DOCTOR OF PHILOSOPHY

The measurement and decomposition of economy-wide productivity growth

Assessing the accuracy and selecting between different approaches

Dimitris Giraleas

2013

Aston University

The measurement and decomposition of economy-wide productivity growth: Assessing the accuracy and selecting between different approaches

Dimitris Giraleas

Doctor of Philosophy

Aston University

May 2013

Aston University

The measurement and decomposition of economy-wide productivity growth: Assessing the accuracy and selecting between different approaches

Dimitris Giraleas

2013

Productivity at the macro level is a complex concept but also arguably the most appropriate measure of economic welfare. Currently, there is limited research available on the various approaches that can be used to measure it and especially on the relative accuracy of said approaches.

This thesis has two main objectives: firstly, to detail some of the most common productivity measurement approaches and assess their accuracy under a number of conditions and secondly, to present an up-to-date application of productivity measurement and provide some guidance on selecting between sometimes conflicting productivity estimates.

With regards to the first objective, the thesis provides a discussion on the issues specific to macro-level productivity measurement and on the strengths and weaknesses of the three main types of approaches available, namely index-number approaches (represented by Growth Accounting), non-parametric distance functions (DEA-based Malmquist indices) and parametric production functions (COLS- and SFA-based Malmquist indices). The accuracy of these approaches is assessed through simulation analysis, which provided some interesting findings. Probably the most important were that deterministic approaches are quite accurate even when the data is moderately noisy, that no approaches were accurate when noise was more extensive, that functional form misspecification has a severe negative effect in the accuracy of the parametric approaches and finally that increased volatility in inputs and prices from one period to the next adversely affects all approaches examined.

The application was based on the EU KLEMS (2008) dataset and revealed that the different approaches do in fact result in different productivity change estimates, at least for some of the countries assessed. To assist researchers in selecting between conflicting estimates, a new, three step selection framework is proposed, based on findings of simulation analyses and established diagnostics/indicators. An application of this framework is also provided, based on the EU KLEMS dataset.

Keywords: Data Envelopment Analysis, Stochastic Frontier Analysis, Growth Accounting, Monte Carlo simulations

This thesis is dedicated to Anna

Acknowledgements

This work would not have been possible without the help and support provided by my supervisors, Prof Emmanuel Thanassoulis and Dr Ali Emrouznejad. I am especially grateful to Prof Thanassoulis, who introduced me to the field of Efficiency and Productivity measurement all those years ago and later gave me this great opportunity to enter fully into the world of academia.

This work has been greatly enhanced by a number of discussions with academics in various conferences over the years. I would particularly like to mention the helpful comments provided by Prof Bert Balk, Prof William Greene and Dr Mika Kortelainen.

Lastly I would like to thank my friends and family for their moral support and encouragement throughout this journey.

This work is dedicated to Anna. Without her, this would be neither possible nor meaningful.

Table of Contents

List of tables

List of figures

Abbreviations

BC	Battese and Coelli (1988) estimator for the conditional mean of technical inefficiency
COLS	Corrected Ordinary Least Squares
CRS	Constant Returns to Scale
DEA	Data Envelopment Analysis
DMU	Decision Making Unit
EC	Efficiency Change
ESA	European System of Accounts
GA	Growth Accounting
GDP	Gross Domestic Product
GO	gross total Output
JMLS	Jondrow, Lovell, Materov and Schmidt (1982) estimator for the conditional mean of technical inefficiency
MAD	Mean Absolute Deviation
MI	Malmquist productivity Index
MLE	Maximum Likelihood Estimation
MSE	Mean Square Error
NACE	Nomenclature generale des Activites economiques dans les Communautes Europeennes; the industry classification standard adopted by the EU
OECD	Organisation for Economic Co-operation and Development
ONS	UK Office of National Statistics
PLF	Piecewise-Linear Function
PPPs	Purchasing Power Parities
RESET	Ramsey's Regression Equation Specification Error Test
SE	Scale Efficiency Change
SFA	Stochastic Frontier Analysis
SNA	United Nations System of Accounts
TC	TECHNOLOGICAL CHANGE
TFP	Total Factor Productivity
TMAD	Mean Absolute Deviation of the 25th percentile
VA	gross Value Added
VRS	Variable Returns to Scale

Chapter 1. Introduction and aims of the research

1.1. Productivity growth: a short introduction

> Productivity isn't everything, but in the long run it is almost everything. A country's ability to improve its standard of living over time depends almost entirely on its ability to raise its output per worker.
>
> Paul Krugman, The Age of Diminishing Expectations (1994)

> Over long periods of time, small differences in rates of productivity growth compound, like interest in a bank account, and can make an enormous difference to a society's prosperity. Nothing contributes more to reduction of poverty, to increases in leisure, and to the country's ability to finance education, public health, environment and the arts.
>
> Alan Blinder and William Baumol (1993), Economics: Principles and Policy, Harcourt Brace Jovanovich, San Diego

Arguably, the main objective of all economic activity is to increase the economic welfare of the participating actors. In the broad production framework, welfare is closely related to the output of the production process. This is true both in the micro setting, ie when studying individual firms, as well as the macro setting, ie when studying whole sectors of the economy or the economy of an individual country. At the micro setting, output strongly influences both costs and revenues, and by extension profits, which is what for-profit firms seek to maximise. In the macro setting, output plays an even more important role; Gross Domestic Product (GDP), a measure of aggregate output, is by far the most commonly used indicator of economic welfare at the national level.

In addition to output produced, there is a further, equally important dimension when considering the notion of economic welfare. This relates to the amount of effort needed to produce said output. 'Effort', as used here, is a broad concept and includes all factors that allow the production of output. These factors are all *inputs* to the production process and in broad terms include the necessary labour, plant and equipment (capital), raw materials or unfinished goods, externally procured services and energy resources required to produce said output.

By accounting for this additional dimension, a more complete measure of economic welfare can be devised, one that includes in its definition both the outputs and the inputs of the production process. The notion of Productivity, which is defined as the ratio of outputs resulting from a production process to the inputs used to generate such outputs, provides a very good fit to this ideal measure. In this context, both productivity and economic welfare are maximised by maximising outputs while at the same time minimising inputs.

Advances in productivity are important because they are the only true source of growth in an economy without the need of input expansion. Indeed, the simple neoclassical growth models (see Solow (1957)) argue that economic growth is only sustainable through productivity expansion, since economic expansion through continuous accumulation of inputs would soon exhibit diminishing returns. As such, the notion of productivity is important because productivity growth represents improvements in the production process, which directly translate to increases in welfare and subsequently living standards.

1.2. Aims of the research

Measuring productivity is not a straightforward process. As a first prerequisite for the creation of a relevant productivity measure, all economic inputs and outputs of the production process need to be accounted for and accurately measured. After collating this data, a method needs to be devised to allow the aggregation of different types of output and different kinds of input into two single, aggregate quantity measures, namely aggregate output and aggregate input.

There are a number of such measurement methods one could employ; in the micro setting, probably the most commonly used are the so-called frontier-based approaches. There are a number of different frontier-based approaches, but the common characteristic amongst them is that they rely on multiple observations of similar units (eg firms) to derive a common frontier that describes the relationship between outputs and inputs; this relationship forms the basis for estimating productivity growth. The most commonly used index for measuring productivity through frontier-based approaches is the Malmquist index (Caves, Chirstensen and Diewert (1982)). The most common frontier-based approaches for constructing this index are Data Envelopment Analysis (DEA), Corrected Ordinary Least Squares (COLS) and Stochastic Frontier Analysis (SFA). These approaches are discussed in more detail in chapter 3.

In the macro setting, the predominant method of measuring productivity growth is through Growth Accounting (GA), an approach that relies on index number theory. GA has its roots to the work of Jan Tinbergen (1942) and independently, to Robert Solow (1957). Both Solow and Tinbergen attempted to explain the total growth in production using data on labour and capital growth and found that input growth could explain only a portion of the observed output growth. They reasoned that for the identity of the production function to hold, an element of technological progress should be explicitly included in the production function. By rearranging the production function, this element is defined as the ratio of outputs to inputs, and, as such, explicitly measures the rate of productivity growth (GA is discussed in more detail in chapter 3).

In order to parameterise the GA production function, two general conditions need to be met:

- the researcher has access to both quantity and price data for all the relevant inputs and outputs; and, more importantly,
- the researcher needs to adopt a number of strict assumptions with regards both to the nature of the production function (constant returns to scale Cobb-Douglas production function) and the market structure in which the production takes place (perfectly competitive markets).

Frontier-based methods offer an attractive alternative for the measurement of productivity change at the macro setting, since their use can provide a number of advantages in the analysis of aggregate productivity. Firstly, frontier-based approaches do not require information on prices to generate estimates of productivity growth (although it should be mentioned that prices of lower-level aggregates for outputs and some of the inputs maybe required to calculate the higher-level aggregates used in the analysis). Secondly, unlike the more traditional GA methods, they allow for the production to occur inside the frontier, thereby relaxing the stringent assumptions of perfect competition, necessary for GA. Lastly, frontier-based methods also allow for the decomposition of productivity growth, which could be of great interest to the users of productivity change estimates (these issues are discussed in more detail in chapter 3).

The main disadvantage of the frontier-based approaches is that all of them require information on a suitable set of comparators (eg the economies of a number of countries) in order to generate the production frontier. In addition, the analysis needs to ensure that the inputs and outputs of each individual unit of assessment (eg the economy of a country) are collated and expressed in a manner that ensures the comparability between the various assessed units. On the other hand, GA estimates can be produced using country- or sector-specific National Accounts data (input and output quantities and prices), without recourse to information from outside the country or the sector examined; this is probably the largest contributor to the wide adoption of GA amongst policy makers.

The information and comparability requirements of the frontier-based approaches can sometimes complicate the analysis, but they are both issues that can be dealt with. On the issue of the availability of information, there have been a number of initiatives, sponsored by global or multinational organisations, to collate and provide National Accounts data from a number of countries at a centralised location (see for example the EU KLEMS[1] database). On the issue of comparability, the vast majority of the developed countries have adopted accounting standards, designed to ensure that National Accounts across countries are, or at least can be made, comparable (see for example the United Nations System of Accounts-SNA[2]). As such, frontier-based approaches can also be used in the macro setting.

In fact, there are a number of applications of frontier-based methods for the measurement of aggregate productivity growth in the academic literature. Färe et al. (1994) was one of the first studies that utilised Data Envelopment Analysis (DEA), the more widely-used non-parametric frontier based approach, to construct Malmquist indices of productivity growth; the approach has since been adopted in numerous other studies (for a comprehensive list of applications of DEA-based Malmquist indices see

[1] See http://www.euklems.net/ (accessed 17 May 2013).
[2] See http://unstats.un.org/unsd/nationalaccount/ (accessed 17 May 2013).

Fried et al. (2008)). Kumbhakar and Lovell (2000) introduced another way to construct a Malmquist index of productivity growth that relies on parametric frontier models, such as Corrected Ordinary Least Squares (COLS) and Stochastic Frontier Analysis (SFA); such models have also been widely used in the literature (see Sharma et al. (2007) for a list of sample studies).

However, despite the adoption of frontier-based methods in the academic literature and the theoretical advantages offered by frontier-based methods compared to the more traditional GA approach, there has been limited research on quantifying how these advantages translate into improved accuracy of the resulting productivity change estimates. In addition, there has also been limited research on the relative accuracy of the productivity estimates derived from frontier-based approaches, under different conditions.

The main contribution of this thesis is to provide quantitative evidence of the relative accuracy of different productivity measurement approaches under different conditions and to propose a framework that could be used to select the most appropriate approach for the application at hand. In more detail, the main aims of this thesis are:

– to provide an overview of the information required to assess aggregate productivity, with a discussion on the issues and the difficulties arising when collating such information;

– to introduce the three main approaches that can be used to measure productivity change at the aggregate level (ie at the economy and industry level), namely index-number approaches and frontier-based non-parametric distance functions and parametric production functions, and discuss their relative strengths and weaknesses;

– to assess the relative accuracy of the most common productivity measurement approaches, namely GA, DEA, COLS and SFA, under different conditions (different shapes of the production possibility set, different input volatility, technical efficiency and noise levels) using Monte Carlo simulations;

– to measure, compare and decompose (if possible) aggregate productivity change using both index-number and frontier-based approaches for a number of countries using a modern dataset created with the main purpose of productivity measurement (EU KLEMS dataset); and

– to suggest a framework that can be used to select the approach that is likely to be most accurate, in terms of estimating productivity growth, for a specific application and to demonstrate how it can be implemented in practice.

The findings of this thesis would be of great interest to both researchers in the field of productivity measurement and users of productivity estimates. For example:

– Policy makers at a national level: The thesis provides a detailed discussion on the merits of adopting frontier-based approaches for the measurement of aggregate productivity and on how these approaches can be implemented in practice, utilising data that are readily available through the system of National Accounts and international databases. As mentioned above, frontier-based approaches have the potential to provide more accurate estimates of productivity change, but also allow for the decomposition of productivity change; both of these 'features' would be very relevant to such users, as they can be of great value to analyses that aims on identifying and assessing possible productivity enhancement initiatives.

– Policy makers at the transnational level: More accurate productivity estimates and the ability to decompose them would also be of great interest to policy makers at the transnational level, ie policy makers at organisations such as the European Commission and the OECD. The more accurate estimates can in turn have significant policy implications; for example the disbursement of EU or national funds for development is informed by measures of productivity of regions and countries. Additionally, the thesis provides productivity estimates from a number of approaches utilising one of the most complete and up-to-date international datasets; these can provide an alternative view of the established global productivity performance picture, which so far has mostly been informed through GA-based estimates.

– Researchers: The results of the simulation analysis and the proposed selection framework (which heavily relies on these results) are likely to be of great interest to researchers (academics and practitioners) in the field of Productivity and Efficiency Analysis. The simulation analysis helps identify a number of characteristics inherent in the dataset that can have a significant impact on the accuracy of the resulting estimates, while the selection framework provides tools that can be used to assess and quantify these characteristics in practical applications. The findings of this thesis could help researchers make better informed choices when selecting between productivity measurement approaches that rely both on robust theoretical arguments and quantitative evidence. Additionally, the framework of the simulation analysis itself could be a useful guide to researchers that want to undertake similar analysis in the field of Production Theory.

Lastly, the findings of the analysis undertaken for this thesis are not necessarily restricted to the macro setting. In fact, the findings of the simulation analysis that assesses the relative accuracy of the different productivity measurement approaches are applicable in studies both on the macro but also in the micro setting, assuming of course that the conditions of said studies are similar to those assessed in the simulations. The same also applies to the proposed selection framework; despite the fact that the practice case revolves around assessing aggregate productivity growth, the selection framework can be just as easily applied in a productivity growth assessment from the micro setting.

1.3. Structure of the thesis

The thesis is structured as follows:

Chapter 2 introduces some key concepts when examining the issue of aggregate productivity growth and discusses the issues peripheral to the analysis. In more detail, chapter 2:

– introduces the neoclassical model of growth;

– discusses the system of National Accounts, which provides the necessary data to estimate productivity growth;

– provides a more detailed discussion on the aggregate outputs (and in particular the notions of Gross Output and Value-Added), aggregate inputs and their prices, as well as how they are constructed from the National Accounts; and lastly;

– discusses the issue of comparability of National Account information between different countries (in particular the most common Accounting Standards and Purchasing Power Parities).

Chapter 3 introduces the most common approaches that can be used to measure productivity growth and, for each approach, provides their theoretical background, discusses their strengths and weaknesses and provides some examples of how these approaches have been applied, drawn from the academic literature. The approaches covered in chapter 3 include: Growth Accounting, DEA-based Malmquist indices, COLS- and SFA-based Malmquist indices.

Chapter 4 presents and discusses the results from the Monte Carlo simulations on the accuracy of the productivity growth estimates derived from Growth Accounting and the assessed frontier-based methods (detailed in chapter 3) under various conditions. These conditions include the presence of technical inefficiency, measurement error, misspecification of the production function (for the GA and parametric approaches) and increased input and price volatility from one period to the next. The results of this analysis have been peer-reviewed and published in the *European Journal of Operational Research*[3].

Chapter 5 provides an up-to-date productivity assessment and decomposition for a number of mostly EU countries utilising a number of approaches and information sourced from the EU KLEMS dataset. Additionally, it introduces a framework that can be used to identify the approach that is likely to provide the most accurate estimates for the current application. The selection framework is based on assessing and quantifying a number of characteristics specific to the application/dataset at hand, selected based on the results of the Monte Carlo analysis undertaken in chapter 4. The characteristics in question include input volatility through time, the extent of technical inefficiency and noise present in the dataset and whether the parametric approaches are likely to suffer from functional form misspecification and are examined using a number of well-established diagnostics and indicators. The use of the proposed selection framework is demonstrated using the analysis of the EU KLEMS data as a case study.

[3] Giraleas, D., Emrouznejad, A., & Thanassoulis, E. (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.

Lastly chapter 6 summarises the main contributions of the thesis, discusses possible topics for future research and concludes.

Chapter 2. An introduction to aggregate productivity growth: Concepts and measures

2.1. Introduction

The aim of this chapter is to provide a short introduction on the issues surrounding the measurement of aggregate productivity growth. The concepts and issues discussed in this chapter are universal to all measurement approaches, ie they are pertinent regardless of the measurement approach (or approaches) selected for the analysis. In more detail, the chapter is structured as follows:

– Section 2.2 provides a short discussion on the importance of productivity growth to an economy, both from a theoretical perspective but also with regards to how official bodies, such as the European Commission, Eurostat and the OECD perceive the issue.

– Section 2.3 provides a short introduction to the standard economic model of production, which was first developed for the micro setting and has long since been adopted in the macro setting for the purposes of clearly defining and measuring productivity change. The discussion provided in this section is important because all productivity measurement approaches rely on the same theoretical underpinnings. This section also briefly discusses the differences and similarities of the notions of productivity and efficiency.

– Section 2.4 provides an extensive discussion on the various factors of production that are used in the macro setting, otherwise known as the inputs and outputs of the aggregate production process. The focus of this section is on the definition and the discussion on issues that affect measurement of the output and the primary inputs (namely labour and capital stock or capital services). This section also briefly discusses the process of aggregating similar units of production to higher economic aggregates.

– Section 2.5 discusses issues of comparability of aggregate data, both across time but mainly between different countries. In more detail, this section presents a brief overview on the various international standards for the creation of National Accounts data, which are central to the measurement of aggregate productivity change and briefly discusses the concept of Purchasing Power Parities (PPPs), which are used to ensure that price- and value-based data from different countries are comparable.

– Section 2.6 summarises and concludes.

2.2. The importance of productivity growth in the macro setting

Productivity is commonly defined as the ratio of a volume measure of outputs, be it goods or services, to a volume measure of the inputs used in their production. Measures of productivity form a key part of many international comparisons of economic performance. OECD[4] lists a number of cases where the use of productivity growth measures is essential:

> '…productivity data are used to investigate the impact of product and labour market regulations on economic performance. Productivity growth constitutes an important element for modelling the productive capacity of economies. It also allows analysts to determine capacity utilisation, which in turn allows one to gauge the position of economies in the business cycle and to forecast economic growth. In addition, production capacity is used to assess demand and inflationary pressures.'

Furthermore, the overall productivity improvement of the whole economy is a central measure of economic performance, as it is an essential factor for determining the growth potential of the economy. According to Eurostat[5],

> '…productivity indicators make it possible to draw conclusions about the growth potential, the associated inflationary risks and the resulting implications for national revenue. In other words, productivity sets decisive markers for economic and monetary policy.'

---

4 See Schreyer (2001).

5 The European Advisory Committee on Statistical Information In the Economic and Social Spheres, "Are we measuring productivity correctly?", Background paper for the 31st CEIES SEMINAR Rome, Italy 12 – 13 October 2006

More importantly, productivity growth is on the forefront of European Union (EU) policy making. The European Council meeting in Lisbon (March 2000) launched the "Lisbon Strategy" aimed at making the European Union (EU) the most competitive economy in the world and achieving full employment by 2010. According to the European Commission[6]:

> ‘Raising the long–term economic growth potential by increasing productivity growth is one of the fundamental objectives of the renewed Lisbon strategy and an important response to the challenges of globalisation, ageing, the rapid pace of technological progress and the need to combat climate change.’

The continuation of the 2010 "Lisbon Strategy" is “Europe 2020”, 10-year strategic plan set out by the European Commission (March 2010)[7]. The Europe 2020 strategy’s primary goal is to support employment, productivity and social cohesion in Europe. To achieve that, the Commission has proposed seven ‘Flagship Initiatives’, which aim to assist the member states in achieving their productivity potential. Most of the initiatives proposed by the broad strategy are linked to improving productivity, especially those relating to investing in innovation[8] and competitiveness[9].

With heavy emphasis being placed on achieving productivity growth, it is of utmost importance for governments and international development organisations to understand the drivers of productivity growth and their likely interactions, so that policies that foster the development of such drivers in the economy can be developed. Before attempting to identify such drivers however, the first necessary step in this process is to create accurate measures of productivity growth.

---

[6] Commission Of The European Communities (2007), ‘Communication from the Commission, Raising productivity growth: key messages from the European Competitiveness Report’, Brussels, 31.10.2007, p. 4

[7] Communication from the Commission (2010), “Europe 2020: A strategy for smart, sustainable and inclusive growth’, Brussels, 3.3.2010

[8] In “Innovation Union”, one of the main issues examined is whether Europe is improving its productivity and competitiveness and what are the links between total factor productivity growth and R&D intensity. See http://ec.europa.eu/research/innovation-union/pdf/competitiveness-report/2011/ (Accessed 10 May 2013)

[9] ‘Ultimately, competitiveness is about stepping up productivity, as this is the only way to achieve sustained growth in per capita income — which, in turn, raises living standards.’ Sourced from http://ec.europa.eu/enterprise/policies/industrial-competitiveness/competitiveness-analysis/european-competitiveness-report/files/ecr2011_full_en.pdf (Accessed 10 May 2013)

2.3. An introduction to productivity

2.3.1. The standard model of production

The starting point of the discussion on the issue of productivity has to be the definition and description of the production process. The production process is simply the transformation of the various factors of production (usually referred to as inputs) to a set of economic commodities (usually referred to as outputs). The entity (or entities) that facilitates this transformation process is called a production unit (sometimes also referred to as a decision making unit (DMU)), and can take the form of a single plant, a firm, a sector of the economy (all firms that produce similar outputs), or even a whole economy. The logical construct that describes the production process is the production function:

$$Y_{i,t} = f(X_{i,t}) \qquad \text{Eq 2.3.1}$$

, where $Y_{i,t}$ is a vector of outputs *Y* of a production unit *i* in time period *t* which is produced using a vector of inputs $X_{i,t}$ within the confines of the production technology described by $f(\cdot)$.

In the macro setting, $Y_{i,t}$ is converted to a single measure of total output (Y_t), also referred to simply as output, which is defined as the sum of the value of the goods or services that are produced in order to be made available for use in the wider market. This is a gross measure in the sense that it represents the value of sales, net of additions to inventories. Aggregate output is discussed in more detail in section 2.4.1.

The various inputs used in the production process are usually grouped into aggregate categories when economy- and industry-wide levels of aggregation are examined. The basic classification is comprised of capital (K_t), labour (L_t) and raw material inputs (M_t). In some cases, additional aggregate input categories are considered, such as expenditure on services (S_t) and energy consumption (E_{t},); when all the five input categories (capital, labour, energy, materials and services) are considered, the resulting analysis is said to have adopted the KLEMS framework. In all cases though, an important distinction between input types is made; labour and capital inputs are classed as primary inputs, while materials, services and energy are classed as intermediate inputs. Intermediate inputs are defined by the OECD as goods and services, other than fixed assets, used as inputs into the production process of an establishment that are produced elsewhere in the economy or are imported. They are either transformed or used up by the production process. Aggregate inputs are discussed in more detail in section 2.4.2.

When purchases of intermediate inputs are deducted from gross output, the output measure becomes a measure of 'Value Added'. So, in the aggregate setting, the production function is given by:

$$Y_{it} = h(K_{it}, L_{it}, E_{it}, M_{it}, S_{it}) \qquad \text{Eq 2.3.2}$$

or

$$VA_{it} = f(K_{it}, L_{it}) \qquad \text{Eq 2.3.3}$$

An important issue to address is how the production function changes over time. Solow (1957) proposed that the technological changes that happen over time can be incorporated into the production function by the inclusion of a simple scaling factor $A_{i,t}$ that affects the production function multiplicatively. In that case, equations 2.3.2 and 2.3.3 become:

$$Y_{it} = A_{it}^{Y}\, h(K_{it}, L_{it}, E_{it}, M_{it}, S_{it}) \qquad \text{Eq 2.3.4}$$

$$VA_{it} = A_{it}^{VA} f(K_{it}, L_{it}) \qquad \text{Eq 2.3.5}$$

By rearranging the above equations, one can derive the formal definition of productivity according to the neoclassical theory of growth:

Productivity (Gross Output) = $A_{it}^{Y} = \dfrac{Y_{it}}{h(K_{it}, L_{it}, E_{it}, M_{it}, S_{it})}$ Eq 2.3.6

Productivity (Value Added) = $A_{it}^{VA} = \dfrac{VA_{it}}{f(K_{it}, L_{it})}$ Eq 2.3.7

The productivity measures above are commonly referred to as multifactor or total factor productivity (TFP) measures, due to the fact that they take into account a number of different inputs used in production. The measures are also referred to as ‘disembodied’ or ‘Hicks-neutral’ productivity because they are not specifically related to any one individual factor of production. In other words, changes in the productivity measure result in increasing levels of output or decreasing levels of inputs, without affecting the parameters of the production function (ie, without changing the relationships between inputs and output). Note however that at least some of the frontier-based approaches examined in this thesis can relax this assumption (details on how this is achieved can be found in chapter 3).

Some of the most widely quoted measures of aggregate productivity are partial productivity indicators. Partial productivity measures are very similar to the productivity measures of equations 2.3.6-7, except the denominator is just a single input (usually labour), rather than the aggregate of all inputs used in the production possess. In fact, one of the most widely used partial productivity measures is output per employee, or labour productivity (LP_t).

$$LP_{it} = \frac{Y_{it}}{L_{it}} \qquad \text{Eq 2.3.9}$$

This measure is relatively straight-forward to calculate and offers the advantage of readability but, as all partial productivity measures, it does not reflects the joint influence of a number of factors used in the production process and thus does not provide a comprehensive estimate of productivity change. For a more detailed discussion on the problem arising from the use of partial productivity measures, see Thanassoulis et al. (1996).

2.3.2. Productivity and efficiency

There are two main dimensions in which productivity performance can be measured. The first is the measurement of productivity across time which answers the question of how is a particular economy performing year after year. The second dimension measures performance across a peer set, which is usually referred to as efficiency benchmarking. This attempts to answer the question of how is a particular economy performing in a single instance in time relative to other economies?

The measurement of productivity across time is usually achieved by the construction of a productivity index; changes in the index between time periods provide an estimate of productivity change, which usually the aim of the analysis. How the index is constructed depends on the productivity measurement approach adopted by the analysis. The most common approaches are discussed in detail in chapter 3 of this thesis.

The measurement and comparison of performance across a peer set is the goal of efficiency analysis. The notions of productivity and efficiency are related but also separate. Productivity is the ratio of outputs to inputs, while efficiency is the ratio of actual outputs to 'optimal' outputs, for set input levels, or the ratio of 'optimal' inputs to actual inputs, for set output levels. Both notions are measured residually, but in order to assess efficiency, one must first construct a benchmark that would reveal the 'optimal' outputs for given inputs or vice versa.

A combination of both approaches is also possible and in fact allows for a number of refinements to the estimation procedure and increased granulation of the measure produced, but with the cost of additional complexity. In fact, this is one of the main advantages afforded by the use of frontier-based approaches, as discussed in chapter 3 of this thesis. Productivity measurement in aggregate levels over time is usually examined by national statistical agencies and international organisations such as the OECD, using data from the National Accounts and Growth Accounting theory. International benchmarking of sectors and economies is not as common and usually undertaken by academics and more rarely by international organisations.

2.4. Inputs and outputs of production

This section discusses in more detail the various factors of production (ie the inputs) and the aggregate outputs that are commonly used in the assessment of aggregate productivity growth. The main sources of information for such analysis are the National Accounts of the countries (or sectors) that take part in the assessment. The UK Office of National Statistics[10] (ONS (2007)) provides the following definition of the National Accounts.

[10] Office for National statistics (2007), 'The ONS productivity handbook – A statistical overview and guide', edited by Dawn Camus, p.2

> ‘The National Accounts are a set of current values, volume measures and volume indices which, together, summarise all the economic activity of a nation. This can be defined as a central framework for the presentation and measurement of the stocks and flows within the economy. In the UK this framework provides many key economic statistics including gross domestic product (GDP) and gross national income (GNI) as well as information on, for example, saving, disposable income and investment. (p. 3)’

In addition, this section also examines some of the more important issues with regards to the definition and measurement of aggregate inputs and outputs, drawing heavily from the OECD productivity manual (Schreyer (2001)), referred to from now on as the OECD manual.

2.4.1. Output

The notion of the aggregate output is central in the measurement of productivity in the macro setting. This section provides the definition of three common measures of aggregate output, discusses their application in productivity analysis, lists some of the issues faced when measuring the output of non-market services and briefly discusses how to deal with changes in the quality of the goods and services that comprise aggregate output.

Gross Output and Value Added: Definitions

As mentioned in section 2.2.1, there are two major types of output measures traditionally used in aggregate productivity measurement, gross output and value added. GDP per capita is also used sometimes as a simple productivity indicator, where GDP is simply defined as Gross Value Added plus taxes minus subsidies. The definitions of the three major output measures are given below:

$$\text{Gross total output} \approx \text{Total Sales} + \text{Increase in Inventories and WIP} \qquad \text{Eq 2.4.1}$$

$$\text{Gross Value Added} = \text{Gross total output} - \text{Intermediate consumption} \qquad \text{Eq 2.4.2}$$

Intermediate consumption includes all non-primary factors of production (such as materials, services, energy, etc) that are either consumed or transformed by the production process.

$$\text{Gross Domestic Product} = \text{Gross Value Added} + \text{Taxes - Subsidies} \qquad \text{Eq 2.4.3}$$

GDP is not the preferred output measure when measuring aggregate, multi-factor productivity growth due to the distortionary effects the inclusion of taxes and subsidies would have on productivity estimates, both across time and across countries. Instead, the majority of productivity assessments rely either on gross Value Added (VA) or gross total Output (GO).

VA and GO are ‘gross’ measures because they are calculated before the depreciation of capital assets is accounted for. While it may be the case that, when comparing such measures directly, the concept of output net of economic depreciation should provide a better representation of the overall changes in welfare, production theory demands that output measures used in the production function are gross of any measure of depreciation. The logic behind this strong statement is related to the measurement of capital services in the macroeconomic setting. As will be discussed in more detail in section 2.4.3, capital inputs are measured either directly as the stock of capital assets or as flows of services that stem from such assets and depreciation is a major component in the calculation of both measures. In simple terms, the effect of depreciation is already accounted for in the measure of capital inputs; to also include it in the output measure would constitute a double-counting.

One important issue to note with regards to output is how to properly measure it when assessing aggregate units of production (sectors or whole economies). When the notion of output moves from the firm level to higher aggregates (eg sectors), it should be adjusted in such a way as to net-off flows of products and services within the aggregate in question. This is especially important in sectors that include relatively long supply chains.

For example, suppose that we want to examine a hypothetical wood products manufacturing industry, which includes two companies (for the sake of this example), a lumber mill and a furniture manufacturer. The lumber mill buys felled trees (intermediate input) and converts them into processed lumber (output), while the furniture manufactured buys the processed lumber from the sawmill (intermediate input) and uses it to produce furniture (final output). When examining the wood products manufacturing industry as a whole, there is only one intermediate input, ie the felled trees, and only one final output, ie the furniture produced. Since we moved up a level in aggregation, the wood products manufacturing industry is treated as a single, integrated entity and as such all transactions that happen within that entity (providing the output of the lumber mill as an intermediate input for the furniture manufacturer) are netted off.

This process of netting-off all intra-industry transactions is critical; otherwise both output and intermediate inputs would be double-counted and the importance of intermediate inputs would be disproportionate to their actual value. In addition, the consistency of both measures would suffer, since it would be possible to artificially inflate both measures if one would base the aggregation on increasingly smaller units of production (eg plant or even division level versus firm or group level).

An industry's gross output net of all intra-industry transactions is defined as sectoral output. Similarly, sectoral intermediate input is defined as the industry's intermediate inputs minus all purchases that took place within the boundaries of the industry. The issue of aggregation is discussed in more detail in section 2.4.3.

What is the most appropriate output when measuring productivity?

Productivity measurement can be based on either of the two output measures (VA and GO); the issue is that the resulting productivity change estimates will not be necessarily equal. The VA and GO productivity estimates are however related. As first demonstrated by Bruno (1978) and later by Balk (2009), under the so-called neoclassical assumptions, the rate of change of value-added based productivity growth equals the rate of change of gross-output based productivity growth, multiplied by the inverse of the share of value added in gross output:

$$TFP(VA)_i^t = TFP(Y)_i^t \times \frac{Y_i^t}{VA_i^t} \qquad \text{Eq 2.4.4}$$

where TFP(VA) is total factor productivity growth based on value added, TFP(Y) is total factor productivity growth based on gross output, Y is nominal gross output and VA is nominal gross value added. Since normally the share of gross value added to gross output is smaller than unity, estimates of value added productivity growth in a specific industry or country will normally be larger than their gross output based counterparts.

The aforementioned neoclassical assumptions are:

– all production is efficient (no technical inefficiency);

– all producers always choose the optimal mix of inputs to produce said output (no allocative inefficiency);

– all assessed units operate in a perfectly competitive environment (all prices are exogenous);

– the technology exhibits globally constant returns to scale.

The neoclassical assumptions are central to the measurement of productivity under Growth Accounting, an index number-based approach that is most commonly used for the measurement of aggregate productivity change. Growth Accounting and the neoclassical assumptions are discussed in more detail in chapter 3.

Given that the two output measures result in different productivity estimates, there is the issue of which output measure should be adopted by the analysis. As noted by Balk (2009), productivity assessments that utilise micro- or meso-data, in other words where the unit of assessment is either a firm or a larger sectoral aggregate, tend to use GO-based TFP measures, while productivity assessments that focus of more high level aggregates (groups of sectors or whole economies), tend to use VA-based measures.

In general, the use of VA-based productivity measures has a theoretical drawback in lower-level aggregates; they are derived from a restricted form of the production function which does not include intermediate inputs. Previous researchers assumed that productivity assessments that utilise VA output implicitly require that the underlying production function is separable with respect to intermediate inputs, but Diewert and Lawrence (2006) demonstrated that neither GO- nor VA-based productivity measures actually require this somewhat restrictive assumption.

Regardless, the condition of separability is not necessary for the VA-based measures to accurately represent productivity change, as shown by Balk (2009). And although gross output is the natural output concept, VA-based productivity measures are important when examining higher-level aggregates for a number of reasons. Balk (2008) states:

> ‘Gross output consists of deliveries to final demand and intermediate destinations. The split between these two output categories depends very much on the level of aggregation. Value added is immune to this problem. It enables one to compare (units belonging to) different industries.

> From a welfare-theoretic point of view the value-added concept is important because value added can be conceived as the income (from production) that flows into society.'

VA-based productivity measures are the most appropriate indicators of an industry's or economy's capacity to translate technical change into income and contribution to final demand, since intermediate consumption is not, by itself, wealth-creating. In effect, value-added productivity measures provide an indication on the additional output over and above the expected level of output a unit of primary inputs can generate for the economy. This, it could be argued, is a more appropriate indicator to base economic policy on, especially when considering economy-wide aggregates, since it translates directly to additional income generated for the economy.

Diewert (2008) is of the view that the final selection of mostly depends on ease-of-use of the resulting productivity estimate; he states:

> 'If we are studying the productivity performance of a particular firm or industry, then perhaps the gross output formulation is most suitable since it will be easier to explain to users. If we are attempting to analyze the productivity performance of an entire economy or an aggregate of industries, then the gross or net value added approaches seem preferable since economy wide growth in TFP will be approximately equal to a share weighted average of the industry growth rates in value added TFP. Thus the contribution of each industry's TFP growth to over all TFP growth is a bit easier to explain to users if we use the gross or net value added approaches.'

In conclusion, the use of both GO and VA can result in two dual, but numerically different productivity measures. GO-based measures are more appropriate for sectoral analysis, while VA-based measures are preferred when examining higher-level aggregates or whole economies. Measuring productivity is a difficult process that requires the mapping of a multidimensional process into one-dimensional space, and as such, there are a number of ways that this can be accomplished. As Balk (2008) concludes:

> 'This does not imply a break-down of measurement, but reflects a structural state of affairs.'

Measuring non-market services

The measures of output defined so far are all based on measures of sales. However, at least part of the economy produces goods or services that are either free at the point or delivery or supplied at prices that are not economically significant. The obvious example of such non-market output is the majority of the goods and services produced by the public sector. In the absence of any sort of meaningful market transactions, the traditional output measurement methods cannot be applied.

In general, in National Accounts the total value of output of a non-market producer is defined by convention as the total costs of production (i.e. the operating surplus is assumed to be zero). However, using this approach, often referred to as the (input=output) approach, makes the measurement of productivity growth impossible through traditional index number approaches (such as Growth Accounting), since it explicitly assumes that any and all change in output volumes is the result of changes in input volumes. For this reason, Eurostat has been advising all member states to develop processes for the direct measurement of non-market outputs based on *outcomes*, rather than on inputs. Although some progress has been made toward developing a system of accounts for non-market activities that could allow for the measurement of productivity (see Afonso et al. (2005) for an application and Diewert (2011) for a methodological discussion), most global agencies, including Eurostat, still employ the (input=output) approach (see for example the methodological paper on Health Accounting, OECD et al. (2011)). Even if a consistent accounting system based on outcomes can be implemented, the use of index number approaches for measuring aggregate productivity of non-market goods would still be problematic, due to the aggregation issues related to directly measured, non-value based output measures (Diewert (2011)). It should be mentioned here that frontier-based approaches can be readily used to measure non-market activities, even if the available data are based on the (input=output) approach; the only change from the methods that measure productivity for market activities is that the analysis moves from the production framework (ie estimating production functions) to the cost framework (ie estimating cost functions). Note though that in this setting, data on outcomes would also be required even by frontier-based approaches.

How to account for changes in quality?

Changes in the quality or the introduction of new products or services (collectively referred to as goods) can have significant effects on the measures of all factors of production and aggregate output and are especially pertinent when considering the issue of productivity change measurement.

The introduction of new goods or a change in quality of existing goods affects both the 'productive capacity' of the factor of production in question and also its price. If the changes in productive capacity were to be immediately and correctly reflected to the price of the factor in question, then changes in quality would not be a major concern for the analysis, at least for the factors of production whose volumes are constructed using price information (such as aggregate output, capital stock and intermediate inputs)[11]. There are a number of goods however where changes in quality are not immediately reflected in changes in prices; a good example of this is Information and Communication Technology (ICT) goods, where in the recent past quality is improving at an exponential rate while prices remain relatively stable.

The most common method of valuating new goods or quality change in the National Accounts is hedonic pricing. Hedonic pricing seeks to place a value on a particular good based solely on its characteristics. It usually involves utilising some form of regression analysis to explain the observed prices of a good or service based on its characteristics. For example, suppose that the price of a property depends on the characteristics of that property, ie type of property, number of bedrooms and bathrooms, location, garden size, neighbourhood, etc. By observing a number of property interactions in an area, one could construct an econometric model that links property prices to those characteristics. So when a new house becomes available, one could produce an estimated price for that house by measuring its characteristics and applying the estimated parameters of the econometric model. The resulting estimated price is referred to as the hedonic price of the property. For a more extensive discussion on the issue of hedonic pricing in the measurement of aggregate economic activity, see Triplett (2004).

2.4.2. Inputs

Aggregate inputs are divided into two major categories; primary inputs, namely labour and capital, and intermediate inputs, which can include materials, energy and services solicited outside of the production unit. This section provides the common definition of the different input categories and discusses some of the issues with regards to their measurement, focusing in particular to the primary inputs of production.

[11] For labour inputs, the issue of quality change is more nuanced and is discussed in more detail in section 2.4.2.1.

2.4.2.1 Labour inputs

For the majority of economic sectors, labour is the single most important input to the production processes. The most commonly used measure of labour input is ‘total hours worked’; it is also arguably the most appropriate measure, since it provides a clear correspondence to the output produced. However, data requirements to construct ‘total hours worked’ are quite significant and as such in some cases labour input is calculated based on the number of employed persons and an estimate of average hours worked.

Additional issues arise when considering the labour input from self-employed persons and part-time employees. The labour input of the latter is sometimes calculated as being half of the average input of a similar full-time position, due to lack of more detailed data. Notwithstanding some of the measurement issues, the OECD manual recommends that ‘hours worked’ should be the measure of choice for labour inputs, as opposed to simply using numbers of employed persons. If a measure of ‘hours worked’ is not directly available, the OECD manual states that ‘hours paid and full-time equivalent persons can provide reasonable alternatives’.

Aggregate labour input can be calculated by simply adding the ‘total hours worked’ for the assessed aggregate (sector or economy). This practice has the advantage of simplicity, but also implicitly assumes that labour is homogeneous. This is an oversimplification, since a number of factors can have a significant impact on the efficacy of labour, such as effort and the skills of the workforce. In effect, labour has both a time and a quality dimension; using ‘total hours worked’ captures the time dimension, but ignores any changes in quality. If this quality dimension is not captured in the input factor, its effects would be included in the productivity estimate. So, all improvements in labour quality that lead to increasing output would be interpreted as increases in productivity, if a simple ‘total hours worked’ labour input was used in the analysis. Whether this is desirable or not, would depend on the use of the final productivity measure. However, most practical applications try to incorporate this quality dimension to the labour input.

The way the quality dimension is incorporated into the labour measure usually depends on how ‘skills’ are measured. Some studies assume a direct relationship between ‘skill’ and occupation, while others try to include a number of additional differentiating characteristics. The OECD manual reports that factors that the relevant research has found to be important in labour differentiation are skill levels, usually expressed as a function of education attainment and relevant work experience, and other characteristics such as age, sex and health status. Therefore, to arrive at a consistent and accurate measure of labour input, all of the relevant factors that affect the contribution of labour to output need to be accounted for, in order to derive relative weights that could then be used to construct an aggregate measure.

The aforementioned quality adjustment can be implemented using hedonic pricing as mentioned above. However, since the data considerations for such an exercise are significant, aggregate labour input is commonly calculated based on some simplifying assumptions. Notably, if the analysis adopts the standard neoclassical assumptions, ie it assumes that the production unit is a price-taker, cost-minimiser and both input and output markets are perfectly competitive, then each production unit would employ additional labour input up to the point where the cost of an additional hour worked would be equal to the additional revenue that this input generates (for a brief overview of these assumptions, see section 2.4.1; these are discussed in greater detail in chapter 3). This implies that the price of labour, ie the wage rate, equals the marginal revenue of the production unit and as such is a good indicator for its relative importance in the production process. An aggregate measure of labour input can then be constructed by calculating the weighted average of ‘hours worked’ (L) using the share (w) of each type of labour (l) to total labour compensation (W) as the weights.

$$L = \sum_{i=1}^{N} \frac{w_i}{W} l_i \qquad \text{Eq 2.4.5}$$

Note that even when such simplifying assumptions are used, the information requirements for the calculation of the labour input measure are quite significant. Data is needed for the number of ‘hours worked’ by wage rate, by industry and by year. In addition, labour price indices are also required in order to control for annual changes in the general price levels and to reflect the labour market supply. As stated in the OECD manual, such rich data sets are usually both difficult and costly to collect and may therefore not be readily available in practice. In this situation, the use of GA methods when measuring economy-wide productivity may offer a usable alternative through the process of implicit labour differentiation.

Implicit labour differentiation is simply an artefact of the way the weights of each of the inputs is calculated under Growth Accounting, when data is available on ‘hours worked’ by industry without any distinction between different types of labour within each industry. When ‘hours worked’ measures are aggregated to the economy-wide level, the aggregation weight is based on each industry’s share in total labour compensation. These weights will be comparatively large for industries that pay above-average wages and relatively small for industries with below-average wages. If one assumes that higher skill levels demand higher wages, then the aggregation process implicitly takes into account the quality of the labour input.

Implicit differentiation is not available if the focus of the research is on industry-level productivity growth. However, this is less of a problem, if one assumes that productivity growth is output augmenting and as such has no impact on the output elasticities of the production factors. When taken together with the necessary assumption of the stability of the production function, the use of ‘hours worked’ measures unadjusted for differences in wage levels, but adjusted for changes in relative price levels would lead to a consistent productivity growth measure that would incorporate any unobserved changes in the quality of the workforce. In other words, any and all improvements in labour quality would be captured by the productivity measure, as mentioned above. This can be an advantage, since the resulting productivity measure could be used as the dependent variables in a second-stage regression analysis that examines likely determinants of productivity growth, with levels of education and/or skill/experience entering the analysis as independent variables.

2.4.2.2 Capital inputs

The measurement of capital inputs is one of the most difficult issues in the National Accounts and in productivity measurement in general. The problem stems from the fact that the capital services that flow from an asset are unobserved, both in terms of their ‘quantity’ but also in terms of their value, since when an asset delivers its services, no market transaction is recorded. In fact, the only market transaction related to capital assets that is readily available to researchers is the actual purchase of the asset and, possibly, expenditure undertaken to enhance the services of the given asset that take place at irregular intervals.

Defining capital services – the productive capital stock measure

Commonly in productivity measurement, capital inputs are defined as the flow of services that become available from the capital stock that a producer has accumulated via past investments. Since these services are not directly observable (eg the services provided by an office building or a computer network), both their quantity and their value need to be explicitly estimated. The basis of such estimation is what the OECD manual refers to as the ‘productive stock’ of each asset type (*K*), which is a function of cumulative past investment (*IN*), a price index of investment (*q*), a retirement function (*F*) and the age-efficiency profile of the asset (*h*):

$$K_{i,t} = \sum_{t=0}^{2} \left(h_{i,t} \times F_{i,t} \frac{IN_{i,t-T}}{q_{i,t-T,0}} \right) \qquad \text{Eq 2.4.6}$$

Past investment is discounted based on the asset-specific producer price index in order to derive real investment. Since even in a narrow definition of an asset type, technological progress may increase the quality of that particular asset at a later period, the relevant price index should ideally capture any impact of quality change. This can be achieved utilising hedonic pricing.

Real investment is then adjusted for assets that have run past their useful lives and have been retired (scrapped). The asset-specific retirement function is used to determine the share of cumulative assets that are still in service in a particular period, and can vary greatly by asset type, depending on the longevity of the asset in question (eg structures and buildings versus IT equipment).

The age-efficiency profile of the asset type represents the (possible) loss in the flow of capital services a typical asset experiences as it ages. Therefore, for new assets the age-efficiency profile takes the value of unity. It is possible that an asset is retired before its age-efficiency profile reaches zero, due to the effects of obsolescence. There are numerous ways to profile the productive efficiency loss of an asset and all of them require a degree of judgement. The most common methods include the use of a linear declining balance, a hyperbolic profile or a geometric one, where the productive efficiency of an asset declines at an ever increasing rate.

An assets' productive stock offers a workable representation of the capital services provided by said asset, since it accounts for retirements, the decline of productive capability due to age (wear and tear) and quality changes. Note that this value is not necessarily comparable with the notion of accounting depreciation, which represents the amount by which the value of the net (or 'wealth') capital stock, ie the current market valuation of all capital assets, of a country or an industry is decreased.

An asset's productive stock corresponds by definition to a specific asset type and for the production of any given type of output, numerous different assets are typically utilised. Therefore, a method to aggregate the capital stocks of different types of assets is required. Aggregation can be achieved by simple addition and the resulting measure can be used directly by frontier-based approaches for the purpose of productivity measurement. However, if the productivity analysis utilises index-number approaches, such as Growth Accounting, the aggregation process also needs to incorporate the relative value of the capital stock to production, to ensure the consistency of the resulting index. This aggregation process in effect converts the measure of capital stock into a measure of capital services that flow from each particular stock.

The relative value of the capital stock is most accurately represented by the notion of the user cost of capital, or the imputed rental price of capital, or its marginal cost. In simple terms, the user cost of capital represents the amount of rent that would have been required in order to secure the productive services of an asset for a single period.

Calculating the user cost of capital

User cost of capital is not directly observable and is thus usually calculated based on the market price of an asset over its useful life. The calculation requires the adoption of the standard neoclassical assumptions, similar to those required for the creation of a labour input index (production unit is a price-taker, cost-minimiser and operates in perfectly competitive markets). Under these conditions, the marginal cost of an asset (μ), ie the user cost of capital, equals its marginal revenue. Additionally, the price of an asset (q) at any point of time should be equal to the discounted sum of future revenues generated by said asset. These two relationships allow us to link the notion of a user cost of capital of an asset to the market price of the asset:

$$q_{i,t,s} = \sum_{\tau=0}^{T} \frac{\mu_{i,t+\tau,\tau+s}}{(1+r)^{\tau+1}}$$ Eq 2.4.7

,where q is the market value of the asset i at age s, μ is the user cost of capital (or the marginal revenue) of asset i which is summed and discounted by an r rate of discount over the current period until the end of its useful life.

Equation 2.4.7 can be refined by the inclusion of the age-efficiency profiles and the retirement function. The underlying assumption one needs to make is that the differences in user cost of capital between two assets of the same type but of different age is solely due to the relative productive efficiency decline, weighted by the probability that they are still functional. This is equivalent to accepting that two of the same type assets of different ages are perfectly substitutable, or, as the OECD puts it, ‘different vintages of the same asset type are perfect substitutes for each other’:

$$\mu_{i,t,s} = \mu_{i,t,0} \times h_{i,s} \times F_{i,s}$$ Eq 2.4.8

Note that this assumption ignores the potential quality change that could take place within the specific asset type. However, if the market price of the asset type is reflective of such quality changes (or is adjusted to take them into account), the final imputed user cost of capital measure will also incorporate them.

Applying equation 2.4.7 to 2.4.8, one can derive:

$$q_{i,t,s} = \sum_{\tau=0}^{T} \frac{\mu_{i,t+\tau,0} \times h_{i,\tau+s} \times F_{i,\tau+s}}{(1+r)^{\tau+1}}$$ Eq 2.4.9

which can be solved for the user cost of capital as[12]:

$$\mu_{i,t,s} = q_{i,t,s} \times r + (q_{i,t,s} - q_{i,t,s+1}) - (q_{i,t+1,s+1} - q_{i,t,s+1})$$ Eq 2.4.10

According to equation Eq 2.4.10, the user cost of capital is determined by:

– the financing cost of capital $q_{i,t,s} \times r$

[12] See Jorgenson (1963)

– the loss of value of an asset due to ageing and possible replacement, ie the rate of *economic* depreciation, $(q_{i,t,s} - q_{i,t,s+1})$ and,

– the difference in the general movement of asset prices, which also incorporates quality change if asset prices are quality adjusted $(q_{i,t+1,s+1} - q_{i,t,s+1})$

Despite the apparent simplicity of Eq 2.4.10, calculating the user cost of capital is very difficult to do in practice due to both the extensive information requirements (data for calculating both economic depreciation and the general movement in asset prices) but mainly due to difficulties finding an appropriate discount rate (or expected rate of return for the owner of the asset) to calculate the financing cost of capital.

Due to these difficulties, the majority of GA applications (which require that prices of all factors of production are available), use an endogenous, sometimes also called balancing, rate of return. The endogenous rate of return is set in such a way so that the total cost of capital plus the total cost of all other inputs equals exactly the final revenue (ie the product of output times its price). As such, the endogenous rate of return is an ex-post measure; its use also explicitly assumes that profit is always equal to zero[13].

The use of an endogenous rate of return is consistent with the neoclassical theory that views the user cost of capital as the marginal revenue for the owner of the capital. However, this relationship only applies if it is assumed that the neoclassical assumptions of perfect competition hold. Since these assumptions are quite restrictive, the use of an endogenous rate of return is theoretically unsatisfying. This issue is discussed in more detail in chapter 3.

Aggregation across different types of capital goods

Having defined and measured the productive capital stock and the user cost of capital, the process of aggregating several discrete asset types into a collective measure for capital services is relatively straightforward. The aggregation is achieved by using the relative user cost of capital as weight of an individual asset to create an aggregate measure. By weighting the index by the user cost of capital, more weight is placed on rapidly depreciating asset types, which reflects the reality of investors demanding

[13] It also requires the adoption of the standard neoclassical assumptions, such as the existence of perfect competition and a technology that displays constant returns to scale.

relatively higher rents for short-lived investments, to compensate for relatively higher depreciation costs.

Utilisation rate

One final important issue in the measurement of capital services relates to capital utilisation. The methodology described so far implicitly assumes that the rates of the utilisation of all primary production factors are constant throughout the timeframe of the analysis. This is unrealistic, given the numerous external factors that come into play in the demand for output or the supply of input in a production process. Unexpected machinery failure, interruptions in the supply of intermediate inputs, a general slowdown in demand due to an economic downturn could all lead to interruptions in the production process and capital-and possibly labour-under-utilisation. For countries that have more flexible labour markets, the impact of either output demand or input supply shocks will be less pronounced, since companies could, in theory, respond quickly to reduced demand by shedding excess labour. In the case of assets however, reducing capacity is a much slower process, if it is at all possible.

Most of production interruptions are random, and so it is expected that their effects would be normally distributed over the timeframe of the analysis and over the industries or countries considered. In that case, they would not introduce any bias in the estimated productivity growth, provided that the analysis adopts a medium to long-term view. Economic downturns however are an exception; the cyclical nature of overall demand is thoroughly documented and economic downturns are considered an integral part of modern economies.

If the effects of a downturn are not directly accounted for in the productivity growth measure, the measure will be biased, since productivity growth could likely be higher in years of increasing demand, as primary input utilisation increases, and lower in the years of the downturn. A simple way to correct for this possible bias is to examine productivity growth over a complete business cycle, or at least in similar points of the cycle. However, since not all business cycles are alike, the selection of an appropriate period requires a degree of judgement. In addition, business cycles do not always coincide when different countries are considered, which could cause additional issues in cross-country comparisons.

Given the above, adjusting for general capacity utilisation appears to be desirable if one seeks to measure productivity in the sense of technical change or outward shifts of a production function. However, an alternative viewpoint is that capacity utilisation should not be adjusted for if the aim of the productivity measure is to inform on real cost changes in the production process. The argument goes that, the full user cost of capital needs to be born by the producer irrespective of whether it is fully utilised or not; as the OECD manual states:

> ‘in times of recession, user costs of capital are spread over a smaller number of actual machine hours and consequently, real cost savings are limited. In times of cyclical upswings, the same user costs are spread over a larger number of machine hours and lead to more rapid real cost savings.‘[14]

If an adjustment for capital or general capacity utilisation is desirable, there are a number of approaches available which aim to develop an external measure of utilisation. Most of them rely on econometric models using an instrumental variable approach, with possible instruments ranging from business confidence surveys to intermediate input consumption. However, as the OECD manual states:

> ‘There have been several attempts to deal with this issue, but a generally accepted solution – if desirable – has yet to crystallise. In practice, statistical offices make no attempt to adjust their standard productivity measures for changes in the rate of capital and capacity utilisation.’

2.4.2.3 Intermediate inputs

The final component required for a productivity measurement analysis is information on intermediate input consumption. This information is used either directly in the analysis of Gross Output, or indirectly in the creation of a Value-Added measure of output. Intermediate inputs are either aggregated in a single input measure (often referred to as ‘materials’) or, in KLEMS analysis, are available in the more discrete measures of energy, materials and services. Aggregation is achieved using the relative value of each intermediate input to total intermediate input expenditure, similar to the aggregation methodology used for the primary inputs.

[14] The difference between the income generated by the asset in question and its opportunity cost is formally known as a ‘quasi-rent’. Quasi-rents are variable throughout the life of the asset and are affected by overall demand and rates of obsolescence. Quasi rents were first observed by Alfred Marshal (1842-1924).

2.4.3. Aggregating the units of production

The process of aggregating units of production (referred to here as simply aggregation) consists of grouping parts of similar economic activity to increasingly larger constituents, such as from department-level to firm-level to industry level and finally to economy-wide level. Aggregation is also accompanied by a process of integration, meaning that all intra-industry flows are netted out and the new aggregate is treated as a single, autonomous unit.

The core principles of productivity measurement hold for every level of aggregation, from individual firms to whole economies. The only requirement for the use of such an aggregation process is the assumption that each level of aggregation can be represented by a discrete production function. The process of aggregation itself requires the adoption of a system that can classify companies and other production units by the type of economic activity in which they are engaged; this provides a stable and uniform framework for the collection, categorisation and ultimately analysis and presentation of aggregate economic data. The industry classification standard adopted by the European Union is formally known as NACE[15].

In general, input and output data, both in terms of values and prices, are considered to be more accurate at more disaggregate levels, due to the greater granularity of the individual products and services and their prices. (see Eurostat (2001)). However, as disaggregation passes a certain point, reliability of data starts to deteriorate. There are a number of reasons why this is the case; firstly, at the very disaggregate levels, the available population of homogeneous units becomes ever smaller, which adversely affects the accurate construction of the price indices. In addition, data collation and aggregation in ever more detailed levels is expensive and gives the opportunity for more errors to creep in the measurement process. Comparability across time and units of assessment also becomes an issue, as at lower levels of aggregation the scope for product differentiation and development is significantly increased. As such, the relationship between data reliability and level of disaggregation is not straightforward.

[15] Nomenclature generale des Activites economiques dans les Communautes Europeennes.

2.5. On ensuring data comparability

As mentioned in section 2.3.2, there are two main dimensions over which performance could be measured, namely across time and/or across a peer set, ie a sample of units that undertake similar activities. If the goal is to measure the productivity of a single unit over a period, the minimum requirement would be that the data necessary for the analysis is consistent over time. If the goal is to measure efficiency (ie performance across a peer set), the minimum requirement would be that the data necessary for the analysis is comparable across the peer set. Lastly when performance is measured both across time and with a reference to a peer set, as is the case with the frontier-based approaches that will be detailed in chapter 3, the analysis needs to ensure that the available data is consistent and comparable in both dimensions, across time and peer set.

In the macro setting, comparability across time can be achieved through the use of the appropriate price indices, adjustments to capital stock values to account for different capital vintages and adjustments to both inputs and outputs to account for quality changes; these have all been already addressed in the sections above. However, the first requirement is to have in place a set of rules that define the various inputs and outputs, how they are measured and how firm-level data are aggregated to higher sector and industry classifications. Drawing up this set of rules, also referred to as standards for National Accounts, is usually the responsibility of the individual national statistics agencies; however, as economic activity became ever more globalised, there has been an increasing need for a common set of National Accounting guidelines at the international level, which would facilitate international comparisons amongst different economies. The result was the creation a unified, international Standard of National Accounts (SNA) that was developed under the aegis of the United Nations, which is currently enjoying widespread acceptance.

The adoption of international standards for National Accounts is one of the two main prerequisites of measuring productivity both across time and across different countries. The second prerequisite relates to the fact that different countries display different price levels; since prices are critical in the aggregation of outputs and inputs, the analysis also needs to ensure that such differences in price levels across countries are accounted for.

The most common methodology used for that purpose is the adoption of Purchasing Power Parities (PPPs). The remaining of this section will briefly discuss these two topics.

2.5.1. International standards for National Accounts

For productivity statistics to show dependable trends over time, they need to be produced from consistent measures of outputs and inputs. When comparisons across countries are required, care must be taken so that data consistency transcends national borders. Obtaining such consistent measures of output and input is one of the main challenges in estimating productivity. As mentioned above, this process is greatly facilitated by the adoption of international standards for National Accounts.

The main sources of guidance on international standards are listed below:

– United Nations System of National Accounts (SNA): This is the most widely adopted international standard and focuses more on providing high level guidance.
– European System of Accounts (ESA): Fully compliant with SNA 1993, provides the legal basis for harmonised accounts within the EU and is more prescriptive than SNA. The current implementation of ESA is ESA95.
– Eurostat Handbook on Price and Volume Measures in National Accounts: Expands on ESA 1995
– OECD Productivity manual: offers a comprehensive guide to aggregate productivity measurement based on Growth Accounting. Although it is not compulsory for the OECD member states to adopt its suggestions, it is currently considered as ‘the authoritative international source on methodology for productivity analysis’ (ONS (2007)).

2.5.2. Purchasing Power Parities

Productivity measurement in the macro setting relies on information on aggregate outputs and inputs. As mentioned in section 2.4, aggregate output is normally expressed in terms of value; the same holds for capital services, one of the primary inputs of the aggregate production process, as well as the various intermediate inputs used by the analysis. When the productivity analysis focuses only on the time dimension, the analysis only needs to apply a set of price deflators to these value-based indicators to achieve their conversion into the appropriate quantity measures. However, if the analysis is interested in measuring productivity across time and also relative to a peer set, additional adjustments are required to achieve like-for-like comparisons.

There are two main issues when it comes to comparing value-based indicators from two different economies. The first relates to the issue of the adopted currency; different countries usually have different currencies whose relative value may demonstrate large variation (for example, the exchange rate of 1 British pound was 124 Japanese Yen at the time of writing). The second issue relates to the fact that different outputs and inputs may well display different price levels in different economies; for example, a hamburger may be more expensive in the USA than it is in China, simply because the purchasing power of the average USA citizen is higher (additionally, the costs to produce one, especially the labour costs, are higher in the USA, but this fact is also closely linked to purchasing power).

One way to overcome the first issue is to use exchange rates to convert all value indicators to a common, base currency (this is usually the US dollar, but the choice of the base currency is not important). However, the issue with using exchange rates is that the adjusted indicators are still valued at the prevailing national price levels and do not accurately reflect the purchasing power of the currencies in their national markets. There are two main reasons for this:

– Exchange rates are greatly influenced by the supply and demand of the currencies in question, which is in turn influenced by factors such as capital flows, adopted monetary policy and currency speculation.

– More importantly, each economy produces a number of goods (products and services) solely for internal consumption; these goods are produced, traded and consumed domestically. Some examples of such goods include public services and residential housing. Since these goods are not traded in international markets, their impact in the setting of exchange rates is very difficult to estimate.

Given that exchange rates are not ideally suited to international comparisons of aggregates, another measure that can both convert values to a single currency unit and also accurately reflect the purchasing power of a currency in its native market is required. This can be achieved through the use of Purchasing Power Parities (PPPs).

Definition and estimation

Purchasing Power Parities are the rates of currency conversion that equalize the purchasing power of different currencies by eliminating the differences in price levels between countries[16]. Their goal is to reveal the relative purchasing power differential between two economies, or in simpler terms the difference in monetary terms of buying the same good (or basket of goods) in two different countries. The notion can be easier demonstrated using an example.

Assume that we want to examine the price relative, ie the ratio of observed prices, of an identical hamburger sold by an international fast food chain in the USA and China. The first step would be to use the exchange rate to convert the price of the hamburger in China in US dollars. For simplicity, let's assume that we find that the same hamburger costs 2 dollars in the USA but only 1 dollar in China; this means that for every one dollar spend in hamburgers in China, two dollars would have to be spend in the USA. As such, the price relative between the USA and China is 2, when using the USA as the base (or 0.5 when using China as the base), after controlling for differences in exchange rate. This prices relative reveals the difference in purchasing power and is the PPPs for hamburgers between the USA and China.

The estimation of PPPs usually involves three stages. The goal is to estimate price relatives similar to the example provided above. However, the calculation of price relatives for each individual product and service produced in an economy is not feasible. Therefore, price relatives are normally calculated for 'basic headings', ie baskets of similar goods for which information on final expenditure is available so that explicit expenditure weights can be estimated. As such, the first stage of the analysis involves the definition of these 'basic headings', a process that depends on the granularity of available data. At the second stage, price relatives are calculated for each individual 'basic heading'. The third and final stage involves the aggregation of these 'basic headings' into economic aggregates comparable with those included in the National Accounts. The PPPs for these economic aggregates are the weighted average of the 'basic heading' price relatives; the weights used in this process are the shares of each 'basic headings' on total expenditure (ie the total value of gross output).

---

[16] http://www.oecd.org/std/prices-ppp/ (accessed 13/07/2013)

To facilitate international comparisons, the basic heading' groupings used in the calculation of PPPs are individual to each country, so that they can better reflect the differences in spending patterns that result from economic, social and cultural differences.

The whole process of PPP estimation heavily relies on index number theory. There are a number of methods developed for this purpose over the years; for a more comprehensive review, see Hill (1997). For a more thorough discussion on PPPs, see the Eurostat-OECD manual on PPPs.

2.6. Summary and Conclusions

Productivity improvements are very important in the macro setting, since they represent the only way to increase economic prosperity without additional input accumulation. This is a view shared by the European Commission, which has put productivity growth in the forefront of EU policy making, as stated in the 'Lisbon Strategy' – EU's previous long-term strategic goals - and in 'Europe 2020', EU's current 10-year strategic plan. Other economic organisations, such as the OECD and Eurostat, are also contingent on the importance of productivity, not only for its end results, but also as a tool for measuring economic performance.

Given the importance of productivity, this chapter was devoted to discussing some of the more universal issues around the concept of productivity, how it can be measured and some of the issues that need to be addressed with regards to the data required for its estimation. In summary, productivity is defined as the ratio of a volume measure of outputs, goods and services, to a volume measure of the inputs used in their production. The central concept here is the production process and the logical construct used to describe it, namely the production function[17]. The production function links aggregate outputs to aggregate inputs; in the macro setting, inputs are grouped into the major categories of Capital, Labour and Material and services, with Capital and Labour classed as primary inputs and Materials and services as intermediate inputs, because these are consumed in the production process. Aggregate output is represented either as Gross

[17] Duality theory has shown that productivity measures can also be derived from the duals of the production function, namely the cost or revenue functions. These will be briefly discussed in chapter 3.

Output (GO), which is the sum of the value of the goods or services produced, or Value-Added (VA), which is simply Gross Output minus intermediate inputs.

Over time, the relationship between inputs and outputs changes; this can happen for a variety of reasons, the most important of which is technological change. Technological change can be incorporated into the production function as scalar that affects the production function multiplicatively; this scalar then becomes the formal definition of productivity since it directly represents the ratio of aggregate outputs to aggregate inputs (see equations 2.3.6 and 2.3.7 in the main text of this chapter).

So, in order to measure productivity, one has to have access to measures for the factors of production, namely inputs and outputs. Outputs are normally aggregated into a single measure, based on the value of each individual output produced (essentially, output is represented by sales). When aggregating output measures from different units of production to higher economic aggregates (eg from companies to sectors), care must be taken to net-off all intra-industry transactions, so that the resulting aggregate output does not double-count the contribution of intermediate inputs (this is discussed in detail in section 2.4.1). For the purposes of productivity measurement, both definitions of output (ie GO and VA) are valid for the analysis; it should be noted however that the resulting productivity measures will not be equal. In general, the consensus is that GO-based productivity measures are more appropriate for sectoral analysis, while VA-based measures are preferred when examining higher-level aggregates or whole economies. Additionally, care must be taken when measuring the output of non-market activities, since prices for those activities are either not available or they do not reflect their costs. In general National Accounts report the total costs of production of a non-market producer as the total value of its output. This treatment is unsatisfactory however for the purposes of productivity measurement, since it explicitly assumes that any changes in output are the result of changes in inputs. Due to this, Eurostat and OECD have advised its members to report outcomes for non-market activities, rather than outputs; progress however in this area is slow. Another issue with the measurement of output, and indeed all factors of production, is how to incorporate quality change in the resulting measure. The most common method of valuating quality change is through hedonic pricing, a technique that employs regression analysis to determine the price of a good based on its characteristics. The disadvantage of adopting this technique is that it greatly increases the data requirements and the analytical burden for the creation of National Account data. A more extensive discussion on the above issues is provided in section 2.4.1.

Labour is arguably the most important input of production and the most appropriate measure of labour input is ‘total hours worked’. However, direct data for this measure is sometimes difficult to obtain, especially for the self-employed and part-time workers. As such, the measure is sometimes imputed, based on information on the number of employed persons and an estimate of average hours worked. Measurement becomes more complicated when the quality dimension needs to be incorporated in the estimation; in this instance, quality represents the overall skill level and experience of the workforce. If this is required, quality can be incorporated using hedonic pricing, which further increases the data requirements of the analysis. An alternative to hedonic pricing is to aggregate labour inputs using the price of labour, ie wages, as weights; if the standard neoclassical assumptions hold, the production unit is a price-taker, cost-minimiser and both input and output markets are perfectly competitive, then the wage rate equals the marginal revenue of the labour resource and as such is a good indicator for its relative importance in the production process. It should be noted here however that incorporating quality into the labour measure is not necessary for the purposes of productivity measurement. If the quality dimension is ignored, then all changes in output that result from labour quality change will be captured by the productivity measure. If the issue of quality needs to be further examined, a second stage analysis (such as second stage regressions) can be attempted, which examines how the rate of productivity change is influenced by changes in the factors that are deemed to be appropriate indicators of labour quality. Labour inputs are discussed in more detail in section 2.4.2.

Capital is the second primary input and arguably the most difficult to measure. The difficulty stems from the fact that a production unit will acquire a capital asset in a certain point of time and use its services over a long time period; the problem is that the analysis can observe only the actual purchase of the asset and not its flow of services. The actual input to the production process however is the flow of capital services; as such they need to be estimated. This is usually done by building up a measure of capital stock that ideally takes into account the cumulative past investment in the general asset category in question, the price movements of the asset over time, the possibilities for asset retirement and the age-efficiency profile of the asset. Although this is a data intensive and quite complicated process, the resulting measure of capital stock provides a good representation of the value of services it delivers. However, index-based productivity measurement approaches also require data on the price of the estimated capital stock; such prices are also unobserved and need to be estimated using information on the rates of economic depreciation, asset price differentials between different groups of assets and the expected rate of return of the asset in question. This last element, the rate of return, is also observable and to this date there is no sufficiently robust method to estimate it. To circumvent this issue, the majority of index-based applications use an endogenous rate of return, set in such a way so that the total cost of capital plus the total cost of all other inputs equals exactly the final revenue. This is in essence an ex-post measure and its use explicitly assumes that profit is always equal to zero. This is theoretically unsatisfying and one of the most significant weaknesses of the index-based approaches, as will be further discussed in chapter 3. The above and other issues relating to the definition and measurement of capital inputs are discussed in more detail in section 2.4.3.

This chapter also briefly discussed the mechanisms and standards used to ensure that the data for the analysis are comparable, both across time but also across different economies (countries). Data comparability across time is achieved partly though the adjustments already discussed (for example, incorporating quality change and accounting for differences in relative prices) but also through the adoption of International standards of National Accounts. Arguably the most important and widely adopted standard is SNA, which is overseen by the United Nations. SNA forms the basis of the European System of Accounts (ESA), which is more prescriptive and is adopted by all EU member states.

The adoption of International standards of National Accounts Data also facilitates comparability across different countries. However, full comparability also requires a mechanism to convert the different inputs and outputs that are expressed in monetary terms to a single, common currency. This could be achieved using standard exchange rates; doing so however would fail to fully capture the differences in the purchasing power of the currency in question. In essence, using the traditional exchange rates to convert a value-based economic aggregate to a base/common currency still values said aggregate at national price levels; therefore such aggregates reflect both differences in the volumes produced in the countries and differences in their price levels, and as such they are measures of both value and quantity. So instead of relying on exchange rates, productivity analyses that rely on international comparisons commonly utilise indices of Purchasing Power Parities (PPPs), which are designed in such a way that they only reflect the differences in the volumes of the outputs and inputs in question. PPPs and the International standards of National Accounts Data are discussed in more detail in section 2.5.

From the above, it should be obvious that the process of collating the information necessary for productivity measurement is both data intensive and laborious. Furthermore, there are a lot of instances where a number of imputations or simplifying adjustment is required. Therefore, it is important that the approaches used to measure productivity should be flexible enough to still provide accurate estimates even if some of the assumptions required in constructing the data are violated.

Chapter 3. Common approaches for the measurement of multi-factor productivity

3.1. Introduction

The aim of the chapter is to provide a brief introduction to the most common approaches that can be used to measure productivity. The discussion in this chapter will focus on how these approaches can be used in the macro setting (ie when measuring productivity change in economic aggregates, such as aggregate industries or whole economies), but all of the approaches covered here can also be applied in the micro setting (ie when measuring the productivity performance of a single or a group of comparable units that engage in similar activities).

As a reminder, productivity is defined as the ratio of outputs produced by the assessed transformation process relative to the inputs used in said transformation process. When the transformation process produces just a single output and utilises just a single input, measuring productivity is a very simple process. However, the majority of transformation processes utilise a number of discrete input types and sometimes result in the production of multiple types of output. In these situations, the measurement of productivity becomes more complicated, since in order to calculate the outputs-to-inputs ratio, the analysis needs to create a single aggregate measure of output and a single aggregate measure of input.

Over time, a number of different methods have been proposed that attempt to deal directly or indirectly with this issue. The most commonly used productivity measurement methods can be divided into three main categories:

– Index-number approaches and Growth Accounting: These approaches adopt a number of assumptions that allow the analysis to directly aggregate inputs and outputs into single measures (indices) based on their relative prices, using a variety of index number formulae. As noted by Diewert (1992), the most commonly used indices are the Laspeyres, Paasche, Fisher and Törnqvist quantity indices. Each of those indices uses a different functional form to aggregate the various inputs and outputs of the transformation process. The most common index number approach for measuring aggregate productivity growth, according to the OECD Productivity manual (2001), is Growth Accounting (GA). Index number approaches and GA in particular are discussed in more detail in section 3.2.

– Non-parametric distance functions: These approaches manage the aggregation process based directly on quantity information on inputs and outputs and some minimal assumptions about the general shape of the technology (ie the transformation process). They require no information on prices and allow for the decomposition of the productivity measure into discrete components attributable to inefficiency, changes in scale, technological change and other effects that contain valuable information on how productivity changes over time. The most common distance functions for measuring productivity growth are based on the notion of the Malmquist productivity index, which was introduced as a theoretical concept in this setting by Caves, Chirstensen and Diewert (1982). Later Färe et al. (1992) demonstrated how the Malmquist productivity index can be estimated by Data Envelopment Analysis (DEA). Malmquist productivity indices and their estimation using DEA are discussed in more detail in section 3.4.

– Econometric approaches: Similar to the non-parametric distance functions, econometric approaches can also estimate a Malmquist productivity index, using only information on input and output quantities and a limited set of assumptions, mainly about the general shape of the technology and the distribution of the noise and inefficiency terms. Since they adopt the same productivity index as the non-parametric distance functions, econometric approaches can also decompose productivity into similar discrete components. The most common econometric approaches for estimating productivity change use either Corrected Ordinary Least Squares (COLS) or, more commonly, Stochastic Frontier Analysis (SFA) models. The estimation of Malmquist productivity indices under these approaches is discussed in more detail in section 3.5.

3.2. Index-number approaches and Growth Accounting

The previous chapter discussed in general terms how value (cost and revenue) information from National Accounts can be converted into quantity indices that can facilitate the measurement of productivity. In summary, there are two main issues that need to be addressed. Firstly, the various production units utilise a large number of diverse inputs and usually also produce a large number of different outputs. Secondly, the main aim of the analysis is to examine productivity performance over time; this raises issues with regards to the selection of the period which should be adopted as the basis for the calculation of the quantity indices. The need for inter-temporal comparisons together with the fact that both outputs produced by an assessed unit and the inputs used in the production process can hardly be considered homogeneous, especially at higher levels of aggregation, raises some important issues in the measurement of the factors of production.

Index number theory, combined with economic theory, provides the tools to collate the various discrete outputs (inputs) into a single, consistent-over-time, aggregate output (input) measure. Therefore, index number theory addresses two main issues:

- how to aggregate the discrete outputs (inputs) into a single measure
- how to consistently measure changes over time for these aggregates

The discussion in this section will focus on output aggregation; input aggregation will be discussed in detail in the following section that describes the Growth Accounting approach.

With regards to the first issue, since the various economic outputs are heterogeneous, one cannot simply add them together. So, to facilitate the aggregation process, the analysis would first need to ‘express’ the various outputs in comparable units of measurement and secondly find a way to combine them in such a way as to account for the temporal dimension of the measures. These two issues are interrelated, due to the fact that the aggregation formula selected will affect the process of converting to a single (or comparable) unit of measurement and vice versa. In other words, the analysis needs to first select a function for the aggregation and then create weights that correspond to the theoretical properties of the selected function so that they accurately reflect the relative importance of each output. The selection of the aggregation function is very important for making inter-temporal comparisons, since it will also have implications on how changes over time in the final aggregate measure are interpreted.

With regards to the measuring changes over time, the main question is whether to compare two non-consecutive periods directly (eg between period 0 and period 2) or indirectly (in which case the change between period 0 and 2 is derived from the change between period 0 and 1 and the change from period 1 to 2). As the OECD manual states (p. 83):

> 'The economics literature as well as the SNA 93 are quite unanimous in this respect: for inter-temporal comparisons, changes over longer periods should be obtained by chaining: i.e. by linking the year-to-year movements.'

The use of a chaining approach also simplifies the choice with regards to the price indices used. The question one needs to answer here is, which prices should the analysis adopt when examining inter-temporal changes, bearing in mind that the resulting quantity indices need to be consistent and comparable and that both prices and quantities are observable only at discrete intervals. There is no straightforward answer to this question, because consistency could either be achieved by keeping prices constant or by keeping quantities constant; the choice depends on the index number formulae adopted for the creation of the index. The issue is that by adopting either approach, accuracy is lost due to the simple fact that both prices and quantities are likely to change simultaneously over the year as a result of the substitution effect (as the price of an input increases, demand for that input decreases as the producer seeks to substitute the more expensive input with less expensive options). The main advantage of using chained indices is that they minimize the substitution bias that is potentially present in direct comparisons, since they utilize the highest frequency data available (ie the data at the highest level of granularity); so for example, if monthly data are available, an index of yearly change should ideally be created by chaining the monthly changes, rather than using just the information on the first and last month of the series.

Traditionally, there have been two main approaches on selecting between index number formulae, the axiomatic and the economic approaches (Diewert (1992))[18]. The axiomatic approach, which dates back to Walsh (1901) and Fisher and Brown (1911), identifies a number of desirable properties that the final index numbers should exhibit and

[18] Note that the two approaches are not mutually exclusive, as is discussed later.

mathematically tests whether the various index number formulae can lead to the creation of index numbers that display said properties. The economic approach examines economic theory and/or empirical evidence regarding producer behaviour and attempts to select an index number formula that satisfies the theory and/or the empirical evidence.

In more detail, the economic approach starts by postulating that the transformation process can be represented by a function that characterises the transformation technology (ie by a production, cost, revenue or profit function) and the producers are exhibiting competitive optimising behaviour. In other words, the producers seek to maximise their profits or minimise their costs according to the available technology (transformation function), while operating in markets that display the characteristics of perfect competition. These two conditions are commonly referred to as the standard or neoclassical assumptions. These neoclassical assumptions were briefly described in chapter 2 and will be discussed in more detail in section 3.2.1, since they provide the theoretical foundations of GA and are indeed one of the main motivating factors of this thesis.

It should be mentioned here that the two general approaches, axiomatic and economic, are not clearly distinct from one another. The axiomatic approach requires some assumptions based on economic theory in order to achieve output aggregation (these relate to the calculation of output shares); for input aggregation, the full set of neoclassical assumptions is required. The economic approach also relies on some of the tests developed over the years from the axiomatic methodology in order to determine the exact properties of the resulting productivity index. As such, modern applications of index number theory for the measurement of productivity growth rely on a combination of the axiomatic and economic approaches when selecting the most appropriate index number formulation for the application at hand.

The more common index number formulae are presented below. To assist in this discussion, a brief description of the notation: suppose that the number of discrete outputs is N in periods t, where t=0,…,T, and that there information available both for the price of each output p_t^n and its quantity q_t^n for all the periods that are relevant to the analysis.

The Laspeyres output index is given by:

$$L_t^{output} = \frac{\sum_{n=1}^{N} p_0^n q_t^n}{\sum_{n=1}^{N} p_0^n q_0^n} \qquad \text{Eq. 3.2.1}$$

In other words, the Laspeyres output index is the total value of output produced in time t as measured by the prices in period 0 divided by the total value of output produced in time 0 as measured by the prices in period 0. So, the Laspeyres index captures the changes in the quantity of output, based on base (starting) period prices.

The above formula can be rewritten as:

$$L_t^{output} = \sum_{n=1}^{N} s_0^n \frac{q_t^n}{q_0^n} \qquad \text{Eq. 3.2.2}$$

,where

$$s_t^n = \frac{p_t^n q_t^n}{\sum_{n=1}^{N} p_t^n q_t^n} \qquad \text{Eq. 3.2.3}$$

In other words, s_t^n is the share of the value of each particular output to the total value of all outputs and is the weighting mechanism that determines the relative contribution of each output to the final index. All of the index number formulae described in this section use s_t^n as the main weighting mechanism for aggregating output. In the Laspeyres index, the output shares used are the base (starting) period shares.

The Paasche output index is similar to the Laspeyres, except that it relies in end period prices, rather than start period prices for its calculations. In more detail:

$$P_t^{output} = \frac{\sum_{n=1}^{N} p_t^n q_t^n}{\sum_{n=1}^{N} p_t^n q_0^n} \qquad \text{Eq. 3.2.4}$$

The above formula can be rewritten as:

$$L_t^{output} = \left[\sum_{n=1}^{N} s_t^n \left(\frac{q_t^n}{q_0^n} \right)^{-1} \right]^{-1} \qquad \text{Eq. 3.2.5}$$

Note that the Paasche is calculated based on the share of each output to total output based on end period prices; as such, the Paasche index captures the changes in the quantity of output, based on end period prices.

Both the Laspeyres or Paasche indices have been widely used in both output and input aggregation when drafting National Accounts data, since they are both quite easy to calculate and have limited data requirements. They both however suffer from a serious limitation in that they rely on the share calculated for just a single period to carry out the aggregation. As the OECD manual states (para 157, p.89), this

> '…implies an underlying fixed-coefficient technology for the production structure – clearly a strongly simplifying assumption because it excludes the possibility of substitution between inputs or outputs, and implies constant marginal products throughout.'

Due to this limitation, Diewert (1976) argues for the adoption of more flexible index numbers, such as the Fisher and Törnqvist indices. He defines these indices as 'superlative', because they can be directly derived from flexible functional forms, such as the translog or the quadratic functional forms. These functional forms are deemed to be flexible because they can provide a 'second-order approximation to an arbitrary, twice differentiable linear homogeneous function' (Diewert (1976)). The use of such indices can greatly increase the accuracy of the analysis, if the underlying production function is displaying the characteristics of a flexible functional form. For example, if one believes that the production technology, or in other words the production function, can be described by the quadratic functional form, then the Fisher output index provides an 'exact' representation for the aggregate output for this particular production function, *under standard (neoclassical) assumptions* (Balk (1998)).

The Fisher quantity index can be directly derived from a quadratic functional form and is simply the geometric average of the Laspeyres and Paasche indices, ie:

$$F_t^{output} = (P_t^{output} L_t^{output})^{1/2} \qquad \text{Eq. 3.2.6}$$

Lastly, the Törnqvist index, which is the exact approximation of the translog functional form, is given by:

$$T_t^{output} = \prod_{n=1}^{N} \left(\frac{q_t^n}{q_0^n} \right)^{\frac{1}{2}(s_0^n + s_t^n)} \qquad \text{Eq. 3.2.7}$$

The Törnqvist index plays a major role in the GA framework; in fact the GA formulation for aggregating inputs, as adopted by EU KLEMS can also be seen as a Törnqvist input index, which is used as an approximation of the Divisia index; section 3.2.1 provides additional discussion on this topic.

Another way to measure change in TFP is through the use of Divisia index numbers, which treat both inputs and outputs as continuous time variables.

A Divisia index is a theoretical construct that can be used to generate indices from continuous time variables (components). It is defined as the weighted sum of the growth rates of the various components, where the weights are the components' shares in total value. Since it assumes continuous time, each component's share in total value will always be a function of time and therefore there is no need to choose whether they should be measured relative to base or to current value. As Balk (2005) notes:

> The novelty of Divisia's indices was that, as functions of continuous time, they take into account the prices and quantities of all, infinitely many, intermediate periods. Thus a Divisia index number is not only dependent on the initial and final points of the time interval considered, but will as a rule depend on the entire path that the prices and quantities belonging to an economic aggregate under consideration have taken.

This property is very helpful in theory, but the fact remains that we can only observe input and output price and quantities in discrete intervals. Therefore, in order to use a Divisia index, we need to find an index number approach that can provide a discrete approximation to a continuous Divisia index.

Balk (2005) has demonstrated that under certain conditions, virtually all chained indices can be conceived as a particular approximation to a Divisia index. The OECD Productivity manual (2001) recommends the use of the Törnqvist index, mainly due to the fact that it provides an exact approximation of a flexible functional form (in this case, translog). In practice, the OECD Productivity manual (2001) observes that the differences in the resulting Divisia TFP growth estimate under different index approaches are marginal at best, so long as chained indices are used and the intervening periods between observations are reasonable (ie annual). Nevertheless, the use of a superlative index number approach is recommended.

As mentioned above, the Fisher and Törnqvist index formulations are generally preferred when aggregating outputs, since they are the exact representations of 'flexible' transformation functions. According to Diewert (1992), the choice between the two is largely down to preference. The relative strengths of the Fisher index are that it has a larger number of desirable properties, according to the axiomatic approach, relative to the Törnqvist index; it also has potentially greater intuitive appeal, since it is a combination of the two most common index formulae, namely the Laspeyres and Paasche indices. On the other hand, the Törnqvist index offers an exact representation of the translog production function (assuming that the neoclassical assumptions hold), which is arguably the most commonly adopted functional form used in econometric analysis of transformation functions[19]. It should be noted however that in most empirical applications, the choice between the two indices is unlikely to cause issues, since the differences in the resulting aggregates are minor (see OECD Productivity manual (2001) for more discussion).

3.2.1. Growth Accounting and total factor productivity

The section above briefly demonstrated how index number theory can be used to combine in a consistent manner discrete outputs into a single, aggregate measure of output. It also mentioned that a similar procedure can be adopted so that discrete inputs can be aggregated into a single measure, which would allow for the calculation of the productivity ratio. According to Diewert and Nakamura (2009), input aggregation can be facilitated by incorporating information about the price of inputs into the analysis. This is done in a similar manner to the output indices, where price information is used to construct the share of the value of each particular output to the total value of all outputs. When aggregating inputs, input prices can be used to construct the share of the value of each particular input to the total *cost* of all inputs used in the production function; the share of each input to the total costs of production would then constitute the weighting mechanism that determines the relative 'value' of each individual input to total input.

[19] The translog is discussed in more detail in section 3.5.1.

The use of cost shares for input aggregation is appealing since it is internally consistent with the methodology used for output aggregation, but also creates a lot of theoretical ambiguities, such as:

– What are the implications of using costs as the basis for creating the input shares in terms of economic theory?

– How should the various price-weighted inputs be aggregated? Can the analysis just add them together?

The seminal work of Tinbergen (1942) and Solow (1957) provided the needed structural framework that is firmly based on economic theory to explain and justify this type of input aggregation. Their work formed the basis of what is now commonly known as the Growth Accounting approach for productivity measurement.

Both Solow's and Tinbergen's work builds upon the economic theory of production, where output is expressed as a function of all the inputs used in the production process. As a reminder, a general production function is given by:

$$Y_t = f(X_{i,t}) \qquad \text{Eq. 3.2.8}$$

, where Y_t is (aggregate) output of a production unit *i* in time period *t* which is produced using a vector of inputs $X_{i,t}$ within the confines of the production technology described by $f(\cdot)$. When early researchers attempted to explain the total growth in aggregate production using data on input growth (namely labour and capital growth), they found that input growth could explain only a relatively small portion of the observed output growth (see Abramovitz (1956)). This prompted the inclusion of an additional element into the production function, whose sole purpose was to account for differences in performance across time, what Abramovitz (1956) called 'a measure of our ignorance'. The inclusion of this element changes Eq. 3.2.8 to

$$Y_t = A_{i,t} f(X_{i,t}) \qquad \text{Eq. 3.2.9}$$

In economic terms $A_{i,t}$ captures the impact of changes in the technology that are not accounted for by changes in the volumes of inputs used in the production. As it will be discussed later in this chapter, these changes can be due to a number of factors; however, early researchers attributed these changes to technological progress and thus called $A_{i,t}$ an estimate of technological (technical) change. Rearranging the equation 3.2.9 yields:

$$A_{i,t} = \frac{Y_t}{f(X_{i,t})} \qquad \text{Eq. 3.2.10}$$

Equation 3.2.10 demonstrates that $A_{i,t}$ is the ratio of aggregate output to a function of inputs used in the production of said output.

As mentioned above, Solow's seminal contribution was to provide the link between the previously developed index number approaches and economic theory to successfully describe and parameterise equation 3.2.9. Solow found that if certain assumptions hold, the index number approach for input aggregation (ie based on the share of the costs of each input to total costs) can be supported by economic theory. Even more so, index number theory provides an analytical framework that requires no estimation and as such produces a completely accurate measure of productivity, assuming that none of the initial assumptions are violated. These assumptions, often called 'standard' or neoclassical assumptions are (OECD (2001)):

– There exists a production technology that can be represented by a production function, relating gross output (*Y*), to primary inputs labour (*L*) and capital services (*K*) as well as intermediate inputs such as material, services or energy (*M*).

$$Y_t = F(K_{i,t}, L_{i,t}, M_{i,t}) \qquad \text{Eq. 3.2.11}$$

– The production function exhibits constant returns to scale.

– The production function is stable over time, in the sense that no additional categories of input are used in output generation and the functional form used to characterise the production function remains the same over the period of the analysis.

– Productivity changes are Hicks-neutral type, i.e. they correspond to an outward shift of the production function, captured by a parameter *A*:

$$Y_t = F(K_{i,t}, L_{i,t}, M_{i,t})A_{i,t} \qquad \text{Eq. 3.2.12}$$

- For any desired level of output, the firm minimises costs of inputs, subject to the production technology shown above. Factor input markets are competitive, so that the firm takes factor prices as given and adjusts quantities of factor inputs to minimise costs.
- The firm and all relevant actors have complete information regarding input prices.
- Labour and intermediate inputs can be hired at any moment at the market rates.
- There are no adjustment costs associated with investment. Alternatively, all adjustment costs are strictly proportional to the volume of investment.

If the above assumptions hold, then according to economic theory, the marginal revenue generated by each input factor is equal to its price.[20] So, for the production function in Eq 3.2.12:

- the marginal revenue of each unit of labour is equal to the wage rate;
- additional capital investment is only undertaken up to the point where its marginal revenue is equal to the (user) cost of capital, and
- intermediate input consumption only takes place up to the point where the marginal revenue it generates is equal to its purchasing cost.

Under such conditions, the output elasticity of each factor has to be equal to its share in the total value of production; if not, then the firm is not optimising outcomes, ie minimising costs or maximising output. This last finding forms the basis for all the calculations required to develop Growth Accounting productivity measures, since it provides the economic justification to use index number aggregation approaches for inputs.

The second and final step in developing the Growth Accounting approach is selecting the most appropriate index number formulae for the production function. As mentioned in

[20] This is because if the firm is a price taker and there is perfect information, each factor of production will demand full remuneration to take part in the production process. So the input factor market will reach equilibrium only when the price of each input is an accurate reflection of its contribution to the production process; ie its price is equal to its marginal revenue

the previous section, Diewert (1976) proved that each index number formulae can be seen as an approximation of a number of different production functions. As such, the choice of an index number formula is ultimately equivalent to an ex-ante assumption about the underlying shape of the production function; in other words, the analysis needs to adopt a functional form for the production function and then utilise the index number formula that provides the best approximation to that functional form.

As of yet, there is no concrete method to analytically determine the underlying form of the production function, at least without implementing some form of econometric analysis that utilises information from a large number of different production units. As such, the choice of functional form is by necessity part of the list of assumptions necessary for the implementation of Growth Accounting. That is not to say that all possible index number formulae are equally valid. Diewert (1976) demonstrated that there is a family of functional forms that he called 'flexible aggregators'. These functional forms are flexible in a sense that they can provide a second order approximation to a twice differentiable linear homogeneous function[21]. In other words, these 'flexible aggregators' can be made to fit a wide variety of possible production functions[22]. For this reason, flexible functional forms such as the quadratic functional form and especially the translog[23] are some of the most common functional forms adopted for a wide range of econometric analyses, including analysis of production functions. Due to these reasons, Growth Accounting has widely adopted the use of the Törnqvist index for both output and input aggregation, since the Törnqvist index provides an 'exact' representation of the translog, assuming that the neoclassical assumptions hold.

To summarise, the adoption of neoclassical assumptions provides the needed economic justification for the use of index number formulae in both input and output aggregation. Given that, when the production function in Eq 3.2.12 is differentiated with respect to time, the rate of growth in output is equal to the weighted average of the growth in inputs and the growth in productivity. The input weights are the output elasticities of each factor of production and although these elasticities cannot be directly observed, it can be proven that, under neoclassical assumptions, the output elasticity of each factor is equal

[21] If a function f is (positively) linear homogeneous, then $f(\lambda x)= \lambda f(x)$ for all $x>>0$ and $\lambda>0$.

[22] 'Flexible aggregators' or more commonly, flexible functional forms are widely utilised in the study of not only production functions, but also utility, cost, revenue, profit and more generally, distance functions.

[23] The translog functional form is discussed in more detail in section 6 of this chapter.

to its share in the total value of production. As such, the change in productivity $A_{i,t}$ can be analytically derived by:

$$\frac{d \ln A_{i,t}}{dt} = \frac{d \ln Y_t}{dt} - S_i^L \frac{d \ln L_{i,t}}{dt} - S_i^K \frac{d \ln K_{i,t}}{dt} - S_i^M \frac{d \ln M_{i,t}}{dt} \qquad \text{Eq 3.2.13}$$

, S_i^L is the average over *t and t-1* share of labour L_i , S_i^K is the average share of capital services K_i *and* S_i^M is the average share of materials M_i. The average shares of the various inputs are given by:

$$S_i^X = \left(\frac{c_{it}^X X_{it}}{p_{it} Y_{it}} + \frac{c_{it-1}^X X_{it-1}}{p_{it-1} Y_{it-1}} \right) \Big/ 2 \qquad \text{Eq 3.2.14}$$

, where c_{it}^X is the price of input X utilised by unit *i* at period *t* and p_{it} is the price of output Y produced by unit *i* at period *t*. The use of inter-period averages conforms to the Törnqvist index number formula, as per equation 3.2.7.

In the KLEMS productivity setting, materials are separated into three components, namely energy (E), services (S) and other materials (M), which are included as separate factors in the production function. This can provide more depth in the analysis, assuming that data are of sufficient quality to accurately undertake this disaggregation.

3.3. Frontier-based approaches for the measurement of productivity

The rest of this chapter focuses on the use of frontier-based approaches for measuring productivity. Frontier-based approaches have been used extensively in the measurement of efficiency and productivity in the micro setting; for a list of applications, see Fried et al. (2008). The same approaches can also measure aggregate productivity growth, and their use can in fact provide a number of advantages to the analysis, which will be discussed in the relevant sections. The discussion provided here will focus on the two most common frontier-based approaches, namely non-parametric DEA-based productivity measurement and parametric, COLS- and SFA-based, productivity measurement.

Similar to Growth Accounting, productivity is also defined here as the ratio of outputs to inputs. As, such the central issue with the frontier-based measures of productivity change is the same as with GA and index number approaches; namely, how to derive a meaningful, accurate and consistent measure of aggregate input and output. GA accomplishes that by postulating a production function and assuming that certain conditions (the so-called neoclassical assumptions) hold; in this case, the production function can be parameterised solely on the basis of economic theory and information on input and output prices. Frontier approaches also postulate a link between inputs and outputs, which is commonly referred to as the production technology (often referred to simply as technology). The notion of technology is similar to the production function, but more general; its formal definition is the set of feasible outputs that can be produced by a combination of inputs. Note that when the transformation process results in just a single output[24], the realisation of technology can be represented as a production function.

The curve (or hull) that envelopes the production technology is referred to as the frontier of the technology, or simply the frontier, and is central to measurement of both efficiency and productivity under all frontier-based approaches. The notion of the frontier is important, because it reveals the maximum amount of output that can be produced by a set of inputs, or the minimum amount of inputs needed to produce a set of outputs. As will be discussed later, the frontier also provides reference points that can be used to measure productivity change over time.

Frontier-based approaches can utilise a number of methods to measure the distance of any observed input/output combination to the frontier. This distance provides an estimate of productive (or technical) efficiency (for a discussion on the notion of efficiency, see section 2.3.2). Most commonly, this distance is measured through the use of distance functions, for non-parametric approaches, or estimated production frontiers, for parametric approaches[25].

The use of distance functions or estimated production frontiers offers a number of advantages for the measurement of productivity. Firstly, they provide a natural

[24] Or there is a simple and effective method for aggregating multiple outputs into a single output measure.

[25] Parametric approaches can also utilise cost, revenue or, more rarely profit functions to derive the characteristics of the technology and thus estimate the technology frontier. This can be achieved through the application of duality theory (Shephard (1953)).

aggregation method for both inputs and outputs without the need for information on prices. Secondly, they do not assume that all producers are perfect optimisers, in the sense that all producers produce the maximum amount of output given set inputs. In other words, these approaches allow for production that is inefficient; this flexibility allows the relaxation of the majority of the neoclassical assumptions that are required for the use of GA.

To summarise, the notion of the frontier is central for the measurement of productivity under these approaches and as such, the critical issue is how each approach derives the frontier. This will be explored further in the following sections.

3.4. Non-parametric measures of productivity change

This section provides a brief discussion on some of the concepts and methods that can be applied to measure productivity in the non-parametric setting. Specifically, the focus is on the use of economic aggregators derived directly through the use of non-parametric techniques. These aggregators are firmly based on economic theory and provide a solid theoretical relationship between inputs and outputs which is based on the concept of the production technology, without the need to rely on information about input and output prices. For a more detailed discussion on the measurement of productivity by non-parametric approaches, please refer to Färe et al. (2008).

Non-parametric approaches define the frontier as the outer boundary of the technology; before discussing the notion of the frontier in more detail, it would be useful define the notion of technology. Technology, or the production possibility set $T \subset R$ [26]is mathematically described as the set that contains all possible production possibilities, ie

$$T \equiv \{(X,Y) | Y \geq 0 \text{ can be produced by } X \geq 0\}$$

Additionally, technology is assumed to display some specific properties[27]:

– Production is possible and currently happening (the technology set is non-empty) - $T \neq \varnothing$

[26] R here denotes a Euclidian vector space with dimensions equal to the sum of inputs and outputs used in the characterisation of the technology.
[27] Also, see Kuosmanen (2003).

– Inactivity of a single producer is possible - $T \cap R = \{0\}$

– No output can be produced if no inputs are utilised (no free lunch)

– Technology can be represented as a closed set (ie the technology can characterise all points that belong to its boundary)

– Inputs are scarce and costly to acquire - $\{y \in T | y \geq y'\}$ is a bounded set for every $y' \in T$

Its very common to adopt the assumptions above when describing the concept of technology, since they conform to the vast majority of production processes; as such they are sometimes referred to as the 'maintained axioms' of production technology (Färe and Primont (1995)).

The frontier is most commonly defined as the convex (or conical), monotonic hull that envelops the technology. Monotonicity (non-decreasing in inputs) means that when inputs are increased, outputs are also increased or stay constant; it is mainly applied so that the frontier is drawn in such a way that an increase in inputs cannot result in a decrease in outputs. As with the 'maintained axioms' above, the monotonicity assumption is very common in economic theory and fairly benign[28]. Convexity (or concavity, when the analysis considers input/output correspondences) in inputs implies that the rate of change of output relative to inputs (ie the marginal product) should be non-increasing. Convexity is needed to enforce the law of diminishing returns and to simulate 'rational' (in economic theory terms) producer behaviour. The convexity assumption, although common in economic theory, can be relatively controversial.

According to Kuosmanen (2003), convexity is sometimes justified by practitioners since it can be derived from a set of elementary and intuitive axioms, like additivity and divisibility of inputs and outputs. However, there are some industries where divisibility might not be possible; there are groups of companies, such as regulated utility companies (eg water companies), that cannot scale back or intensify their operations due to legal and/or natural constraints. There are other industries where economies of

[28] There are situations when the assumption of monotonicity is not appropriate, such as when input congestion is a possibility or the production process generates undesirable outputs; however, these situations are the exception rather than the rule in the realm of economic activities.

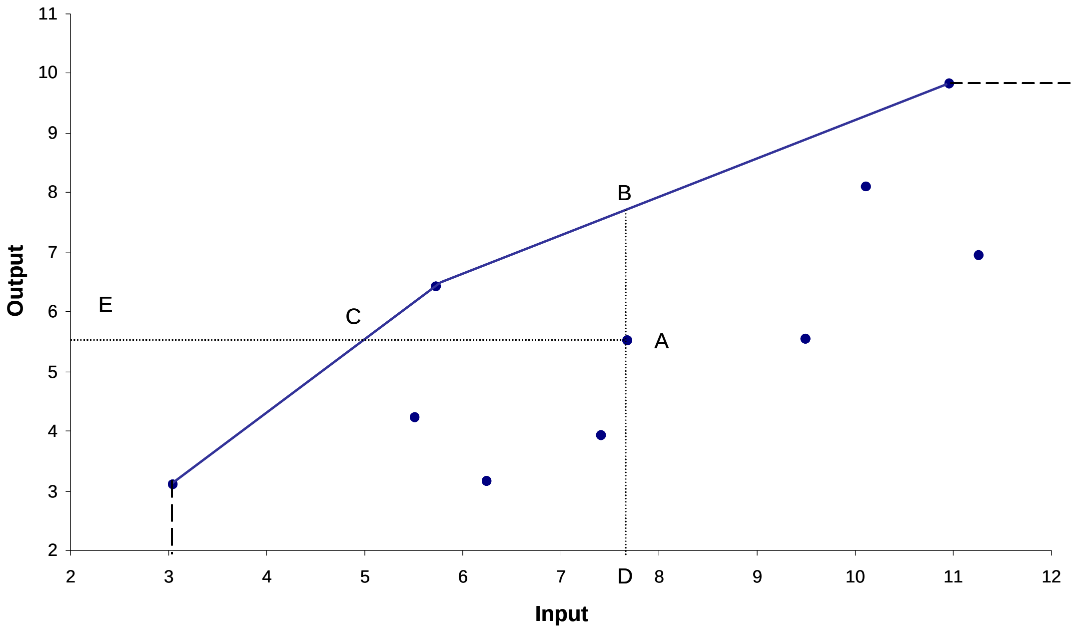
scope (or economies of specialisation) can have a significant impact in the production process. In both instances, the assumption of convexity could be violated. Due to this, a number of approaches that do not require the convexity assumption have been proposed (such as Free Disposal Hull)[29]. Even so, the convexity assumption currently remains central to the most common and most easily applied non-parametric approaches for measuring efficiency and productivity; as such, the approaches discussed on this chapter and employed throughout this thesis will assume that the technology is convex.

3.4.1. Data Envelopment Analysis

Given the above, the frontier can be derived from the observed technology by drawing a monotonic, convex hull over the available observations. This can be achieved using Data Envelopment Analysis (DEA), a non-parametric linear programming technique. Figure 3.1 provides a graphical illustration of the DEA frontier for a convex technology utilising a single input to produce a single output.

[29] For a current review on the possible approaches available that do not rely on the assumption of strict convexity, see Emrouznejad and Amin (2009).

Figure 3.1: Example of a DEA frontier with convex technology



The full lines in the figure above represent the DEA frontier, ie the maximum amount of output that can be achieved at different input levels. AB and AC measure the distance of unit A from the frontier and are both measures of units' A efficiency. More specifically, AB represents units' A output-orientated efficiency, as it measures the additional output that would be required for A to reach the frontier, while keeping its input constant. Similarly, AC represents units' A input-orientated efficiency as it measures the reduction in input that would be required for A to reach the frontier, while keeping its output constant. The distance functions are normally represented in ratio form, using distances from the axis relative to the observed input/output correspondences and to the projected-to the frontier, efficient input/output correspondences. So in the above example, the distance function that measures efficiency is:

$$D_o(x,y) = \frac{DA}{DB} \qquad \text{Eq 3.4.1}$$

for the output-oriented measure, and

$$D_i(x,y) = \frac{EC}{EA} \qquad \text{Eq 3.4.2}$$

for the input-oriented measure.

For multiple output- multiple input technologies, the input and output distance functions can be measured by solving their respective DEA models, ie either the input- or output-oriented DEA model. The general form of the DEA input oriented model that can accommodate multiple inputs and multiple outputs is given below:

$$
\begin{array}{lll}
\min \,{}^{\text{¯}}{}_{j_0} & & \\
\sum_{\mathrm{j=1}}^{N} \mathrm{a}_{\mathrm{j}}\, x_{ij} \leq {}^{\text{¯}}{}_{j_0}\, x_{ij_0} & (i=1,2,...,m) & \\
\sum_{\mathrm{j=1}}^{N} \mathrm{a}_{\mathrm{j}}\, y_{rj} \geq y_{j_0} & (r=1,2,...,s) & \text{Model 3.1} \\
\sum_{\mathrm{j=1}}^{N} \mathrm{a}_{\mathrm{j}} = 1 & (j=1,2,...,n) & \\
\mathrm{a}_{\mathrm{j}} \geq 0 & &
\end{array}
$$

, where ${}^{\text{¯}}{}_{j_0}$ is the input efficiency of unit j_0 that utilises m inputs to produce s outputs, x_{ij} is the observed level of input i of unit j, y_{rj} is the observed level of output r of unit j and a_{j} is the interpolation multiplier applied to unit j. Similarly, the general form of the DEA output-oriented model is:

$$
\begin{array}{lll}
\max \,\text{¤}_{j_0} & & \\
\sum_{\mathrm{j=1}}^{N} \mathrm{a}_{\mathrm{j}}\, x_{ij} \leq x_{ij_0} & (i=1,2,...,m) & \\
\sum_{\mathrm{j=1}}^{N} \mathrm{a}_{\mathrm{j}}\, y_{rj} \geq \text{¤}_{j_0}\, y_{rj_0} & (r=1,2,...,s) & \text{Model 3.2} \\
\sum_{\mathrm{j=1}}^{N} \mathrm{a}_{\mathrm{j}} = 1 & (j=1,2,...,n) & \\
\mathrm{a}_{\mathrm{j}} \geq 0 & &
\end{array}
$$

, where ¤_{j_0} is the output inefficiency of unit j_0.[30] The two models presented above were first developed by Banker, Charnes, and Cooper (1984) and provide the representation of

[30] By construction, $\text{¤}_{j_0} \geq 1$. A measure of output efficiency is the inverse of ¤_{j_0}.

a convex technology; in other words, they assume variable returns to scale (VRS) for the production technology. These VRS DEA models provide a closer envelopment of the data than the original DEA models developed by Charnes, Cooper and Rhodes (1978), which provide the representation of a conical technology, ie a technology that assumes constant returns to scale. The CRS models are exactly the same as the VRS models, except they omit the convexity constraint, ie $\sum_{j=1}^{N} a_j = 1$.

The notion of the frontier provides a natural mechanism for aggregating both inputs and outputs. This can be easily demonstrated by the dual of the DEA envelopment (primal) model, which is usually referred to as the value-based, or shadow-price DEA model. The general form of the CRS value-based DEA model is given below:

$$\max \frac{\sum_{r=1}^{s} u_r y_{rj_0}}{\sum_{i=1}^{m} v_i x_{ij_0}}$$

$s.t.$

Model 3.3

$$\frac{\sum_{r=1}^{s} u_r y_{rj}}{\sum_{i=1}^{m} v_i x_{ij}} \leq 1 \qquad (j = 1, 2, \ldots, n)$$

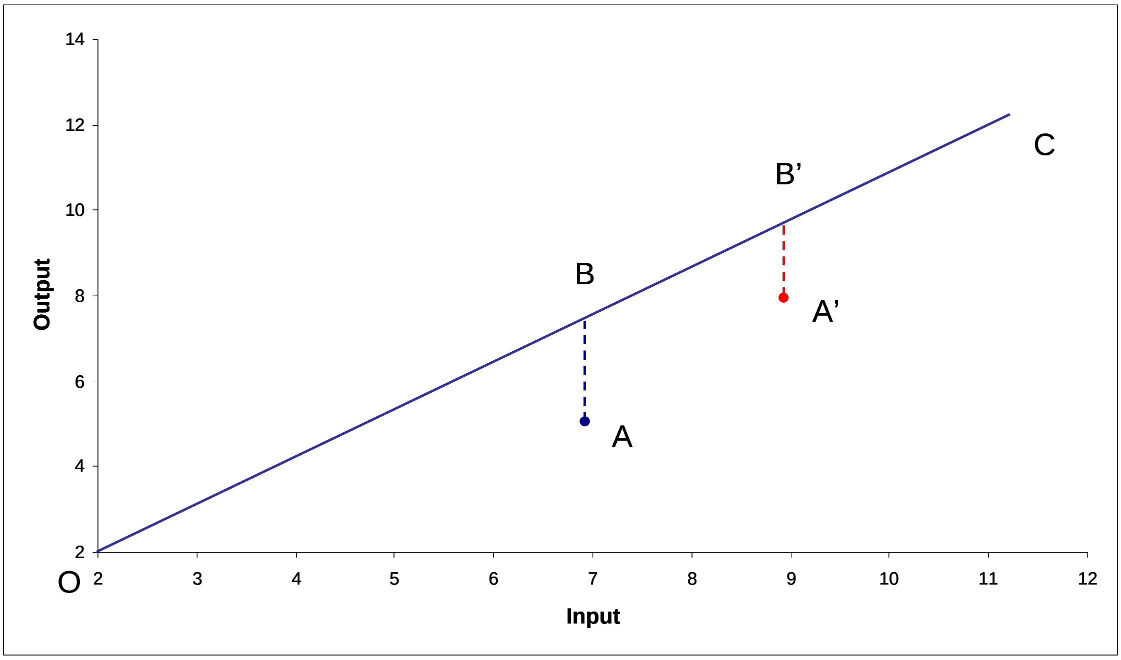
$$u_r, v_i \geq 0 \qquad (r = 1, 2, \ldots, s;\ i = 1, 2, \ldots, m)$$

The objective function in the value-based DEA model is a natural measure of productivity, since it maximises the ratio of aggregate outputs to aggregate inputs. Aggregation is achieved by allowing the assessed unit to select weights for its inputs and outputs in such a way as to maximise the ratio of aggregate outputs to aggregate inputs, with the restriction that when the selected weights are used by another unit, its output-to-input ratio cannot exceed a predetermined upper bound (this is normally set to 1, as per model 3.3). In other words, aggregation is done with reference to the technology frontier; the assessed unit will adopt the weights that apply to the section of the frontier is projected to (either point C or point B in the example presented in figure 3.1).

3.4.2. Malmquist indices

A simple one input/one output example of how productivity change can be measured using DEA is given in the figure below.

Figure 3.2: Productivity measurement using a static conical (CRS) frontier



In the example above, OC is the static CRS DEA frontier and A is the input/output correspondence of the assessed unit in time 0, while A' is the input/output correspondence of the assessed unit in time 1. The output distances AB and A'B' can be used as a natural way to measure productivity change[31]; in time 1, the assessed unit produces more output but also utilises more input. However, since it is closer to the frontier in time 1, its productivity has improved.

In their seminal work, Caves, Christensen and Diewert (1982) used this concept of the output distance function to define the Malmquist productivity index (MI) as:

$$M_o^0 = \frac{D_o^0(x^1, y^1)}{D_o^0(x^0, y^0)} \qquad \text{Eq 3.4.3}$$

[31] This example utilises the output distance function to measure efficiency. However, the same concepts fold for the input distance function. This is especially true in this example, since the example assumes constant returns to scale. Under CRS, the output distance function will be equal to the input distance function for all assessed units.

, where $D_o^0(x^1, y^1)$ is the output distance function of the (x^1, y^1) input/output correspondence relative to the frontier in time 0 and $D_o^0(x^0, y^0)$ is the output distance function of the (x^0, y^0) input/output correspondence, also relative to the frontier in time 0, since this example assumes that the frontier is static. The more formal definition of the output distance function as used in the context of the Malmquist index is that it is the reciprocal of the maximum proportional expansion of output given inputs. As such, it always holds that $D_o^t(x^t, y^t) \leq 1$ and $D_o^t(x^t, y^t) = 1$ only if (x^t, y^t) is on the technology frontier.

The above formulation for the Malmquist productivity index is also valid for input distance functions; in fact, if the technology displays constant returns to scale, the output-oriented MI will be equal to the input oriented MI, since when the frontier is a conical hull (CRS technology), the input-oriented and output-oriented distances to the frontier are the same.

If the frontier is allowed to move, a similar Malmquist index can be defined relative to period 1 technology:

$$M_o^1 = \frac{D_o^1(x^1, y^1)}{D_o^1(x^0, y^0)} \quad \text{Eq 3.4.4}$$

, where the output distance functions $D_o^1(x^1, y^1)$ and $D_o^1(x^0, y^0)$ are relative to the frontier in time 1.

Färe et al. (2008) have shown that generally, M_o^1 and M_o^0 will result in different productivity change estimates since the technologies that define the frontier in the two periods differ. The only time when M_o^1 and M_o^0 will result in the same productivity change estimates is when the overall technology is Hicks-output neutral, or in other words, the technology in all periods is the outward shift of the technology of the base period (period 0) by a function A(t), where t denotes time.

These likely differences between M_o^1 and M_o^0 create a complication for the analysis; either the analysis assumes that the technology is Hicks-output neutral, or it has to make an a-priori choice between using period 0 or period 1 technology as the reference frontier.

In their seminal paper, Färe et al. (1992), proposed a different formulation of the Malmquist index that could remove this complication. Instead of choosing blindly between M_o^1 and M_o^0, they reformulated the Malmquist productivity index to be the geometric average of M_o^1 and M_o^0, similar to the way the Fisher index (see equation 3.2.6) is defined as the geometric average of the Paasche and Laspeyres indices:

$$M_o = (M_o^0 M_o^1)^{1/2}$$

$$M_o = \left(\frac{D_o^0(x^1,y^1)}{D_o^0(x^0,y^0)} \frac{D_o^1(x^1,y^1)}{D_o^1(x^0,y^0)} \right)^{1/2} \qquad \text{Eq 3.4.5}$$

This formulation of the Malmquist productivity index has been widely adopted in the academic literature and has become the standard index of productivity change, both in the non-parametric but also in the parametric setting, as will be discussed in the following section. One of the major strengths of this formulation is that it can be used to identify the various ‘sources’ of productivity change, in terms of the movement of the assessed unit within the production possibility space defined by the frontier and the movement of the frontier itself.

Färe, Grosskopf, Norris and Zhang (1994) – referred as FGNZ from now on – provide one of the most well known decompositions of the MI:

$$\begin{aligned} M_o &= \left(\frac{D_o^0(x^1,y^1)}{D_o^0(x^0,y^0)} \frac{D_o^1(x^1,y^1)}{D_o^1(x^0,y^0)} \right)^{1/2} \\ &= \frac{D_o^1(x^1,y^1)}{D_o^0(x^0,y^0)} \left(\frac{D_o^0(x^1,y^1)}{D_o^1(x^1,y^1)} \frac{D_o^0(x^0,y^0)}{D_o^1(x^0,y^0)} \right)^{1/2} \\ &= EC \times TC \end{aligned} \qquad \text{Eq 3.4.6}$$

, where efficiency change (also referred to as catch-up), $EC = \frac{D_o^1(x^1,y^1)}{D_o^0(x^0,y^0)}$, captures the changes in the efficiency of the assessed unit from period 0 to period 1 and technological change (also referred to as frontier shift) $TC = \left(\frac{D_o^0(x^1, y^1)}{D_o^1(x^1, y^1)} \frac{D_o^0(x^0, y^0)}{D_o^1(x^0, y^0)} \right)^{1/2}$,

captures the movement of the frontier itself, from period 0 to period 1.

As with all index numbers, a MI greater than 1 indicates that productivity has improved between periods 0 and 1, and this could be due to the assessed unit becoming more efficient (ie moving closer to the frontier) and/or the frontier moving outward, which would signify that the technology has improved, in a sense that the most efficient units can produce higher aggregate output in period 1 relative to period 0, using the same aggregate inputs in both periods[32].

Another interesting feature of the Malmquist productivity index is that it is consistent with the neoclassical measure of productivity $A_{i,t}$ as described by Abramovitz (1956) and Solow (1957) (see equation 3.2.9, section 3.2.1). Färe et al. (2008) demonstrates that for production functions with one output, constant returns to scale, and Hicks-neutral productivity[33], the MI will result in the same productivity estimate as $A_{i,t}$, assuming that the efficiency of the assessed unit remains exactly the same between the periods of the assessment (ie *EC=1*).

The discussion so far utilises the conical, monotonic frontier usually associated with a technology that displays constant returns to scale. FGNZ (1994) refer to distance functions estimated relative to the conical frontier as ‘benchmark’ distances, presumably since they represent what is economically optimal. However, technologies that display variable returns to scale are not uncommon, especially in situations where the scale of the production is not under the direct control of the assessed unit. In situations such as these, the use of ‘best practise’ distances, ie distances estimated relative to the convex frontier, are preferable when measuring efficiency. Productivity however can still be measured relative to the ‘benchmark’ (conical) frontier, even when the actual technology is VRS. Based on this property, FGNZ (1994) proposed an alternative decomposition of the Malmquist index that also identifies the changes in productivity resulting from the

---

[32] This holds for the output-oriented MI; for the input oriented case, the most efficient units can produce the same aggregate output in period 1, while utilising lower volumes of aggregate input in period 1 relative to period 0.

[33] As a reminder, a production function with one output, constant returns to scale, and Hicks-neutral productivity is a subset of the so-called neoclassical assumptions required for the measurement of productivity under Growth Accounting.

assessed unit moving towards a more productive scale size, ie closer to the conical part of the frontier. This decomposition is given by:

$$M_C = \frac{D_V^1(x^1,y^1)}{D_V^0(x^0,y^0)} \frac{S^1(x^1,y^1)}{S^0(x^0,y^0)} \left(\frac{D_C^0(x^1,y^1)}{D_C^1(x^1,y^1)} \frac{D_C^0(x^0,y^0)}{D_C^1(x^0,y^0)} \right)^{1/2} \qquad \text{Eq 3.4.7}$$
$$= EC_V \times SC_{FGNZ} \times TC_C$$

,where $D_V^t(x^t,y^t)$ denotes distance relative to the VRS frontier and $D_C^t(x^t,y^t)$ denotes distance relative to the CRS frontier. SC denotes scale efficiency change and is the ratio of scale efficiency in period 1 relative to period 2:

$$SC_{FGNZ} = \frac{S^1(x^1,y^1)}{S^0(x^0,y^0)} = \frac{D_C^1(x^1,y^1)}{D_V^1(x^1,y^1)} \Bigg/ \frac{D_C^0(x^0,y^0)}{D_V^1(x^0,y^0)} \qquad \text{Eq 3.4.8}$$

Although this decomposition is both theoretically and analytically correct, Ray and Desli (1997) noted that it is inconsistent in that the efficiency change component is measured relative to the VRS frontier while the technological change component measures the movement of the CRS frontier. To amend that, they proposed an alternative decomposition of the MI:

$$M_C = \frac{D_V^1(x^1,y^1)}{D_V^0(x^0,y^0)} \left(\frac{D_V^0(x^1,y^1)}{D_V^1(x^1,y^1)} \frac{D_V^0(x^0,y^0)}{D_V^1(x^0,y^0)} \right)^{1/2} \left(\frac{S^0(x^1,y^1)}{S^0(x^0,y^0)} \frac{S^1(x^1,y^1)}{S^1(x^0,y^0)} \right)^{1/2} \qquad \text{Eq 3.4.9}$$
$$= EC_V \times TC_V \times SC_{RD}$$

, where both the efficiency change and technological change components are measured relative to the VRS frontier and the scale efficiency change component is the geometric mean of the scale efficiency ratio measured relative to period 0 and the scale efficiency ratio measured relative to period 1.

The Ray and Desli (1997) decomposition is theoretically correct and internally consistent, but may not always be feasible in practice. The issue rests with the calculation of the cross-period distance functions required for both the technological change and scale efficiency change components; sometimes, the observed input/output correspondences in one period may not be fully enveloped by the frontier in another period. If this is the case, the DEA models used to measure these distances have no solutions and as such, the decomposition of productivity change for some units may not be possible.

An issue that often arises in practical applications of the Malmquist index is how to deal with unbalanced panel data, ie when the analysis is missing observations for some of the assessed units for some time periods. Some authors state that the calculation of MI requires the availability of a balanced panel and apply a number of adjustments to convert the available unbalanced panel to a balanced one[34]. However, the availability of a balanced panel is not strictly a prerequisite of the MI; unbalanced panel datasets will not lead to infeasibilities similar to those found in the Ray and Desli (1997) decomposition[35]. This does not mean that unbalanced panels are not problematic in this setting, since in order to calculate the MI one requires information for both the base and the subsequent period (ie input/output correspondences are required both for period 0 and 1); if information on one of those periods is missing, then the MI for this period cannot be calculated. More important however are the possible effects of an entry or exit of a unit to the MI estimates of all the other assessed units for that period. For example, if the unit that exits was used to define the frontier in period 0 and there are no other units with similar input/output mix close to the frontier at the moment of exit, the resulting frontier in the next period can be very different to what was previously. This in turn can cause significant volatility in the MI estimates of all other units that were projected to this particular segment of the frontier.

Another potential shortcoming of the 'traditional' Malmquist index formulation is that the index is non-circular. In general, circularity is a desirable property of index numbers; a circular index *I* measured over three consecutive periods (t_1, t_3, t_3) is one that satisfies the following:

$$I(t_1,t_3) = I(t_1,t_2)I(t_2,t_3) \qquad \text{Eq 3.4.10}$$

For the 'traditional' Malmquist index formulations, this implies that the analysis may not derive the change in productivity between periods 1 and 3 even when the change in productivity between years 1 and 2 and between years 2 and 3 is known. Färe and Grosskopf (1996) demonstrated that the only time the Malmquist index will be circular is

[34] See for example Hollingsworth and Wildman (2003).

[35] FGNZ (1994) also state that explicitly in footnote 14 of their paper.

when the overall technology is Hicks-output neutral. They also argued that circularity is not necessarily an important property of productivity indices; they site Fisher, whose productivity index is also not circular, who argues that ‘there is a natural order of time that makes productivity change inherently path-dependant’[36], and thus non-circular.

3.4.3. A circular Malmquist index

As mentioned above, there are two main issues with the ‘traditional’ Malmquist index:

- The first is the ambiguous treatment with regards to the definition, measurement and decomposition of the scale efficiency component. If the Ray and Desli (1997) decomposition is adopted, then there is the danger that, at least for some of the assessed units, the DEA models used to estimate productivity change can be unsolvable.

- Secondly, ‘traditional’ Malmquist indices might result in significant volatility of the resulting productivity estimates when using unbalanced panel data. This has serious implications in studies that utilise unbalanced panel datasets, such as the EU KLEMS dataset, since it necessitates the exclusion of some of the available information in order to balance the dataset.

These issues can be resolved by adopting a circular Malmquist index formulation, similar to the formulation proposed by Pastor and Lovell (2005) and refined by Portela and Thanassoulis (2010).

The circular Malmquist index described here is based on the observation that, a distance function can be measured indirectly, by comparing the multidimensional points of the two periods relative to a common reference point, or in this case, to a common frontier. This is a departure from the ‘traditional’ Malmquist index, which defines the distance functions with reference to two frontiers, each based on the start and end periods of the analysis.

This common frontier can be defined as the ‘meta-frontier’, which envelopes all data points from all periods; this also allows for the formulation of a Malmquist-type index that is circular. To draw this ‘meta-frontier’, one must assume that convexity holds for all data points across different time-periods. This actually translates to the assumption that what

[36] See Färe et al. (2008) op cit., p.551

was technologically feasible in a given time period will always be feasible in any future time period. This assumption can be somewhat restrictive, especially in cases where changes in the legal or regulatory environment make restrict the adoption of production plans that were available in the past (for example emission restrictions in the automotive industry). However, the severity of this restriction is debatable, since it is not necessary for the meta-frontier to include of all current (at the time of the analysis) production possibility sets. In fact, the 'traditional' Malmquist index is also based on two different frontiers and one of them may include 'non-feasible' production possibility sets.

Using the notion of the meta-frontier, a unit's efficiency in time *t* relative to the meta-frontier (referred to as the meta-efficiency) can be written as:

$$-^{m}_{it} = -^{T}_{it} TG_{it} \qquad \text{Eq 3.4.11}$$

Where, $-^{m}_{it}$ is unit's *i* meta-efficiency in time *t*=*T*, $-^{T}_{it}$ is unit's *i* cross-sectional efficiency in time *t*=*T* (ie relative to the frontier in time *T*) and TG_{it} is the technological gap between the frontier in time *t*=*T* and the meta-frontier. $-^{m}_{it}$ is straightforward to estimate by solving a DEA model that includes all observations from all the assessed units in all available periods. TG_{it} is residually estimated as:

$$TG_{it} = \frac{-^{m}_{it}}{-^{T}_{it}} \qquad \text{Eq 3.4.12}$$

Since the meta-frontier is fixed in the timeframe selected for the analysis, the productivity change of a unit between any two time periods can be measured using the ratio of the meta-efficiencies of these two periods:

$$CM^{i}_{t,t+1} = \frac{-^{m}_{it+1}}{-^{m}_{it}} \qquad \text{Eq 3.4.13}$$

Using equation 3.4.13, the circular Malmquist index can be decomposed such that:

$$CM^{i}_{t,t+1} = \frac{-^{T}_{it+1}}{-^{T}_{it}} \frac{TG_{it+1}}{TG_{it}} \qquad \text{Eq 3.4.14}$$

The first term in the right hand side of equation 3.4.14 captures the efficiency change of unit *i* from period *t* to period *t+1*, similar to the 'traditional' Malmquist index. The second term captures the movement of the frontier where unit *i* is projected to, between period *t* and period *t+1*; in other words, the second term represents an estimate of frontier shift or technological change between periods t and t+1.

The way frontier shift is estimated is the main point of departure between the 'traditional' Malmquist index and the circular Malmquist index. In the 'traditional' Malmquist index, frontier shift is defined as the geometric mean of the distance of the frontiers in periods *t* and *t+1*, measured at the point where the unit with a particular input-output mix is projected to in periods *t* and *t+1*. The circular Malmquist index is defined as the ratio of the distance of the frontier in t+1 from the meta-frontier to the distance of the period t frontier from that same meta-frontier; although the assessed unit has the same input-output mix as in the 'traditional' Malmquist setting, since it projects itself to the meta-frontier, rather than the period *t* or *t+1* frontiers, the underlying distance functions of the 'traditional' Malmquist index and the circular Malmquist index may well differ. In fact, the only instance where the frontier shift estimates from 'traditional' and circular Malmquist indices are the same is when the input-output mix of the assessed unit stays constant in the two periods of assessment. This usually has a marginal impact on assessments involving the measurement of aggregate productivity growth, especially when the assessment is done in annual intervals; this is because whole industries or economies generally display relative stable input-output mixes in the short-term.

The circular Malmquist index in equation 3.4.14 can be further decomposed if the assumption of constant returns to scale is relaxed. As was mentioned before, there are various decompositions of the 'traditional' Malmquist index with respect to the scale component proposed in the literature. Each of those offers a different treatment and definition of the scale component and all could be considered complimentary. When examining aggregate productivity change, the most common objective is to derive an estimate of efficiency and technological change, free of bias from the effects of scale size, since the overall scale of the economy is depended on a number of factors outside of the control of any policy maker. If this is indeed the underlying objective, the decomposition of the Malmquist index should probably be attempted using a VRS frontier as the base, similar to the approach suggested by Ray and Desli (1997), but adapted for the circular Malmquist index.

Portela and Thanassoulis (2010) suggest the following decomposition of the circular Malmquist index:

$$CM^{i}_{t,t+1} = \frac{-^{T(VRS)}_{it+1}}{-^{T(VRS)}_{it}} \frac{TGV_{it+1}}{TGV_{it}} \frac{MSE_{it+1}}{MSE_{it}} \qquad \text{Eq 3.4.15}$$

, where $-^{T(VRS)}_{it}$ is unit's *i* cross-sectional VRS efficiency in time *t* (ie relative to the frontier in time *t*), TGV_{it} is the technological gap between the VRS frontier in time t and the VRS meta-frontier and MSE_{it} is unit's *i* meta-scale efficiency in time *t.*

The first component of equation 3.4.15 represents the VRS efficiency change of unit *i* from period *t* to period *t+1*, the second component is the frontier shift of unit *i* from period *t* to period *t+1* and the last component is the change in scale efficiency of unit *i* from period *t* to period *t+1*. The VRS technological gap and meta-scale efficiency scores are easy to calculate, since:

$$TGV_{it} = \frac{-^{m(VRS)}_{it}}{-^{T(VRS)}_{it}} \qquad \text{Eq 3.4.16}$$

, and

$$MSE_{it} = \frac{-^{m(CRS)}_{it}}{-^{T(VRS)}_{it}} \qquad \text{Eq 3.4.17}$$

As is apparent from equations 3.4.15 and 3.4.16, the decomposition of the circular Malmquist index is internally consistent in the manner of the Ray and Desli (1997) decomposition, since both the efficiency and technological change components use the VRS frontier as reference. This is due to the fact that the circular Malmquist index does not require the calculation of cross-period efficiencies, which are the source of the potential DEA infeasibilities in the 'traditional' Malmquist index formulations. The utilisation of the meta-frontier as the frontier of reference also allows the analysis of unbalanced panels, without the fear of possible volatility in the resulting productivity change estimates, since the meta-frontier is fixed. The use of the meta-frontier also results in an index that satisfies the circularity property, as expressed in equation 3.4.10.

3.5. Econometric approaches to productivity measurement

This section provides a brief discussion on two of the most common econometric/parametric approaches used for the measurement of productivity change, namely Corrected Ordinary Least Squares (COLS) and Stochastic Frontier Analysis (SFA). These approaches are firmly rooted in economic theory and, as is the case with the non-parametric distance functions discussed in the previous section, can produce productivity change estimates without requiring information on input and output prices. For a more detailed discussion on the measurement of productivity utilising econometric approaches, see Kumbhakar and Lovell (2000).

The econometric approaches discussed in this section belong in the more general family of frontier-based approaches. Similar to the non-parametric distance functions, they measure productivity change with a reference to the frontier, ie the outer boundary of the technology. Also similar to the non-parametric distance functions, they also assume that production is possible inside the frontier, ie not all of the assessed units need to be technically efficient. In addition, the frontier can display a wide range of return to scale properties and as such, the analysis is not restricted by assuming that the production process displays solely constant returns to scale. However, probably the most important feature of some of these approaches, namely SFA, is that they can take into account the stochastic nature of the production process.

The utilisation of the notion of the frontier, which can display the whole range of returns to scale properties, allows for the estimation of a productivity change measure that can be decomposed into different elements similar to those examined when discussing the non-parametric Malmquist index. As such, the econometric approaches described in this section can measure productivity change and also decompose it into efficiency change, technological change and scale efficiency change. In essence, the econometric approaches produce a Malmquist-type productivity index, although the methodology used to derive this index is quite different to the non-parametric distance functions discussed in the previous section.

3.5.1. General theory and decomposition

At the heart of the econometric approaches for measuring efficiency and productivity lies the notion of production technology and its frontier. The main difference with the non-parametric distance functions is that the production technology is usually represented by a production function, or, as is the case in the approaches examined here, by a production frontier. The reason for this restriction is that the econometric approaches rely on some form of regression analysis for the estimation of the characteristics of the transformation function. This limits the analysis in the sense that only one factor can be considered in the left-hand side of the regression equation. As such, in order to parameterise the production function or production frontier, output must be expressed by a single measure[37].

Although this might seem to be a significant shortcoming, this is not necessarily the case. For many studies in the micro setting, information on costs and quantity of outputs is easier to secure than information on input and output quantities; in such cases, econometric approaches can take advantage of the duality theory (Samuelson (1947); Shephard (1953)) and estimate the cost frontier, rather than a production frontier. Sometimes, sufficient information is even available for the estimation of the profit frontier. As noted by Kumbhakar and Lovell (2000), the estimation and decomposition of productivity change is well developed within the cost function framework; more recently, Kumbhakar (2002) also provided the analytical framework for measuring and decomposing productivity change within the profit function setting.

In this section, the focus will remain in the primal approaches, ie those that utilise the notion of the production frontier, since all of the approaches reviewed so far for the measurement of aggregate productivity growth were based on the production process. The requirement of a single output measure is not limiting in this setting, given that the National Accounts framework already represents aggregate output as a single measure. As a reminder, output aggregation is achieved through the use of index number approaches and is relatively relaxed in terms of required assumptions (see section 3.2.1 for more details).

[37] This limitation can be alleviated by adopting parametric input- or output-distance functions rather than production functions; however, parametric distance functions face a number of theoretical and estimation issues when used to measure efficiency and productivity. For additional discussion see Kumbhakar and Lovell (2000), op cit.

The starting point of the discussion is the specification of the deterministic production frontier:

$$y_{it} = f(x_{it}^{j};t)TE_{it} \qquad \text{Eq 3.5.1}$$

,where y_{it} is the output of unit i in time t, x_{it}^{j} is the vector of *j* inputs of unit *i* in time *t*, *t* is a time trend and TE_{it} is the measure of technical efficiency of unit *i* in time *t* . Equation 3.5.1 characterises the deterministic production frontier because all of the elements that influence output are deterministic in nature. To specify the stochastic production frontier, a stochastic element needs to be included in equation 3.5.1:

$$y_{it} = f(x_{it}^{j};t)TE_{it}e^{v_{it}} \qquad \text{Eq 3.5.2}$$

, where *v* represents a two-sided error term that captures statistical noise and as such is identically and symmetrically distributed. Stochastic production frontiers will be discussed in more detail in section 3.5.2 below. Overall, the general methodology with regards to the estimation of productivity change is the same, regardless of whether the production frontier is deterministic or not; of course, that is not to say that the productivity change estimates will be the same under both approaches.

Note that the stochastic component, ie the two-sided error term *v*, was included in equation 3.5.2 multiplicatively and as an exponent. The reason for that is that the most common functional forms adopted for the production frontier are linearised by logarithmic transformations of the underlying data. Including the stochastic component as an exponent will also ensure that it will be in linear form after the logarithmic transformation of the production frontier.

The above is also the reason why most econometric approaches also include the technical efficiency element as an exponent:

$$y_{it} = f(x_{it}^{j};t)e^{-u_{it}}e^{v_{it}} \qquad \text{Eq 3.5.3}$$

, where

$$TE_{it} = e^{-u_{it}} \qquad \text{Eq 3.5.4}$$

and $u_{it} \geq 0$ represents the output-oriented inefficiency of unit *i* in time *t*.

Using equation 3.5.3, Kumbhakar and Lovell (2000) construct the productivity change of a unit by identifying and adding together its three main components, namely efficiency change, technological change and scale efficiency change.

The element of technological change, or frontier shift, is captured in the time trend *t*. Technological change, ΔTC_{it} itself can be expressed as the partial derivative of equation 3.5.3 with respect to the time *t*.

$$\Delta TC_{it} = \frac{\partial \ln f(x_{it}^{j};t)}{\partial t} \qquad \text{Eq 3.5.5}$$

Technological change, ΔTC_{it}, can be positive or negative, reflecting upwards or downwards shifts in the production frontier. It can also be zero, when the frontier remains constant over the periods examined. Note that the above formulation does not restrict the technological change component to be output Hicks-neutral.

Similarly, the efficiency change component can also be expressed as the partial derivative of technical inefficiency with respect to the time *t*.

$$\Delta TE_{it} = -\frac{\partial u_{it}}{\partial t} \qquad \text{Eq. 3.5.6}$$

The interpretation of the efficiency change component is exactly the same as that in the non-parametric Malmquist index; it captures the movement of the unit *i* either towards the frontier, when $\Delta TE_{it} > 0$, or away from the frontier, when $\Delta TE_{it} < 0$. If $\Delta TE_{it} = 0$, then the units' efficiency is unchanged between the periods of the analysis.

The scale efficiency change component can be derived from totally differentiating equation 3.5.3, after accounting for the general structure of the production frontier:

$$\Delta SE_{it} = (£ - 1)\sum_{j} \frac{£_{j}}{£} \Delta x_{j} \qquad \text{Eq. 3.5.7}$$

,where $\pounds_j$ are the elasticities of output with respect to each of the *j* inputs, $\pounds = \sum \pounds_j$ and Δx_j is the rate of change of input *j* between the periods of the analysis for the assessed unit.

Kumbhakar and Lovell (2000) also include a fourth component of productivity change that is meant to capture the effects of allocative efficiency change. Allocative efficiency is a measure of a unit's ability to select its input mix in such a way as to minimise costs. This component is given by:

$$\Delta AE_{it} = \sum_j \left(\frac{\pounds_j}{\pounds} - S_j \right) \Delta x_j \qquad \text{Eq. 3.5.8}$$

, where $S_j = \frac{w_j x_j}{\sum w_j x_j}$ is the share of observed expenditure of input x_j with price w_j to total expenditure. To assess changes in allocative efficiency in this setting, the analysis needs to have access to input prices. If input price information is unavailable, this component cannot be estimated. In such instances, Kumbhakar and Lovell (2000) suggest that the expenditure shares should be implicitly assumed to be equal to the elasticities of output, namely $S_j = \frac{\pounds_j}{\pounds}$, which results in $\Delta AE_{it} = 0$.

If it is assumed that information on input prices is not available, productivity change, ΔTFP_{it}, is defined as the sum of the three aforementioned components:

$$\Delta TFP_{it} = \Delta TE_{it} + \Delta TC_{it} + \Delta SE_{it} \qquad \text{Eq. 3.5.9}$$

It should be mentioned here that according to the framework put forward by Kumbhakar and Lovell (2000), an estimate of productivity change can be produced through either a production function or a production frontier. The difference between the two is that by employing the production function, the analysis assumes that all production takes place within the confines set out by the production function itself and that all observed deviations are due to stochastic variation (ie statistical noise). In other words, the analysis explicitly assumes that all assessed units are fully efficient. In this case, the efficiency change component drops out of the equation 3.5.8 and productivity change is defined as:

$$\Delta TFP_{it}^{prodfunction} = \Delta TC_{it} + \Delta SE_{it} \qquad \text{Eq. 3.5.10}$$

, ie productivity change is defined as the sum of technological change and scale efficiency change.

Note that the estimates of technological change and scale efficiency change are the same, regardless on whether the analysis utilises a production function or a production frontier. The reason for that is twofold: firstly, as is apparent from equations 3.5.4 and 3.5.6, the estimation of both technological change and scale efficiency change is independent from the estimate of (in)efficiency. Secondly, as will become apparent in the following sections, the production frontier is simply an upwards parallel shift of the production function. As such, all coefficients of the production function are the same to those of the production frontier, with the coefficient of the constant being the sole exception[38]. Therefore, the only difference between a productivity change estimate from a production function and a production frontier is the effect of efficiency change, assuming of course that they adopt the same functional form and include the same inputs in the assessment.

Selection of the functional form

As mentioned above, in order to parameterise the production frontier, the analysis needs to select the functional form that will link the output to the inputs, ie $f(\cdot)$ in equation 3.5.3. This selection takes place prior to the start of the analysis and as such reflects the assumptions of the analysis with regards to the underlying functional form of the true production frontier. However, that does not mean that all possible functional forms that are used in econometric analyses are appropriate for the formulation of a production frontier. Instead, this selection process can, and should, be informed by the general theoretical properties of production frontiers.

These properties have already been discussed in section 3.4, when talking about the notion of production technology; the most pertinent of them are also summarised here:

Production frontier – Standard assumptions

[38] In practice, there might be some minor differences in the coefficients produced by COLS (and all other approaches that utilise OLS) and SFA (and all other approaches that utilise an estimation method other than OLS). In the case of SFA, which is estimated through maximum likelihood (MLE), the estimated coefficients have been proven to be identical to the OLS coefficients asymptotically (ie in large samples).

– Non-negativity: Output should always be positive or zero

– Weak Essentiality: To produce any amount of output, at least some amount of input should be used

– Monotonicity (non-decreasing on inputs): when inputs are increased, output should either increase or stay constant.

– Concavity in inputs (technology is a convex set): the rate of change of output relative to inputs (ie the marginal product) should be non-increasing.

Concavity is usually assumed because it is required to enforce the law of diminishing returns and allow the production frontier to simulate 'rational' producer behaviour under competitive markets[39]. As mentioned before, the assumption of concavity is not always desirable, especially when assessing units that operate in markets where competition is restricted. For the purposes of this discussion however, it is assumed that concavity is a desirable property.

Given the above, any analysis that utilises production frontiers needs to adopt a functional forms that:

– conforms to law of diminishing returns (assuming that some inputs are difficult to vary in the short-run);

– is monotonic in its continuity, ie when inputs are increased, output either increases or stays constant;

– adheres to concavity constraints globally, and

– can be empirically estimated (ie applied in practical settings)

Additionally, the functional form would ideally be flexible enough to allow for a production frontier that displays different returns to scale at different input/outputs correspondences, similar to the production frontier estimated by DEA. Another desirable property is that the selected functional would allow for variation in the rates of substitution between inputs

[39] 'Rational' producer behaviour in this instance is used to describe the situation where no production will take place in the area of the production function that is non-concave, ie the area that displays increasing marginal products of the inputs.

(ie variable marginal rates of technical substitution for inputs and variable output elasticities).

Probably the most common functional form applied in econometric analysis is the Cobb-Douglas, named after Charles Cobb and Paul Douglas who presented it in their seminal study in 1928. This functional form is especially relevant in this case, since it was the functional form used in the study that modelled an aggregate, economy-wide production function. Since then, it has become the staple functional form to represent both production, but also cost functions. A Cobb-Douglas production function with two inputs, labour *L* and capital *K* is given below:

$$Y_i = AK_i{}^{\ 1}L_i{}^{\ 2} \qquad \text{Eq. 3.5.11}$$

, where A is a constant and $\ _1$ and $\ _2$ are the elasticities of capital and labour respectively. The Cobb-Douglas is not linear, but can be easily transformed into a linear function by logarithmic transformation:

$$\ln Y_i = \ln A + \ _1 \ln K_i + \ _2 \ln L_i \qquad \text{Eq. 3.5.12}$$

The advantages of adopting a Cobb-Douglas production frontier are:

– very easy to estimate by simply logarithmically transforming the data and applying Ordinary Least Squares;

– exhibits decreasing marginal productivity and so can model diminishing returns, and

– allows for either increasing, decreasing or constant returns to scale.

The disadvantages of the Cobb-Douglas production frontier are:

– It results in returns to scale estimates that are global (ie not allowed to vary from unit to unit).

– The elasticity of substitution for inputs is always equal to unity[40].

[40] The elasticity of substitution measures the possibility of substitution between inputs within the confines of the production technology. It can take only non-negative values. An elasticity of substitution of zero indicates that no substitution is possible between the chosen inputs, while an elasticity of substitution of infinity indicates that the inputs are

– Does not incorporate second-order effects, in the form of interaction terms and quadratic transformations of the inputs.

– It can only produce a technological change component that is Hicks-neutral.

These issues can be resolved by adopting a so-called ‘flexible’ functional form, such as the translog. The translog functional form was developed by Christensen, Jorgenson, and Lau (1973), with the explicit purpose of overcoming some of the shortcoming of the Cobb-Douglas functional form. A translog production function with two inputs, labour *L* and capital *K* is given below:

$$\ln Y_{it} = a_i + {}_{L} \ln L_{it} + {}_{K} \ln K_{it} + \frac{1}{2} {}_{LL} \left(\ln L_{it}\right)^2 + \frac{1}{2} {}_{KK} \left(\ln K_{it}\right)^2 + {}_{KL} \ln K_{it} \ln L_{it}$$

Eq. 3.5.13

The adoption of a translog functional form offers a number of theoretical and practical advantages in this setting:

– The translog is considered a ‘flexible’ functional form since it can provide a local second order approximation to any arbitrary functional form – other common functional forms used in productivity and efficiency analysis such as the Cobb-Douglas and CES can be considered as a special case of the translog.

– It can exhibit decreasing marginal productivity and is as such is consistent with the diminishing returns assumption.

– The elasticity of substitution for inputs is fully flexible.

– The flexibility of the translog allows the model to display non-constant returns to scale, both in the sense that the sum of output elasticities can be different than unity and also that each assessed unit can display unique elasticity estimates depending on its mix of inputs.

– The translog allows for both time- and unit-variant technical change, which is not restricted to be Hicks-neutral.

---

perfect substitutes. An elasticity of substitution of one implies that the use of one input is independent of the use of the other input.

These advantages however come at a cost:

– The translog is in general more difficult to parameterize in practice, relative to the Cobb-Douglas. The estimation process takes up more degrees of freedom and as such requires relatively large samples. Additionally, the inclusion of interaction terms and second-order effects makes it difficult to evaluate marginal effects, mainly due to the fact that they introduce multi-collinearity in the estimation process, which can also affect the stability of the coefficients of the translog model[41].

– More importantly however, a translog production function is not guaranteed by construction to be globally monotonic or concave. If this is the case, the standard assumptions on the technology would be violated, which would have adverse effects on the estimation of productivity change.

There are a host of other possible functional forms that can be used to estimate a production frontier, but are not as common in the literature. Two of them are briefly summarised below:

Constant elasticity of substitution (CES)

An early alternative to the Cobb-Douglas production function, designed to display elasticities of substitution for inputs other than unity.

$$Y = A\left({}_1K^{\circ} + {}_2L^{\circ} \right)^{\frac{1}{\circ}} \qquad \text{Eq. 3.5.14}$$

– more flexible than Cobb-Douglas (Cobb-Douglas is nested in CES when ρ approaches 0);

– more tractable than the translog, in the sense that it includes fewer parameters to estimate; but,

– difficult to estimate in practice since it requires statistical approximation methods or non-linear least squares, rather than standard OLS or MLE; and

– not common in performance measurement.

[41] Note that multicollinearity by itself will not lead to biased coefficients, since it only affects their standard errors. However, if the model is somehow misspecified (for example due to omitted variables), the presence of collinear variables can compound the bias introduced by the misspecification.

Quadratic functional form

$$Y = A + {}_1K + {}_2L + {}_{11}K^2 + {}_{12}KL + {}_{22}L^2 \qquad \text{Eq. 3.5.15}$$

- very similar to translog, both in terms of its theoretical properties and also with regards to estimation;
- is usually adopted when some inputs are equal to zero, since the log of 0 is undefined;
- similar issues with the translog, with regards to ease of use, interpretation of marginal effects and violation of monotonicity and concavity.

3.5.2. Corrected OLS (COLS)

Corrected OLS is a deterministic, econometric approach developed for the measurement of efficiency. It is probably one of the first approaches that have been created to ‘correct’ the inconsistency of the OLS-derived constant term of the regression, when technical inefficiency is present in the production process.

COLS explicitly assumes that the production function estimated through standard OLS provides a good representation of the technology, especially with regards to the output elasticities of the inputs utilised in production. Where the production function falters is in that it does not provide a full envelopment of the data and as such, cannot be considered as an appropriate representation of the technology frontier. To correct for this deficiency, COLS shifts the production function upward so that it envelopes all observed input/output correspondences. In other words, the production function is shifted in such a manner that all observed units lie either on the shifted function or below it. Since the parameters of the production function are deemed to be a good representation of the technology, this shift is parallel, so that the output elasticities of the inputs remain unchanged. The parameters that are changed are the constant and the error terms of the original production function.

In essence, the model is asked to estimate[42]:

$$\ln y_i = \sim + \; {}^{k} x_i^k - u_i \qquad \text{Eq. 3.5.16}$$

[42] The examples here assume Cobb-Douglas production technology.

, where $\ln y_i$ is the log of output, x_i^k is a vector of *k* inputs (also in logarithms) and u_i is technical inefficiency. OLS can be used to estimate:

$$\ln y_i = \sim + \quad {}^k x_i^k + £_i \qquad \text{Eq. 3.5.17}$$

, where $£_i$ is the residual. So, equation 3.5.16 can be converted to equation 3.5.17 just by setting $£_i = -u_i$. The problem however is that $£_i$ has the properties of the classical residual, namely it is normally distributed with a zero mean; u_i on the other hand is always non-negative parameter, with an unknown distribution. This can be corrected by standardising the estimate of technical inefficiency:

$$£_i^* = -u_i + E(u_i) \qquad \text{Eq. 3.5.18}$$

$£_i^*$ is a parameter that has zero mean and an unknown distribution. COLS uses equation 3.5.16 as a starting point and converts it to:

$$\begin{aligned} \ln y_i &= [\sim - E(u_i)] + \quad {}^k x_i^k - u_i + E(u_i) \\ \ln y_i &= a^* + \quad {}^k x_i^k + £_i^* \end{aligned} \qquad \text{Eq. 3.5.19}$$

, which can be estimated through OLS.

COLS sets the expected value of technical inefficiency to be equal to the maximum residual of the OLS-estimated production function, since this would result in a production frontier that envelops all observations. In short:

$$\begin{aligned} E(u_i) &= \max(£_i^*) \\ a &= a^* + \max(£_i^*) \\ -u_i &= £_i^* - \max(£_i^*) \end{aligned} \qquad \text{Eq. 3.5.20}$$

As is apparent from the above, COLS models are quite easy to specify and since the estimation is based on a simple OLS regression, estimates of efficiency and productivity change can be easily derived, even by non-specialists using basic software[43]. The drawback of COLS however is that it is does not account for the stochastic nature of the

[43] A basic version of Microsoft Excel with no add-ins is sufficient for the application of COLS.

production process; the next approach examined here attempts to address this specific oversight.

3.5.3. Stochastic Frontier Analysis

Stochastic Frontier Analysis is the pre-eminent parametric frontier-based approach, developed independently by Aigner, Lovell, and Schmidt (1977) and by Meeusen and van Den Broeck (1977). The approach relies on the notion that not all deviations from the frontier are the result of the decisions made by the assessed unit. Under a deterministic approach, such as DEA, GA and COLS, any external event that has an impact on the production process, such as bad weather, but also any non-systematic error in the measurement of the relevant components of the approach (be it inputs, outputs, prices or other contextual variables) would have a direct impact on the estimated efficiency or productivity measure. SFA attempts to disentangle those random effects by decomposing the residual of the parametric formulation of the production function into noise (random error) and inefficiency.

In more detail, SFA attempts to estimate[44]:

$$\ln y_i = \sim + \quad {}^{k} x_i^k + v_i - u_i \qquad \text{Eq. 3.5.21}$$

, where v_i is the standard two-sided, normally distributed error term, which is also assumed to be independently distributed of technical inefficiency, u_i.

Equation 3.5.21 cannot be estimated through OLS; however, if technical inefficiency is independently distributed relative to the inputs, the OLS coefficients are statistically consistent, except for the constant and the estimates of technical inefficiency. Therefore, OLS can be used as a first step to estimate the slope parameters (coefficients of all the inputs) and then a second method can be used to estimate the constant and the two residual components, namely the stochastic element and technical inefficiency.

In essence, what is required here is an approach that allows for the decomposition of the classical regression residual into an inefficiency term and a stochastic (noise) element:

$$£_i = v_i - u_i \qquad \text{Eq. 3.5.22}$$

44 The examples here assume Cobb-Douglas production technology.

This decomposition is informed by the general properties of these two elements:

– v_i is a standard, two-sided, normally distributed error term with zero mean;

– u_i is strictly positive and has an unknown distribution.

The above does not provide sufficient information to attempt the decomposition; to do so, the analysis requires more detailed information on the distribution of the technical inefficiency component in the form of its general shape and its first, second and third central moments. Therefore, in order to proceed with the decomposition, the analysis needs to assume **a-priori** the distribution of u_i. The potential distributions should have the following characteristics:

– They must be one-sided, since technical inefficiency takes only non-negative values, ie $u_i \geq 0$.

– They should conform to the assumption that higher inefficiency is less likely to be observed, ie, the chosen distribution should be decreasing at higher values.

Based on the above, a number of distributional assumptions are possible for the decomposition. However, the most common distributions for the technical inefficiency component used in SFA are the half-normal and the exponential. The main reason for the prevalence of these distributions is that they are both one-parameter distributions; in other words, if one moment of the distribution is known, the analysis can use this information to derive all the other moments of interest[45]. Other, two-parameter distributions can also be employed, such as the truncated normal or the gamma; their use can provide additional flexibility but at the cost of increased difficulty of estimation. For a comprehensive discussion on the possible distributions that can be used for the decomposition, see Greene (2007).

After selecting the distribution of the inefficiency term, SFA estimates and decomposes the composed error term in a single step using Maximum Likelihood Estimation (MLE). MLE estimates the parameters of the production function in such a way so they provide

[45] In other words, if the analysis can provide an estimate of the variance, this can be used to calculate the mean and the skewness statistics of the one-parameter distributions.

the highest joint probability of observing the current sample. There are three main steps in the estimation procedure:

- First, the analysis needs to construct a joint probability distribution for the composed error term; this will be based on the normal distribution (for the stochastic element) and the selected distribution of the technical inefficiency term.
- The second step is to use this information to construct the log likelihood function that corresponds to the sample under analysis, or in this case, to the input/output correspondences of the assessed units.
- The third and final step is to iterate through different values for the estimated parameters in order to find the set that maximizes the log likelihood.

The MLE process itself does not provide a direct estimate of technical inefficiency, since the inefficiency parameter is unobservable. However, it provides sufficient information to generate an estimate of the conditional mean of inefficiency $E(u|\varepsilon)$ which can be used to generate estimates of technical inefficiency for all assessed units based on the distributional assumption for the term. The issue here is that there is no single way to generate this conditional mean. In the SFA literature, the two most common estimators used to generate the conditional mean of technical inefficiency are the JMLS (Jondrow, Lovell, Materov and Schmidt (1982)) and the BC (Battese and Coelli (1988)) estimators. Both have their strengths and weakness and both estimator usually produce efficiency estimates that are very similar, in terms of absolute values, and highly correlated. As such, the decision to select one estimator over the other has, in the majority of cases, a negligible impact to the final productivity change estimates.

What is probably more concerning is that, regardless of the estimator chosen, the nature of the SFA approach is such that the final estimates of efficiency are statistically inconsistent, when the analysis utilises cross-sectional data or pooled data. In other words, the estimate of u_i does not necessarily converge to the true value of u_i. Statistically consistent estimates of technical efficiency are possible in the panel setting, but only under the assumption that technical efficiency remains unchanged for the duration of the analysis. Arguably, this is a very restrictive assumption when the aim is to measure aggregate productivity growth, especially when the available data span multiple decades.

For a more detailed discussion on those topics, see Kumbhakar and Lovell (2000).

3.6. Discussion and conclusions

This chapter provided an introduction to some of the most widely adopted approaches for measuring productivity change. The approaches discussed were classified in three distinct groups: index number approaches (detailed in section 3.2), non-parametric distance functions (detailed in section 3.4) and econometric approaches (detailed in section 3.5). The discussion focused on:

– Growth Accounting, as the most common representative of the index number approaches;

– DEA-based Malmquist indices, as the most common representative of the non-parametric distance functions; and

– COLS- and SFA-based Malmquist-type indices, as the most common representative of the econometric/parametric approaches.

DEA-based Malmquist indices and COLS- and SFA-based Malmquist-type indices also belong to the wider family of frontier-based approaches, since they measure productivity change relative to the production frontier.

The discussion provided details on the reasoning behind each particular group and approach and discussed how each approach derives the final productivity estimate and what this estimate contains. This concluding section summarises some of the more pertinent points and discusses the relative strengths and weaknesses of each approach, starting with Growth Accounting.

3.6.1. Growth Accounting

Growth Accounting is currently the most common method for measuring productivity. It is adopted by most national statistical agencies, such as the US Bureau of Labor Statistics, Eurostat (which sets policy for all national statistical agencies in the European Union) and the UK Office of National Statistics, as well as a number of global organisations, such as the OECD.

Probably the most important reason for the widespread adoption of GA is its ease of use and its close correspondence to the National Accounting framework adopted by all developed and most developing countries world-wide. GA has relatively modest data requirements; when measuring VA-based productivity growth, the required data include:

– the overall value of the output used in the analysis (economy- or industry-level) and its aggregate price level,

– the overall cost of intermediate inputs, so that VA-based output can be calculated,

– the overall cost of labour employed, together with its aggregate price level; and lastly,

– data of sufficient quality to calculate a quantity index of capital services employed (admittedly, collating the necessary information to generate a relevant and accurate capitals services measure is quite challenging, as discussed in chapter 2). Information on the price of capital services is usually not required, as GA normally relies on the neoclassical assumptions that allow for the residual calculation of the share of capital services.

When the analysis requires a measure of GO-based productivity growth, the required data include the above, plus information on the aggregate price level of the intermediate inputs employed in the production process.

It should be mentioned that collating the above information in not an easy task in itself. However, the various National Statistical Agencies already gather and aggregate the majority of the data needed for the generation of the required input and output price and quantity measures for the purposes of generating National Accounts. As such, one of the most significant hurdles when undertaking any quantitative analysis, namely data availability, is not a major issue when measuring productivity growth using Growth Accounting. Note though that this same information can also be utilised by the frontier-based approaches for the purposes of measuring productivity change; in fact, they have lower data requirements relative to GA, since they do not require information on prices.

Probably the major advantage of GA, and its main strength relative to the aforementioned frontier-based approaches is that GA does not require information on a comparator set in order to assess the productivity change of the assessed unit. As discussed in the previous sections, the frontier-based approaches discussed here all require information on a number of comparable units to derive a representation of the technology that would inform the estimation of the production frontier. Comparable units is defined here as production units that utilise similar inputs to produce similar outputs. In the context of aggregate productivity growth, GA would require data on inputs, outputs and their respective price for just a single economy, while the frontier-based approaches would require data for a number of economies, in order to estimate economy-wide productivity growth. So, although GA does require information on prices, while frontier-based approaches do not, GA can still be considered as having lower data requirements relative to these approaches, since it can be used to measure aggregate productivity change without requiring information on other economies.

The main issue with GA however is with how aggregate productivity change is measured. As discussed in section 3.2.1, in order to parameterise the aggregate production function, the analysis needs to adopt the so-called neoclassical assumptions, which dictate some important characteristics of the production process and the input markets. Some of these assumptions are by their very nature quite restrictive; especially those that state that the production function can only display constant returns to scale, that there is perfect information both on the side of the producers and on the side of the input markets and that producers and input markets have perfect foresight. Taken together, the neoclassical assumptions limit GA to modelling production processes that:

– are fully deterministic,
– can only exhibit constant returns to scale,
– assume that all information is measured with perfect accuracy, and
– assume that all production is efficient.

Given the above, one could argue that the neoclassical assumptions do not necessarily provide a fair representation of most production processes. Numerous studies in the micro setting have demonstrated that there are production processes that display variable returns to scale and that production is not always efficient. Intuition suggests that production is also not necessarily deterministic, due to the existence of unforeseen factors that lie outside of the strict input-output production framework and that can affect the results of the production process.

Even more to the point, the data provided by the National Accounts cannot be considered as perfect representations of reality. There are numerous categories of data that require a number of estimations or imputations from the national statistical agencies, simply because there are not sufficient primary data or collating primary data is impossible (Balk (2008) provides as an example the labour input that is due to self-employed workers).

In addition to the difficulties surrounding the creation of National Accounts, which can lead to imperfect measures of quantities and prices in general, there is also the well-documented issue of measuring the price of capital services, which was discussed in some detail in section 2.2.2. In summary, one of the defining factors used in the calculation of the price of capital services, namely the user cost of capital, is unobservable. GA sets the overall price of capital in such a way so that the total cost of capital plus the total cost of all other inputs equals exactly the final revenue (ie the product of output times its price). In other words, GA implicitly assumes that the profit of all productive activities is, on aggregate level, equal to zero. Balk (2008) states that:

> 'This procedure is usually rationalized by the assumption of perfect foresight, which in this case means that the ex-post calculated capital input prices can be assumed as ex-ante given to the production unit, so that they can be considered as exogenous data for the unit's profit maximization problem.'

In other words, the price of capital services, ie capital input, which is required in order to calculate a robust measure of the elasticity of output with respect to capital services, is not directly available to the analysis.

Due to the above limitations, GA does not necessarily produce a measure that provides a clear representation of changes in productive capability. This is something that practitioners of GA are well aware of; Timmer et al. (2007), when describing the methodology adopted for the estimation of productivity change in the KLEMS Growth and Productivity Accounts project, state in a footnote:

> 'Under strict neo-classical assumptions, MFP [multifactor productivity] growth measures disembodied technological change. In practice, MFP is derived as a residual and includes a host of effects such as improvements in allocative and technical efficiency, changes in returns to scale and mark-ups as well as technological change proper. All these effects can be broadly summarized as "improvements in efficiency", as they improve the productivity with which inputs are being used in the production process. In addition, being a residual measure MFP

> growth also includes measurement errors and the effects from unmeasured output and inputs.’

A further disadvantage of GA is that the estimated productivity change cannot be decomposed. Since the approach assumes that there is no inefficiency and that production displays only constant returns to scale, all productivity change is assumed to be due to technological change.

There are possible modifications that can be made to GA to remove the need for the adoption of the so-called neoclassical assumptions. Balk (2008), suggests the use of an exogenous user cost of capital, set based on a benchmark value such as the official Central Bank interest rate. Under this treatment, profits are not set to zero by the measurement approach and as such changes in profitability can be treated as the natural measure of productivity change. While such modifications are promising, they have not been, as of yet, widely adopted in practice and are not further considered here.

3.6.2. DEA-based Malmquist indices

This is a frontier-based approach that measures productivity change through the use of distance functions. The distance functions are estimated through the use of standard, non-parametric DEA models. DEA itself is one of the most common approaches in the measurement of efficiency and productivity change; Emrouznejad, Parker and Tavares (2008) list more than 4,000 published examples of DEA appearing in the academic literature. DEA-based Malmquist indices have also been used in the past for the measurement of aggregate productivity growth; in fact, one of the earliest applications of the approach has been on the measurement and decomposition of aggregate productivity change for 17 OECD countries over the period of 1979 to 1988, using data from the Penn World Tables (one of the largest databases on National Accounts information sourced from a large number of counties worldwide).

The use of DEA-based Malmquist indices offers a number of advantages to the analysis of productivity change:

– **Minimal assumptions**: In order to specify the frontier, the analysis needs to impose minimal regularity conditions on the description of technology. Probably the most restrictive assumption is the requirement that the underlying technology is convex. Normally, this used to ensure that the technology conforms to the law of diminishing returns and to simulate 'rational' producer behaviour. It can be violated when producers cannot freely choose their scale of production and when there are significant economies of scope. However, for most productive activities that take place in competitive or semi-competitive markets, the convexity assumption is justified.

– **Frontier-based**: The approach allows for production to take place below the frontier and as such does not automatically assume that all producers are efficient. This allows for the simulation of diverse behaviour and is one of the main advantages of all frontier-based approaches relative to GA.

– **Can easily accommodate variable returns to scale**: The frontier can be constructed in such a manner that it displays the full range of returns to scale (ie constant, increasing and decreasing). Furthermore, each unit will display its own, individual, returns to scale characteristics, depending on where it is projected on the frontier.

– **Decomposition of productivity change**: Since the approach allows for inefficient behaviour and for unit-specific and time-variable returns to scale, productivity change can be decomposed to (at least) three main elements. Efficiency change, which measures whether the assessed unit became more or less efficient between the periods of the assessment, technological change, which measures the movement of the frontier and scale efficiency change, which measures whether the assessed unit moved towards or away from its most productive scale size.

– **Information on prices is not required**: The construction of the frontier requires only information on input and output quantities. As such, there are no issues with the measurement of the price of capital (especially the estimation of the user cost of capital), as is the case with GA.

– **No requirement for the a-priori specification of the functional form of the production function**: The frontier is constructed as a convex (concave) or conical hull that envelops all data points. As such, the approach does not need to select a functional form that determines the general shape of the production function or production frontier prior to estimation. As a reminder, both GA and the econometric approaches discussed here require this assumption.

– **Technological change is not restricted to be Hicks-neutral**: Due to the flexibility of the non-parametric frontier, the resulting estimate of technological change is not restricted to be Hicks-neutral, ie a parallel shift of the frontier, and common to all assessed units. In other words, as is the case with the returns to scale properties, the technological change component is both unit- and time-variant. This is not the case with GA and with the econometric approaches that assume a Cobb-Douglas functional form for the production frontier.

In summary, the adoption of DEA-based Malmquist indices requires minimal assumptions with regards to the underlying technology and producer behaviour, while at the same times provides a rich decomposition of the resulting productivity change estimate. Nevertheless, the approach also has some weaknesses, namely:

– **Requires information on comparators**: In order to construct the frontier, all of the relevant approaches discussed in this chapter require data on comparable units. For aggregate productivity change measurement, comparable units would be either other economies or industries with the same classification operating on other economies, depending on the level of aggregation adopted by the analysis. This can be a significant limitation, since if this data is not available, frontier-based approaches simply cannot be applied. Fortunately, there are a number of databases that collate National Accounts information from a large number of economies/countries, with the main purpose of facilitating international comparisons[46]. The wider adoption of global National Accounts Standards (discussed in chapter 2) also helps ensure that the available data are largely consistent, both across time and across different economies. Nevertheless, absolute consistency may not always be entirely possible. In addition, the requirement of international comparators necessities the use of deflators that can convert value-based indices denominated into national currencies into indices expressed at a common currency. As mentioned in chapter 2, this is achieved by the use of PPP-adjusted exchange rates; however, as is the case with the underlying National Accounts data, these adjustments cannot be considered faultless. Any errors in imputation or omissions in the calculation of the PPPs, which

[46] See for example the World Penn Tables, the OECD Productivity database and the EU KLEMS project.

are themselves based on information sourced from the National Accounts, can have detrimental effects on the accuracy of the final productivity change estimates.

– **Deterministic in nature**: The formulation of the technology frontier and the measurement of the distance functions necessary for the calculation on the Malmquist index are all deterministic in nature. In other words, the approach implicitly assumes that there are no factors external to the adopted specification of the technology frontier that influence the performance of the assessed units. Additionally, the approach implicitly assumes that the data used to construct the frontier and produce estimates of performance (efficiency and productivity change) are measured without error. Both of these assumptions are unlikely to hold in practice; the issue here is whether the violation of these assumptions is likely to have a significant detrimental effect in the accuracy of the final productivity change estimates. This will be examined in the next chapter. More recently, there have been a number of approaches appearing in the literature that attempt to take into account the stochastic nature of the production function within the non-parametric or semi-parametric framework.[47] These approaches appear promising but have not as yet gained sufficient traction in the academic community in order to be widely adopted in practical settings.

– **Decomposition issues**: As mentioned in section 3.4, there are a number of possible decompositions of the DEA-based Malmquist index when the analysis wants to also assess the impact of scale efficiency change. The 'traditional' Malmquist index (the formulation suggested by Färe et al. (1994), presented in equation 3.4.7) was criticised as inconsistent because it measured efficiency change relative to the VRS frontier but technological change relative to the CRS frontier. The alternative formulation presented by Ray and Desli (1997) (presented in equation 3.4.9), provides a more consistent treatment but may result in cases where productivity change measurement may not be possible for some units, due to issues having to do with the calculation of the required cross-efficiencies.

– **'Traditional' Malmquist indices are likely to be more volatile in unbalanced panels**: The exit of a unit that was defining the frontier prior to the exit can cause

[47] Some examples: Stochastic DEA as a chance constrained DEA model (Olesen and Petersen (1995), MLE-based stochastic DEA (Kumbhakar, et al. (2007)), and StoNED (Kuosmanen and Kortelainen (2012)).

significant volatility in the productivity changes estimates of all other units at the time of exit. Similarly, the entry of a unit that is more efficient than any other thus far can result in a similar effect.

– **‘Traditional’ Malmquist indices are not circular:** In general, circularity is a desirable property of index numbers; however, Färe et al. (2008) suggested that it is not necessarily an important property of productivity indices, due to the fact that productivity change can be perceived as inherently path-dependant.

Some of the weakness of the ‘traditional’ Malmquist index detailed above can be remedied if the analysis adopts an alternative formulation, such as the ‘circular’ Malmquist index (Portela and Thanassoulis (2010)), which measures productivity change with reference to the meta-frontier. In more detail:

– The circular Malmquist index provides an easy and practical way of measuring scale efficiency change, with no possibility of undefined efficiency scores. Since the circular Malmquist index assesses productivity relative to the meta-frontier, which by definition offers full envelopment of all available data points, there is no possibility for infeasible solutions.

– The circular Malmquist index is immune to sudden changes in the shape of the frontier that can be caused by a unit entering or exiting the dataset. As such, it allows the use of unbalanced panel datasets, without the need to discard any observations.

– Lastly, the circular Malmquist index satisfies the circularity condition for both the headline index and its decompositions.

Nevertheless, the main weaknesses of the approach, namely that the technology is assumed to be deterministic in nature and that the approach relies on the availability of data on comparable units, remain, regardless of the index formulation adopted for the analysis.

3.6.3. COLS- and SFA-based Malmquist-type indices

These approaches belong to the more general family of econometric (or parametric) approaches for performance measurement, be it either efficiency or productivity change.

They are both frontier-based approaches and as such have a lot of similarities with the DEA-based Malmquist indices discussed above. In summary, both approaches have the following strengths, irrespective of their actual specification:

– They are frontier-based and as such do not require the assumption that all production is efficient.

– They can decompose productivity change; the actual decomposition will depend on the adopted functional form of the production frontier. If a Cobb-Douglas functional form is adopted, productivity change can be decomposed into efficiency and technological change. If a flexible functional form, such as the translog, is used then scale efficiency change can also be assessed.

– They do not require information on prices. As is the case with DEA-based Malmquist indices, the frontier can be derived using solely information on input and output quantities.

When a flexible functional form is adopted for the estimation of the production frontier, the analysis also shares the following strengths with DEA-based Malmquist indices:

– The production frontier can display the full range of returns to scale; the analysis will result in returns to scale estimates that can be both unit- and time-variable.

– The estimate of technological change is not restricted to be Hicks-neutral. As above, the analysis will result in technological change estimates that can be both unit- and time-variable.

Overall, econometric approaches are seen as having two main advantages over the non-parametric distance functions:

– firstly, they can utilise a barrage of easy-to-implement statistical tests for inference, drawn from the long and varied history of econometric analysis; and,

– secondly and more importantly, they can model a production process that is stochastic in nature.

With regards to the first issue, relatively recent advances in the area of DEA and non-parametric distance functions have revealed that statistical inference in these models is justified based on the statistical properties of the non-parametric distance functions.. This area of research started with the work of Banker (1993) and has been build on and expanded by a number of academics since then; for a more comprehensive review, see Simar and Wilson (2008). Note however that there is, as of yet, no broad agreement on how to assess these properties and what type of statistical tests would be appropriate. As such, the advantage that the econometric approaches have with regards to statistical inference is still relevant.

The second advantage, that econometric approaches can treat production as a stochastic process, applies to only the second of the approaches discussed in this chapter, namely SFA, but is arguably much more important. GA and DEA-based Malmquist indices, but also COLS – a parametric approach – assume that all deviations from the frontier are the result of the decisions made by the assessment unit. As mentioned in the section above, this could be a quite restrictive assumption, especially in the macro setting. SFA is the only approach examined in this thesis that explicitly takes into account the stochastic nature of the production process, by attempting the decomposition of the residual of the parametric formulation of the production process into noise (the stochastic element) and inefficiency.

However, the ability to disentangle noise from inefficiency and the ease-of-use when it comes to statistical inference comes at a cost.

– Non-parametric distance functions and DEA in particular construct the frontier using a minimal set of assumptions and the observed input/output correspondences of the units under assessment. Econometric approaches require the same assumptions as DEA, but in addition to those, also require the specification of the functional form of the frontier. In this, econometric approaches are similar to the index number approaches; in order to derive a productivity change estimate, the analysis must pre-specify the general form of the transformation function.

– In addition to the above, SFA requires some additional assumptions regarding the distribution of the inefficiency component in order to attempt the decomposition of the classical error term into noise and inefficiency. Fortunately, the ultimate choice of a distribution does not appear to be overly significant; a number of studies demonstrate that the choice of a specific distribution over another has very little impact on the rank correlations of the different efficiency measures (see for example Kumbhakar and Lovell (2000) and Greene (2008)).

– The decomposition of the parametric residual into noise and inefficiency even with these assumptions is not an easy task for the estimation approach. Various simulation studies[48] have demonstrated that the accuracy of the SFA estimates, particularly that of technical efficiency, suffers when the sample available for analysis is small (ie n<100). In such cases, the determinist approaches outperform SFA in terms of the accuracy of the resulting efficiency estimates, even when the production process is assumed to be stochastic.

– Lastly, even with the imposition of these additional assumptions, the estimates of technical inefficiency and, by extension, the estimates of efficiency change are statistically inconsistent when the analysis is based on a pooled dataset. This issue can be corrected if the analysis utilised the panel structure of the dataset and assumes that efficiency is constant over the period of the assessment. This last assumption is problematic, since most studies of aggregate productivity growth utilise fairly long panels, often spanning more than a couple of decades. The assumption that technical efficiency stays constant over long periods of time is fairly unrealistic.

3.6.4. Concluding remarks

The above discussion reveals that each approach has its relative strengths and weaknesses. GA has been traditionally the preferred approach mainly due to its ease-of-use, since the necessary data are readily available through the system of National Accounts and there is no requirement for information on comparable units, ie from other economies. However, the assumptions needed for its implementation are quite restrictive, one could argue to the point of being unrealistic.

Due to the wider adoption of global National Accounts Standards but also due to increasing interest in international comparisons of macroeconomic indicators, there are currently at least three high quality databases that aggregate and also standardise National Accounts information from a large number of economies worldwide. The pan-

[48] See for example Ruggiero (1999); for a more thorough review, see chapter 4 and appendix 1.

European database, EU KLEMS, is of particular relevance, given that it was designed with the explicit aim of measuring and comparing aggregate productivity growth. The availability of data on international scale means that other approaches that require comparable data sourced from more than a single economy can now also be more widely employed. Frontier-based approaches have been widely utilised in past, but mainly in the micro setting and in applications where data on comparable units are available. Since data availability in no longer a major issue, there is no reason why they cannot also be employed for the measurement of aggregate productivity change.

In fact, the adoption of such approaches can offer significant advantages to the analysis, given that frontier-based approaches uniformly require fewer, and arguably less restrictive, assumptions.

In particular, DEA-based Malmquist indices can both estimate and decompose aggregate productivity growth using minimal regulatory assumption for the nature of the technology and its frontier. They can accommodate production processes that display variable returns to scale, non Hicks-neutral technological change and require no information on input prices, which can be a significant advantage given the difficulties in the construction of a price index for capital inputs. Even more crucially, they do not automatically assume that all producers are efficient and can therefore model a wider range of productive activities.

The major drawback of DEA-based Malmquist indices is that they are deterministic in nature. There is however another common frontier-based approach, namely SFA, that can model aggregate production as a stochastic process. SFA-based Malmquist-type indices share many of the strengths of the DEA-based Malmquist indices; variable returns to scale, non Hicks-neutral technological change and inefficient production can all be modelled, without the need for information on input prices. However, they are also more restrictive, since they need to pre-specify a functional form of the production frontier and the distribution function for the inefficiency component. Furthermore, they require large sample sizes and also suffer from the fact that the final estimates of efficiency are statistically inconsistent.

According to the above, it appears that frontier-based approaches are likely to produce richer, more accurate estimates of productivity than GA, mainly due to the fact that they rely on fewer assumptions and can thus model a wider range of production processes.

However, it is impossible to actually assess or quantify these differences without undertaking some form of controlled experiment. The process of constructing and undertaking such an experiment is the subject of the next chapter of this thesis.

Chapter 4. Measuring productivity change using GA and frontier-based approaches – Evidence from a Monte Carlo analysis

4.1. Introduction

The previous chapter detailed a number of approaches that can be used in the measurement of productivity change; each has its strengths and weaknesses and it is sometimes difficult for the analyst to choose between them, based solely on their theoretical properties. Despite the availability of numerous alternatives, Growth Accounting (GA) has been, and still remains, the method of choice when measuring aggregate (ie country- or sector-wide) productivity growth for most interested agents, namely statistical agencies (national and international), central banks and government bodies (see for example the US Bureau of Labor Statistics technical note on multifactor productivity[49], the UK Office for National Statistics Productivity handbook (2007), and the EC-sponsored EU KLEMS project[50]). A major factor in the widespread adoption of GA is the fact that estimates can be (relatively) easily produced using country- or sector-specific National Accounts data, without recourse to information from outside the country or the sector examined; on the other hand, GA requires the adoption of a number of simplistic (potentially unrealistic) assumptions, most notably those relying on the existence of perfect competition, which could lead to unreliable estimates.

The aim of this chapter is to provide quantitative evidence through the use of Monte Carlo simulation experiments on the performance of both GA and the various frontier-based approaches under a number of conditions for the estimation of productivity change. The first goal is to assess the relative accuracy of the GA produced estimates when some of the standard neoclassical assumptions are violated. The second goal is to

---

[49] http://www.bls.gov/mfp/mprtech.pdf, accessed 14 January 2011

[50] See O'Mahony and Timmer (2009).

compare the performance of the different frontier-based approaches discussed in the previous chapter under a number of conditions[51].

Simulation studies, also called Monte Carlo experiments have been used many times in the past in the Efficiency and Productivity literature, although the main focus of such experiments has been on examining the accuracy of the efficiency estimates. One of the main reasons such simulation analyses have been so prominent in the field is that both efficiency and productivity are not quantities that can be directly observed in real world applications and as such, they can never be directly measured. This is a significant issue in a field where there are a number of competing approaches for measuring these residual values, each with its own theoretical strengths and weaknesses. This is further compounded by the fact that both efficiency and productivity change estimates produced by these approaches can sometimes be quite dissimilar. If this is the case, the applied researcher is faced with the difficult choice of selecting which set of estimates to recommend as more accurate. This selection process could be informed by the theoretical properties of the estimation techniques, and sometimes, in situations where the assumptions made by a single estimation technique clearly match the real-world data, this is sufficient. However, the situations where the applied researcher does not have a clear picture of the characteristics of the industry under examination or the sample of units under assessment are probably more common in practice[52]. Also, while theory might suggest that some approaches would be more appropriate under some conditions, it may well be the case that adopting these approaches might introduce additional complexity in the analysis, which in turn could introduce bias, without officering any significant advantages in accuracy.

Due to these considerations, simulation experiments that test the robustness of various approaches under different conditions can provide valuable quantitative evidence that can be used to facilitate the selection process. The goal of such experiments is to provide a tool that helps the researcher study and understand a system better.

At the centre of all such experiments lies the data generating process (DGP), ie the methodology for generating the parameters of interest. Since these parameters are generated from a known process, their values are known (or can be calculated) a priori.

[51] As mentioned in the introduction, an extract of this chapter has been published in the European Journal of Operational Research (see Giraleas et al. (2012))
[52] This is in most cases due to lack of data or issues regarding the accuracy of the data available.

This is what makes simulation experiments so attractive in the Efficiency and Productivity field; they allow for the measurement of the overall accuracy of the estimates produced by each approach in an objective, quantifiable way, under a wide range of conditions.

This chapter is devoted to carrying out a number of simulation experiments to assess the accuracy of a number of approaches for measuring productivity change. The chapter is structured as follows:

– Section 4.2 is an overview of similar simulation studies that have been undertaken in the filed.

– Section 4.3 presents the methodology for the simulation experiments undertaken for this research, including the data generation processes used and the productivity change measurement approaches considered.

– Section 4.4 presents and discusses the results of the simulation experiments

– Section 4.5 summarises the analysis undertaken for this chapter and its findings of this research, compares them with the findings of the studies examined in the literature review and concludes.

4.2. A brief overview of simulation studies in efficiency and productivity analysis

There have been a number of studies that employ simulations to test the accuracy and robustness of various performance measurement approaches. The majority focus on the accuracy of the estimated efficiency measures and take place in the cross-sectional setting. Only a few studies utilise panel data and from those only one focuses on the accuracy of productivity growth estimates (Van Biesebroeck 2007)[53]. Nevertheless, even though the aim of this chapter is to assess the accuracy of productivity change estimates, it is useful to examine the findings of the studies that focus on efficiency estimates in more detail, since changes in efficiency can have a significant effect in

[53] However, this study adopts non-standard definitions of productivity change and makes some strong assumptions about the nature of the technology that ultimately makes its findings of little use for the current research. This is further elaborated on in the relevant subsection of this chapter.

productivity change[54]. This section presents only a brief overview of their findings; for a more detailed literature review, refer to appendix 1.

The literature review revealed a number of interesting points:

– DEA on the whole produces more accurate estimates of efficiency relative to COLS when the sample size is small, but COLS accuracy improves in larger samples. Additionally, when a significant element of noise is included in the data generating process (similar to that used in the analysis undertaken in this chapter), neither approach produces accurate efficiency estimates. See Banker et al. (1987), Banker et al. (1993) and Thanassoulis (1993).

– DEA and COLS efficiency estimates are more highly correlated with true efficiency values (based on rank correlations) than SFA when sample size is limited, even when noise is present in the dataset. SFA approaches and eventually surpasses the correlation scores of the deterministic approaches at larger sample sizes (greater than 100 observations), although this finding appears to depend on the levels of noise included in the DGP[55]. SFA performance also improves when the data display larger ratios of overall inefficiency (technical and allocative) to noise. See Ruggiero (1999), Resti (2000) and Banker et al. (2004).

– Misspecification of SFA in the form of making a wrong assumption on the distribution of the inefficiency component has only a marginal impact on the rank correlations of the efficiency estimates with their true values. See Ruggiero (1999) and Ruggiero (2007).

The only study that specifically stated that its goal was to examine the robustness of ‘productivity’ estimates derived from a number of techniques, namely the Van Biesebroeck (2007) study, is, unfortunately of limited usefulness, due to a number of incompatibilities in definitions and experiment design. In more detail:

[54] This the case for all frontier-based approaches, since efficiency change is a major component of productivity change, but also for GA and other index number approaches, since the resulting productivity change estimate will incorporate any effects that are due to changes in efficiency. See chapter 3 for more details.

[55] Banker et al. (2004) concluded that DEA is the most accurate approach regardless of sample size, but the DGP included only modest levels of noise (standard deviation of 0.05) and the highest sample size examined included just 100 observations.

- The study assumes that all firms are technically efficient; in fact, the concept of inefficiency, which is central in the present research, plays only a very small part in this study. An element of firm heterogeneity that could be attributed to allocative inefficiency is introduced in some experiments, but due to way this enters the DGP, its effects are unclear.

- The study utilises what could be considered as non-standard and incomplete definitions of productivity change, at least when it comes to frontier approaches. For DEA, what the study labels as productivity change is actually only one component of the Malmquist index of productivity, namely efficiency change. For the parametric approaches, the 'productivity' term includes both the effects of the estimated heterogeneity component (which can be viewed as allocative inefficiency) and the error term; also, the estimated production function does not include a time trend and thus the rate of technological change is not measured.

- The 'productivity' term included in the data generation process is either fixed or follows an autoregressive (AR(1)) process. Furthermore, since the study assumes that all firms are technically efficient, this productivity term represents an inherent characteristic of each individual firm that is **known** by that firm before production begins in each time period (but is not observable by the researcher) and as such is accounted for in the optimisation decision at the start of each production period (this is significant, since this feature gives an advantage to the index numbers approaches considered in this study). As such, this productivity term could also be interpreted as a firm-specific effect that changes over time, but is always known to the firm in question before production takes place in each time period.

Due to the above, the findings of the Van Biesebroeck (2007) study are of limited usefulness as a cross-check to the analysis undertaken in this chapter and as such, they are not reported here.

Summing up, none of these articles demonstrates that any one efficiency measurement approach has an absolute advantage over another. Nevertheless, the findings of such simulation studies can prove very useful to the applied researcher, in that they can identify a range of specific situations where some estimation technique proves superior (for example when measurement error is expected to be low or high, or depending on the number of units in the sample).

4.3. Methodology of the current research

This section presents the methodology adopted for generating the data for the simulations, the assumptions about the scale of the parameters of interest and the models used for the estimation of productivity change.

4.3.1. Data generating process

There are a number of elements pertaining to the construction of the data generating process (DGP). These include issues such as:

– How should inputs be generated?
– How to link inputs to outputs, ie what form should the production function adopt?
– Should the analysis take into account prices (both input and output) and how does this affect the assessment framework?
– How to generate efficiency values and in what manner should efficiency impact the DGP?
– How to generate a noise parameter to capture the stochastic element of production and how should this element enter the DGP?
– How to generate values that represent technological change and how should this element impact the DGP?
– How many repetitions of the simulations would be required to derive robust results?
– How large should the sample size drawn for this DGP be?

Given that the aim of this research is to compare the various productivity measurement approaches when technical inefficiency is present in the sample, the simulation study focuses on the production side of the economic process, ie the transformation of physical inputs into physical outputs. As such, information on inputs and output(s) is sufficient for the estimation of productivity change under the frontier-based approaches. However, GA requires information on prices for both inputs and output(s) in order to parameterise the production function[56] (see chapter 3), so price information that is consistent with the quantities of inputs used and outputs produced by each assessed unit would also need to be generated.

[56] The input factor elasticities under GA are calculated as the share of each input to total costs; to calculate this share, it is necessary to be able to calculate the cost of each input, which is equal to its price times quantity.

Since the simulations will focus on the measurement of productivity change in the production setting, one of the first and most important decisions that needs to be made is how should this production process be represented mathematically, ie what functional form to adopt. The measurement approaches that will be examined are both parametric and non-parametric, and as such the choice of the functional form can have a significant impact on the accuracy of the results, as is apparent from the literature review. In order to provide a level playing field for the comparisons, the analysis will examine the accuracy of each approach under two functional forms: a Cobb-Douglas production function and a piecewise-linear production function. In addition, since DEA is the only approach that can easily accommodate multiple inputs and multiple outputs[57], the production framework will focus on the case of a single output produced using two inputs. This case is also the norm when measuring aggregate productivity change, as chapter 2 demonstrated, with value added as the output and labour and capital as the inputs.

The Cobb-Douglas production function employed here is:

$$Y_{it} = L_{it}^{\sim} K_{it} t^{¥} \exp(\neg_{it} - u_{it}) \qquad \text{Eq 4.3.1}$$

,where Y_{it} is the output of unit *i* in time *t*, L_{it} is the labour input of unit *i* in time *t*, K_{it} is the capital input of unit *i* in time *t,* $\neg_{it}$ is the noise component and u_{it} is the technical efficiency of unit *i* in time *t*. An element of technological change is also included in the form of the time trend *t*. Output elasticities are given by the parameters α and β for labour and capital, while γ reflects technological change. The values for the elasticity parameters are α=β=0.5 and γ=0.0198 (which, as noted below, corresponds to 2% p.a. increase in output due to technological change). The noise component (also referred to as measurement error) is normally distributed with zero mean and variance that changes according to aims of each simulation experiment (some simulation experiments assume no noise, while other assume varying degrees of noise, to better assess the impact of this parameter in the accuracy of the productivity estimates.)

[57] The parametric frontier approaches can also in theory accommodate multiple outputs, but this requires the estimation of a distance function with a composite error term, which presents a whole new set of complications, both computational (since a system of equations is usually required) and theoretical (how to deal with endogeneity concerns and how to select whether to use an input distance or an output distance function). For a brief review, see (Greene 2008)

The second set of experiments are based on a DGP that adopts a piecewise linear production function. Generating a piecewise linear production function is not as easy as generating a Cobb-Douglas production function. In the latter case, the functional form itself ensures that the resulting production function has the desirable properties set out by the production theory, namely convexity (as in all input/output correspondences belong to a convex set) and monotonicity (these properties are discussed in more detail in chapter 3). In a piecewise linear function, these properties are not guaranteed; therefore, when designing a piecewise linear production function, the research needs to work within certain confines that would result in a monotonic, convex technology. To do so, one needs to ensure that the marginal product - given by the ‘slope’ of each facet (or ‘piece’) - is non-increasing and that output is non-decreasing on inputs. This is trivial to do in a single-output/single-input case, but it becomes more complicated when additional inputs are introduced since there is no easily identified ‘slope’ for each facet. Adding more inputs also increases the complexity in that the breakpoints in the function that separate the facets need to be selected in such a manner as to ensure that the function is monotonic in its entirety.

For this research, an original process was developed that allows the generation of semi-random piecewise linear production functions (single-output/two-inputs) that display CRS and the desirable properties of monotonicity and convexity. The approach relies on the use of input ratios to determine the breakpoints and provides a methodology to consistently generate the parameters of the each facet in such a manner that each facet displays a progressively decreasing marginal product. The process used is described in brief in Appendix 2.

The piecewise-linear function used for the simulation experiments is given below:

$$y^*_i = \begin{cases} 1.84L_i+0.13K_i \text{ for } L_i/K_i<=0.3 \\ 1.53L_i+0.22K_i \text{ for } L_i/K_i>0.3 \text{ and } L_i/K_i<=0.73 \\ 1.04L_i+0.59K_i \text{ for } L_i/K_i>0.73 \text{ and } L_i/K_i<=1.03 \\ 0.46L_i+1.18K_i \text{ for } L_i/K_i>1.03 \text{ and } L_i/K_i<=1.14 \\ 0.40L_i+1.25K_i \text{ for } L_i/K_i>1.14 \text{ and } L_i/K_i<=2.01 \\ 0.19L_i+1.67K_i \text{ for } L_i/K_i>2.01 \text{ and } L_i/K_i<=3.22 \\ 0.06L_i+2.09K_i \text{ for } L_i/K_i>3.22 \end{cases} \qquad \text{Eq 4.3.2}$$

It should be noted that equation 4.3.2 represents a production function under constant returns to scale (this is easy to check, since all linear functions of eq. 4.3.2 pass through the origin).

The y^*_i parameter represents ‘clean’ output, ie before the effects of inefficiency, technical change and possible measurement error are included. The output value used in the simulation experiments includes all those elements and is given by:

$$y_{it} = y^*_{it} TE_{it} TC_t \exp(v_{it}) \qquad \text{Eq 4.3.3}$$

,where TE_{it} represents technical efficiency and is given by

$$TE_{it} = \exp(-u_{it}) \qquad \text{Eq 4.3.4}$$

,TC_t represents technological change and is a function of time (t) and a constant *y* and is given by:

$$TC_t = t^{¥} \qquad \text{Eq 4.3.5}$$

and v_{it} represents measurement error, which is normally distributed with zero mean and variance that changes according to aims of each simulation experiment. As with the Cobb-Douglas specification, *y*=0.0198, which corresponds to 2% p.a. increase in output due to technological change.

Note that the observed output of the piecewise linear function is consistent with the observed output of the Cobb-Douglas function, in that it includes both elements of technical inefficiency and technical change and in that these elements affect the ‘clean’ output parameter in exactly the same way.

Since the main focus of this analysis is assessing the accuracy of productivity change estimates, the generated data forms a panel dataset. More specifically, all simulation experiments assess the productivity performance of 20 units, observed over a period of five years (ie the total number of observations is 100).

Inputs in the first period are randomly generated following U[0,1]; in subsequent periods, they are scaled by a random, normally-distributed number (the default assumption is that this scaling factor follows N(0,0.10), but some experiments also examine cases where the standard deviation is set to 0.25). The reason for the use of the scaling factor in generating later period data is that one of the goals of the analysis is to assess the impact of input volatility to the productivity change estimates. This is discussed in more detail in later sections.

Efficiency is also randomly generated and follows the exponential distribution; the variance of the efficiency component differs for some experiments, to assess how different levels of prevailing inefficiency affect the productivity change estimates, but generally efficiency follows Exp(1/7), which results in an average inefficiency of approximately 12%. Some simulation experiments also assume that a proportion of the assessed units are fully efficient. This is achieved by generating a set of random numbers following U[0,1], one for each observation in the dataset; if the value of the random number is higher than 0.9, then the efficiency score for that observation is set to 100%. So, for the simulation experiments with fully efficient units, approximately 10% of the observations are fully efficient and the average inefficiency in the sample is approximately 10.8%.

The definition of productivity change used for this analysis relies on the notion of what has come to be known as the Malmquist productivity index. As mentioned in chapter 3, this is probably the most widespread definition of productivity change in the literature and has been used extensively in both the parametric and the non-parametric setting. Furthermore, as discussed in chapter 3, the productivity index produced by GA can be considered as a special case of the Malmquist productivity index.

Under this definition, productivity change is the sum of efficiency change EC_{it}, scale efficiency change SE_{it} and technological change TC_{it}:

$$d \ln TFP_{it}^{true} / dt = d \ln EC_{it} / dt + d \ln SE_{it} / dt + d \ln TC_{it} / dt \qquad \text{Eq 4.3.6}$$

All of the simulation experiments undertaken for this chapter are based on production function that display constant returns to scale; as such scale efficiency change always takes zero values and can thus be ignored. Therefore, productivity change in the DGP is calculated by[58]:

$$\begin{aligned} d\ln TFP_{it}^{true}/dt &= d\ln EC_{it}/dt + d\ln TC_{it}/dt \\ &= (u_{it-1} - u_{it}) + ¥ \end{aligned}$$ Eq 4.3.7

So, given that the efficiency scores in two consecutive periods and the parameter of technical change are known in the generated dataset, the calculation of true productivity change (in the context of the generated dataset) is trivial using equation 4.3.7.

Price data

As was mentioned above, the simulated quantities of inputs and the derived output are sufficient to estimate production-based productivity change under both parametric and non-parametric frontier approaches. However, GA requires information on both input and output prices, in order to create the input shares that serve as the output elasticity parameters of the GA production function. So, in order to include GA in the set of approaches under assessment, price data would need to be generated.

Such price information cannot be just generated at random, since prices play an important role in the optimisation process undertaken by producers. In the basic microeconomic model describing the production process, both output and factor input prices are set exogenously and the producer, based on this information, must decide how much output to produce and what mix of inputs should be employed, given the available production technology.

In the current analysis, the production technology is represented by the production function and since the input quantities that are utilised have already been generated, the related input prices need to be generated in such a way as to correspond to input demand characteristics compatible with the parameters of the production function. Therefore, a structural relationship is needed that links the production function and its parameters to the input demand functions.

---

[58] Note that u_{it} represents technical inefficiency and that $u_i t = -\ln TE_{it}$, where TE_{it} represents technical efficiency. As such $d\ln EC_{it}/dt = \ln(TE_{it}/TE_{it-1})$, so technical efficiency change is measured as the ratio of technical efficiency in period *t* to technical efficiency in period *t-1*.

Fortunately, such structural relationships are readily available in the theory of the producer's optimisation problem. First however, some behavioural assumptions are required in order to make the whole issue tractable. The earliest behavioural assumption put forward by economic theory was that the goal of each producer was to maximise profits (profit maximisation). However, the profit maximisation problem requires a wealth of information, both in terms of price data but also in terms of specifying the demand for output, in order to be solvable; in addition, the analysis carried out in this chapter mainly deals with notions of production, and as such profit maximisation is unnecessarily complex for the purposes of this research. A less demanding assumption with regard to producer behaviour is the assumption of cost minimisation, which was developed by Shephard (1953).

The cost minimisation assumption links the production function to the costs of the production. Under the assumption of constant returns to scale, the production function allows for various capital-labour combinations all of which lie in the same isoquant (ie result in the same amount of output being produced). Thus, the producer could, by simply changing the proportions of input factors, decrease total costs without affecting total revenue - and thus increase profits. This cost minimisation assumption is crucial, in that it identifies the aforementioned structural relationships that link input prices to input quantities.

In more detail, according to the cost minimisation assumption, the goal of each producer is to minimise costs, subject to a given production function, ie:

$$\begin{aligned} &\min \ C_i = w_i^L L + w_i^K K \\ &s.t. \\ &y_i = f(L_i, K_i) \end{aligned} \qquad \text{Eq 4.3.8}$$

, assuming that producers utilise two inputs, capital and labour, with prices w_i^K and w_i^L respectively to produce a given level of a single output. To explore the optimal solution for equation 4.3.9, the Lagrangian form is required, ie:

$$E = w_i^L L + w_i^K K + \text{ª}(y_i - f(L_i, K_i)) \qquad \text{Eq 4.3.9}$$

, where *λ* is the Lagrangian multiplier. The first order conditions for a maximum (or minimum) are given by the partial derivatives of the Lagrangian form, ie:

$$\frac{\partial E}{\partial L} = w_i^L + \mathrm{a}\,\frac{\partial f(L_i, K_i)}{\partial L} = 0$$

$$\frac{\partial E}{\partial K} = w_i^K + \mathrm{a}\,\frac{\partial f(L_i, K_i)}{\partial K} = 0 \qquad \text{Eq 4.3.10}$$

$$\frac{\partial E}{\partial \mathrm{a}} = y_i - f(L_i, K_i)) = 0$$

Combining the first two equations yields:

$$\frac{w_i^L}{w_i^K} = \frac{\dfrac{\partial f(L_i, K_i)}{\partial L}}{\dfrac{\partial f(L_i, K_i)}{\partial K}} \qquad \text{Eq 4.3.11}$$

, which provides the sought-after structural relationship that links input prices to production characteristics. Note that due to the duality theory, the same relationship applies even if the producer is assumed to be output maximising, ie producing the maximum amount of output for a given cost level. Also of note is that the above relationship does not require any assumptions about returns to scale characteristics; additionally, it is easy to demonstrate that this relationship can accommodate inefficient production, in the form of technical inefficiency, regardless of whether the inefficiency term enters the production function in an additive of multiplicative manner. In fact, the only assumptions required for duality theory to hold is that producers are minimising costs based on exogenously-determined input prices (ie producers are assumed to be price takers).

For the simulation experiments that assume a Cobb-Douglas function, equation 4.3.11 becomes:

$$\ln\left(\frac{L_{it}}{K_{it}}\right) = \ln\left(\frac{\sim w_{it}^K}{w_{it}^L}\right) \qquad \text{Eq 4.3.12}$$

where α is the output elasticity of labour, β is the output elasticity of capital, w_{it}^{K} and w_{it}^{L} are the prices of capital and labour respectively for unit *i* in time *t.*

For the simulation experiments that assume a piecewise-linear function, equation 4.3.11 becomes:

$$\frac{w_i^L}{w_i^K} = \frac{a_j}{{}_j} \qquad \text{Eq 4.3.13}$$

where α_j is the output elasticity of labour for the j[th] 'piece' of the piecewise linear function, β_j is the output elasticity of capital for the j[th] 'piece' of the piecewise linear function and w_{it}^{K} and w_{it}^{L} are the prices of capital and labour respectively for unit *i* in time *t.*

Note that equation 4.3.11 does not consider the effects of allocative efficiency. Although it is possible to account for allocative inefficiency within this setting, the simulation experiments undertaken in this chapter assume that the units are always allocatively efficient, and as such, technical efficiency is equal to overall efficiency. This is mainly done to ensure comparability between the GA and frontier-based estimates, since if allocative inefficiency is present, the GA-based productivity estimates will also include its effects, while the frontier-based estimates will ignore them. This could make the interpretation of the productivity measure even more complicated than it already is, given that, as was mentioned in chapter 3, the GA productivity measure also includes the effects of technical change, efficiency change and scale efficiency change.

Given the above, input prices are generated using the following approach:

- First, unique prices for labour are generated for each individual unit for the first period of the analysis as random draws from a uniform distribution (U(0,0.1]).

- These values are then scaled by a random, normally-distributed number to generate values for the subsequent periods, similar to approach used for the generation of the input quantities. The default setting is that the scaling factor follows N(0,0.10), but a second set of experiments also examine the case where the standard deviation is set to 0.25.

– Equations 4.3.12 or 4.3.14 are then used to calculate the true price of capital input, depending on whether the simulation experiments assume a Cobb-Douglas or a piecewise linear production function. Note that the true price of capital is not observable by the researcher (as discussed in chapter 2 and 3) and as such, it is not used directly in the simulation experiments (more detail on this is provided below).

Output prices also need to be generated in a manner that makes them consistent to the already generated input prices. To achieve this, a further assumption is required about the level of profits achievable by the assessed units. Throughout this analysis the assumption is that ‘excess’ profits are zero for an efficient company, which implies that ‘excess’ profits are negative for inefficient companies. In general, the zero profit assumption is implicit (and necessary) in the GA setting. Since one of the goals of this analysis is to examine the accuracy of GA estimates, a zero profit assumption for the efficient units is necessary in order to ensure that no additional sources of potential bias are introduced in the GA estimates (other than the ones we introduce in each simulation runs by relaxing the perfect competition assumptions). In addition, the zero profit assumption is not unrealistic, considering that ‘normal’ profits can be assumed to be achieved through an appropriate rate of return on capital, which is assumed to be already included in the price of capital (see OECD Manual (2001)).

The zero profit assumption allows the analysis to derive a consistent price of output, since costs are equal to efficient revenue for all units:

$$p_{it}Y_{it}^{*} = w_{it}^{K}K_{it} + w_{it}^{L}L_{it} \qquad \text{Eq 4.3.14}$$

$$p_{it} = \frac{w_{it}^{K}K_{it} + w_{it}^{L}L_{it}}{Y_{it}^{*}} \qquad \text{Eq 4.3.15}$$

Where p_{it} is the price of output of unit *i* in period *t* and Y_{it}^{*} is the efficient level of output of unit *i* in period *t*.

Information on inputs, output and their respective prices is sufficient to produce productivity change estimates under GA. However, the methodology so far has relied on the standard neoclassical assumptions (detailed in chapter 3), which do not allow for production to display any type of inefficiency (be it technical or allocative). If technical inefficiency were to be introduced, a unit that is technically inefficient would fail to produce the maximum possible amount of output and as such its revenues will not be able to cover all its costs. These units represent a challenge to the GA framework, since the input shares for those units will not add up to one. A possible way to get around this issue is to adapt the methodology proposed by Hall (1991) to control for the effects of increasing returns to scale and market power to the Solow residual, so that it can apply to inefficient production. However, this would require additional assumptions about input price mark-ups and the allocation of profits (or losses) amongst the factors of production. Such additional information is not present in the national accounts and it is unclear whether reliable sources which could be used to estimate such parameters of interest exist, at least in the economy- or industry-wide setting.

It could be argued that, since the data generation process provides the true prices for labour and capital, the GA weights for the inputs could be calculated based on total costs rather than total revenue. In fact, this seems to be the default option adopted by the Van Biesebroeck (2007) study. However, this treatment would be highly unrealistic, since, as was mentioned in chapter 2, the true price of capital is not observable. Given these difficulties, the usual practice within the GA framework is to derive the price of capital residually. This is achieved by setting capital compensation (ie the cost of capital) to be equal to Value Added (which is equivalent to revenue in the setting of these simulations) minus the labour compensation (ie the cost of labour). Since the quantity of capital can be estimated using national account data, the price of capital can be derived by:

$$w_{it}^{K,GA} = \frac{p_{it}Y_{it} - w_{it}^{L}L_{it}}{K_{it}} \qquad \text{Eq 4.3.16}$$

It should also be mentioned that if the true price of capital was available, GA could easily assess the performance of units that were technically inefficient; since total costs are equal to revenues only for the efficient units, technical efficiency equals the ratio of revenues to total costs. Note that this is not an estimate, but rather an accounting identity; as such, if no other factors that could confound the relationship between costs and revenues are present (such as variable returns to scale and measurement error), GA could be used to measure productivity change with total accuracy in this setting.

However, since the true price of capital is unobservable, the above discussion has little value in the current setting.

Given the above discussion, the analysis undertaken here adopts the more common treatment of calculating the price of capital endogenously, ie as per equation 4.3.16. It should be noted that this approach was also adopted by EU KLEMS (2008) in order to derive their productivity change estimates. The use of this GA-adjusted price of capital ($w_{it}^{K,GA}$) ensures that input shares add up to one and thus allows the use of GA in such a way that is consistent with EU KLEMS and the methodology proposed by the OECD.

A final issue that needs to be addressed is that, in a very few cases, the cost of labour could exceed total revenue, and as such the GA-adjusted price of capital is negative. Although negative capital prices are not inconsistent with theory (see for example Berndt and Fuss (1986)), they are incompatible with the standard GA framework, since they result in negative input shares. To avoid this, the analysis follows the EU KLEMS practice of setting all instances of negative prices to zero.

4.3.2. Productivity measurement approaches considered

Each simulation experiment examines the performance of the following approaches:

– GA,
– DEA,
– Corrected OLS (COLS) and
– SFA (only when measurement noise is included in the experiment).

All of the above approaches are discussed in some detail in chapter 3. In this section, a brief description of each approach is provided together with specification of the various models used in the experiments.

Growth Accounting

Growth Accounting (GA) is the most common index number-based approach for measuring aggregate productivity change. The simulations in this chapter (and indeed throughout the thesis) use the following formulae to estimate the GA-based productivity change measure:

$$\frac{d\ln TFP_i^{GA}}{dt} = \frac{d\ln y_i}{dt} - S_i^L \frac{d\ln L_i}{dt} - S_i^K \frac{d\ln K_i}{dt}$$ Eq 4.3.17

,where S_i^L is the average share of labour in *t* and *t-1*, S_i^k is the average share of capital in *t* and *t-1* given by:

$$S_i^L = \left(\frac{w_{it}^L L_{it}}{p_{it} Y_{it}} + \frac{w_{it-1}^L L_{it-1}}{p_{it-1} Y_{it-1}} \right) \Big/ 2$$ Eq 4.3.18

$$S_i^K = \left(\frac{w_{it}^{K,GA} K_{it}}{p_{it} Y_{it}} + \frac{w_{it-1}^{K,GA} K_{it-1}}{p_{it-1} Y_{it-1}} \right) \Big/ 2$$ Eq 4.3.19

Data Envelopment Analysis and the circular Malmquist index

The approach adopted for this chapter utilises the notion of a circular Malmquist-type index (thereafter referred to as circular Malmquist), which relies on the notion of the 'meta-frontier'. The use of the circular Malmquist makes the estimation of DEA-based productivity change relatively straightforward. Since the meta-frontier is fixed in the timeframe selected for the analysis, the productivity change of a unit between any two time periods can be measured using the ratio of the meta-efficiencies of these two periods:

$$TFP_{it,it+1}^{DEA} = \frac{\bar{\ }^{m}_{it+1}}{\bar{\ }^{m}_{it}}$$ Eq 4.3.20

, where $TFP_{it,it+1}^{DEA}$ is the DEA-based productivity change index and $\bar{\ }^{m}_{it}$ is unit's *i* meta-efficiency in time *t*. As a reminder, $\bar{\ }^{m}_{it}$ is estimated by solving a DEA model that includes all observations from all the assessed units in all available periods. The index is converted to estimates of annual change by a simple logarithmic transformation:

$$\Delta TFP_{it,it+1}^{DEA} = \ln(TFP_{it,it+1}^{DEA})$$ Eq 4.3.21

It should be noted that one of the main advantages of estimating a Malmquist (or Malmquist-type index) using DEA is the ease of decomposition of the index into its components. However, since such decomposition is not available by the standard GA approaches, it is not examined in this chapter.

Corrected OLS

Corrected OLS is a deterministic, parametric approach and one of the numerous ways that have been suggested to ‘correct’ the inconsistency of the OLS-derived constant term of the regression, when technical inefficiency is present in the production process.

COLS was selected for these simulations for a number of reasons:

– The model is very easy to specify and, since it is based on a simple OLS regression, COLS-based estimates can be easily derived, even by non-specialists using basic software (eg. a basic version of Microsoft Excel with no add-ins is enough).

– The approach is deterministic, ie does not take into account measurement error. This is also the case with the DEA-based Malmquist indices and GA and as such, comparisons of accuracy between these approaches would be interesting, especially since COLS is a parametric technique.

– Despite its simplicity, the approach provides surprisingly accurate estimates of efficiency as the literature review revealed, especially when sample sizes are low. It would be interesting to examine whether this good performance persists when measuring productivity change.

Two different COLS model specifications are tested in this chapter. Both are based on a pooled regression model (ie all observations are included in the same model with no unit-specific effect), which is consistent with the adopted data generating process. The first model assumes a Cobb-Douglas functional form and is used for those experiments where the data is generated using the Cobb-Douglas production function. In more detail, the functional form used is:

$$Y_{it} = L_{it}^{a*} K_{it}^{\;*} t^{¥*} \exp(£ *_{it}) \qquad \text{Eq 4.3.22}$$

, where $£*_{it}$ are the estimated OLS residuals. The standard logarithmic transformation converts equation 4.3.22 into[59]:

$$\ln Y_{it} = \sim^{*} \ln L_{it} + \quad ^{*} \ln \mathrm{K}_{it} + ¥^{*} t + £*_{it} \qquad \text{Eq 4.3.23}$$

It should be noted that the above specification matches perfectly the data generating process, when measurement error is not included in the experiments. As such, it is expected that the COLS-derived estimates will be very accurate at least for those experiments.

The second COLS model specification assumes a translog functional form and is used, alongside the Cobb-Douglas models, for those simulation experiments where the data is generated using the piecewise-linear production function. Under the piecewise-linear generated datasets, the Cobb-Douglas COLS model will be misspecified, but it would still be interesting to examine how damaging to the overall accuracy of the estimates this functional form misspecification proves to be. In addition, adopting a more flexible functional form, such as the translog functional form, will reveal whether, and/or by how much, the biases resulting from functional form misspecification can be mitigated.

The translog COLS model is given by:

$$\ln Y_{it} = a_i + \quad_{L} \ln L_{it} + \quad_{K} \ln K_{it} + ¥_{t} t + \frac{1}{2} \quad_{LL} (\ln L_{it})^2 + \frac{1}{2} \quad_{KK} (\ln K_{it})^2 + \frac{1}{2} ¥_{tt} t^2$$
$$+ \quad_{KL} \ln K_{it} \ln L_{it} + ¥_{Kt} \ln K_{it} t + ¥_{Lt} \ln L_{it} t + £^{*}_{it}$$

Eq 4.3.24

In all cases, inefficiency estimates are derived by:

$$-u^{*}_{it} = £^{*}_{it} - \max(£^{*}_{it}) \qquad \text{Eq 4.3.25}$$

Productivity change is calculated based on the same formula as used for the calculation of true productivity change, substituting the true parameters with the various parametric estimates. So, productivity change is given by:

$$d \ln TFP^{COLS}_{it} / dt = d \ln EC^{COLS}_{it} / dt + d \ln SE^{COLS}_{it} / dt + d \ln TC^{COLS}_{it} / dt \qquad \text{Eq 4.3.26}$$

---

[59] It should be noted that Equation 4.3.23 uses the standard practice of not logarithmically transforming the time variable (see Kumbhakar and Lovell (2000)).

, where EC_{it}^{COLS} is the COLS-estimated efficiency change, SE_{it}^{COLS} is the COLS-estimated scale efficiency change and TC_{it}^{COLS} is the COLS-estimated technological change. Efficiency change and technological change are given by:

$$d\ln EC_{it}^{COLS} / dt = (u_{it-1}^{*} - u_{it}^{*}) \qquad \text{Eq 4.3.27}$$

$$d\ln TC_{it}^{COLS} / dt = ¥^{*} \qquad \text{Eq 4.3.28}$$

,for the Cobb-Douglas function and

$$d\ln TC_{it}^{COLS} / dt = ¥_{t} + t¥_{tt} + ¥_{Kt}\ln K_{it} + ¥_{Lt}\ln L_{it} \qquad \text{Eq 4.3.29}$$

, for the translog function.

Scale efficiency change is given by:

$$d\ln SE_{it}^{COLS} / dt = (£_{it}^{COLS} - 1)\sum_{j}\frac{£_{j,it}^{COLS}}{£_{it}^{COLS}}\Delta x_{jt,jt-1} \qquad \text{Eq 4.3.30}$$

, and since the data generation process utilise only two inputs, for the Cobb-Doulgas functional form:

$$\begin{aligned} £_{L,it}^{COLS} &= \sim^{*} \\ £_{\mathrm{K},it}^{COLS} &= \ ^{*} \end{aligned} \qquad \text{Eq 4.3.31}$$

, while for the translog function:

$$\begin{aligned} £_{L,it}^{COLS} &= {}_{K} + {}_{KK}\left(\ln K_{it}\right) + {}_{KL}\ln L_{it} + {}_{Kt}t \\ £_{\mathrm{K},it}^{COLS} &= {}_{L} + {}_{LL}\left(\ln L_{it}\right) + {}_{KL}\ln K_{it} + {}_{Lt}t \end{aligned} \qquad \text{Eq 4.3.32}$$

Stochastic frontier analysis

Stochastic frontier analysis similar to COLS, in that both approaches are parametric, but SFA is stochastic rather than deterministic, in that it attempts to decompose the regression residual into an estimate of noise and inefficiency. It is the only stochastic approach examined in this chapter, and as such is expected to provide the most accurate estimates in the simulation experiments that include measurement error.

As is the case with the COLS approach, two separate SFA model specifications are used: a Cobb-Douglas functional form is employed for those experiments where the data are generated through a Cobb-Douglas production function, and both a Cobb-Douglas and translog functional form for those experiments where the data are generated through a piecewise linear production function. The models are very similar to those used under COLS; in fact, the only difference lies in the specification of the residual.

In more detail, the Cobb-Douglas model is given by:

$$\ln Y_{it} = \sim^{*} \ln L_{it} + \quad {}^{*} \ln \mathrm{K}_{it} + \yen^{*} t + v_{it} - u_{it} \qquad \text{Eq 4.3.33}$$

, whereas the translog model is given by:

$$\ln Y_{it} = a_i + \quad {}_{L} \ln L_{it} + \quad {}_{K} \ln K_{it} + \yen_{t} t + \frac{1}{2} \quad {}_{LL} \left(\ln L_{it}\right)^2 + \frac{1}{2} \quad {}_{KK} \left(\ln K_{it}\right)^2 + \frac{1}{2} \yen_{tt} t^2$$
$$+ \quad {}_{KL} \ln K_{it} \ln L_{it} + \yen_{Kt} \ln K_{it} t + \yen_{Lt} \ln L_{it} t + v_{it} - u_{it}$$

Eq 4.3.34

where u_{it} represents the inefficiency component (and as such $u_{it} \geq 0$) and v_{it} represents measurement error ($v_{it} \sim N(0, \pm_{v}^{2})$). The inefficiency component is estimated based on the JMLS estimator.

Two different distributions for the inefficiency component are tested:

– the exponential distribution, $u_{it} \sim Exp(\pm_{u})$

– the half-normal distribution, $u_{it} \sim N^{+}(0, \pm_{u}^{2})$

When the data is generated using the Cobb-Douglas production function, the exponential Cobb-Douglas SFA model is perfectly specified, since the data generation process also generates the inefficiency values from an exponential distribution. The estimates from the half-normal distribution are included in the experiments to examine the impact of misspecification in the inefficiency distribution to the SFA productivity change estimates. Productivity change is measured in exactly the same way as with COLS, ie using equations 4.3.26 to 4.3.32.

4.3.3. How to measure the accuracy of each estimate

The productivity estimates produced by each approach are compared to the true rate of productivity change, which is derived by equation 4.3.6. Three different measures are employed to judge the accuracy of the estimates under each approach:

The **mean absolute deviation (MAD)** of productivity change, given by:

$$MAD = \sum_{i=1,t=1}^{n,5} \left| TFP_{it}^{TRUE} - TFP_{it}^{EST} \right| / N \qquad \text{Eq 4.3.35}$$

, where TFP_{it}^{TRUE} is true productivity change and TFP_{it}^{EST} is the estimated productivity change derived from the approach under examination. The MAD measure provides a robust central estimate of the overall accuracy of each approach, regardless of the sign of the deviation between the true and the estimated value.

The **mean square error (MSE)** of productivity change, given by:

$$MSE = \sum_{i=1,t=1}^{n,5} (TFP_{it}^{TRUE} - TFP_{it}^{EST})^2 / N \qquad \text{Eq 4.3.36}$$

The MSE measure plays a complimentary role to the MAD measure and provides an estimate of the dispersion of the estimates relative to the true values. Due to its nature, the MSE measure is quite sensitive to estimates that deviate significantly from the true value; as such, larger MSE values are suggestive of more extreme deviations from the true values, other things being equal.

The **mean absolute deviation of the 25th percentile ('top' MAD or TMAD)** of productivity change, which is the MAD of the top 25% of observations when sorted in descending order according to the absolute deviation from the true value. In other words, the analysis calculates the absolute deviation of all observations and then takes into account only the top 25% of those, in order to calculate the TMAD measure. This results in a measure that is quite similar to the MSE measure, with the notable exception that it uses the same units as the MAD measure (absolute deviations rather than squared deviations), and is thus easier to interpret.

Lastly, the analysis also calculated the Pearson correlation between the estimated and true values of productivity change together with the Spearman (rank) correlation statistic, as an additional measure of the overall similarities between the estimated and true values.

In addition to calculating the above measures, the analysis also undertakes statistical testing to determine whether the pairwise differences in those measures between approaches are statistically significant, for all combinations. For example, the average MAD score of the DEA estimates over all simulation runs in a single experiment is tested against the average MAD score of the GA, COLS and SFA (where applicable) estimates. The analysis adopts both a standard pair-wise Student's t-test (assuming unequal variance) for testing the difference in means and the signed-rank test (otherwise known as the Wilcoxon Signed-Rank test), which is usually employed as an alternative to the pairwise t-test when the underlying population cannot be assumed to be normally distributed.

Of the three main accuracy measures, the most important is MAD, which reveals the average absolute deviation from the true estimates. MSE places greater emphasis on instances where the difference between the estimated and the true productivity change is large, and as such is a compliment to the MAD measure, especially if it is felt that larger deviations between the estimate and the true measure in a few units are less desirable than smaller deviations in a greater number of units. The TMAD measure is complimentary to the MSE measure, and has the advantage that it is presented in the same units as the MAD measure, which makes for easier interpretation.

4.3.4. The simulation experiments

As was previously mentioned, the main goal of the simulation analysis is to compare the relative accuracy of the GA and frontier-based estimates under various assumptions that diverge from the long-run, perfect competition optimum. In more detail, this research aims to examine the accuracy of the estimates:

– when technical inefficiency, in various degrees of severity, is present;

– when the production function includes a stochastic element (noise), that could result from measurement error in the factors of production; again this element enters the analysis at various degrees of severity;

– when inputs and prices are volatile from one period to the next, and finally

– when the production function is misspecified.

To that effect, a number of simulation experiments are undertaken, based on the same underlying data generation methodology, while varying the parameters of interest relevant to each experiment. The simulation experiments are subdivided into two major categories. Those in the first category (denoted as S1) use the data generating process that assumes a Cobb-Douglas production function, while those in the second category (denoted as S2) use the data generating process that assumes a piecewise linear production function. As mentioned above, the adoption of two different functional forms is done so that the analysis can test the impact of functional form misspecification for the parametric approaches; it also allows the analysis to assess the accuracy of the DEA- and GA-based estimates when production is represented by a classically smooth function and also when the technology is represented by a more general (less restrictive) convex and monotonic hull.

In both sets of experiments, elements of technical inefficiency and noise (measurement error) are gradually introduced to the production function used to generate the simulated output. More specifically, the experiments undertaken here assume two different levels of technical inefficiency: 'average' levels ($u_{it} \sim Exp(1/7)$) and 'higher' levels ($u_{it} \sim Exp(1/2)$) and three different levels of noise: zero noise ($\neg_{it} = 0$ for all *i* and *t*), 'extensive' noise relative to inefficiency ($\neg_{it} \sim N(0,0.2)$) and 'modest' noise relative to inefficiency ($\neg_{it} \sim N(0,0.05)$). As a reminder, both of these elements enter the DGP individually and both affect output in a multiplicative manner (the inefficiency element enters the DGP as $\exp(-u_{it})$ and the noise element as $\exp(\neg_{it})$). For more detail, see equations 4.3.1 and 4.3.3 and the accompanying discussion in section 4.3.1.

The simulations also examine the effects of input and price volatility from one period to the next; this is achieved by generating a new set of data for each simulation experiment which is based on the more volatile scaling factor ('default' scaling factor is randomly generated and follows N(0,0.10), while the more 'volatile' scaling factor follows N(0,0.25)).

The way all of the above parameters enter into the production function is described in detail in section 4.3. As a reminder, all data generated come from production functions that display constant returns to scale and also include and element of time-invariant technical change (which corresponds to a 2% p.a. increase in output).

All the simulations undertaken in this study are presented in table 4.1:

Table 4.1: Simulation experiments

	Production function	Technical inefficiency	Noise	Input and price Volatility assumptions
S1.1	Cobb-Douglas	'average' levels	zero	both 'default' and 'higher' volatility
S1.2	Cobb-Douglas	'higher' levels	zero	both 'default' and 'higher' volatility
S1.3	Cobb-Douglas	'average' levels	'extensive'	both 'default' and 'higher' volatility
S1.4	Cobb-Douglas	'average' levels	'modest'	both 'default' and 'higher' volatility
S2.1	Piece-wise linear	'average' levels	zero	both 'default' and 'higher' volatility
S2.2	Piece-wise linear	'higher' levels	zero	both 'default' and 'higher' volatility
S2.3	Piece-wise linear	'average' levels	'extensive'	both 'default' and 'higher' volatility
S2.4	Piece-wise linear	'average' levels	'modest'	both 'default' and 'higher' volatility
S2.5	Piece-wise linear	'higher' levels	'extensive'	both 'default' and 'higher' volatility
S2.6	Piece-wise linear	'higher' levels	'modest'	both 'default' and 'higher' volatility

The analysis also tested whether the inclusion of fully efficient units would have any impact on the summary accuracy measures[60]. Overall, the analysis found that the accuracy measures from the simulations which included fully efficient units are almost indistinguishable from the base case; nevertheless, the results from this analysis are also reported in the following sections for completeness.

All results are based on simulation experiments that were repeated 100 times. The decision to use 100 repetitions came through sensitivity analysis; for all the experiments,

[60] The data generation methodology implemented for these simulations ensures that no unit is fully, ie 100% technically efficient.

it was found that after a maximum of 50 to 60 repetitions, the various accuracy measures showed very little variation. Based on this, using 100 repetitions would ensure that the final accuracy estimates provide robust estimates of central tendencies.

It should be mentioned here that the results from the simulation analysis that adopts a piecewise linear production function appear to be robust to different specifications for the production function. This was tested by generating a new piecewise linear function (based on the methodology detailed in appendix 2) and replicating some of the simulation experiments; the differences in the MAD scores from the two different specifications were minor and there were no cases were the relative accuracy rankings differed and as such, the results are not reported in the following sections.

4.4. Results

This section is separated in two main parts:

– the first part provides a summary and a brief discussion of the results for each individual experiment

– the second part examines the results as a whole and provides some discussion on the relative strengths and weaknesses of each assessed approach under the various scenarios examined. This section concludes by providing recommendations on which approach is more appropriate under certain conditions.

Note that the analysis undertaken for these simulations has been quite extensive and as such, only summary results are provided in this section.

As was previously mentioned, the main accuracy indicators are the MAD, MSE and TMAD scores relating to each approach. The MSE measure is difficult to interpret, since it is not expressed in the same units used to measure productivity change[61]. Nevertheless, lower values are desirable. In order to put the various MAD and TMAD measures into context, note that the data generation process adopted, both for the

[61] It is the same issue when trying to interpret different values of variance. Since variation is also not expressed in the same units as the variable it describes.

Cobb-Douglas and the piecewise-linear function, results in an average true productivity change of 2% p.a. but with a standard deviation of approximately 20%[62].

4.4.1. S1 simulation experiments

These experiments adopt a Cobb-Douglas production function for the data generating process. The MAD and TMAD scores are in the same unit of measurement as productivity change, ie percentage change relative to the previous value of the productivity index.

S1.1: 12% average inefficiency, no noise

The most accurate approach for these experiments is clearly COLS, i.e. the parametric, deterministic approach. This comes as no surprise, given that the COLS specification is an exact match to the data generation process adopted for this experiment. The second most accurate approach is DEA, while GA comes third, although even this approach demonstrates only a modest deviation from the true productivity change estimates, as the average MAD value of 0.9% reveals. The tests for the statistical significance of the difference in the accuracy estimates between the approaches reveal that the MAD scores for each approach are statistically different from each other, while the differences in MSE and TMAD scores between the DEA and GA approaches are statistically insignificant as shown in Table 4.2.

Table 4.2: S1.1 summary accuracy scores for the default volatility assumptions

	MAD	MSE	TMAD	Correlation	Rank correlation
GA	0.9%	0.21	4.1%	99.7%	99.6%
COLS	0.4%	0.04	1.3%	99.9%	99.8%
DEA	0.7%	0.20	4.1%	99.7%	99.4%

These results serve to confirm the dominance of COLS in this experiment and the fact that when one of the standard neoclassical assumptions is violated, in the form of introducing the possibility of technical inefficiency in the production process, the frontier-

[62] Both the mean and standard deviation of the true productivity change measure will vary slightly in each simulation run, due to the random nature of the data generating process.

based approaches outperform, on the whole, the more traditional GA[63]. It should be noted however, that GA still provides what could be considered as a quite accurate estimates of true productivity change, although this is not very surprising, given that only the assumption of no technical inefficiency was violated and that the actual level of technical inefficiency in these experiments is relatively low.

When both inputs and input prices are assumed to be more volatile from one period to the next, the relative accuracy rankings between approaches remains the same, but the overall accuracy of all approaches decreases. As a reminder, the analysis randomly generates inputs and price values for the first period while for the subsequent periods input and price values are generated by multiplying the input values of the previous period by a randomly generated scaling factor. The default assumption is that this scaling factor follows N~(0,0.1); to increase volatility, the standard deviation of the normal distribution is increased from 0.1 to 0.25.

Table 4.3 summarises the results of the experiment that assumes increased input and price volatility:

Table 4.3: S1.1 summary accuracy scores for the increased volatility assumptions

	MAD	MSE	TMAD	Correlation	Rank correlation
GA	2.5%	1.70	11.8%	97.9%	97.9%
COLS	0.8%	0.16	2.7%	99.8%	99.5%
DEA	1.2%	0.50	6.1%	99.3%	98.7%

The differences in overall accuracy between approaches become more apparent under conditions of increased volatility in inputs and input prices. As mentioned above, the relative accuracy rankings remain unchanged; COLS still provides the most accurate estimates under the S1.1 conditions, followed by DEA and then GA. It should be noted that the differences in the all three main accuracy scores between all approaches are statistically significant.

63 It should be noted that SFA was also tested under these conditions, but in the vast majority of cases the approach (correctly) failed to identify any significant amount of noise in the data and reverted to using the OLS estimates.

Including fully efficient units in the analysis, also did not affect the overall rankings. COLS still outperformed both DEA and GA, while DEA outperformed GA in all measures.

Table 4.4: S1.1 summary accuracy scores with fully efficient DMUs included

	MAD	MSE	TMAD	Correlation	Rank correlation
GA	0.8%	0.20	4.1%	99.7%	99.7%
COLS	0.5%	0.05	1.3%	99.9%	99.8%
DEA	0.7%	0.20	4.0%	99.7%	99.4%

S1.2: 32% average inefficiency, no noise

The second experiment increases the amount of technical inefficiency in the data, to examine how this would impact the estimation of productivity change. The summary accuracy scores are given below:

Table 4.5: S1.2 summary accuracy scores for the default volatility assumptions

	MAD	MSE	TMAD	Correlation	Rank correlation
GA	2.8%	1.57	10.1%	99.8%	99.6%
COLS	1.7%	0.62	4.8%	99.9%	99.8%
DEA	1.2%	0.46	5.9%	100.0%	99.8%

The findings are quite surprising; despite the fact the COLS model used for this experiment is perfectly specified, the DEA MAD and MSE scores are smaller and the difference is statistically significant. The COLS TMAD score is lower though, relative to the DEA TMAD score, which suggests that although DEA performs better on average, the COLS productivity change estimates are more accurate at the extremes. The difference in these scores is actually quite small, but still statistically significant. The GA scores are the least accurate of the three, by a (relatively) wide margin.

When the analysis is replicated using more volatile input and price values, the overall ranking of the approaches remains the same. Table 4.6 demonstrates:

Table 4.6: S1.2 summary accuracy scores for the increased volatility assumptions

	MAD	MSE	TMAD	Correlation	Rank correlation
GA	7.6%	12.92	30.6%	98.8%	97.9%
COLS	3.0%	1.92	9.3%	99.8%	99.5%
DEA	2.2%	1.99	12.7%	99.8%	99.5%

As with the previous experiment (S1.1), while the relative accuracy rankings between approaches do not change, the increased volatility also results in reduced overall accuracy for all the estimates. All of the assessed approaches are affected, but it appears that the most negatively affected approach is GA, which displays MAD scores that are almost three times higher relative to the default volatility conditions, while both the DEA and COLS MAD scores are less than twice as high.

The relative rankings amongst the three approaches also remain the same when fully efficient units are included in the analysis. In fact, both the DEA and GA accuracy scores improve, which is somewhat expected given that the inclusion of the fully efficient units raises the average efficiency in the sample. However, the accuracy scores of the COLS estimates do not show any improvement, relative to the results of the experiment summarised in Table 4.5; this is probably due to the fact that units that are fully efficient in these experiments have a limited influence on the construction of the COLS frontier. Lastly, the DEA accuracy scores appear to have improved by a larger margin than the GA scores, but on the whole, the difference is minor.

Table 4.7: S1.2 summary scores with fully efficient DMUs

	MAD	MSE	TMAD
GA	2.5%	1.34	9.6%
COLS	1.7%	0.60	4.7%
DEA	0.9%	0.34	5.5%

S1.3: 12% average inefficiency, noise~N(0,0.2)

The purpose of this experiment is to determine the relative accuracy of each approach when a substantial amount of noise is present in the data. Since a stochastic element is now part of the data generating process, the analysis also assesses the performance of the SFA-derived productivity estimates. As the table below illustrates, the findings of this analysis are not very encouraging for any of the approaches considered.

Table 4.8: S1.3 summary accuracy scores for the default volatility assumptions

	MAD	MSE	TMAD	Correlation	Rank correlation
GA	22.5%	79.52	62.0%	57.1%	51.9%
COLS	22.4%	78.87	61.7%	57.2%	51.9%
DEA	22.5%	79.08	61.4%	57.1%	51.7%
SFA (exponential)	12.5%	27.80	40.0%	50.2%	43.9%
SFA (half-normal)	13.7%	32.13	41.3%	52.9%	44.8%

All of the deterministic approaches are revealed to be quite inaccurate under conditions of severe noise. COLS appears to be the most accurate of all the deterministic approaches, but the difference in the three main accuracy scores between both GA and DEA is marginal and assessed as statistically insignificant. In any case, all deterministic approaches display an average MAD score of approximately 22.5%, which could be considered unacceptably high for policy purposes.

The stochastic approaches perform significantly better than their deterministic counterparts, which was not unexpected. Even so, the perfectly specified SFA-exponential model (which correctly assumes that technical inefficiency is exponentially distributed) displays a MAD score of 12.5%, which could also be considered too high for the purposes of productivity change measurement. It is interesting to note that the incorrectly specified SFA-half-normal model comes second in overall accuracy. While the difference in the MAD scores between the SFA exponential and half-normal models is statistically significant, it is quite small, which suggests that the impact of misspecification in the inefficiency distribution could be relatively mild.

Another interesting point arising from these experiments is that the correlations between the estimates of productivity change and their true values are quite low, compared to those observed in the S1 experiments. This is true for all approaches examined, including the better performing SFA models; in fact, the correlation coefficients for the SFA models are actually smaller than those derived from the deterministic approaches, although by only a relatively small margin.

An almost identical picture emerges when fully efficient units are included in the analysis, with no statistical significance in the accuracy scores of the GA, DEA and COLS approaches.

Table 4.9: S1.3 model summary scores with fully efficient DMUs

	MAD	MSE	TMAD	Correlation	Rank correlation
GA	22.4%	79.00	61.7%	58.5%	53.2%
COLS	22.4%	78.75	61.7%	58.6%	53.4%
DEA	22.5%	79.09	61.5%	58.6%	53.1%
SFA (exponential)	12.5%	28.37	41.0%	54.9%	47.7%
SFA (half-normal)	14.3%	35.47	43.2%	55.9%	47.1%

And the results remain stable even when the volatility in inputs and prices is increased.

Table 4.10: S1.3 summary accuracy scores for the increased volatility assumptions

	MAD	MSE	TMAD	Correlation	Rank correlation
GA	22.9%	82.76	63.9%	55.4%	49.2%
COLS	22.5%	79.75	62.5%	56.3%	49.8%
DEA	23.1%	83.16	63.2%	55.2%	48.7%
SFA (exponential)	12.5%	28.11	40.0%	59.9%	50.8%
SFA (half-normal)	13.9%	33.43	42.0%	61.9%	50.9%

Once again, the scores of GA, DEA and COLS in the three main accuracy measures are statistically indistinguishable.

S1.4: 12% average inefficiency, noise~N(0,0.05)

The previous experiment introduced a substantial amount of noise in the data. For this experiment, the overall accuracy of the approaches is examined when noise is less severe.

Table 4.11: S1.4 summary accuracy scores for the default volatility assumptions

	MAD	MSE	TMAD	Correlation	Rank correlation
GA	5.8%	5.30	16.3%	93.8%	89.9%
COLS	5.7%	5.06	15.9%	94.0%	90.1%
DEA	5.8%	5.34	16.2%	93.7%	89.5%
SFA (exponential)	5.0%	4.04	14.2%	94.3%	90.3%
SFA (half-normal)	5.4%	4.64	15.1%	94.8%	90.6%

As with the previous experiment, both SFA models are assessed as the most accurate when a stochastic element is present in the data generating process. It is interesting to note however that the performance of the deterministic approaches is not lagging far behind relatively to the stochastic approaches, although the differences in all three main accuracy measures considered are statistically significant. COLS estimates are still the more accurate from the three determinist approaches, although the differences in scores between COLS and the DEA and GA approaches are only statistically significant under the non-parametric statistical tests. The DEA and GA scores are very similar in all measures and are in fact statistically indistinguishable under both parametric and non-parametric tests.

The effect of misspecification in the efficiency distribution for the SFA models is still present (similar to the previous experiment), and although the difference in the accuracy scores between the SFA exponential and half-normal models is statistically significant, it could still be considered relatively small.

Once more, there is no change in the overall findings when fully efficient units are included in the dataset.

Table 4.12: S1.4 summary accuracy scores with fully efficient DMUs

	MAD	MSE	TMAD	Correlation	Rank correlation
GA	5.8%	5.24	15.9%	93.8%	89.9%
COLS	5.6%	4.99	15.5%	94.1%	90.1%
DEA	5.8%	5.27	16.0%	93.8%	89.5%
SFA (exponential)	4.9%	3.81	13.8%	94.4%	90.4%
SFA (half-normal)	5.4%	4.54	14.9%	95.0%	90.8%

When the volatility in inputs and prices is increased for one period to the next, the overall findings change slightly.

Table 4.13: S1.4 summary accuracy scores for the increased volatility assumptions

	MAD	MSE	TMAD	Correlation	Rank correlation
GA	6.3%	6.53	18.5%	92.2%	88.5%
COLS	5.8%	5.17	15.9%	93.7%	89.7%
DEA	6.0%	5.66	16.7%	93.0%	88.8%
SFA (exponential)	5.0%	3.96	14.1%	94.1%	90.1%
SFA (half-normal)	5.4%	4.67	15.2%	94.7%	90.6%

Both stochastic models are again assessed as being the most accurate under all measures, with the correctly specified SFA exponential model being ranked as the most accurate. However, there is a somewhat clearer differentiation between the deterministic approaches; COLS is found to be the most accurate of the deterministic approaches in all measures and the statistical tests show that the difference in accuracy scores is statistically significant. DEA is the second most accurate of the deterministic approaches and again, the statistical tests confirm that the difference between DEA and GA scores is statistically significant. In summary, although the overall rankings have not changed substantially, there is a clearer hierarchy amongst the approaches considered.

Summary results for the S1 simulation experiments

In general, the analysis found that the most accurate approaches in the simulation experiments that adopted a Cobb-Douglas production to generate output values are the parametric approaches, namely COLS when a stochastic element was not included in the analysis and SFA otherwise. This is not an unexpected result, since the parametric models that are ranked highest in each experiment are perfectly specified, in that they utilise the same functional form as the adopted production function of the DGP and, in the case of the best-performing SFA models, assume the correct distribution for the inefficiency term.

What is likely of more interest is how the two non-parametric, determinist approaches have fared over this set of experiment. Firstly, the overall performance of GA was surprisingly robust, even if the approach displayed the worst (or joint worst accuracy) in the majority of the experiments. In most cases, the difference in accuracy scores between GA and DEA was quite small and for the experiments that included measurement error, the differences were statistically insignificant. The analysis however identified some conditions where the accuracy of the GA quickly deteriorates:

- It appears that as technical inefficiency becomes more prevalent in the data that include no noise, the accuracy of the GA estimates rapidly deteriorates. In the S1.2 experiment, GA ranked last, with inaccuracy scores that were substantially higher than both COLS and DEA.

- When volatility in inputs and input prices increases from one period to the next, it appears that the accuracy of the GA deteriorates at a faster rate than the other approaches. This features in almost all the experiments that utilise a Cobb-Douglas production function (the exception is experiment S1.3, which includes both relatively high technical inefficiency and measurement error levels, where the overall accuracy of all approaches considered does not change when volatility is increased).

As for the performance of DEA, the analysis raises two major points:

- DEA is the most accurate approach based on the MAD measure when technical inefficiency is found at relatively high levels in the data that also do not include any noise. This is a rather surprising result, since as was mentioned above, the COLS model that is also assessed in the relevant experiment (S1.2) is perfectly specified, given that the S1.2 data are constructed using a Cobb-Douglas functional form and contain no noise. And indeed, the COLS approach is more accurate than DEA in this experiment based on the TMAD measure and equally accurate based on the MSE measure, which suggests that the performance of COLS is better for the units that occupy outlying positions in the dataset.

– The accuracy of the DEA-based estimates decreases at a lower pace relatively to the accuracy of the other deterministic approaches when inputs and input prices become more volatile from one period to the next, in the experiments that do not include any measurement error (ie S1.1 and S1.2).

In addition to the points made above, some more general comments can be made when considering the analysis as a whole:

– When technical inefficiency is modest, there is no stochastic element in the DGP and the input levels and prices between subsequent periods are relatively stable, all approaches provide quite accurate estimates of productivity change.

– Increased volatility in inputs and prices in subsequent periods adversely affects the accuracy of all approaches, when no stochastic element is included in the DGP. The DEA estimates are the least affected, while the GA estimates are the most affected. Interestingly, when measurement error is introduced in the analysis, the increased volatility appears to have very little impact on the accuracy of the deterministic approaches and almost no impact at all on the accuracy of the stochastic approaches.

– When noise is present in the dataset, the SFA approaches provide the most accurate estimates. However, when measurement error is more severe, even the best performing SFA model demonstrates quite large deviations from the true productivity change values (MAD scores of approximately 12.5%). In addition, when measurement error is moderate, the gains in accuracy achieved by the SFA models are quite modest compared to the deterministic approaches (eg GA and DEA MAD scores are 5.8%, while the best performing SFA model has a MAD score of 5% in S1.4).

The tables below provide a summary of the three main accuracy measures from the S1 experiments for all the assessed approaches, as well as the relative accuracy rankings of each approach, taking into account the results of the statistical tests for the difference in mean accuracy estimates.

Table 4.14: Summary results for the first category of experiments (S1), default volatility assumptions

	Mean absolute deviation (in %)					Mean square error					'Top' Mean absolute deviation (in %)				
	GA	COLS	DEA	SFA (exponential)	SFA (half-normal)	GA	COLS	DEA	SFA (exponential)	SFA (half-normal)	GA	COLS	DEA	SFA (exponential)	SFA (half-normal)
S1.1: 12% average inefficiency, no noise	0.9%	0.4%	0.7%			0.21	0.04	0.20			4.1%	1.3%	4.1%		
S1.2: 32% average inefficiency, no noise	2.8%	1.7%	1.2%			1.57	0.62	0.46			10.1%	4.8%	5.9%		
S1.3: 12% average inefficiency, noise~N(0,0.2)	22.5%	22.4%	22.5%	12.5%	13.7%	79.52	78.87	79.08	27.80	32.13	62.0%	61.7%	61.4%	40.0%	41.3%
S1.4: 12% average inefficiency, noise~N(0,0.05)	5.8%	5.7%	5.8%	5.0%	5.4%	5.30	5.06	5.34	4.04	4.64	16.3%	15.9%	16.2%	14.2%	15.1%

Table 4.15: Relative accuracy rankings for the first category of experiments (S1), default volatility assumptions

	Mean absolute deviation (in %)					Mean square error					'Top' Mean absolute deviation (in %)				
	GA	COLS	DEA	SFA (exponential)	SFA (half-normal)	GA	COLS	DEA	SFA (exponential)	SFA (half-normal)	GA	COLS	DEA	SFA (exponential)	SFA (half-normal)
S1.1: 12% average inefficiency, no noise	3	1	2			3	1	2			2	1	2		
S1.2: 32% average inefficiency ,no noise	3	2	1			3	1	1			3	1	2		
S1.3: 12% average inefficiency, noise~N(0,0.2)	4	4	4	1	2	4	4	4	1	2	4	4	4	1	2
S1.4: 12% average inefficiency, noise~N(0,0.05)	4	3	4	1	2	4	3	4	1	2	4	3	4	1	2

Note: The above rankings take into consideration the results of the statistical tests for the difference in mean accuracy scores

Table 4.16: Simulation summary results for the first category of experiments (S1), increased volatility assumptions

	Mean absolute deviation (in %)					Mean square error					'Top' Mean absolute deviation (in %)				
	GA	COLS	DEA	SFA (exponential)	SFA (half-normal)	GA	COLS	DEA	SFA (exponential)	SFA (half-normal)	GA	COLS	DEA	SFA (exponential)	SFA (half-normal)
S1.1: 12% average inefficiency, no noise	2.5%	0.8%	1.2%			1.70	0.16	0.50			11.8%	2.7%	6.1%		
S1.2: 32% average inefficiency ,no noise	7.6%	3.0%	2.2%			12.92	1.92	1.99			30.6%	9.3%	12.7%		
S1.3: 12% average inefficiency, noise~N(0,0.2)	22.9%	22.5%	23.1%	12.5%	13.9%	82.76	79.75	83.16	28.11	33.43	63.9%	62.5%	63.2%	40.0%	42.0%
S1.4: 12% average inefficiency, noise~N(0,0.05)	6.3%	5.8%	6.0%	5.0%	5.4%	6.53	5.17	5.66	3.96	4.67	18.5%	15.9%	16.7%	14.1%	15.2%

Table 4.17: Relative accuracy rankings for the first category of experiments (S1), increased volatility assumptions

	Mean absolute deviation (in %)					Mean square error					'Top' Mean absolute deviation (in %)				
	GA	COLS	DEA	SFA (exponential)	SFA (half-normal)	GA	COLS	DEA	SFA (exponential)	SFA (half-normal)	GA	COLS	DEA	SFA (exponential)	SFA (half-normal)
S1.1: 12% average inefficiency, no noise	3	1	2			3	1	2			3	1	2		
S1.2: 32% average inefficiency ,no noise	3	2	1			3	1	2			3	1	2		
S1.3: 12% average inefficiency, noise~N(0,0.2)	4	4	4	1	2	4	4	4	1	2	4	4	4	1	2
S1.4: 12% average inefficiency, noise~N(0,0.05)	5	3	4	1	2	5	3	4	1	2	5	3	4	1	2

Note: The above rankings take into consideration the results of the statistical tests for the difference in mean accuracy scores

4.4.2. S2 simulation experiments

As was previously mentioned, these experiments adopt a piecewise-linear production function for the data generating process.

S2.1: 12% average inefficiency, no noise

This experiment is identical in its parameters to S1.1, except that now a piecewise-linear function is used to generate output values. The summary findings from this experiment are presented in the table below:

Table 4.18: S2.1 summary accuracy scores for the default volatility assumptions

	MAD	MSE	TMAD	Correlation	Rank correlation
GA	0.9%	0.20	4.1%	99.7%	99.5%
COLS	2.4%	1.08	8.2%	98.6%	97.2%
COLS (translog)	2.9%	70.31	21.6%	95.4%	96.8%
DEA	0.8%	0.22	4.1%	99.7%	99.4%

The analysis reveals that the DEA and GA estimates are the most accurate, according to the three main measures, from all the approaches examined by a comfortable margin (that is also statistically significant). The DEA and GA MAD and TMAD scores are almost identical and the slightly better performance of GA in the MSE measure is too small to be statistically significant.

Probably the most interesting finding of the analysis however is that functional form misspecification can have a very significant impact on the accuracy of the parametric approaches. The COLS estimates, which were the most accurate in the S1.1 experiment are now revealed to be the least accurate; since the only thing that is different from the S1.1 experiment is the functional form used to generate output values, this appears to be the most likely source of the deterioration of the COLS overall accuracy. The accuracy scores for S1.1 are shown in the following table, to facilitate the comparisons.

Table 4.18b: Combined S1.1 and S2.1 summary accuracy scores for the default volatility assumptions

		MAD	MSE	TMAD
	S2.1	0.9%	0.20	4.1%
GA	S1.1	0.9%	0.21	4.1%
	S2.1	2.4%	1.08	8.2%
COLS	S1.1	0.4%	0.04	1.3%
	S2.1	0.8%	0.22	4.1%
DEA	S1.1	0.7%	0.20	4.1%

So, while the accuracy of the DEA and GA estimates has remained relatively constant, the MAD scores for the COLS estimates have increased from 0.4% to 2.4%, while TMAD scores have increased from 1.3% to 8.5%.

Another interesting finding from this analysis is that the COLS translog model overall accuracy scores are worse than those from the COLS Cobb-Douglas model. This is more apparent when examining the MSE and TMAD scores which measure the accuracy of the approach at the 'edges' of the sample. Based on these scores, it can be stated that the translog model has difficulties in correctly measuring productivity growth for at least a subset of the assessed units. It should be repeated here that the analysis examined the performance of the translog model on the hypothesis that the translog is a flexible functional form and could, in theory, provide a better fit to the underlying piecewise-linear function that was used to generate the output. As will be demonstrated in the following sections, the relative inaccuracy of the translog specification persists in most of the S2 simulations. Since this is a significant finding (given that the translog specification is one of the most widely adopted functional forms in the econometric analysis of efficiency and productivity), a more thorough discussion on this issue is be provided at the end of this section.

Including fully efficient units in the dataset has no material impact on the results, as the following table demonstrates.

Table 4.19: S2.1 summary accuracy scores with fully efficient DMUs

	MAD	MSE	TMAD	Correlation	Rank correlation
GA	0.8%	0.21	4.1%	99.7%	99.5%
COLS	2.5%	1.19	8.9%	98.1%	97.1%
COLS (translog)	3.1%	90.89	29.7%	93.1%	94.4%
DEA	0.7%	0.22	4.0%	99.6%	99.4%

In general, the inclusion of fully efficient units has very little impact on the results of all of the S2 simulation experiments; as such, the results from this permutation will not be reported from now on.

When volatility in input and price levels is increased, all accuracy measures are inflated (signifying an overall decrease in accuracy) by a significant margin, as the table below demonstrates.

Table 4.20: S2.1 summary accuracy scores for the increased volatility assumptions

	MAD	MSE	TMAD	Correlation	Rank correlation
GA	2.5%	1.79	12.2%	97.8%	97.0%
COLS	6.4%	8.31	23.2%	90.9%	87.2%
COLS (translog)	7.0%	145.60	38.9%	88.2%	90.2%
DEA	1.5%	0.68	6.8%	99.1%	98.3%

The results show that the DEA-based estimates are the most accurate in all three main measures from all the approaches examined by a comfortable margin (that is also statistically significant). The GA estimates are ranked second in overall accuracy, trailing the DEA estimates in all accuracy measures considered. The COLS translog model still produces relatively inaccurate estimates.

S2.2: 32% average inefficiency, no noise

This experiment is identical in its parameters to a previous experiment, S1.2, except that a piecewise-linear function is used for the DGP rather than a Cobb-Douglas function. The summary findings from this experiment with the default input volatility assumptions are presented in the table below:

Table 4.21: S2.2 summary accuracy scores for the default volatility assumptions

	MAD	MSE	TMAD	Correlation	Rank correlation
GA	2.2%	1.09	8.8%	99.9%	99.7%
COLS	2.4%	1.20	8.8%	99.9%	99.6%
COLS (translog)	4.9%	132.12	32.1%	97.2%	98.8%
DEA	1.1%	0.39	5.4%	100.0%	99.8%

The table shows that the DEA-based estimates are the most accurate according to all the measures, followed by the GA estimates. The performance of the Cobb-Douglas COLS model is quite close to GA, although the small differences are in fact found to be statistically significant. It is also interesting to note that the accuracy scores of this model are very similar to those from the previous simulation experiment, despite the larger amount of inefficiency in the DGP included in the current experiment. The DEA-based productivity estimates also seem to be only marginally affected by the larger inefficiency in the dataset; on the other hand, the GA estimates are quite heavily affected, which is expected given that GA assumes no inefficiency in the production process.

Similar to the previous experiment, the COLS translog model is clearly the worst performer, with MAD scores almost twice as high as those from the COLS Cobb-Douglas model and TMAD scores that are almost three times as high.

The overall accuracy rankings of each approach do not change, when the increased volatility assumptions are introduced to the data generation process, as the table below demonstrates:

Table 4.22: S2.2 summary accuracy scores for the increased volatility assumptions

	MAD	MSE	TMAD	Correlation	Rank correlation
GA	5.7%	8.04	24.7%	99.2%	98.7%
COLS	6.3%	8.14	23.4%	99.1%	98.2%
COLS (translog)	8.3%	98.90	49.1%	96.1%	97.2%
DEA	2.2%	1.45	10.2%	99.8%	99.6%

As with almost all previous experiments, increased volatility reduces the accuracy of all estimates, but to a different degree for each approach, with DEA appearing to produce the more robust estimates. DEA is ranked first in terms of overall accuracy in this experiment, followed by GA. Relative to the S2.1 experiment, the difference in accuracy is even more pronounced between DEA and GA. As with the default assumptions, the accuracy scores for COLS are relatively stable, in comparison with the S2.1 experiment (with increased volatility assumptions).

S2.3: 12% average inefficiency, noise~N(0,0.05)

This and the following experiments introduce a stochastic element in the data generating process that can represent the effects of luck or measurement error in the available data. Since a stochastic element is present in the data, these experiments also assess the accuracy of the SFA-based productivity change estimates.

For this experiment, both the noise element and technical inefficiency are at moderate levels; later experiments increase either or both to higher levels. The table below presents the summary accuracy measures for the approaches considered.

Table 4.23: S2.3 summary accuracy scores for the default volatility assumptions

	MAD	MSE	TMAD	Correlation	Rank correlation
GA	5.8%	5.33	16.2%	94.0%	89.9%
COLS	6.3%	6.24	17.6%	93.0%	88.5%
COLS (translog)	6.5%	7.81	20.6%	91.8%	88.1%
DEA	5.8%	5.23	16.0%	94.0%	89.8%
SFA	7.1%	9.11	21.6%	88.2%	84.1%
SFA (translog)	6.1%	6.12	17.7%	92.2%	87.3%
SFA (half-normal)	6.4%	6.96	18.5%	90.3%	85.6%

The most striking finding from this experiment is that the stochastic models assessed are not the most accurate ones. Instead, the most accurate approaches appear to be DEA and GA, with DEA doing marginally better, although the difference in accuracy scores is not statistically significant. Even the deterministic COLS Cobb-Douglas model is assessed as being more accurate than the SFA Cobb-Douglas model, which indicates that functional form misspecification has a more severe impact on the accuracy of the stochastic parametric approaches relative to their deterministic counterpart.

Interestingly, the translog SFA (exponential) model appears to be the most accurate of all the parametric models; this is a departure from the previous results, which found that the translog specification can sometimes cause the model to produce widely inaccurate productivity change estimates. The statistical tests find that this model's accuracy scores are indeed statistically significantly smaller than those from the Cobb-Douglas SFA (exponential) model, but not from the Cobb-Douglas COLS model or from the wrongly specified Cobb-Douglas SFA model that assumes a half-normal distribution for the inefficiency component.

In fact this last point is another interesting finding that comes from this set of experiments; the miss-specified, in terms of inefficiency distribution assumptions, half-normal SFA model outperforms the correctly specified exponential SFA model, by a wide margin that is statistically significant for the three main accuracy measures. Again, this can be taken as additional evidence of how important is the functional form assumption to the overall accuracy of the SFA models.

A very similar picture emerges when the volatility in inputs and prices is increased in consequent periods.

Table 4.24: S2.3 summary accuracy scores for the increased volatility assumptions

	MAD	MSE	TMAD	Correlation	Rank correlation
GA	6.3%	6.79	18.9%	92.4%	89.2%
COLS	8.5%	12.36	26.6%	87.2%	82.5%
COLS (translog)	8.9%	292.19	46.0%	85.6%	84.8%
DEA	5.8%	5.45	16.6%	93.5%	89.8%
SFA	9.1%	14.91	29.0%	79.0%	74.2%
SFA (translog)	7.4%	10.09	23.6%	86.9%	81.9%
SFA (half-normal)	8.4%	12.42	26.4%	82.9%	78.0%

Once again, the SFA models display lower overall accuracy relative to the non-parametric deterministic approaches. In this setting, the most accurate approach appears to be DEA, followed by GA, and in this case, the difference in the all of the three main accuracy criteria is statistically significant between these two approaches. It is interesting to note the resilience of the DEA accuracy scores under conditions of greater input volatility in this setting; while all other approaches see increased MAD scores by approximately two percentage points, the DEA score remains practically unchanged between the two experiments.

For the parametric approaches, the translog SFA (exponential) model provides the most accurate estimates. As with the experiment with default input volatility, the Cobb-Douglas half-normal SFA model performs better than the Cobb-Douglas exponential SFA, despite the fact that the latter model assumes the correct inefficiency distribution.

S2.4: 12% average inefficiency, noise~N(0,0.2)

For this experiment the stochastic element is more prominent, representing severe measurement error, while technical inefficiency is kept at moderate levels.

Table 4.25: S2.4 summary accuracy scores for the default volatility assumptions

	MAD	MSE	TMAD	Correlation	Rank correlation
GA	22.9%	86.21	67.7%	56.6%	50.8%
COLS	23.0%	86.93	67.8%	56.6%	50.6%
COLS (translog)	23.4%	95.96	71.3%	55.4%	50.2%
DEA	22.8%	85.38	67.4%	56.6%	50.8%
SFA	14.3%	38.41	48.8%	50.4%	43.6%
SFA (translog)	15.4%	43.70	51.4%	53.6%	46.1%
SFA (half-normal)	16.6%	49.76	52.2%	52.0%	45.4%

The findings of this analysis show that under severe measurement error conditions, the SFA models produce the more accurate productivity change estimates. All three SFA models are found in the top three places in the accuracy rankings, which suggests that severe measurement error is a stronger source of bias than any form of misspecification. DEA, GA and the Cobb-Douglas COLS specification display similar scores in all accuracy measures and the statistical tests reveal that the marginal differences between them are too small to be considered significant. Similar to the other S2 experiments, the COLS translog estimates are ranked last in terms of overall accuracy.

With regards to the three SFA models, the correctly specified (in terms of the inefficiency distribution assumption) Cobb-Douglas exponential SFA model is the most accurate, followed by the also correctly specified translog SFA model. The differences in the accuracy scores between these approaches suggests that under conditions of severe measurement error, the misspecification of the inefficiency distribution in the SFA models can have a significant impact to the overall accuracy of the resulting estimates. Also of note here are the relatively low correlations between the estimates of productivity change from all approaches and true productivity change; this was also the case for the S1.3 experiment, which suggests that the ratio of noise to inefficiency also significantly affects the relative accuracy of the approaches examined.

The relative accuracy of the examined approaches remains unchanged when the volatility in inputs and prices is increased, as the following table demonstrates:

Table 4.26: S2.4 summary accuracy scores for the increased volatility assumptions

	MAD	MSE	TMAD	Correlation	Rank correlation
GA	23.2%	89.70	69.8%	54.3%	50.0%
COLS	24.0%	95.39	71.9%	53.7%	49.8%
COLS (translog)	24.6%	137.09	80.6%	51.8%	49.7%
DEA	23.0%	87.73	68.8%	54.7%	50.4%
SFA	14.6%	40.00	50.2%	48.1%	44.5%
SFA (translog)	15.5%	44.74	52.2%	50.0%	45.4%
SFA (half-normal)	17.0%	52.04	54.3%	51.3%	47.0%

As with the default conditions, the SFA models dominate in this setting, followed by DEA and GA (which are statistically indistinguishable), while both COLS specifications occupy the last places in the rankings.

It is interesting to note that the increased volatility in this experiment does not have the usual negative effect on the overall accuracy of the approaches, as evident in the experiments described so far. This is probably due to the fact that the variance introduced in the analysis from the presence of substantial levels of noise greatly overshadows the effect of the increased input volatility. This is another indication of how significant the presence of noise is to the accuracy of all the approaches examined.

Overall, when the stochastic element dominates, all approaches produce productivity change estimates that could be considered unacceptably inaccurate, as evidenced by the high scores in all three main accuracy measures and the low correlations with the true values of productivity change.

S2.5: 32% average inefficiency, noise~N(0,0.2)

In this experiment, the data include substantial levels of both the stochastic element and technical inefficiency.

It should be mentioned here that the analysis encountered some difficulties with the SFA estimation. In a number of cases the initial exploratory analysis before the estimation of the SFA model revealed that the residuals had the wrong skew, which SFA interprets as strong evidence that there is no technical inefficiency in the assess units. In these instances, the software used to run the analysis abandons the SFA estimation and reverts to using a simple OLS regression. This error was relatively frequent, as it was encountered in approximately 1 in 20 simulation runs. Since this analysis is more interested in the accuracy of the SFA estimates, these results were discarded and the experiment was modified so that enough simulation runs were undertaken in order to have a full set of 100 runs in which an SFA model was able to be estimated.

Table 4.27: S2.5 summary accuracy scores for the default volatility assumptions

	MAD	MSE	TMAD	Correlation	Rank correlation
GA	23.6%	91.33	69.1%	91.5%	87.0%
COLS	23.6%	91.50	69.5%	91.5%	87.0%
COLS (translog)	24.8%	148.54	82.0%	89.4%	86.4%
DEA	23.4%	89.94	68.7%	91.5%	87.1%
SFA	21.5%	76.76	63.6%	91.9%	87.1%
SFA (translog)	22.2%	82.49	66.4%	91.8%	87.1%
SFA (half-normal)	22.9%	86.35	67.4%	91.7%	87.1%

The table above demonstrates that the Cobb-Douglas exponential SFA model is assessed as the most accurate approach in this experiment, followed by the other SFA models. The statistical tests however reveal that the accuracy scores of all three SFA models are not statistically significantly different. The same applies for the accuracy scores for the deterministic GA, DEA and Cobb-Douglas COLS models; their differences are too small to be assessed as statistically significant. And although the difference between the best performing SFA model and the best performing deterministic model (which is DEA in this case) is statistically significant, it is also quite small (just 2 percentage points) given the overall inaccuracy of the produced estimates.

Overall, the performance of the deterministic approaches is very similar to the performance observed in the previous experiment (S2.4), which suggests that increased levels of technical inefficiency do not introduce additional bias in the estimates under conditions of severe noise. This is not the case or the SFA specifications, as there is a marked decrease in overall accuracy relative to the previous experiment.

Another interesting point is that while the deviations between the estimates and the true values of productivity change are quite substantial, the correlations between the estimates and the true values are relatively high, for all approaches. Similarly high deviations from the true values were observed in the previous experiment (S2.4), but both mean correlations measures were also relatively low.

The reason these differences between the deviation and correlation measures in this experiment is unclear at this stage; given that the findings of the previous experiments suggests that increasing levels of technical inefficiency negatively impact the overall accuracy of the measurement approaches, it would be reasonable to expect a deterioration of performance in this experiment relative to the previous one, which included lower levels of technical inefficiency on average.

The overall findings are very similar when the volatility in inputs and prices is increased:

Table 4.28: S2.5 summary accuracy scores for the increased volatility assumptions

	MAD	MSE	TMAD	Correlation	Rank correlation
GA	24.5%	100.24	73.2%	90.7%	86.3%
COLS	24.8%	101.41	72.9%	90.5%	86.0%
COLS (translog)	26.2%	157.38	89.5%	88.0%	85.1%
DEA	23.8%	93.87	70.3%	91.1%	86.6%
SFA	22.7%	85.44	67.3%	90.9%	85.9%
SFA (translog)	23.4%	92.37	70.2%	90.7%	85.7%
SFA (half-normal)	23.8%	94.64	70.7%	90.7%	86.1%

The estimates from the SFA exponential specification remain the most accurate, with the rest of the SFA specifications and the DEA model following. In fact, the DEA-based estimates are statistically indistinguishable from the translog SFA and half-normal SFA; this once again demonstrates the resilience of the DEA estimates in high input volatility conditions. Overall, the accuracy scores of all approaches under increased volatility assumptions are not very different in absolute values to those from the experiments with default volatility assumptions. This is similar to the behaviour observed in the S2.4 experiment and suggests that in dataset with high noise levels, the negative effect of input volatility is greatly diminished.

Probably the most important finding of this round of simulations is that no approach is able to produce reliable estimates when both noise and technical inefficiency are prominent in the data.

S2.6: 32% average inefficiency, noise~N(0,0.05)

For this last experiment, the levels of technical inefficiency are kept at relatively high levels, while the impact of the stochastic element is reduced to more moderate levels. As with the previous experiment, similar issues arose with the SFA estimation; as before, the problematic datasets where discarded and the analysis redrew additional datasets so that the experiment included SFA results from 100 successful simulation runs.

Table 4.29: S2.6 summary accuracy scores for the default volatility assumptions

	MAD	MSE	TMAD	Correlation	Rank correlation
GA	6.2%	6.09	17.5%	99.3%	98.6%
COLS	6.2%	6.12	17.4%	99.3%	98.6%
COLS (translog)	8.7%	287.07	47.2%	95.9%	97.8%
DEA	5.8%	5.33	16.3%	99.4%	98.7%
SFA	7.1%	8.56	20.4%	99.1%	98.4%
SFA (translog)	6.5%	6.77	18.6%	99.3%	98.5%
SFA (half-normal)	6.2%	6.16	17.4%	99.3%	98.5%

Despite the fact that the DGP contains a stochastic element, it is the deterministic approaches that are assessed as being more accurate in this experiment. More specifically, DEA is revealed to be the most accurate approach under these conditions, followed by GA, the COLS Cobb-Douglas model and the SFA half-normal model. The statistical tests carried out reveal that the very small differences in accuracy scores between GA, COLS and SFA (half-normal) specifications are not statistically significant. These results are consistent with the findings from the S2.3 experiment, which assumed the same level of noise but lower levels of technical inefficiency. It is interesting to note that the miss-specified (in terms of the inefficiency distribution assumption) half-normal SFA model is deemed to be more accurate than the correctly specified SFA exponential model; this was also the case in the S2.3 experiment.

The overall accuracy rankings remain relatively stable when volatility in inputs and prices is increased.

Table 4.30: S2.6 summary accuracy scores for the increased volatility assumptions

	MAD	MSE	TMAD	Correlation	Rank correlation
GA	8.4%	12.93	28.2%	98.7%	97.8%
COLS	8.7%	13.04	27.2%	98.7%	97.3%
COLS (translog)	10.6%	202.80	55.4%	95.5%	96.8%
DEA	6.2%	6.36	18.5%	99.3%	98.5%
SFA	10.0%	17.27	30.3%	98.3%	97.0%
SFA (translog)	8.4%	13.28	28.6%	98.7%	97.3%
SFA (half-normal)	8.8%	13.34	27.8%	98.6%	97.3%

DEA still remains the most accurate approach overall. GA is the second most accurate approach based on the MAD measure, but its MSE and TMAD scores are statistically indistinguishable from the COLS and SFA (half-normal) specifications. The COLS Cobb-Douglas specification together with the SFA half-normal specification are ranked jointly third, followed by the translog SFA exponential model. As with almost all previous experiments, increased volatility leads to an overall deterioration of accuracy for all the examined approaches, but to a different degree; DEA remains the most robust approach under conditions of increased volatility.

Perhaps the most important finding from this experiment is that when the stochastic element is relatively modest, the deterministic approaches provide more (or at least as) accurate estimates of productivity change relative to the various SFA specifications examined for this experiment.

Summary results for the S2 simulation experiments

While the parametric approaches dominated in terms of accuracy in the first set of experiments, where the datasets were generated based on a Cobb-Douglas production function, it is the non-parametric approaches that appear as more accurate on the whole when the dataset is generated using a piecewise-linear function. This was not unexpected, given that the underlying production function in the second set of experiments is not a perfect match with the functional form adopted by the parametric approaches examined; this functional form misspecification was bound to have a negative effect on the overall accuracy of the parametrically-derived estimates.

What was probably unexpected was the magnitude of this negative effect. When no measurement error is present, the COLS Cobb-Douglas specification displays MAD scores that are at least twice as large as those displayed by the DEA estimates and the discrepancy in MSE scores is significantly bigger (at least three times higher).

Furthermore, the overall accuracy of the COLS specification that adopts a translog functional form is even worse; in fact, in all experiments, the COLS translog specification was ranked last in terms of overall accuracy. On the other hand, the translog specification yields more accurate results when paired with SFA, but only for the experiments that included a 'moderate' noise component. When noise levels were elevated, the Cobb-Douglas SFA models were more accurate across all measures.

The overall performance of the translog specification is a somewhat puzzling result, since it was expected that the additional flexibility it afforded to the parametric models would allow for a better fit to the underlying piecewise-linear function. There are a number of possible reasons for the underperformance of the translog specifications:

– The sample size was too small; the translog specification takes up an additional 6 degrees of freedom relative to the Cobb-Douglas specification, which reduces the accuracy of the estimated coefficients.

– The translog might be too flexible; the underlying piecewise-linear production function is convex (in the sense that all input-output combinations belong to a convex set) and monotonic; the translog on the other hand is flexible in a sense that parts of the production function can be non-convex and non-monotonic. It is possible that the combined effects of inefficiency and noise in the data lead the estimation to wrongfully conclude that there are indeed parts of the production function that are either non-monotonic or non-convex, simply because they would provide a better fit to the data.

– The way the experiments applied the translog functional form was rather basic; the analysis did not apply any variable selection techniques, such as general-to-specific, so that variable that were deemed to be statistically insignificant remained in the model. Although econometric theory suggests that the presence of insignificant variables should not result in any form of bias[64], issues might arise due to the multi-collinearity that is frequently displayed between the translog variables.

Further research would be required to test whether any of the above factors is the actual reason for the relatively bad accuracy performance of the translog specifications. It is important to note however that the adoption of a flexible functional form does not always improve the fit of the parametric models. It is important to note however that the SFA models were found to be more accommodating to the translog specification, relative to the COLS models. This is probably due to the fact that the SFA models can 'assign' part of the distance of each unit to the apparent frontier to noise, thus reducing the probability of observing very large variations in the productivity estimates and/or lessening the impact of such variations if they happen to exist in a particular simulation run.

Another important issue revealed by the S2 simulations is the relative underperformance of the SFA models under conditions of moderate noise. As the S2.3 and S2.6 experiments revealed, when the standard deviation of the normally distributed stochastic element is 0.05, the non-parametric deterministic approaches (ie DEA and GA) perform better that the stochastic models, while COLS is almost as accurate as the most accurate of the SFA specifications. In addition, the SFA specification that (incorrectly) assumes that the inefficiency is half-normally distributed is more accurate than the correctly specified SFA exponential model, at least when the models adopted a Cobb-Douglas functional form. Only when measurement error is more severe (is the standard deviation is increased from 0.05 to 0.20), is the correctly specified SFA exponential model deemed to be most accurate. Even in these cases, the performance of said specification is significantly better than the next best deterministic approach only when the levels of technical inefficiency are relatively modest (see S2.4). And even so, with a MAD score of 14.7%, it would be difficult for any outside observer to label the performance of the SFA model as 'accurate'.

Another point that should be repeated here is the difficulties faced by the analysis when estimating the SFA models under conditions of relatively large technical inefficiency levels. The problem was that the skew of the residual of the affected models was wrong, which meant that the SFA estimation could not proceed.

---

[64] The estimated parameters would simply display larger confidence intervals, since the insignificant variables reduce the degrees-of-freedom available to the estimation.

Although the analysis circumvented this issue by discarding the problematic datasets, this would not be possible in real-world applications. As such, it should be noted as a possible weakness of the SFA approach in general that is not reflected in the quantitative results of this analysis.

In general, when measurement error becomes more prevalent in the data, no approach can produce accurate productivity estimates. This was also the case for the S1 experiments and suggests that additional research would be required to identify approaches that can produce robust estimates under these seemingly adverse conditions.

Overall, the S2 experiments revealed that the non-parametric, deterministic approaches provide reasonably accurate estimates under various conditions. In the case of GA, this is a somewhat surprising result, given that the approach does not acknowledge the presence of technical inefficiency, which is a not inconsiderable component of productivity change in these experiments. However, DEA is revealed to be the more accurate approach of the two, and, one could argue, the more accurate approach on the whole, in the S2 experiments.

Another advantage of the DEA-derived estimates is their apparent robustness under conditions of increased volatility in inputs and prices. The S2 simulations showed that increased volatility reduces the accuracy of all estimates, but to a different degree for each approach; the same experiments also showed that the DEA-based estimates are in the majority of cases the ones that are least affected.

A summary of the results from all S2 experiments under the default assumptions and their relative accuracy rankings are presented in the tables below:

Table 4.31: Simulation summary results for the second category of experiments (S2), default volatility assumptions

	Measure	GA	COLS	COLS (translog)	DEA	SFA	SFA (translog)	SFA (half-normal)
S2.1: 12% average inefficiency, no noise	MAD	0.90%	2.40%	2.90%	0.80%			
	MSE	0.2	1.08	70.31	0.22			
	TMAD	4.10%	8.20%	21.60%	4.10%			
S2.2: 32% average inefficiency, no noise	MAD	2.20%	2.40%	4.90%	1.10%			
	MSE	1.09	1.2	132.12	0.39			
	TMAD	8.80%	8.80%	32.10%	5.40%			
S2.3: 12% average inefficiency, noise~N(0,0.05)	MAD	5.80%	6.30%	6.50%	5.80%	7.10%	6.10%	6.40%
	MSE	5.33	6.24	7.81	5.23	9.11	6.12	6.96
	TMAD	16.20%	17.60%	20.60%	16.00%	21.60%	17.70%	18.50%
S2.4:,12% average inefficiency, noise~N(0,0.2)	MAD	22.90%	23.00%	23.40%	22.80%	14.30%	15.40%	16.60%
	MSE	86.21	86.93	95.96	85.38	38.41	43.7	49.76
	TMAD	67.70%	67.80%	71.30%	67.40%	48.80%	51.40%	52.20%
S2.5:, 32% average inefficiency, noise~N(0,0.2)	MAD	23.60%	23.60%	24.80%	23.40%	21.50%	22.20%	22.90%
	MSE	91.33	91.5	148.54	89.94	76.76	82.49	86.35
	TMAD	69.10%	69.50%	82.00%	68.70%	63.60%	66.40%	67.40%
S2.6:, 32% average inefficiency, noise~N(0,0.05)	MAD	6.20%	6.20%	8.70%	5.80%	7.10%	6.50%	6.20%
	MSE	6.09	6.12	287.07	5.33	8.56	6.77	6.16
	TMAD	17.50%	17.40%	47.20%	16.30%	20.40%	18.60%	17.40%

Table 4.32: Relative accuracy rankings for the second category of experiments (S2), default volatility assumptions

	Measure	GA	COLS	COLS (translog)	DEA	SFA	SFA (translog)	SFA (half-normal)
S2.1: 12% average inefficiency, no noise	MAD	1	3	4	1			
	MSE	1	3	4	1			
	TMAD	1	3	4	1			
S2.2: 32% average inefficiency, no noise	MAD	2	3	4	1			
	MSE	2	3	4	1			
	TMAD	2	2	4	1			
S2.3: 12% average inefficiency, noise~N(0,0.05)	MAD	1	5	5	1	7	3	5
	MSE	2	4	6	1	7	3	4
	TMAD	1	4	6	1	7	4	4
S2.4:,12% average inefficiency, noise~N(0,0.2)	MAD	5	5	7	4	1	2	3
	MSE	5	5	7	4	1	2	2
	TMAD	5	5	7	5	1	2	2
S2.5:, 32% average inefficiency, noise~N(0,0.2)	MAD	5	5	7	3	1	2	3
	MSE	5	5	7	3	1	2	3
	TMAD	5	5	7	3	1	2	3
S2.6:, 32% average inefficiency, noise~N(0,0.05)	MAD	3	3	7	1	6	5	3
	MSE	3	3	7	1	6	5	3
	TMAD	3	3	7	1	6	5	3

Table 4.33: Simulation summary results for the second category of experiments (S2), increased volatility assumptions

	Measure	GA	COLS	COLS (translog)	DEA	SFA	SFA (translog)	SFA (half-normal)
S2.1: 12% average inefficiency, no noise	MAD	2.5%	6.4%	6.0%	1.5%			
	MSE	1.79	8.31	145.60	0.68			
	TMAD	12.2%	23.2%	38.9%	6.8%			
S2.2: 32% average inefficiency, no noise	MAD	5.7%	6.3%	8.3%	2.2%			
	MSE	8.04	8.14	98.90	1.45			
	TMAD	24.7%	23.4%	49.1%	10.2%			
S2.3: 12% average inefficiency, noise~N(0,0.05)	MAD	6.3%	8.5%	8.9%	5.8%	9.1%	7.4%	8.4%
	MSE	6.79	12.36	92.19	5.45	14.91	10.09	12.42
	TMAD	18.9%	26.6%	46.0%	16.6%	29.0%	23.6%	26.4%
S2.4:,12% average inefficiency, noise~N(0,0.2)	MAD	23.2%	24.0%	24.6%	23.0%	14.6%	15.5%	17.0%
	MSE	89.70	95.39	137.09	87.73	40.00	44.74	52.04
	TMAD	69.8%	71.9%	80.6%	68.8%	50.2%	52.2%	54.3%
S2.5:, 32% average inefficiency, noise~N(0,0.2)	MAD	24.5%	24.8%	26.2%	23.8%	22.7%	23.4%	23.8%
	MSE	100.24	101.41	157.38	93.87	85.44	92.37	94.64
	TMAD	73.2%	72.9%	89.5%	70.3%	67.3%	70.2%	70.7%
S2.6:, 32% average inefficiency, noise~N(0,0.05)	MAD	8.4%	8.7%	10.6%	6.2%	10.0%	8.4%	8.8%
	MSE	12.93	13.04	202.80	6.36	17.27	13.28	13.34
	TMAD	28.2%	27.2%	55.4%	18.5%	30.3%	28.6%	27.8%

Table 4.34: Relative accuracy rankings for the second category of experiments (S2), increased volatility assumptions

	Measure	GA	COLS	COLS (translog)	DEA	SFA	SFA (translog)	SFA (half-normal)
S2.1: 12% average inefficiency, no noise	MAD	2	4	3	1			
	MSE	2	3	4	1			
	TMAD	2	3	4	1			
S2.2: 32% average inefficiency, no noise	MAD	2	3	4	1			
	MSE	2	2	4	1			
	TMAD	3	2	4	1			
S2.3: 12% average inefficiency, noise~N(0,0.05)	MAD	2	5	6	1	7	3	4
	MSE	2	4	7	1	6	3	5
	TMAD	2	5	7	1	6	3	4
S2.4:,12% average inefficiency, noise~N(0,0.2)	MAD	5	6	6	4	1	2	3
	MSE	5	6	6	4	1	2	3
	TMAD	4	6	7	4	1	2	2
S2.5:, 32% average inefficiency, noise~N(0,0.2)	MAD	5	5	7	3	1	3	3
	MSE	5	5	7	3	1	3	3
	TMAD	5	5	7	3	1	3	3
S2.6:, 32% average inefficiency, noise~N(0,0.05)	MAD	2	4	7	1	6	2	4
	MSE	4	4	7	1	6	4	4
	TMAD	4	4	7	1	6	4	4

4.5. Summary and conclusions

As was stated in the introduction of this chapter, the aim of this analysis is to provide quantitative evidence on the performance of GA and various frontier-based approaches under a number of conditions, the most important being the presence of inefficiency and noise (or measurement error), for the purposes of productivity growth measurement. To achieve this, the analysis undertook a number of Monte Carlo simulation experiments that were based on different conditions with regards to the distribution and severity of technical inefficiency in the assessed units and the presence and severity of noise in the data.

The starting point of the analysis was to formulate the production function(s) that form the backbone of data generation process (DGP). Two production functions were employed for this research, a two-input, one-output Cobb-Douglas function and a two-input, one-output piecewise-linear function (the two inputs are referred to as labour and capital, to ensure consistency with the GA framework[65]). The analysis used two production functions firstly to assess the effects that functional form misspecification in the parametric approaches and secondly to determine whether the departure from a more ‘standard’, smooth production function to a piecewise technology would affect the GA estimates.

For the various frontier-based approaches, information on inputs and output(s) is sufficient for the estimation of productivity change; however, GA requires information on prices for both inputs and output(s) in order to parameterise the production function, so price information consistent with the quantities of inputs used and outputs produced by each assessed unit was also generated.

One of the contributions of this research was to identify the exact structural relationship that provides this consistent link between the production function and its parameters to the input demand functions. This was done by relying on the findings of the duality theory of production, specifically the duality between the production function and the cost function If the analysis assumes that the assessed units are either pursuing cost minimisation or output maximisation (input minimisation), the duality theory provides the framework that links the production function and the cost

[65] The results of the analysis apply for any general two-input, one output production process with the same characteristics as those employed in each individual experiment, when considering the various frontier-based approaches. However, GA requires that the inputs of the analysis include at least the two primary factors of production, namely labour and capital.

function; further manipulating this relationship, the analysis was able to derive consistent formulas that link the parameters of the production function with the prices of the different inputs. This derivation is described in more detail in section 4.3.1.

Each simulation experiment examined the performance of the following approaches:

– GA,
– DEA-based circular Malmquist indices,
– pooled Corrected OLS (COLS) and
– pooled SFA (only when measurement noise is included in the experiment).

Productivity change for the frontier-based approaches is estimated on the basis of the Malmquist productivity index. The models used and the method of estimation are discussed in more detail in section 4.3.2.

For the S1 category of experiments (ie those that adopt a Cobb-Douglas production function in the DGP), the analysis found that the parametric approaches are in general more accurate than GA and DEA. This is not an unexpected result, since the parametric models that are ranked highest in each experiment are perfectly specified, in that they utilise the same functional form as the adopted production function and, in the case of the best-performing SFA models, assume the correct distribution for the inefficiency term. In the S2 category of experiments (ie those that adopt a piecewise-linear production function) however, it is the non-parametric approaches (GA and DEA) that appear as more accurate on the whole, sometimes by a large margin.

On the whole, some general statements can be made examining the results of the simulation analysis:

– All approaches could be considered relatively accurate when there is no noise in the data and average technical inefficiency is modest. In these cases, COLS is the better performing approach when there is no functional form misspecification, while DEA and GA are equally valid choices when the COLS model is misspecified.

– When the technical inefficiency is more prominent and the data are not measured with any error (no noise component in the DGP), DEA provides the most accurate estimates overall, even when the COLS model is perfectly specified.

- All approaches could also be considered relatively accurate when a modest amount of noise is introduced in the dataset. Under such conditions, the most accurate approach is SFA when the DGP utilises a Cobb-Douglas production function, while DEA is deemed the most accurate approach when adopting a piecewise-linear production function. Probably the most interesting finding here is that the deterministic approaches perform only slightly worse than the stochastic models in the S1 experiments and slightly better in the S2 experiments. This suggests that the use of deterministic approaches should not be dismissed in applications, just because some (modest) amount of measurement error is very likely to be present in the data.

- When measurement error becomes larger, the accuracy of all approaches deteriorates rapidly, to the point that their estimates could be considered unreliable for policy purposes. This is observed in both S1 and S2 sets of experiments, and although the estimates of the SFA models are significantly more accurate when technical inefficiency is modest, they are still quite dissimilar from the true productivity change values (average MAD scores of 12.5% in the best of cases).

- A surprising result is the relatively bad performance of the COLS models that adopted the translog specification. In al experiments, the Cobb-Douglas COLS models perform significantly better than their translog counterparts, despite the fact that, according to theory, the more flexible nature of the translog would result in a better overall fit to the data. This is partly reversed for the translog SFA models, which are assessed as being more accurate than their Cobb-Douglas counterparts when noise is modest. However, when noise becomes more prominent, the Cobb-Douglas SFA models outperform the translog SFA specifications. Possible reasons for this behaviour could be the relatively small sample size adopted for this analysis (100 obs), the fact that the translog models may have included statistically insignificant variables or the fact that the adopting a translog specification might lead to the violation of the monotonicity and convexity conditions that are inherent in the adopted piecewise-linear production function. The above are just untested hypotheses at this stage and additional research would be required to examine this issue in more detail.

- Increased volatility in inputs from one period to the next adversely affects the accuracy of all approaches, in almost all experiments. The DEA estimates are the least affected, while the GA estimates are the most affected. Interestingly, when noise is introduced in the analysis, the increased volatility appears to have very little impact on the accuracy of the deterministic approaches and almost no impact at all on the accuracy of the stochastic approaches in the S1 experiments, while in the S2 experiments, all approaches are affected except when a more extended noise component is introduced in the data.

– The inclusion of fully efficient units has no perceptible impact on the accuracy of the results, other than a slight improvement in some accuracy indicators. This slight improvement is most likely due to the fact that the inclusion of fully efficient units decreases the average technical inefficiency in the data, which is as was previously noted, has a beneficial effect on the overall accuracy of all approaches.

– While it was expected that functional form misspecification would adversely affect the accuracy of the parametric models, the simulation experiments revealed that the magnitude of this negative effect can be quite significant. In the S2 experiments, even when no noise is present, the COLS Cobb-Douglas specification displays MAD scores that are at least twice as large as those displayed by the DEA estimates and the discrepancy in MSE scores is significantly bigger (at least three times higher). When modest noise was included in the data generation process, the non-parametric deterministic approaches perform better that the stochastic models. In addition, the SFA specification that (incorrectly) assumes that the inefficiency is half-normally distributed is more accurate than the correctly specified SFA exponential model. Only when the noise component becomes more prominent is the correctly specified SFA exponential model deemed to be most accurate. Even in these cases, the performance of said specification is significantly better than the next best deterministic approach only when the levels of technical inefficiency are relatively modest (see S2.4). And even so, with a MAD score of 14.7%, it would be difficult for any outside observer to label the performance of the SFA model as ‘accurate’.

– In addition, the analysis faced difficulties when estimating the SFA models under conditions of relatively large technical inefficiency levels, because the MLE procedure could not converge to a solution. Although the analysis circumvented this issue by discarding the problematic datasets, this does not mean that the problem is not there and cannot arise in real-world applications. As such, it should be noted as a possible weakness of the SFA approach in similar situations (functional form misspecification with relatively large inefficiency levels) that is not reflected in the quantitative results of this analysis.

- The impact of misspecification in the inefficiency distribution for the SFA models to the accuracy of the productivity change estimates is quite small. Although in the majority of the experiments, the correctly specified SFA exponential model performed better than the SFA half-normal model, the differences in all accuracy measures were quite small. In fact, in the S2 experiments with modest noise (ie S2.3 and S2.6), the SFA half-normal model was assessed as more accurate than the correctly specified SFA exponential model. The fact that the adoption of the wrong inefficiency distribution has only a small impact in the overall accuracy of the results strengthens the case for the use of SFA in the context of productivity change measurement, since one of the main theoretical disadvantages of the approach (ie the fact that a distribution for the inefficiency term had to be specified a priori, with no way of testing whether the choice was the right one) does not seem able to introduce significant bias to the results.

To summarise, this analysis demonstrates that no productivity change measurement approach has an absolute advantage over another, but rather under some specific circumstances, a specific approach is likely to be more accurate than its counterparts. Probably the most significant findings are that the non-parametric deterministic approaches offer very robust estimates even when noise is present in the data in moderate levels and that functional form misspecification has a large negative effect in the accuracy of the parametric approaches even when a flexible functional form is adopted.

The analysis also clearly demonstrates that frontier-based approaches can usually produce at least as accurate, and in the majority of cases more accurate, productivity change estimates than the more traditional GA approach. And given that high quality databases on measures of economic growth, employment creation and capital formation are becoming increasingly available, the adoption of frontier-based approaches when measuring aggregate productivity growth can only help improve our understanding of this elusive and complex topic.

Chapter 5. Selecting between different productivity measurement approaches: An application using EU KLEMS data

5.1. Introduction

This chapter has two main aims: first, to provide up-to-date application of aggregate productivity measurement using both GA and frontier-based approaches and secondly to discuss how the findings of the simulation analysis undertaken in the previous chapter can be applied in practice, with reference to this application.

There are a number of approaches that could be used to measure productivity change, both in the macro (ie economy- or industry-wide) and micro (eg company or department) level, and each has its own strengths and weaknesses. Some of the most common approaches were discussed in chapter 3 and their relative accuracy was assessed through simulation analysis in chapter 4. This chapter provides a practical application of these approaches using the EU KLEMS dataset. The EU KLEMS dataset was originally designed so it can provide estimates of productivity change using Growth Accounting (GA); in fact, the main goal of the EU KLEMS project is to measure productivity change in the EU countries. However, the available data can also be readily used to measure productivity change through frontier-based approaches. This can be very valuable to the users of this analysis, since frontier based approaches can provide a number of advantages over GA, such as the ability to decompose productivity change and the incorporation of variable returns to scale in the analysis. More importantly however, frontier-based approaches can potentially provide more accurate estimates of productivity change itself, as the simulation analysis in chapter 4 revealed. The productivity analysis is detailed in section 5.2.

The issue of the accuracy and the overall reliability of the productivity estimates can be quite complex. It is not uncommon for productivity analyses that adopt a number of different approaches to derive productivity estimates that are quite different from each other; in some cases, these differences can be quite substantial[66]. In those instances, the analysis would need to be able to put forward an informed view on

[66] See for example Coelli (2002).

why the various approaches come up with such divergent views and, more importantly, which estimates are likely to be more accurate. To that end, this chapter proposes the use of a simple framework which could be used to select between the competing estimates, based on the likely accuracy of the adopted approaches when taking into account the characteristics/conditions *specific to the application at hand*. Although such a framework is also applicable in the micro setting and when the analysis is interested in measuring efficiency, the focus here is on assessing the relative accuracy of the different productivity estimates in the aggregate, economy-wide setting. The selection framework is detailed in section 5.3.

5.2. Productivity change in the EU KLEMS dataset

Throughout this chapter, productivity change is measured through the use of the classic model of aggregate production, detailed in chapter 2. In its simplest form, the aggregate production model stipulates that aggregate output is a function of aggregate inputs. As demonstrated in chapter 3, this model forms the basis of all major productivity measurement approaches (namely index-number approaches, parametric frontiers and non-parametric distance functions).

Section 5.2 is structured as follows:

- Section 5.2.1 provides an overview of the data (sourced from EU KLEMS) used in the analysis, specifically focusing on how they can be utilised to measure aggregate productivity change by both index-based (GA) and frontier-based approaches.
- Section 5.2.2 provides a brief discussion of the approaches used to measure aggregate productivity growth in this application. Additional details on the theory behind these approaches and their relative strengths and weaknesses can be found in chapter 3, while the technical details on the formulation of the models can be found in section 4.3.2.
- Section 5.2.3 presents and compares the estimates of productivity change from all adopted approaches.
- Section 5.2.4 provides a more in depth view of the estimated productivity performance in a sub-sample of the assessed countries, focusing on a selected number of key approaches, with the aim of highlighting the differences and similarities of the produced estimates.

– Section 5.2.5 presents and discusses the decomposition of productivity change into efficiency change, technological change and scale efficiency change, derived from the key approaches identified in section 5.2.4.

5.2.1. Data

The analysis carried out in this chapter uses the dataset constructed by the EU KLEMS (2008) project. The EU KLEMS project aims to provide a harmonised set of indicators for the measurement and comparison of productivity performance for a large number of, mostly, EU countries. The dataset provides information, among others, on economy-wide level aggregates from 1970 to 2007[67], based on information from each country's national accounts but adjusted for comparability across time and countries. The dataset also includes GA-based total factor productivity growth estimates derived from the primary data, which are used in this chapter as is, together with the originally-derived frontier-based estimates.

Probably the most important feature of the EU KLEMS dataset is the work done to ensure the comparability of the National Accounts data drawn from a large number of countries. As mentioned in chapter 2, the adoption of international Standards of National Accounts greatly enhances data comparability, but the adoption of these standards is a relatively recent development; for earlier periods (years prior to 1985, when NACE[68] was adopted by the majority of EU countries), EU KLEMS utilised additional data to ensure that industrial classifications, aggregation levels and price concepts are consistent across all surveyed countries. Additional details for this process can be found in the EU KLEMS Methodology paper (O'Mahony and Timmer (2009)). As mentioned in the introduction, the existence of a dataset of harmonised National Accounts allows the utilisation of frontier-based approaches for the measurement of aggregate productivity change.

This application focuses on assessing productivity at the economy-wide level. The main reason for this choice of aggregation is that economy-level data are likely to be more robust, due to the fact that any errors when estimating inter-industry transactions, necessary for the creation of industry-level data, have no impact at the economy-wide level of aggregation.

---

67 For some countries the start and end data of the period for which data is available differs, resulting in a unbalanced panel.

68 As a reminder, Nomenclature generale des Activites economiques dans les Communautes Europeennes (NACE) is the industry classification standard adopted by the European Union.

5.2.1.1 Output

The output measure of choice at this level of aggregation is Gross Value Added (VA), as discussed in some detail in chapter 2. EU KLEMS provides both economy-wide nominal VA as well as its price index, which can be used to calculate real VA, ie a volume measure of output.

This information is sufficient to estimate productivity change through GA, but frontier-based approaches also require that output volumes are expressed in the same unit of measurement for all units included in the analysis. This is achieved by further adjusting VA to account for differences in Purchasing Power Parities (PPPs). As a reminder, PPP adjustment ensures that the volume in question is converted to a common, base currency that also reflects the purchasing power of each individual currency in its perspective national market (more information on PPPs is provided in chapter 2). The PPPs used for this particular conversion are output-specific (in this case, calculated on the basis of VA) and are also sourced from the EU KLEMS dataset.

5.2.1.2 Inputs

Since the analysis utilises VA as the output of choice, the inputs required are measures of labour and capital services used.

In terms of the labour input, the analysis utilises aggregate 'hours worked', adjusted by a number of factors to take into account the differences in the composition of labour input. Also of note here is that EU KLEMS have made a number of adjustments to primary (National Accounts) data to ensure comparability between the different countries; these adjustments mainly concern the broad categories of 'full-time equivalency' (FTE) definitions, using actual rather than paid hours worked and collating additional data to estimate self-employed hours worked. A more detailed discussion on these issues is provided in the EU KLEMS Methodology paper (op cit.). What should be noted here is that these adjustments were individual to each country examined, in order to take into consideration the unique characteristics of each assessed economy.

The resulting data on hours worked were further adjusted to take into account the differences in labour composition (low, medium and high skill labour). This adjustment was in turn based on data on the overall educational attainment of the labour force. Since this data is not available in the National Accounts, EU KLEMS employed additional data sources (mainly Labour Force Surveys) for this adjustment; more details can be found in the EU KLEMS Methodology paper.

This final adjustment is not strictly necessary for the measurement of productivity change; its main purpose is to allow the creation of a productivity change measure that is net of the effects of changes in the labour force composition. Nevertheless, the adjusted-working hours measure was also adopted for this analysis, mainly to ensure comparability between the GA estimates sourced from EU KLEMS and the frontier-based estimates calculated in this chapter.

The second input of the analysis is a measure of capital. As discussed in some detail in chapter 2 of this thesis, this is a very difficult measure to construct and requires a number of assumptions, both in the construction of the capital stock (ie the ‘pool’ of capital assets available to an economy) and in the construction of the flow of services derived from this stock. The first measure, capital stock, was constructed by EU KLEMS using the Perpetual Inventory Method (PIM); an overview of this method can be found in chapter 2 and additional discussion is provided in the EU KLEMS methodology paper and the OECD manual (OECD (2001)).

The creation of the capital stock measure is a laborious process that requires data from National Accounts and capital formation matrices for long timeframes; EU KLEMS had to deal with a large number of implementation issues, details of which can be found in their methodology paper. Furthermore, to ensure comparability between the countries in the sample, EU KLEMS used harmonised depreciation rates and applied consistent capital accounting procedures to deal with issues such as weighting between various asset categories and rental rates. According to EU KLEMS, ‘this treatment results in a data series that is both consistent across time and across countries and at the same time includes substantially more information than those utilised in previous studies’ (see Inklaar and Timmer (2009)). The use of harmonised depreciation rates however may not reflect economic reality. There can be a variety of reasons why the depreciation rates between countries can be different, especially in the time period covered by this dataset; countries that undergo rapid structural changes[69] may well display higher rates of depreciation relative to more stable economies as older, less productive capital assets are rapidly replaced. This is also mentioned in the EU KLEMS methodology paper, but not discussed further. Despite all the implementation difficulties, EU KLEMS was able to construct

[69] For example, Eastern European countries after the collapse of the Eastern block, but also European Mediterranean countries after joining the European Union and/or the Euro and Germany after the unification of East and West Germany.

consistent time series of capital stock at both industry- and economy-wide levels, by allocating capital investment into seven main asset categories (3 ICT and 4 non-ICT assets)[70].

EU KLEMS does not use the capital stock measure for those assets directly; rather, capital stock is used as the basis of constructing a measure of the flow of capital services arising from this stock. The capital services measure used by EU KLEMS is simply a weighted sum of the different assets that comprise the capital stock. The weights used for this aggregation are based on the average shares of each asset in the value of total capital compensation, which is calculated in this instance as VA minus labour compensation[71]. In order to calculate average shares for each asset, the analysis also requires the estimation of the price of capital stock, which in turn means that an estimate of the user cost of capital is also required. As was discussed in previous chapters, the user cost of capital is not directly observable, but if the neoclassical assumptions hold, it can be estimated residually (see sections 3.6.1 and 4.3.1 for additional discussion on this issue). It should be noted here that the most common applications of GA require this type of aggregation of capital stock; otherwise, the final weights for the aggregate capital measure in the GA formula (see equations 5.2.1 to 5.2.3 below) will be inconsistent.

One of the main strengths of the frontier-based approaches is that they do not rely on the neoclassical assumptions and they can estimate productivity change without requiring information on input prices. In other words, there is no need to convert capital stock into a measure of capital services when using frontier-based approaches. Furthermore, since capital stock for all assets is expressed in the monetary currency of each individual economy, it can be easily aggregated to a single measure by a simple addition. Note that this does not reduce the information context of the measure relative to capital services. In fact, capital stock offers a complete representation of the capital services provided by said asset, since it accounts for retirements, the decline of productive capability due to age (wear and tear) and quality changes. In addition, since it is based on investment expenditure, it also reflects the relative importance of each asset, assuming that the price of each asset is a good indicator of its importance to production.

[70] The ICT (Information and Communication Technology) assets are office and computing equipment, communication equipment and software while the non-ICT assets are transport equipment, other machinery and equipment, residential buildings and non-residential structures.

[71] This definition of capital compensation implicitly assumes that only ‘normal’ profits are possible, ie profits that correspond directly to the user cost of capital. This was briefly discussed in chapter 4 (section 3.1).

One possible complication of using capital stock directly in the analysis is that, the measure expressed in the currency of each individual country; the problem arises from the fact that the frontier-based approaches require all inputs to be expressed in the same units of measurement. This issue can be resolved by adjusting the available capital stock information based on PPP indices, similar to the treatment of aggregate output (VA in this application). For this analysis, capital stock-specific PPPs are used, sourced from the GGDC database[72]. It should be noted that previous studies of aggregate productivity change have sometimes used output-based PPPs in order to convert inputs expressed in national momentary terms to a single currency[73]; this could potentially introduce bias to the results, since PPPs are based on 'baskets' of goods and services and it is quite likely that the group of goods and services that comprise aggregate output will be different to those that comprise aggregate capital stock.

The countries that are included in the analysis together with the time periods for which data is available and the average values of the inputs and output are given in table 5.1 below. Overall, the productivity growth estimates are produced for 14 different countries, over a number of years starting from 1970 and ending in 2007; on the whole the analysis includes 375 observations (each country in each time period as a different observation).

[72] Groningen Growth and Development Centre (GGDC) Productivity level database: Inklaar and Timmer (2008).
[73] For a discussion, see Inklaar and Timmer (2009), op cit.

Table 5.1: Descriptive statistics of the dataset

Country	Short code	Observations	Start date	End date	PPP-adjusted Value added (average)	Adjusted Hours worked (average)	PPP-adjusted Capital stock (average)
Australia	AUS	26	1982	2007	334,732	14,996	1,012,975
Austria	AUT	28	1980	2007	135,593	6,411	600,872
Czech Republic	CZE	13	1995	2007	115,130	10,278	215,538
Denmark	DNK	28	1980	2007	98,943	4,052	401,680
Spain	ESP	28	1980	2007	533,431	24,859	2,046,906
Finland	FIN	38	1970	2007	79,438	3,754	258,258
Germany	GER	17	1991	2007	1,660,380	57,623	5,705,057
Italy	ITA	38	1970	2007	836,748	39,704	3,221,461
Japan	JPN	34	1973	2006	1,831,401	119,325	9,767,948
Netherlands	NLD	29	1979	2007	282,496	10,205	987,615
Slovenia	SVN	12	1995	2006	20,850	1,712	29,957
Sweden	SWE	15	1993	2007	180,068	6,996	355,325
United Kingdom	UK	38	1970	2007	827,492	45,309	1,890,611
United States of America	USA	31	1977	2007	6,867,596	233,426	18,108,226

Note: PPP-adjusted Value added and capital stock is in millions of PPP-adjusted Euros (German Euros as base), adjusted Hours worked is also in millions.

5.2.2. Methods

Productivity change in this application is assessed using the same approaches adopted for the simulation analysis undertaken in the previous chapter, namely:

- GA (productivity change estimates are sourced directly from the EU KLEMS database),
- DEA-based circular Malmquist indices,
- COLS-based Malmquist indices, and
- SFA-based Malmquist indices.

The approaches are discussed in detail in chapter 3; what follows is a brief overview of the formulations adopted for each approach.

Growth Accounting

EU KLEMS adopts the 'standard' GA framework[74] for its productivity analysis. In brief, GA postulates the existence of an aggregate production function that can be parameterised if the 'neoclassical' assumptions hold. These assumptions include assertions such as production is always at constant returns to scale, markets in general are perfectly competitive and all actors have perfect information on prices and marginal products, there are not delay costs and that productivity is Hicks-neutral.

If these, admittedly restrictive, assumptions hold, once the production function is differentiated with respect to time, the rate of change in output is equal to the sum of the weighted average of the change in inputs and the change in productivity. The input weights are the output elasticities of each factor of production, which are derived as the share of each input to the total value of production. Therefore, productivity change is estimated by:

$$\frac{d\ln TFP_i^{GA}}{dt} = \frac{d\ln Y_i}{dt} - S_i^L \frac{d\ln L_i}{dt} - S_i^K \frac{d\ln K_i}{dt} \qquad \text{Eq. 5.2.1}$$

where S_i^L is the average share of labour in periods t and t-1, S_i^L is the average share of capital in t and t-1 given by:

$$S_i^L = \left(\frac{w_{it}^L L_{it}}{p_{it} Y_{it}} + \frac{w_{it-1}^L L_{it-1}}{p_{it-1} Y_{it-1}} \right) \Big/ 2 \qquad \text{Eq. 5.2.2}$$

$$S_i^K = \left(\frac{w_{it}^{K,GA} K_{it}}{p_{it} Y_{it}} + \frac{w_{it-1}^{K,GA} K_{it-1}}{p_{it-1} Y_{it-1}} \right) \Big/ 2 \qquad \text{Eq. 5.2.3}$$

As previously noted, the price of capital is not observable; as such, EU KLEMS, like the majority of GA applications (see for example OECD (OECD, 2001)), uses an endogenous 'user cost of capital' to estimate the final price of capital.

DEA-based circular Malmquist index

The DEA-based circular Malmquist-type index (thereafter referred to as circular MI), is based on the notion of the 'meta-frontier', a single frontier that envelops all data points from all periods.

[74] Detailed in chapter 3 of this thesis.

For this application, the meta-frontier was constructed based on a technology that utilised two inputs (skill-adjusted Labour hours and PPP-adjusted Capital stock) to produce a single output (PPP-adjusted real VA). As with the standard MI, productivity change was measured with a reference to technology at constant returns to scale. For the decomposition of the circular MI, the scale efficiency change calculations were based on the output oriented VRS efficiency scores.

Corrected OLS

Corrected OLS (COLS) is a deterministic, parametric approach and one of the numerous ways that have been suggested to 'correct' the inconsistency of the OLS-derived constant term of the regression when technical inefficiency is present in the production process.

Two different COLS model specifications are used for this application. Both are based on a pooled regression model (ie all observations are included in the same model with no unit-specific effect). The first model assumes a Cobb-Douglas functional form and is given by:

$$\ln Y_{it} = \sim^{*} \ln L_{it} + {}^{*}\ln \mathrm{K}_{it} + \yen^{*} t + \pounds *_{it} \qquad \text{Eq. 5.2.4}$$

Where Y_{it} is PPP-adjusted real VA, L_{it} is skill-adjusted Labour hours, K_{it} is PPP-adjusted Capital Stock, t is the time variable and $\pounds *_{it}$ are the estimated OLS residuals.

The second COLS model specification assumes a translog functional form and is given by:

$$\begin{aligned}\ln Y_{it} = a_i + {}_{L}\ln L_{it} + {}_{K}\ln K_{it} + \yen_t t + \frac{1}{2}\,{}_{LL}\left(\ln L_{it}\right)^2 + \frac{1}{2}\,{}_{KK}\left(\ln K_{it}\right)^2 + \frac{1}{2}\yen_{tt} t^2 \\ + {}_{KL}\ln K_{it}\ln L_{it} + \yen_{Kt}\ln K_{it} t + \yen_{Lt}\ln L_{it} t + \pounds^{*}_{it}\end{aligned} \qquad \text{Eq. 5.2.5}$$

Inefficiency estimates are derived by:

$$-u^{*}_{it} = \pounds^{*}_{it} - \max(\pounds^{*}_{it}) \qquad \text{Eq. 5.2.6}$$

Productivity change is calculated by adding the different components of the Malmquist productivity index:

$$d\ln TFP^{COLS}_{it} / dt = d\ln EC^{COLS}_{it} / dt + d\ln TC^{COLS}_{it} / dt + d\ln SEC^{COLS}_{it} / dt \qquad \text{Eq. 5.2.7}$$

where EC_{it}^{COLS} is the COLS-estimated efficiency change, TC_{it}^{COLS} is the COLS-estimated technical change and SEC_{it}^{COLS} is the COLS-estimated scale efficiency change. The formulae for calculating the three components of the productivity index are given in detail in section 4.3.2 (Equations 4.3.27 to 4.3.32).

Stochastic Frontier Analysis

Stochastic Frontier Analysis (SFA) is a stochastic parametric approach that relies on the notion that the observed deviation from the frontier could be due to both genuine inefficiency but also random effects, including measurement error. SFA attempts to disentangle those random effects by decomposing the residual of the parametric formulation of the production process into noise (random error) and inefficiency.

As is the case with the COLS approach, two separate SFA model specifications are used in this application: one that adopts a Cobb-Douglas functional form and a second that adopts the translog. The models are very similar to those used under COLS; the only difference lies in the specification of the error term.

In more detail, the Cobb-Douglas model is given by:

$$\ln Y_{it} = \sim^{*} \ln L_{it} + \ ^{*} \ln \mathrm{K}_{it} + ¥^{*} t + v_{it} - u_{it} \qquad \text{Eq. 5.2.8}$$

whereas the translog model is given by:

$$\begin{aligned}\ln Y_{it} = a_i + \ _{L} \ln L_{it} + \ _{K} \ln K_{it} + ¥_{t} t + \frac{1}{2} \ _{LL} \left(\ln L_{it}\right)^2 + \frac{1}{2} \ _{KK} \left(\ln K_{it}\right)^2 + \frac{1}{2} ¥_{tt} t^2 \\ + \ _{KL} \ln K_{it} \ln L_{it} + ¥_{Kt} \ln K_{it} t + ¥_{Lt} \ln L_{it} t + v_{it} - u_{it}\end{aligned} \qquad \text{Eq. 5.2.9}$$

where u_{it} represents the inefficiency component (and as such $u_{it} \geq 0$) and v_{it} represents measurement error ($v_{it} \sim N(0, \pm_{v}^{2})$). The inefficiency component is estimated based on the JMLS[75] estimator.

Two different distributions for the inefficiency component are tested:

– the exponential distribution, $u_{it} \sim Exp(\pm_{u})$

– the half-normal distribution, $u_{it} \sim N^{+}(0, \pm_{u}^{2})$

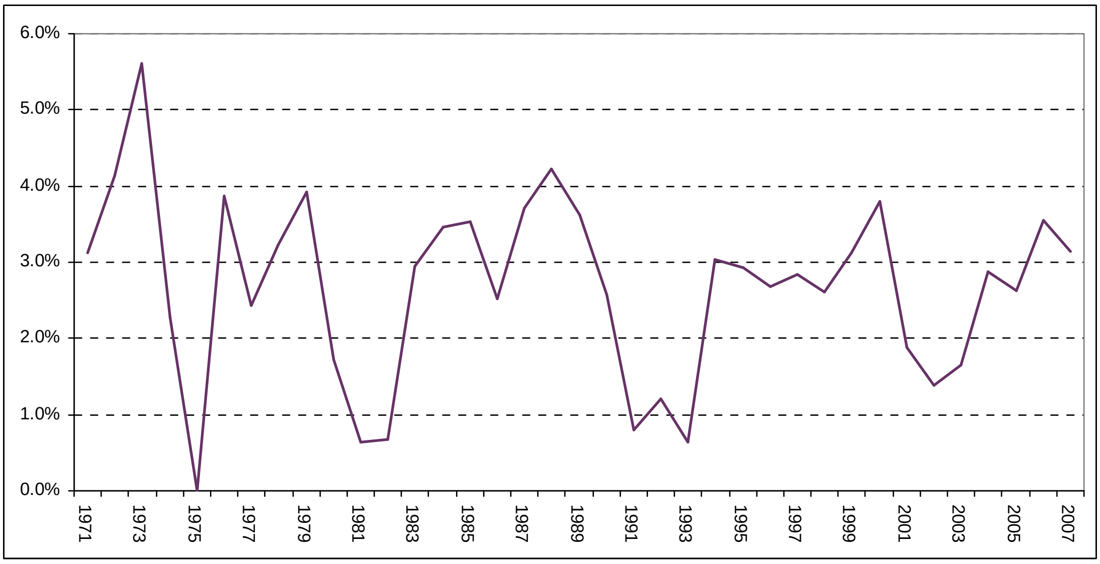
Productivity change is measured in exactly the same way as with COLS.

[75] See Jondrow, Knox Lovell, Materov, & Schmidt (1982), op cit.

5.2.3. Overview of the results on productivity change

Due to the relatively long timeframe of this analysis, the various estimates of productivity change are summarised on the basis of a number of distinct periods. In more detail, summary results aggregated across all countries are presented for the full period, ie 1970-2007, for the period when all countries included in the analysis provide useable data, ie 1995-2007 and finally for smaller timeframes that attempt to simulate 'business cycles'. 'Business cycles' are defined in this analysis as periods that begin with years of increased rate of average VA growth[76] and end with years where the average VA rate of growth is declining. The reason why the analysis utilises these 'business cycles' (as defined above) is that they provide a view of productivity growth that is less contaminated by the cyclical movement of output. These cyclical movements are demonstrated in Figure 5.2.

Figure 5.1: Annual change in VA (PPP-adjusted, real), averaged across all countries



The 'business cycle' periods selected for this analysis are:

– from 1970 to 1982 (12 years in total);

– from 1983 to 1993 (10 years in total);

– from 1994 to 2002 (8 years in total) and

– from 2003 to 2007 (4 years in total).

76 Average VA growth rates were calculated by averaging VA across all observations by period and then calculating the relative change of this indicator on an annual basis.

Figure 5.1 shows that the first period (1970-1982) might actually contain two ‘business cycles’, but due to the small number of economies for which data was available in the 1970-1980 period and the relative short timeframe between the economic downturns, this analysis chose to present its results based on this longer timeframe.

Before presenting the results, it should be noted that the analysis also run a number of diagnostic tests for the parametric models (the COLS and SFA models). The results of these tests and how they can be used to establish a selection framework for productivity measurement will be discussed in the next section of this chapter.

Table 5.3 presents a summary of the annual productivity change estimates by approach for the full period and the 1995-2007 period, ie the period in which all countries included in the analysis provide useable data.

Table 5.2: Annual productivity change estimates for the full period

Productivity measure	DEA	COLS	COLS translog	SFA (half-normal)	SFA (exponential)	SFA translog (half-normal)	SFA translog (exponential)	GA
Full period (361 obs)								
Mean	0.52%	0.67%	0.86%	0.82%	0.82%	0.77%	0.88%	0.54%
Std. Dev.	1.66%	1.44%	1.70%	1.14%	0.99%	1.69%	1.15%	1.44%
1995-2007 (178 obs)								
Mean	0.65%	0.65%	0.75%	0.75%	0.75%	0.57%	0.68%	0.45%
Std. Dev.	1.24%	1.21%	1.57%	0.99%	0.92%	1.60%	1.08%	1.26%

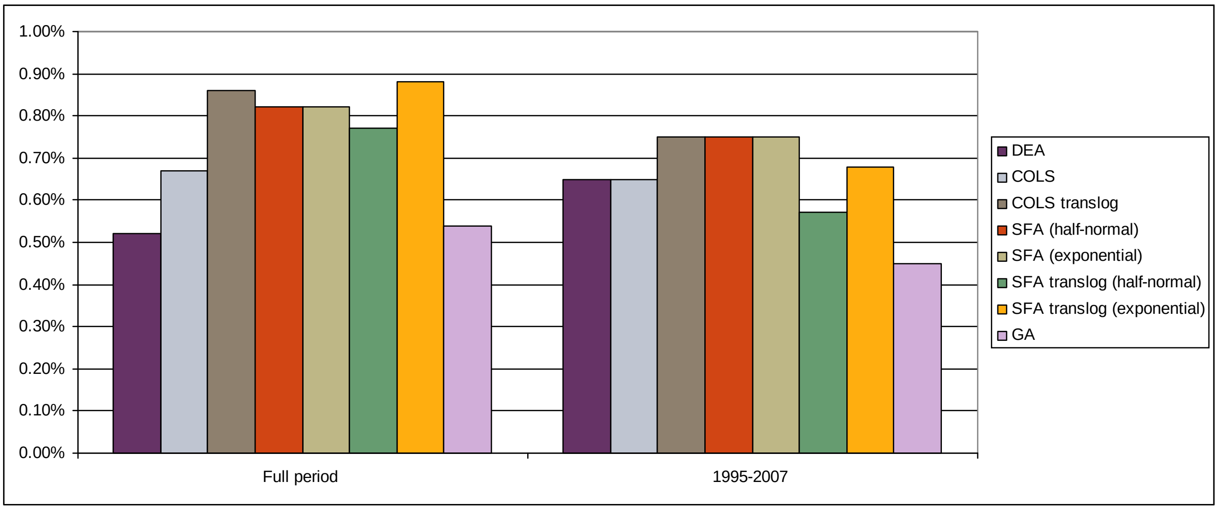
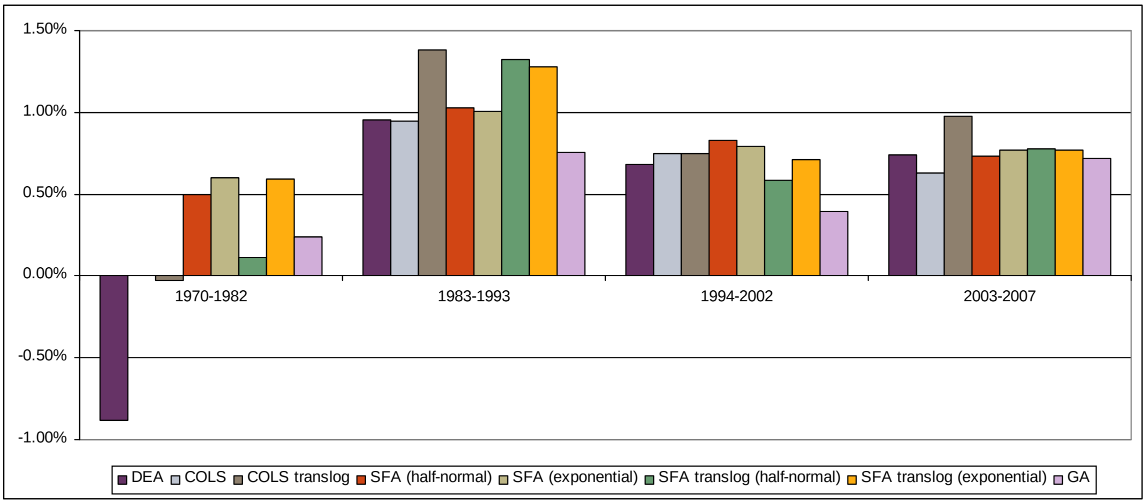


Table 5.2 shows that for the full period, productivity has been growing at an average rate of between 0.52% and 0.88%. The lowest average growth comes from the DEA-based circular Malmquist index, while the highest estimate comes from the translog SFA model that assumes an exponential distribution of inefficiency. Overall, the deterministic approaches, with the exception of the translog COLS, provide lower productivity change estimates relative to the stochastic models.

Average productivity growth during the more recent 1995-2007 period is quite similar with the average growth displayed in the full sample period, with only some small variations by approach. In general, the SFA- and GA-based estimates of average growth are slightly lower, while the DEA-based estimates are slightly higher. The translog SFA-based estimates display the largest change, with average TFP growth being approximately 0.2 percentage points smaller in the 1995-2007 period relatively to the full sample period. The estimates from the deterministic approaches are still lower relative to those derived from the SFA models, but the differences are less pronounced relative to the full period results.

Table 5.3: Average annual productivity change estimates for the different 'business' cycles

Productivity measure	1970-1982	1983-1993	1994-2002	2003-2007
DEA	-0.88%	0.95%	0.68%	0.74%
COLS	0.00%	0.95%	0.75%	0.63%
COLS translog	-0.02%	1.38%	0.75%	0.97%
SFA (half-normal)	0.49%	1.02%	0.83%	0.73%
SFA (exponential)	0.60%	1.01%	0.79%	0.77%
SFA translog (half-normal)	0.11%	1.32%	0.59%	0.78%
SFA translog (exponential)	0.59%	1.28%	0.71%	0.77%
GA	0.24%	0.75%	0.40%	0.72%



Overall, the variation in average productivity growth between the various estimates is relatively small, when examining performance over the various 'business cycles'. Only in the 1970-1982 period do the estimates of some approaches (namely DEA and both COLS specifications) diverge in a more pronounced fashion; this may be due to the fact that although this period covers almost 12 years – more than any other 'business cycle' in this analysis – it only includes 59 observations, which are less than in any other 'business cycle' (the 2003-2007 period is only four year long but still contains 68 observations). Another reason of the variability of the productivity growth estimates of this first period is the apparent volatility in the output observed during this time. This is expected given that during that time the global economy experienced the so-called ‘oil crises’ of the 1970’s (the first crisis in 1973 and the second crisis in 1979).

In terms of the trend in average productivity growth, it appears that the countries in the sample experienced the fastest productivity improvements during the second business cycle. The rate of productivity growth receded during the 1994-2002 period according to all approaches and then slightly accelerated again during the 2003-2007 period, according to the majority of the estimates (with the exception of the Cobb-Douglass parametric models).

The overall similarity of the average productivity growth estimates between the examined approaches is also apparent in the correlations between the estimates. The following table presents both Pearson’s and Spearman’s (rank) correlation coefficients for the full period.

Table 5.4: Correlation coefficients for annual productivity growth

Approach	Correlation measure	DEA	COLS	COLS translog	SFA (half-normal)	SFA (exponential)	SFA translog (half-normal)	SFA translog (exponential)
COLS	Pearson’s	89%						
	Spearman’s	88%						
COLS translog	Pearson’s	88%	81%					
	Spearman’s	90%	84%					
SFA (half-normal)	Pearson’s	84%	96%	77%				
	Spearman’s	85%	98%	81%				
SFA (exponential)	Pearson’s	80%	91%	72%	98%			
	Spearman’s	83%	96%	78%	99%			
SFA translog (half-normal)	Pearson’s	86%	80%	99%	76%	72%		
	Spearman’s	90%	85%	99%	82%	80%		
SFA translog (exponential)	Pearson’s	82%	76%	92%	78%	79%	94%	
	Spearman’s	86%	79%	93%	81%	80%	95%	
GA	Pearson’s	80%	95%	79%	92%	88%	79%	75%
	Spearman’s	80%	94%	83%	93%	91%	84%	78%

According the table above, GA estimates are more highly correlated with the Cobb-Douglass COLS and SFA (half-normal) estimates and less highly correlated with the DEA and translog-specified parametric approaches. DEA estimates are more highly correlated with the COLS estimates, a finding which is expected due to the deterministic nature of both approaches. It is interesting to note that the translog SFA estimates are more highly correlated with each other and their translog COLS counterparts, while they display the smallest correlation coefficients with the estimates from the Cobb-Douglas parametric models (COLS and SFA). This demonstrates that the selection of the functional form to parameterise the models can have a large effect on the TFP growth estimates, which is consistent with the findings of the simulation analysis undertaken in the previous chapter. It is also an indicator that in this particular application, the selection of the functional form for the parametric approaches has a discernible impact to the resulting estimates.

The results so far suggest that there appears to be a broad consensus between the various approaches. However average TFP change estimates across all countries masks the underlying variation observed at the (individual) country level.

Table 5.5: Average annual productivity estimates by country, full period

Country	DEA	COLS	COLS translog	SFA (half-normal)	SFA (exponential)	SFA translog (half-normal)	SFA translog (exponential)	GA
AUS	0.8%	0.7%	1.0%	0.8%	0.8%	0.9%	0.9%	0.5%
AUT	1.2%	0.9%	1.5%	1.0%	1.0%	1.4%	1.4%	1.0%
CZE	0.9%	1.4%	0.3%	1.5%	1.5%	0.0%	0.1%	0.6%
DNK	1.0%	0.8%	1.1%	0.9%	0.9%	1.1%	1.2%	0.3%
ESP	0.3%	0.3%	0.6%	0.4%	0.4%	0.6%	0.6%	0.0%
FIN	0.1%	0.7%	0.8%	0.9%	0.9%	0.9%	1.1%	1.0%
GER	1.3%	0.7%	1.8%	0.9%	0.9%	1.6%	1.2%	0.7%
ITA	0.6%	0.6%	1.0%	0.7%	0.6%	1.0%	0.9%	0.4%
JPN	0.5%	0.6%	1.5%	0.8%	0.8%	1.3%	1.2%	0.8%
NLD	0.8%	0.6%	1.0%	0.8%	0.9%	0.9%	1.3%	0.4%
SVN	-0.1%	1.0%	-2.6%	1.2%	1.1%	-3.0%	-1.6%	0.9%
SWE	0.3%	1.1%	1.2%	1.0%	1.0%	1.0%	0.9%	0.8%
UK	-0.4%	0.4%	0.2%	0.6%	0.7%	0.1%	0.5%	0.4%
USA	0.6%	0.5%	1.0%	0.9%	0.9%	0.8%	0.9%	0.2%

Note: A key of the short codes used by EU KLEMS and adopted in this analysis is provided in Table 5.1.

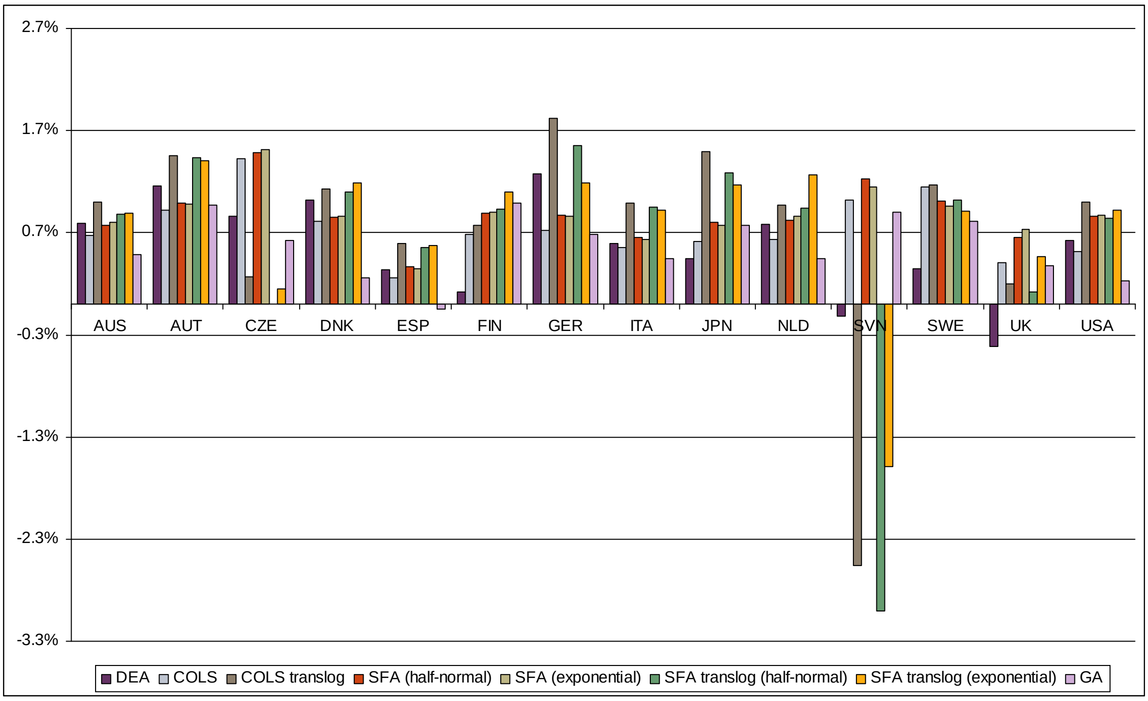
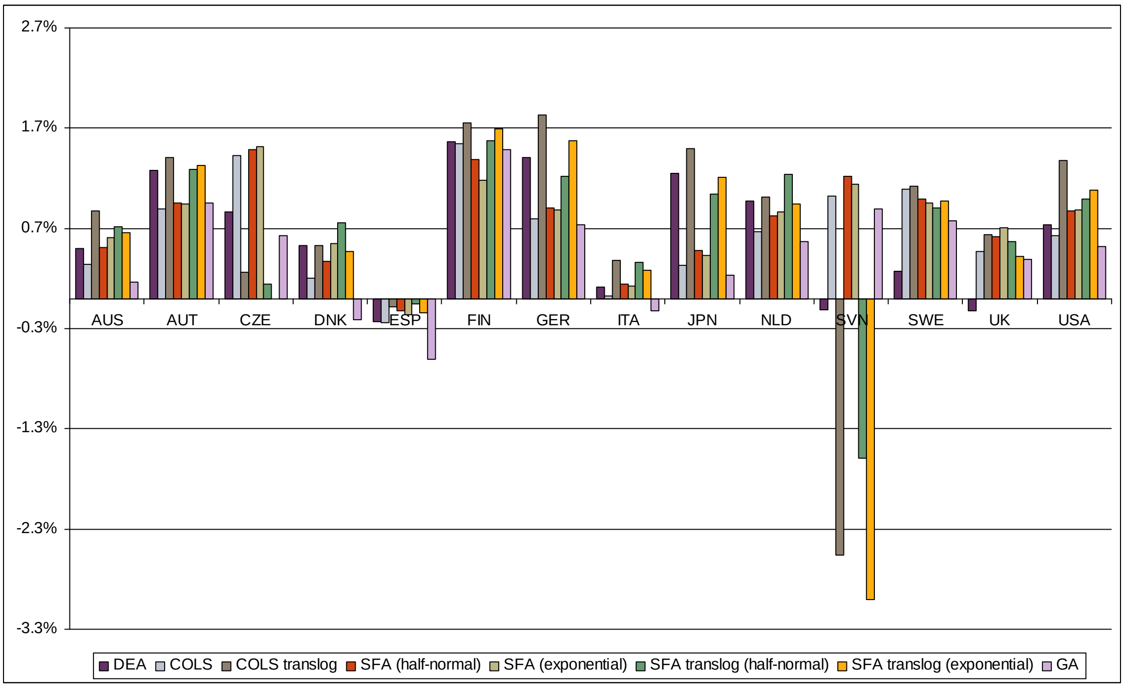


Table 5.5 reveals that the various productivity change estimates at country level appear to be quite different, for some countries at least. This is despite the fact that correlations of the different estimates are still relatively high when comparing productivity growth estimates within an individual country[77]. On average, the difference between the smallest and the largest estimate in levels is approximately 1.1 percentage points and for some countries the difference can be much larger (eg the spread is 4.2 and 1.5 percentage points for SVN and CZE respectively).

[77] The tables of within-country correlations of the different productivity growth estimates can be found in appendix 3, for a sub-set of the approaches.

Table 5.6: Average annual productivity change estimates by country, 1995-2007

Country	DEA	COLS	COLS translog	SFA (half-normal)	SFA (exponential)	SFA translog (half-normal)	SFA translog (exponential)	GA
AUS	0.5%	0.3%	0.9%	0.5%	0.6%	0.7%	0.7%	0.2%
AUT	1.3%	0.9%	1.4%	0.9%	0.9%	1.3%	1.3%	1.0%
CZE	0.9%	1.4%	0.3%	1.5%	1.5%	0.1%	0.0%	0.6%
DNK	0.5%	0.2%	0.5%	0.4%	0.5%	0.8%	0.5%	-0.2%
ESP	-0.2%	-0.2%	-0.1%	-0.1%	-0.2%	-0.1%	-0.1%	-0.6%
FIN	1.6%	1.5%	1.7%	1.4%	1.2%	1.6%	1.7%	1.5%
GER	1.4%	0.8%	1.8%	0.9%	0.9%	1.2%	1.6%	0.7%
ITA	0.1%	0.0%	0.4%	0.1%	0.1%	0.4%	0.3%	-0.1%
JPN	1.2%	0.3%	1.5%	0.5%	0.4%	1.0%	1.2%	0.2%
NLD	1.0%	0.7%	1.0%	0.8%	0.9%	1.2%	0.9%	0.6%
SVN	-0.1%	1.0%	-2.6%	1.2%	1.1%	-1.6%	-3.0%	0.9%
SWE	0.3%	1.1%	1.1%	1.0%	1.0%	0.9%	1.0%	0.8%
UK	-0.1%	0.5%	0.6%	0.6%	0.7%	0.6%	0.4%	0.4%
USA	0.7%	0.6%	1.4%	0.9%	0.9%	1.0%	1.1%	0.5%



The variation between the different estimates at the country level is also present when the focus is only in the 1995-2007 period, as the table above demonstrates. As with the full sample, the average difference between the largest and the smallest average productivity change estimate at the country level is approximately 1.1 percentage points and once more, SVN and CZE display even larger variations.

5.2.4. Productivity change in selected economies

Given these wide variations in average productivity change estimates, it is useful to examine in more detail the productivity performance of a sub-sample of the assessed economies. This sub-sample includes economies that display interesting variations in estimates, namely UK, SVN, GER and NLD, and will also present the case of an economy (ITA) where the different approaches provide an almost identical view of productivity change. Note that detailed figures and tables for the productivity performance of all assessed countries can be found in appendix 3.

To improve readability, the following analysis examines the estimates of four models (out of the eight utilised in this section), namely DEA-based MI, GA, translog SFA (exponential) and Cobb-Douglas COLS. For COLS, the Cobb-Douglas specification was selected due to its overall better accuracy performance in the simulations undertaken in the pervious chapter, relative to the translog COLS models. The same reasoning applies for the selection of the SFA functional form specification; the exponential model was selected based on the results of the diagnostic analysis, which will be discussed in detail in the next section of this chapter.

5.2.4.1 United Kingdom (UK)

The first economy assessed is the United Kingdom and is selected for closer examination mainly due to the relatively large differences in average productivity change estimates between DEA and the other three selected approaches.

Figure 5.2: UK productivity change (%), by selected approaches

Note: dTFP_sfa_e_translog = translog SFA (exponential) MI, dTFP_cols = Cobb-Douglas COLS MI, dTFP_dea = DEA MI and dTFP_va = GA-based productivity change (VA as output).

The first thing to notice from the above figure is that the estimates of productivity change from all four approaches are quite highly correlated. This is in fact true for the majority of the assessed counties and it was also observed in the simulation analysis undertaken in the previous chapter. Another feature that remains constant throughout all the countries assessed is that the SFA[78] estimates consistently display lower variability relative to those derived from the other three assessed approaches. This is predominantly due to the efficiency change component of the Malmquist index; as will be discussed later on this chapter, SFA produces less variable efficiency estimates from one period to the next and this in turn results in lower variability in the efficiency change component of the SFA-based MI; this is in turn most likely due to the fact that SFA classifies part of variation of the residual as noise, thus resulting in less volatile estimates of efficiency change.

With regards to the levels of the estimates, GA and COLS produce the highest productivity change estimates until 1993; from that year onwards the highest estimates are provided by COLS and SFA, with the exception of the final 4 years of

[78] The SFA estimates discussed in these sections refer specifically to the estimates derived from the translog SFA (exponential) model.

the analysis (2003 onward), when GA once again produces the highest estimates. On the other hand, DEA produces the lowest productivity estimates of the approaches considered, for the majority of the period of the analysis. This is also apparent in the summary statistics, presented in the tables below.

Table 5.7: Summary statistics of TFP estimates for UK, full period

TFP measure	Average	Standard Deviation	Minimum (%)	Minimum (year)	Maximum (%)	Maximum (year)
DEA	-0.4%	1.7%	-5.7%	1974	2.1%	1987
COLS	0.4%	1.7%	-6.3%	1974	3.0%	1983
SFA translog (exponential)	0.5%	0.7%	-2.0%	1974	1.7%	1983
GA	0.4%	1.8%	-6.8%	1974	3.0%	1983

Table 5.8: Correlations of TFP estimates for UK, full period

	DEA	COLS	SFA translog (exponential)	GA
DEA	1			
COLS	0.92	1		
SFA translog (exponential)	0.92	0.92	1	
GA	0.83	0.97	0.85	1

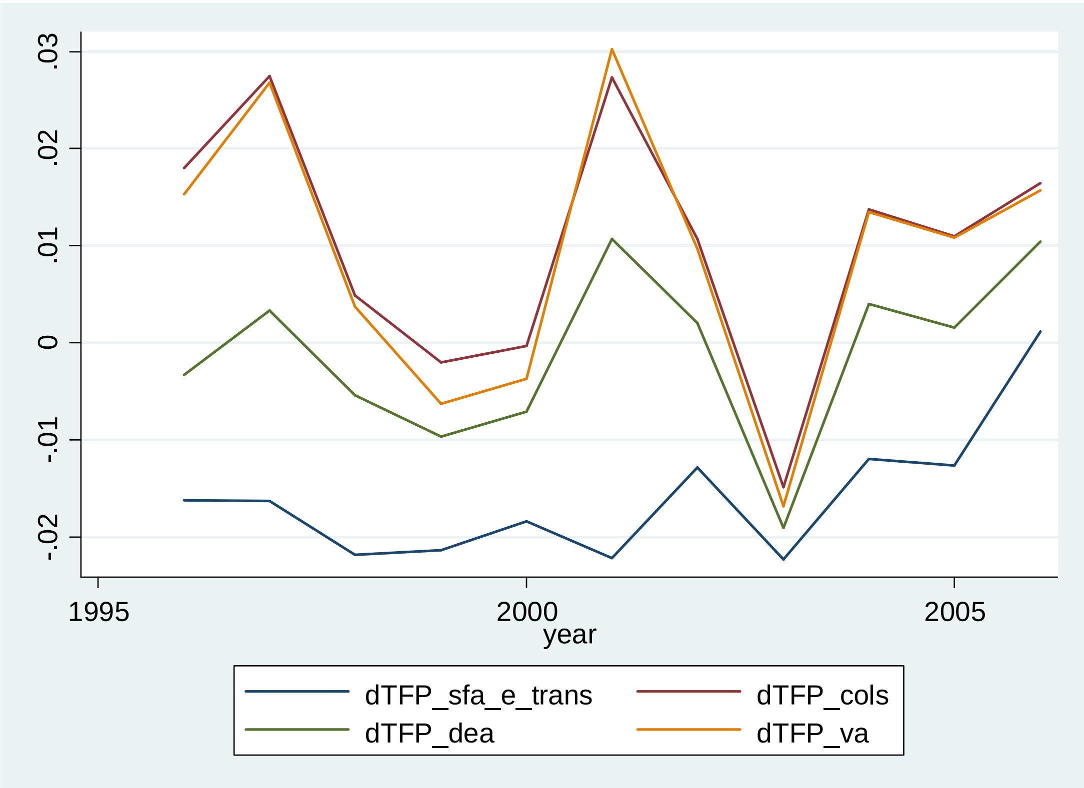
All four approaches agree that the largest productivity contraction that happened in the UK was in the 1973-1974 period, which corresponds to the first Oil Crisis of 1973, when a number of large oil-producing countries initiated an oil embargo. COLS, SFA and GA also agree that the period with the largest productivity expansion was 1982-1983, which follows the recession caused by the global financial crisis of the early 1980’s and the instances of high price inflation and economic and social unrest that took place in the UK in the late 70’s and early 80’s. DEA finds that the period with the highest productivity expansion was 1986-1987; this is not in complete disagreement with the other approaches, since they too display estimates of high productivity growth for this period. In fact, for COLS and SFA, the productivity change in the 1986-1987 period is the second largest observed and for GA the third largest.

Overall, it is interesting to note that all four approaches demonstrate similar patterns in the rate of change of productivity, which roughly correspond to periods of economic cycles; this is true even for DEA, which produces quite lower productivity change estimates. The main disagreement lies in the level of the actual productivity change measure, and again the differences are mainly between DEA and the other three approaches.

5.2.4.2 Slovenia

Slovenia is an interesting case for a number of reasons; it is the smallest economy examined in this chapter (approximately 40 times smaller than the UK) and it is also one of the two ‘transition’ economies (ie countries that moved from centralised to market economies) that provide sufficient data for the purposes of total factor productivity measurement.[79] It should be noted here that despite SVN’s very small scale, the econometric outlier diagnostics undertaken for this analysis did not classify it as a possible outlier and was thus included in the final set of assessed economies. Data for SVN are available from 1995 to 2006.

Figure 5.3: SVN’s productivity change (%), by selected approaches



Note: dTFP_sfa_e_translog = translog SFA (exponential) MI, dTFP_cols = Cobb-Douglas COLS MI, dTFP_dea = DEA MI and dTFP_va = GA-based productivity change (VA as output).

Similar to the UK, SVN’s productivity change estimates from the three deterministic approaches are quite highly correlated. However, this is not the case for the SFA estimates, which display relatively low correlations with all three deterministic

[79] The other ‘transition’ economy is the Czech Republic and it could also be argued that Germany shares a lot of common issues, given that the German reunification between the former West and East Germany happened at the same period and resulted in similar economic circumstances.

approaches. In fact, SVN is the only country in the sample that displays relative low correlations between any of the adopted approaches. In addition to the low correlations, SFA also consistently produces the lowest productivity estimates of the four selected approaches, while COLS and GA produce the highest.

Table 5.9: Summary statistics of TFP estimates for SVN, full period

TFP measure	Average	Standard Deviation	Minimum (%)	Minimum (year)	Maximum (%)	Maximum (year)
DEA	-0.1%	0.9%	-1.9%	2003	1.1%	2001
COLS	1.0%	1.3%	-1.5%	2003	2.7%	1997
SFA translog (exponential)	-1.6%	0.7%	-2.2%	2003	0.1%	2006
GA	0.9%	1.4%	-1.7%	2003	3.0%	2001

Table 5.10: Correlations of TFP estimates for SVN, full period

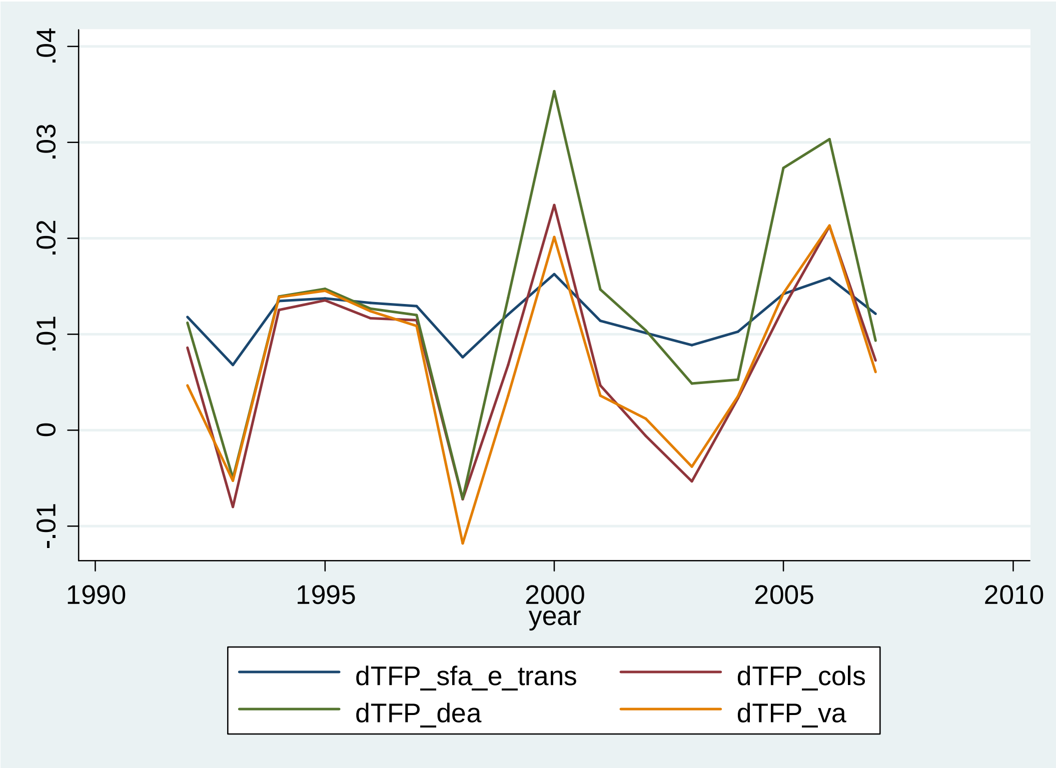
	DEA	COLS	SFA translog (exponential)	GA
DEA	1			
COLS	0.88	1		
SFA translog (exponential)	0.59	0.33	1	
GA	0.90	0.99	0.32	1

All four approaches however agree that the period of the greatest productivity contraction was 2002-2003. The productivity contraction of this period was mainly due to a steep decline in output. This was probably fuelled by the global economic slowdown in the early part of the 21st century, given that Slovenia is a very small economy that is dominated by an export-led manufacturing sector. With regards to the period with the highest productivity expansion, all four approaches provide different answers. What is more troubling is the extent of disagreement between the approaches in the level of productivity change; taking 2000-2001 as an example, COLS and GA estimate that productivity change is between 2.7 and 3%, while under DEA it is 1.1%, a difference of almost 2 percentage points. The differences are even more pronounced when comparing the GA and COLS estimates to SFA estimates; for the 2000-2001, SFA estimates productivity change to be -2.2%, for a total absolute difference of approximately 5 percentage points. The above example may be somewhat extreme, in that the absolute differences between COLS/GA estimates and SFA estimates are largest in this particular period. Nevertheless, it helps illustrate the fact that different approaches can produce divergent estimates and as such, a framework that can be used to select between conflicting estimates can be of great value to the analysis of productivity change.

5.2.4.3 Germany

The case of Germany is another interesting example, mainly presented here to demonstrate the low variability of the SFA productivity change estimates.

Figure 5.4: GER's productivity change (%), by selected approaches



Note: dTFP_sfa_e_translog = translog SFA (exponential) MI, dTFP_cols = Cobb-Douglas COLS MI, dTFP_dea = DEA MI and dTFP_va = GA-based productivity change (VA as output).

As the above figure reveals, the description of Germany's productivity performance over the assessed period greatly depends on the approach selected to measure productivity change.

All three deterministic approaches provide a relatively consistent picture. They detect two instances of productivity contraction in the 90's, the first of which corresponds to macroeconomic adjustments prior to the introduction of the Euro and microeconomic pressures from collectively bargained wage increases, while the second contraction was probably due to the effects of the Asian financial crisis that happened during the 1997-1999 period. These contractions were followed by a brief period of productivity expansion, probably driven again by collective bargaining agreements designed to reduce the rate of wage increases, followed again by a contraction which roughly corresponds to the first financial crisis of the 21st century and the spike in oil prices that was observed in during this period[80].

SFA also maps the effects of these events on GER's productivity performance, but the magnitude of said effects is substantially muted, compared with the estimates from the deterministic approaches.

Table 5.11: Summary statistics of TFP estimates for GER, full period

TFP measure	Average	Standard Deviation	Minimum (%)	Minimum (year)	Maximum (%)	Maximum (year)
DEA	1.3%	1.1%	-0.7%	1998	3.5%	2000
COLS	0.7%	0.9%	-0.8%	1993	2.3%	2000
SFA translog (exponential)	1.2%	0.3%	0.7%	1993	1.6%	2000
GA	0.7%	0.9%	-1.2%	1998	2.1%	2006

Table 5.12: Correlations of TFP estimates for GER, full period

	DEA	COLS	SFA translog (exponential)	GA
DEA	1			
COLS	0.90	1		
SFA translog (exponential)	0.93	0.99	1	
GA	0.89	0.97	0.97	1

As can be seen from the table above, the estimates from all four selected approaches are very highly correlated and all display relatively similar values for average productivity change over the assessed period; COLS and GA find that average productivity change was 0.7% p.a. while DEA and SFA find that it was approximately 1.2%-1.3% p.a. This similarity in the aggregate, together with the high correlation of the estimates might lead to the conclusion that all four approaches provide a similar view of productivity performance. However, this is not completely true in this case; SFA finds that GER's productivity performance is quite stable, while the deterministic approaches provide estimates that are much more varying.

If the analysis is only interested in deriving a long-tem estimate of productivity change, the choice of approach would not have a great impact on the final result, given that all four approaches provide relatively similar average estimates. However,

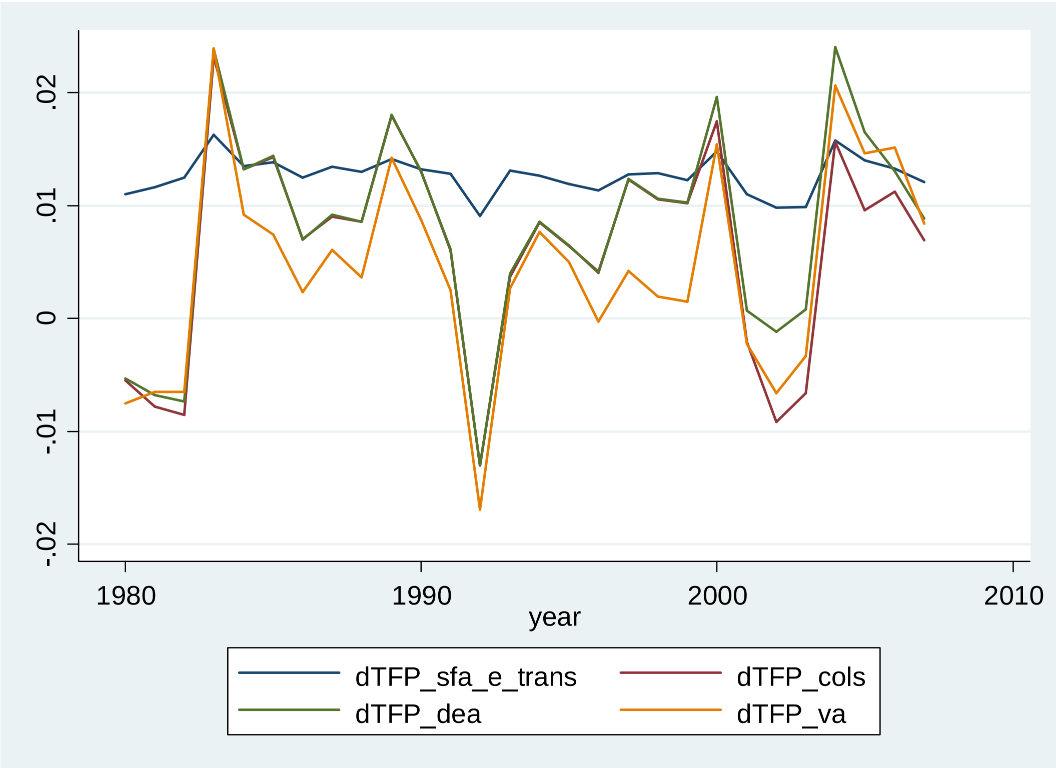
80 For an extensive discussion on Germany's economic performance in this period, see DG-Economic and Finance (2002), 'Germany's growth performance in the 1990's', http://ec.europa.eu/economy_finance/publications/publication1878_en.pdf (accessed 12 April 2013).

if the analysis is more concerned about shorter timeframes (e.g. when the goal is to analyse the effects of a certain policy or a change in the economic environment), then the choice of an approach would in fact be critical to the final outcome.

5.2.4.4 Netherlands

The case of the Netherlands is similar to Germany, in that all three deterministic approaches produce similar productivity estimates, but the SFA estimates are quite different. The difference with the case of Germany is that the SFA estimates are not only less volatile here but are also significantly higher relative to their deterministic counterparts.

Figure 5.5: NLD's productivity change (%), by selected approaches



Note: dTFP_sfa_e_translog = translog SFA (exponential) MI, dTFP_cols = Cobb-Douglas COLS MI, dTFP_dea = DEA MI and dTFP_va = GA-based productivity change (VA as output).

Overall, the economic performance of the Netherlands in the 80's is consistent with global economic developments; the country experienced a spike in productivity growth after the second Oil Crisis in the late 70's and had a relative stable period of productivity growth throughout the decade. There were two major periods of productivity contraction in the later two decades: The first was in the 1991-1992 period which was probably due to the global economic recession that happened at this time; the large productivity contraction is not surprising, given that during this time the economy of the Netherlands was 'outward-facing', with significant contributions to output coming from international trade (Europe's largest cargo port is in Rotterdam and Amsterdam's airport was at the time the second largest in Europe) and foreign investment (favourable corporate tax incentives for foreign investments). The second period of major productivity contraction was in the 2001-2002 period, which again corresponds to a period of global economic contraction. The productivity decline in this period lasted longer than the majority of the other assessed countries, probably due to structural issues that the Dutch economy was experiencing at the time[81].

Table 5.13: Summary statistics of TFP estimates for NLD, full period

TFP measure	Average	Standard Deviation	Minimum (%)	Minimum (year)	Maximum (%)	Maximum (year)
DEA	0.8%	0.9%	-1.3%	1992	2.4%	2004
COLS	0.6%	0.9%	-1.3%	1992	2.3%	1983
SFA translog (exponential)	1.3%	0.2%	0.9%	1992	1.6%	1983
GA	0.4%	0.9%	-1.7%	1992	2.4%	1983

Table 5.14: Correlations of TFP estimates for NLD, full period

	DEA	COLS	SFA translog (exponential)	GA
DEA	1			
COLS	0.96	1		
SFA translog (exponential)	0.87	0.87	1	
GA	0.96	0.91	0.90	1

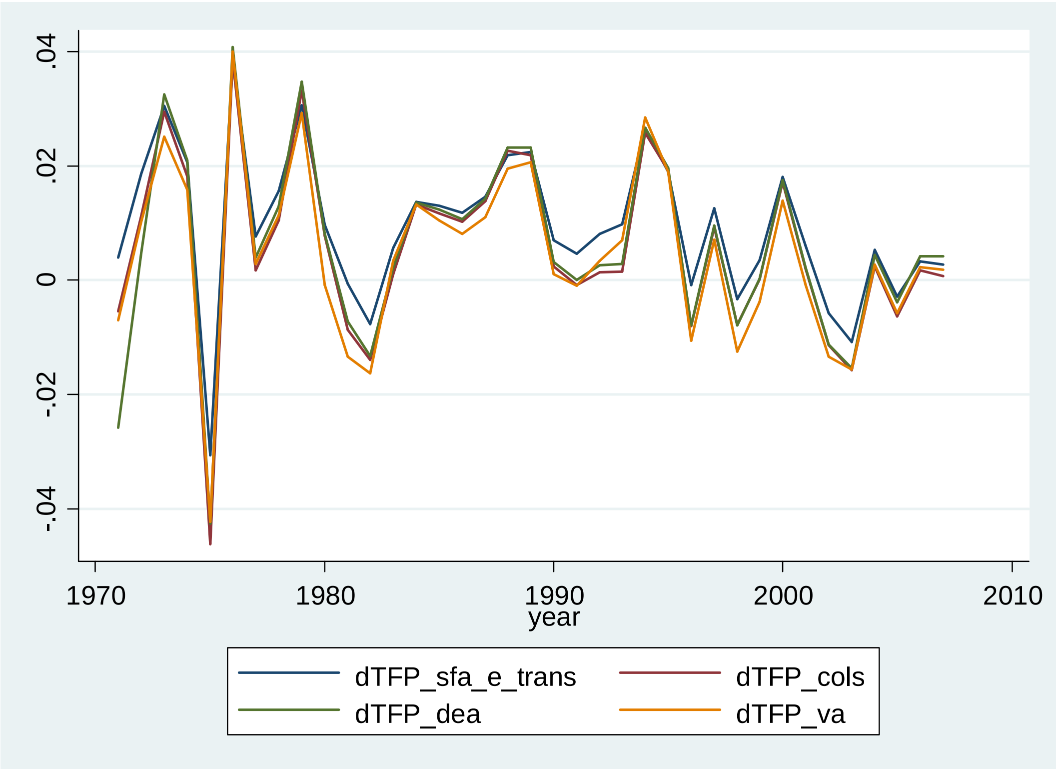
5.2.4.5 Italy

The productivity performance of Italy is presented here to provide a counter-point to the case studies examined thus far.

As was previously mentioned, the countries that were examined in more detail here were selected in such a way as to demonstrate that the choice of a productivity measurement approach can have a significant impact in the productivity estimates, even if this not immediately apparent by looking at summary statistics and correlation measures for the whole panel. On the other hand, there are also countries, such as Italy, for which all approaches provide a consistent view of their productivity performance.

[81] For a more detailed look at the recent economic performance of the Netherlands, see Albers, R. and Langedijk, S. (2004), 'The Netherlands: from riches to rags?', ECFIN Country Focus series

Figure 5.6: ITA's productivity change (%), by selected approaches



Note: dTFP_sfa_e_translog = translog SFA (exponential) MI, dTFP_cols = Cobb-Douglas COLS MI, dTFP_dea = DEA MI and dTFP_va = GA-based productivity change (VA as output).

The similarity in the estimates of all four approaches is evident in the summary statistics of the estimates and in their detailed annual movements, as demonstrated by the figure above and is also confirmed by the very high correlations between the examined approaches. Overall, Italy's productivity performance in the 70's and early 80's is greatly influenced by the two major oil crises that took place in this period; a similar effect was also observed in the UK's productivity performance and indeed all of the developed economies included in this analysis for which data was available for this period. Productivity recovered in the mid 80's but suffered a blow in the start of the 90's, probably due to Italy's adoption of the European Exchange Rate Mechanism and internal fiscal adjustments and structural changes designed to reduce public deficit[82]. Productivity briefly recovered during the early 90's, but from then on Italy's productivity performance has been quite volatile (although not at the scale observed in the 70's), with average productivity change in the 1995-2007 period between -0.1% (GA) to 0.1% (DEA).

[82] For a discussion on Italy's macroeconomic performance in the 90's and the early part of the 21st century, see DG-Economic and Finance (1999), 'Italy's slow growth in the 1990s: Facts, explanations and prospects', available at http://ec.europa.eu/economy_finance/publications/publication8097_en.pdf (accessed 12 April 2013).

Table 5.15: Summary statistics of TFP estimates for ITA, full period

TFP measure	Average	Standard Deviation	Minimum (%)	Minimum (year)	Maximum (%)	Maximum (year)
DEA	0.6%	1.7%	-4.3%	1975	4.1%	1976
COLS	0.6%	1.6%	-4.6%	1975	3.9%	1976
SFA translog (exponential)	0.9%	1.3%	-3.1%	1975	3.8%	1976
GA	0.4%	1.6%	-4.2%	1975	4.0%	1976

Table 5.16: Correlations of TFP estimates for NLD, full period

	DEA	COLS	SFA translog (exponential)	GA
DEA	1			
COLS	0.97	1		
SFA translog (exponential)	0.95	0.99	1	
GA	0.96	0.98	0.98	1

5.2.4.6 Summary of country analysis

As demonstrated above, the various measurement approaches can sometimes produce quite dissimilar productivity change estimates. These, sometimes pronounced, differences can be problematic, if such analyses were to be used to inform policy. It is quite likely that a policy maker, upon being presented such results would enquire as to why do the various estimates differ and, more importantly, which estimate is likely to be more accurate. The framework described in section 5.3 aims to facilitate this selection process.

5.2.5. Decomposition of productivity change

This section provides a brief discussion on the results of the analysis with regards to the individual components of the productivity change estimates derived from the various approaches. As mentioned in chapter 3, one of the main advantages of the frontier-based approaches over GA is that they can all provide robust methods that can be used to identify the ‘sources’, or components, of productivity change. The analysis undertaken here decomposes the MI productivity index into its three main components, namely efficiency change (EC), technological change (TC) and scale efficiency change (SE). For a detailed description of how this decomposition is achieved for each approach, refer to chapter 3.

The following table presents the summary statistics of the three productivity change components for three approaches; the DEA-based circular Malmquist index (DEA), the Malmquist index derived from the translog, exponential SFA model and the Malmquist index derived from the Cobb-Douglas COLS model. Similarly to the individual case studies examined in the pervious section, the decision to limit the reporting of the results to these particular approaches was based on the diagnostic analysis for the econometric models.

Table 5.17: Summary statistics of productivity change components

		Full sample period (1970-2007)			1995-2007		
		DEA	SFA	COLS	DEA	SFA	COLS
Average Efficiency Change	Obs	361	361	361	178	178	178
	Mean	-0.3%	-0.1%	-0.1%	-0.3%	-0.3%	-0.2%
	Standard Dev	3.7%	1.0%	1.4%	1.2%	0.9%	1.2%
	Min	-57.8%	-4.4%	-7.1%	-3.6%	-3.4%	-4.2%
	Max	4.2%	3.2%	3.5%	3.4%	2.9%	3.5%
Average Technological change	Obs	361	361	361	178	178	178
	Mean	0.7%	1.0%	0.8%	0.7%	1.1%	0.8%
	Standard Dev	3.8%	0.3%	0.0%	1.3%	0.3%	0.0%
	Min	-8.1%	0.2%	0.8%	-4.6%	0.3%	0.8%
	Max	56.3%	1.6%	0.8%	4.0%	1.6%	0.8%
Average Scale efficiency change	Obs	361	361	361	178	178	178
	Mean	0.2%	0.0%	0.0%	0.2%	-0.1%	0.0%
	Standard Dev	0.8%	0.2%	0.0%	0.5%	0.3%	0.0%
	Min	-4.6%	-1.3%	0.0%	-1.0%	-1.3%	0.0%
	Max	2.5%	0.7%	0.0%	2.3%	0.5%	0.0%

As was the case with the aggregate productivity index, the three selected approaches seem to provide a uniform view of productivity performance in the sample. All find that over both the full period of the analysis and during the shorter 1995-2007 period, there has been a slight decrease in overall efficiency, almost no variation in scale efficiency and thus the main driver of productivity change has been technological change.

The fact that efficiency change has been slightly negative or stable suggests that the performance of the assessed countries on average is sufficient to keep up with the improvements in technology, but not enough to push inefficient countries closer to the frontier. In other words, the results suggest that over the assessed period there has not been any substantial convergence towards the frontier; it should be stressed however that additional analysis would be required in order to provide firm evidence on the issue of convergence.

The most important driver of the productivity growth on aggregate has been technological change, ie improvement in productivity achieved by countries that were already assessed as fully efficient or close to that. In more detail, the countries that are either fully efficient or close to that under both DEA and SFA and also display significant improvements in technological change are Sweden, the Netherlands and USA.

Scale efficiency change has little impact on overall productivity change, at least when considering the sample in aggregate. This is not surprising, given that the units of the analysis are entire economies; even smaller countries cannot easily, and certainly not quickly, change their scale. This is in fact supported by the data, which show quite low variation both in the levels of inputs and the levels of output from one year to the next for an individual country. This is be demonstrated in more detail in section 3.4.1 of this chapter.

It should be noted that COLS shows no variation in either the technological change or the scale efficiency change component; this is due to the fact that the COLS model presented here assumes Cobb-Douglas technology, which by construction exhibits global returns to scale and unit/time-invariant technological change. On the other hand, both DEA and SFA utilise frontiers that allow for both unit- and time-variable scale effects and technological change; DEA does this by construction and SFA achieves that by assuming that the technology is translog. Due to this limitation of the selected COLS model, the discussion in this section will focus on the DEA and SFA results from here on.

The productivity components derived from the SFA model are significantly less volatile compared to their DEA counterparts, at least when the focus is on the full period of the analysis; similar behaviour was also observed in the estimates for the productivity index itself. It is interesting to note that this disparity in the variation of the estimates lessens quite considerably when considering the 1995-2007 period. This suggests that a possible reason for the difference in the variability of the components is the frequent economic shocks that the assessed countries experienced in the earlier period of the analysis. DEA is a deterministic approach and as such the effects of these shocks are starkly reflected in the productivity index and its components; SFA on the other hand moderates the impact of these shocks, by assigning some of their effects to the stochastic element. Examining the more recent 1995-2007 period, when economic conditions were more stable, the analysis finds that the differences in the volatility between the SFA and DEA estimates are less pronounced.

Another interesting point is the extreme minimum and maximum value of the efficiency change and technological change component respectively under DEA. Further examination reveals that these values are both estimates for Japan and that they both correspond to the 1977-78 period. In general, the DEA efficiency change and technological change estimates for that particular year are quite inflated for all assessed countries; the reason for that is that the 1977-78 period was the first period for which data for USA are available. This is important because the introduction of USA data cause a considerable upset in output VRS efficiency scores (which inform the DEA efficiency change estimates); for example, Japan in 1977 was assessed as 100% output VRS efficient but with the introduction of USA data the VRS frontier shifted outward by such a margin that Japan in 1978 was assessed as only 57% output VRS efficient. This outward shift of the frontier is also the cause for the extreme positive technological change estimates observed in this particular period.

Despite the similarity of the estimated components of productivity change from both approaches in the aggregate, closer examination of the results of the analysis shows that the patterns of change are significantly different. The table below presents the correlation coefficients between the DEA- and SFA-based estimates.

Table 5.18: Correlation coefficients between the DEA- and SFA-derived productivity change components

Efficiency Change	Full period	0.44
	1995-2007	0.58
Technological change	Full period	0.09
	1995-2007	0.32
Scale efficiency change	Full period	0.00
	1995-2007	0.02
Efficiency plus Scale efficiency change	Full period	0.39
	1995-2007	0.47

The table above clearly demonstrates that there is relatively low correlation between the DEA- and SFA-based estimates of the productivity change components. The only component with a correlation coefficient that approaches 0.5 is efficiency change; for technological change and scale efficiency change, there is almost no correlation between the different approaches, at least when considering the estimates for the full period. This is a very interesting result, especially when the correlation in the overall productivity change estimates from the two approaches was quite high; 82% for the full period and 81% for the 1995-2007 period. The correlation coefficients of the components are higher for 1995-2007 period (probably due to the lower variation of the DEA estimates in that period), but are still markedly lower than the correlation coefficients of the overall productivity change estimates.

One hypothesis for the low correlations between the components is that DEA allows for a production frontier that is more sensitive to changes in productivity, since it is informed only by the performance of the assessed units in a single period. Due to this sensitivity, it could be possible that part of the productivity change that is attributed to scale efficiency change under DEA might be detected as efficiency change in the translog SFA models. To assess if this is the case, the analysis also examined the correlation of an aggregate measure of efficiency and scale efficiency change (defined as the sum of efficiency and scale efficiency change estimates). The results of this simple crosscheck do not support the initial hypothesis; the correlation of the aggregate measure from the DEA and SFA approaches is remains quite low, lower in fact than the correlation of the efficiency change measure.

The relatively low similarity of the estimates for the productivity components is also apparent when examining the evolution of these measures over time:

Figure 5.7: DEA- and SFA-based Efficiency change (%), by year

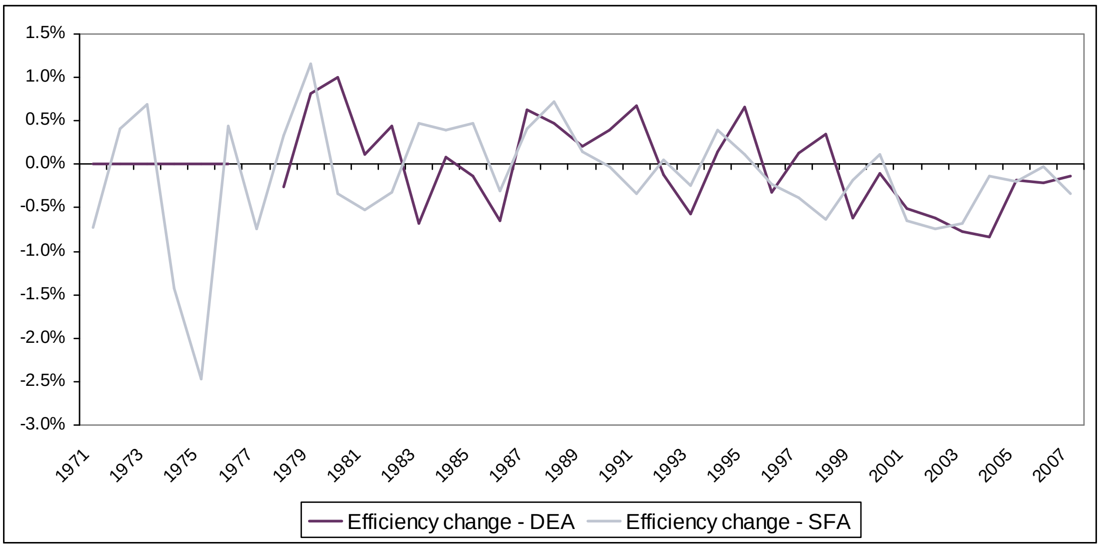


Figure 5.8: DEA- and SFA-based Technological change (%), by year

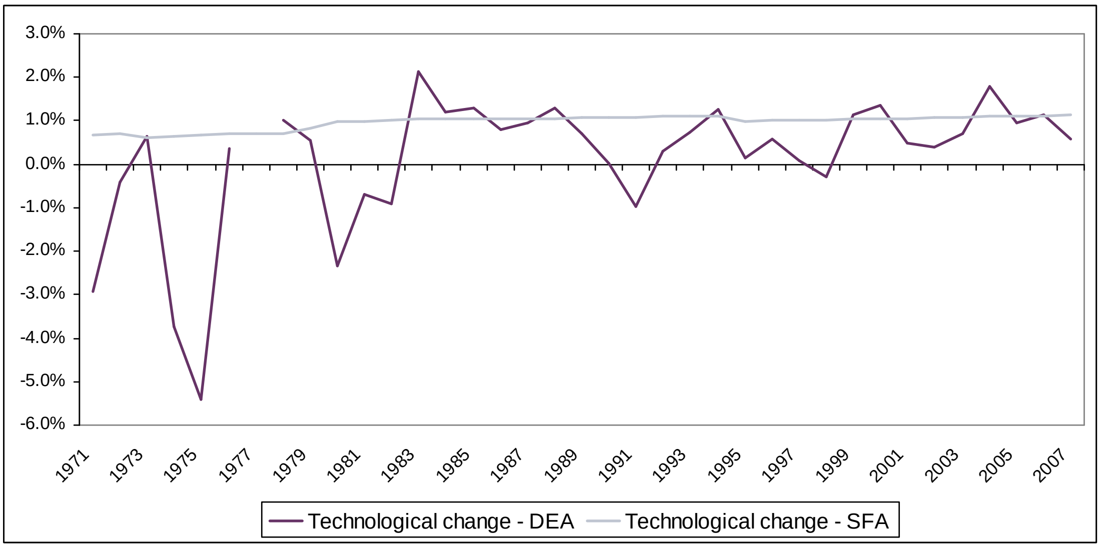


Figure 5.9: DEA- and SFA-based Scale efficiency change (%), by year

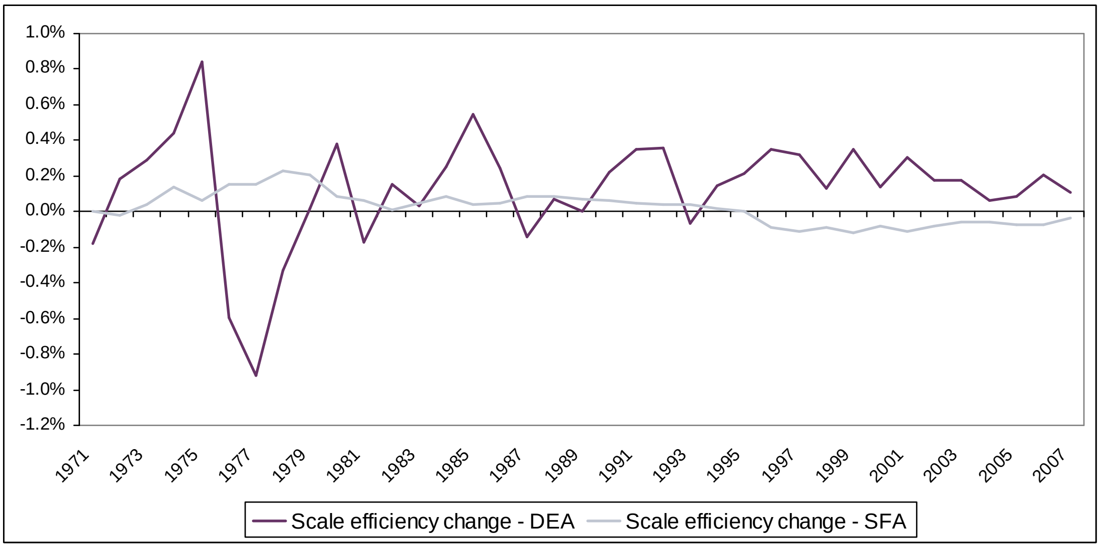
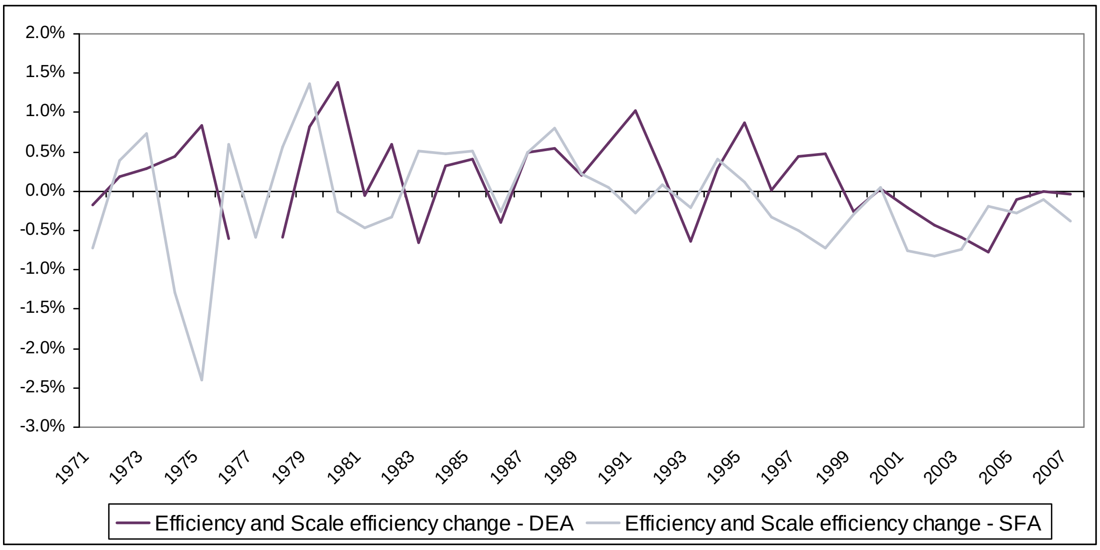


Figure 5.10: DEA- and SFA-based Efficiency and Scale efficiency change (%), by year



Note that for the above figures, the DEA-based estimates of efficiency and technological change for the 1977-78 period are omitted, due to the extreme values for these estimates in that particular period; this is done here to increase the readability of the figures.

The differences between to two approaches are even more pronounced when examining the performance of individual countries.

Table 5.19: Averages of the components of productivity change, by country

	Efficiency Change		Technological change		Scale efficiency change		Productivity change	
	DEA	SFA	DEA	SFA	DEA	SFA	DEA	SFA
AUS	**-0.5%**	**-0.1%**	**1.1%**	**1.1%**	**0.2%**	**0.0%**	**0.8%**	**0.9%**
AUT	0.2%	0.0%	1.1%	1.4%	-0.2%	0.0%	1.2%	1.4%
CZE	**0.0%**	**-0.1%**	**0.1%**	**0.4%**	**0.7%**	**-0.2%**	**0.9%**	**0.1%**
DNK	-0.3%	-0.3%	1.2%	1.5%	0.1%	-0.1%	1.0%	1.2%
ESP	-0.6%	-0.6%	0.6%	1.1%	0.3%	0.1%	0.3%	0.6%
FIN	**0.0%**	**-0.1%**	**0.6%**	**1.3%**	**-0.4%**	**-0.1%**	**0.1%**	**1.1%**
GER	**0.5%**	**0.1%**	**0.4%**	**1.1%**	**0.4%**	**0.0%**	**1.3%**	**1.2%**
ITA	-0.1%	-0.2%	0.4%	1.0%	0.4%	0.1%	0.6%	0.9%
JPN	**0.1%**	**0.2%**	**-0.3%**	**0.8%**	**0.7%**	**0.2%**	**0.5%**	**1.2%**
NLD	-0.2%	-0.1%	1.0%	1.4%	0.0%	0.0%	0.8%	1.3%
SVN	0.0%	-1.1%	-0.1%	0.6%	0.0%	-1.1%	-0.1%	-1.6%
SWE	0.0%	0.0%	0.3%	1.0%	0.1%	-0.1%	0.3%	0.9%
UK	**0.0%**	**-0.2%**	**-0.2%**	**0.6%**	**-0.2%**	**0.1%**	**-0.4%**	**0.5%**
USA	0.0%	-0.1%	0.0%	0.7%	0.6%	0.3%	0.6%	0.9%

The countries in bold signify situations where the aggregate results (averaged estimates) from each approach are noticeably different, for at least one of the components of the productivity index:

– For Austria (AUS), DEA finds that efficiency change had a relatively large negative effect on productivity growth; SFA also finds a negative effect, but at a lower magnitude. For DEA, this negative effect of efficiency change is somewhat lessened by the positive contribution of the scale efficiency component. The net result is that both approaches result in a very similar estimate for overall productivity change.

– For the Czech Republic (CZE), the area of major disagreement between the two approaches is scale efficiency change; the DEA results suggest a large positive contribution for this component, while SFA finds that scale efficiency change negatively impacts the overall productivity change estimate. This, together with the smaller differences in the technological change component result in DEA concluding that CZE's productivity growth is above average, while SFA finds that it is below average (average productivity change under both approaches is between 0.6% and 0.7%).

- For Finland (FIN), the two approaches mainly agree on the direction of the effect for all three components, but there is significant disagreement with regards to the magnitude of the effects. The end effect is observed in the substantial differences in the estimates for overall productivity change (0.1% for DEA, 1.1% for SFA).

- The case for the Germany (GER) is somewhat similar to Austria, in that both approaches result in productivity growth estimates that are very close in level terms. However, they are not in agreement on how this productivity growth was achieved; DEA assigns equal importance on all three components, while SFA finds that GER's productivity improvement was almost solely due to technological change.

- Japan (JPN) is similar to the Czech Republic, in that there is substantial disagreement in the levels of estimated scale efficiency change; DEA finds a significant positive contribution, while SFA finds a much smaller positive effect. Additionally there is also disagreement in the levels and the direction of the effect of technological change; DEA suggests that the effect is negative while SFA finds a relatively large positive effect. The end result is a somewhat large discrepancy in overall productivity change estimates.

- Lastly, the estimates for the United Kingdom show that there is disagreement in the direction of the effect for all three components, although this is more pronounced in the estimates for technological change. This results in productivity change estimates that are very dissimilar (positive for SFA and negative for DEA).

To conclude, the estimates of the components of the productivity indices derived from the two selected approaches examined in this section are quite different in the majority of the countries assessed. This is an interesting finding because the overall productivity change estimates from these two approaches are quite similar in levels and are also quite highly correlated. A closer examination of the rates of change of the components in individual countries revealed that there is no apparent systematic pattern that could help explain the differences in the findings. As such additional analysis would be required to examine the likely causes of these differences; this could be the subject of further research into this area.

5.3. Selecting between approaches

As the previous section demonstrated, the various productivity measurement approaches can produce estimates that are quite different form each other, especially when considering the results at an individual country or economy level.

There are a number of possible explanations as to why the various productivity change estimates differ; after all, they are produced by different approaches that adopt different perspectives and sets of assumptions to measure productivity change (see chapter 3 for discussion). Therefore, small, and in some cases not-so-small, differences in the various estimates can be quite common in real-life productivity measurement studies[83]. In other words, although it would be ideal if the analysis revealed that there is a strong consensus between the different estimates, at least some degree of disagreement is to be expected. If the level of disagreement is large, as is the case for some of the countries assessed in the previous section of this chapter, the key question is which estimates are likely to be more accurate.

The issue of selecting a set of estimates that are likely to be more accurate is quite complex. One could select a set of estimates based on arguments grounded on the theory behind the adopted approaches; for example, one could discount the use of GA because it adopts a set of assumptions that are, at first glance, significantly more restrictive than the frontier-based approaches. The problem with this type of argumentation is that all measurement approaches rely on assumptions and that it is not known beforehand what would be the impact to the accuracy of the estimates when one or more of those assumptions are violated. To assess that impact, one would need additional evidence, such as those provided by simulations. Even that avenue of research however might not be able to provide clear-cut answers; the simulation analysis in the previous chapter revealed that the accuracy of the assessed approaches heavily depends on a number of characteristics inherent in the data generation process that manifest directly in the dataset.

Therefore, the key issue here is how to identify and assess the relevant conditions/characteristics prevalent in the dataset at hand. If this can be done with some degree of certainty, the findings of the simulation analysis can be more readily applied in a practical setting, so that the analysis can make an informed decision on

[83] See for example Coelli et al. (2005).

the approach, or a subset of approaches, that are likely to provide the most accurate TFP change estimates for the current analysis.

This reasoning can provide a blueprint of a methodological framework for selecting between competing estimates. Such a framework would require three main steps:

– First, identify the characteristics, features and/or conditions (from here on referred to as simply the characteristics of the dataset) pertinent to the analysis that can have a significant impact on the relative accuracy of the various productivity measurement approaches. As mentioned above, this can be based on theory, but would ideally also utilise additional evidence from controlled experiments (ie simulations), as these provide a more complete, quantified view of the issue.

– Second, identify methods and techniques that can be used to detect the presence of such characteristics in the application at hand and quantify them if possible.

– Lastly, combine the results of the second step with prior knowledge on the performance of the various approaches under said characteristics, in order to select the productivity estimates that are likely to be more accurate in the current application.

This selection framework relies heavily on findings of simulation experiments, both for the first step (identifying the pertinent characteristics) and for the third step (selecting the most likely accurate approach). Note that this general blueprint of a selection framework is not necessarily confined in applications of productivity measurement; indeed, it can be easily modified to assist in the selection of efficiency measurement approaches. In such cases, the analysis needs to choose characteristics that affect the estimates of efficiency (rather than productivity) and rely on simulation findings that focus on this aspect of performance measurement (efficiency measurement rather than productivity measurement). However, since the focus of this thesis is productivity measurement, the discussion here will be limited to this particular area. Further more, since simulation analyses focusing on productivity measures are sparse in the academic literature, the practical application of the first and third step of the proposed framework will be based on the findings of the simulation analysis undertaken in chapter 4. For the second step (identification and quantification of characteristics in the dataset), this section proposes the use of a set of readily available diagnostic tests and/or indicators. These are discussed in more detail below.

The rest of this section is structured as follows:

- Section 5.3.1 provides a brief discussion on identifying the characteristics of interest.
- Section 5.3.2 suggests the use of a number of diagnostic tests and indicators sourced from the productivity performance analysis to detect and quantify said characteristics.
- Section 5.3.3 discusses how the findings of the first two steps and can be used to select between competing estimates.
- Section 5.3.4 demonstrates the use of this framework on the estimates from the EU KLEMS dataset, from section 5.2 of this chapter.

5.3.1. First step: Identifying the characteristics of interest

The simulation analysis undertaken in the previous chapter revealed that the most influential characteristics of the data generating process (DGP) to the accuracy of the examined approaches are:

- the extent of volatility in inputs from one year to the next. Increased volatility adversely affects the accuracy of all approaches, but DEA-based estimates are the least affected, while the GA estimates are the most affected;

- the extent of inefficiency present in the sample. Increased levels of technical inefficiency have only a small negative effect on the accuracy of COLS- and DEA-derived productivity estimates, but a larger impact on GA and SFA-based estimates (for the SFA estimates the change in accuracy is co-dependent on the extent of measurement error-noise in the data);

- the extent of measurement error/noise in the data. Increased levels of noise have detrimental effect on the accuracy of all approaches, but the overall effect depends on the extent of technical inefficiency also present and on whether the parametric approaches are likely to suffer from functional form misspecification;

- whether the parametric approaches are likely to suffer from functional form misspecification; this has a severe negative impact on the accuracy of all parametric approaches.

5.3.2. Second step: Assessing the characteristics of interest

The second step of the proposed selection framework is to assess how prevalent the above characteristics are (if at all) in the application/dataset at hand. This assessment is not straightforward; in fact, it is impossible to determine with certainty at least some of the characteristics in question, given that all productivity measurement approaches examined rely on certain implicit or explicit assumptions, which are made prior to the actual analysis and directly influence the estimates used to assess said characteristics. For example, both COLS and DEA are deterministic approaches, in a sense that they do not include a stochastic element directly in the estimation process. Another example is that most frontier-based approaches automatically assume that there is some inefficiency (SFA is the exception, as it can test for the presence inefficiency), while GA assumes that there is no inefficiency in the sample data.

Despite the above concerns, there are a number of simple diagnostic tests/indicators that can provide useful information on the presence or prevalence of the characteristics in question. These are described below and their efficacy is tested (through simulation analysis) in section 5.3.4.

5.3.2.1 Input volatility

This is relatively easy to assess by simply examining the summary statistics (namely average values, standard deviations and coefficients of variation) of the annual change in inputs of each assessed unit.

5.3.2.2 Technical inefficiency

The various frontier-based approaches can readily provide estimates of technical inefficiency even when the focus of the analysis is to examine productivity change; the analysis can make use of these estimates to assess the possible extent of technical inefficiency in the dataset/application at hand. Since it is not known which approach is likely to provide the most accurate efficiency estimates, this characteristic can be quantified by simply averaging the different efficiency estimates derived from all adopted approaches.

In general, it is difficult to assess the extent of technical inefficiency with a high degree of accuracy, since the performance measurement approaches examined measure it residually. As such, the way they construct the efficiency frontier (or the production possibility set) will always have an impact on the final performance measure (efficiency or productivity estimate). Nevertheless, although perfect accuracy is out of reach, there is a large and growing pool of evidence in the literature that suggests that technical inefficiency is present in the economy (see for example, Fried et al.(2008), op cit.) and that frontier-based approaches can, in most cases, measure such inefficiency with a degree of accuracy sufficient for the purposes of the selection framework (the literature review on simulation studies for efficiency estimates in Chapter 4 and in Appendix 1 provides an overview of the efficacy these approaches).

5.3.2.3 Noise levels

From the approaches examined, only SFA directly incorporates a stochastic element in the estimation process, intended to capture the impact of measurement error/statistical noise; as such, overall noise levels in the DGP could be measured by the estimated standard deviation of the noise component, denoted as σ_v, which can easily be extracted from the SFA models[84].

The issue with relying on this estimate is that although σ_v is unbiased, it is also inconsistent in the pooled setting (because it is independent of *i*, ie the observation whose technical efficiency is to be estimated, see Kumbhakar and Lovell (2000), op cit.). It is not clear whether this issue would materially affect the accuracy of the estimator, at least for the purposes of this selection framework; to explore this further, additional simulations are undertaken in this chapter to examine how reliable are the estimates of σ_v under certain conditions (see section 3.4.3). The simulation analysis revealed that the estimated σ_v is reasonably accurate under conditions similar to those observed in the EU KLEMS dataset and can thus be used as an indicator of the overall noise levels in this particular application.

5.3.2.4 Functional form misspecification

This issue relates only to the parametric frontier-based approaches and could be examined through the use of RESET (Ramsey's Regression Equation Specification Error Test) and by investigating the statistical significance of the input coefficients.

RESET is one of the most widely-used tests to detect the presence of functional form misspecification. The test examines whether the inclusion of non-linear combinations of either the fitted values of the regression model or the model's explanatory variables are statistically significant when included in the original regression model[85].

---

[84] The noise component/variable is assumed to be normally distributed with a mean value of zero, by construction.
[85] For the EU KLEMS application, the analysis adopted the standard quadratic combinations.

If they are, then the regression model is likely to suffer from some form of misspecification. RESET is quite powerful and can offer compelling evidence, but can only be applied when the regression model is estimated using OLS (ordinary least squares). However, since the input coefficients from the SFA models are consistent estimates of the respective OLS input coefficients, the findings of RESET as applied in the OLS regression model also apply for the SFA model. In fact, it is not uncommon in SFA studies to first estimate the equivalent OLS models solely for the purpose of applying RESET to test for misspecification.

Examining the statistical significance of the input coefficients offers a more qualitative assessment on the possible existence of functional form misspecification; the intuition behind it is that if some of the input coefficients are found to be statistically insignificant, the adopted functional form does not match exactly to the underlying data generation process and as such the parametric model in question **could** be misspecified. It should be mentioned that there could be a number of reasons why a variable could be assessed as being statistically insignificant even though it is in fact part of the DGP; these include extensive noise in the data or multi-collinearity amongst the various explanatory variables. Therefore, statistically insignificant variables in this context do not necessarily imply that the model is misspecified; they are however an indicator that the current parametric model might suffer from a number of possible shortcomings that could affect the accuracy of the derived productivity estimates.

To ensure that the results of RESET and the more qualitative assessment based on the statistical significance of the model variables are good indicators of functional form misspecification, a new round of simulations is undertaken for this chapter. The full results of this analysis can be found in section 3.4.4; in summary, the analysis found that these simple tests can indeed provide valuable insight regarding this issue.

5.3.3. Third step: Selecting between approaches

After assessing the prevalence of the above characteristics in the current dataset/application, the third and final step of the selection framework is to determine which of the assessed approaches offers the more accurate productivity change estimates under these specific conditions. This can be achieved in two ways: the analysis could either rely on the findings of previous simulation studies that specifically assess the overall accuracy of different approaches under these conditions, or an original simulation analysis could be undertaken, which uses a DGP specifically tailored to the application currently considered. The advantage of relying on already existing studies is simplicity and ease of implementation; however, this might come at the cost of accuracy, in the event that the DGPs adopted by the existing studies do not closely match the characteristics of the current application. The next section demonstrates how the findings of a simulation analysis with a DGP constructed to be similar to the observed characteristics of the application in hand (in other words, the simulation analysis of chapter 4 and the EU KLEMS dataset) can be used to inform the selection process.

It should be mentioned that even the DGP of an original simulation analysis will not be able to capture all of the peculiarities of the current application. The aim should of course be to construct it in such a way as to be as similar as possible with the current application; so, the simulations' DGP should include the same number of inputs and outputs, similar number of available observations (units and time periods) and similar volatility, noise and inefficiency characteristics as the current dataset. In addition, if the diagnostics find that the parametric approaches show evidence of functional form misspecification even when flexible functional forms are adopted, the analysis should use non-smooth functional forms (such as piecewise-linear functions[86]) for the simulation DGP to ensure that the parametric approaches in the simulation analysis also suffer from functional form misspecification. Nevertheless, there will always be some degree of uncertainty, since the analysis cannot have full knowledge of the underlying DGP of the current application (if it did, it wouldn't need to estimate it). For example, relatively accurate estimates of the mean and standard deviation of technical inefficiency of the assessed units might be achievable, but its actual distribution is unknown and cannot be derived from the available models; similarly, if there is evidence of misspecification and a non-smooth function is used for the simulation DGP, the adopted function will not necessarily be representative of the true underlying DGP of the current application.

It should be stressed however that these gaps in our understanding of the underlying DGP of our application are only an issue if they negatively affect our ability to draw useful conclusions from the simulation analysis, be it either original or sourced from previously published studies. In other words, these characteristics are only important in so much as they affect the accuracy of the resulting productivity estimates. According to the findings of chapter 4, neither of the two examples given above were

[86] When creating these functions, the analysis also needs to consider whether to impose the various restrictions suggested by theory, such as monotonicity and concavity.

found to have a material effect in the relative accuracy of the approaches examined; the SFA-based estimates were very similar regardless of the distributional assumptions made by the models, while the parametric models displayed similar loss in accuracy under a number of different piecewise-linear DGPs.

That is not to say that the four characteristics included in the proposed framework are the only characteristics that are likely to significantly affect the relative accuracy of the productivity estimates. In fact, issues such as latent heterogeneity in the assessed units (which could manifest as heteroskedasticity in the parametric models) and variable returns to scale could also be significant. However, the simulations undertaken in chapter 4 did not examine how such factors affect the relative accuracy of the various approaches; as such, the assessment of those characteristics is left for future research.

5.3.4. Applying the selection framework to the EU KLEMS dataset

5.3.4.1 Assessing input volatility

Input volatility is the easiest characteristic to assess; this is achieved by simply examining the annual change in inputs by country. Average input growth and its standard deviation is summarised in the table below.

Table 5.20: Average annual growth in inputs

Country	Average Growth in Labour	Standard deviation of Labour growth	Average Growth in Capital	Standard deviation of Capital growth
AUS	2.29%	1.75%	3.81%	1.44%
AUT	0.71%	1.24%	2.37%	0.29%
CZE	0.32%	1.73%	2.84%	0.33%
DNK	0.72%	1.57%	1.55%	0.92%
ESP	2.19%	2.59%	3.44%	0.87%
FIN	0.84%	2.17%	3.94%	2.26%
GER	-0.35%	1.21%	2.54%	0.64%
ITA	1.04%	1.02%	2.74%	1.13%
JPN	0.64%	1.28%	4.66%	1.93%
NLD	1.42%	1.38%	2.35%	0.43%
SVN	0.89%	2.18%	6.17%	0.76%
SWE	1.15%	1.39%	3.23%	0.49%
UK	0.64%	2.20%	3.14%	0.79%
USA	1.70%	1.61%	3.29%	0.62%

Table 5.20 demonstrates that almost all countries (GER is the only exception) have been increasing the quantities of labour inputs used in the production of aggregate output, although the rate of increase is relatively modest. The relative volatility of labour input growth, measured as the ratio of standard deviation to average, is approximately 2.1 on average, while labour growth volatility in absolute terms, measured only by examining the standard deviation of the growth measure, is relatively small, averaging in approximately 1.7%.

Most countries have also been increasing their capital stock over the period of the analysis, with an average growth in capital inputs of 3.3%. Both relative and absolute volatility in capital input growth is quite low (compared with labour inputs), averaging at 0.3 and 0.9% respectively.

5.3.4.2 Assessing the extent of technical inefficiency

In order to provide an indication of how widespread technical inefficiency is in the countries in EU KLEMS dataset, this analysis examines the various estimates for the different approaches (and models) adopted; these are summarised in the following table.

Table 5.21: Average technical efficiency estimates, by approach

Approach	Number of observations	Average	Standard deviation	Minimum	Maximum
DEA meta-frontier CRS [1]	375	73.1%	13.8%	41.9%	100.0%
DEA meta-frontier VRS (output oriented) [1]	375	79.7%	14.6%	49.5%	100.0%
DEA CRS	375	83.7%	13.2%	52.7%	100.0%
DEA VRS (output oriented)	375	89.6%	13.6%	52.7%	100.0%
COLS (Cobb-Douglas)	375	72.5%	11.9%	47.3%	100.0%
COLS (translog)	375	72.3%	10.8%	46.5%	100.0%
SFA (Cobb-Douglas, half-normal)	375	82.9%	10.8%	54.7%	96.8%
SFA (Cobb-Douglas, exponential)	375	86.6%	10.5%	56.1%	97.3%
SFA (translog, half-normal)	375	81.9%	12.1%	48.5%	100.0%
SFA (translog, exponential)	375	88.0%	10.2%	53.5%	97.4%

Note: [1] DEA meta-frontier efficiency estimates do not take into account the time dimension (technological change and scale efficiency change) and as such are likely to be biased (downward if we assume positive technological change). They are presented here for completeness.

Direct tests for the existence of technical inefficiency are only possible for the SFA models; with regards to this application, these tests resulted in the rejection of the null hypothesis of no technical inefficiency in all four SFA specifications examined.

Table 5.21 reveals a relative small spread of average efficiency in all the approaches examined. The two COLS specifications display the smallest average efficiency (approximately 72%), while the DEA output oriented VRS models display the largest average efficiency scores (approximately 90%). Average efficiency across all models is estimated at approximately 81% or 82% if the DEA meta-frontier efficiency scores are excluded (see note to table 5.21).

5.3.4.3 Assessing the extent of noise in the data

The relevant estimates of σ_v, the standard deviation of the noise component, from all the SFA models adopted for this application are provided in the table below.

Table 5.22: Summary statistics of the σ_v estimate from the SFA models

SFA model	Estimate of σ_v	Standard deviation of the σ_v estimate	Minimum	Maximum
Cobb-Douglas, half-normal	0.075	0.010	0.058	0.098
Cobb-Douglas, exponential	0.086	0.007	0.073	0.101
Translog, half-normal	0.000	0.000	0.000	0.000
Translog, exponential	0.074	0.006	0.063	0.087

The two Cobb-Douglas models and the translog model that assumes technical inefficiency is exponentially distributed find that the standard deviation of the normally-distributed error term is between 0.05 to 0.1. On the other hand, the translog SFA model that assumes half-normally distributed technical inefficiency finds that the amount of noise in the current dataset is negligible (σ_v is approximately equal to zero). This last finding appears quite improbable; while it is true that EU KLEMS collated the various country data in such a way as to ensure the greatest possible compatibility between the different countries, the underlying data are still based on National Accounts information. Since the process of data collation and aggregation required to draw-up the National Accounts rests on a number of assumptions and imputations[87], it is expected that the data would almost always incorporate some degree of inaccuracy[88]. As such, it is unlikely that the EU KLEMS dataset is completely free of measurement error and/or statistical noise.

Since the estimate of σ_v is inconsistent in the pooled setting, in order to provide some clarity on whether the use of the σ_v estimate is valid in this instance, it would be helpful to observe the behaviour of the estimate under controlled conditions. To that purpose, a new round of simulations is undertaken for this chapter.

This analysis utilises the same simulation framework adopted in chapter 4[89]. To enhance readability, only a single DGP is considered; however, to ensure the relevance of the results for this particular application, the DGP is constructed in such

[87] See for example the requirement to incorporate imputed rents for owners/occupiers and the methodology used to estimate VA from privately held corporations and unincorporated enterprises (see Office of National Statistics, (2008), op cit).
[88] This is also evident from the number of times that National Account information is updated, sometimes quite a few years after the original estimates were first published.
[89] As a reminder, the simulation framework in question uses 100 observations (20 DMU observed over a 5 periods) and summarises the findings of 100 experiments.

a way that it displays similar characteristics as those observed in the EU KLEMS dataset. In more detail, the utilised DGP:

– is a piece-wise linear production function, since the analysis in section 3.4.4 below suggests that the underlying production function in the current dataset is neither Cobb-Douglas nor translog[90];

– utilises input and price data that were constructed so that they are consistent with the level of input volatility observed in the EU KLEMS dataset (section 3.4.1). In summary, input quantities and price are randomly generated for the first period and then scaled by a random factor that follows N~(0.0.1);

– includes a technical inefficiency component, $u_{it} \sim Exp(1/5.5)$, which results in average technical efficiency levels in the simulations of appr. 81%. This is consistent with the estimates of technical inefficiency observed in the EU KLEMS dataset, as detailed in section 3.4.2 of this chapter;

– and lastly, includes a noise component that is randomly generated following N~(0, 0.05), consistent with the estimates presented in table 5.22;

The summary findings of the simulation analysis are given below:

Table 5.23: Summary statistics of the σ_v estimate from the simulation analysis

	SFA translog (exponential)	SFA Cobb-Douglas (exponential)
Average of σ_v across all simulations	0.054	0.108
Standard deviation of σ_v across all simulations	0.040	0.054
Instances of zero σ_v	21	0
MAD scores (for reference)	0.061	0.073
MSE scores (for reference)	6.49	9.86

The results show that the translog SFA model, which is the most accurate of the SFA models under these conditions with regards to productivity change estimates, displays an average estimate of σ_v that is very close to its true value. However, the standard deviation of this average measure is quite large; the 95% upper confidence interval is approximately 0.135, which is more than twice as large as the true value. The simulation analysis also finds that out of the 100 simulation experiments, in 21 of

[90] The piece-wise linear production function employed here is monotonic and concave; it is described fully in chapter 4.

those the translog SFA models displayed an estimated σ_v that was approximately equal to zero. This suggests that sometimes even the more accurate SFA model is not able to detect the presence of noise, even though modest levels of noise are part of the DGP. For the Cobb-Douglass SFA model, there were no instances where σ_v approached zero, but the σ_v estimate was also twice as large on average as the true standard deviation of the noise component.

Overall, the results from the simulations demonstrate that in conditions that approximate those found in the current analysis, the estimate of σ_v can provide an overall indication of the extend of measurement error/noise in the data, with the caveat that high levels of precision should not be expected.

5.3.4.4 Are the parametric models misspecified?

Table 5.24 below provides the results of the RESET test and the p-values of the coefficients from the parametric models.

Table 5.24: Statistical significance of the variables in the parametric models and RESET test results from the application

	COLS (Cobb-Douglas)	COLS (translog)	SFA (Cobb-Douglas, half-normal)	SFA (Cobb-Douglas, exponential)	SFA (translog, half-normal)	SFA (translog, exponential)
L	0.00	0.00	0.00	0.00	0.00	0.00
K	0.00	0.00	0.00	0.00	0.00	0.00
t	0.00	**0.17***	0.00	0.00	**0.65***	**0.71***
L^2		0.00			0.00	0.00
K^2		0.00			0.00	0.00
t^2		**0.30***			**0.32***	**0.33***
LK		0.00			0.00	0.00
Kt		0.01			0.00	0.00
Lt		0.04			0.00	0.00
Insignificant variables	0	2	0	0	2	2
RESET	0.00	0.00				

Note: The values corresponding to the model variables represent the p-values of the t-tests for statistical significance. The values corresponding to the row labelled RESET represent the p-values of the F-test for statistical significance.

The analysis found that both the Cobb-Douglas and the translog models failed the RESET test; in addition, all translog models found that the time variable and its square displayed coefficients that were statistically insignificant. Both of these factors suggest that the parametric models could suffer from some form of misspecification.

The next step is to test whether parametric models that are known to be misspecified also display similar symptoms; this is again achieved by a new round of simulation experiments that use the same assumptions as those in section 3.4.3. The following table provides a summary of the instances of statistically insignificant variables and failed RESET tests from the simulations.

Table 5.25: Summary of statistical significance of the variables in the parametric models of the simulation analysis

	COLS (Cobb-Douglas)	COLS (translog)	SFA (Cobb-Douglas, exponential)	SFA (translog, exponential)
L	0	3	0	0
K	0	1	0	1
t	79	96	59	68
L^2		40		20
K^2		27		12
t^2		95		71
LK		17		8
Kt		94		68
Lt		91		65
Average number of insignificant variables	0.79	4.64	0.59	3.13
Cases where all variables were significant	21	0	41	22
Cases where RESET failed	40	51	N/A	N/A

Note: The values corresponding to the model variables represent the number of instances where the variable in question was found to be statistically insignificant (note that the simulation were run 100 times).

The simulation analysis shows that the RESET test found evidence of miss-specification in almost half of the simulation experiments. In addition, there were instances of insignificant variables in the majority of the experiments undertaken; the translog COLS specification had no cases where all variables were significant, while the Cobb-Douglas SFA model that (correctly) assumed exponentially-distributed inefficiency was the better performing model in this measure, with just 41 cases where all variables were statistically significant.

Overall, these results suggest that when the parametric models suffer from functional form misspecification, it is quite common to observe statistical insignificant variables and failures in the RESET test. Given that similar symptoms were observed in the current application, one could conclude that the parametric models in this application are likely to suffer from some form of misspecification, which would negatively impact the accuracy of their productivity change estimates.

5.3.4.5 Selecting the most appropriate estimation approach

With regards to the characteristics of the current dataset, this analysis found that:

- input volatility is quite low, averaging just 1.7% p.a. for the labour input and 0.9% p.a. for the capital input (section 4.3.1);
- average technical inefficiency across all approaches in this application is approximately 82% (section 4.3.2);
- the SFA models suggest that the standard deviation of the normally-distributed noise component (σ_v) probably takes a value between 0.05 and 0.1 (section 4.3.3);
- the parametric models are likely to suffer from some form of misspecification, which could be due to the adopted functional form not being an appropriate representation of the underlying DGP section 4.3.4);

According to the above findings, the simulation experiment from chapter 4 that more closely matches the characteristics of the current dataset is S2.3 with 'default' input volatility. In more detail, for the S2.3 simulation experiment:

- the underlying DGP is piecewise-linear, since the current analysis found that neither the Cobb-Douglas nor the more flexible translog functional forms provide a close approximation to the underlying DGP.
- inputs are scaled from one year to the next by a random factors that follows N~(0,0.1), which results in input volatility similar the EU KLEMS dataset.
- average technical efficiency in the simulations is designed to be approximately 87% on average - the current analysis found that average technical efficiency across all approaches in the EU KLEMS dataset is 82%.
- includes a noise component in the DGP, which is randomly generated and follows N~(0,0.05). The decision to adopt this level of noise could be considered conservative, since the mid-point between the various chosen estimates of σ_v is closer to 0.075.

The summary accuracy measures of the above experiment are replicated in the table below:

Table 5.26: Summary accuracy results

	Measure	GA	COLS	COLS (translog)	DEA	SFA	SFA (translog)	SFA (half-normal)
Accuracy scores	MAD (%)	5.80	6.30	6.50	5.80	7.10	6.10	6.40
	MSE	5.33	6.24	7.81	5.23	9.11	6.12	6.96
Accuracy rankings	MAD	1	5	5	1	7	3	5
	MSE	2	4	6	1	7	3	4

Note: MAD = Mean Absolute deviation, MSE = Mean Square Error. Smaller values represent higher accuracy, both for the scores and the rankings, ie an approach with accuracy ranking 1 is more accurate than and an approach ranked as 2.

As table 5.26 demonstrates, the two most accurate approaches in this simulation experiment were DEA and GA, closely followed by the translog SFA model. The DEA and GA accuracy scores are almost identical; it should be noted however that the simulation analysis is designed such that the relevant input and output prices indices required by GA are measured with no error, while also explicitly assuming that there is no element of allocative inefficiency in the analysis. The reason for designing the experiment in such a way was that it allowed for a level playing field when comparing the GA with the frontier-based estimates, which do not rely on price information. In a real life application such as the current analysis however, some amount of measurement error is expected to be present in the price data. In addition, the GA estimates would also be influenced by changes in allocative inefficiency in the countries examined. Given that the impact of those factors to the relative accuracy of the GA estimates under the current conditions is unknown, it would be more prudent to rely mostly on the DEA-based productivity estimates.

5.4. Summary and conclusions

The aim of this chapter was two-fold. First, to provide a comparison of the different measurement approaches in a real-life aggregate productivity change measurement application and second, to devise a selection framework to help policy makers choose the productivity measurement approach that is likely to produce the most accurate estimates relative to the application in hand.

The real-life application was based on the EU KLEMS dataset, which collated information on aggregate inputs and outputs for a number of economies over a relatively long timeframe (1970-2007, although for some countries information is not available for the full timeframe). The measurement approaches considered were those detailed in chapter 3, namely Growth Accounting, Circular DEA-based

Malmquist indices and COLS- and SFA-based Malmquist indices. The analysis found that although at first glance all assessed approaches produce very similar productivity change estimates on average, the productivity estimates from the various approaches can be quite dissimilar at the individual country level.

To examine this issue further, the analysis provided a more in-depth look at the productivity performance of individual countries. This revealed that for most cases the productivity estimates from the various approaches are quite highly correlated at the individual country level; only in few cases the estimates from the SFA models display relatively lower correlation coefficients to their deterministic counterparts. Where there is disagreement it is in the levels of the estimates themselves, which can be quite pronounced in some cases (such as the estimated productivity performance of the UK and Slovenia). The analysis also revealed that the SFA estimates are generally much less variable relative to the estimates from the deterministic approaches. This is probably due to the fact that some of the variability of the ratio of outputs to inputs that is detected as productivity change by the deterministic approaches is assigned to the stochastic element under SFA. This reasoning is supported by the fact that the variability of the productivity estimates from the deterministic approaches is considerably smaller in the latter period of the analysis (from 1995 onward), when the economic performance of the assessed countries was more stable.

Related to the above, the analysis also found that productivity performance appears to be heavily influenced by economic cycles. In periods of economic recession, productivity is declining, only to pick up again when the economy starts to grow again. In other words, the estimates of productivity growth appear to be pro-cyclical, which is consistent with findings from other studies that examined this issue in more detail[91].

In addition to the above, this chapter also provided a brief discussion on the decomposition of the estimated productivity change for the various frontier-based approaches (the GA productivity estimates were not considered here, since they cannot be decomposed). The productivity components used in this analysis are efficiency change, technological change and scale efficiency change and the analysis focused its attention mainly on two approaches, namely DEA and SFA. When looking at the summary statistics of the component, there appears to be a consensus by the various approaches; they all find that efficiency change had a small, negative effect

[91] See for example Boisso et al. (2000).

on productivity growth, scale efficiency change had almost no impact while productivity growth came almost exclusively through technological change. However, when examining the results at the level of individual countries, it is clear that there are a lot of differences in the estimates of the components derived from DEA and SFA models, both in their levels but also in their patterns of change (ie low correlations between the two approaches). This is a significant finding, since the overall productivity change estimates from these two approaches are quite similar in levels and are also quite highly correlated. A closer examination of the results found no easily detected patterns that could be helpful in explaining the observed differences; as such, a more detailed examination of this issue is left for future research.

Overall, the analysis of the EU KLEMS data found that the different approaches can often lead to different views on productivity performance. These differences are problematic from a policy perspective, since policy decisions on the issue of economic growth rely on having accurate productivity estimates at the national level. As such, there is a need for a process or a mechanism that can be used to select between the various approaches, in the event of such disagreements. To that end, this chapter proposes the use of a selection framework, designed to detect the set of productivity estimates that are likely to be more accurate for the conditions/characteristics prevalent in the dataset/application at hand.

This selection framework includes three steps:

– First, determine those conditions/characteristics inherent in the DGP that can have a significant influence in the relative accuracy of the assessed productivity measurement approaches.

– Secondly, examine the current dataset and try to quantify said conditions/characteristics.

– Finally, examine the relative accuracy of the adopted approaches in datasets specifically designed to display those characteristics/factors found the real-life data of the current application.

With regards to the fist step, the analysis relied on the findings of the simulation analysis undertaken in chapter 4, which identified that the characteristics of the DGP that are most influential the overall accuracy of the most common productivity measurement approaches include: input volatility, technical inefficiency, noise and whether the parametric approaches are likely to suffer from functional form misspecification. The above list is not necessarily exhaustive and there may well be additional characteristics that have a significant impact on the overall accuracy of productivity change estimates, such as latent heterogeneity amongst the assessed units and variable returns to scale. However, since these are quite complex issues, the assessment of these and other potentially significant characteristics is left for future research.

At the second step, the goal is to detect the presence of the characteristics in the application at hand and to quantify their effects. To do so, the use of a number of well-established diagnostics and indicators is suggested so that the proposed selection framework can be easily implementable; these diagnostics included tests such as RESET for assessing functional form misspecification and the utilisation of estimates of technical efficiency derived from the assessed approaches. To assess whether the proposed diagnostics can lead to reliable results, a new round of simulations was undertaken, based on the characteristics of the EU KLEMS dataset. The analysis found that the proposed diagnostics and indicators can indeed provide relatively reliable estimates of the characteristics in question; however, care is advised in not to take these findings as absolute, since some of the characteristics are, by their very nature, very difficult to quantify (for example, detecting and quantifying the extend of noise in the data is particularly difficult). Hopefully, more focused diagnostics/indicators can be developed and refined in the future.

The third and final step of the selection framework is to determine which of the adopted approaches is more accurate overall, under the conditions prevalent in the application in hand. For the application examined here, this was achieved by relying on the findings of the simulation analysis of chapter 4. However, if the application at hand is quite dissimilar to the various DGP adopted in past simulation studies, it would be more appropriate for the analysis to construct a new DGP that more closely matches with the current conditions and use that as the basis of a new round of simulations; the results of this analysis would provide a better indicator on the suitability of the assessed approaches.

Applying the above framework in the EU KLEMS dataset revealed that input volatility is low, technical efficiency is approximately 82%, noise levels are also relatively low and the parametric models are likely to suffer from some form of misspecification. Under such conditions, the approaches that are most accurate, according to the findings of the simulation analysis of chapter 4, are DEA-based Malmquist indices and Growth Accounting, closely followed by the Malmquist indices derived form the translog, exponential SFA model. If the analysis was asked to recommend one of the above approaches, it would be DEA-based Malmquist indices based on its performance in the simulation experiments and also due to the fact that it is not as sensitive as GA to other potentials issues that were not included in the simulations, such as measurement error in the price of inputs and the presence of allocative inefficiency.

Chapter 6. Summary and conclusions

6.1. Productivity growth in the macro setting: why is it important and how to measure it

Productivity is a complex concept but also arguably the most appropriate measure of changes in economic welfare. Despite its importance, changes in productivity are usually not one of the main topics in macroeconomic debates. Indeed, most mainstream economic publications on the macro setting focus more on the actual growth of output (usually expressed as growth in GDP), considering this as the main indicator of economic welfare. While it is true that output is an important indicator, this measure on its own cannot provide a detailed description of economic welfare, since it does not take into consideration the necessary ‘effort’ required to produce said output. Even when the issue of productivity is directly discussed, oftentimes the measure of productivity adopted is labour productivity, which is defined as a ratio of output to labour input. While labour productivity is indeed a valid measure of economic welfare, since it accounts for both the output and the ‘effort’ of producing such output, it provides only a partial view of productivity performance because it only accounts for one input of the production process and ignores everything else; it is a single factor productivity measure and is therefore limited by construction. In order to better understand the productivity phenomenon, we require a measure that takes into account all, or at least the most important of, inputs; these measures are usually referred to as total-factor or multi-factor measures of productivity (TFP and MFP respectively). Overreliance on simple output growth rates or partial productivity measures can lead to a distorted view of productivity and more importantly, this distorted view can lead to adoption of misguided economic policies as discussed by Krugman (1994).

So, to discuss a topic as complex as productivity, it is critical to have access to measures that can capture this complexity; in other words, economists need robust ways to measure productivity in a multi-input, multi-output setting that reflects economic reality. Arguably, the most widespread method for measuring productivity growth in the macro setting, ie when considering the productivity performance of entire industries or countries, is Growth Accounting (GA). GA is an index number-based approach that relies on the neo-classical production framework, and seeks to estimate the rate of productivity change residually, ie by examining how much of an observed rate of change of a unit's output can be explained by the rate of change of the combined inputs used in the production process. It is the productivity measurement method of choice for most interested agents, namely statistical agencies (national and international), central banks and government bodies. The main strength of the approach comes from its relative accessibility; the required data can be collated from National Accounts (a framework that measures and presents all factors of economic activity in the macro level) and productivity estimates can be easily calculated using simple algebra. This ease of implementation sometimes masks the complexity and depth of the approach. GA is based on robust (if simplistic) economic principles and mathematical logic, which provide a strong theoretical underpinning to justify its use.

Nevertheless, the approach is not without its weaknesses; although the majority of the required primary data can be sourced from National Accounts, there are a number of additional adjustments that need to be applied to ensure inter-temporal comparability and consistency. In short, care must be taken to net-off all intra-industry transactions, to account for non-market output (the discussion on how to do so is still in progress) and to account for changes in quality. Creating a measure of capital services is particularly problematic, since the process is quite laborious and requires data not available through the National Accounts and a number of additional imputations (these issues were discussed in more detail in chapter 2). Even so, data considerations are not an issue necessarily restricted to GA; all other approaches require similar data for the estimation of productivity, although the data collation process for some factors of production (for the capital services input in particular) is less convoluted. Arguably, the biggest drawback of GA is that it requires the adoption of the so-called standard neo-classical assumptions. These assumptions require that the production process is fully deterministic, exhibits only constant returns to scale, all information relating to it is measured with perfect accuracy and that all producers are fully efficient. The above assumptions are very restrictive and it is not difficult to argue that they do not provide a fair representation of most production processes (GA was discussed in some detail in chapter 3).

The issues with the available data and especially the requirement for the adoption of the neo-classical assumptions can cast doubts on the overall validity of the resulting productivity estimates. After all, if the approach relies on assumptions that do not reflect reality, how can one trust the resulting estimates? So the question is, are there alternative approaches that are likely to be more accurate than GA, and if so, by how much? This question was the main motivation of this thesis.

Frontier-based approaches for productivity measurement have been extensively used in the micro setting and offer an interesting alternative to GA. Although the parametric and non-parametric frontier-based approaches differ substantially in how they estimate productivity change, they both rely on the same underlying principles; that there exists a production technology defined as the set of feasible outputs that can be produced by a combination of inputs and that all production units operate within this technology set. The technology frontier is convex (or conical), monotonic hull that envelops the technology; the assumptions of monotonicity and convexity are relatively benign and are used to ensure that the production frontier conforms to common production processes that result in no undesirable goods and obey the law of diminishing returns. As such, it is not hard to argue that frontier-based approaches require less restrictive assumptions compared to GA. In addition, frontier-based approaches do not require any information on input prices, while GA cannot be used without such information. This is important because, although some price information is available in the National Accounts, the price of capital services is unobservable and very difficult to estimate in practice without adopting the neo-classical assumptions. Another major advantage of the frontier-based approaches is that they allow the decomposition of the productivity estimates into discreet components, such as efficiency change, scale efficiency change and technological change; this additional granularity of the estimates is of great interest to the users of such analysis, since it can improve the development, targeting and assessment of productivity improvement policies (frontier-based approaches were discussed in chapter 3).

The main drawback of the frontier-based approaches is that all of them require information on a suitable set of comparators (eg the economies of a number of countries) in order to estimate the production frontier. Furthermore, the analysis needs to ensure that the inputs and outputs of each individual unit of assessment (eg the economy of a country) are collated and expressed in a manner that ensures the comparability between the various assessed units. However, there are currently a number of databases that collate National Account data from a large group of countries and they also include the information necessary to make the data comparable; as such, data limitations are not a significant hurdle for the adoption of these approaches.

The above discussion demonstrates that frontier-based approaches are likely to produce richer, more accurate estimates of productivity than GA, mainly due to the fact that they rely on fewer assumptions and can thus model a wider range of production processes, at least according to theory. The question now becomes, do these theoretical advantages translate to practical improvements in the accuracy of the productivity estimates and if so, by how much?

6.2. Exploring and quantifying accuracy: Simulation analysis

To assess or quantify the differences in accuracy between approaches, one needs to rely on some form of controlled experiment, since productivity is not a quantity that can be directly observed in real-world applications. This thesis utilised Monte Carlo simulations to examine the behaviour of GA and some of the most common frontier-based approaches under a number of conditions that violate the standard neoclassical assumptions. The simulation analysis and its findings are detailed in chapter 4 and have also been published in the European Journal of Operational Research (see Giraleas et al. (2012)).

In more detail, the simulation analysis examined the accuracy of GA and three of the most common frontier-based approaches for the measurement of productivity change namely:

- DEA-based circular Malmquist indices: Data Envelopment Analysis (DEA) is the most common non-parametric approach for efficiency and productivity measurement. The strengths of the approach are that it requires minimal assumptions for the estimation of the frontier and no a-priori specification of its functional form. Its main weakness is that it is deterministic in nature, in that it does not explicitly take into account the stochastic nature of the production process. This thesis utilises the circular Malmquist index, as opposed to the 'traditional' (Färe et al. (1994)) Malmquist index, mainly for the ease of computation and the ability to accommodate unbalanced panel data.
- COLS-based Malmquist indices: Corrected OLS is a deterministic, parametric approach and one of the numerous ways that have been suggested to 'correct' the inconsistency of the OLS-derived constant term of the regression when technical inefficiency is present in the production process. COLS, similar to DEA,

is deterministic in nature and also requires the prior specification of the functional form of the estimated frontier; it was examined here since it is the easiest approach to implement and as such likely to be used by non-specialists.

– SFA-based Malmquist indices: Stochastic Frontier analysis (SFA) is also a parametric approach similar to COLS, but it is stochastic in nature, in the sense that it explicitly attempts to account for the stochastic nature of the production process. This comes at the cost of increased complexity and in the fact that it requires both the prior specification of the functional form of the estimated frontier and also the specification of the inefficiency distribution.

The starting point of any simulation analysis is the data generating process (DGP), which sets out the framework used to generate the parameters of value. Given that the aim of the simulations was to examine the accuracy of the different approaches when the various neoclassical assumptions were violated, the DGP was set up in such a way as to include both elements of technical inefficiency and stochastic noise. The analysis also examined whether increased volatility in inputs has any effect in accuracy and also how impactful would be if the a-priori assumptions about the functional form and the distribution of inefficiency for the parametric approaches were wrong. The latter was achieved by running two sets of simulations; the first set adopted a smooth, Cobb-Douglas production function, while the second set utilised a piecewise-linear production function.

Piecewise-linear production functions have been utilised in previous simulation studies, but they were usually restricted to modelling a production process with just single input and a single output. One of the contributions of this thesis was the creation of a framework that allows the generation of a random, convex and monotonic piecewise-linear production function with a single output and two inputs, which is sufficient for the purposes of measuring aggregate productivity change.

The second contribution of this thesis with regards to the DGP was to provide a methodology that can be used to generate price information consistent with the characteristics of the production function (the availability of price information is necessary for GA in order to parameterise the production function). This methodology is based on the duality between the production function and the cost function, which links the parameters of the production function to the costs faced by the producers; by manipulating this relationship, the analysis was able to derive consistent formulas that link the parameters of the production function with the prices of the different inputs. This derivation is described in more detail in section 4.3.1.

The resulting DGP enabled the creation of a number of experiments to test for the effects of the violation of the neoclassical assumptions. The results of the analysis were quite interesting and sometimes unexpected. In summary the analysis found that:

– The deterministic approaches perform adequately even under conditions of (modest) measurement error. This also includes GA, although the frontier-based deterministic approaches were more accurate in the majority of the experiments.

– Functional form misspecification has a severe negative impact on the accuracy of all parametric approaches. It was anticipated prior to the analysis that functional form misspecification is likely to lead to a general loss of accuracy; what was unexpected was the magnitude of the effect.

– When measurement error becomes larger, the accuracy of all approaches (including SFA) deteriorates rapidly, to the point that their estimates could be considered unreliable for policy purposes. Again, some loss in accuracy was expected, but probably not to the point observed in the experiments.

– The SFA models that adopt a translog specification appear to be more accurate in general than the Cobb-Douglas SFA models when the underlying (true) production function is piecewise linear. On the other hand, it is the Cobb-Douglas COLS models that are assessed as more accurate relative to their translog counterparts when the underlying production function is piecewise linear. Prior to the analysis, the hypothesis was that the adoption of a flexible functional form would reduce the impact of functional form misspecification for all parametric approaches; the experiments revealed that this was not the case, at least for the COLS models.

– Misspecification of the inefficiency distribution in the SFA models does not appear to have a significant effect on the overall accuracy of said approach. This confirms previous findings in the literature, when the focus of the analysis was in assessing the accuracy of efficiency estimates. It should be mentioned however that the inefficiency distribution adopted by the DGP (exponential distribution) is similar to the alternative distribution adopted in the simulation analysis (half-normal distribution).

– Increased volatility in inputs and prices from one period to the next adversely affects the accuracy of all approaches, in almost all experiments. The DEA estimates are the least affected, while the GA estimates are the most affected. It was interesting to observe how input volatility affects the accuracy of the estimates, especially when noise was introduced in the DGP.

Overall, the simulation analysis demonstrated that no productivity change measurement approach has an absolute advantage over another, but rather under some specific circumstances, a specific approach is likely to be more accurate than its counterparts. The analysis also demonstrates that frontier-based approaches can usually produce at least as accurate, and in most cases more accurate, productivity change estimates than the more traditional GA approach.

It should be noted here that the findings of the simulation analysis of chapter 4 are also applicable to the micro setting (ie when assessing the productivity performance of single production units), especially with regards to the accuracy of the frontier-based approaches, since they do not rely on assumptions and/or conditions that are restricted to the macro setting.

6.3. Practical application: Frontier-based approaches using EU KLEMS

The simulation experiments presented compelling evidence on the suitability of frontier-based approaches for the measurement of productivity change, but also found that there is no single approach that is likely to be more accurate under a range of different conditions. Given these findings, it would be interesting to compare and contrast the productivity estimates from the different approaches in a practical application. Additionally, the recent EU KLEMS project, which provides access to harmonised National Accounts information for a large number of, mostly, EU countries, provides a great opportunity to re-examine this issue using a relatively up-to-date dataset.

The EU KLEMS project collated information on aggregate inputs and outputs for a number of economies over a relatively long timeframe (covering the 1970-2007 period), for the express purpose of productivity measurement under GA. The productivity measurement application presented in chapter 5 utilised the same information to also estimate (and decompose) productivity change using a number of the most common frontier-based approaches.

The results were mixed; although at first glance, all assessed approaches produce very similar productivity change estimates on average, a more detailed analysis of the findings revealed that the productivity estimates from the various approaches can be quite dissimilar at the individual country level. In the majority of the countries examined, the different productivity estimates were highly (positively) correlated with each other. However, the levels of the estimates could be significantly different, a fact that was particular evident for the UK and Slovenia, but present in all assessed countries; the differences between the maximum and minimum average productivity change estimate was never less than 0.5 percentage points, which is significant when the range of estimated average productivity change across all countries from all approaches was between approximately 0.5% and 0.9%. The results of the analysis also revealed that:

- The SFA estimates are generally much less variable relative to the estimates from the deterministic approaches. The most likely reason for this reduced variability is that that some of the changes in the ratio of outputs to inputs that is detected as productivity change by the deterministic approaches are interpreted as noise by SFA.

- Productivity performance appears to be heavily influenced by economic cycles; in fact, the estimates of productivity growth appear to be pro-cyclical, which is consistent with findings of other studies.

- When productivity is decomposed, all frontier-based approaches show that the largest contributor to productivity growth has been technological change. However, at country level there are significant differences in both the levels of the estimated components and in their patterns (demonstrated by low correlation coefficients) between the assessed approaches; this is a significant finding because the differences in the overall productivity change estimates are much less pronounced.

6.4. Selecting between conflicting estimates

The observed differences in productivity change estimates from the different approaches can be problematic from a policy perspective, especially for those countries where they provide conflicting views of the aggregate productivity performance. Disagreements in the results from different approaches are, unfortunately, not uncommon in the field of efficiency and productivity analysis. When they arise, the usefulness of the analysis diminishes, since it cannot provide a clear answer to the main issue; in some cases, conflicting estimates could cause the validity of the whole analysis to be cast into doubt. As such, it is critical to be able to explain any such differences and more importantly sort through conflicting estimates to select those that are likely to be more accurate.

The simulation analysis undertaken in chapter 4 revealed that there is no single approach that is likely to be accurate under all circumstances; however, when those circumstances are known, the results of the analysis can be utilised to make an informed decision with regards to the likely accuracy of the different estimates. This argument forms the basis of the selection framework proposed in chapter 5 of this thesis. In short, the selection framework involves three steps:

– The first step is to identify those conditions/characteristics inherent in the DGP that can have a significant influence in the relative accuracy of the assessed productivity measurement approaches. This can be achieved by examining the findings of previous simulations studies; chapter 4 identified four conditions/characteristics that can have a significant effect, which are input volatility, technical inefficiency, noise and whether the parametric approaches are likely to suffer from functional form misspecification.

– The second step is to assess whether these conditions/characteristics are present in the application at hand and if so, at what levels. The proposed selection framework suggests the use of well-established diagnostics and indicators to that purpose, such as RESET for assessing functional form misspecification and the utilisation of estimates of technical efficiency derived from the assessed approaches. Since some of the conditions/characteristics are difficult to measure (or are unobservable), the efficacy of the proposed diagnostics and indicators was tested using additional simulation analysis, which confirmed that these can indeed provide relatively reliable estimates.

– The third and final step of the selection framework is to determine which of the adopted approaches is more accurate overall, under the conditions prevalent in the application in hand. Since estimates of the conditions/characteristics of the particular application are now available, the researcher can utilise the findings of previous simulation analyses to identify the approaches that are likely to provide the most accurate estimates. If the simulation analyses in the literature do not provide a good fit for the application at hand (due to the DGP being significantly different relative to model used in the current assessment), a new round of simulations could be undertaken using a DGP specifically designed to emulate the characteristics of the application.

Applying the above framework in the EU KLEMS dataset revealed that the two approaches that are likely to be most accurate based on the observed input volatility and estimated technical inefficiency and noise are the DEA-based circular Malmquist indices and GA.

6.5. Concluding remarks and further research

The first aim of this thesis was to examine and assess the different approaches that policy makers have at their disposal for measuring aggregate productivity change. This was achieved by discussing the strengths and weaknesses of the different approaches from a theoretical perspective, but more importantly examining the accuracy of the resulting estimates under different conditions. The simulation analysis undertaken for this reason was relatively extensive but can be expanded in many directions. The parameters of the DGP could be extended in such a way as to offer a greater granularity of results, for example by varying the number of observations available (both cross-sectional and across time) or adopting a wider range of average technical efficiency and noise parameters. The DGP could also be expanded to include additional characteristics that could have a significant impact on the overall accuracy of productivity change estimates, such as latent heterogeneity amongst the assessed units, variable returns to scale and allocative inefficiency. The earlier two characteristics were not assessed here due to time and space limitations, while allocative inefficiency was purposely excluded to improve the comparability of the frontier-based estimates relative to GA. Nevertheless, it would be interesting to examine how this latter factor in particular affects the accuracy of the estimates.

The scope of the analysis could also be broadened by searching for ways to improve the quality and accuracy of the various frontier-based estimates by incorporating price information directly into the analysis. Incorporating information on prices is indeed possible in the framework of frontier-based approaches by utilising duality theory, but research relevant to how this would impact the accuracy of the resulting estimates is limited.

The second main aim of the thesis was to provide an up-to-date application of aggregate productivity measurement using both GA and frontier-based approaches and provide some guidance to applied researchers when asked to choose between sometimes conflicting estimates. One of the findings of the application was that the different approaches provide significantly different views on the components of productivity change, even in cases where the primary productivity change estimates are quite similar. It would be interesting to see if this is an issue in general and if so, what are the causes of such differences.

With regards to the proposed selection framework, there are a number of potential enhancements that could be implemented. The framework could examine additional conditions/characteristics of interest, based on new findings from simulation analysis and the efficacy of the proposed diagnostics and indicators could be tested over an expanded range of conditions. More importantly, additional diagnostics and indicators could be adopted or developed to provide a more comprehensive assessment of the characteristics of interest.

In conclusion, the issue of productivity measurement is quite complex and despite the extensive research undertaken so far in the area, there are still gaps in our knowledge and thus a lot of opportunities for interesting and useful research for the future.

References

Abramovitz, M. (1956). Resource and Output Trends in the United States Since 1870. The American Economic Review, 46(2), 5-23.

Afonso, A., Schuknecht, L. & Tanzi, V. (2005). Public sector efficiency: An international comparison, Public Choice, 123: 321-47.

Aigner, D. J., & Chu, S. F. (1968). On estimating the industry production function. The American Economic Review, 58(4), 826-839.

Aigner, D., Lovell, C. A. L., & Schmidt, P. (1977). Formulation and estimation of stochastic frontier production function models. Journal of Econometrics, 6(1), 21-37.

Balk, B.M. (2008). Measuring productivity change without neoclassical assumptions: a conceptual analysis. Price And Productivity Measurement: Volume 6--Index Number Theory, 7, 133-154.

Balk, B. M. (2005). Divisia price and quantity indices: 80 years after. Statistica Neerlandica, 59(2), 119-158.

Balk, B. M. (2009). On the relation between gross output–and value added–based productivity measures: The importance of the Domar factor. Macroeconomic Dynamics, 13(S2), 241-267.

Balk, B. M., & Institut national de la statistique et des études économique (Paris). (1998). Industrial price, quantity, and productivity indices: the micro-economic theory and an application. Dordrecht: Kluwer Academic.

Banker, R. D. (1993). Maximum likelihood, consistency and data envelopment analysis: a statistical foundation, Management Science 39, 1265–1273.

Banker, R. D., Chang, H., & Cooper, W. W. (2004). A simulation study of DEA and parametric frontier models in the presence of heteroscedasticity. European Journal of Operational Research, 153(3), 624-640.

Banker, R. D., Charnes, A., & Cooper, W. W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. Management Science, 30(9), 1078-1092.

Banker, R. D., Charnes, A., Cooper, W. W., & Maindiratta, A. (1987). A comparison of DEA and translog estimates of production frontiers using simulated observations from a known technology. In Applications of modern production theory: Efficiency and Productivity (pp. 33-55). Springer Netherlands.

Banker, R. D., Gadh, V. M., & Gorr, W. L. (1993). A Monte Carlo comparison of two production frontier estimation methods: corrected ordinary least squares and data envelopment analysis. European Journal of Operational Research, 67(3), 332-343.

Battese, G. E., & Coelli, T. J. (1988). Prediction of firm-level technical efficiencies with a generalized frontier production function and panel data. Journal of Econometrics, 38(3), 387-399.

Battese, G. E., & Coelli, T. J. (1992). Frontier production functions, technical efficiency and panel data: with application to paddy farmers in India. Journal of Productivity Analysis, 3(1-2), 153-169.

Berndt, E. R., & Fuss, M. A. (1986). Productivity measurement with adjustments for variations in capacity utilization and other forms of temporary equilibrium. Journal of Econometrics, 33(1), 7-29.

Blundell, R., & Bond, S. (2000). GMM estimation with persistent panel data: an application to production functions. Econometric Reviews, 19(3), 321-340.

Boisso, D., Grosskopf, S., & Hayes, K. (2000). Productivity and efficiency in the US: effects of business cycles and public capital. Regional Science and Urban Economics, 30(6), 663-681

Bruno, M. (1978), "Duality, Intermediate Inputs and Value Added", in Fuss, Melvyn and Daniel McFadden (eds.), Production Economics: A Dual Approach to Theory and Applications, North Holland.

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.

Christensen, L. R., Jorgenson, D. W., & Lau, L. J. (1973). Transcendental logarithmic production frontiers. The Review of Economics and Statistics, 55(1), 28-45.

Cobb, C. W.; Douglas, P. H. (1928). A Theory of Production. American Economic Review 18 (Supplement): 139–165

Coelli, T. (2002). “A comparison of alternative productivity growth measures: with application to electricity generation”, in “Efficiency in the Public Sector” Fox K.(ed), 2002, Springer.

Coelli, T. J., Rao, D. P., O’Donnell, C. J., & Battese, G. E. (2005). An introduction to efficiency and productivity analysis. Springer Science+ Business Media.

Communication from the Commission (2010), Europe 2020: A strategy for smart, sustainable and inclusive growth, Brussels, 3.3.2010

Diewert, W. (2011), ‘Measuring productivity in the public sector: some conceptual problems’, Journal of Productivity Analysis. 36,2: 177-91

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

Diewert, W. E. (1992). Fisher ideal output, input, and productivity indexes revisited. Journal of Productivity Analysis, 3(3), 211-248.

Diewert, W.E (2008), ‘OECD Workshops on Productivity Analysis and Measurement: Conclusions and Future Directions’, in OECD, ‘Productivity Measurement and Analysis’.

Diewert, W.E. & Nakamura, A.O. (2009), “Accounting for Housing in a CPI,” chapter 2, in Diewert, W.E., B.M. Balk, D. Fixler, K.J. Fox and A.O. Nakamura (eds), Price and Productivity Measurement: Volume 1 -- Housing. Trafford Press.

Diewert, W.E. & Lawrence, D. (2006), ‘Measuring the Contributions of Productivity and Terms of Trade to Australia’s Economic Welfare, Report by Meyrick and Associates to the Australian Government’, Productivity Commission, Canberra, Australia, 106 pp. http://www.oecd.org/dataoecd/7/19/37503743.pdf (accessed on 12 February 2013)

Emrouznejad, A. Parker, B. & Tavares, G. (2008): Evaluation of research in efficiency and productivity: A survey and analysis of the first 30 years of scholarly literature in DEA Journal of Socio-Economics Planning Science, 42(3) 151-157.

Emrouznejad, A., & Amin, G. R. (2009), DEA models for ratio data: Convexity consideration. Applied Mathematical Modelling, 33(1), 486-498.

Eurostat (2001), ‘Handbook on price and volume measures in national accounts’, Luxembourg: Office for Official Publications of the European Communities, 2001

Eurostat, OECD (2012). "Methodological manual on purchasing power parities."

Färe R., Grosskopf S. & Margaritis D. (2008) “Efficiency and Productivity: Malmquist and more” in Fried, H. O., Lovell, C. K., & Schmidt, S. S. eds “The measurement of productive efficiency and productivity growth”. Oxford University Press, USA.

Färe, R., & Grosskopf, S. (1996). Productivity and intermediate products: a frontier approach. Economics Letters, 50(1), 65-70.

Färe, R., & Primont, D. (1995). Multi-output production and duality: theory and applications. Kluwer Academic Pub.

Färe, R., Grosskopf, S., Lindgren, B., & Roos, P. (1992). Productivity changes in Swedish pharamacies 1980–1989: A non-parametric Malmquist approach. Journal of Productivity Analysis, 3(1-2), 85-101.

Färe, R., Grosskopf, S., Norris, M., & Zhang, Z. (1994). Productivity growth, technical progress, and efficiency change in industrialized countries. The American Economic Review, 66-83.

Färe, R., Grosskopf, S., Lovell, C.A.K., 1985. The Measurement of Efficiency of Production. Kluwer–Nijhoff Publishing, Dordrecht

Fisher, I., & Brown, H. G. (1911). The Purchasing Power of Money: Its Determination and Relation to Credit, interest and Crises: By Irving Fisher, Assisted by Harry G. Brown. Macmillan.

Fried, H. O., Lovell, C. K., & Schmidt, S. S. (2008), “Efficiency and Productivity”, Chapter 1 in “The measurement of productive efficiency and productivity growth”. Oxford University Press, USA.

Giraleas D., Emrouznejad A. & Thanassoulis E. (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

Greene, W., (2008) “The econometric approach to efficiency analysis” in Fried, H. O., Lovell, C. K., & Schmidt, S. S. (eds) “The measurement of productive efficiency and productivity growth”. Oxford University Press, USA.

Hall, R. E. (1991). Invariance properties of Solow’s productivity residual (No. w3034). National Bureau of Economic Research.

Hill, R. (2007), A taxonomy of multilateral methods for making international comparisons of prices and quantities, Review of Income and Wealth, 43(1), 49-69.

Hollingsworth, B., & Wildman, J. (2003). The efficiency of health production: re-estimating the WHO panel data using parametric and non-parametric approaches to provide additional information. Health Economics, 12(6), 493-504.

Inklaar, R. & Timmer, M.P. (2008). GGDC Productivity Level Database: International Comparisons of Output, Inputs and Productivity at the Industry Level, Groningen Growth and Development Centre Research Memorandum GD-104

Inklaar, R., & Timmer, M. P. (2009). Productivity convergence across industries and countries: the importance of theory-based measurement. Macroeconomic Dynamics, 13(S2), 218-240.

Jondrow, J., Knox Lovell, C. A., Materov, I. S., & Schmidt, P. (1982). On the estimation of technical inefficiency in the stochastic frontier production function model. Journal of Econometrics, 19(2), 233-238.

Jorgenson, D. W. (1963). Capital Theory and Investment Behavior. American Economic Review 53, no. 2: 247-59

Kopp, R. J., & Diewert, W. E. (1982). The decomposition of frontier cost function deviations into measures of technical and allocative efficiency. Journal of Econometrics, 19(2), 319-331.

Krugman, P. (1994). The myth of Asia’s miracle. Foreign affairs, 73(6), 62-78.

Kumbhakar, S. C. (2002). Productivity measurement: A profit function approach. Applied Economics Letters, 9(5), 331-334.

Kumbhakar, S. C., & Lovell, C. K. (2000). Stochastic frontier analysis. Cambridge University Press.

Kumbhakar, S.C. , Park, B.U., Simar, L. & E.G. Tsionas (2007), Nonparametric stochastic frontiers: a local likelihood approach, Journal of Econometrics, 137, 1, 1–27

Kuosmanen, T. (2003). Duality theory of non-convex technologies. Journal of Productivity Analysis, 20(3), 273-304.

Kuosmanen, T., & Kortelainen, M. (2012). Stochastic non-smooth envelopment of data: semi-parametric frontier estimation subject to shape constraints. Journal of Productivity Analysis, 38(1), 11-28.

Meeusen, W., & van Den Broeck, J. (1977). Efficiency estimation from Cobb-Douglas production functions with composed error. International Economic Review, 18(2), 435-444.

OECD (2001), 'Measuring Productivity, Measurement of aggregate and industry-level productivity growth.'

OECD, Eurostat, WHO (2011), A System of Health Accounts, OECD Publishing.

Office for National statistics (2007), The ONS productivity handbook – A statistical overview and guide, edited by Dawn Camus

Olesen, O.B. & Petersen, N.C., Chance constrained efficiency evaluation, Management Science, 1995, 41, 442-457

Olley, G. S., & Pakes, A. (1996). The Dynamics of Productivity in the Telecommunications Equipment Industry. Econometrica, 64(6), 1263-1297

O'Mahony, M., & Timmer, M. P. (2009). Output, input and productivity measures at the industry level: The eu klems database*. The Economic Journal, 119(538), 374-403.

Pastor, J. T., & Lovell, C. A. (2005). A global Malmquist productivity index. Economics Letters, 88(2), 266-271.

Portela, M. C., & Thanassoulis, E. (2010). Malmquist-type indices in the presence of negative data: An application to bank branches. Journal of Banking & Finance, 34(7), 1472-1483.

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.

Resti, A. (2000). Efficiency measurement for multi-product industries: A comparison of classic and recent techniques based on simulated data. European Journal of Operational Research, 121(3), 559-578.

Retzlaff-Roberts, D. L., & Morey, R. C. (1993). A goal-programming method of stochastic allocative data envelopment analysis. European Journal of Operational Research, 71(3), 379-397.

Ruggiero, J. (1999). Efficiency estimation and error decomposition in the stochastic frontier model: A Monte Carlo analysis. European Journal of Operational Research 115, no. 3 (June 16): 555-563.

Ruggiero, J. (2007). A comparison of DEA and the stochastic frontier model using panel data. International Transactions in Operational Research, 14(3), 259-266.

Samuelson, P. A. (1947). Foundations of Economic Analysis, Harvard University Press.

Schreyer, P. (2001). OECD Productivity manual: A guide to the measurement of industry-level and aggregate productivity growth. Organisation for Economic Co-operation and Development.

Sharma, S. C., Sylwester, K., & Margono, H. (2007). Decomposition of total factor productivity growth in US states. The Quarterly Review of Economics and Finance, 47(2), 215-241.

Shephard, R. W. (1953). Cost and production functions. Princeton University NJ.

Simar, L. and Wilson, P.W., (2008) “Statistical inference in non-parametric frontier models” in Fried, H. O., Lovell, C. K., & Schmidt, S. S. (eds) “The measurement of productive efficiency and productivity growth”. Oxford University Press, USA.

Solow, R. (1957): "Technical Change and the Aggregate Production Function." Review of Economics and Statistics, 39:312-320.

Thanassoulis, E. (1993). A comparison of regression analysis and data envelopment analysis as alternative methods for performance assessments. Journal of the Operational Research Society, 1129-1144.

Thanassoulis, E., Boussofiane, A., & Dyson, R. G. (1996). A comparison of data envelopment analysis and ratio analysis as tools for performance assessment. Omega, 24(3), 229-244.

The European Advisory Committee on Statistical Information In the Economic and Social Spheres, “Are we measuring productivity correctly?”, Background paper for the 31st CEIES SEMINAR Rome, Italy 12 – 13 October 2006

Timmer, M., O’Mahony, M., & Van Ark, B. (2007). EU KLEMS Growth and Productivity Accounts: Overview November 2007 Release. Groningen: University of Groningen. Online at www.euklems.net/data/overview_07ii.pdf (accessed 10th July 2009).

Tinbergen, J. (1942), Zur Theorie der Langfristigen Wirtschaftsentwicklung, (On the Theory of Long-Term Economic Growth), Weltwirtschaftliches Archiv, 55: 511-549.

Triplett, J. (2004), ‘Handbook on Hedonic Indexes and Quality Adjustments in Price Indexes’, OECD STI WORKING PAPER 2004/9, OECD http://www.oecd.org/dataoecd/37/31/33789552.pdf (accessed on 12 May 2013)

Van Biesebroeck, J. (2007). Robustness of Productivity Estimates. The Journal of Industrial Economics, 55(3), 529-569.

Walsh, C. M. (1901). The measurement of general exchange-value (Vol. 25). The Macmillan Company.

A1 Literature review of simulation studies in efficiency and productivity analysis

There are a number of elements in other simulation studies in the field of efficiency and productivity analysis that are of particular interest to this research, namely:

- the data generating process for:
  - inputs,
  - efficiency scores,
  - measurement error and
  - other factors relevant to the analysis (such as input and output prices and the degree of multi-collinearity between the various variables),
- the functional form used in the construction of output,
- the sample size used for the simulation,
- the number of simulation runs undertaken,
- the measures utilised to judge the overall accuracy of the estimates,
- the efficiency measurement approaches examined and finally
- the relative accuracy of the approaches.

This literature review section will briefly summarise the methods and results from a number of similar simulation studies and will examine in more detail three studies that were identified as being both especially relevant and relatively recently published. These are:

- the Resti (2000) study
- the Banker et al. (2004) study; and finally
- the Van Biesebroeck (2007) study.

Brief summary of less recent studies

One of the older studies in this literature review is by Banker et al. (1987). The aim of this study was to compare the accuracy of the efficiency estimates under DEA and COLS. The study employed a piecewise log-linear production function to generate values for a single output using two discrete inputs for two sets of 100 and 500 observations. The output was multiplied by an efficiency score; approximately 30% of the observations had an efficiency score of 1 and the rest displayed efficiency scores that followed U~[0.65,1). The study employed a CRS DEA model and a translog production function for the COLS methodology. The study found that the mean absolute deviation (MAD) between the 'true' efficiency scores and the estimated efficiency scores was smaller for the DEA-based estimates[92] relative to the COLS estimates. DEA was also able to correctly identify almost all observations that were truly efficient as DEA efficient (ie having an efficiency score of 1); only 'corner' observations (ie that displayed either very large or very small values in either the inputs or the outputs) were misclassified. The study also considered the effects of using a translog production function to generate the output; it found that DEA-based efficiency estimates were more accurate than the COLS-based estimates for both sample sizes, despite the fact that the parametric approach utilised the same underlying functional form for the production function. It should be noted that the findings of this study are based on a single simulation run (ie the experiment was not repeated) and, as such, it is not clear whether these results are robust to different starting input data.

A similar study was undertaken at a later date by Banker, Cadh and Gorr (1993) again comparing DEA and COLS efficiency estimates, but this time also examining the effects of including an error component in the generation of output. This error component, u, followed N~(0,σ_u) and was directly multiplied to the output derived from the DGP. Two values for σ_u where adopted, 1.05 and 1.20. The study also examined the sensitivity of the efficiency estimates under different underlying efficiency distributions, sample sizes and log-linear production technologies. The main findings of this study were:

– there can be substantial variation between 'true' and estimated efficiency scores;

– COLS performs better than DEA when sample size increases and inefficiency is exponentially distributed;

– DEA performs better than COLS when inefficiency is distributed according to the half-normal, truncated-normal and or other two-parameter distributions;

– DEA is also more accurate than COLS when sample size is small (ie 25 obs);

– both methods fail to provide accurate estimates when measurement error is high.

[92] The reported MAD scores for the 500 obs and 100 obs cases, respectively where: 0.003 and 0.010 for the DEA model and 0.024 and 0.044 for the COLS model.

A similar study was also undertaken by Thanassoulis (1993), using a linear cost function (one output and three inputs) to assess the accuracy between DEA and regression analysis efficiency estimates in a small sample (15 obs). This study also found that DEA provides more accurate efficiency estimates on average, although for some DMUs with extreme values (either very large or small inputs or output), the estimated inefficiency was highly inaccurate. On the other hand, although regression analysis-derived efficiency estimates were less accurate on the whole, for the aforementioned ‘corner’ cases the estimated efficiency scores were closer to the true values.

A more recent study by Ruggiero (1999) focused on the accuracy of the parametric approaches and more specifically on the ability of the SFA model to correctly decompose the residual into estimates of inefficiency and noise in the cross-sectional setting. The data used for the simulations was generated using a simple Cobb-Douglas functional form with one output and two inputs and constant returns to scale. Output values were further modified by adding an element of technical inefficiency *u* to the model ($u \sim N^{+}(0,0.20$ or $0.25)$, leading to average efficiency scores of 86% to 83%) and an element of measurement error or noise *ε* ($\varepsilon \sim N(0,0.05$ or 0.10 or $0.15)$, to reflect ‘low’, ‘medium’ and ‘large’ measurement error). The study used five different sample sizes (of 25, 50, 75, 100 and 200 observations) over 100 simulation runs to test the accuracy of efficiency estimates for a number of SFA models with different functional form specifications and assumptions for the distribution of the inefficiency terms. The study also examined the performance of a simple, Cobb-Douglas deterministic COLS model, as a contrast to the more complex, stochastic SFA specifications. The performance criterion was rank correlation statistics.

The results of the Ruggiero (1999) study were unexpected. The correctly specified SFA model (both in terms of the functional form and the distribution of inefficiency) performed quite poorly in the majority of the situations and especially in small sample sizes. The deterministic COLS model performed better than the SFA models in almost all cases; only when the sample size was large (200 observations), did the correctly specified SFA model match the performance of the COLS model. In addition, the standard deviations of the rank correlations for the SFA models are quite large, which suggests of a general inconsistency on the part of the SFA estimates. Other interesting findings include:

– When measurement error is large, both approaches perform relatively poorly. The rank correlation between true and estimated efficiency scores were above

85% for the best models in ‘low’ measurement error scenarios and approximately 65% in ‘high’ measurement error scenarios.

– Functional form misspecification can have a significant impact to accuracy when true inefficiency is expected to be relatively high.

– Omitted variables have a significant negative impact to accuracy under all conditions and approaches.

– Misspecification of the inefficiency distribution has a relative small negative impact on the accuracy of the SFA models, especially when sample size increases.

In another study, Ruggiero (2007) examined the performance of SFA and DEA when panel data is available. This study was partly motivated by the fact that the cross-sectional SFA efficiency estimates are statistically inconsistent (see chapter 3 for discussion on this issue). This shortcoming can be overcome when panel data is available, since the panel data incorporate additional information from the time-series nature of the data. The panel data SFA model (with time-invariant inefficiency) was compared to a COLS model, and two DEA models, the classic CRS model and the slack-based DEA model that estimates the Russell measure of efficiency[93]. For all determinist models, the panel dataset was collapsed to a single cross-section of the average values of both inputs and output (ie the data was averaged across time).

Data was generated using a simple, two-input, one-output, Cobb-Douglas production function with constant returns to scale. The inefficiency component *u* was time-invariant and was half-normally distributed ($u \sim N^{+}(0,0.20)$, leading to average efficiency scores of 86%) and measurement error *ε* was normally distributed, with variable variance ($\varepsilon \sim N(0,0.05$ or 0.10 or 0.15 or 0.20 or $0.30)$). The simulations observe 100 units over a varying amount of time periods, namely 5, 10, 15 and 20 time periods. Each scenario was replicated 100 times and the results were averaged. The accuracy of the efficiency estimates were judged based on both Pearson’s and rank correlations with the true values.

The study found that:

– As the number of time periods considered increases, so does the accuracy of the estimates for all approaches.

[93] Defined in Färe, Grosskopf, and Lovell (1985).

– As the variance of the measurement error decreases, the accuracy of all approaches improves.

– In almost all cases, the SFA estimates outperform the other measures. However, the difference in performance is relatively small. The study does point out that the SFA model is perfectly specified in all respects (functional form, inefficiency distribution and the fact that inefficiency is time-invariant).

– The Russell measure outperforms the more standard CRS DEA efficiency measure in the majority of the cases considered.

Resti (2000) study

Resti (2000) examines the performance of DEA, SFA and 'stochastic' DEA (as proposed by Retzlaff-Roberts and Morey (1993)), when data is generated such as they are more 'realistic'. In fact, one of the central aims of this study was to devise a more complex data generating process that produces data that are similar to those found in real-world applications in the banking sector. To that aim, the study utilises three inputs and two outputs together with their respective prices, which potentially allow for the estimation of both technical and allocative efficiency in the cross-sectional setting. Such efficiencies are estimated using:

– a translog cost function to create a measure of overall efficiency, which is then decomposed into a measure of technical and allocative efficiency following the methodology set out by Kopp and Diewert (1982);
– a 'standard' DEA model for the estimation of technical efficiency and the allocative DEA model for the estimation of overall efficiency, which allows for the calculation of the allocative efficiency component;
– a series of stochastic DEA models to measure overall efficiency (namely an additive model, a multiplicative model and a model with a heteroskedastic error component).

Input data were constructed using a multiple log-normal distribution, so that the values are skewed to the right (indicating the presence of large-scale producers) and to ensure some degree of correlation between the inputs, given that this is feature that was deemed to be commonly observed in the banking sector. Price data were created using a multiple normal distribution and displayed a slight negative correlation with each other. The two outputs were generated using two distinct piecewise log-linear functions[94]. As for the efficiency score and the noise component:

– technical inefficiency scores were generated using a truncated-normal distribution with an implied mean of appr. 7% and a standard deviation of appr. 5% and were applied to both outputs,

– allocative inefficiency scores are generated using the normal distribution with a varying standard deviation (ranging from 0.05 to 0.25) and were applied to all inputs (the final input is generated in such a way as to ensure that allocative inefficiency does not have an impact on the unit's technical efficiency score) and

– the noise component was also generated using the normal distribution with a standard deviation of 1%.

Using the above data generating process, the study constructed six different samples, changing the number of observations (50 or 500) and the assumptions about the degree of allocative inefficiency present in the data. Accuracy was assessed using MAD scores, correlations between estimates and true values and the average bias (true values minus estimates). No repeated simulation runs were undertaken.

The results of this study can be summarised as follows:

– Both SFA and DEA estimates for overall efficiency are quite close to the true values; SFA narrowly outperforms DEA in the larger sample, while DEA significantly outperforms SFA in the smaller sample.

– In the large sample, the larger the ratio of overall efficiency to measurement error, the better the performance of SFA.

– In the smaller sample size, CRS DEA estimates were more accurate than VRS DEA estimates.

– When decomposing overall efficiency into allocative and technical efficiency, the DEA estimates are significantly more accurate than the SFA estimates for all cases.

[94] This design does not allow for any interactions in the production of the two outputs-thus the data generation process assumes that there is complete separability in output production, so no trade-offs between outputs and no economies of scope are possible.

– All stochastic DEA models mostly underperformed relative to both SFA and DEA when sample size was small, with few minor exceptions. The experiment was not attempted for the larger sample size due to computational concerns.

Banker, Chang, and Cooper (2004) study

The main focus of this study is to examine the effects of the presence of heteroskedasticity in the inefficiency and/or noise component in the context of the production function. However, the study also provides accuracy estimates, measured as mean and median absolute deviations of the estimated efficient output relative to the true efficient output, for cases with no heteroskedasticity for a wide range of approaches and thus is relevant for this research. More specifically, the study examines the accuracy of:

– The standard VRS DEA model and two second-stage DEA-adjusted regression models (at the first stage efficiency scores are estimated using a VRS DEA model and the second stage a log-linear and log-quadratic model is fitted using the efficient output as the dependent variable)

– A COLS model, adopting a log-linear and log-quadratic functional form (two specifications)

– A MOLS (Modified OLS) model (similar to a COLS model but the ‘correction’ of the OLS model is based on the model’s mean square error rather than the maximum residual), adopting a log-linear and log-quadratic functional form (two specifications).

– An MLE model (equivalent to an SFA model which assumes that the inefficiency term is exponentially distributed), adopting a log-linear and log-quadratic functional form (two specifications).

– The Aigner and Chu (A&C) goal programming model (Aigner and Chu (1968)), adopting a log-linear and log-quadratic functional form (two specifications).

Output data was generated using a log-linear piecewise production function with a single output. Output values were further modified by adding a noise component, $\varepsilon \sim N(0,0.05)$ and an inefficiency component $u \sim \exp(7)$, which translates to an average efficiency of 84%. For some cases, the study also introduced an element of heteroskedasticity in the measurement error and/or the inefficiency term; however, since the issue of heteroskedasticity is not examined in the current research, these cases are only briefly discussed here. Three different sample sizes were used, of 25, 50 and 100 units and each simulation case was replicated 100 times.

The main results of this study were as follows:

– When no measurement error is present, the standard DEA model provides the most accurate estimates of efficient output, regardless of the sample size. The second most accurate approach was COLS using the log-linear specification, followed by the MLE model; for this model, the log-quadratic specification appeared to be performing marginally better than the log-linear specification. Although the differences in MAD scores between the DEA and COLS estimates were quite small, ranging from 0.06 to 0.47[95], the differences between the COLS and MLE MAD scores were substantially larger, ranging from approximately 7.6 to 8.2. By far the worst performer was the A&C model, with differences in MAD scores in relation to the DEA model ranging from approximately 35 to 45 units.

– When both measurement error and inefficiency are present, the DEA model still outperformed every other approach, regardless of the sample size, although absolute performance was worse than the case with no measurement error. It's interesting to note however that the two-stage, log-linear DEA-adjusted model performed marginally better than the standard DEA model in all sample sizes. It is not clear why this is the case, but it might be due to the smoothing effect that the second stage regression can have on the estimated efficient output. The log-linear COLS model was the second most accurate approach in sample sizes smaller than 100; for the larger sample sizes, MLE proved to be slightly more accurate (by about 0.5 units), for both parametric functional form specifications. In addition, the performance of the MOLS model was quite close to the performance of the MLE model. The A&C model remained the worst performer by far.

– When heteroskedasticty was introduced, the relative rankings of the approaches did not change; the DEA estimators (specifically the second-stage DEA adjusted regression) were still the more accurate, followed by COLS and MLE approaches. The results of the study also indicated that the presence of heteroskedasticty in either the measurement error or the inefficiency term have no material impact on the performance of the efficiency estimators of all the approaches examined.

[95] It is unclear from the study what the average output is, so these values cannot easily be put into context.

Van Biesebroeck (2007) study

The study by Van Biesebroeck appears to be of particular relevance to this research, since its stated aim is to access the robustness of 'productivity' estimates derived from a number of techniques which include index numbers (IN), DEA, OLS, SFA, the GMM system estimator (Blundell and Bond (2000)) and the semi-parametric approaches originally proposed by Olley and Pakes (1996). However, as will become apparent, there are a number of incompatibilities in critical definitions and in general experiment design that limit the usefulness of this study for this research.

The study examines the accuracy of each estimator, measured by the correlation of the estimated values to the true values, under these conditions:

– when a noise component is introduced in the price of labour, which this study refers to as 'heterogeneity in factor input prices';

– when measurement error is introduced in the measurement of input and output quantities; and

– when the production function is firm-specific (ie input elasticities are different for each firm).

The study assumes that all firms are technically efficient; in fact, the concept of inefficiency, which is central in the present research, plays only a very small part in this study. Specifically, any deviation from maximum output is solely the result of the first condition mentioned above. The introduction of the noise component in the price of labour can be interpreted as a way to simulate optimisation mistakes on the part of the producers; an efficient producer will optimise according to the price of labour that includes the noise component, while 'true' productivity change will be determined by the optimisation process that includes the price of labour without noise. Therefore, the introduction of the noise component could be interpreted as introducing an element of allocative inefficiency to the production process.

The study focuses on 'productivity levels', productivity growth and bias in the estimated factor elasticities. A 'productivity level' is defined as the natural logarithm of the productivity of a firm at a point in time relative to the average productivity of the sample at the same point in time, ie

$$\ln(A_{it}) - \overline{\ln(A_t)}, \qquad \text{Eq A.1.1}$$

where A_{it} is the unobservable productivity term found in the standard model of production with two inputs, labour and capital, and one output, namely value added:

$$Y_{it} = A_{it} F_{(it)}(l_{it}, k_{it}) \qquad \text{Eq A.1.2}$$

In essence, this productivity term appears to be what is more commonly referred to in the Efficiency and Productivity literature as technical efficiency. However, since this study assumes that all firms are technically efficient, this productivity term represents an inherent characteristic of each individual firm. Furthermore, the study assumes that this characteristic is **known** to the firm before production begins in each time period (both in terms that this effect exists and also its exact level and its impact to production), **but is not observable by the researcher**. Since this characteristic and its effects are known to each individual producer before production takes place, each producer will take it into account when optimising its production process. This is significant, since this feature gives a significant advantage to the index numbers approaches considered in this study. As such, this productivity term could also be interpreted as a firm-specific effect that can change over time, but is always known to the firm in question before production takes place in each time period.

Productivity growth is defined as the natural logarithm of productivity of a firm at time *t* relative to the productivity of the same firm at time *t-1*, ie

$$\ln(A_{it}) - \ln(A_{it-1}) \qquad \text{Eq A.1.3}$$

For DEA, productivity levels are defined as:

$$\ln(A_{it}^{DEA}) - \overline{\ln(A_{t}^{DEA})} = \ln\theta_{it} - \overline{\ln(\theta_{t})} \qquad \text{Eq A.1.4}$$

where θ is the DEA efficiency measure and productivity growth is defined as:

$$\ln(A_{it}^{DEA}) - \ln(A_{it-1}^{DEA}) = \ln\theta_{it} - \ln\theta_{it-1} \qquad \text{Eq A.1.5}$$

,which is more commonly referred by the Efficiency and Productivity literature as efficiency change (which is in turn only one of the components of the Malmquist productivity index).

For OLS and SFA (denoted as PAR in the equations below), productivity levels are defined as:

$$\ln(A_{it}^{PAR}) - \overline{\ln(A_{t}^{PAR})} = (q_{it} - \overline{q}_t) - \hat{a}_l^{PAR}(l_{it} - \bar{l}_t) - \hat{a}_k^{PAR}(k_{it} - \bar{k}_t) \qquad \text{Eq A.1.6}$$

, where both inputs and the output are in logarithms and $\hat{a}_l^{PAR}$ and $\hat{a}_k^{PAR}$ are the estimated elasticities of labour and capital. Productivity growth is defined as:

$$\ln(A_{it}^{PAR}) - \ln(A_{it-1}^{PAR}) = (q_{it} - q_{t-1}) - \hat{a}_l^{PAR}(l_{it} - l_{t-1}) - \hat{a}_k^{PAR}(k_{it} - k_{t-1}) \qquad \text{Eq A.1.7}$$

It is clear from the above definition that the ‘productivity’ term includes both the effects of the ‘heterogeniety’ in factor inputs and the error term; also, the estimated production function does not include a time trend and thus the rate of technological change is not measured. The estimated production function is given by:

$$q_{it} = a_0 + a_l l_{it} + a_k k_{it} + \mu_{it} + \text{£}_{it} \qquad \text{Eq A.1.8}$$

where all parameters are in logarithms, μ_{it} represents the productivity term that is known to the to the firm but is unobservable to the researcher and £_{it} represents the measurement error.

The productivity term is modelled according to Battese and Coelli (1992) as:

$$\mu_{it} = -e^{-\text{¦}(t-T)}\mu_i \text{ with } \mu_i \sim N^{+}(\text{¥}, \pm^2) \qquad \text{Eq A.1.9}$$

ie productivity is modelled as a time-invariant draw from a truncated normal distribution and increases (or decreases) over time if *η* is positive (or negative) at the same rate for all firms.

The study generates output using a Cobb-Douglas production function with constant returns to scale (although this changes in the third set of modelled conditions, as mentioned above), while inputs are generated as a function of input prices and the firm-specific productivity effect[96]. Results are based on 50 simulation runs of 200 firms observed over 10 years.

The use of non-standard definitions of productivity and productivity growth as well as the assumptions on the absence of technical efficiency, the fact that the productivity term is fixed (or follows an AR-1 process) and is known to the firm, all serve to make the findings of this study incompatible with the simulation methodology adopted for the present research. Therefore, the findings of the Van Biesebroeck study are of

[96] To achieve this, the study assumes that all firms are revenue maximising and have perfect information on both current **and future** levels of input prices and firm-specific productivity effects.

limited usefulness as a cross-check and as such, they are not reported here. This is quite unfortunate, since the Van Biesebroeck study is the only simulation-based study that specifically examined the robustness of ‘productivity growth’ estimates, which is the central aim of the current research.

A2 Generating a piecewise-linear function

There are three main issues that need to be considered when generating a convex, monotonic, piecewise-linear function (abbreviated here as PLF):

- The number of ‘facets’, or pieces it will contain;
- The breakpoints that will determine the start and end points of each piece; and lastly,
- The parameters of each piece itself, ie the coefficients of the right-hand side variables.

The starting point of the generation process is to randomly determine the number of the efficient facets that describe the frontier. These facets correspond to the number of ‘pieces’ or linear equations that, when combined, constitute the production function. In general, but not necessarily always, the higher the number of assessed units, the higher the number of efficient facets, assuming that the ranges of input and output values remain relatively constant.

For the PLF detailed in section 4.3.1, the number of facets was randomly generated from a uniform distribution (U[3,10]).The upper and lower limits of the distribution were chosen based on a number of simulation experiments designed to provide an estimate of the number of facets that are likely to be observed in applications with one output and two inputs in samples of 100 observations (conditions similar to those used in the simulation analysis of chapter 4). The experiments first utilised a Cobb-Douglas production function to generate output values for randomly generated inputs, following the methodology described in chapter 4; an element of moderate inefficiency was then included in the generated output (average efficiency was approximately 85%[97]) and the resulting 100 units were assessed using DEA. DEA was used in this case, since the approach formulates the frontier as a PLF and it is a relatively straight-forward process to count the facets of the frontier in such application.

The next stage of the generation process is to identify the breakpoints of each piece and parameterise the resulting linear functions. However, this is not a straightforward task, because the resulting PLF needs to be continuous, convex and monotonic, so that it conforms to standard production theory. The key characteristic that can be

[97] These are the conditions that are prevalent in the simulation experiments that were undertaken in chapter 4.

used to ensure that the resulting function displays the desirable properties is the input mix; since the PLF needed here uses just two inputs, this mix is easily represented as the ratio of the two inputs. This input ratio is critical to both ensuring monotonicity and as a marker to identify the breakpoints of each ‘piece’-ie the points in the continuous linear function where the parameters change.

It turns out that monotonicity is relatively straightforward to impose; it only requires that the breakpoints of each piece are generated such that the input ratio is increasing in each subsequent piece. So what is needed now is a way to generate these breakpoints, so that they are consistent with the underlying data generation process that is used in the simulation analysis undertaken in chapter 4, specifically so that the breakpoint values are similar to what are likely to be observed in datasets generated using the same parameters as those used in the chapter 4 simulations.

This was achieved here by randomly generating sets of two inputs using the same process as the one employed for the simulation exercises (ie 100 observations of two inputs, each input drawn from a uniform distribution following U[0,1]) and calculating their input ratios; this provided an indication of the range of input ratios that are likely to be observed in such conditions. To identify the actual breakpoints, the input ratios were sorted in ascending order (to ensure monotonicity) and a number of ratios equal to the number of facets that was determined in the previous step were selected based on their rank. The actual values of the chosen ranks were randomly determined, but using a methodology that would ensure that the selected breakpoints were neither too close together, nor too far apart. However, since the analysis utilises ratios, it is difficult to define a reasonable measure of distance, ie it is difficult to say what is too close or too far apart. So instead of finding a way to judge these distances, the analysis instead adopted the notion that each individual piece should represent an almost equal number of input ratios, as these were observed in the generated data. This is akin to the notion of a DEA frontier, where each identified facet has an almost equal number of DMUs projected to it.

To implement this, the input ratios previously calculated were each ‘assigned’ to an individual piece. To accomplish this, the analysis randomly determined a frequency statistic, ie how many observations out of the 100 were to be included in each piece. The frequency statistic is a random number generated from a normal distribution with a mean of the expected number of observations by piece (calculated as the ratio of available observations to the number of desired breakpoints), and a standard deviation equal to one half of the mean. After the numbers were rounded and corrected so that they add up to 100 (ie, the number of observations in the dataset), each individual observation was assigned to a piece, based on its rank. In other words, the observations were sorted in an ascending order based on their input ratios and assigned to each piece based on the previously generated frequency; so, assuming that the frequency of the first piece was 12, the first 12 observations in the sorted list were assigned to the first piece, and so on. Then, the midpoint of the input ratios for the 12th and 13th observations was deemed to be the breakpoint when the PLF switches to a new linear function.

The above can become clearer when demonstrated in an example:

Assume that the number of observations is set to 30 and they need to be assigned to two pieces. Also assume that the frequency of the first piece is 12; this means that the first 12 sorted observations will be assigned to the first piece and the remaining 18 observations will be assigned to the second facet. The sample input data are given in the table below:

Table A2.1: Piecewise linear function generation example data

Rank	L	K	L/K	Facet	Threshold
1	0.004975	0.926145	0.01	1	
2	0.014496	0.407422	0.04	1	
3	0.037965	0.796258	0.05	1	
4	0.042482	0.518143	0.08	1	
5	0.095798	0.777642	0.12	1	
6	0.082278	0.575915	0.14	1	
7	0.040712	0.23072	0.18	1	
8	0.064058	0.358348	0.18	1	
9	0.03238	0.164129	0.20	1	
10	0.177953	0.866756	0.21	1	
11	0.21775	0.859035	0.25	1	
12	0.211097	0.740471	0.29	1	
					=(0.29+0.31)/2
13	0.255623	0.817896	0.31	2	=0.30
14	0.291299	0.802149	0.36	2	
15	0.300211	0.802179	0.37	2	
16	0.074343	0.198431	0.37	2	
17	0.373455	0.9859	0.38	2	
18	0.270943	0.699698	0.39	2	
19	0.100314	0.256691	0.39	2	
20	0.280343	0.703299	0.40	2	
21	0.300211	0.750206	0.40	2	
22	0.17365	0.404798	0.43	2	
23	0.351482	0.775658	0.45	2	
24	0.496048	0.850642	0.58	2	
25	0.523453	0.896786	0.58	2	
26	0.085055	0.132267	0.64	2	
27	0.601886	0.930052	0.65	2	
28	0.596484	0.899106	0.66	2	
29	0.528794	0.796686	0.66	2	
30	0.668111	0.926939	0.72	2	

The value of the breakpoint will be between the input ratio of the 12th and 13th observation. So, according to the data presented in the table above, the generated PLF will take the form of:

$$y_i = \begin{cases} a_1+b_{L1}L_i+b_{K1}K_i \text{ for } L_i/K_i<=0.3 \\ a_2+b_{L2}L_i+b_{K2}K_i \text{ for } L_i/K_i>0.3 \end{cases} \quad \text{Eq A2.1}$$

Since chapter 4 did not consider the issue of variable returns to scale the values of the parameters $a_{\text{facet number}}$ were all set to zero. Thus all linear functions pass through the origin, ensuring that the resulting production function displays constant returns to scale globally.

As a reminder, the only purpose of this exercise is to derive breakpoints that are appropriately spaced and that conform to monotonicity constraints. The process might seem somewhat convoluted, but it does ensure that the generated breakpoints are appropriately spaced from one another and as such, no single piece of the final PLF dominates.

The third and final step is to generate the input coefficients for each linear piece. To accomplish that, first the coefficients for the first input (the one used as the numerator of the input ratio) were randomly generated for all pieces. For the analysis undertaken in chapter 4, these coefficients were generated as random draws from a uniform distribution (U[0,2]). The generated coefficients were then assigned to each piece, *in descending order*, ie the largest coefficient was assigned to the first piece and so on, until the smallest coefficient was assigned to the last piece. This sorting of the coefficients is critical, since it ensures the convexity of the resulting piecewise-linear function.

The 'starting' coefficient of the second input (ie the coefficient of the second input corresponding to the first piece) was also randomly generated[98]; this was achieved by drawing a sample of random numbers equal to the number of facets using the same uniform distribution as before and selecting the *smallest* value. So, the first facet has the largest coefficient for one input and the smallest coefficient for the second input; as was previously mentioned, this ensures that the resulting piecewise-linear function is convex. The rest of the coefficients for the second output are set in such a manner as to ensure that the piecewise-linear function is continuous. Continuity is achieved by ensuring that at each breakpoint, the linear functions corresponding to each consecutive piece result in the same value for y. So, assuming CRS (ie the values of all *a* set to zero), at each threshold to following should hold:

$$b_{Lj}L_i + b_{Kj}K_i = b_{Lj+1}L_i + b_{Kj+1}K_i \qquad \text{Eq A2.2}$$

[98] Or it could be set to be equal to a fraction of the corresponding coefficient for the first input, say 10%.

Rearranging this equation gives:

$$(b_{Lj} - b_{Lj+1})\frac{L_i}{K_i}b_{Kj} = b_{Kj+1} \qquad \text{Eq A2.3}$$

and since the $\frac{L_i}{K_i}$ is known (it is the threshold value), b_{Kj+1} can be calculated.

Extending the example presented above, assume that:

$b_{L1} = 1.84$

$b_{L2} = 1.53$

$b_{K1} = 0.13$

threshold value = 0.30

$a_1 = a_2 = 0$

Then:

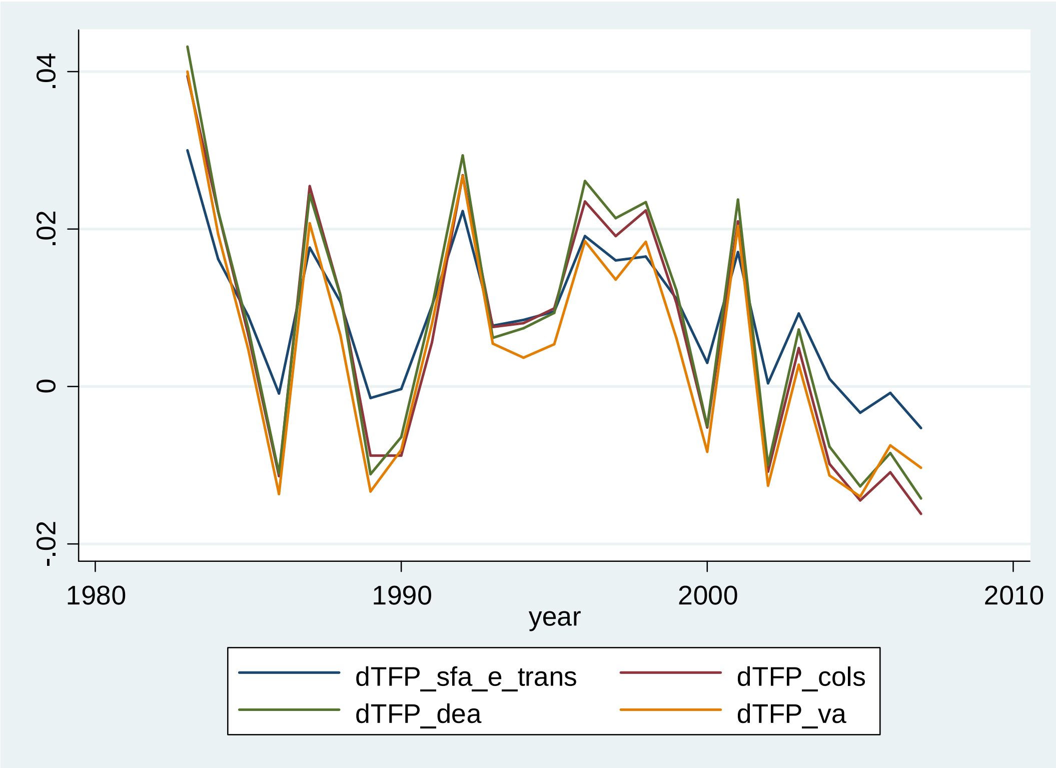
$b_{K1} = (1.84 - 1.53)*0.3 + 0.13 = 0.22$

So the final piecewise-linear production function is:

$$y_i = \begin{cases} 1.84L_i + 0.13K_i \text{ for } L_i/K_i <= 0.3 \\ 1.53L_i + 0.22K_i \text{ for } L_i/K_i > 0.3 \end{cases} \qquad \text{Eq A2.4}$$

A3 Detailed Graphs and tables of productivity performance by country

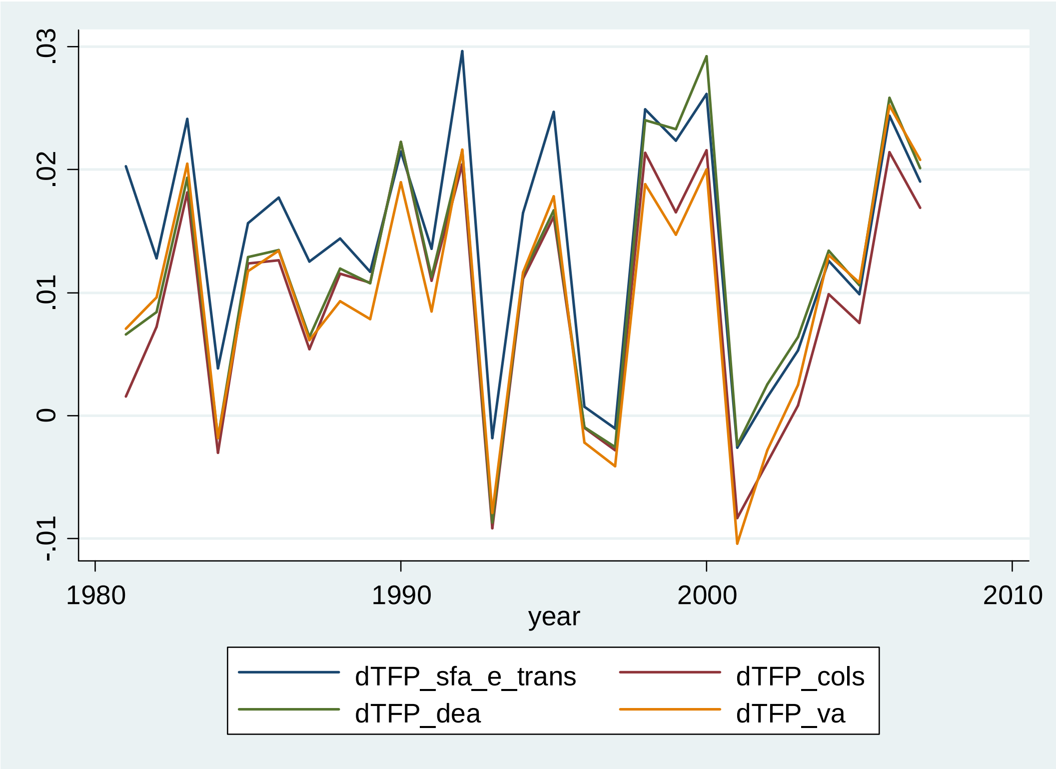
Australia



	DEA	COLS	SFA translog (exponential)	GA
DEA	1			
COLS	0.99	1		
SFA translog (exponential)	1.00	0.99	1	
GA	0.99	0.98	0.98	1

TFP measure	Average	Standard Deviation	Minimum (%)	Minimum (year)	Maximum (%)	Maximum (year)
DEA	0.8%	1.6%	-1.4%	2007	4.3%	1983
COLS	0.7%	1.6%	-1.6%	2007	3.9%	1983
SFA translog (exponential)	0.9%	0.9%	-0.5%	2007	3.0%	1983
GA	0.5%	1.5%	-1.4%	2005	4.0%	1983

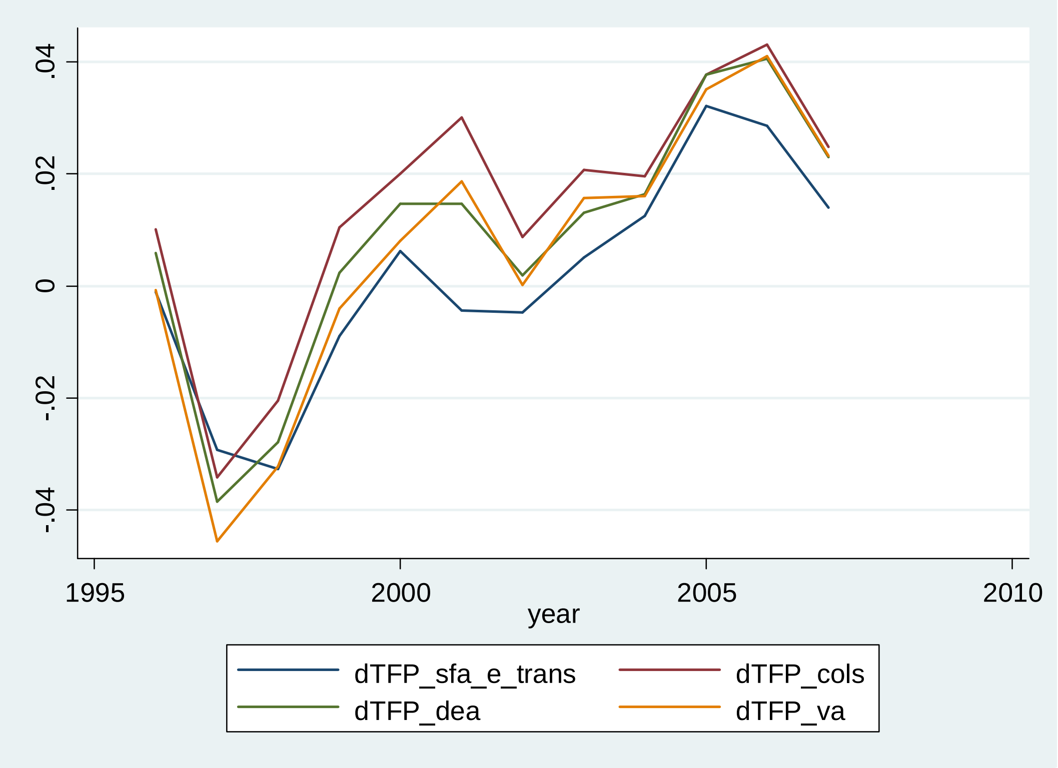
Austria



	DEA	COLS	SFA translog (exponential)	GA
DEA	1			
COLS	0.97	1		
SFA translog (exponential)	0.92	0.93	1	
GA	0.95	0.97	0.94	1

TFP measure	Average	Standard Deviation	Minimum (%)	Minimum (year)	Maximum (%)	Maximum (year)
DEA	1.2%	1.0%	-0.9%	1993	2.9%	2000
COLS	0.9%	1.0%	-0.9%	1993	2.2%	1990
SFA translog (exponential)	1.4%	0.9%	-0.3%	2001	3.0%	1992
GA	1.0%	1.0%	-1.0%	2001	2.5%	2006

Czech Republic



	DEA	COLS	SFA translog (exponential)	GA
DEA	1			
COLS	0.99	1		
SFA translog (exponential)	0.97	0.91	1	
GA	0.99	0.99	0.94	1

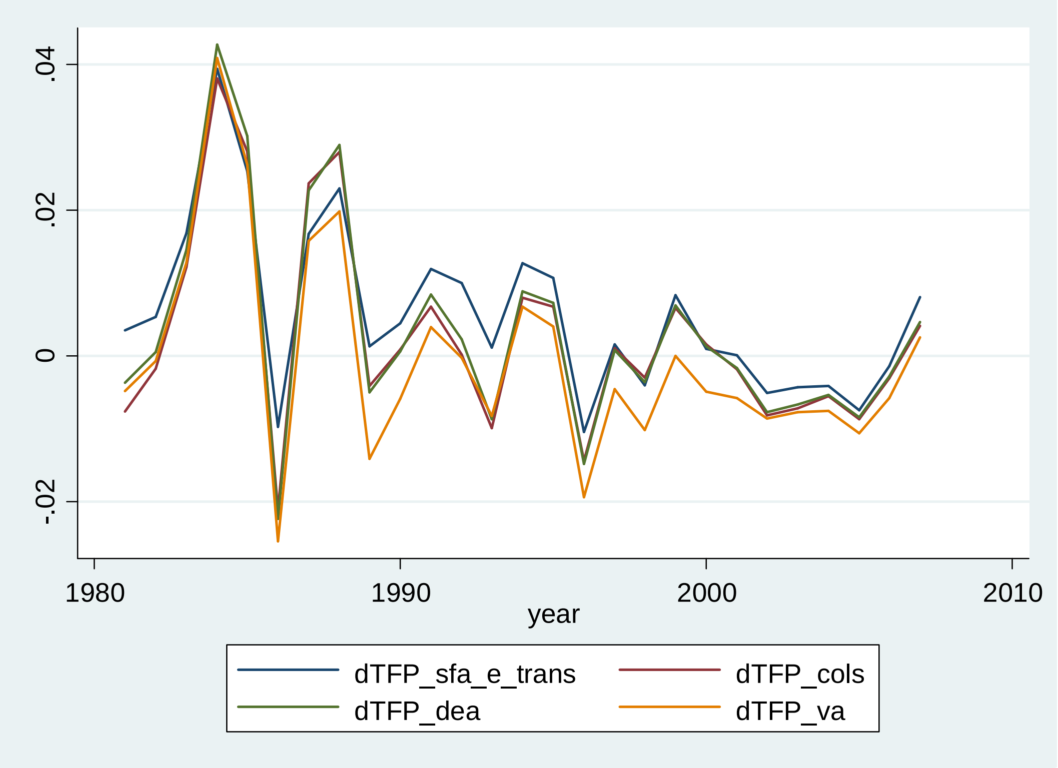
TFP measure	Average	Standard Deviation	Minimum (%)	Minimum (year)	Maximum (%)	Maximum (year)
DEA	0.9%	2.3%	-3.9%	1997	4.1%	2006
COLS	1.4%	2.2%	-3.4%	1997	4.3%	2006
SFA translog (exponential)	0.1%	2.0%	-3.3%	1998	3.2%	2005
GA	0.6%	2.5%	-4.6%	1997	4.1%	2006

Denmark

	DEA	COLS	SFA translog (exponential)	GA
DEA	1			
COLS	0.94	1		
SFA translog (exponential)	0.90	0.91	1	
GA	0.95	0.91	0.92	1

TFP measure	Average	Standard Deviation	Minimum (%)	Minimum (year)	Maximum (%)	Maximum (year)
DEA	1.0%	1.2%	-1.3%	1998	4.3%	1994
COLS	0.8%	1.2%	-1.3%	2001	4.0%	1994
SFA translog (exponential)	1.2%	0.8%	-0.4%	2007	2.9%	1994
GA	0.3%	1.2%	-2.2%	1998	3.3%	1994

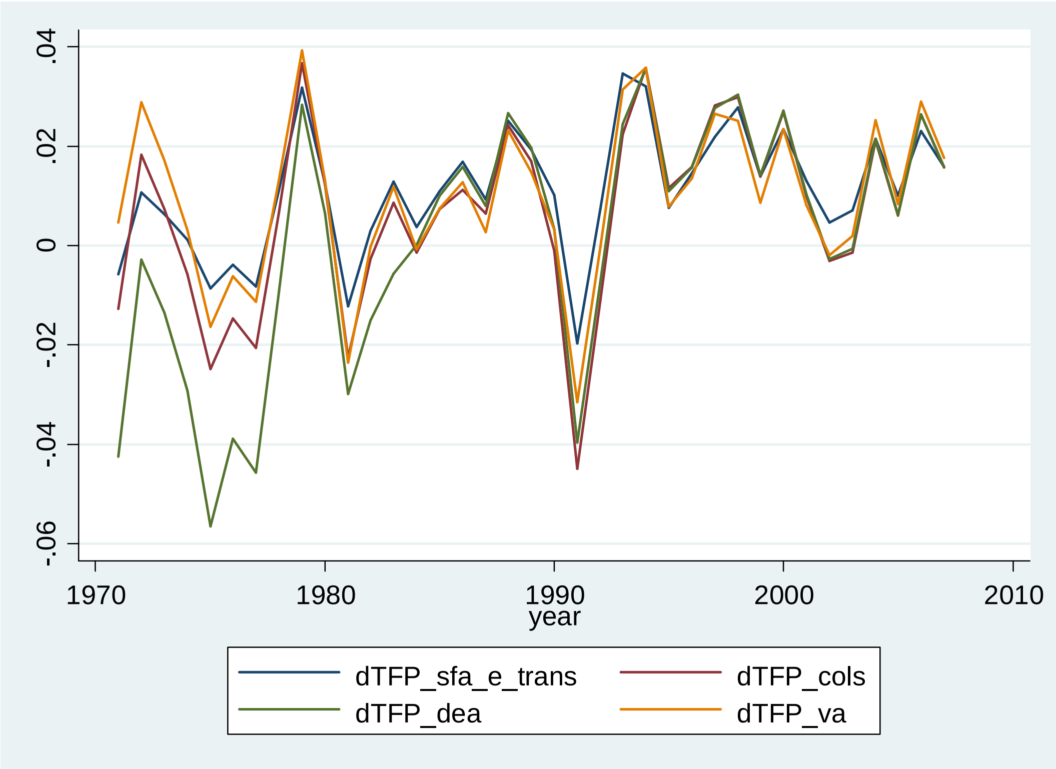
Spain



	DEA	COLS	SFA translog (exponential)	GA
DEA	1			
COLS	1.00	1		
SFA translog (exponential)	0.97	0.95	1	
GA	0.98	0.97	0.97	1

TFP measure	Average	Standard Deviation	Minimum (%)	Minimum (year)	Maximum (%)	Maximum (year)
DEA	0.3%	1.4%	-2.2%	1986	4.3%	1984
COLS	0.3%	1.4%	-2.1%	1986	3.8%	1984
SFA translog (exponential)	0.6%	1.2%	-1.0%	1996	3.9%	1984
GA	0.0%	1.4%	-2.5%	1986	4.1%	1984

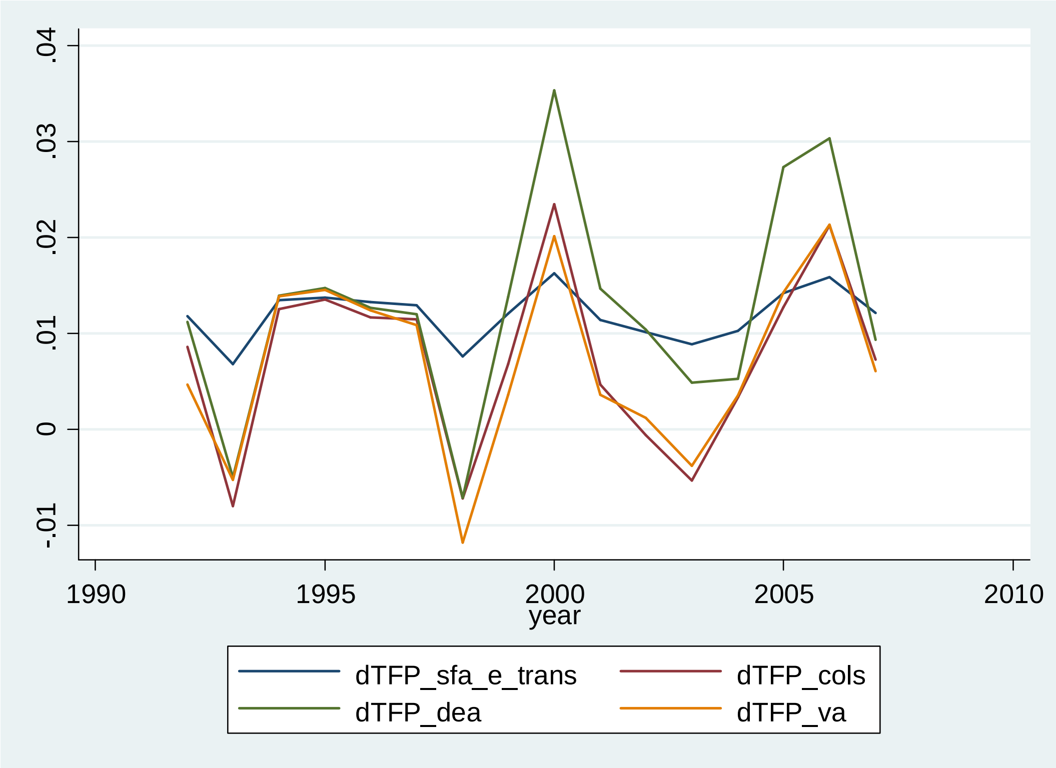
Finland



	DEA	COLS	SFA translog (exponential)	GA
DEA	1			
COLS	0.91	1		
SFA translog (exponential)	0.92	0.95	1	
GA	0.81	0.96	0.93	1

TFP measure	Average	Standard Deviation	Minimum (%)	Minimum (year)	Maximum (%)	Maximum (year)
DEA	0.1%	2.4%	-5.7%	1975	3.6%	1994
COLS	0.7%	1.8%	-4.5%	1975	3.7%	1994
SFA translog (exponential)	1.1%	1.3%	-2.0%	1975	3.5%	1994
GA	1.0%	1.6%	-3.1%	1975	3.9%	1994

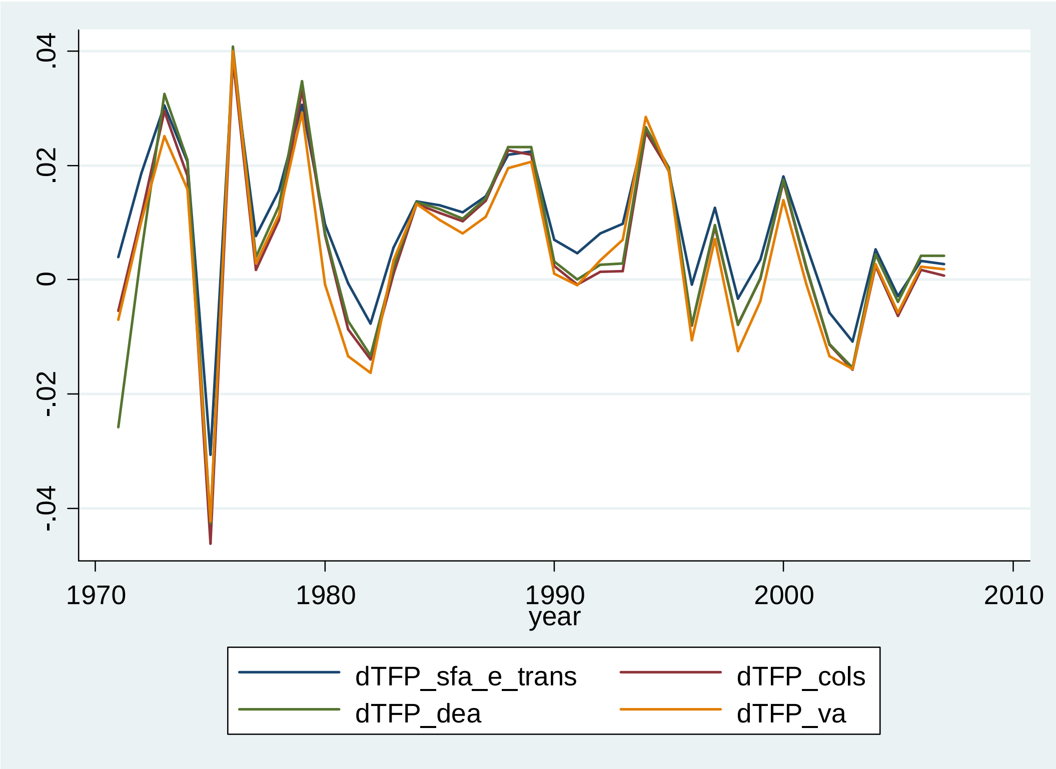
Germany



	DEA	COLS	SFA translog (exponential)	GA
DEA	1			
COLS	0.90	1		
SFA translog (exponential)	0.93	0.99	1	
GA	0.89	0.97	0.97	1

TFP measure	Average	Standard Deviation	Minimum (%)	Minimum (year)	Maximum (%)	Maximum (year)
DEA	1.3%	1.1%	-0.7%	1998	3.5%	2000
COLS	0.7%	0.9%	-0.8%	1993	2.3%	2000
SFA translog (exponential)	1.2%	0.3%	0.7%	1993	1.6%	2000
GA	0.7%	0.9%	-1.2%	1998	2.1%	2006

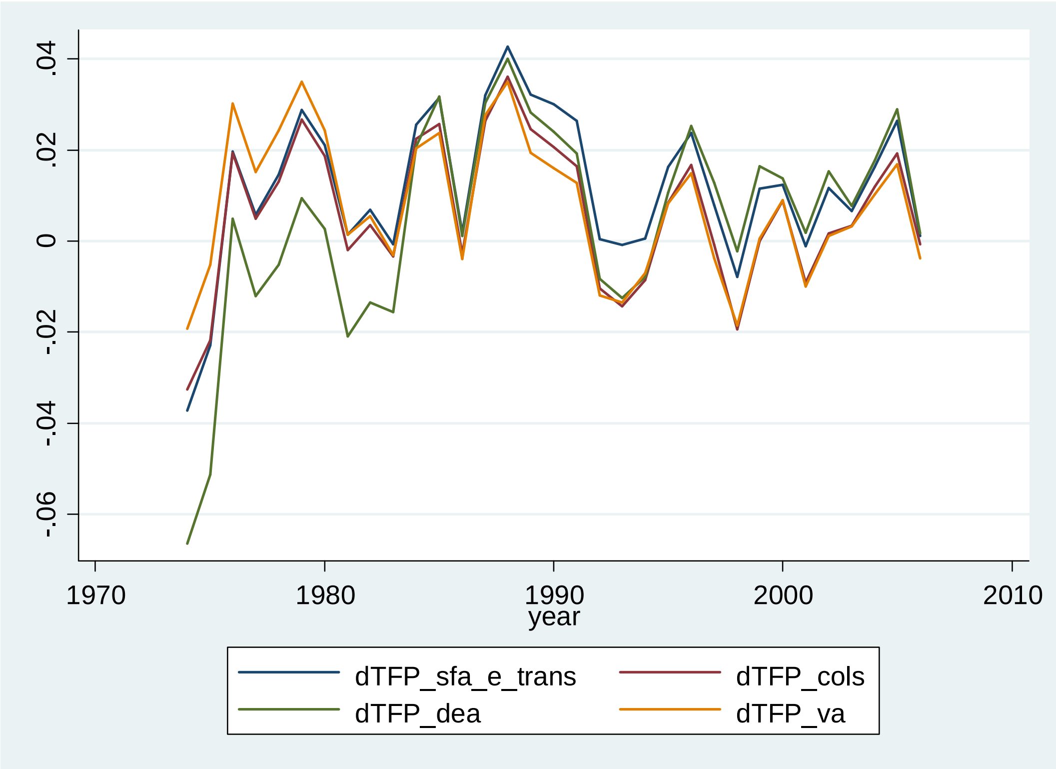
Italy



	DEA	COLS	SFA translog (exponential)	GA
DEA	1			
COLS	0.97	1		
SFA translog (exponential)	0.95	0.99	1	
GA	0.96	0.98	0.98	1

TFP measure	Average	Standard Deviation	Minimum (%)	Minimum (year)	Maximum (%)	Maximum (year)
DEA	0.6%	1.7%	-4.3%	1975	4.1%	1976
COLS	0.6%	1.6%	-4.6%	1975	3.9%	1976
SFA translog (exponential)	0.9%	1.3%	-3.1%	1975	3.8%	1976
GA	0.4%	1.6%	-4.2%	1975	4.0%	1976

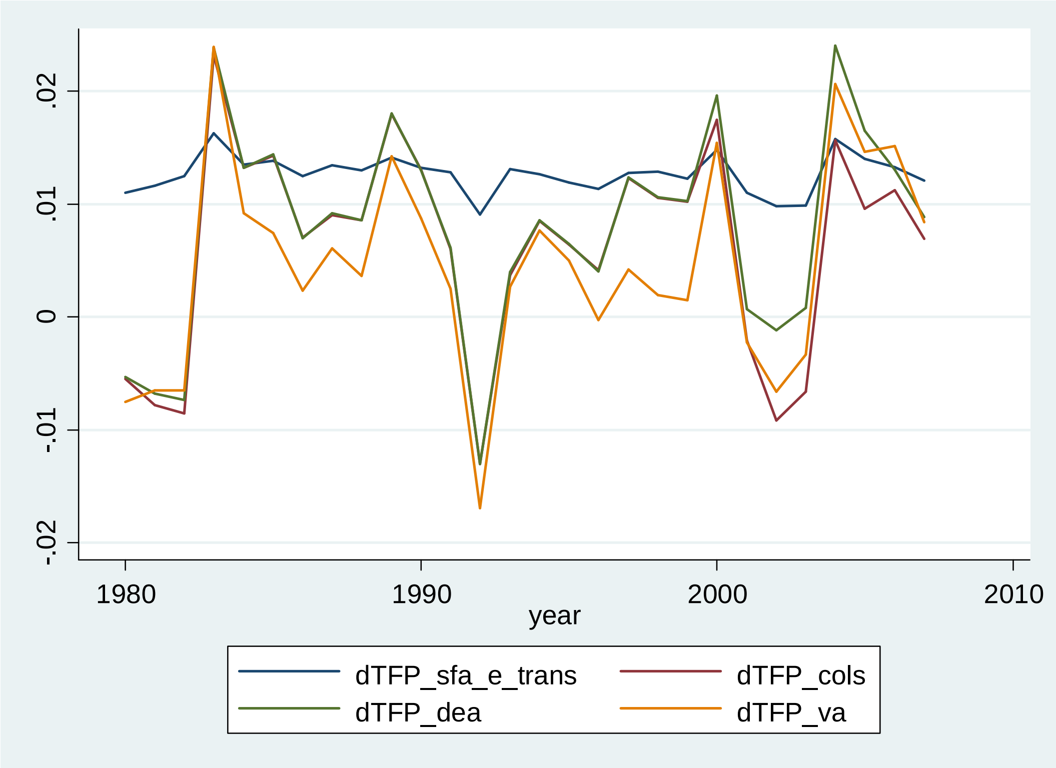
Japan



	DEA	COLS	SFA translog (exponential)	GA
DEA	1			
COLS	0.82	1		
SFA translog (exponential)	0.91	0.97	1	
GA	0.61	0.94	0.85	1

TFP measure	Average	Standard Deviation	Minimum (%)	Minimum (year)	Maximum (%)	Maximum (year)
DEA	0.5%	2.3%	-6.7%	1974	4.0%	1988
COLS	0.6%	1.6%	-3.3%	1974	3.6%	1988
SFA translog (exponential)	1.2%	1.7%	-3.7%	1974	4.3%	1988
GA	0.8%	1.5%	-1.9%	1974	3.5%	1988

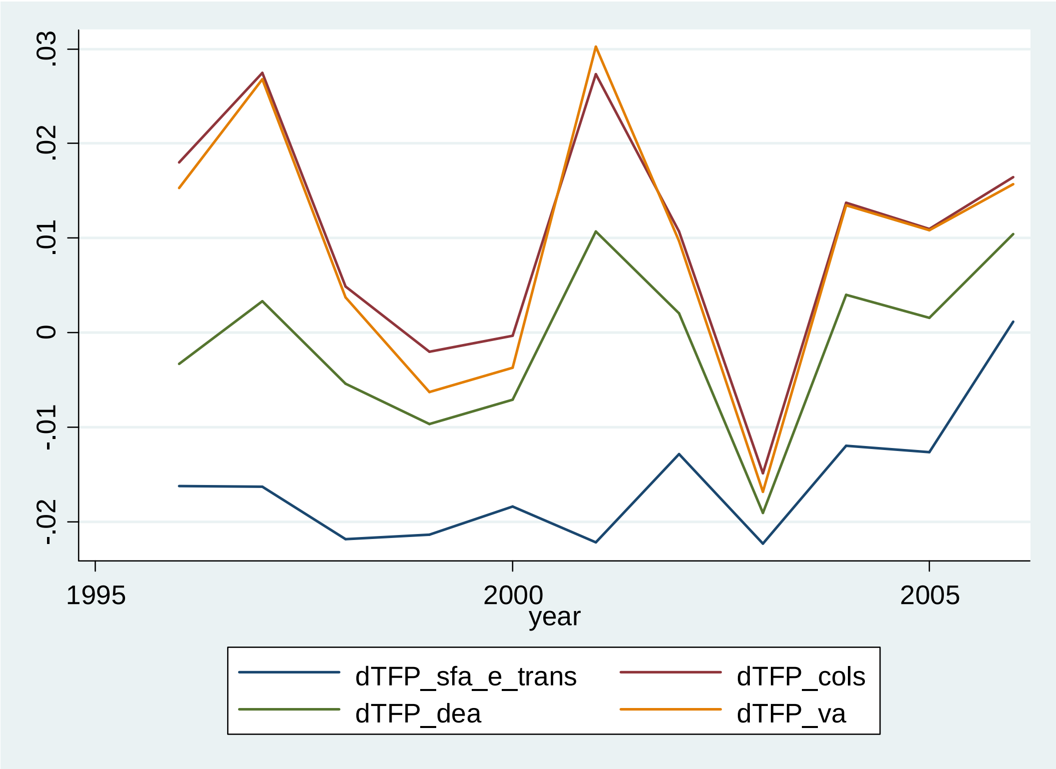
Netherlands



	DEA	COLS	SFA translog (exponential)	GA
DEA	1			
COLS	0.96	1		
SFA translog (exponential)	0.87	0.87	1	
GA	0.96	0.91	0.90	1

TFP measure	Average	Standard Deviation	Minimum (%)	Minimum (year)	Maximum (%)	Maximum (year)
DEA	0.8%	0.9%	-1.3%	1992	2.4%	2004
COLS	0.6%	0.9%	-1.3%	1992	2.3%	1983
SFA translog (exponential)	1.3%	0.2%	0.9%	1992	1.6%	1983
GA	0.4%	0.9%	-1.7%	1992	2.4%	1983

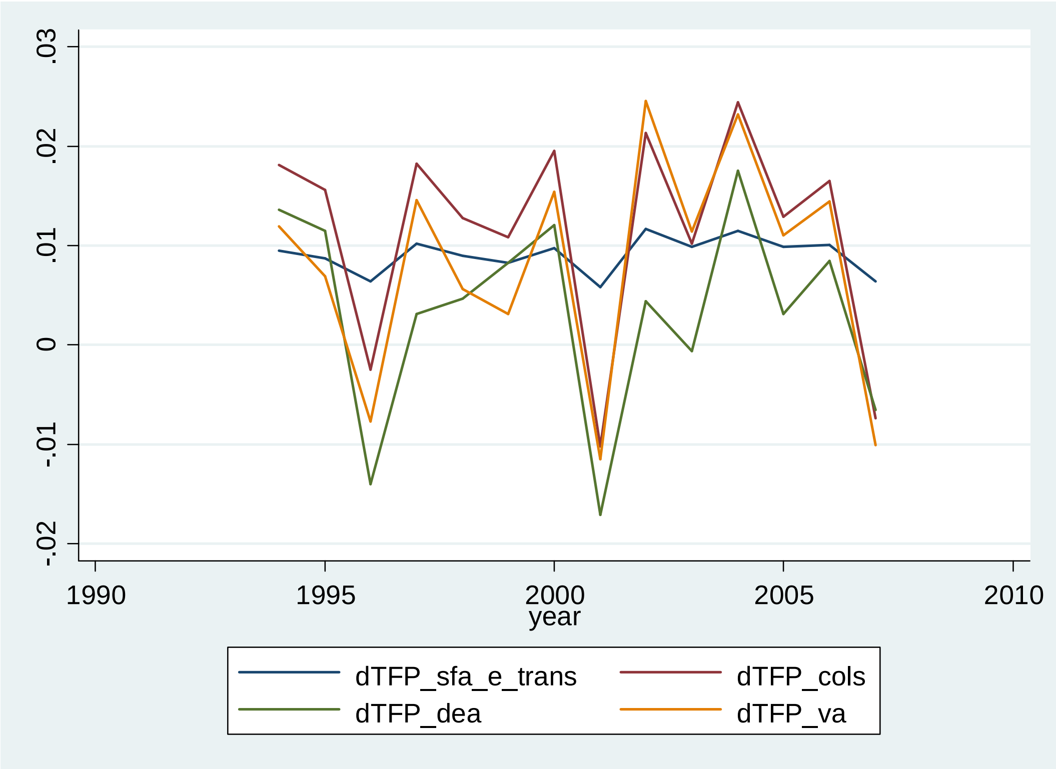
Slovenia



	DEA	COLS	SFA translog (exponential)	GA
DEA	1			
COLS	0.88	1		
SFA translog (exponential)	0.59	0.33	1	
GA	0.90	0.99	0.32	1

TFP measure	Average	Standard Deviation	Minimum (%)	Minimum (year)	Maximum (%)	Maximum (year)
DEA	-0.1%	0.9%	-1.9%	2003	1.1%	2001
COLS	1.0%	1.3%	-1.5%	2003	2.7%	1997
SFA translog (exponential)	-1.6%	0.7%	-2.2%	2003	0.1%	2006
GA	0.9%	1.4%	-1.7%	2003	3.0%	2001

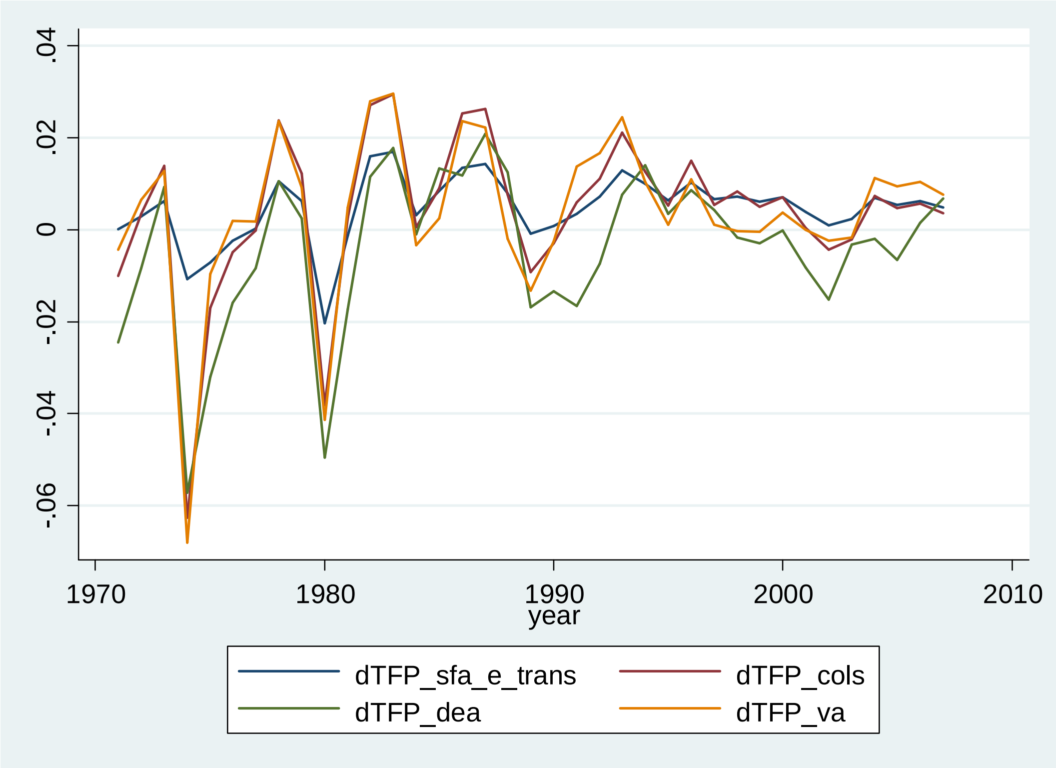
Sweden



	DEA	COLS	SFA translog (exponential)	GA
DEA	1			
COLS	0.90	1		
SFA translog (exponential)	0.76	0.94	1	
GA	0.78	0.95	0.99	1

TFP measure	Average	Standard Deviation	Minimum (%)	Minimum (year)	Maximum (%)	Maximum (year)
DEA	0.3%	1.0%	-1.7%	2001	1.8%	2004
COLS	1.1%	1.1%	-1.0%	2001	2.4%	2004
SFA translog (exponential)	0.9%	0.2%	0.6%	2001	1.2%	2002
GA	0.8%	1.1%	-1.1%	2001	2.5%	2002

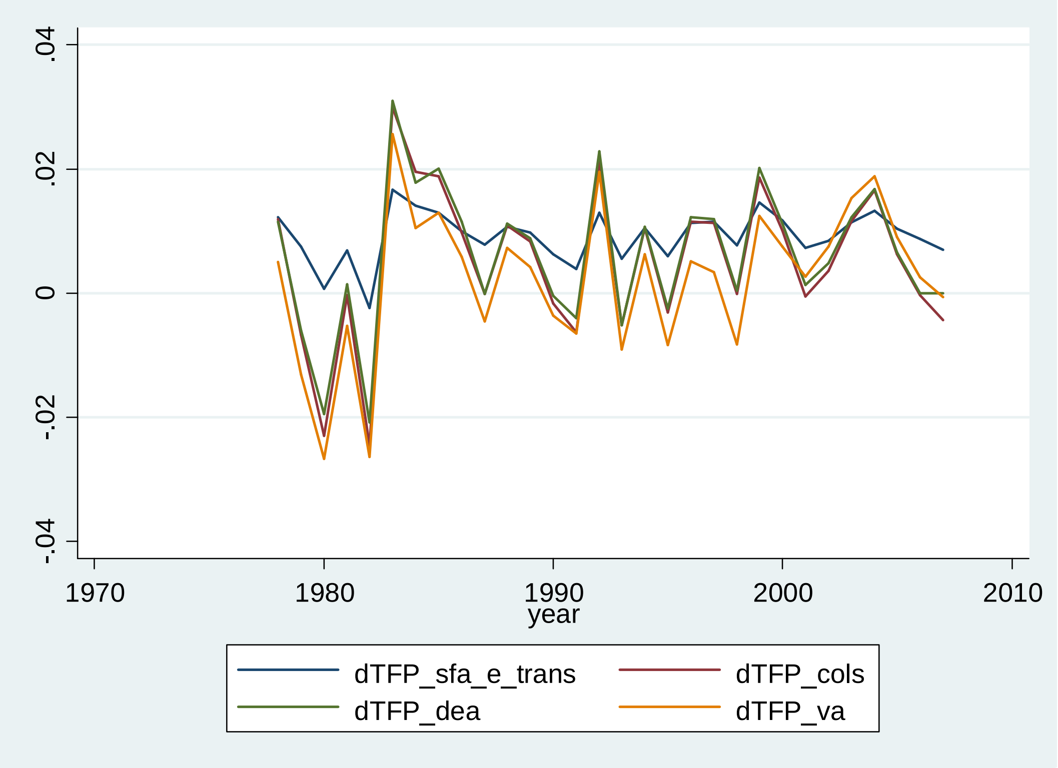
United Kingdom



	DEA	COLS	SFA translog (exponential)	GA
DEA	1			
COLS	0.92	1		
SFA translog (exponential)	0.92	0.92	1	
GA	0.83	0.97	0.85	1

TFP measure	Average	Standard Deviation	Minimum (%)	Minimum (year)	Maximum (%)	Maximum (year)
DEA	-0.4%	1.7%	-5.7%	1974	2.1%	1987
COLS	0.4%	1.7%	-6.3%	1974	3.0%	1983
SFA translog (exponential)	0.5%	0.7%	-2.0%	1974	1.7%	1983
GA	0.4%	1.8%	-5.7%	1980	3.0%	1983

United States of America



	DEA	COLS	SFA translog (exponential)	GA
DEA	1			
COLS	1.00	1		
SFA translog (exponential)	0.97	0.98	1	
GA	0.95	0.95	0.92	1

TFP measure	Average	Standard Deviation	Minimum (%)	Minimum (year)	Maximum (%)	Maximum (year)
DEA	0.6%	1.2%	-2.1%	1982	3.1%	1983
COLS	0.5%	1.2%	-2.5%	1982	3.0%	1983
SFA translog (exponential)	0.9%	0.4%	-0.2%	1982	1.7%	1983
GA	0.2%	1.2%	-2.7%	1980	2.6%	1983