scarce developing economy where such machinery may be limited. This approach has however had limited impact yet on the efficiency and productivity analysis modelling that we have been surveying here, and it poses problems for the way in which stochastic frontier analysis models can be specified.

8. Estimation issues 3: spillovers and spatial effects

In further research one can consider similarities across countries since the performance of one national economy cannot be easily separated from that of its closest neighbours to assess how efficiency and productivity analysis using the newest developments in spatial econometrics can contribute to this question. In other words: why might spatial spillovers be important in understanding the performance of national economies?

Spatial analysis in general has a long history in statistical modelling, and spatial econometrics has become recognised in recent years as an important new field of applications. In this survey of work on the performance of national economies, we do not have space to give a full summary of spatial econometric applications but we can indicate briefly a relatively recent development which is the specification of a spatial econometric model with stochastic frontier analysis. We do this because there is a small emerging literature on using this approach to begin to understand the role of spatial spillovers on the performance of national economies with stochastic frontier analysis. Consider the aggregate production function relating output y to inputs $\mathbf{x}$, other exogenous variables $\mathbf{z}$ and time t together with the usual composed error term from stochastic frontier analysis as shown in table 1 above:

$$\ln y = \ln f(\mathbf{x}', \mathbf{z}', t) - u + v$$

[36]

Conventionally this is fitted as a Cobb-Douglas or translog functional form, but at present there is no allowance for spillovers onto the technology of one country from the technological advances in another neighbouring country. Spatial econometrics repairs this gap and we can make a simple start with the following Cobb-Douglas specification adapted from Glass et al (2016) for a cross section of countries labelled i or j over time periods labelled t.

$$\ln y_{it} = \alpha + \sum_{k} \beta_k \ln x_{kit} + \sum_{l} \gamma_l \ln z_{lit} + \rho t + \delta \sum_{j \neq i} w_{ij} \ln y_{jt} - u_{it} + v_{it}$$

[37]

This production function is very familiar in the first three expressions representing inputs and other exogenous variables with constant output elasticities together with Hicks neutral technological progress. The fourth expression however adds a weighted summation of the output levels in other countries to the explanation of production in country i. This spatial autoregressive model SAR represents the effect of accumulated spatial lags which are one way of modelling spillovers from one technology to another. With appropriate numerical values for the spatial weights matrix $[w_{ij}] = \mathbf{W}$ we can devise a set of explanatory variables whose spill over effects are captured by the estimated parameter δ. These imposed numerical values permit the researcher to investigate a variety of nearby neighbour effects based on geographical dispersion to capture possible spillovers. A multiplicity of extensions to this idea can be devised, see for example Greene (2017), which include the application of the spatial weights matrix to the explanatory variables and to the error components.

There are issues of interpretation of the results that require careful analysis, as Glass et al (2016) point out, “a unit in a spatial model is therefore simultaneously exporting and importing spillovers to and from its neighbours. The indirect marginal effects from a spatial model measure the magnitude of the spillovers which are imported and exported in the sample”. More interesting from our point of view of the estimation issues in the modelling of national performance by stochastic frontier analysis is how the composed error term can be addressed. Use can be made of a concentrated likelihood function approach first suggested by Fan et al (1996). This was used in the spatial autoregressive model by Glass et al (2016) and its use in non-parametric estimation is recommended by Kuosmanen et al (2015). We can show here the Glass et al (2016) procedure adapted to the problem in hand.

The first order conditions for the log-likelihood function for the stochastic frontier analysis model are still valid even if the frontier is unknown and estimated separately, provided it does not depend on $\sigma^2 = \sigma_v^2 + \sigma_u^2$ and $\lambda = \sigma_u / \sigma_v$.

Following Glass et al (2016), a two-step procedure is available:

Step 1: Solve the spatial regression model estimators retaining the SAR residuals

Step 2: Now use these to obtain the concentrated log-likelihood in terms of $\lambda = \sigma_u / \sigma_v$ only. Maximise this by grid search for $\hat{\lambda}$ and iterate jointly with $\hat{\sigma}^2$ to convergence

With the estimators obtained, a transformation of the usual measures of conditional efficiency can be derived and these depend on turn on the spatial lag effects; these results can then be written as direct, indirect and total efficiency measures. Glass et al (2016) used this approach to estimate a stochastic frontier analysis aggregate production function using aggregate data for 41 European countries for the period 1990–2011 with a dense spatial weights matrix based on distances. The output variable was real aggregate value added and the inputs were capital and labour with additional variables of export openness relative to GDP and Government expenditure relative to GDP. A key finding was that on average, countries are more adept at importing efficiency than they are at exporting efficiency. This finding is consistent with the diffusion of knowledge embodied in imports of hi-tech goods and services from a relatively small number of technological leaders in the sample (e.g. Germany).

9 Summary and conclusions

This chapter serves as an introduction to the issues of comparing the performance of national economies. We could not hope to survey in detail the massive number of empirical papers that have accumulated on this topic and of course the methodologies of data envelopment analysis and stochastic frontier analysis are well covered in other chapters therefore we could have little to add on strict methodology. Instead, we have opted to present a broad overview of a wide range of different topics of relevance to the general idea of comparing the national performance of countries.

We began with basic historical ideas that are still important for researchers coming new to the topic. Productivity comparisons are critical and are made every day in the media and in political and economic commentary. We showed how a myriad of different ideas have evolved from the original growth model of Solow and the identification of TFP with a residual. We questioned whether the key variable of real value-added GDP tells us anything about economic welfare and presented a few ideas on this from the current wealth of contributions that are available, including the suggestion of measuring the 'happiness frontier'. Then we investigated the roles of data envelopment analysis and stochastic frontier analysis in making efficiency and productivity comparisons amongst countries. Our key argument here is that the data envelopment analysis and stochastic frontier analysis approaches permit the relaxation of the major assumptions associated with TFP measures reported in the media and which are usually the basis of policy making. We explored ways in which the data envelopment analysis and the stochastic frontier analysis permit the researcher to relax the assumptions of allocative efficiency, constant returns to scale, absence of exogenous variable effects (the *ceteris paribus* assumption) and the absence of inefficient performance that characterise the conventional growth accounting or neoclassical approach to the comparison of national economic performance.

From there we investigated a number of estimation issues both settled and unsettled in the efficiency and productivity analysis approach to national TFP measurement.

We considered the ideas about convergence of national performance and how this might be measured, and we saw the contrast between the convergence in national economic performance and TFP rates predicted by the neoclassical model and the lack of convergence due to the endogeneity innovations associated with the endogenous growth model. A second estimation issue concerned the modelling of technological change and whether this applied to the whole representation of the frontier as is conventional in stochastic frontier analysis or whether we could consider localised technical change as initially suggested by Atkinson and Stiglitz. Intuitively it seemed as if data envelopment analysis or other non-parametric approaches could be more fruitful than stochastic frontier analysis in this context but researcher ingenuity will no doubt overcome this. The final estimation issue that we examined was the interface between spatial econometrics and stochastic frontier analysis and we gave an example of comparison of national economic performance in which the composed error term of stochastic frontier analysis was incorporated in a spatial autoregressive model using a concentrated likelihood estimation approach.

In many ways this chapter differs from the other technical chapters in this book. However, this is deliberate. Our intention has been to provide a broad overview of the whole context in which we can compare, as economists particularly interested in efficiency and productivity analysis, the performance of national economies. We have deliberately not attempted the impossible task of summarising the empirical literature on international differences in TFP, even those using only efficiency and productivity analysis since there are literally tens of thousands of such papers. Instead we have consciously taken a wide and eclectic view about what constitutes international economic performance, in the belief that the powerful tools of efficiency and productivity analysis will successfully address these massive issues gaining an accurate picture of how different countries compare with each other using the widest range of concepts of what constitutes a country's economic performance.

References

Acemoglu, D., Johnson, S., & Robinson, J. A. (2005). Institutions as a fundamental cause of long-run growth. *Handbook of economic growth*, Volume 1, 385-472. Edited by Philippe Aghion and Steven N. Durlauf, Elsevier B.V..

Acemoglu, Daron. (2015) "Localised and biased technologies: Atkinson and Stiglitz's new view, induced innovations, and directed technological change." *The Economic Journal* 125 (583) 443-463.

Aigner, D., Lovell, C. K., & Schmidt, P. (1977). "Formulation and estimation of stochastic frontier production function models". *Journal of econometrics*, 6(1), 21-37.

Atkinson, Anthony B. and Joseph E. Stiglitz , (1969) "A New View of Technological Change", *The Economic Journal*, Vol. 79, No. 315 (Sep., 1969), pp. 573-578

Baker, Dean, J. Bradford DeLong, and Paul Krugman (2005), "Asset Returns and Economic Growth," *Brookings Papers on Economic Activity* 2005:1.

Barro R. and X. Sala-i-Martin (2004) *Economic Growth*, The MIT Press

Bauer P. (1990) "Recent developments in the econometric estimation of frontiers", *Journal of Econometrics*, 1990, vol. 46, issue 1-2, 39-56

Baumol, William J, Robert E. Litan, Carl J. Schramm (2007) *Good capitalism, bad capitalism, and the economics of growth and prosperity*, Yale University Press

Baumol, William J. (1986) "Productivity growth, convergence, and welfare: what the long-run data show." *The American Economic Review* (1986): 1072-1085.

Bettio, Francesca & Janneke Plantenga (2004) Comparing Care Regimes in Europe, *Feminist Economics*, 10:1, 85-113, DOI: 10.1080/1354570042000198245

Bogetoft P. and L. Otto (2011) *Benchmarking with DEA, SFA, and R*. Springer-Verlag New York

Casu, B., & Girardone, C. (2010). "Integration and efficiency convergence in EU banking markets." *Omega*, 38(5), 260-267.

Caves, D. W., Christensen, L. R., & Diewert, W. E. (1982). "The economic theory of index numbers and the measurement of input, output, and productivity". *Econometrica*: Journal of the Econometric Society, 1393-1414.

Charnes, A., Cooper, W. W., & Rhodes, E. (1978). "Measuring the efficiency of decision making units". *European journal of operational research*, 2(6), 429-444.

Chen, Hao-Tsung and Fu, Tsu-Tan (2018) "Source of Growth Analysis at the industry-level for selected Asian Economies using DEA Malmquist and Asia KLEMS data bases", *Asia-Pacific Productivity Conference*, APPC2018 at Seoul University, Korea, July 2018.

Coelli T, A Estache, S Perelman and L Trujillo (2003). *A Primer on Efficiency Measurement for Utilities and Transport Regulators*, World Bank.

Coyle, Diane (2017) “Do-it-yourself digital: the production boundary and the productivity puzzle” *ESCoE Discussion Paper 2017-01*, Economic Statistics Centre of Excellence June 2017

Coyle, Diane, (2014) *GDP: A Brief but Affectionate History*, Princeton University Press, Princeton NJ, 2014.

Diewert, W. E. (1976). “Exact and superlative index numbers”, *Journal of Econometrics*, 4(2), 115-145.

Easterlin, R. (1974). “Does economic growth improve the human lot? Some empirical evidence.” In P. A. David & M. W. Reder (Eds.), *Nations and Households in Economic Growth: Essays in Honor of Moses Abramovitz* (pp. 89-125). New York: Academic Press.

Fan, Yanqin, Qi Li, and Alfons Weersink (1996) "Semiparametric estimation of stochastic production frontier models." *Journal of Business & Economic Statistics* 14.4 (1996): 460-468.

Fare, R., S. Grosskopf, M. Norris, and Z. Zhang. "Productivity Growth, Technical Progress, and Efficiency Change in Industrialized Countries." *American Economic Review* 84 (1994), 66-83.

Färe, R., & Primont, D. (2012*). Multi-output production and duality: theory and applications*. Springer Science & Business Media.

Farrell, Michael (1957) “The Measurement of Productive Efficiency”, *Journal of the Royal Statistical Society*, 120, 253-81

Folbre, Nancy and Julie A. Nelson For Love or Money -- Or Both, *The Journal of Economic Perspectives*, Vol. 14, No. 4, (Autumn, 2000), pp. 123-140

Giraleas, Dimitris, Ali Emrouznejad, and Emmanuel Thanassoulis. (2012) "Productivity change using growth accounting and frontier-based approaches–Evidence from a Monte Carlo analysis." *European Journal of Operational Research* 222.3: 673-683.

Glass, Anthony J., Karligash Kenjegalieva, and Robin C. Sickles. (2016) "A spatial autoregressive stochastic frontier model for panel data with asymmetric efficiency spillovers." *Journal of Econometrics* 190.2 (2016): 289-300.

Goodridge, Peter, Jonathan Haskel and Gavin Wallis (2016) “Accounting for the UK Productivity Puzzle: A Decomposition and Predictions” *Economica*, December DOI:10.1111/ecca.12219

Gordon, R. J. (2003). “Exploding productivity growth: context, causes, and implications.” *Brookings Papers on Economic Activity*, 2003(2), 207-298.

Gordon, Robert J, (2016) *The Rise and Fall of American Growth: The U.S. Standard of Living since the Civil War* Princeton University Press,.

Greene, William H. (2017) *Econometric Analysis* 8e Pearson

Haskel, Jonathan and Stian Westlake (2017) *Capitalism without Capital: the rise of the intangible economy*, Princeton University Press

Helliwell, John and Layard, Richard and Sachs, Jeffrey (eds.) (2012) *World happiness report*, The Earth Institute, Columbia University, New York, USA.

Jones, C. I., & Romer, P. M. (2010) "The new Kaldor facts: ideas, institutions, population, and human capital", *American Economic Journal: Macroeconomics*, 2(1), 224-45.

Jones, Charles I. and Peter J. Klenow (2016) "Beyond GDP? Welfare across Countries and Time" *American Economic Review*, 106(9): 2426–2457

Jorgenson, Dale, and Zvi Griliches. (1967) "The explanation of productivity change." *The Review of Economic Studies* 34.3: 249-283.

Kuosmanen, Timo, Andrew Johnson, and Antti Saastamoinen. (2015) "Stochastic nonparametric approach to efficiency analysis: A unified framework." *Data envelopment analysis*. Springer, Boston, MA, 2015. 191-244.

Krugman, P. (1994) "The Myth of Asia's Miracle." *Foreign Affairs* 62-78.

Lovell, C.A. Knox (2003) "The Decomposition of Malmquist Productivity Indexes", *Journal of Productivity Analysis*, 20, 437-58

Lukasz Rachel and Thomas D Smith (2015) "Secular drivers of the global real interest rate" *Staff Working Paper No. 571*, Bank of England

Milner, C., & Weyman–Jones, T. (2003). "Relative national efficiency and country size: evidence for developing countries". *Review of development economics*, 7(1), 1-14.

North, D. C. (1991). "Institutions". *Journal of economic perspectives*, 5(1), 97-112.

Orea, L. (2002): `Parametric decomposition of a generalized Malmquist productivity index'. *Journal of Productivity Analysis*, vol. 18, pp. 5-22.

Panzar, J. C., & Willig, R. D. (1977). "Economies of scale in multi-output production". *The Quarterly Journal of Economics*, 481-493.

Ray, S. C., & Desli, E. (1997). "Productivity growth, technical progress, and efficiency change in industrialized countries: comment." *The American Economic Review*, 87(5), 1033-1039.

Romer, P. M. (1986). "Increasing returns and long-run growth". *Journal of political economy*, 94(5), 1002-1037.

Romer, P. M. (1994). "The origins of endogenous growth." *Journal of Economic Perspectives*, 8(1), 3-22

Sickles, Robin C., Jiaqi Hao, Chenjun Shang, (2017) *Panel Data and Productivity Measurement* Chapter 17 of *Oxford Handbook of Panel Data*, Badi Baltagi (ed.) New York: Oxford University Press.

Simar, L. and Wilson, P. W. (2007). "Estimation and inference in two-stage semi-parametric models of production processes", *Journal of Econometrics* 136: 31–64.

Skidelsky, Robert. (2003) *John Maynard Keynes1883-1946: economist, philosopher, statesman.* London: Macmillan, 2003.

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

Solow, R. M. (1957). “Technical change and the aggregate production function”. *The Review of Economics and Statistics*, 39(3), 312-320.

Smith, Adam (1776) *An Enquiry into the Nature and Causes of the Wealth of Nations*, edited by R.H. Campbell, A.S. Skinner and W.B. Todd, 2 vols., Liberty Fund 1981

Stiroh, K. J. (2001) “What Drives Productivity Growth?” *Economic Policy Review*, 37, Federal Reserve Bank of New York, March 2001

Summers, Robert and Heston, Alan. "The Penn World Table (Mark 5): An Expanded Set of International Comparisons, 1950- 1988." *Quarterly Journal of Economics*, May 1991, 106(2), pp. 327-68

Swan, T. W. (1956). “Economic growth and capital accumulation”. *Economic record*, 32(2), 334-361.

ten Raa, T., & Mohnen, P. (2002). “Neoclassical growth accounting and frontier analysis: a synthesis.” *Journal of Productivity analysis*, 18(2), 111-128.

ten Raa, T., & Shestalova, V. (2011). “The Solow residual, Domar aggregation, and inefficiency: a synthesis of TFP measures.” *Journal of Productivity Analysis*, 36(1), 71-77.

ten Raa, T., & Wolff, E. N. (2000). “Engines of growth in the US economy.” *Structural Change and Economic Dynamics*, 11(4), 473-489.

Thaler, R. H. (2018). “From Cashews to Nudges: The Evolution of Behavioral Economics”, *American Economic Review*, 108(6), 1265-87.

Varian, Hal (2016) “A Microeconomist looks at productivity: a view from the valley” The Brookings Institution, https://www.brookings.edu/wp-content/uploads/2016/08/varian.pdf URL accessed February 2018

Winsten, C. B. (1957). “Discussion on Mr. Farrell’s paper.” *Journal of the Royal Statistical Society*, 120(3), 282-284.

Young, Andrew T., Matthew J. Higgins, and Daniel Levy (2008). "Sigma convergence versus beta convergence: Evidence from US county-level data.″ *Journal of Money, Credit and Banking* 40.5 (2008): 1083-1093.

Young, A. (1995). “The Tyranny of Numbers: Confronting the Statistical Realities of the East Asian Growth Experience." *The Quarterly Journal of Economics* 110(3):641-680.