Università degli Studi Di Roma, Tor Vergata

FACOLTÀ DI INGEGNERIA
DIPARTIMENTO DI INGEGNERIA DELL'IMPRESA
Dottorato di Ricerca in "Ingegneria Economico-Gestionale"
XX ciclo - 2007

An Improved Procedure for Bootstrapping Malmquist Indices and its applications on the Regional Productivity Growth

Advisors:

Prof. Paolo Mancuso
Prof. Léopold Simar

Ph.D. Candidate:

Claudia Curi

Author's address:

Claudia Curi

Università degli Studi Roma "Tor Vergata"
Dipartimento di Ingegneria dell'Impresa
Via del Politecnico, 1
00133 Roma Italy
curi@disp.uniroma2.it
claudia.curi@gmail.com

Shosholoza
Ku lezontaba
Stimela siphum' eSouth Africa
Wen' uyabaleka
Wen' uyabaleka
Ku lezontaba
Stimela siphum' eSouth Africa

*(**Shosholoza**, a traditional Southern African folk song)*

To my Grandfather Giovanni

Contents

List of Figures

List of Tables

Acknowledgements

I am in the habit of writing the acknowledgements at the end of writing my thesis: it is wrong! If any value is to be recognized to this work, much of it is due to the endless support that I received from many people.

First of all, I would like to thank my doctoral advisor Prof. Paolo Mancuso for accepting me as her Ph.D. student (the first one!!!) and introducing me with shyness to the wonderful field of the research, but mainly of the Productivity and Efficiency Analysis. I must express my appreciation for his "open mind" thinking, able to stimulate my scientific interests everyday and to promote my intellectual traveling and networking as the basis of my research and professional career. The shyness by Paolo introduced me to this field, today it is for me a real honor!!! A honor to work with special people in a special environment.

Indeed, I met Prof. Léopold Simar and very soon I started working with him and his assistant, Cinzia Daraio. It was a real pleasure and a unique experience; I must thank him for opening the door and plainly introducing me to this field, with his enthusiasm and optimism of everyday, even when I left mine. I must express my appreciation for his unceasing efforts at providing an academically inspiring and prolific environment.

I owe a special note of gratitude to Cinzia for her continuous academic support during my most difficult periods in Louvain la Neuve and for encouraging me to achieve always the best!

I would like to thank also Prof. Domenico Campisi for adding managerial skill to my professional growth.

I am very grateful to Bank of Italy; in particular to Dr. Bronzini and Dr. Piselli for sharing their knowledge in the regional productivity analysis and providing the database

on the Italian regions.

Then all my family! Yes, them for their everytime and everywhere support, advice and kindness without which this work would have never been finished , nor even started.

I want to thank all my friends to feel with me and trust me, even when I did not; and a special thank to Carla for jointly enjoying both the much darkness days and the most successful days.

Finally, I would like to thank all my new friends of Louvain-la-Neuve for their help and making wonderful days, even if outside was raining and my best Ph.D. colleagues for sharing with me most of my time spent to our common second house, the Ph.D. Office!

Abstract

Improving the Fare et al. (1992) approach on Malmquist index of productivity, which can be decomposed into indices describing changes in technology and changes in efficiency, Simar and Wilson (1999) provided a statistical interpretation to their Malmquist productivity index and its components, and presented a bootstrap algorithm to estimate confidence intervals for the indices.

Extending the recent developments introduced by Simar and Wilson (2007) in the bandwidth specification in the univariate case, we propose new methods of density estimation, based on more accurate bandwidth specification. Monte Carlo experiments has been computed for the first time in this context. They have shown a low quality of performance of the Simar and Wilson (1999)'s bootstrap approximations, and high level of quality for the proposed methods. In particular, they have found out as best performer method the procedure based on the density estimation without considering the ones, revealing the severe problem of deteriorating the estimation of the continuous density of the efficiency scores. Moreover, data driven methods have been applied to the Malmquist Index framework and at this stage of research they have shown different results from those provided by Simar and Wilson (1999).

From an empirical point of view, Total Factor Productivity (TFP) growth of the Italian regions over the period 1980-2001 has been analyzed. Malmquist Productivity Index (MPI) and its components (namely Efficiency Change and Technical Change) as well as confidence intervals have been estimated by applying the best performed procedure, previously determinated. Including human capital among inputs, we estimated an overall bias-corrected productivity gain of 2.1 percent, an efficiency gain of 0.5 and a technical gain of 1.6 percent. The bootstrap analysis revealed that for most Italian regions efficiency and technical changes did not show a statistically significant change. According to these results, the inferential approach has provided a more rigorous and accurate insights on the Italian regional TFP than the traditional Data Envelopment Analysis (DEA) estimation carried out by Leonida et al.(2004 ,Table 1, pg. 2190) in

which all the estimated values are interpreted as progress or regress without taking into account the bias of the estimated values and their statistical significance.

Chapter 1

Introduction

1.1 Motivation

Measures of productivity growth constitute core indicators for the analysis of economic growth. Productivity is commonly defined as a ratio of a volume measure of output to a volume of input use. While there is no disagreement on this general notion, a deeper analysis of the problem of productivity measurement reveals very quickly that the objectives of any productivity measurement include many concepts which can not be neglect, such as the *Technology*, that describes as the currently known ways of converting resources into outputs desired by the economy, *Efficiency*, that describes a production process compared with the maximum amount of output that is physically achievable with current technology, *Real cost savings* that captures changes in capacity utilisation, learning-by-doing and measurement errors of all kinds. In this context, there are many different productivity measure and the choice between them depends on the purpose of productivity measurement. Nonetheless, the Malmquist Productivity Indices are the most popular statistical tool since its appealing nature. Their potential richness makes it a simple tool to measure the Total Factor Productivity as it allows to capture the most sources of its variations, by means of its decomposition, provides an easy economic interpretability and rely on several simplifying assumptions.

It is as appealing as it is discussed by the scientific community. The debate on Malmquist Indices arise as soon as it came in the productivity research field and, basically, it has concerned the issues related to its decomposition. Few years ago, Simar and

Wilson widened the debate focusing on more accurate estimation based on the statistical inference, overcoming the most important drawbacks of the estimate based on the Data Envelopment Analysis (DEA). They contributed to achieve a consistent and unbiased productivity measurement. Despite the positive consent received by the community, so far, it is possible to point out a huge gap in literature between this two research avenues. Actually, the debate is still bordered to the decomposition issue as only Simar and Wilson are handling with the statistical issues.

My research belongs to the Simar and Wilson's stream and aims to take part and enrich the debate on the statistical issue on the Malmquist Indices. Simar and Wilson proposed a consistent bootstrap estimation procedure for obtaining confidence intervals for Malmquist Index and its components: economical speaking, this method can be used to determine whether indicated changes in productivity, efficiency or technology are really determined by the analyzed process or merely artifact, due to the sampling variation. Therefore, this method has high relevance on the empirical analysis since it avoids a distort measurement of the Total Factor Productivity, otherwise achievable by means of the simple decomposition.

Although its importance, so far, it is the only tool to carry out the statistical inference of the Malmquist Indices and its components. Moreover, its quality of the performance, defined as the achieved rate coverages of the confidence intervals regards with the real efficiency scores, has never been established. This might explain why the bootstrap procedure has so rarely used in applied works.

1.2 Scope and Contributions of the Research

The main scope of this dissertation is twofold. Firstly, the theoretical aim is to improve the statistical inference on the Malmquist Indices, introducing new methods of efficiency density estimate; secondly, the empirical aim is to investigate the Italian regional productivity over the period 1980-2001, by means of the best improved method, determinated in the theoretical phase of my research.

Deeply, from a theoretical point of view, the aim is to investigate the quality of the performance of the original bootstrap procedure, based on the main drawback of

using the "normal rule of thumb" for selecting the bandwidth in the efficiency density estimate. Monte Carlo experiments will be carried out to statistically test the power of the procedure. Secondly, we would like to suggest some improvements of the statistical inference on the Malmquist Indices, attacking the main problem shown by any efficiency analysis, here more severe as the multivariate context: the presence of spurious mass at one. Indeed, it damages the estimation of the continuous density of the efficiency score. So, we will adapt the new developments proposed by Simar and Wilson in the univariate case to the bivariate case. Moreover, we will prove the performance of each new development using the Monte Carlo experiments, employing the same data generating process used for testing the Simar and Wilson's procedure, in order to compare the obtained results. Finally, w develop some methodologies to provide data-driven methods for selecting the bandwidth.

From an empirical point of view, we investigate the economic performance of Italian regions using the best theoretical method, previously identified. The use of Malmquist Indices allow us the decomposition of the productivity growth into technological progress and technical efficiency change as well as the analysis of the influence of these productivity growth sources on the observed gap across Italian regions. Moreover, the use of the best "performed" method previously determinate, allows for pointing out the accuracy and the higher precision given by our sensitivity analysis, by comparing the results to those of a research, carried out by Leonida et al., based on the traditional non parametric estimate (lacking in statistical inference on the indices).

1.3 Organization

This dissertation comprises two main parts. The first part deals with the problem of efficiency density estimate in the context of statistical inference on Malmquist Index whereas the second part deals with the problem of the Italian regional productivity, treated using the methodology described in the previous part. In particular, in Chapter 2 we present a complete overview for the productivity measurement concept, taking into account the most advanced theory. In Chapter 3, we introduce the Malmquist Indices, discussing from its origin to its decomposition, passing through the well-known

debate; in Chapter 4, we deal with the inferential aspect of the Malmquist Indices, highlighting the main drawback of the only procedure so far available. In Chapter 5, we propose new methods for making inference and we propose a Monte Carlo scenario for establishing the quality of the performance. Moreover, for the first time, we test the performance of the three methods. In Chapter 6, an adaptation of the data driven methods is shown as well as the need to further investigation, provided by the different (and so, good) results by comparing the "old" inferential results with the new ones. Chapter 7 conclude the dissertation dealing with the economic problem of the Italian regional productivity analysis; it is also an empirical proof of the goodness of the best theoretical method, previously found.

Chapter 2

Productivity and Efficiency Measurement Concepts

2.1 The Frontier Model: Technology and Distance Measures

In this chapter, the main concepts rely on the productivity and efficiency analysis are described. It is very important for the aim of this thesis to clarify the different between these two terms, *productivity* and *efficiency*, since they are extensively used in the forthcoming sections and are often employed interchangeably and mixed up, also by the scientific communities. On the contrary, they have different meanings, even though in the efficiency literature many authors do not make any difference between productivity and efficiency (*e.g.* Sengupta, 1995 [73]; Cooper *et al.*, 2000 [18]). According to a classic definition (see e.g. Vincent 1968[88]) *productivity* is the ratio between an output and the factors that made it possible. In the same way, Lovell (1993)[52] defines the productivity of a production unit as the ratio of its output to its input. This ratio is easy to compute if the unit uses a single input to produce a single output. On the contrary, if the production unit uses several inputs to produce several outputs, then the inputs and outputs have to be aggregated so that productivity remains the ratio of two scalars. This implies that *productivity* is a measure concerning only the analyzed unit, without proving any information of it comparing with the remaining. Hence, the measure of

productivity assumes a ratio form and may be distinguish in:

- *Partial Factor Productivity*: it associates output to a sole input factor (*e.g.* either capital or labor)
- *Total Factor Productivity*: it associates output to all input factors.

Moving from Partial to Total Factor Productivity measures by combining all inputs and all outputs to obtain a single ratio helps to avoid imputing gains to one factor (or one output) that are really attributable to some other input (or output). However, an attempt to move from the Partial to the Total Factor Productivity measures encounters difficulties such as choosing the inputs and outputs to be considered and the weights to be used in order to obtain an aggregate index of output and input,respectively.

The definition of *efficiency* or *productivity efficiency* is dated to the works proposed by Koopmans (1951)[72], Debreu (1951)[22], and Farrell (1957)[30], who defined the concept of *productive efficiency* within an economic environment composed by a cluster of observed units (*e.g.* firms, countries,...). Generally speaking, it is defined in relative terms as distance between the two measures, respectively, the quantity of input and output of one unit and the quantity of input and output, defined by the best possible achievable frontier (on turn, composed by the best observations belong to the cluster). In details, any author provided a different definition based on the same concept, definition that has been evolving over time. For a complete overview, it follows a short summary:

- 1951-Koopmans defines that an input-output vector is technically efficient if, and only if, increasing any output or decreasing any input is possible only by decreasing some other output or increasing some other input;
- 1957-Farrell and later 1985-Charnes *et al.*[16] retrieve the Koopmans definition of technical efficiency as a relative notion, providing a way of differentiating efficient from inefficient production units, but neglecting either the degree of inefficiency of an inefficient vector or the identification of an efficient vector or combination of efficient vectors against which comparing an inefficient vector.
- 1951-Debreu provides the notion of *productive efficiency* in terms of the coefficient of resource utilization. This coefficient is computed by means of the radial mea-

sure which allows for measuring maximum feasible equiproportionate reduction in all variable inputs, or the maximum feasible equiproportionate expansion of all outputs, differently from Koopmans. This criteron of computing the measures allow the indipendence from the unit of measurement (*scale invariant*). One important remark is that the radial measures may suggest the presence of technical efficiency for an observed unit, even though there may remain slacks in inputs or outputs. This remark points out the difference from the *productive efficiency* and the concept of *Pareto optimality*. Indeed, an allocation is Pareto efficient or Pareto optimal when any unit can move from one allocation to another that can make at least one unit better off without making any other unit worse off. This implies that one unit can be productive efficiency but is not Pareto optimal since it has the opportunity to increase in output or decrease in inputs without modifying the allocative configuration of the remain units.

Moreover, Farrell(1957)[30] extended the notion of *productive efficiency* toward a more general notion in oder to define an overall economic productive efficiency able to take into account the more realist behaviors of the units in the market, incorporating the information on prices. Indeed, he defined the new concept of productivity efficiency trying to reflect the ability of producers to select the right technically efficient input-output vector in light of prevailing input and output prices. This led to define the overall productive efficiency as the product of *technical* and *allocative efficiency*. This new definition allows for modeling the efficiency measure under the typical goal of the producers of selecting the best mix (that permits to achiest the highest level of technical efficiency) of input (*e.g.* labour and capital) and output (*e.g.* value added), minimizing the cost function of the inputs in the competitive markets. Further decompositions of the efficiency measure are followed in the literature with the aim of investigating the sources of its generation. Here, a brief summary of the main definition used in the scientific literature:

- *Technical Efficiency*: it measures a firms success in producing maximum output from a given set of inputs;
- *Allocative Efficiency*: it measures a firms success in choosing an optimal set of

inputs with a given set of input prices;

- *Scale Efficiency*: it is defined as the potential productivity gain from achieving optimal size of firm;

- *Structural Efficiency*: it measures in what extent an industry keeps up with the performance of its own best practice firms. Its computation (Bjurek, Hjalmarsson and Forsund, 1990)[12] derives from the aggregation of the Technical Efficiency measures of the firms belong to the analyzed cluster in an average unit and then estimating the individual measure of the technical efficiency for this average unit.

- textitEconomic Efficiency: it measures the ratio of minimum cost to observed cost.

Currently, the main purpose of productivity analysis studies is to evaluate the performance numerically and, hence, make benchmark among the observed units from the point of view of *technical efficiency* . From an analytical point of view, any productivity and efficiency analysis may face the following issues:

- Economic Problem Definition: numerically determination of the economic performance of each observed unit, according to the "Activity Analysis" (*What I want to measure ?*);

- Data Selection: identification and gathering of a cross-sectional data (i.e. panel of unit observed over a limited period of time) (*Which are the available data that best describe the analyzed production process ?*)

- Methodology Definition: choice of the estimation procedure which allow for the definition of the boundary of the production set (*Which approach I want to use to estimate the productivity and efficiency ?*)

- Economic Interpretation of the Results: Interpretation of the mathematical measurement in economic terms (*What can I learn from the results ?*)

The main focus in this section is to provide useful definitions and axioms which the economic theory is based on, needed in the forthcoming sections.

A convenient way to describe a multi-output and multi-input production technology is to use the technology set, Ψ that characterized the analyzed production process.

Consider a sample of N production units. Let $x \in \mathbb{R}_+^p$ the vector of inputs and $y \in \mathbb{R}_+^q$ the vectors of output. The technology set Ψ is the production possibilities set defined as:

$$\Psi = \{(x, y) \mid x \text{ can produce } y\} \tag{2.1}$$

This set consists of all input-output vector such that x can produce y. For the efficiency measurement, the interest is concentrated on the upper boundary of the set Ψ. The efficient boundary (frontier) of Ψ is the locus of optimal production plans (e.g., minimal achievable input level for a given output, or maximal achievable output given the level of the inputs), and is defined as:

$$\vartheta\Psi = \{(x, y) \in \Psi \mid (\theta x, y) \notin \Psi, \forall 0 < \theta < 1, \quad (x, \lambda y) \notin \Psi, \forall \lambda > 1\} \tag{2.2}$$

It is sometimes referred to as the *technology* or the *production frontier*, and is given by the intersection of Ψ and the closure of its complement. Economically speaking, it describes the technical relationship between the inputs and the outputs of a production process. In particular, firms that are technically inefficient operate at points in the interior of Ψ, while those that are technically efficient operate somewhere along the technology defined by $\vartheta\Psi$. Sometimes, it is useful to define the set Ψ in terms of its sections. The input requirement set is defined for all $y \in \mathbb{R}_+^q$ by:

$$X(y) = \left\{x \in \mathbb{R}_+^p \mid (x, y) \in \Psi\right\} \tag{2.3}$$

and represents all the inputs vectors that can produce the output vector $y \in \mathbb{R}_+^q$. On the other hand, the output requirement set is defined for all $x \in \mathbb{R}_+^p$ by:

$$Y(x) = \left\{y \in \mathbb{R}_+^q \mid (x, y) \in \Psi\right\} \tag{2.4}$$

It follows that the efficiency boundary input and output oriented are defined,respectively, as:

$$\vartheta X(y) = \{x \mid x \in X(y), \theta x \notin X(y), \forall 0 < \theta < 1\} \tag{2.5}$$

and

$$\vartheta Y(x) = \{y \mid y \in Y(x), \lambda y \notin Y(x), \forall \lambda > 1\} \tag{2.6}$$

where θ and λ are, respectively, the Debreu-Farrell input and output measure of efficiency for a given point (x, y). They are radial measures of the distance between the point (x, y) and the frontier $\vartheta X(y)$ and $\vartheta Y(x)$. Before going on with the description on the Debreu-Farrell measures, I would like to add a further characterization of the shape of the frontier, certainly essential for a complete picture of the *production frontier* concept.

Any production function Ψ is a graphical representation of the technology and hence embraced the technical property of the production process, that economically speaking are referred as Return to Scale (RTS). According to a standard definition in economics, RTS express the relation between a proportional change in inputs to a productive process and the resulting proportional change in output. If an x per cent rise in all inputs produces an x per cent increase in output, there are constant returns to scale (CRS). If output rises by a larger percentage than inputs, there are increasing returns to scale (IRS). If output rises by a smaller percentage than inputs, there are decreasing returns to scale (DRS). There are many reasons why a particular firm may possess certain RTS properties but the most common shape of the production frontier is the Variable Returns to scale (VRS) characterized by the (a) IRC, (b) DRS and (c) CRS. The economic interpretation of ICR concerns to the possibility of a firm, *e.g.*, to gain more efficiency by having additional staff specialised in particular tasks. Instead, one possible situation of DRC is when a firm has become so large that management is not able to exercise a close control over the all aspects of the production process so that it should be better downsize the firm. From a benchmarking standpoint, the CRS frontier allows small firms to be compared to big firms, and vice versa, while the VRS frontier will tend to only compare firms of similar sizes. It highlights this point on the nature of the technology since whether the underlying technology exhibits increasing, constant or decreasing return to scale is a crucial question in any study of productive efficiency. Indeed many empirical studies still face this preliminary question just comparing the alternative assumptions. In these cases a priori imposed restrictions of the characteristics

of the production frontier may seriously distort measures of the efficiency, in particular when CRT are imposed and the true technology display non-constant return to scale. This problem can be overcome following the approach proposed by Simar and Wilson (2002)[81]: a bootstrap procedure for testing hypothesis regarding retuns to scale.

Turning back to the distance functions, there are two direction of looking at the efficient frontier: either the input direction (where the efficiency subset is characterized by the $\vartheta X(y)$) or the output direction (where the efficiency subset is characterized by the $\vartheta Y(x)$). Mathematically, the Debreu-Farrell input measure of efficiency for a given point operating at level (x, y) is define as:

$$\begin{aligned} \theta(x,y) &= inf\{\theta \mid \theta x \in X(y)\} \\ &= inf\{\theta \mid \theta x \in \Psi\} \end{aligned} \tag{2.7}$$

whereas the Debreu-Farrell output measure is given by:

$$\begin{aligned} \lambda(x,y) &= sup\{\lambda \mid \lambda y \in Y(x)\} \\ &= sup\{\lambda \mid (x, \lambda y) \in \Psi\} \end{aligned} \tag{2.8}$$

So, the **input efficiency** $\theta(x, y) \leq 1$ measures the radial feasible proportional reduction of inputs that an observed unit should perform to be considered as *input technically efficient*, in the sense that $(\theta x, y)$ belong to the frontier. In the same way, the **output efficiency**$\lambda(x, y) \geq 1$ measures the radial feasible proportional increase in outputs that an observed unit should perform to be considered as *output technically efficient*, in the sense that $(x, \lambda y)$ belong to the frontier. It is worth to note that all efficient units exhibit $\theta = \lambda = 1$, independently from the direction of the distance measurement. Indeed, the frontier is one but it can be describe in terms of its sections $X(y)$ and $Y(x)$. The economic interpretation of the distance functions gains more importance if we consider the fact that it allows to measure the efficiency looking only at the physical aspects of production (input and output of the analyzed production function). Indeed, so far I have discussed only the production function and the physical aspects but, it is worth to note, that any observed units (*e.g.* firms) assumes an economic behavior in the market. On turn, it considers strategical decisions on the optimal quantity of input and

output under four different sets of behavioral assumptions, that can be summarized in the concept of the *economical duality*:

- Short-Run (SR) cost minimization;
- Long-Run (LR) cost minimization;
- SR profit maximization;
- LR profit maximization

This consideration gets necessary the analysis of these economic aspects but, meanwhile, imposes some difficulties for the mathematically formulation of these behavioral objectives since the lack of a complete information on prices. The introduction of the distance function in this field has allowed to overcome the formulation and computational problem of any productivity and efficiency analysis since it does not require the specification of the behavioral objective (such as the cost minimization or the profit maximization). It powerful is on the possibility to identify "how far is the observed unit form the frontier", departing only from production aspects. More precisely, the **input efficiency** $\lambda(x, y) \leq 1$ measures the radial feasible proportional reduction of inputs that an observed unit should perform to be considered as *input technically efficient*. The **output efficiency** $\theta(x, y) \geq 1$ measures the radial feasible proportional increase in outputs that an observed unit should perform to be considered as *output technically efficient*.

Sometimes, it is easier to compute the radial distance of a given unit using the Shepard distance (1970) that is a normalized measured of Euclidean distance from the point $(x, y) \in \mathbb{R}_+^{p+q}$ to the boundary of Ψ in a radial direction orthogonal to y and is defined, respectively, in the input and output direction as:

$$D_{in}(x, y) = sup\left\{\beta \mid \frac{x}{\beta} \in X(y)\right\} \tag{2.9}$$

$$D_{out}(x, y) = inf\left\{\beta \mid \frac{y}{\beta} \in Y(x)\right\} \tag{2.10}$$

As remarked above, any behavioral assumption are necessary and the direction of measurement depends on the context of applications. Since their powerfuls, the out-

put and input distance have a number of uses in the efficiency analysis. They are so important since both their application and easily estimation by either econometric or mathematical programming methods. As it will discuss later, they will be intensely used in the definition of the Malmquist Productivity Index. Hence, the main concepts for carrying out an efficiency and productivity analysis are the production set, defined by (2.1) and the Shepard (or Farrell)[1], defined by (2.9) or (2.8).

2.2 The Statistical Model: Estimators and DGP

In this section, I introduce the concept of *Statistical Model* in the context of the Frontier Model. The shift from the Frontier Model to the Statistical one is fundamental whether the efficiency analysis has to be meaningful. Indeed, Efficiency (whatever the type) must be estimated, as opposed to being measured or deduced. As a matter of fact, in any analysis of a particular production process the attainable set Ψ is ***unknown*** as well as $X(y), \vartheta X(y), Y(x)$, and $\vartheta Y(x)$. In their turn, the efficiency scores $\theta(x,y)$ and $\lambda(x,y)$ are unknown. The only information available for a researcher is the random sample of production units (*i.e.* the input and output data related to the analyzed production process), given by:

$$\chi_n = \{(x_i, y_i), \quad i = 1, ..., n\} \tag{2.11}$$

In this thesis, I will focus on a particular type of frontier, called **deterministic frontier**,[2] where all observations are assumed to be technically attainable in the sense there are no measurement errors, no noise, *ect.*.

Formally,

$$Prob((x_i, y_i) \in \Psi) = 1 \tag{2.12}$$

Consequently, the problem of efficiency measurement becomes a problem of estimation the unknown quantities listed above using χ_n. So, two questions are posed:

[1] The use of one rather than the other one is based only to needs of simpler mathematical assumption and computation. Indeed an efficiency score less than 1 admits two boundaries at 0 and at 1, that can have negative influences on the estimations

[2] The stochastic frontier approach allows some observed points to be outside the attainable set, but there, only parametric restrictions on the shape of the frontier allows identification of the frontier from the sample. In addition, the stochastic frontier approach requires parametric assumptions regarding the distribution of both the inefficiency process as well as the noise process to recover estimates of firm-specific inefficiencies using cross-sectional data.

- Where do the data come from?

- What about the possibility of new data?

In other words, "How can we can use the information contained in the χ_n to estimate $\theta(x,y)$ and $\lambda(x,y)$, or Ψ hence $X(y), \vartheta X(y), Y(x)$ and $\vartheta Y(x)$?".

This problem has been risen and solved by the theory proposed by Simar and Wilson (1998a [77], 2000a [79], 2006[82]), according to which the efficiency analysis calls for more than a linear programming problem, involving methods of statistical analysis to understand the properties of whatever estimators have been used to obtained estimates of the things of interest. Indeed, their considerations get the placement to provide a statistical nature to the efficiency and productivity analysis, distinguishing between the concept of *estimator* and *estimate* in the efficiency context. It is useful remind that an *estimator* is a random variable, while an *estimate* is a realization of an estimator. In this context, as reported by Simar and Wilson (2006)[82], the efficiency score are estimates while the production set and the requirement set are the estimators.

Simar and Wilson transposed the statistical estimation theory to this context putting their attention on the issues related to consistency and unbiased: after analysing the suitable properties, they constructed a statistical framework which constitutes the reference paradigm from which my thesis has been developed. Summing up briefly, the logical steps to follows in an efficiency analysis are the following:

1. what is estimated (?)

2. how to carry out a consistence and unbiased estimation (?)

So, a statistical model must be defined in order to describe the process that yields the data in the sample χ_n, called Data Generating Process (DGP). As mention before, any efficiency analysis is based on the convergence of many different theories such as the "Activity Analysis Model" (Debreu, 1951 [22]) that imposes the respect of certain properties or axioms to the technology (and so to the production model), the Linear Programming (*e.g.* Charnes *et al.*, (1978) [17]; Banker et al., (1984)[7]), based on the mathematical computation of a problem of production plan and the more recent theories

developed by Simar and Wilson ,(2000a, 2006)[79] [82] and Kneip *et al.*,(1998)[45], Kneip *et al.*, (2003)[46].

Below, I describe the Statistical Model by means of a set of axioms that constitutes the **Data Generating Process**. Their definition is of crucial importance, since any inference that might be made will typically be valid only if the assumptions are in fact true. The first group (1-4) concerns Ψ and are the standard assumptions in the microeconomic theory of the firm, the 5th concerns the statistical nature of the analyzed sample of observed units and, together with the 6th and 7th,constitute the regularity conditions, that assure the sufficient condition for proving the consistency of all the non parametric estimators.

Axiom 1 *: All production requires use of some input: (x,y)* $\notin \Psi$ *if* $x = 0, y \geq y \neq 0$ *(No free lunch condition)*

This admits the inactivity of a production process but the impossibility to produce output without input.

Axiom 2 *: Both inputs and outputs are freely disposable: if (x,y)* $\in \Psi$*, then for any (x',y') such that x'*$\geq$ *x and y'*$\leq$ *y, (x',y')*$\in \Psi$ *(Strong disposability condition).*

This characterizes the technology function as a monotone function and admits the inefficiency of the production process

Axiom 3 *:* Ψ *is convex: if* $(x_1, y_1), (x_2, y_2) \in \Psi$*, then* $(x, y) \in \Psi$ *for* $(x, y) = \alpha(x_1, y_1) + (1 - \alpha)(x_2, y_2), \forall \alpha \in$ *[0,1].*

Axiom 4 *:* Ψ *is closed.*

This assumption is a necessary condition to avoid mathematical problems for infinite production plans.

Axiom 5 *: The sample observation* (x_i, y_i) *in* χ_n *are realizations of identically, independently distributed (iid) random variables (X,Y) with probability density function f(x,y), which has support over* $\Psi \subset \mathbb{R}_+^{N+M}$*, the production set defined as* $Prob((X, Y)) \in \Psi) = 1$ *(Random sample condition).*

It states that the observations are considered as random draws from a population of firms.

Axiom 6 *: The density $f(x,y)$ is strictly positive on the boundary of Ψ and is continuous in any direction toward the interior of Ψ (Positiveness condition).*

It states that the probability of observing firms in an open neighborhood of the frontier is strictly positive. It assures the consistency of nonparametric estimators, hence it could be relaxed at a cost of losing performance in the rate of convergence.

Axiom 7 *: For all (x,y) in the interior of Ψ, the functions $\theta(x,y)$ and $\lambda(x,y)$ are differentiable in both arguments (Smoothness condition).*

It is a sufficient condition used by Kneip, Simar and Wilson (2003)[46] to derive the asymptotic condition of the DEA estimator. Summing up, these axioms provide a completely characterization of the production set Ψ and of the DPG by means of the $f(x,y)$ and of Ψ itself with the regularity conditions $(5-7)$.

The above considerations are equally applied to the described parametric approach as well as to the nonparametric approaches for making estimation. It is useful to imagine a spectrum of estimation approaches, ranging from fully parametric (most restrictive) to fully non-parametric (least restrictive). Fully parametric estimation strategies necessarily involve stronger assumptions on the probability model, which is completely specified in terms of a specific probability distribution function, structural equations, etc. Semi-parametric strategies are less restrictive; in these approaches, some (but not all) features of the probability model are left unspecified (for example, in a regression setting one might specify parametric forms for some, but not all, of the moments of a distribution function in the probability model). Fully non-parametric approaches assume no parametric forms for any features of the probability model. Instead, only (relatively) mild assumptions on broad features of the probability distribution are made, usually involving assumptions of various types of continuity, degrees of smoothness, etc. With nonparametric approaches to efficiency estimation, no specific analytical function describing the frontier is assumed. In addition, (too) restrictive assumptions on the stochastic part of

the model, describing the probabilistic behavior of the observations in the sample with the respect to the efficiency boundary of Ψ, are also avoided.

Hence, in this thesis I will focus on non parametric estimators of the efficient frontier, based on Envelopment Techniques, in particular on the Data Envelopment Analysis (DEA) that is one of the most extensively used for estimating efficiency of firms as it relays only on very few assumptions for Ψ.

2.3 The Envelopment Estimator: Data Envelopment Analysis (DEA)

In this section, I introduce the most popular nonparametric estimator, the Data Envelopment Analysis (DEA). DEA belongs to the set of estimators that envelops the observed data. The set is composed by two estimators that differ in the axioms they rely on:

- Free Disposal Hull (FDH): it is based only on free disposal axiom on Ψ (Deprins Simar and Tulkens, 1984)[24];
- Data Envelopment Analysis (DEA): it is based on free disposal axiom *and* convexity of Ψ (Farrell, 1957)[30].

The main advantages they provide are:

- find the $\hat{\Psi}(X)$ which envelops the data in some best way;
- the frontier is not defined by some parametric model (no parametric assumption for the frontier);
- no particular assumption on the density $f(x, y)$ of (x, y) on Ψ (no particular assumption for the DGP).

They confer great flexibility to the analysis since a multi-input and multi-output scenario can be analyzed with simple computation efforts. On the other hand, the main drawbacks are:

- very sensitive to error in measurements due to extreme values, outliers or missing explaining variables (as any envelopment estimators);
- no noise is allowed (deterministic model).

With respect to them, the new theory on the frontier estimation departs and aims to overcome them since their negligence brings to distortions in the estimates, and hence distortion in the performance analysis.

However, from a estimate computation point of view, the DEA estimator of the production set was initiated by Farrel (1957)[30] and later operationalized as a Problem of Linear Programming by Charnes, Cooper and Rhodes (1978)[17] which assumed the free disposability and the convexity of the production set Ψ. The main idea is:

- estimating the attainable set Ψ by the smallest set $\hat{\Psi}$ that represents the smallest free disposal convex set covering all the data;
- measurement of the distance for a given unit (x, y) relative to the boundary of the convex hull of $\chi_n = \{x_i, y_i\}$ for $i = 1, ..., n$.

The DEA estimator of the attainable set Ψ is defined as:

$$\begin{aligned}\hat{\Psi}_{DEA}(\chi) = \{(x,y) \in \mathbb{R}_+^{p+q} \mid y \leq \sum_{i=1}^{n} \gamma_i y_i; \quad x \geq \sum_{i=1}^{n} \gamma_i x_i \quad \text{for}(\gamma_1, ..., \gamma_n) \\ s.t. \sum_{i=1}^{n} \gamma_i = 1; \gamma_i \geq 0, i = 1, ..., n\}.\end{aligned} \tag{2.13}$$

The $\hat{\Psi}_{DEA}(\chi)$ given by (2.13) refers to the VRS (see Banker, Charnes and Cooper, 1984 [7]). From (2.13), other estimators can be derived by modifying the constraints; in particular:

- $\hat{\Psi}_{CRS}$ (Constant Returns to Scale) is obtained dropping the equality constraints by (2.13);
- $\hat{\Psi}_{NIRS}$ (Non Increasing Returns to Scale) is obtained replacing the equality constraints in (2.13) with $\sum_{i=1}^{n} \gamma_i \leq 1$;
- $\hat{\Psi}_{NDRS}$ (Non Decreasing Returns to Scale) is obtained replacing the equality constraints in (2.13) with $\sum_{i=1}^{n} \gamma_i \geq 1$;

On turn, the estimation of the input and output requirement set $X(y)$ and $Y(x)$ as well as of the $\vartheta X(y)$ and $\vartheta Y(x)$ are achievable just plugging-in the true, but unknown, production set Ψ with the one of the estimators $\hat{\Psi}$. It follows that the estimation of the efficiency scores for a firm operating at level (x_0, y_0) are computed by solving the linear program, respectively, for input and output case:

$$\begin{aligned}\hat{\theta}(x_0, y_0) = min\{\theta \mid y_0 \leq \sum_{i=1}^{n} \gamma_i y_i; \quad \theta x_0 \geq \sum_{i=1}^{n} \gamma_i x_i \quad \text{for}(\gamma_1, ..., \gamma_n) \\ s.t. \sum_{i=1}^{n} \gamma_i = 1; \gamma_i \geq 0, i = 1, ..., n\}.\end{aligned} \tag{2.14}$$

where $\hat{\theta}(x_0, y_0)$ measures the radial distance between (x_0, y_0) and the $(\hat{\theta}(x_0, y_0), x_0, y_0)$. Similarly in the output orientation, the efficiency score is estimated solving the following linear program:

$$\begin{aligned}\hat{\lambda}(x_0, y_0) = max\{\lambda \mid \lambda y_0 \leq \sum_{i=1}^{n} \gamma_i y_i; \quad x_0 \geq \sum_{i=1}^{n} \gamma_i x_i \quad \text{for}(\gamma_1, ..., \gamma_n) \\ s.t. \sum_{i=1}^{n} \gamma_i = 1; \gamma_i \geq 0, i = 1, ..., n\}.\end{aligned} \tag{2.15}$$

and the technically efficient level of output for a given level of input x_0 is estimated by $(x_0, \hat{\lambda}(x_0, y_0)y_0)$. This framework, based on computation of linear programs, is definitely not enough to conduct the efficiency estimates due to the lack of relationship between the unknown characteristic of the model that it is supposed to estimate and the estimated. As describe in Section 2.2, in any efficiency analysis it is working in a statistical model that describes a DGP that, on turn, completely characterized by the knowledge of the joint probability density function $f(x, y)$ and of it support Ψ under the condition imposed by axioms $(5 - 7)$.

Hence, as reported by Simar and Wilson (2000, 2006)[79] [82] is fundamental to investigate the statistical properties (consistency and unbiased) of the estimators to make inference and whether these estimators can reveal appropriate information about efficiency and under what conditions. As reported early, the DEA estimators are biased;

how about the consistency property[3] ?

In nonparametric frontier estimation, for decades nobody worried about these issues, neither for the DEA nor for the FDH estimators. The first result which appeared was due to Banker (1993)[6], who proved the consistency of the DEA efficiency scores in the very particular univariate case (one input for input orientation, or one output in the output orientation) but nothing was given about the rate of convergence. Later, many papers were published on the rate of convergence [4](Korostelev *et al.*,1995[49]; Kneip *et al.*, 1998 [45]; Park *et al.*, 2000)[61])[5]proving the consistency of the estimator but also the *curse of dimensionality* on the rate of convergence. These results, on the other hand, gave little practical importance for doing inference since its achievement it possible only by means the knowledge of the sampling distributions of the estimators. Indeed, the sampling distribution of the estimators allows to derive the bias, to compute its standard deviation and mainly to build confidence intervals for individual efficiency scores $\theta(x, y)$. In this complex situation, the only possible solution to obtain the minimal theoretical property of consistency has been to look at the asymptotic results, i.e. a reasonable approximation of the sampling distribution of the estimator when n is large enough (in the same spirit that a Central Limit Theorem gives an approximate normal distribution of a sample mean when n is large enough). It has been proved (see for detail Simar and Wilson, 2007[82]), that when the limiting distribution is available and tractable, then it is possible to estimated the *bias* of DEA estimators (as well as FHD estimators) and build *confidence intervals*. The only issue to handle is the additional noise introduced by estimating the unknown constants appearing in the limiting distributions. To date, the only solution available to date as the unique attractive practical way of making inference in both univariate and multivariate DEA estimation is the ***Bootstrap Technique*** (Efron, 1979[27]; Efron and Tibshirani, 1993[28]).

[3]Consistency means that if the sample size increases, an estimator $\hat{\theta}$ will converge to the true but unknown value θ it is supposed to estimate.

[4]It means that when the number of input and output increase, the rate of convergence decrease.

[5] For details, see Daraio and Simar, 2007[20].

2.4 A More Accurate Estimator: Bootstrapped DEA (B-DEA)

In this section, an overview of *bootstrap* in the context of the estimation of the frontier model is described. The first use of the bootstrap in the frontier model is attributable to Simar (1998c)[75] while further developments were introduced later by Simar and Wilson and Kneip, Simar and Wilson (2003)[46]. The reason underlying its adaption in this context relies on the fact that bootstrap allows for approximating the sampling distributions of interest by simulating, or mimicking the DGP. Indeed, the bootstrap features match with the essential goal of any efficiency analysis, as reported above, that is to identify the distributions of the estimators, either $\theta(x,y)$ or $\lambda(x,y)$ in order to identify the bias and construct the confidence intervals.

The presentation below is in terms of the input oriented case, but it can easily be translated to the output one. Hence, the bootstrap is intended to provide approximations of the sampling distributions of the difference $\hat{\theta}(x,y) - \theta(x,y)$ or, equally, ratio $\frac{\hat{\theta}(x,y)}{\theta(x,y)}$. Roughly speaking, when the bootstrap is employed the entire analysis is carried out in a *virtual scenario* that mimics the *true scenario*. Indeed, the virtual scenario has those desirable information than real one missed. It follows that both the estimation analysis and the inference procedure are carried out in this virtual place and the final results are transposed, by means of asymptotic approximations, in the real scenario. The shift between the two scenarios passes using, as channel, the consistent estimators of the DGP, $\hat{P}$, the production frontier $\hat{\Psi}$ and the efficiency estimates $\hat{\theta}(x,y)$ that are unknown in the real scenario whereas are assumed known in the bootstrap scenario. At the starting point of the analysis, the two scenarios differ in the information they have:

- real scenario is characterized by the sample of observations on input and output level for the set of production units $\left(\chi_n = \{(x_i, y_i), i...n\}\right)$
- virtual scenario is characterized by the $\hat{P}$, since it is a known estimate.

Therefore, in the virtual scenario a new data set $(\chi_n^* = \{(x_i^*, y_i^*), i...n\}$ can be generated by $\hat{P}$. On turn, this pseudo-sample defines the corresponding quantities $\hat{\Psi}^*$ and $\hat{\theta}^*(x,y)$ that can be viewed as estimators of the corresponding quantities $\hat{P}$ and $\hat{\theta}(x,y)$. They

are defined by:

$$\hat{\Psi}_{DEA}(\chi) = \{\,(x,y) \in \mathbb{R}_+^{p+q} \mid y \leq \sum_{i=1}^{n} \gamma_i y_i^*; \quad x \geq \sum_{i=1}^{n} \gamma_i x_i^* \quad \text{for}(\gamma_1, ..., \gamma_n) \\ s.t. \sum_{i=1}^{n} \gamma_i = 1; \gamma_i \geq 0, i = 1, ..., n. \tag{2.16}$$

For a fixed point (x_0, y_0), an estimator of $\hat{\theta}(x,y)$ is provided by

$$\hat{\theta}^*(x_0, y_0) = min\,\{\,\theta \mid y_0 \leq \sum_{i=1}^{n} \gamma_i y_i^*; \quad \theta x_0 \geq \sum_{i=1}^{n} \gamma_i x_i^* \\ s.t. \sum_{i=1}^{n} \gamma_i = 1; \gamma_i \geq 0, \forall i = 1, ..., n. \tag{2.17}$$

Conditionally on χ_n, the sampling distribution of $\hat{\theta}^*(x,y)$ is (in principle) known since $\hat{P}$ is known, although, in practice, it is impossible to compute analytically.

Monte Carlo methods can be used to easily approximate the sampling distribution of $\hat{\theta}^*(x,y)$. Using $\hat{P}$ to generate B samples $\hat{\chi}_b^*$ for $b = 1, ..., B$ and applying the linear program described above, for the given unit (x,y), a set of pseudo estimates $\left\{\hat{\theta}_{b=1}^{B}\right\}$ is obtained. The empirical distribution of these bootstrapped values gives a Monte Carlo approximation of the sampling distribution of $\hat{\theta}^*(x,y)$, conditional on $\hat{P}(\chi_n)$. If the bootstrap method is ***consistent***, the available bootstrap distribution of $\hat{\theta}^*(x,y)$ will "mimic" the original unknown sampling distribution of the estimator of interest $\hat{\theta}(x,y)$. Formally,

$$(\hat{\theta}^*(x,y) - \hat{\theta}(x,y))|\hat{P} \sim (\hat{\theta}(x,y) - \theta(x,y))|P \tag{2.18}$$

or equivalently

$$\frac{\hat{\theta}^*(x,y)}{\hat{\theta}(x,y)}|\hat{P} \sim \frac{\hat{\theta}(x,y)}{\theta(x,y)}|P \tag{2.19}$$

Since the left hand side of (2.18) or (2.19) is available (though the Monte-Carlo exercise), it can be used to provide properties usually obtained from the right-hand side. Hence, the bootstrap is an asymptotic procedure, whose quality of approximation depends on both the number of replications B (by the law of large numbers) and the sample size n. When the bootstrap is consistent, the approximation becomes exact as $B \to \infty$ and $n \to \infty$.

The main aspect underlying the adaptation of the bootstrap is that the consistency condition must have been verify in order to hold the condition expressed by (2.18). Basically, the bootstrap consistency depends on how the pseudo-sample are generated. It has been proved (Simar and Wilson, 1998a [77]) that the simple *naive bootstrap*[6] provides an inconsistent approximation of the desired sampling distribution. A more sophisticated bootstrap procedure based on smoothing techniques, instead on discrete empirical technique, has been proposed by Simar and Wilson.

The main idea is to generate the pseudo observation (x^*, y^*) from a *smooth* estimate of the density $f(x, y)$ (Silverman and Young, 1987[13]). The smooth non parametric density estimators are the most common used since they:

- provide a smooth density estimates;
- get easy the drawing from the kernel density estimates.
- the problem involves multiple $(p + q)$ dimensions;
- the boundary of support, $\vartheta\Psi$, is unknown;
- the boundary is non-linear.

These problems have been solved by Simar and Wilson (2000a [79], 2007[82]) by means of exploiting the radial nature of the efficiencies $\theta(x, y)$ and $\lambda(x, y)$, and, in particular, the third problem has been dealt with by transforming the problem from Cartesian coordinates to spherical coordinates. They propose a complicate algorithm, based on the kernel estimation, to simulate the pseudo-data, taking into account the polar coordinates and then transform them back in Cartesian ones to obtain the interested bootstrap sample. Since its complexity, they propose a simplified version that requires an additional assumption on the DGP. In particular, they assume a *homogeneous distribution* of the inefficiency over the input-output space[7]. The criterion is to create a bootstrap sample by projecting each observation (x_i, y_i) onto the estimated frontier, and the projecting this point away from the frontier randomly.

[6] It consists in the simple re sampling with replacement from the original $\chi(x, y)$

[7] This is analog to the assumption of homoskedasticity in regression model.

For practical purposes[8], it is advantageous to express the input-oriented efficiency in terms of the Shepard (1970) input-distance function introduced in Section 2.1 and defined in (2.9) as $\delta(x, y) = (\theta(x, y))^{-1}$). From now on, the Shepard distance will be used as the reference notation.

So, briefly the main aspects of the new procedure based on the *homogeneous bootstrap*:

- *Hypothesis*:Homogeneous Distribution of the efficiency scores;
- *Efficiency Density Estimator*: Smoothed Kernel Estimator (Silverman and Young, 1987[13]);
- *Boundary Constrain*: Reflection Method (Silverman, 1986)[74].

A dissertation on the kernel estimator will be provided in next Section; here it is enough to report that the problem of generation of the pseudo data sample from the estimates obtained using the kernel density estimator on the $2n$ points (the sample size becomes $2n$ as the reflection of the original set composed by n points) does not require computation of the density estimate (by the convolution theorem; see Efron and Tibshirani, 1993[28]) but only a value for the selected bandwidth h is need. Here the need to the best choice of the bandwidth since its influence on the inference of the estimates. For the last version of the **Homogeous Bootstrap Algorithm**, see Simar and Wilson, 2006[82].

2.5 The Inference: Bootstrapped Bias and Confidence Intervals

In this section, I illustrate how the bootstrap allows to correct for the bias (bias that appears by construction) of the DEA estimators and how to make inference, and hence, how to construct the confidence interval of the point estimated. Recalling the basic idea of the mechanism of the bootstrap technique, in the bias estimate too we are used to

[8]It is convenient working with efficiency score characterized by only one boundary than two. Indeed, the input distance expressed in Shepard distance is greater than one, contrary to the Farrell distance

work in the virtual world. Generally speaking, the bias is defined as:

$$bias[\hat{\delta}(x,y)] = E[\hat{\delta}(x,y)] - \delta(x,y) \tag{2.20}$$

Of course, its estimate can be compute only in the virtual world since more information is available. It is defined as:

$$\hat{bias}_B[\hat{\delta}(x,y)] = B^{-1}\sum_{b=1}^{B}\hat{\delta}_b^*(x,y) - \delta(x,y) \tag{2.21}$$

Then, a biased-corrected estimator of $\delta(x,y)$ can be computed as:

$$\begin{aligned}\bar{\delta}(x,y) &= \hat{\delta}(x,y) - \hat{bias}_B[\hat{\delta}(x,y)] \\ &= 2\hat{\delta}(x,y) - B^{-1}\sum_{b=1}^{B}\hat{\delta}_b^*(x,y) - \delta(x,y)\end{aligned} \tag{2.22}$$

In practice, the correction for the bias is not always recommended. One should avoid to use $\bar{\delta}(x,y)$ if it has higher mean-square error than $\hat{\delta}(x,y)$. The mean-square error of $\hat{\delta}(x,y)$ is defined by:

$$\hat{std}(\hat{\delta}(x,y)) = B^{-1}\sum_{b=1}^{B}[\hat{\delta}_b^*(x,y) - \tilde{\delta}(x,y)] \tag{2.23}$$

where $\tilde{\delta}(x,y) = B^{-1}\sum_{b=1}^{B}\hat{\delta}_b^*(x,y)$. Efron and Tibshirani(1993)[28] recommend the correction if:

$$\frac{|\hat{bias}_B(\delta(x,y))|}{\hat{std}_B(\delta(x,y))} > 0.25. \tag{2.24}$$

With respect to the *Bootstrapped Confidence Intervals* , the aim is know the distribution of $\delta(x,y)$ in order to derive the value of the quantiles. It is may be possible and if the distribution of $(\hat{\delta}(x,y) - \delta(x,y))$ is known since it would be trivial to find the values a_α and b_β such that:

$$\text{Prob}(-b_\alpha \leq \hat{\delta}(x,y) - \delta(x,y) \leq -a_\alpha) = 1 - \alpha \tag{2.25}$$

where a_α denotes the α^{th}-quantile of the sampling distribution of $(\hat{\delta}(x,y) - \delta(x,y))$. Of course, the quantile a_α is unknown but it can be derived from the empirical bootstrap distribution of the pseudo estimates $\hat{\delta}_b^*$ for $b = 1, ..., B$. Indeed, it is possible to exploit the approximation given by () and hence compute the quantile estimates such that:

$$\text{Prob}(-\hat{b}_\alpha \leq \hat{\delta}^*(x,y) - \hat{\delta}(x,y) \leq -\hat{a}_\alpha \mid \chi) = 1 - \alpha \tag{2.26}$$

Practically, it means:

1. sorting the value $\hat{\delta}^*(x,y) - \hat{\delta}(x,y), b = 1, ..., B$ by algebric value;
2. deleting $(\frac{\alpha}{2} \times 100)$-percent of the elements at either end and of this sorted array;
3. setting $-\hat{b}_\alpha$ and $-\hat{a}_\alpha$ equal to the end points of the resulting sorted array, with $\hat{a}_\alpha \leq \hat{b}_\alpha$

The bootstrap approximation is obtained by:

$$\text{Prob}(-\hat{b}_\alpha \leq \hat{\delta}(x,y) - \delta(x,y) \leq -\hat{a}_\alpha \mid \chi)1 - \alpha \tag{2.27}$$

and thus, the $(1-\alpha)$-percent confidence interval is:

$$\hat{\delta}(x,y) + \hat{a}_\alpha \leq \delta(x,y) \leq \hat{\delta}(x,y) + \hat{b}_\alpha. \tag{2.28}$$

Chapter 3

The Malmquist Productivity Index and the point estimates

3.1 A Literature Overview: The Historical Debate

In this section, I will introduce a literature overview concerns the most popular, as well as discussed, indicator: the **Malmquist Productivity Index (MPI)**. Nowadays, measures of productivity growth constitute core indicators for the analysis of economic growth. However, there are many different approaches to productivity measurement and their calculation and interpretation requires careful consideration, in particular when undertaking international comparisons.

The MPI is the most appealing productivity indicator in the performance analysis due to its flexible nature that get it preferable to the remaining number indexes (*e.g.* T*ö*rquist index and Fisher indices) as its nature of measuring of productivity change over time. Indeed, in the general formulation, it requires neither assumptions about efficient producer behaviour nor about constant returns to scale. Then, the main question posed by researchers has been and, still is, "How is it possible to empirically implement a Malmquist quantity index, if all one has are quantity observations for two or more periods ?" (lack of information on prices).

The first inspiration for addressing this issue was taken by F*ä*re and Grosskopf when they were invited by Lindegren and Roos to work on a project to measure the

productivity of Swedish hospitals. They decided to accomplish this task on the base of the their experience in the "Activity Analysis", the knowledge of the paper proposed by Caves, Christen, and Diewert (1982) and the theory of the Törquist index. Since the lack of information on the prices or explicit assumption about optimization, they abandon the Törquist index for inspiration and merge the ideas from the measurement of efficiency by Farrell (1957)[30] and the embryonic versions on the MPI propose by Caves *et al.*. Indeed, Caves *et al.* theoretically introduced the now popular Malmquist Index (MI) as the ratio between two distance functions that compares a firms productivity with that of an alternative firm and, in a straightforward dynamic extension, over time. The assumption of the no technical inefficiency was very restricting as well as the assumption on the translog form for the technology function.

They overcame these shortcomings, starting to work empirically on the MPI implementation by means of Data Envelopment Analysis (DEA) techniques, modeling the technology as piecewise linear and allowing for inefficiencies. Indeed, the introduction of the concepts of efficiency measurements based on the distance functions according to the Farrell terminology, allow for the exploration of the multi-faced efficiency phenomenon.

In the first moment, Färe *et al.* (1992) [32] broke down the productivity change into two components: change in efficiency, called *catching term* and change in technology, called *innovation*. The distinction between technological change and efficiency change can be made by conceiving the firm as operating in an exogenously determined environment, the technology defined by (2.2), which is the set of all at a given period feasible combinations of input and output quantities. A firm which operates on the boundary of this set is called technically efficient, whereas a firm which operates in the interior of this set is called technically inefficient. Technological change then means that the set of feasible combinations expands or contracts, while technical efficiency change means that the firm moves closer to or further away from the boundary. These two kinds of movement are clearly independent of each other: there can be technological change without efficiency change, and efficiency change without technological change. More usual, however, is a combination of both factors.

But the key question regarding the developments were, and still are, whether it

is possible to propose a Malmquist productivity index decomposition that provides all relevant information regarding technological and efficiency change, and whose terms can be interpreted in a meaningful manner. Very soon, the scientific community posed the following questions:

- how to mathematically disentangle the MPI in many "building blocks"?
- how important is the definition of the underlying technology when defining MPI?
- how to economically interpret these "building blocks"?

Färe *et al.* (1994) [33] proposed the first extended decomposition following a *top-bottom approach*, in which the Malmquist Index is factorized in four blocks, originally included results for both CRS and VRS version technology. Indeed, this decomposition takes into account the *scale component*, broking down the technical change into: pure efficiency change and scale efficiency change.

This paper was objective of flourishing debate among many authors and was followed by many others papers on alternative extended decompositions. Nowadays, it does not exist an unique decomposition of MPI universally accepted by the scientific community (see for details, the paper recently published by Zofio (2007)[91] for an complete overview on the MPI decomposition) since each of them could be interpreted in a meaningful way. Choosing one of them required to sacrificed some information provided by the remaining, and *viceversa*. The debate was surrounded on the *trade-off* between technical change and scale efficiency change and they did not found a common ground that would render all terms meaningful from a theory of production context.

Färe *et al.* individualized the scale contribution in this way:

1. defining the technical change corresponding to the potential productivity change at optimal scale.

   This assumption was inherited from Färe *et al.* (1989) [31], but it was always not well accepted by the scientific community that prefers the usual definition based on the concept of how a given unit varies its productive performance in time in reference to the technology changes. In these terms, Färe *et al.* (1994)[33]

measured the technical change on the benchmark cone technology, pointing out the independence of the technical efficiency from the scale effect.

2. attributing the scale effect only to the efficiency change.

Unconvinced by the existing definition of technical change, two groups of authors (Grifell-Tatjé and Lovell, 1996 [39]; Balk, 2001[5]; Lovell, 2003[53] on one hand and on the other hand Ray and Desly[1],that worked independently of each other, rejected the proposed decomposition claiming :

1. the lack of correctness in the technical change;

2. the lack of meaning of the productivity change at the potential optimal scale.

Contemporaneously, they proposed an alternative decomposition, based on additional component: return to scale.[2] It was suddenly criticized by Färe *et al.* (1997) [35] as containing the mix-period terms (not suitable for accounting the scale effect).

At this point of the debate, the main question was "Is it possible to suggest a decomposition which includes *the scale bias of the technical change?*" The debate attracted the attention of other researchers since from a theoretical and empirical point of view. The empirical application of the previous decompositions could not be undervaluated since they exhibited striking difference in the estimates of the annual productivity growth rate for productivity in the US: Färe *et al.* estimated an average annual growth rate of 0.85 percent over the period 1979-1988 against a rate of -5.56 percent over the period 1979-90 estimated by Ray and Desli. Moreover, it was not plausible that US faced a productivity decline of 5.56 percent per annum during 1979-90, particularly since US gross domestic product increased by 2.3 percent per year over this period while the rate of growth in the ratio of gross domestic product to labor exceeded the growth rate of the capital/labor ratio in the US over this period.

[1]They were rather critical and stated that the decomposition proposed raises the problem of "internal consistency" in the VRS. Indeed, under the CRS condition, the technical change term correctly shows the shift in the frontier but, no scale effect exist at all. On the other hand, if VRS is hold, the Färe *et al.*, (1994) technical change does not show the maximum producible output changes due to technical change holding the input bundle constant. Hence, they judged the previous extended decomposition misleading

[2]They offered an alternative decomposition in which the pure efficiency change appeared as in the paper Färe *et al.* (1994)[33], but both the scale efficiency change and technical change measures differed, since they chose the VRS frontier as benchmark, rather than the CRS frontier.

At the end of the story, the extended decomposition progresses following a more rigorous direction, combining the main standpoints previously neglected, the economical and statistical ones. Indeed, Simar and Wilson (1998b) [76] and Zofio and Lovell (1998)[92] uncovered the concept of scale bias of technical change offering a decomposition that, to date, is the best both in terms of economically interpretability in the context of the production theory and in information provided. In particular, Simar and Wilson pointed out that even in the Ray and Desli's decomposition there was something misleading since the component which is supposed to measure changes in returns to scale confounds the different effects of the movement of production unit in input/output space and changes in the shape of the technology over time. Färe *et al.*(1997)[35] recognized the inappropriateness of the decomposition but neglected the measurement confusions among the two effects.

Then, Simar and Wilson (1998)[76] proposed for the first time a **more accurate Malmquist Indexes framework**, based on the to-down approach, departing from the simple version of the MPI decomposition proposed by Färe *et al.*. They uncovered the main issues raised by the hsitorical debate as well as further drawbacks, completely neglected. Their paper is still unpublished but is often cited by many other authors (*e.g.* Grosskopf, 2003[40]).

The decomposition debate is still open in literature and further decompositions are continually proposed in the scientific community (*e.g.* Thanassoulis and Karagiannis, Conference EURO XXII, Prague, 2007[87]). Conversely to these huge efforts to improve the decomposition, few attention is given to the statistical nature of the estimates and the only researchers to be involved in this field, at the best of my knowledge, are Simar and Wilson (1998b [76], 1999b [78]), now supported by Curi, Daraio, Mancuso (EWEPAX, Lille, 2007)[19] and Daskovska and Van Bellegem (EWEPAX, Lille, 2007)[21].

3.2 A More Accurate Estimation: Economic and Statistical Aspects

This section summarizes some remarks concerning the productivity measurement and its decomposition, limiting the focus only to a particular stream of research since the

fragmentation of the topic toward different directions. As demonstrated from the origin of theoretical and empirical studies on MPI, the productivity change has been always handled jointly with the concept of its decomposition.

Economically speaking, part of the current interest in productivity overall is due to the revival of the economic interest in the sources of economic growth; and at the center of the revived debate are productivity, technical change and more recently deviations from best practice which we can refer to generally as inefficiency, variously attributed to corruption, bad institutions, distorted incentives, market power, and so on (Grosskopf, 2003)[40]. What we know as neoclassical growth accounting implemented the SolowSwan aggregate production function approach which showed that output growth is due to growth of inputs, but also due to growth in productivity/technical change, which is exogenous in the Solow model. New growth theory seeks to reconcile patterns of economic growth across countries with these models by modifying the neoclassical Solow Swan growth model to allow for endogenous technical change. Human capital, RD and spillovers play key roles in this literature as sources of endogeneity. The ability to disentangle productivity changes into components associated with these factors may prove to be the most important application of these indexes. For example, we could pay more attention to the relationship between efficiency change (as a measure of diffusion or spillovers) and technical change to RD and human capital (see, for instance, Henderson and Russell [43]). Thus, the motivation for finding the main source of the productivity change is deep-seated in economics.

Mathematically speaking, the MPI is a very appealing tool to measure productivity change since it embraces better advantages relative to the other productivity index. The main advantages are:

- it does not require input or output prices, which makes it particularly useful in situations where prices are misrepresented or non-existent;
- it does not require the profit maximization or cost minimization assumption, which makes it useful in situation where the objectives o the producers suffer or are unknown or not achieved;
- it allows for the decomposition at least two components: *technical change* and

efficiency change;

- it provides an easy economic interpretability (*e.g.*[3] $MPI > 1 \rightarrow$ *regress* , $MPI = 1 \rightarrow$ *no change* , $MPI < 1 \rightarrow$ *improvement*).

The main disadvantage is the necessity to compute the distance function, defined by (2.7) (or 2.8, 2.9, 2.10) in section 2.1.

In this thesis, I will follow the decomposition proposed by Simar and Wilson (1998b[76], 1999b[78]) since the attractive nature of taking into account the inferential aspects of the estimation of MPI and its components and, with respect to, I handle some improvements in the inferential algorithm, as shown in the coming chapters. Hence, this choice of which type of decomposition employed is dictated by reasons of completeness and consistency of the entire research.

Basically, in the context of productivity growth, a different notation from those used in Chapter 2 is employed since:

- any estimator (*i.e* Production set, Production frontier, Distance functions) defined by, respectively, (2.1), (2.2), (2.7) or (2.9), take a additional dimension: the *time dimension*.

- it is necessary to work, simultaneously, with the Production Frontier Estimates as well as its Convex Cones.

Of course, the analysis gets more difficult since the statistical framework shift from a univariate setting (**Cross-Sectional Data**)[4] to a bivariate setting **(Time-Series Cross-Sectional Data)**[5] and involves the computation of distance functions from two different shapes of the technology (CRS and VRS).

So briefly, firstly I re-define the **The Frontier Model** in the **input case**, taking into account the time dimension as:

- The *production possibilities set* at time $t = 1, ..., T$, is given by:

$$\Psi^t = \{(x, y) \mid x \text{ can produce } y \quad \text{at time} t\} \tag{3.1}$$

[3] Output oriented case.

[4] Cross-Sectional Data refers to data collected by observing many subjects (such as firms, countries,or regions) at the same point of time, or without regard to differences in time.

[5] They are called also Panel Data.

where $x \in \mathbb{R}^p_+$ and $y \in \mathbb{R}^q_+$ denote the vectors of input and output, respectively.

- The *production frontier* is given by:

$$\vartheta\Psi^t = \left\{(x,y) \in \Psi^t \mid (x/\lambda, y) \notin \Psi^t \forall \lambda > 1\, (x, y/\theta) \notin \Psi^t, \forall \theta < 1\right\} \tag{3.2}$$

- The Shepard input distance function for the i-th production unit at time t_j, relative to the technology existing at time t_k:

$$D_i^{t_j|t_k} \equiv \sup\left\{\beta > 0 \mid x_i^{t_j}/\beta, y_i^{t_j} \in \Psi^{t_k}\right\}. \tag{3.3}$$

- The set Ω^t is the convex cone (with the vertex at the origin) spanned by $\Psi^t : \Psi^t \subseteq \Omega^t$. If the technology exhibits CRS everywhere, then the technology implying a mapping $\mathbf{x} \rightarrow \mathbf{y}$ that is homogeneous of degree 1 (*i.e.*, $(\mathbf{x}, \mathbf{y}) \in \vartheta\Psi^t$ implies $(\lambda\ \mathbf{x},\ \lambda\mathbf{y}) \in \vartheta\Psi^t \quad \forall \lambda > 0$. In this case $\Psi^t = \Omega^t$. This property refers to the well-know *Proportionality Axiom* (Forsud,1997[37]),that states that if outputs are to be increased in the same proportion from one period to the next while inputs remain the same, then the productivity index is to increase in the same proportion. Correspondingly, if inputs are reduced in the same proportion while outputs remain the same, then the productivity index should increase in such proportion. With regard to the specific MPI, this property requires that the benchmark technology is characterized by constant returns to scale. However, the fact that the supporting technology to correctly define productivity indices correspond to constant returns to scale does not mean that the underlying technology may not exhibit variable returns to scale[6]. In fact, when identifying the contribution of returns to scale and scale efficiency one implicitly assumes that these terms have a role to play driving productivity change and, therefore, must be included in the analysis.

- The Shepard input distance function from Ω^{t_k} is given by:

$$\Delta_i^{t_j|t_k} \equiv \Delta^{t_k}(x_i^{t_j}, y_i^{t_j}) \equiv \left\{\beta > 0 \mid x_i^{t_j}/\beta, y_i^{t_j} \in \Omega^{t_k}\right\} \tag{3.4}$$

[6]According to Grosskopf (2003)[40] the proportionality axiom should not be strictly enforced if the users were aware that it does not have an average product interpretation

- By construction, the following relations are hold:

  1. if $t_j = t_k = t$, then $\Delta^t(x_i^t, y_i^t) \geq D^t(x_i^t, y_i^t) \geq 1$;
  2. if $\Psi^t = \Omega^t$, then $\Delta^t(x_i^t, y_i^t) = D^t(x_i^t, y_i^t) \geq 1$ and the CRS prevail everywhere on the technology.

Secondly, I re-call all the **Axioms** on the Data-Generating Process provided in Section (2.2), for the definition of **The Statistical Model**. In this case, the random sample of production units is given by:

$$\chi_n^t = \{(x_{i,t}, y_{i,t}), i = 1, ..., n\} \text{for} t = t_1, t_2. \tag{3.5}$$

Moreover, the DEA estimator of Ψ^t and Ω^t, defined in Section (2.2) are re-defined, plug-in the time dimension, so that:

- convex hull estimator (*i.e.* VRS) is defined as:

$$\begin{aligned} \hat{\Psi}_{DEA}^t(\chi) = \{(x, y) \in \mathbb{R}_+^{p+q} \mid y \leq \sum_{i=1}^n \gamma_i y_{it}; \quad x \geq \sum_{i=1}^n \gamma_i x_{it} \\ s.t. \sum_{i=1}^n \gamma_i = 1; \gamma_i \geq 0, i = 1, ..., n\} \end{aligned} \tag{3.6}$$

  with the following estimates of the distance functions:

$$\hat{D}_i^{t|t}(x, y) = sup\left\{\beta \mid (x_{it}/\beta, y_{it}) \in \hat{\Psi}^t\right\} \tag{3.7}$$

$$\hat{D}_i^{t_1|t_2}(x, y) = sup\left\{\beta \mid (x_{it_1}/\beta, y_{it_1}) \in \hat{\Psi}^{t_2}\right\} \tag{3.8}$$

- conical hull estimator (*i.e.* CRS) is defined as:

$$\begin{aligned} \hat{\Omega}_{DEA}^t(\chi) = \{(x, y) \in \mathbb{R}_+^{p+q} \mid y \leq \sum_{i=1}^n \gamma_i y_{it}; \quad x \geq \sum_{i=1}^n \gamma_i x_{it} \\ s.t. \gamma_i \geq 0, i = 1, ..., n\} \end{aligned} \tag{3.9}$$

with the following estimates of the distance functions:

$$\hat{\Delta}_i^{t|t}(x,y) = sup\left\{\beta \mid (x_{it}/\beta, y_{it}) \in \hat{\Omega}^t\right\} \tag{3.10}$$

$$\hat{\Delta}_i^{t_1|t_2}(x,y) = sup\left\{\beta \mid (x_{it_1}/\beta, y_{it_1}) \in \hat{\Omega}^{t_2}\right\} \tag{3.11}$$

Mutatis mutandis, both $\hat{D}_i^{t_2|t_1}(x,y)$ and $\hat{\Delta}_i^{t_2|t_1}(x,y)$ can be obtained by permuting the time indices. From a computation point of view, $\hat{D}_i^{t_1|t_2}(x,y)$ and $\hat{\Delta}_i^{t_1|t_2}(x,y)$ can be obtained by solving, respectively, the following linear programs:

$$\begin{aligned}(\hat{D}_i^{t_1|t_2})^{-1} = min\{\theta > 0 \mid y_{it_1} \leq \sum_{i=1}^{n} \gamma_i y_{it_2}; \quad \theta x_{it_1} \geq \sum_{i=1}^{n} \gamma_i x_{it_2} \\ s.t. \sum_{i=1}^{n} \gamma_i = 1; \gamma_i \geq 0, i = 1, ..., n\}\end{aligned} \tag{3.12}$$

and

$$\begin{aligned}(\hat{\Delta}_i^{t_1|t_2})^{-1} = min\{\theta > 0 \mid y_{it_1} \leq \sum_{i=1}^{n} \gamma_i y_{it_2}; \quad \theta x_{it_1} \geq \sum_{i=1}^{n} \gamma_i x_{it_2} \\ s.t. \sum_{i=1}^{n} \gamma_i = 1; \gamma_i \geq 0, i = 1, ..., n\}\end{aligned} \tag{3.13}$$

3.3 Point Estimation: MPI and its Decomposition

In this section I will present the MPI and its basic decomposition. From the previous section, we have learnt that we are working in a "workplace of estimates" and not in a "'workplace of measures" of the true values, since the the lack of possibility of observation of the production set as the requirement sets, as the values of the distances, incorrectly though for many years. In this Section, only for reason of notation, I do not put the symbol "hat" (previously used to differ an estimate from a measure)on the estimators. This notation will be assumed in Chapter 4.

After giving a preliminary framework within the analysis will be carried out, I would like to remarks three aspects on the productivity change measurement. In particular, it is worth to note that:

– the measurement and the explanation of productivity change are like two side of the same coin;

– in the case of $p = q = 1$ (one input and one output), the MPI over two different periods t_1 and t_2 for a specif unit operating at level (x, y) does not require the knowledge and, hence, the estimation of the frontier. It is given by:

$$MPI^{input} = \frac{x^{t_2}/y^{t_2}}{x^{t_1}/y^{t_1}} \tag{3.14}$$

Here, a gain o or a progress is when $MPI < 1$ and a regress or loss in productivity is when $MPI > 1$. Straightforwardly, MPI^{output} is given by:

$$MPI^{output} = \frac{y^{t_2}/x^{t_2}}{y^{t_1}/x^{t_1}} \tag{3.15}$$

Although traditionally productivity has the intuitive definition in the single input-single output case given by the ratio of input and output (output and input) as given by (3.5), for general purposes of economic evaluation, it is necessary to extend the analysis toward a multiple variable production process, which leads defining a multivariate setting for the measurement of the Total Factor Productivity. Contrary to the second remark stated above, in this context more information must be derived, so that the *aggregate input and output radial distance functions* must be computed relatively to the upper boundary of Ω^{t_k}. Following the definition provided by Simar and Wilson (1998[76], 1999[78])[7], the $MPI_i(t_1, t_2)$ for a specif unit i-th, operating at level (x, y) over two different periods t_1 t_2, is measured to the convex cones, Ω^{t_1} and Ω^{t_2}, and it is given by:

$$MPI_i(t_1, t_2) = \left(\frac{\Delta_i^{t_2|t_1}}{\Delta_i^{t_1|t_1}} \times \frac{\Delta_i^{t_2|t_2}}{\Delta_i^{t_1|t_2}} \right)^{1/2} \tag{3.16}$$

where the geometric mean is used to avoid choosing an arbitrary technology benchmark (as in the Fisher ideal indexes) and the two indexes recalled the form used by Caves *et al.* (1982) to prove the Törnquist exactness. Differently by the Färe et

[7]The structure of the decomposition is the same of the one proposed by Färe et al. (1992, 1994) but differs in the distance function estimates. Indeed, they were aware of that relation is hold true only if the underlying, true technology (that is unknown) is constant returns to scale everywhere.

al. (1992[32], 1994[33], Simar and Wilson (1998b)[76] correct the measure changes regardless of the shape of the true technology (they substituted distance given by (3.3) with (3.4)) since distances are measured relative to the boundary of the conical hull of the production set. The updated version of the MPI index is given by:

$$MPI_i(t_1, t_2) = \frac{\Delta_i^{t_2|t_2}}{\Delta_i^{t_1|t_1}} \times \left(\frac{\Delta_i^{t_2|t_1}}{\Delta_i^{t_2|t_2}} \times \frac{\Delta_i^{t_1|t_1}}{\Delta_i^{t_1|t_2}} \right)^{1/2} \tag{3.17}$$

where $t_2 > t_1$. Following the equation (7) in Färe *et al.* (1994), the first terms defined an input-based index of efficiency change, given by:

$$MPI_i(t_1, t_2 \mid t_1) = \frac{\Delta_i^{t_2|t_2}}{\Delta_i^{t_1|t_1}} \tag{3.18}$$

and it is relative to the technology at time t_1. Similarly, the remaining part of the right-hand side of the Eq. (3.8) defines an input based measure of technical change:

$$MPI_i(t_1, t_2 \mid t_2) = \frac{\Delta_i^{t_2|t_2}}{\Delta_i^{t_1|t_1}} \tag{3.19}$$

and it is relative to the technology at time t_2. By simple algebra, the MPI can be rewritten is its decomposed version as:

$$MPI_i(t_1, t_2) = \frac{\Delta_i^{t_2|t_2}}{\Delta_i^{t_1|t_1}} \times \left(\frac{\Delta_i^{t_2|t_1}}{\Delta_i^{t_2|t_2}} \times \frac{\Delta_i^{t_1|t_1}}{\Delta_i^{t_1|t_2}} \right)^{1/2} \tag{3.20}$$

As said above, the overall index is consistent if we suppose the true underlying technology exhibits CRS. In this particular case, the factors can be interpreted as follows:

- input-based index of efficiency change:

$$\Delta Eff_i(t_1, t_2) = \frac{\Delta_i^{t_2|t_2}}{\Delta_i^{t_1|t_1}} \tag{3.21}$$

  It measures the change in input efficiency between two periods t_1 and t_2. Values of $\Delta Eff_i(t_1, t_2)$ less than (greater than) 1 indicate improvements (decreases) efficiency between between t_1 and t_2. From an economical point of

view , they represents, respectively, whether the observation has gotten closer (or farther from) to the frontier over time.

- input-based index of technical change:

$$\Delta Tech_i(t_1, t_2) = \left(\frac{\Delta_i^{t_2|t_1}}{\Delta_i^{t_2|t_2}} \times \frac{\Delta_i^{t_1|t_1}}{\Delta_i^{t_1|t_2}} \right)^{1/2} \tag{3.22}$$

It measures the change in input technical between two periods t_1 and t_2. As with the $\Delta Eff_i(t_1, t_2)$, values of $\Delta Tech_i(t_1, t_2)$ less than (greater than) 1 indicate improvements (decreases) efficiency between between t_1 and t_2. Deeply, the first ratio inside the bracket evaluates the shift in the frontier at the data observed in period t_2, whereas the second captures that shift evaluated at the data observed in period t_1. From an economical point of view , they represent, respectively, whether technology frontier shifts is upwards (downward), indicating improving (regression) in innovation over time. As well is components, $MPI_i(t_1, t_2) < 1$ indicate improvement or progress in productivity between t_1 and t_2, whereas $MPI_i(t_1, t_2) > 1$ regress. $MPI_i(t_1, t_2) = 1$ states for no change in productivity.

3.4 Point estimation: MPI and its Extended Decomposition

In this section, as mentioned in section (3.2), I will follow the recent decomposition proposed by Simar and Wilson since it is the most reliable version of the MPI extended decomposition as well as it enjoys the attractive nature of including inferential aspects of the estimation of MPI and its components and, with respect to, I handle some improvements in the inferential algorithm, as shown in the coming chapters. Hence, this choice of which type of decomposition employed is dictated by reasons of completeness and consistency of the entire research. The main "source" of the decomposition debate is surrounding the strong hypothesis of CRT on the true, but unknown, technology. When the underlying technology does not exhibit CRS, the previous definition of the MPI is meaningless, even if in literature they are many authors which, careless of the consistency of the

indicator, used it indifferently. So, the entire measurement framework requires some adjustment to allows for other types of returns to scale.

Simar and Wilson,(1998b)[76] construct a extended decomposition that allows for including in a reasonable way the contribution of returns to scale, jointly looking at the scale efficiency change and the scale bias of technical change, following the distinction of distance functions, defined by (3.3) and (3.4). Hence, the uncovered extended decomposition is defined as the combination of four "'blocks":

- *pure efficiency change* ($\Delta PureEff$);
- *change in the scale efficiency* ($\Delta Scale$);
- *pure technology change* ($\Delta PureTech$);
- *change in the scale of technology* ($\Delta ScaleTech$)

Formally, for each i-th unit between t_1 and t_2, they are defined as, respectively, :

$$\Delta PureEff = \left(\frac{D_i^{t_2|t_1}}{D_i^{t_{11}}} \right) \tag{3.23}$$

It captures the *true* change in efficiency.

$$\Delta Scale = \left(\frac{\Delta_i^{t_2|t_2} / Di^{t_2|t_2}}{\Delta_i^{t_1|t_1} / Di^{t_1|t_1}} \right) \tag{3.24}$$

It captures the change in scale efficiency.

$$\Delta PureTech = \left(\frac{D_i^{t_2|t_2}}{Di^{t_2|t_2}} \times \frac{D_i^{t_1|t_1}}{Di^{t_1|t_2}}) \right)^{1/2} \tag{3.25}$$

It measures the real shift of the technology, being the geometric mean of two ratio of technological changes from the perspective of the firm at time t_1 and t_2

$$\Delta ScaleTech = \left(\frac{\Delta_i^{t_2|t_1} / D_i^{t_2|t_1}}{\Delta_i^{t_2|t_2} / D_i^{t_2|t_2}} \times \frac{\Delta_i^{t_1|t_1} / D_i^{t_1|t_1}}{\Delta_i^{t_1|t_2} / D_i^{t_1|t_2}} \right)^{1/2} \tag{3.26}$$

This term is more difficult to interpret, but facing it ratio by ratio, it is possible to stem:

- first ratio → change in shape of the technology between t_1 and t_2 relatively to the position of the i-th production unit at time t_2
  - the denominator measures the distance between the production frontier at time $t = 2$, from the perspective of $(x_i^{t_2}, y_i^{t_2})$;
  - the numerator measures the distance between the production frontier at time $t = 1$, from the perspective of $(x_i^{t_2}, y_i^{t_2})$;
- second ratio → change in shape of the technology between t_1 and t_2 relatively to the position of the i-th production unit at time t_1 (it measures the same thing of the first ratio but from the perspective of the i-th production unit at time t_1).

Hence, some conclusions can be withdraw:

1. $\Delta ScaleTech$ can only be due to changes in the shape of the technology, since the reference points are supposed fixed;
2. $\Delta Scale$ can be caused either by (i) changes in the shape of the technology, (ii) changes in the location of the production unit in input/output space between t_1 and t_2, or (iii) some combination of (i) and (ii).

Clearly, these four blocks can be combined couple by couple to provide the initial terms of the simple decomposition given by (3.11). In particular:

$$\Delta Eff = \Delta PureEff \times \Delta Scale; \tag{3.27}$$

and

$$\Delta Tech = \Delta PureTech \times \Delta ScaleTech. \tag{3.28}$$

Chapter 4

Inference on MPI and its components

4.1 A literature overview: an elitist research field

One decade ago, Färe *et al.* (1992)[32] arise the fundamental issue on how mathematically formulating and measuring the productivity change over time and how decomposing it to identify its main driven forces, in the context of the economic production theory. As reported in Chapter 3, a vital debate arise through the scientific community and began with addressing the issue of decomposition of MPI, trying to uncover the concept of *scale bias in the technical change*. Since the publication of "Productivity Growth, Technical Progress, and Efficiency Change in Industrialized Countries", the scientific efforts for later years were focused on the same goal but stood from two different perspectives, attracting a considerable different number of researchers. Indeed, it is evident the *gap* between:

- huge variety of contributions on different decompositions on the MPI and their empirical applications (*e.g.* for an overview: Färe, Grosskopf and Russell, 1998[34]; Grosskopf, 2003[40]; Lovell, 2003[53]);
- few attention devoted to the inferential aspects of the estimated effects, proposed by the first time by Simar and Wilson, (1998)[76].

As the Färe *et al.*'s paper, dated 1994, is a mainstay in the history of the MPI and its decompositions since it opened the research field on the ***measurement*** of the economic growth, in the same way the Simar and Wilson's working paper constitutes the forerunner of the most advanced studies surrounding the ***estimation*** of Malmquist Indexes. Although it is a widely known but so far unpublished working paper, it allowed Simar and Wilson to loom out of the fog of the debate in 1998 and shine a light on a more robust Malmquist Index Framework. Differently from the Färe *et al.*'s paper, few theoretical authors were attracted by this perspective, in spite of more robustness than the previous one; so far, this dominance of the topic has been confirmed by the sole paper proposed one year later by the same authors. Further contributions are presented at the conference EWEPAX, 2007 in Lille by other their co-authors. Broadly speaking, the contribution of this outstanding working paper, titled "Productivity Growth in Industrialized Countries", which retrieves the title previously attributed, is three-fold since it clarifies some issues of the debate and provides additional mathematically improvements. In particular, it uncovers the following issues:

- the *true*, but ***unknown***, underlying technology measured by Färe *et al.* (1994) [33] could only exhibit CRS (they were not aware of this property);
- information of return to scale could be stemmed by a jointly examination of the scale efficiency change and the scale bias of technical change.

Moreover, they pointed out further inaccuracies, completely neglected by the authors, that turned the way of carrying out a productivity analysis. Indeed, the author made confusion about one more point, besides the points claimed early; they did not make difference between the ***unknown*** quantities and then ***estimates***. Since the MPI defined by (3.16) and its decomposition defined by the combination of (3.27) and (3.28), depends on the concepts of the true, but unknown, distance functions measured to the production frontiers and the the convex cones, given, respectively, by (3.2) and (3.4), they have to be estimated and not merely determinated from the data. On the other hand, they neglected the the statistical nature of the estimation as, they explicitly stated of computing the distance functions

only by means of the "appropriate programming problems". With regards the statistical meaning of the estimation, they introduce a Statistical Model, as described in Section (3.2), or Econometric Model, and re-interpret what the previous authors did (and did not do) to clarify both the economic and the econometric issues in the context of productivity analysis. Hence, a totally different framework was proposed, framework that get more robust with the paper published the next year "Estimating and bootstrapping Malmquist Indices" on the European Journal of Operational Research in 1999, that is the point of departure of my thesis. The same authors developed a consistent bootstrap estimation procedure for obtaining confidence intervals for Malmquist indices of productivity and their decomposition. Since the far 1999, any scientific effort has been made surrounding improved methodology on the inference on the Mamquist Indexes as Simar and Wilson were focused on different research line as any other researcher gets involved in this field. During the last years, a group of new researchers and Ph.D students started to handle some theoretical research: in particular Daskovska,Simar and Van Bellegem are working on the Dynamical Analysis of the Malmquist Productivity Index while Curi, Daraio, Mancuso and Simar on the improvement of its inference.

4.2 Sensitivity Analysis and Inference: The Rule of Thumb

This Section will deal with the inferential setting surrounding the MPI and its components. Simar and Wilson, making use of their competences in *Bootstrap Techniques* in the univariate case (Simar and Wilson, 1998b[76],2007[82]), extended the Färe *et al.* approach by providing a sensitivity analysis to the Malmquist Indeces and its components. The aim of their efforts is to draw meaningful inferences from the point estimates in order to know whether changes in productivity, efficiency, or technology reflect the real economic phenomena or merely artifacts due to the sampling variation. From a statistically standpoint, making inference on Malmquist Productivity Indexes (and its components) means to test if some

of the corresponding measures are significantly different from 1 (the case of "'no change"'), and so indicate regress or progress. More specifically, to obtain the confidence intervals for each estimate is necessary to know the sampling distribution of the estimator. So far, bootstrap sounds the best practice to achieve this objective since it allows to approximate the sampling distribution of interest by mimic the unobserved data generating process, as described in Section (2.4). The only disadvantage of bootstrapping is that it could provide inconsistent estimators of the confidence intervals depending on how the replication of the Data Generating Process is carried out. From their previous research on the bootstrapped efficiency estimators, Simar and Wilson inherit the ***smooth bootstrap*** as consistent technique to simulate the original sample of data, χ_n, so that a sampling from a kernel density estimate of the joint density $f(x, y)$ is employed to simulates the bootstrap sample χ_n^*, avoiding to achieve inconsistency, otherwise obtainable using, for instance, the "naive" bootstrap. Additionally, the possibility of temporal correlation among the estimators, due to the presence of two periods (associated at i-th unit at time t_j and t_{j+1}), leads to a different statistical setting.

So, the mainly statistical difference between the efficiency estimation and the MPI estimation is the shift of the analysis from a univariate (only one period of observation) to a bivariate case (two periods of observation). Broadly speaking, the new setting modifies the estimation procedure including the time-dependence structure of the data; specifically speaking, it employs the bivariate kernel density estimator, instead of the univariate one.

Briefly summarizing, the main principles of the sensitive analysis concern:

- **Data Generating Process**: firms randomly deviate from the true frontier in a radial input direction with density f;
- **Bootstrap**: provide B pseudo samples:

$$\chi^{t,*} = \{(x_{it}^*, y_{it}^* \mid i = 1, ..., n\} \tag{4.1}$$

  $\Rightarrow$ bootstrap estimates

$$\{\hat{D}_i^{t_1|t_1*}, \hat{D}_i^{t_2|t_2*}, \hat{D}_i^{t_1|t_2*}, \hat{D}_i^{t_2|t_1*}\}_{b=1}^{B} \tag{4.2}$$

$$\{\hat{\Delta}_i^{t_1|t_1*}, \hat{\Delta}_i^{t_2|t_2*}, \hat{\Delta}_i^{t_1|t_2*}, \hat{\Delta}_i^{t_2|t_1*}\}_{b=1}^{B} \tag{4.3}$$

bringing the MPI and its main sources:

$$\widehat{MPI}_i^*, \quad \widehat{\Delta Eff}_i^*, \quad \widehat{\Delta Tech}_i^* \tag{4.4}$$

for $i = 1, ..., n$ and $b = 1, ..., B$

$\Rightarrow$ Monte Carlo realizations: eventually correct for bias and build their respective confidence intervals

$\Rightarrow$ Statistical Interpretation: if the value "1" is not in the confidence interval of the point estimates, the hypothesis of no change for the firm i-th between the two periods is rejected.

- **Consistent Bootstrap Estimates**: consistency is achievable $\Longleftrightarrow$ the replication of the Data-Generating-Process is correct.

  $\Rightarrow$ ***Smooth Bootstrap Procedure with Reflection***

- **Time Dependence**: the bootstrap has to take into account the dependence between $D_i^{t_1|t_1}$ and $D_i^{t_2|t_2}$ for the generation of the pseudo sample.

- **Open issue already known by the authors** "The only remaining issue is the *choice of the bandwidth,* h. Tapia and Thompson (1978), Silverman (1978, 1986), and Härdle (1990) discuss considerations relevant to the choice of h; in general, for a given sample size, larger values of h produce more diffuse (*i.e.*, less efficient) estimates of the density, while very small values produce estimated densities with multiple modes.
  In the empirical examples which follow, we use *Silverman's (1986)* suggestion for bivariate data by setting $\mathbf{h = (4/5n)^{1/6}}$ since we are using a bivariate normal kernel scaled to have the same shape as the data" (Simar and Wilson, 1999b [78])

It follows that, basically, the designed procedure turns the sensitive analysis into a problem of kernel density estimation, leaving room of research improvements on the performance of the *kernel density estimator*. Two choice must be made in order to implement the kernel density estimation:

- the kernel function selection (Product Kernel or Radially Symmetric Kernel);
- bandwidth selection (Rule of Thumb or Data Driven methods).

So, let us assume that $Z_1, ..., Z_n$ the given multivariate data set underlying density to be estimated, taken in $\mathbb{R}^d$, with $d = 2$. Broadly, the multivariate kernel density estimator with kernel K is defined by:

$$\hat{f}(z) = \frac{1}{nh^d} \sum_{i=1}^{n} K\{\frac{1}{h}(z - Z_i)\} \tag{4.5}$$

where h is the window width, also called the ***smoothing parameter*** or ***bandwidth*** and K is the multivariate density function. The kernel estimator is a sum of "bumps" centered at the observations whereas the kernel function $K(z) : \mathbb{R}^d \to \mathbb{R}$ is a function that satisfies the following conditions:

$$\int_{\mathbb{R}^d} K(z)dz = 1 \tag{4.6}$$

$$\int_{\mathbb{R}^d} zK(z)dz = 0 \tag{4.7}$$

$$\int_{\mathbb{R}^d} z^2K(z)dz \geq 0 \tag{4.8}$$

It determinates the shape of the bumps while the smoothing parameter h determinates their width. In our framework, we use the most popular function: the ***standard normal (Gaussian) probability density function***, defined as:

$$K(z) = (2\pi)^{-d/2} exp(\frac{1}{2} z^T z) \tag{4.9}$$

On the other hand, the use of a *single* smoothing parameter implies that the version kernel placed on each data point is scaled equally in all directions. In some cases, such as if the spread of the data points is different among the coordinate

directions, it may be more appropriate to use a vector of smoothing parameters or even a *matrix of shrinking coefficients*. This is equivalent to use the following multivariate kernel density estimator:

$$K_H(z) = |H|^{-1} K(H^{-1}z) \tag{4.10}$$

where H is a symmetric positive definite dxd matrix, called ***bandwidth matrix***, and $|H|$ is the determinant of the matrix H.

The choice of the bandwidth matrix H is a crucial aspect of the smoothing parametrisation when constructing a multivariate kernel density estimator, due of its role in controlling both the amount and the direction of multivariate smoothing[1]. Indeed, as we have higher dimensions, of course we have more degree of freedom.

The full bandwidth matrix allows for more flexibility; however it also introduce more complexity into the estimator since more parameters need to be chosen (Wand and Jones, 1995). They are several levels of sophistication when specifying the bandwidth matrix:

- $H = h^2 I \rightarrow$*single smoothing parameter*: same amount of smoothing in each coordinate direction with I the dxd identity matrix ;
- $H = diag(h_1^2, ..., h_d^2) \rightarrow$ *multiple smoothing parameter*: different amount of smoothing in each coordinate direction, with H in the diagonal form;
- smooth in directions different to those of the coordinate axes (arbitrary orientation).

In the framework of the DEA distance estimations, we are interested in taking into account the dependence among the component of Z. So, a simple version of bandwidth matrix with arbitrary orientation is used: the ***scaled bandwidth matrix*** $H = hS^{1/2}$ where S is the empirical covariance matrix of the components of Z (Fukunaga, 1972[38]; Silverman 1986[74]). This approach (also called *sphering* the data) is equivalent to pre-scale the data by a linear transformation to have unit

[1]In the univariate case, the bandwidth controls only for the amount of smoothing.

covariance matrix (it avoids extreme differences of spread in the various coordinate directions), next to smooth using the radially symmetric kernel and finally to transform back to obtain the density estimator of the original data. Formally, the scaled kernel density estimate function of reference, in its extend version, is given by:

$$\hat{f}(z) = \frac{1}{n(2\pi)^{d/2}h^d|S|^{(1/2)}} \sum_{i=1}^{n} exp\{-\frac{1}{2h^2}(Z_i - z)'S^{-1}(Z_i - z)\} \tag{4.11}$$

where n is the number of set points to evaluate and d the dimension of the density.

At the end, it follows that the smoothing parametrisation for the scaled bandwidth matrix H requires **only** for the selection of the bandwidth, as in the univariate case. On turn, the selection of the bandwidth parameter h requires the definition of approach:

- h constant or *global* bandwidth;
- $h(z)$ *local variable* bandwidth ;
- $h(Z_i)$ *global variable* bandwidth.

So, it stems that the selection of the bandwidth is a stone step in Bootstrap Procedure as in usual kernel density estimation, since it strongly influence the shape of the estimator $\hat{f}(z)$. As Silverman (1986)[74] discussed in his monograph, small value of h make the estimate look "'wiggly"' and shoe spurious features, whereas too big values of h will lead to an estimate which is too smooth in the sense that it is too biased and may not reveal structural features, like for example bimodality, of the underlying density f. Because of such reasons, lot of research was done in the last years to find objective, data-driven bandwidth selection methods. In Section (6.2) and (6.3), a set of data driven methods are chosen to be applied in th context of Malmquist Indices estimates whereas here I will discuss the ***Normal Reference bandwidth selection***. Indeed, it is the starting point of my research , regards which I propose some improvements in the forthcoming Chapters.

The normal reference rule of thumb is based on a "plug-in" approach as it derives the optimal bandwidth h making reference to a normal distribution of the density

function. In order to select an "optimal bandwidth", we need to choose a way to measure the closeness of an estimator $\hat{f}_Z(\cdot;h)$ to its target density f_Z. A commonly-used criterion is the Mean Integrated Squared Error (MISE), defined as:

$$MISE\{\hat{f}_Z(\cdot;h)\} = E\int\{\hat{f}_Z(z;h) - f_Z(z)\}^2 dz \tag{4.12}$$

where $f_Z(z)$ is a generic parameter and $\hat{f}_Z(z;h)$ a generic estimator. The optimal bandwidth, h_{MISE}, is then defined as the minimizer of (4.12) quantity with respect to h. By far, it is the more tractable global measure as it can be expressed as:

$$MISE\{\hat{f}_Z(\cdot;h)\} = E\int\{E\hat{f}_Z(z;h) - f_Z(z)\}^2 dz + \int var\hat{f}_Z(z;h) \tag{4.13}$$

which gives the MISE as the sum of the integrated square bias an the integrated variance. The ideal value of h is given by the minimization of this quantity MISE,and so, of the sum of the two terms: bias and variance. On the other hand, the above expression however depends on unknown quantities involving $_Z$, and thus is of no direct practical use for selecting the bandwidth. This shortcoming can be overcome by using the asymptotic approximated MISE (AMISE). Some algebraic manipulations, based on the multidimensional form of Taylor's Theorem and asymptotic properties (see for details Silverman 1986 [74]), yield the approximations for the interested quantities:

$$bias_(z) \approx \frac{1}{2}h^2\alpha\nabla^2 f(z) \tag{4.14}$$

$$var\hat{f}(x) \approx n^{-1}h^{-d}\beta f(z) \tag{4.15}$$

where α and β are two constants. Combining them, the AMISE(h) is given by:

$$AMISE(h) \approx \frac{1}{4}h^4\alpha 2\int\{\nabla^2 f_Z(z)\}^2 dz + n^{-1}h^{-d}\beta \tag{4.16}$$

Hence, the approximately optimal window width, in the sense of minimizing mean

integrated square error, is given by:

$$h_{opt}^{d+4} = d\beta\alpha^{-2}\left[\int(\nabla^2 f_Z(z))^2\right]^{-1} n^{-1} \tag{4.17}$$

From (4.17), it can be observed that the approximately optimal h converges to zero as n increases, but does so extremely slowly, at the rate of $n^{-\frac{1}{d+4}}$ and the appropriate value depends on the unknown density being estimated. Supposing $f_Z(z)$ a standard density, such as the multivariate normal, the optimal bandwidth is given by:

$$h_{opt}^{d+4} = \left[\frac{4}{(2d+1)}^{1/(d+4)}\right] \tag{4.18}$$

The cautionary remark is that if the Fukunaga estimate (4.11)is being used, then the h_{opt} of (4.18) gives a directly appropriate value for the smoothing parameter. In conclusion, the choice of the bandwidth by means of the "Normal Reference Rule" appears to be optimal in the sense of minimizing the AMISE of $\hat{f}z$, so that we can call it *quick and simple* bandwidth selectors, oppositely to the DATA-driven methods. The *quick and simple* bandwidth selectors depend on the true density $f_Z(z)$, assumed to be a bivariate normal density function and does not take into account the structure of the data.

4.3 The Basic Bootstrap Algorithm

In this section, I will introduce the basic bootstrap algorithm proposed by Simar and Wilson (1999b) [78] where the density function estimate is carried out adopting the ***normal reference rule of thumb***, fitted on the Malmquist indices estimate setting. The aim of the procedure is to find out the bootstrap empirical distribution of the Malmquist indices MPI and its components ΔEff and $\Delta Tech$, starting from the sole available information, provided by the original data set defined as:

$$\chi = \{(x_{it}, y_{it}) \mid \quad i = 1, ..., n; t = t_1, t_2\} \tag{4.19}$$

The algorithm entirely works on a virtual scenario, the *bootstrap scenario*, and it is initialized by means of the estimators of the input distance functions, computed from the original data set (4.19). The initialization phase marks the shift from the real to the virtual world within which whole computation will be carried out. Practically, the algorithm takes into account the main principles described in section (4.2) and gives attention to the problem of Statistical Modelization of the available data (*i.e.* a bivariate and bounded sample and temporal correlation). So, let the $\{(\hat{\Delta}_i^{t_1,t_1}, \hat{\Delta}_i^{t_2,t_2})\}_{i=1}^n$ the estimators, supposed known in the bootstrap scenario. Using a vectorial notation, for any firm i-th, with $i = 1, ..., n$, at any time $t = 1$ and $t = 2$, the two set of distance functions are:

$$A = [\hat{\Delta}_i^{t_1,t_1}, ..., \hat{\Delta}_i^{t_1,t_1}]' \tag{4.20}$$

$$B = [\hat{\Delta}_i^{t_2,t_2}, ..., \hat{\Delta}_i^{t_2,t_2}]' \tag{4.21}$$

where A and B are nx1 vectors, bounded below the unity and characterized by temporal correlation.

The algorithm runs according the following steps:

1. **Reflection**: the distance function values are reflected about the boundaries. To reflect we have to take the "mirror imagine" of each point about the horizontal and vertical axis, passing thought (1,1). Indeed, departing from the *original* data set $Z = \{\hat{\Delta}_i^{t_1,t_1} \quad \hat{\Delta}_i^{t_2,t_2}\}_{i=1}^n$ localized in the first quadrant (as the boundaries below the unit), the reflection generates three set of points, that jointly with the original one, constitute four symmetrical sets, widespread over the all quadrants. Formally, the four data points set are expressed by a 4nx2 matrix, represented in partitioned form by:

$$\Delta = \begin{bmatrix} A & B \\ 2-A & B \\ 2-A & 2-B \\ A & 2-B \end{bmatrix} \tag{4.22}$$

2. **Estimate of the covariance structures**: the cloud of points generated by refolding is characterized by an estimated covariance structure defined as:

$$\hat{\Sigma} = \begin{bmatrix} \hat{\sigma}_1^2 & \hat{\sigma}_{12}^2 \\ \hat{\sigma}_{12}^2 & \hat{\sigma}_2^2 \end{bmatrix} \tag{4.23}$$

for the sub-matrix of $[A \quad B]$ and $[2-A \quad 2-B]$, whereas as:

$$\hat{\Sigma}_R = \begin{bmatrix} \hat{\sigma}_1^2 & -\hat{\sigma}_{12}^2 \\ -\hat{\sigma}_{12}^2 & \hat{\sigma}_2^2 \end{bmatrix} \tag{4.24}$$

for the matrix of $[2-A \quad B]$ and $[A \quad 2-B]$.

3. **Cholesky decomposition**: do the Cholesky decomposition of the matrix $\hat{\Sigma}$ and $\hat{\Sigma}_R$ to obtain $LL' = \hat{\Sigma}$ and $L_R L_R' = \hat{\Sigma}_R$.

4. **Consistent Bootstrap Estimates**

   – randomly draw with replacement n rows (naive bootstrap) noted as $\Delta_{i.}^*$, from Δ to construct the bootstrap sample, namely $\Delta^* = [\delta_{ij}]$, $i = 1, ..., n, \quad j = 1, 2$

   – generate ε_i^* matrix containing n independent draws from the bivariate kernel function $\mathrm{K}_j(\cdot)$, computed as:

   * $\varepsilon_i^* = L\varepsilon_i$ if $\Delta_{i.}^*$ is drawn from [A B] or [2-A 2-B];
   * $\varepsilon_i^* = L_R\varepsilon_i$ if $\Delta_{i.}^*$ is drawn from [2-A B] or [A 2-B]

   where ε_i is a (2x1) vector given by the i-th draw from the bivariate standard normal distribution.

5. **Matrix of Bivariate Deviates**: Compute the (nx2) matrix of bivariate deviates, $\Gamma = [\gamma_{ij}]$ from the estimated density function of Δ, as:

$$\Gamma = (1+h^2)^{-1/2}\left(\Delta^* + h\varepsilon^* - C\begin{bmatrix} \bar{\delta}_{.1} & 0 \\ 0 & \bar{\delta}_{.2} \end{bmatrix}\right) + C\begin{bmatrix} \bar{\delta}_{.1} & 0 \\ 0 & \bar{\delta}_{.2} \end{bmatrix} \tag{4.25}$$

where C is an (nx2) matrix of ones, which gives which gives an (nx2) matrix of bivariate deviates from the estimated density of Δ, scaled to have the first

and second moment properties observed in the original sample represented by [A B] and $\bar{\delta}_{.j} = n^{-1} \sum_{i=1}^{n} \delta_{ij}$ for $j = 1, 2$ and h is the bandwidth computes as $(4/5n)^{1/6}$.

6. **Distance Estimated Matrix**: Compute the matrix $\Gamma^* = [\gamma^*_{ij}]$ for $i = 1, ..., n$ and $j = 1, 2$ of the simulated distance function values, as:

$$\gamma^*_{ij} = \begin{cases} \gamma_{ij} & if\ \gamma_{ij} \geq 1, \\ 0 & otherwise \end{cases} \tag{4.26}$$

7. **Bootstrap sample**: Derive the pseudo sample $\chi^* = \chi^{*,t_1}, \chi^{*,t_2}$ [2]through:

$$x^*_{it_j} = \gamma^*_{ij} x_{it_j} / \hat{\Delta}_i^{t_j,t_j} \tag{4.27}$$

$$y^*_{it_j} = y_{it_j} \tag{4.28}$$

for $i = 1, ..., n \quad j = 1, 2$.

8. For the selected firm i-th, the distance function estimates and hence the Malmquist Index and its components are computed using the pseudo samples, previously defined.

To obtain the empirical bootstrap distribution of the Malmquist Index and its components, for each firm we repeat the loop from [4-8] B times.

It is worth to note that the consistent multivariate kernel density estimate of the original data set is computed passing through the estimate of the 4n reflected data points, where each point i is represented by the row $(\Delta_{i.})$ of Δ. Hence, the kernel estimator is defined as:

$$\hat{g}(z) = \frac{1}{4nh^2} \sum_{i=1}^{n} K(\frac{z - \Delta_i}{h}) \tag{4.29}$$

where $z = [z_1 \quad z_2]$, and $K_j(\cdot)$ is the bivariate normal density function with shape $\hat{\Sigma}$ for $j = 1, ..., n$ and $j = 2n + 1, ..., 3n$ or shape $\hat{\Sigma}_R$ for $j = n + 1, ..., 2n$ and

[2]If we were using output distance functions and the output-based Malmquist index, we would retain the original input vector in the pseudo sample and generate a new output vector.

$j = 3n + 1, ..., 4n$. Coming back to the original data set, a consistent estimate is obtained truncating the resulting density estimates on the left at unity for the two dimensions. It is defined as:

$$\hat{g}^*(z) = \begin{cases} 4\hat{g}(z) & \text{for } z_1 \geq 1,\ z_2 \geq 1 \\ 0 & otherwise \end{cases}$$

Practically, **we do not need to compute the bivariate density estimate** but only a naive bootstrap of the scores perturbated by a gaussian noise. So, the estimate is carried out drawing n times (doing naive bootstrap) from the data points contained in Δ, so that each row has the same probability of selection and then correcting them for the mean and variance to assure the same statistical structure of the 4n data points.

Chapter 5

Analysing TFP through the b-DEA: the case of Italian regions

5.1 Modern Growth Theory: a briefly history

In this section, I will introduce some important concepts of the *Economic Growth* by means of a briefly excursus on the main authors involved in this process. Classical economists, such as Adam Smith (1776)[83], David Ricardo (1817)[66], and Thomas Malthus (1798)[56], and, much later, Frank Ramsey (1928)[64], Allyn Young (1928)[90], Frank Knight (1944)[47], and Joseph Schumpeter (1934)[71], provided many of the basic ingredients that appear in modern theories of economic growth. These ideas include the basic approaches of competitive behavior and equilibrium dynamics, the role of diminishing returns and its relation to the accumulation of physical and human capital, the interplay between per capita income and the growth rate of population, the effects of technological progress in the forms of increased specialization of labor and discoveries of new goods and methods of production, and the role of monopoly power as an incentive for technological advance.

From a chronological viewpoint, the starting point for modern growth theory is the classic article of Ramsey (1928)[64], a work that was several decades ahead of its time. Ramseys treatment of household optimization over time goes far beyond its application to growth theory; it is hard now to discuss consumption theory, asset pricing, or

even business-cycle theory without invoking the optimality conditions that Ramsey (and Fisher, 1930[36]) introduced to economists. Ramseys intertemporally separable utility function is as widely used today as the CobbDouglas production function. The economics profession did not, however, accept or widely use Ramseys approach until the 1960s.Between Ramsey and the late 1950s, Harrod (1939) [42] and Domar (1946)[26] attempted to integrate Keynesian analysis with elements of economic growth. Although these contributions triggered a good deal of research at the time, very little of this analysis plays a role in todays thinking.

The next and more important contributions were those of Solow (1956)[84] and Swan (1956)[85]. The key aspect of the SolowSwan model is the neoclassical form of the production function, a specification that assumes constant returns to scale, diminishing returns to each input, and some positive and smooth elasticity of substitution between the inputs. This production function is combined with a constant-saving-rate rule to generate an extremely simple general-equilibrium model of the economy. One prediction from these models, which has been exploited seriously as an empirical hypothesis only in recent years, is conditional convergence. The lower the starting level of per capita GDP, relative to the long-run or steady-state position, the faster the growth rate. This property derives from the assumption of diminishing returns to capital; economies that have less capital per worker (relative to their long-run capital per worker) tend to have higher rates of return and higher growth rates. The convergence is conditional because the steady-state levels of capital and output per worker depend, in the SolowSwan model, on the saving rate, the growth rate of population, and the position of the production function characteristics that might vary across economies. Recent empirical studies indicate that we should include additional sources of cross-country variation, especially differences in government policies and in initial stocks of human capital. The key point, however, is that the concept of conditional convergencea basic property of the SolowSwan modelhas considerable explanatory power for economic growth across countries and regions. Another prediction of the SolowSwan model is that, in the absence of continuing improvements in technology, per capita growth must eventually cease.

The neoclassical growth theorists of the late 1950s and 1960s recognized this model-

ing deficiency and usually patched it up by assuming that technological progress occurred in an exogenous manner. Cass (1965) [15] and Koopmans (1965) [48] brought Ramseys analysis of consumer optimization back into the neoclassical growth model and thereby provided for an endogenous determination of the saving rate. This extension allows for richer transitional dynamics but tends to preserve the hypothesis of conditional convergence. The endogeneity of saving also does not eliminate the dependence of the long-run per capita growth rate on exogenous technological progress.

The inclusion of a theory of technological change in the neoclassical framework is difficult, because the standard competitive assumptions cannot be maintained. Technological advance involves the creation of new ideas, which are partially nonrival and therefore have aspects of public goods. For a given technologythat is, for a given state of knowledgeit is reasonable to assume constant returns to scale in the standard, rival factors of production, such as labor, capital, and land. In other words, given the level of knowledge on how to produce, one would think that it is possible to replicate a firm with the same amount of labor, capital, and land and obtain twice as much output. But then, the returns to scale tend to be increasing if the nonrival ideas are included as factors of production. These increasing returns conflict with perfect competition. In particular, the compensation of nonrival old ideas in accordance with their current marginal cost of productionzerowill not provide the appropriate reward for the research effort that underlies the creation of new ideas.

Romer (1986)[68] showed later that the competitive framework can be retained in this case to determine an equilibrium rate of technological advance, but the resulting growth rate would typically not be Pareto optimal. More generally, the competitive framework breaks down if discoveries depend in part on purposive R&D effort and if an individuals innovations spread only gradually to other producers. In this realistic setting, a decentralized theory of technological progress requires basic changes in the neoclassical growth model to incorporate an analysis of imperfect competition.9 These additions to the theory did not come until Romers (1987)[69], 1990[67] research in the late 1980s.

After the mid-1980s, research on economic growth experienced a boom, beginning

with the work of Romer (1986)[68] and Lucas (1988)[55]. The motivation for this research was the observation (or recollection) that the determinants of long-run economic growth are crucial issues, far more important than the mechanics of business cycles; thus, in one way or another, the recent contributions determine the long-run growth rate within the model, designating the *endogenous growth models*.

The initial wave of the new researcher (Romer (1986)[68], Lucas (1988)[55], Rebelo (1991)[65]) did not really introduce a theory of technological change. In these models, growth may go on indefinitely because the returns to investment in a broad class of capital goodswhich includes human capitaldo not necessarily diminish as economies develop. (This idea goes back to Knight, 1944[47]). Spillovers of knowledge across producers and external benefits from human capital are parts of this process, but only because they help to avoid the tendency for diminishing returns to the accumulation of capital. The incorporation of R&D theories and imperfect competition into the growth framework began with Romer and included significant contributions by Aghion and Howitt (1992)[1] and Grossman and Helpman (1991)[41]. In these models, technological advance results from purposive R&D activity, and this activity is rewarded by some form of ex post monopoly power. If there is no tendency for the economy to run out of ideas, the growth rate can remain positive in the long run. The rate of growth and the underlying amount of inventive activity tend, however, not to be Pareto optimal because of distortions related to the creation of the new goods and methods of production. In these frameworks, the long-term growth rate depends on governmental actions, such as taxation, maintenance of law and order, provision of infrastructure services, protection of intellectual property rights, and regulations of international trade, financial markets, and other aspects of the economy. The government therefore has great potential for good or ill through its influence on the long-term rate of growth. This research program remained active through the 1990s and has been applied, for example, to understanding scale effects in the growth process (Jones, 1999)[44], analyzing whether technological progress will be labor or capital augmenting, and assessing the role of competition in the growth process. The new research also includes models of the diffusion of technology. Whereas the analysis of discovery relates to the rate of technological progress in

leading-edge economies, the study of diffusion pertains to the manner in which follower economies share by imitation in these advances. Since imitation tends to be cheaper than innovation, the diffusion models predict a form of conditional convergence that resembles the predictions of the neoclassical growth model. Some recent empirical work has verified the importance of technological diffusion in the convergence process.

5.2 The Human Capital: A strategical source to growth

In this section, I describe the important role oh ***Human Capital*** in the theory of *Economic Growth*. The theoretical models have incorporated human capital as one of the determining factors of development. Thus, in the case of neo-classical growth models, the study by Mankiw, Romer and Weil (1992)[57] offers the generalisation of the Solow model in this line, including a rate of saving in human capital, and offers evidence to confirm its positive contribution to growth, reconciling the empirical evidence with the neo-classical model of exogenous growth.

Apart from the neo-classical growth models, the so-called endogenous growth models have also used human capital in their analyses. The central idea of some of these models consists of generating the growth from the existence of non-diminishing returns on the accumulable factors. This property is sometimes established through externalities, thus maintaining the coherence with a context of perfect competition. At all events, the incorporation of an added type of capital is appropriate, especially if it is a factor to which positive externalities can be attributed, as for example in Lucas (1988)[54].

Another type models derives endogenous growth as a result of the development of new ideas and new products, a process that need have no limits. In Romer (1990)[67] the existence of a sector of the economy dedicated to research and development is the mechanism through which sustained growth is reached, so that human capital is the most highly-qualified candidate for generator of this type of progress and therefore becomes a determinant of the economic growth rate. Indeed, human capital can not only drive innovation, but also contribute significantly to the imitation and adoption by one economy of the techniques previously developed by more advanced countries. This question is not new, this type of phenomena having already been analysed in Nelson

and Phelps (1966)[59] or Welch (1970)[89].

To sum up, there exist very varied theoretical arguments on which to base the idea that a greater endowment of human capital increases the rate of technical progress by encouraging both innovation and the diffusion of technology and new products. In this sense, any measure that increases human capital would be highly recommendable for its effects on the growth rate. Indeed, this diversity of mechanisms by means of which human capital can influence growth may explain to a large extent its success in the literature. This diversity is an aspect that requires more detailed reflection. Firstly, human capital may contribute to growth in a way analogous to any other factor of production such as the amount of labour or physical capital. In this sense, the higher the level of human capital, ceteris paribus, the greater the production. This, then, is a level effect of human capital as a consequence of which a growth of human capital will generate economic growth. This is the type of effects that are usually considered by the neo-classical growth models and there exists both positive and negative evidence in this respect. Human capital may also contribute to technical progress by driving both innovation and imitation. In this case, the economic growth rate itself will depend on the level of human capital, due to what is called the rate effect of human capital. Endogenous growth models, though not only they, emphasise these aspects. Kyriacou (1991)[50] and Benhabib and Spiegel (1994)[11] point out that this seems to be the channel through which human capital acts, the significance of the level effect being non-existent or debatable. The evidence offered by Barro and Lee (1994)[8] and Engelbrecht (1997)[29] indicates the existence of both types of effects. In general, the results seem to be sensitive to the specification employed, as well as to the indicator of human capital used.

Human capital is also relevant from another standpoint, more concerned with the discontinuity of the processes of development and the existence of poverty traps. These are situations in which for different reasons, e.g. the inability of an economy to access the most developed technologies by itself, a long-term equilibrium with higher per capita income is impossible. The evidence offered by Kyriacou (1991)[50], Benhabib and Spiegel (1994)[11] and Taskin and Zaim (1997)[86] seems to indicate that there are sig-

nificant differences in the configuration of growth when countries are analysed in groups according to level of development. Their results suggest that the incidence of human capital seems to depend of the degree of development attained, driving innovation in the developed countries and technological catching-up in the poorer ones.

5.3 The Italian Case: A Literature Overview

The analysis of Italian economic growth at regional level has been extensively examined in the context of the well known convergence debate between Northern and Southern regions. So far, the convergence process has been widely dealt with using a parametric approach (*e.g.* Barro and Sala-i-Martin,1991[9]; Sala-i-Martin, 1996[70]; Di Liberto, 1994[25]; Paci and Saba, 1998[60]; Carmeci and Mauro, 2002[14]; Aiello and Scoppa, 2005[2]), without identifying its main sources. The only exception, to the best of our knowledge, is Leonida et al. (2004)[51] which estimate the sources of Total Factor Productivity (TFP) growth of Italian regions over the period 1970-1995 by means of the Malmquist Productivity Index decomposition (Färe et al., 1994)[33], using a nonparametric approach based on Data Envelopment Analysis (DEA). However, recent theoretical papers highlight several shortcomings of the DEA methodology (see for a detailed analysis Daraio and Simar, 2007)[20], among which there is the bias in the estimates and the difficulty of carrying out statistical inference in a multivariate nonparametric set up.

The aim of this research[1] is to re-estimate and update TFP growth and its determinants (Efficiency Change and Technical Change) of Italian regions using a rigorous procedure, based on the bootstrap method proposed by Simar and Wilson (1999)[78] for DEA estimations. It allows for the estimation of bias-corrected MPI (and its components) and permits the construction of confidence intervals on the estimated measures. Moreover, we also include human capital in the analysis since it is widely recognized that it generally plays a central role in the economic growth (*e.g.* Del Barrio-Castro et al., 2002[23]; Bassanini and Scarpetta, 2002[10]; Maudos et al., 1999[58]) and also for

[1] The part which follows has been taken from a Discussion Paper "Analysing Total Factor Productivity through a bootstrap-based nonparametric approach: the case of Italian regions" written by Bonaccorsi A., Curi C., Daraio C., Mancuso P., Simar L. and published at the Institute de Statistique, UCL and CORE, Belgium.

the Italian regions (Ascari and Di Cosmo, 2004[3]).

Following the research proposed by Piselli and Bronzini, (2006)[63] in which they estimated the long-run relationship between regional TFP, RD,human capital and public infrastructure between 1980 and 2001, we use the same sample of 19 Italian regions, exploiting the database provided by Bank of Italy. The regions are geographically classified into three main areas (North, Centre and South), according to the Italian Statistical Institute (Istat). One region, Valle d'Aosta, is excluded due to lack of data. The variables used in the analysis are: 1) aggregate output measured by the Value Added (VA), expressed at constant price; 2) aggregate labour measured by total employment; 3) Private Capital Stock calculated using the method introduced by Piselli (2001)[62], which breaks down national capital stock by region, sector and type of capital good; and 4) human capital stock calculated by Bronzini and Piselli (2006)[63]. In particular, in order to calculate regional private capital stock we have used the method introduced by Piselli (2001)[62], whose methodology breaks down national capital stock by region, sector and type of capital good. Piselli's procedures can be divided into four steps: 1) regional gross investment is disaggregated by sector and type of capital good; 2) the national capital stocks, by sector and type of capital good, are split by region in a benchmark year; 3) regional stocks are calculated over the entire period with the capital stock in the benchmark year, the annual investment and the depreciation of capital; 4) total regional capital stocks are calculated by adding sectoral and by-type regional stocks. As far as human capital stock is concerned, it is be approximated by the average years of schooling of employees. In particular, the years of schooling in each region are obtained by the number of years required to reach a certain level of qualification (see below), weighted by the share of workers with that qualification to total employees:

$$\text{Average years of schooling}_R = \frac{1}{N} \sum_Q w(R,Q) \times YS(R,Q) \tag{5.1}$$

with $w(R,Q) = \frac{N(R,Q}{n(R,Q)}$ where $n(R,Q)$ is the number of individuals of the sample in region R with qualification Q, and $N(R,Q)$ is the total number of employees in the region with qualification Q; YS is the years of schooling per employee with qualification Q in region R; the weights w are provided in the survey. Data are from Istat (Indagine sulle forze

lavoro). Before 1993, data are not homogeneous with the present survey and data on age and qualification of employees are taken from the reconstructed dataset by Baffigi (1996)[4]. From 1993 to 2001, we attribute 0 to a person with no qualification, 5 for completing primary school, 8 for lower secondary school, 10.5 for a professional diploma, 12.5 for people completing secondary education, 15.5 for a short degree (laurea breve), 17.5 for a standard degree and finally 21.5 years of schooling to those with a doctoral qualification or specialization. Before 1993, we have only three kinds of qualification: up to primary school, lower secondary school, secondary school diploma or more. We assign 5 and 8 years of schooling to the first two qualifications. In the third class, in order to estimate the share of graduates in the third class in each region i (G_i), we use the average shares from the 1981 and 1991 Census regarding the regional population. Hence, we calculate the average years of schooling in region i as $13 \times (1 - G_i) + 17 \times G_i$. Finally, to detect possible breaks in the series over the entire range 1980-2001, we compare the estimates obtained in the two samples in the overlapping year 1993. Differences turn out to be very small, about 1 per cent or less in all regions, regardless of the variable you taken into account. We calculate a correction coefficient based on the ratio of the data in 1993. This coefficient, which differs by region and variable, is applied to the series before 1993.

5.4 Bootstrapping MPI

The basic idea of the approach proposed by Simar and Wilson (1999b)[78] is to estimate bias corrected estimates and confidence intervals at desidered level of significance for MPI and its components, as shown in Section (4.3). It allows assessing whether productivity growth (or decline) is statistically significant or not. This means to determinate if the changes are really determined by the analyzed process or merely inducted by sampling variations. The same holds for the significance of both efficiency and technical changes. The measurement of productivity is done using the decomposition of MPI proposed by Färe et al. (1994)[33], with an output orientation and assuming Constant Returns to Scale (CRS) technology, according to the standard neoclassical growth model. Hence, the MPI between period t and t+1 is defined as the product of Efficiency Change (EffCh)

and Technical Change (TechCh), as follows:

$$MPI_i^{t_1} = \frac{\Delta_i^{t_2|t_2}}{\Delta_i^{t_1|t_1}} \times \left(\frac{\Delta_i^{t_2|t_1}}{\Delta_i^{t_2|t_2}} \times \frac{\Delta_i^{t_1|t_1}}{\Delta_i^{t_1|t_2}} \right)^{1/2} i = 1, ..., N \tag{5.2}$$

where $\Delta_i^{t_j|t_j}$ is the Shepard (1970) output distance for a generic region $i-th$ at time t_j, relative to the technology existing at time t_j; it is defined (see Section 3.1) as:

$$\Delta_i^{t|t} = inf\{\theta : (x^t, y^t/\theta) \in \Psi^t\} = \{\theta : (x^t, \theta y^t) \in \Psi^t\} \quad i = 1, ..., N \tag{5.3}$$

where Ψ^t is the technology corresponding at time t which models the transformation of inputs $x^t \in \mathbb{R}^p$, into the outputs $y^t \in \mathbb{R}^q$ and θ is the technical efficiency score; and

$$\Delta_i^{t_{j+1}|t_{j+1}} = inf\{\theta : (x^t, y^t/\theta) \in \Psi^{t+1}\} = \{\theta : (x^t, \theta y^{t)\in\Psi^t\} i=1,...,N} \tag{5.4}$$

is the Shepard (1970) output distance for a generic region $i-th$ at time t_j, relative to the technology existing at time t_{j+1}.

For both MPI and its components, values higher than 1 indicate improvements from period t to t+1, values less than 1 indicate regress and values equal to 1 indicate no change. Following the bootstrap-based nonparametric approach by Simar and Wilson (1999b[78], 2000b[80]) the MPI bias is estimated by:

$$BIAS[\widehat{MPI}_i] = B^{-1} \sum_{b=1}^{B} \widehat{MPI}_{i,b}(t, t+1) - \widehat{MPI}_i(t, t+1) \quad i = 1, ..N \tag{5.5}$$

where B is the number of bootstrap replications, set to 2000, $\widehat{MPI}^*_{i,b}(t, t+1)$ is the MPI bootstrapped estimate and $\widehat{MPI}_i(t, t+1)$ the estimate for the generic region $i-th$. The bias correction of the estimate is obtained by subtracting the bias (5.5) from the original estimates. Instead, the main idea behind the construction of bootstrap confidence intervals is to compute the quantiles from the empirical bootstrap distribution of the pseudo estimates, approximating the unknown distribution of $\widehat{MPI}_i(t, t+1) - MPI_i(t, t+1)$ by the known distribution of $\widehat{MPI}^*_i(t, t+1) - MPI_i(t, t+1)$, conditioned on the original Data Generating Process (DGP) where, for a generic region $i-th$, $\widehat{MPI}^*_i(t, t+1)$ is

the MPI bootstrapped estimate, $\widehat{MPI}_i(t,t+1)$ is the MPI estimate and $MPI_i(t,t+1)$ is the real value of MPI which is unknown. Some algebraic transformations (see more details in Simar and Wilson, (1999b)[78], yield an estimated $(1-a)$-percent confidence interval of the real value of MPI for the generic region $i-th$ as:

$$\widehat{MPI}_i(t,t+1) + c_{1-\alpha/2} \leq MPI_i(t,t+1) \leq \widehat{MPI}_i(t,t+1) + c_{\alpha/2} \tag{5.6}$$

where c_α is the α-quantile of the distribution of $\widehat{MPI}_i(t,t+1) - MPI_i(t,t+1)$. It is straightforward to extend this procedure to the MPI's components EffCh and TechCh: just substitute $\widehat{MPI}_i(t,t+1)$ in (5.5) and (5.6) with EffCh (and TechCh).

The main purpose of this approach is to see whether the Bias-Corrected (BC) estimates, namely BC-EffCh, BC-TechCh and BC-MPI, are significantly different from 1. If the interval expressed by (5.6) does not cover 1, then there is a significant change in productivity that can be either a progress (if the lower bound is greater than 1) or regress (if the upper bound is less than 1). A statistically significant BC-EffCh (or BC-TechCh) with bounds greater than 1 is interpreted as evidence of a Catching Up Process towards the frontier (or *Improvements in Innovation*) while with bounds lower than 1 is interpreted as evidence of shifting away from the frontier (*Regress in innovation*).

5.5 Empirical Evidences

Table 1 shows the annual geometric means of Bias Corrected (BC)-MPI and its components (BC-EffCh and BC-TechCh), for each pair of consecutive years in the sample. In the following we report only the bias corrected estimates, since according to Efron and Tibshirani (1993)[28], we have to correct for the bias when the bias in absolute value is higher than of the standard deviation of the estimates; and indeed in all our analysis this is the case.

The results show an overall annual productivity gain (measured by the MPI) of 2.1 percent over the analyzed period since most regions (except Liguria and Toscana) exhibit progress in both components of MPI . Deeply, the results suggest that improvement in productivity is achieved mainly by the innovation driving force (BC-TechCh) instead of

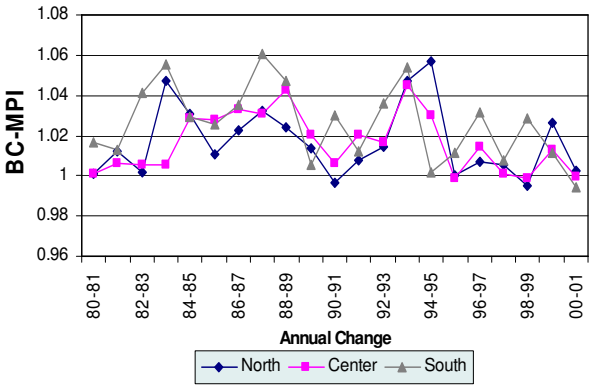
Region	BC-EffCh	BC-TechCh	BC-MPI
Piemonte (PIE)	1.005	1.022	1.026
Lombardia (LOM)	1	1.013	1.007
Trentino-Alto Adige (TAA)	1.002	1.017	1.019
Veneto (VEN)	1.002	1.024	1.025
Friuli-Venezia Giulia (FVG)	1.006	1.019	1.025
Liguria (LIG)	1.006	0.988	0.992
Emilia Romagna (ERO)	1.001	1.027	1.027
Toscana (TOS)	0.998	0.999	0.995
Umbria (UMB)	1.005	1.023	1.027
Marche (MAR)	1.009	1.027	1.034
Lazio (LAZ)	1	1.011	1.009
Abruzzo (ABR)	1.006	1.018	1.024
Molise (MOL)	1.013	1.017	1.03
Campania (CAM)	1.01	1.019	1.028
Puglia (PUG)	1.005	1.02	1.024
Basilicata (BAS)	1.015	1.017	1.032
Calabria (CAL)	1.015	1.017	1.032
Sicilia (SIC)	1.006	1.017	1.022
Sardegna (SAR)	1	1.017	1.016
Overall (geometric mean)	1.005	1.016	1.021

Table 5.1: *BC-MPI and its components. Annual Average Change 1980-2001*

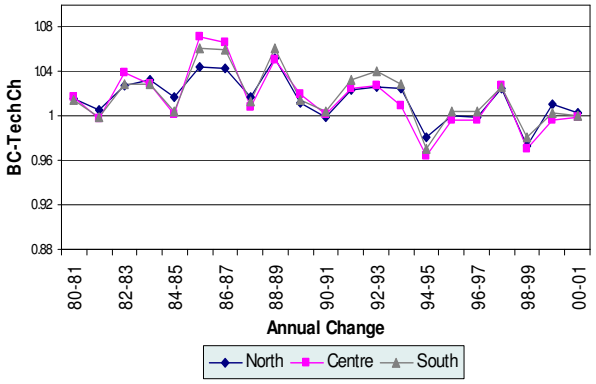
catching up (BC-EffCh). Indeed, the technological progress (upward shift of the frontier) shows an overall annual growth rate of 1.6 percent while the efficiency gain (movement towards the frontier) only of 0.05 percent.

Figure 1 shows the trends of annual geometric means of the estimated MPI and its components for Northern, Central and Southern areas. As far as the efficiency change is concerned, the Southern regions are the best performers of the sample with 0.86 percent average annual change while Northern and Central regions exhibit average annual change, respectively, of 0.33 and 0.34 percent. As far as the technical change is concerned, again Southern regions achieve better results with an annual average rate of 1.80 percent (only in 1994-1995 and 1998-1999, they experienced a regress), followed by Northern regions with 1.60 percent and Central regions with 1.50 percent. Finally, as far as the Malmquist Productivity Index is concerned, all Italian regions experience improvements in productivity and their ranking is unchanged compared with the former ones. Over the period, Southern regions exhibit productivity gains of 2.61 percent, whereas Northern and Central regions of 1.69 and 1.64 percent, respectively. It is worth

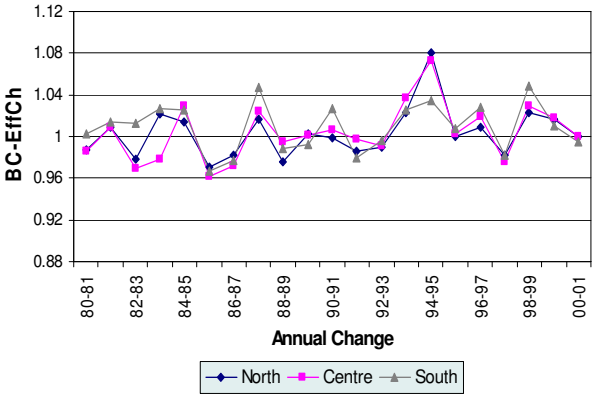
to note the complementarity between the evolutions of BC-EffCh and BC-TechCh. Indeed, higher gains in BC-EffCh (i.e. between 1987-1988 and 1994-1995) are achieved in correspondence to highest lost in innovation (for the same pairs of consecutive years) and vice versa. This may suggest that Italian regions have experienced new technologies, without improving their efficiency by using them.



(a) *Bias-corrected MPI by geographic areas.*



(b) *Bias-corrected TechCh by geographic areas.*



(c) *Bias-corrected EffCh by geographic areas.*

Figure 5.1: Annual growth evolution over the period 1980-2001. Geometric means by areas are reported.

More accurate statistical insights are obtained for each region from the distribution of the bootstrapped estimates of MPI and its components. The distribution of bootstrap estimates, in fact, is used to construct confidence intervals on MPI and its components according to equation (5.6) described above. Hence, it allows the assessment of whether efficiency, technical and productivity change are statistically different from unity. For the sake of brevity, Figure 2 discloses information on the bootstrapped distribution of regional estimates for three annual changes: 1980-1981, 1990-1991 and 2000-2001.

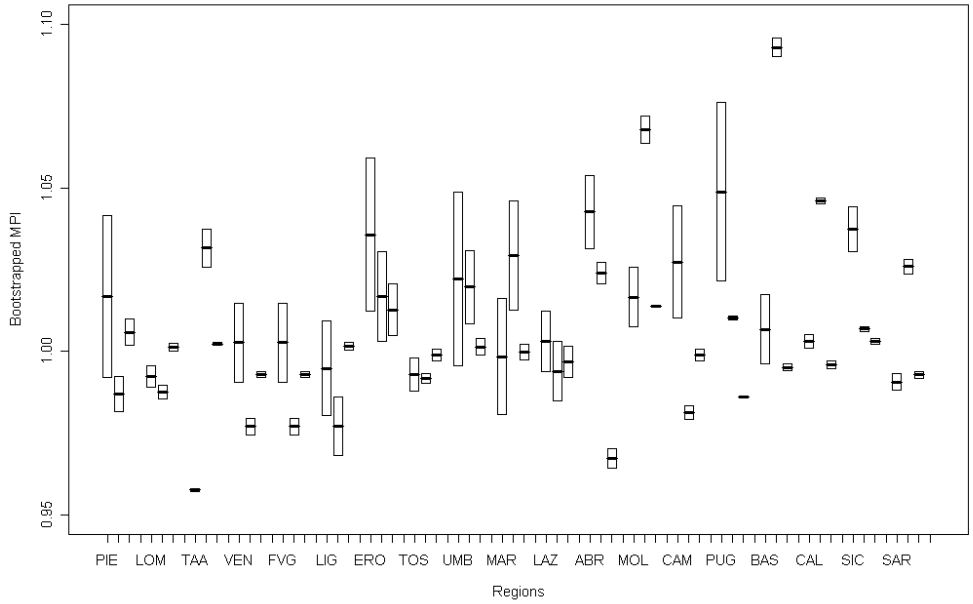


Figure 5.2: *Boxplots of bootstrapped MPI. For each Italian region, the following three pairs of years are reported: 1980-1981 (first box), 1990-1991 (second box) and 2000-2001 (third box)*

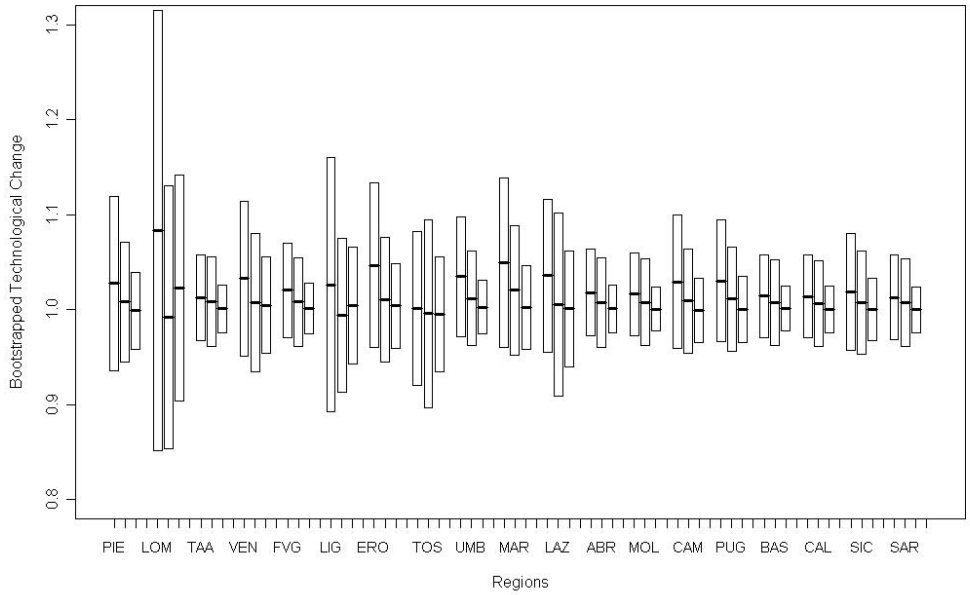


Figure 5.3: *Boxplots of bootstrapped TechCh. For each Italian region, the following three pairs of years are reported: 1980-1981 (first box), 1990-1991 (second box) and 2000-2001 (third box)*

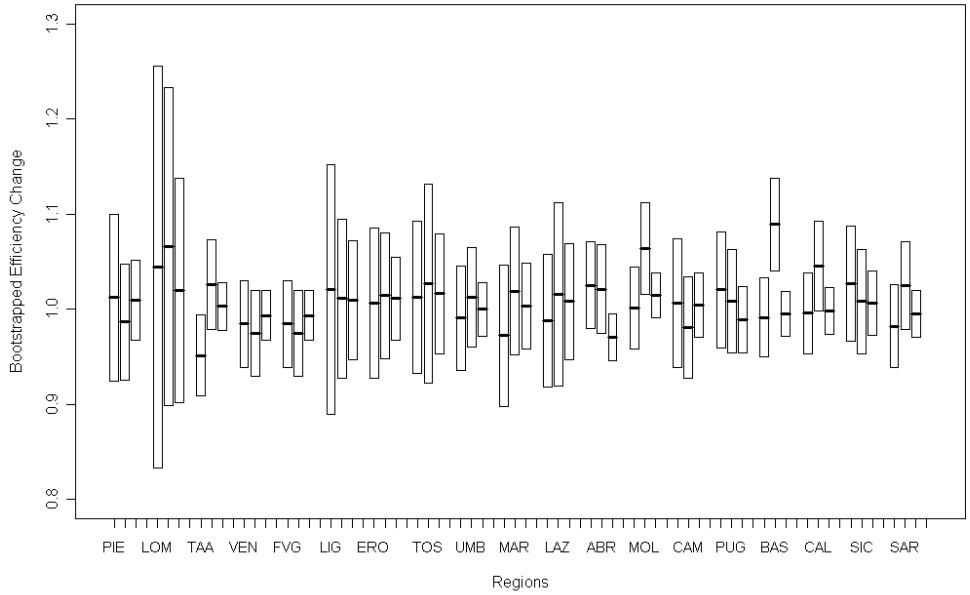


Figure 5.4: *Boxplots of bootstrapped EffCh. For each Italian region, the following three pairs of years are reported: 1980-1981 (first box), 1990-1991 (second box) and 2000-2001 (third box)*

Each box represents the interquartile range (IQR), in which the 0.75 and 0.25 quar-

tiles define the upper and lower segments, respectively. Boxplots of EffCh and TechCh show that IQRs vary around 1, revealing steady levels and isolated cases of statistical significant changes. Looking at the efficiency changes, Trentino Alto Adige (TAA) and Abruzzo (ABR) show significant regress in 1980-1981 and 2000-2001, respectively, and only Basilicata (BAS) shows a significant improvement in 1990-1991. Most regions reveal a density fairly concentrate, instead Lombardia (LOM) and Liguria (LIG) show high dispersion. Looking at the technical changes, a more concentrate density points out the persistence of the lack of innovation over the period. A different picture emerges from the boxplots of MPI. In 1980-81, 47% of regions show no change in productivity against 32% which experience productivity gain and 21% of regions which experience productivity lost. In 1990-91, most regions (58%) experience a gain in productivity while in 2000-2001 only the 21% of Italian regions shows a productivity gain. These descriptive comments are confirmed by the confidence intervals constructed following equation (4). Moreover, confidence intervals of MPI are narrower than those of EffCh and TechCh, indicating more accuracy in the estimation. For all the considered pairs of years, a statistically significant progress is achieved by Emilia Romagna (ERO) and Molise (MOL) whereas a statistically significant regress is experienced by Toscana (TOS).

In conclusion, we applied a rigorous statistical procedure based on the bootstrap in a DEA framework to accurately re-estimate and update the measurement of the overall productivity growth of Italian regions, including human capital among inputs. We found a bias corrected improvement of overall productivity of 2.1 percent over the period 1980-2001, mainly driven by technological improvements in production possibilities (overall annual change: 1.6%), whereas efficiency improvements remained fairly constant (overall annual change: 0.5%). The bootstrap procedure allows us a more careful analysis at regional level. It has confirmed that only a few Italian regions have a Malmquist Productivity Index significantly different from 1 at a significant level of 95%. The percentages of regions which experienced progress are 32%, 58% and 21%, respectively, for the three annual changes (1980-1981, 1990-1991 and 2000-2001). As far as the technical change is concerned, most Italian regions exhibited a significant persistence of lack of innovation. Concerning the efficiency change, isolated cases of significant regional im-

provement are identified (i.e. Basilicata) as well as of regional regress (i.e. Trentino Alto Adige and Abruzzo). According to these results, the inferential approach applied in the paper has provided a more rigorous and accurate insights on the Italian regional TFP than the traditional Data Envelopment Analysis (DEA) estimation carried out by Leonida et al.(2004 ,Table 1, pg. 2190) in which all the estimated values are interpreted as progress or regress without taking into account the bias of the estimated values and their statistical significance.

Chapter 6

Conclusions

The current interest in productivity overall in last decades is due to the revival of the economic professions's interest in the source of economic growth. At the center of the revived debate are productivity, technical change and more recently deviations from the best practice-which we can refer to generally as inefficiency variously attributed to corruption, bad institutions, distorted incentives, market power, and so on. The motivation underlying the matching of the Growth Accounting Theory and the Malmquist Indices Theory is based on the fact that finding sources is deep-stated in economics and Malmquist Index Framework provides an intuitive appeal of identification of sources of productivity growth in terms of "catching up" and innovation, using a simple implementation of nonparametric linear programming techniques.

In the current literature, it is evident a *gap* between: huge variety of contributions on different decompositions on the Malmquist Productivity Indices with their empirical applications (*e.g.* for an overview: Färe, Grosskopf and Russell, 1998[34]; Grosskopf, 2003[40]; Lovell, 2003[53]) and few attention devoted to the inferential aspects of the estimated effects, proposed by the first time by Simar and Wilson, (1998)[76].

In our work we have focused on the inferential aspects of the estimated Malmquist Indices, trying to enrich the debate and propose methods of estimating and bootstrapping Malmquist Indices by means of procedure at highest quality performance.

In particular, we have focused our research on the problem of "bandwidth specification" in the context of the efficiency density estimates. The issued objective has been

twofold: firstly, to investigate, for the first time, the coverage rate of the Simar and Wilson method based on the specification of the bandwidth according the "normal rule of thumb", secondly, to find out further methods to improve the quality of the estimate. Moreover, we have exploited the recent developments proposed by Simar and Wilson in the univariate case about the spurious mass at one when smoothing the distribution of the DEA efficiency score. Indeed, it deteriorates the estimation of the continuous density of the efficiency scores, in the multivariate case more than in the univariate case.

The results of our research have been very encouraging and quite new in the efficiency literature. Indeed, starting from the reference procedure, we have identified two improved methods that sound to overperformance the old one.

The first method consists in correct the moments on the estimated mean and covariance matrix, obtained by deleting all the ones, whereas the other one is based on the elimination of the ones in all the process of density estimation. The Monte Carlo experiments highlighted poor coverage rates of the Simar-Wilson (1999) method as well as for the Adapted Method Simar-Wilson (2007) method with one output; the latter being slightly better with two outpts and for small n. On the other hand, higher coverage rates have been found out for the new method (even in the worse cases of small sample sizes): the achieved coverages (except for the Malmquist index) are of the same order of what is generally obtained in bootstrapping DEA scores. Moreover, further consideration can be drawn. In particular, Malmquist Index is more difficult to estimate than its components. This is not a surprising result since its is the product of the two other components. Indeed, the Malmquist index is build from an estimate of the shift of the efficient frontier and then from estimates of distances to these two frontiers. The new method is always significantly better than the two other methods. No method (all being based on "rule of thumb" bandwidth selection) seems to improve significantly when n increases above, say, $n = 100$. This has indicated a need of exploiting the more elaborate techniques for selecting the bandwidth (data-driven methods).

In this phase of the research, we have adapted some methods to our framework, finding out different values in the estimates, opening a huge research room for our

investigation, of course, more computer intensive (selection of the bandwidth for each Monte-Carlo sample by cross-validation).

In order to reconcile traditional approaches to the analysis of economic growth determinants with the frontier productivity measurement literature, we have employed the new improved method to handle the Italian Regional productivity analysis. Starting from data kindly given by the Bank of Italy, we have deeply investigate the Total Factor Productivity (TFP) growth of the Italian regions over the period 1980-2001. The Malmquist Productivity Index (MPI) and its components (namely Efficiency Change and Technical Change) as well as confidence intervals have been estimated by applying the the developed procedure proposed in this dissertation. Including human capital among inputs, we estimated an overall bias-corrected productivity gain of 2.1 percent, an efficiency gain of 0.5 and a technical gain of 1.6 percent.

The bootstrap analysis revealed that for most Italian regions efficiency and technical changes did not show a statistically significant change. According to these results, the inferential approach applied in the paper has provided a more rigorous and accurate insights on the Italian regional TFP than the traditional Data Envelopment Analysis (DEA) estimation carried out by Leonida et al.(2004 ,Table 1, p. 2190) in which all the estimated values are interpreted as progress or regress without taking into account the bias of the estimated values and their statistical significance.

Bibliography

[1] Aghion P., H. P. A model of growth through creative destruction. *Econometrica 60*, 2 (1992), 323–351.

[2] Aiello, F., and Scoppa, V. Convergence and regional productivity divide in italy: Evidence from panel data.

[3] Ascari, G.and Di Cosmo, V. Determination of total factor productivity in the italian regions.

[4] Baffigi, A. Archivio con serie regionali e provinciali tratte dall'indagine sulle forze del lavoro. *Bank of Italy* (1996).

[5] Balk, B. Scale efficiency and productivity change. *Journal of Productivity Analysis*, 15 (2001), 153–183.

[6] Banker, R. Maximum likelihood, consistency and data envelopment analysis: A statistical foundation. *Management Science 10*, 39 (1993).

[7] Banker, R., Charnes, A., and Cooper, W. Some models for estimating technical and scale inefficient in data envelopment analysis. *Management Science 30*, 9 (1984), 1078–1092.

[8] Barro, R., and Lee, J. Sources of economic growth. *Carnegie-Rochester Conference Series on Public Policy*, 40 (1994), 1–46.

[9] Barro, R., and Sala-i Martin, X. Convergence across states and regions. *Brooking Papers on Economic Activity*, 1 (1991), 107–158.

[10] Bassanini, A., and Scarpetta, S. Does human capital matter for the growth in oecd countries? a pooled mean-group approach. *Economics Letters*, 74 (2002), 399–405.

[11] Benhabib, J., and Spiegel, M. The role of human capital in economic development: evidence from aggregate cross-country data. *Journal of Monetary Economics*, 34 (1994), 143–173.

[12] Bjurek, H., Hjalmarsson, L., and Forsund, F. Deterministic parametric and nonparametric estimation of efficiency in service production: A comparison. *Journal of Econometrics*, 46 (1990), 213–228.

[13] B.W., S., and Young, G. The bootstrap: Smooth or not to smooth? *Biometrika 74* (1987), 469–479.

[14] Carmeci, G., and Mauro, L. The convergence of the italian regions and unemployment. theory and evidence. *Journal of Regional Science 3*, 42 (2002), 509–32.

[15] Cass, D. Optimum growth in an aggregative model of capital accumulation. *The Review of Economic Studies 32*, 3 (1965), 233–240.

[16] Charnes, A., Cooper, W., Lewin, A., and Seiford, L. *Data Envelopment Analysis: Theory, Methodology and Applications*. Kluwer Academic Publishers, Boston, 1995.

[17] Charnes, A., Cooper, W., and Rhodes, E. Measuring the efficiency of decision making units. *European Journal of Operational Research*, 2 (1978), 429–444.

[18] Cooper, W. W., Seiford, L. M., and Tone, K. *Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA-Solver Software*. Kluwer Academic Publishers, Boston, 2000.

[19] Curi, C., Daraio, C., Mancuso, S., and Simar, L. Improving statistical inference on the malmquist productivity index, 2007.

[20] Daraio, C., and Simar, L. *Advanced Robust and Nonparametric Methods in Efficiency Analysis: Methodology and Applications. Studies in Productivity and Efficiency*. Springer, 2007.

[21] Daskovska, A., Simar, L., and Van Bellegem, S. Dynamical analysis of the malmquist productivty index, '2007.

[22] Debreu, G. The coefficient of resource utilization. *Econometrica 19*, 3 (1951), 273–292.

[23] Del Barrio-Castro, T., Lopez-Bazo, E., and Serrano-Domingo, G. New evidence on international rd spillovers, human capital and productivity in the oecd. *Economics Letters*, 77 (2002), 41–45.

[24] Deprins, D., L., S., and Tulkens, H. Measuring labor-efficiency in post offices. In *The Performance of public enterprises:Concepts and Measurement*, P. Marchand, M. andPestieau and H. Tulkens, Eds. Amsterdam, North-Holland, 1974, pp. 243–267.

[25] Di Liberto, A. Convergence across italian regions. *Nota di Lavoro 168, Fondazione ENI Enrico Mattei, Rome* (1994).

[26] Domar, E. D. Capital expansion, rate of growth, and employment. *Econometrica 14*, 2 (1939), 137–147.

[27] Efron, B. Bootstrap methods: another look at the jackknife. *The Annals of Statistics 7*, 1 (1979), 1–26.

[28] Efron, B., and Tibshirani, R. *An Introduction to the Bootstrap*. Chapman and Hall, NY, 1993.

[29] Engelbrecht, H. International rd spillovers, human capital and productivity in oecd economies: An empirical investigation. *European Economic Review*, 41 (1997), 1479–88.

[30] Farrell, M. J. The measurement of productive efficiency. *Journal of the Royal Statistical Society Series A, CXX, Part 3*, 120 (1957), 253–281.

[31] Färe, R., Grosskopf, G., Lindgren, B., and Roos, P. Productivity developments in swedish hospitals: A malmquist output index approach.

[32] Färe, R., Grosskopf, G., Lindgren, B., and Roos, P. Productivity changes in swedish pharmacies 1980-1989: A non parametric malmquist approach. *The Journal of Prodcutivity Analysis*, 3 (1992), 80–101.

[33] Färe, R., Grosskopf, S., Norris, M., and Zhang, Z. Productivity growth, technical progress, and efficiency change in industrialized countries. *American Economic Review 84*, 1 (1994), 66–83.

[34] Färe, R., Grosskopf, S., and Russell, R. *ndex Numbers: Essays in Honour of Sten Malmquist.* Kluwer Academic Publishers: Boston, 1998.

[35] Färe, R., Grosskopf, S., and Zhang, Z. Productivity growth, technical progress, and efficiency change in industrialized countries: Reply. *American Economic Review 87*, 6 (1997), 1040–1043.

[36] Fisher, I. *The Theory of Interest as Determined by Impatience to Spend Income and Opportunity to Invest it.* New York, The Macmillan Company, 1930.

[37] Forsund, F. The malmquist productivity index, tfp and scale. *Memorandum no 233*, School of Economics and Commercial Laws, Unity of Goteborg, Sweden ('1997).

[38] Fukunaga, K. *Introduction to Statistical Pattern Recognition.* Accademic Press, New York, 1972.

[39] Grifell-Tatjé, E., and Lovell, C. A. K. A generalized malmquist productivity index. *Top*, 7 (1996), 101.

[40] Grosskopf, S. Some remarks on productivity and its decompositions. *Journal of Productivity Analysis 20* (2003), 459–474.

[41] Grossman, H. *Innovation and Growth in the Global Economy.* Cambridge, MA, MIT Press, 1991.

[42] Harrod, R. F. An essay in dynamic theory. *Economic Journal*.

[43] Henderson, D., and Russell, R. Human capital and macroeconomic convergence: A production-frontier approach. *Paper presented at the Seventh EWEPA, Oviedo, Spain* (2001).

[44] Jones, F. F. A qualitative study of the managerially controlled attribute of rapid-growth firms.

[45] Kneip, A., Park, B., and Simar, L. A note on the convergence of nonparametric dea estimators for production efficiency scores. *Econometric Theory*, 14 (1998), 783–793.

[46] Kneip, A., Simar, L., and Wilson, P. Asymptotic for dea estimators in non-parametric frontier models. *Discussion Paper 0317, Institute o Statistics, UCLb, 1998* (2003).

[47] Knight, F. *The Journal of Political Economy.* 1944.

[48] Koopmans, T. C. Objectives, constraints and outcomes in optimal growth models. *Econometrica*, 212 (1965).

[49] Korostelev, A., Simar, L., and Tsybakov, A. Maximum likelihood, consistency and data envelopment analysis: A statistical foundation. *The Annals of Statistics*, 23 (1995).

[50] Kyriacou, G. Level and growth effects of human capital. *Working Paper, C.V. Starr Center*, 78 (1991), 91–126.

[51] Leonida, L.and Petraglia, C., and Murillo-Zamorano, L. Total factor productivity and the convergence hypothesis in the italian regions. *Applied Economics*, 36 (2004), 2187–2193.

[52] Lovell, C. Production frontiers and productive efficiency. In *The Measurement of Productive Efficiency:Techniques and Applications*, H. Fried, C. Lovell, and S. Schmidt, Eds. New York, Oxford University Press, 1993, ch. 1, pp. '1–10.

[53] Lovell, C. The decomposition of malmquist productivity indexes. *Journal of Productivity Analysis*, 20 (2003), 437–458.

[54] Lucas, R. On the mechanics of economic development. *Journal of Monetary Economics 1*, 22 (1998), 3–42.

[55] Lucas, R. E. On the mechanics of economic development. *Journal of Monetary Economics 22*, 1 (1988), 3–42.

[56] Malthus, T. *An Essay on the Principle of Population.*

[57] Mankiw, N., Romer, P., and Weil, D. A contribution to the empirics of economic growth. *Quarterly Journal of Economics 2*, 107 (1992), 407–437.

[58] Maudos, J., Pastora, J., and Serrano, L. Total factor productivity measurement and human capital in oecd countries. *Economics Letters*, 63 (1999), 39–44.

[59] Nelson, R., and Phelps, E. Investment in humans, technological diffusion and economic growth. *American Economic Review, Papers and Proceedings 2*, 56 (1966), 69–75.

[60] Paci, R., and F., S. The empirics of regional economic growth in italy 1951-1993. *Rivista Internazionale di Scienze Economiche e Sociali*, 45 (1998), 515–542.

[61] Park, B., Simar, L., and Weiner, C. The fdh estimator for productivity efficiency scores: Asymptotic properties. *Econometric Theory*, 16 (2000).

[62] Piselli, P. Lo stock di capitale nelle regioni italiane: una ricostruzione 1980-1996. *Temi di discussione del Servizio Studi, Bank of Italy* (2001).

[63] Piselli, P., and Bronzini, R. Determinants of long-run regional productivity: the role of rd, human capital and public infrastructure. *Temi di discussione del Servizio Studi, Bank of Italy* (2006).

[64] Ramsey, F. *A Mathematical Theory of Saving.* 1928.

[65] Rebelo, S. Long-run policy analysis and long-run growth. *The Journal of Political Economy 99*, 3 (1991), 500–521.

[66] Ricardo, D. *On The Principles of Political Economy and Taxation.* 1817.

[67] Romer, P. Endogenous technical change. *Journal of Political Economy*, 98 (1990), S7–S102.

[68] Romer, P. M. Increasing returns and long run growth. *Journal of Political Economy 94*, 5 (1986), 1002–37.

[69] Romer, P. M. Growth based on increasing returns due to specialization. *The American Economic Review Papers Proceedings 77*, 2 (1987), 56–62.

[70] Sala-i Martin, X. The classical approach to convergence analysis. *The Economic Journal*, 16 (1996), 1019–1036.

[71] Schumpeter, J. *The Theory of Economic Development.* Library of Congress, 1934.

[72] Sengupta, J. K. An analysis of production as an efficient combination of activities.

[73] Sengupta, J. K. *Dynamics of Data Envelopment Analysis. Theory of Systems Efficiency.* Kluwer Academic Publishers, Dordrecht, 1995.

[74] Silverman, B. *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London, 1986.

[75] Simar, L. Estimating efficiencies from frontier models with panel data: A comparison of parametric, nonparametric and semi-parametric methods with bootstrapping. *Journal of Productivity Analysis 44*, 3 (1998c), 49–61.

[76] Simar, L., and Wilson, P. Productivity growth in industrialized countries. *Discussion paper 9810, Institut de Statistique, UCL and CORE, UCL*.

[77] Simar, L., and Wilson, P. Sensitivity analysis of efficiency scores: How to bootstrap in nonparametric frontier models. *Management Science*, 44 (1998a), 49–61.

[78] Simar, L., and Wilson, P. Estimating and bootstrapping malmquist indices. *European Journal of Operational Research*, 115 (1999b), 459–471.

[79] Simar, L., and Wilson, P. Statistical inference in nonparametric frontier models:the state of the art. *Journal of Productivity Analysis*, 13 (2000a), 49–78.

[80] Simar, L., and Wilson, P. A general methodology for bootstrapping in nonparametric frontier models. *Journal of Applied Statistics 27*, 6 (2000b), 779–802.

[81] Simar, L., and Wilson, P. Nonparametric test of return to scale. *European Journal of Operational Research*, 139 (2002), 115–132.

[82] Simar, L., and Wilson, P. Statistical inference in nonparametric frontier models: recent developments and perspectives. In *The Measurement of Productive Efficiency, 2nd Edition*, H. Fried, C. Lovell, and S. Schmidt, Eds. New York, Oxford University Press, 2007, ch. 4.

[83] Smith, A. *The Wealth of Nations*. Penguin Classics, 1776.

[84] Solow, R. M. A contribution to the theory of economic growth. *Quarterly Journal of Economics 70*, 1 (1956), 65–94.

[85] Swan, T. W. Economic growth and capital accumulation. *Economic Record 32* (1956), 334–61.

[86] Taskin, F., and Zaim, O. Catching-up and innovation in high- and low-income countries. *Economic Letters*, 54 (1997), 93–100.

[87] Thanassoulis, E., and Karagiannis, G. Group-wise decomposition of the malmquist index. *paper presented at the EURO XXII, Prague* (2007).

[88] Vincent, A. *La mesure de la productivite*. Dunod, Paris, 1968.

[89] Welch, F. Education in production. *Journal of Political Economy*, 78 (1970), 35–59.

[90] Young, A. A. Increasing returns and economic progress. *Economic Journal*.

[91] Zofio, J. Malmquist productivity index decompositions: a uniying framework. *Applied economics*, 39 (2007), 2371–2387.

[92] Zofio, J. L., and Lovell, C. A. K. Yet another malmquist productivity index decomposition. *Mimeo*, Department of Economics, Universidad Autonoma de Madrid, Madrid, University of Georgia, Athens.