A RESOURCE-BASED INTERPRETATION OF TECHNICAL EFFICIENCY INDEXES

Eduardo González Fidalgo
Ana Cárcaba García

Abstract

During the last decades, the measurement of technical efficiency indexes has become a very popular field of research. Recent refinements in estimation tools and techniques have contributed to increase the reliability of efficiency analyses. On the other hand, the lack of theory behind the concept of technical inefficiency complicates the interpretation of the empirically estimated indexes. The objective of this article is to outline an interpretation of currently used efficiency indexes within the framework of the resource-based view of the firm. The problems inherent to the definition of technical efficiency as a relevant theoretical concept are examined and, then, the link between technical efficiency and the resource-based view of the firm is discussed.

*JEL Classification:* D24, D23, L2.

*Key words:* technical efficiency, X-inefficiency, resource-based view

---

* This article has benefited from discussion with the participants at the Oviedo permanent seminar in efficiency and productivity analysis.

** Dpt. Admón. Empresas y Contabilidad, Universidad de Oviedo, Av. Cristo s/n 33006, Oviedo (Spain)
Comments are welcome at fidalgo@econo.uniovi.es
1.- INTRODUCTION

Research on efficiency and productivity analysis has been vast during the 1990s. The bulk of the efforts were devoted to further develop the quantitative techniques available for empirical analysis. Many different indexes and computational procedures have been used in an endless list of empirical applications\(^1\). The standard approach undertaken in the empirical literature on efficiency measurement implies fulfilling four steps: 1) collect data on inputs and output(s) from a set of assumed homogeneous decision making units (DMUs), 2) choose the estimation technique—i.e., econometrics or linear programming—that best suits the nature of available data or the type of indexes the author wants to obtain, 3) estimate the efficiency indexes, and 4) explain the indexes through the lens of a second stage regression analysis or analysis of variance. This last step involves searching for variables capable of distinguishing between efficient and inefficient DMUs.

Paradoxically, the huge advance that took place within the measurement field deeply contrasts with the alarming lack of rigorous theoretical background on the notion itself of "technical efficiency". The familiar expression "technical efficiency" refers to a very fuzzy concept, one that is more ambiguous than currently acknowledged. Standard microeconomic theory of production does not even consider the possibility that firm behaviour may be inefficient, at least from a productive or technical point of view\(^2\). Thus, the very concept of technical inefficiency cannot be rationalised with the tools of the neoclassical theory of the firm. It is curious though that the efficiency literature has evolved extremely linked to the analytical framework of the neoclassical theory of production. However, it is common to read efficiency analyses that do not even attempt to briefly discuss the real economic meaning of the indexes reported.

The objective of this article is to explore possible interpretations of the empirically estimated technical efficiency indexes. To accomplish this task, the paper is structured as follows. First, the traditional approach to the concept of efficiency is critically reviewed, including Leibenstein's (1966) theory of X-inefficiency. Then, an

---

\(^1\) See Seiford (1996) for an exhaustive survey of the non-parametric stream of the efficiency measurement literature. Førsund (1999) provides a historical overview of the literature as starting from Farrell's (1957) pathbreaking article.

\(^2\) This is due to the constraints the profit maximisation assumption places on firm behaviour.
alternative resource-based interpretation of empirically estimated indexes of technical efficiency is introduced. This approach is argued to overcome some of the limitations of alternative interpretations. The final discussion suggests the existence of a strong relationship between resources, capabilities, technology, and efficiency.

2.- TRADITIONAL APPROACHES TO THE CONCEPT OF TECHNICAL EFFICIENCY

Following Koopmans (1951), a DMU is technically efficient if and only if it is not possible to increase any of the outputs or to reduce any of the inputs without reducing some other output or increasing some other input. The literature on the measurement of technical efficiency has strongly relied on this definition. Thus, a preliminary step before measuring efficiency indexes is to determine which things are considered to be possible and which things are not. The different techniques at hand characterise the technology by establishing the set of input-output vectors that are considered to be feasible, i.e. the production set. Feasibility is usually established by means of well mathematically defined technological properties. Then, efficiency indexes are obtained by measuring the distance between the observed input-output vector and an empirical feasible benchmark, as defined by the frontier of the production set.

Distances can be measured in many different ways. So is inefficiency, since it is defined as a distance function. The most common approach is to measure the maximum equiproportionate expansion in the observed output vector of the DMU under analysis that is technologically feasible or, alternatively, the maximum equiproportionate contraction in the input vector. Such, radial indexes of technical efficiency were first proposed by Debreu (1951) and Farrell (1957), who developed an algorithm that could be used in empirical estimation. The fact that Farrell’s radial measures have been by far the most widely used in the literature is due to their correspondence with Shephard’s (1953) distance functions, which have a dual interpretation in terms of cost reduction or revenue increase. Other non-radial indexes of technical efficiency are also available\(^3\). Any of these indexes can be

\(^3\) e.g. Bogetoft & Hougaard (1999), Färe & Lovell (1978), Russell (1985).
interpreted as an index of total factor productivity\textsuperscript{4}, with the qualification of being relative measures. This is, they are obtained from multiple comparisons between each DMU and the best observed practices that define the frontier of the production set.

Today, despite there is considerable debate about certain topics in the efficiency measurement literature, it is clear that we know how to measure relative efficiency and we have plenty of techniques and tools available to conduct empirical analyses. In contrast, there is little (if any) guidance regarding \textit{what are we actually measuring?} Empirically estimated indexes reflect the fact that some DMUs seem to perform better than others. But, what's the reason? Putting it in a different way, what is the inefficiency we are measuring, where does it come from? Without answering these basic questions, efficiency measurement would be of little practical use for managerial purposes.

Surprisingly, although efficiency is a central concept in economics and management sciences, it is far from easy to find a theoretical interpretation in standard production theory. The very notion of inefficiency violates central assumptions of economics. Recall, that the neoclassical firm is defined as a very simple production function\textsuperscript{5} that transforms an input vector into an output vector, through an established and well defined technology. Within this theoretical framework, real observed output should always match potential output—the profit maximisation assumption rules out the possibility of resource misuse or suboptimal decision making. When we replace neoclassical assumptions with a more realistic notion of firms the cost minimisation assumption is an oversimplification. Suboptimal decision making and resource waste seem to happen in real production processes. Although traditional production theory has been used to develop the techniques to measure such misperformances it cannot explain why they do occur.

\textsuperscript{4} This point is more evident within the Data Envelopment Analysis (DEA) literature, where the efficiency index is explicitly obtained as a ratio of weighted outputs to weighted inputs.

\textsuperscript{5} Other authors have characterised the neoclassical firm as a black box, a “mysterious thing” (Acs & Gerlowsky, 1996: 146) through which inputs are transformed in outputs in a prestablished way. “It is obvious that in such a theoretical framework there is no place for the firm as an economic institution: the firm is nothing more than an algorithm” (Screpanti & Zamagni, 1993: 372).
3.- X-INEFFICIENCY

The first serious attempt to construct a theory of productive inefficiency is due to Harvey Leibenstein, who clearly departed from neoclassical theory of production. After thorough analysis of empirical evidence, he suggested that, in general, firms do not minimise production costs. According to Leibenstein (1966) inefficiency in production, and not allocative inefficiency, is the principal source of inefficiency in the economy. Unlike Farrell (1957) he did not use the term “technical inefficiency” or “productive inefficiency” to refer to his notion of inefficiency in production. Instead, Leibenstein (1966) coined the term “X-inefficiency” to refer to the amount of forgone output that occurs as a consequence of motivation deficiencies along the firm’s hierarchy.

The hypothesis underlying the notion of X-inefficiency states that the motivation to reduce production costs comes primarily from external pressure. For example, the CEO of a firm in a highly competitive industry would support more pressure to reduce costs than the CEO of a monopoly. This amounts to assume that the reason firms do not maximise profits is because of effort discretion. Effort discretion propagates along the hierarchy because managers do not act in an omniscient way as to minimise costs. Instead, they typically rely on financial reports showing deviations from a priori established targets. Only if these deviations are large enough an energetic response can be expected to control slack. If results are good enough (i.e., as expected or better), the risk that managers start resting in their laurels is considerable, because the motivation needed to search for improvement is scant, even though further improvement may be feasible.

Leibenstein (1966) supported his view that X-inefficiency was an important issue with a large collection of empirical evidence. He cites studies by the International Labour Organisation-ILO (1951, 1956, 1957a, and 1957b) which report cost savings derived from “simple reorganisations of the production process, e.g., plant layout reorganisation, materials handling, waste controls, work methods, and payment by results” (Leibenstein, 1966: 399). More often than not, the savings were as large as 25% of previous production costs. The conclusion is that the nature of the managerial input, the external pressure and the incentive systems have a deep impact on production results. Thus, motivation should be the central theme in the
analysis of the complex set of human and technological relationships that coexist within an organisation.

Leibenstein (1966) stresses the importance of motivation when he analyses the possibility that suboptimal behaviour may be due to a relative lack of knowledge. This explanation of inefficiency would imply that any DMU obtaining more output without using more input than the average do so because it owns a superior knowledge background. Against this interpretation, Leibenstein points out that most of the improvement that seems to be achieved through better knowledge is actually induced by the pressures of motivation. In many instances the knowledge was already there, but not the motivation to exploit it or develop it. Evidence from the ILO productivity missions suggests that, sometimes, managers returned to their old (less productive) techniques, not because they lacked the required knowledge, but because the pressure that motivated the use of the new (and better) techniques had disappeared. In other instances, of course, the motivation may exist but not the knowledge required in order to minimise costs.

The main factors that sustain the theory of X-inefficiency can be summarised as follows (Leibenstein, 1975):

1) Labour contracts are incomplete. It is not possible to completely specify in real contracts all possible contingencies in advance. This implies that an unavoidable degree of effort discretion will be present in the behaviour of workers and managers. Much of this behaviour is then left to custom, authority, moral constraints, incentive systems, and other institutional arrangements.

2) Not all factors of production are marketed. This is an important issue in the case of knowledge. The firm cannot buy all the required knowledge in the optimal quantity at the optimal date. Unfortunately, this was not a central issue in Leibenstein's theory.

3) The production function is not completely specified or known. This is, a given input vector can result in different output vectors, depending on the motivational and organisational schemes. The technology is a complex thing that cannot be represented by a simple functional relationship.
Despite its apparent realism, the theory of X-inefficiency has been strongly criticised. In *The Xistence of X-Efficiency*, George Stigler (1976) firmly criticises Leibenstein's notion of X-inefficiency as an unnecessary and awkward concept. First of all, Stigler denies that motivation has something to do with the quantity of output that is produced by a group of workers. The argument is very simple. Motivation is the same for all individuals: to maximise their respective utility functions. Individuals do not have a particular interest in maximising any output, but in maximising their own utility levels. When the output increase is achieved through a higher effort, efficiency does not improve at all. Rather a different output vector is obtained, one that includes more physical product and less leisure, for instance. But individuals want to obtain the output vector that maximises utility, not the one that maximises physical production. Parish & Yew-Kwang had expressed the same view: "If the monopolist (the inefficient firm) prefers to take it easy, this may just be a form of producers' surplus. Nevertheless, it maybe held that the monopolist is indulging in satisfying non-essential wants. (But...) each man is the best judge of his own interest..." (1972: 302).

On the other hand, it is true that contracts are incomplete, as Leibenstein points out, but a great quantity of (managerial) resources may be required to enforce contractual accomplishment to the point that maximises production (Alchian & Demsetz, 1972). Positive agency theory has called attention upon this fact: the objective of management is not to minimise the residual loss but the sum of all agency costs, which also include formalisation, monitoring, and bonding costs. We cannot seriously argue that a firm is incurring any kind of productive inefficiency when it does not produce the maximum output, given the unavoidable existence of contractual constraints that accompany team production. X-inefficiency would arise if it were possible to produce more at a lower cost and the firm didn't do it. Leibenstein theory falls into the *Nirvana fallacy*, a term coined by Demsetz (1969) to refer to the common practice of comparing the real world with an ideal but non existent world to conclude that the real world is (relatively) inefficient.

Despite the formal elegance of these criticisms, Leibenstein (1978, 1979) did not desist from his view and further developed his theory of X-inefficiency. The contra criticism argued that utility maximisation theory is very attractive but leads to a
dangerous tautology, because, in the limit, it just explains that people do what they do. "There is an extraneous assertion that is added on to the notion that people do what they do, which suggests that when they do what they do they are also maximising utility" (Leibenstein, 1979: 495).

Grounding his theory on the concepts of selective rationality—individuals select the extent to which they deviate from maximising behaviour—and "inert areas"—in which the individual is located once her effort level has been chosen and remain until the external stimuli is enough to counteract the cost of moving to a different "effort position"—, Leibenstein asserts that "...the deviation between the optimal levels of effort from the firm's viewpoint and the actual level that individuals are motivated to put forth determines the degree of X-inefficiency in the system" (1978: 204). Thus, through an appropriate incentive system the utility of all firm members may be increased. However, he does not define what is understood by the "firm's viewpoint", nor where does it come from. Some lines bellow, Leibenstein writes: "Since no one in the firm is presumed to maximise profits, no one is necessarily motivated to try to get the most output from purchased inputs, and hence costs are not minimised. In other words, under this scheme we would expect X-inefficiency to exist" (1978: 205). However, an alternative interpretation would suggest that it seems that under that scheme X-inefficiency is imposed to exist.

The contributions from transaction cost economics, agency theory and property rights theory allow for a generalisation of the neoclassical view of the firm. Within this framework in mind, Leibenstein arguments can be reinterpreted by the statement that individuals react to environmental opportunities and constraints depending on their preferences—i.e., the gain derived from the effort and the gain derived from leisure—and their budget constraints. The budget constraints include the own cognitive ability of people to perceive and scan the state of the environment (De Alessi, 1983; DiLorenzo, 1981). This way, the theory of X-inefficiency can be accommodated within the framework of a more general theory of transaction and/or agency costs, which explicitly considers friction as an essential component of a theory of the firm 6. "Leibenstein's collection of postulates and related variables of X-

---

6 Interestingly, Bogetoft (2000) has developed a model that links agency theory and efficiency measurement showing that DEA based production plans may generate optimal results under
efficiency appears to be a combination of some of the axioms and some of the implications of generalised neoclassical theory” (De Alessi, 1983: 70).

4.- HETEROGENEITY

The theoretical problems associated with the concept of productive inefficiency outlined above have not deterred the growth of empirical studies. Researchers have found important cost differences among firms in almost every sector of the economy and have interpreted these differences as the result of technical inefficiencies. However, empirical measurement refers almost exclusively to relative efficiency indexes, which are obtained through comparisons among firms that are considered to be similar or homogeneous. The relative nature of the efficiency indexes is more obvious in the studies that employ non-parametric techniques. These techniques directly compare the input-output vectors of a set of firms by means of linear programming—under the (strong) assumption that the firms in the sample are comparable, i.e., employ a common technology. More or less the same kind of relative comparisons take place in the applications that estimate a parametric production function, under the (strong) assumption that the function, i.e. the technology, is common to all the firms in the sample. Although, the comparison in this case is conducted in an indirect way, after the production frontier has been estimated.

Although this is the way the empirical literature has developed, it is evident that if some (efficient) firms do better than others (inefficient) it simply happens because "they are different", nor homogeneous as was assumed in the first place. There exist differences among firms that are not registered in accounting states, given the complexity that the evaluation in monetary terms or even the identification of some critical resources would involve. The difference between the firm with the lowest costs and the rest reflects the existence of unobservables that are not being accounted for by the researcher. These differences are what we are commonly calling (relative) “technical inefficiency”. This is equivalent to say that the residual we observe and call inefficiency comes from somewhere, even though we do not know exactly from where—if we did, we would presumably not call it inefficiency. Two conditions of asymmetric information between the principal and the agent, in the sense of minimising
identical firms should always obtain identical results, except for random, and thus unimportant, shocks.

When an engineer asserts that "machine A is more efficient than machine B" he is well aware that he is talking about different (heterogeneous) machines. He does not formulate the *ceteris paribus* clause on the technology, as it is explicitly done in the economic efficiency studies. The word inefficiency is simply employed to summarise differences that exist and that have concrete causes, although sometimes difficult to identify. Leibenstein's attempts to defend a theory of X-inefficiency entirely based in motivational aspects are condemned to incompleteness, because the term just refers to a way of speaking about motivational differences that have concrete sources. The issue of productive inefficiency is, thus, an issue of heterogeneity and therefore investigating a more basic question can approach it more accurately: why are firms different?

5.- A RESOURCE-BASED VIEW OF TECHNOLOGY

The study of technical efficiency has been traditionally formulated on the basis of observable variables—physical inputs and outputs—and assuming an implicit common technology for all the firms that enter the analysis. This implicit technology is an abstraction that represents the possibilities of transformation of physical and observable inputs into physical and observable outputs. But, in reality, the technology (possibilities of transformation, production set) differs across firms even in the same industry, because different firms usually possess some resources and capabilities which are unique, which role is ignored in the estimation of inefficiency. This type of resources includes intangibles, such as knowledge or culture, which are hard to observe, quantify, evaluate, and imitate. Some of these resources are the basis for competitive advantage or disadvantage and, as such, should be also the basis for observed inefficiency indexes.

If we take a resource-based perspective, we must accept Stigler's (1976) explanation of what Leibenstein (1966) called X-inefficiency: if firms obtain different amounts of output from given inputs it is because they are using different transformation technologies. In other words, because they control different sets of the agent's information rents.
intangible resources which are not accounted for in the specification of the efficiency model. These include the firm's knowledge, culture, learning capacity, incentive systems, organisational routines, and other institutional arrangements that evolve over time within the organisation.

The resource-based view of the firm considers that the *ceteris paribus* clause should not be applied across firms, because the level of heterogeneity is typically high. Resource heterogeneity allow different firms to achieve different observable output levels from given observable inputs, generating economic rents that can be sustained from competition (Barney, 1991; Dierickx & Cool, 1989; Lippman & Rumelt, 1982; Peteraf, 1993; Wernerfelt, 1984). The resource-based view provides a satisfactory explanation of the mechanisms that allow for permanent differences in firm performance among direct competitors\(^7\). The control of heterogeneous and hard to imitate resources provides some firms with a competitive advantage.

According to Dierickx & Cool (1989) a useful analytical classification distinguishes *flow* from *stock* resources. Flow resources are those that can be immediately obtained whenever needed. In general terms, flows can be easily identified and a monetary value can be attached to them. Examples of this type of resources are machinery, human force and even market share. In contrast to flows, stock resources generate internally from flows along a period of time through an *accumulation process*. Stocks are idiosyncratic resources deeply embedded in the firm and thus imperfectly mobile. In general it is difficult and, more often than not, impossible to attach stocks a precise monetary value. Dierickx & Cool (1989) point out that a market cannot exist to trade this type of resources\(^8\).

On the other hand, capabilities refer to the firm's ability to accomplish tasks by appropriately combining sets of resources. Scientific terminology has distinguished among *capabilities* (Amit & Schoemaker, 1993; Grant, 1991), *competencies* (Teece, Rumelt, Dosi & Winter, 1994), *core competencies* (Prahalad & Hamel, 1990), and *distinctive competencies* (Andrews, 1971; Ansoff, 1965; Hofer & Schendel, 1978;

---

\(^7\) Industrial organisation, in turn, explains persistent profitability dispersion across industries.

\(^8\) Barney (1989) qualifies this observation by pointing out that the acquisition of stock resources implies incurring costs along a time period. Thus, the distinction between buying in the market and generating internally corresponds mainly to the temporal dimension in which the resource is *bought*. To internally generate the resources, the firm must anticipate their future value. The conditions under which a resource can be traded in the open market are outlined in Chi (1994).
Selznick, 1957). All these terms refer to the same concept, a set of specific firm abilities that are the basis for competitive advantage.

Some characteristics of the accumulation process of stock resources contribute to reduce the degree of imitability and, thus, to the creation and sustainability of rents derived from owning those resources. Among these, two are especially relevant: time compression diseconomies, that retard the possibility of imitating success, and causal ambiguity, that makes it difficult to even identify the causes of success (Lippman & Rumelt, 1982).

Therefore, it is possible to interpret current technical efficiency indexes as indicators of firm heterogeneity. More precisely, observed technical inefficiency arises from heterogeneity in stock resources and essential capabilities that are not included as inputs in the efficiency model. On the other hand, heterogeneity in resources and capabilities can be interpreted as heterogeneity in the underlying technology employed in the firms' production processes. Stigler (1976) had already suggested that the apparent observation of production inefficiencies was the consequence of considering that the firms were using a common or representative technology, while they were in fact using heterogeneous technologies. In the same line, prominent strategy scholars such as Collis & Montgomery point out that "Finely honed capabilities can be a source of competitive advantage. They enable a firm to take the same factor inputs as rivals and convert them into products and services, either with greater efficiency in the process or greater quality in the output " (1997: 29).

Technical efficiency indexes would in fact being measuring distances between different production functions instead of measuring distances between the firm and a hypothetical, but unreal, common production function. That is, given resource heterogeneity, the firms in the sample operate on different production frontiers. A way to relate this view to the theory of X-inefficiency is to consider the different incentive systems employed by the firms as an important part of their production capabilities that shape the actual production technology—the contractual part of the technology.

---

9 Majumdar (1998) has proposed a similar interpretation of the technical efficiency indexes, suggesting the use of DEA to evaluate the differences in firms' capabilities.
Summarising, the inconsistency of the traditional approach to technical efficiency analysis rests on the assumption of a common technology. Taking into account that stock resources and core capabilities are developed through a time consuming and hard to replicate accumulation process seriously challenges the traditional assumption of a common technology. In fact, technical efficiency indexes may be of more practical help if they are interpreted as measures of the relative value of resources and capabilities possessed by the firm but not accounted for in the analysis.

7.- CONCLUSION

Empirically estimated indexes of relative technical efficiency are obtained under the assumption that all the firms in the sample use a common technology. In our view, the homogeneity postulate amounts to assume that all relevant resources and outputs have been taken into account in the efficiency model. This would enable the researcher to perform meaningful interfirm comparisons. However, under the homogeneity assumption no theory is available that explains where do the observed performance differences between otherwise identical firms come from.

The resource-based view of the firm offers a consistent rationale for empirically estimated indexes. The assumption that firms are identical, i.e. that share a common technology and use identical resources as usually stated within the efficiency literature, is unacceptable. Instead, the resource-based view suggests that firms' resources and capabilities are widely heterogeneous, even within the same industry. Critical characteristics of resource accumulation processes, such as time compression diseconomies and causal ambiguity, seriously challenge the assumption of homogeneity. It is precisely resource heterogeneity that enables to explain observed stable differences in total factor productivity.

Empirical analyses of technical efficiency interpret the residual as a measure of inefficiency. But the residual contains two different parts. One of them is statistical noise, product of uncontrollable random shocks. The other part, the one that is usually called technical inefficiency, is in fact a measure of the "error" that the researcher makes when assuming that the firms are homogeneous (comparable). According to Stigler "waste" (inefficiency) is error and “it will not become a useful
concept until we have a theory of error” (1976: 216). We certainly do not have a
theory of error, by we have a resource-based theory that explains most of the
systematic differences captured by the empirically estimated systematic part of the
error term.

Within this framework, estimated indexes of technical efficiency measure the
relative value of resources and capabilities not observed or not included in the
empirical model, that were assumed to be homogeneous across firms. The very fact
that empirical evidence repeatedly shows the existence of large performance
differences in almost every industry analysed, can be interpreted as an excellent test
of resource-based theory. If heterogeneous resources didn't generate competitive
advantages, observed inefficiency would tend to disappear over time, a trend that is
not common in empirical studies. Rather, many "inefficient" firms remain inefficient
because of inferior resources. The empirical observation that some firms always
belong to the higher efficiency groups suggests that their competitive advantages
were sustained along time from competition.

Of course, technical efficiency indexes only measure strictly technical
differences among firms—more precisely, among their productive processes. Competitive advantage has a much wider scope. Resources and capabilities not
directly productive enable the firm to achieve different profitability levels in the
marketplace, by allowing to charge their customers higher prices or to pay lower
prices to their suppliers. Thus, the interpretation of technical efficiency indexes
suggested in this article as the relative value of resources and capabilities, or as a
measure of the competitive advantage of the firm, must be restricted to the strict
productive arena. In this way, we may speak of productive advantage, as a technical
component of competitive advantage.
References


