Abstract

Data Envelopment Analysis (DEA) is an extensively studied, application driven doctrine for analyzing Decision Making Units (DMUs) that consume inputs to produce outputs in a deterministic and quantifiable production process. The essence of DEA is captured in the minimum extrapolation principle, which guides us to envelope the data by the smallest set that satisfies the imposed production assumptions. We thereby argue that the production assumptions lie in the core of DEA, and the specification of assumptions is a determinant practical consideration for all application of DEA. Unfortunately, this important matter is too often confronted with persistent myths, illogical reasoning, or sheer ignorance.

In DEA, the production assumptions typically concern the shape of the production possibility set in terms of such mathematical properties as monotonicity, convexity, and returns to scale. This dissertation discusses the meaning and the interpretation of these properties specifically within the DEA framework. Such aspects as economies of scale and of specialization as well as incomplete or uncertain price information are considered. Our discussion focuses especially on the various convexity properties, which are debated from both technological and economical perspectives. Moreover, some difficulties in the treatment of congestion using the two-stage DEA techniques are pointed out. In addition to the theoretical discussion, we present some readily implementable operational DEA approaches which build more directly on the observed data.

Key Words: Data Envelopment Analysis (DEA), Non-parametric efficiency analysis, Production assumptions, Convexity, Free Disposable Hull (FDH), Economies of scale, Economies of specialization, Price uncertainty

THE ROLE OF PRODUCTION ASSUMPTIONS IN DATA ENVELOPMENT ANALYSIS

Timo Kuosmanen
Helsinki School of Economics and Business Administration
Department of Economics and Management Science
P.O. Box 1210, FIN-00101 Helsinki, **FINLAND**
TEL. +358 9 4313 8537
FAX. +358 9 4313 8535
E-MAIL: kuosmane@hkkk.fi
http://www.hkkk.fi/methods/kuosmanen/

THIS DISSERTATION CONSISTS OF AN INTRODUCTORY SECTION AND THE FOLLOWING ARTICLES:

I. KUOSMANEN, T. (2001): DEA WITH EFFICIENCY CLASSIFICATION PRESERVING CONDITIONAL CONVEXITY, *EUROPEAN JOURNAL OF OPERATIONAL RESEARCH* 132(2), 83-99

II. CHERCHYE, L., T. KUOSMANEN, AND G.T. POST (2001): ALTERNATIVE TREATMENTS OF CONGESTION IN DEA: A REJOINDER TO COOPER, GU, AND LI, *EUROPEAN JOURNAL OF OPERATIONAL RESEARCH* 133(1), 69-74

III. CHERCHYE, L., T. KUOSMANEN, AND G.T. POST (2000): WHAT IS THE ECONOMIC MEANING OF FDH, A REPLY TO THRALL, *JOURNAL OF PRODUCTIVITY ANALYSIS* 13(3), 259-263

IV. KUOSMANEN, T., AND G.T. POST (2001): MEASURING ECONOMIC EFFICIENCY WITH INCOMPLETE PRICE INFORMATION: WITH AN APPLICATION TO EUROPEAN COMMERCIAL BANKS, *EUROPEAN JOURNAL OF OPERATIONAL RESEARCH*, TO APPEAR

V. KUOSMANEN, T., AND G.T. POST (1999): NONPARAMETRIC EFFICIENCY ANALYSIS UNDER UNCERTAINTY: A FIRST ORDER STOCHASTIC DOMINANCE APPROACH, HELSINKI SCHOOL OF ECONOMICS AND BUSINESS ADMINISTRATION, WORKING PAPER W-230

Preface

Now that I have typed the last corrections and refinements to this piece of work, it is time to look back and recall how this dissertation shaped into its present form. At the same time, I have a pleasant opportunity to acknowledge those who have influenced this work and supported my efforts through these years.

The initiative to this dissertation project came from Professor Pekka Korhonen and Professor Jyrki Wallenius who recruited me as a full-time doctoral student to their MCDS - program at Helsinki School of Economics and Business Administration, the Department of Economics and Management Science in 1997. Ever since, Professor Korhonen, the chairman on the department, has been a valuable advisor and supporter for me. As a successful internationally renowned scholar, he has been an excellent example for me to follow. Moreover, he has shown a lot of tolerance as my thesis supervisor, and has allowed me to work independently with my own ideas without putting me too much pressure. Another good academic role model for me has been Professor Jyrki Wallenius, the Vice Rector of our university. I wish to acknowledge with much appreciation the financial support from the Graduate School for Systems Analysis, Decision Making and Risk Management, which has formed the primary source of funding for my dissertation project. Professor Seppo Salo, the former head of our department, also earned my deep appreciation and respect. It is devastatingly sad that he passed away before I had a chance to hand this completed dissertation to him.

Looking back, the excellent background knowledge from my underground studies at the University of Joensuu Economics Department gave me an ideal starting point to this dissertation project. Most notably, Dr. Tarmo Räty is guilty of introducing me to the topics of this thesis in the Econometrics classes on Spring-term 1996, and deepening my knowledge while advising my subsequent Master's Thesis. Tarmo was preparing his own dissertation on the topic at the time, and his genuine fascination of the topic gave the decisive inspiration for me to keep studying this literature. I believe the blueprints of this dissertation began to form already in my undergraduate studies. I became fond of the nonparametric approach since I was fully aware of the difficulties in specifying an appropriate parametric function form for the production function. Moreover, I had already developed healthy skepticism towards the convexity postulate of DEA, which I found insufficiently motivated in most textbooks. (In my Master's Thesis I already applied the non-convex FDH model.) Through these years, I have been in touch with Tarmo on a regular basis. Tarmo has proved a knowledgeable discussant, a good friend, and a valuable source of older JPA papers, to which I am indebted.

In the beginning, the exact subject for my dissertation was not instantly clear, and it took some while before it began to dawn on me. The key insight that eventually lead to the publication of Article I below, and subsequently developed into this entire dissertation, occurred very suddenly and unexpectedly to me in a particularly creative moment during the 5th European Workshop on Efficiency and Productivity Analysis held in Copenhagen on October 1997. I cannot explain how this happened, but the plan for this dissertation began to crystallize, and I became very enthusiastic and motivated to start working.

I feel have been fortunate in many respects, but I should also stress that the luck is simply not enough. Refining even a most inspiring idea first to a journal article and subsequently to a dissertation was a much more demanding process than I ever had expected.

Very soon I had to start revising my plans, as well as my papers. This appeared to go on and on, endlessly, over and over again. In this process I acquired dozens of nice ideas I was keen on presenting, but those ideas were scattered, unstructured, and seemingly unrelated. Unfortunately, none of those nice ideas seemed to please the referees enough to warrant a paper in any decent journal. Thinking of a complete book like this felt very remote, even frustrating at times.

Fortunately, I have had numerous opportunities to attend and present my work in international conferences, workshops and summer schools, thanks to the generous financial support from the Foundation of Helsinki School of Economic and Business Administration, the Foundation for Economic Education (Liikesivistysrahasto), Yrjö Jahnsson's Foundation, and the Finnish Operations Research Society. The input from my 13 trips to 10 different countries on 4 different continents has been an essential factor. These professional meetings have enabled me to adapt to the mindset of the field, keep up with the latest developments, make invaluable contacts with overseas colleagues, and get highly valuable feedback to my own studies, among other benefits. Especially the EURO Summer Institute at the University of Warwick on August 1998, and the Productivity Workshops since 1997 have had a great impact on my work. I wish to forward my deepest thanks all those wonderful scholars who have shown interest in my work and have unselfishly shared their knowledge and insights with me while down on the road. I am afraid the complete list would be too long to be printed here, but I'm sure you know who you are!

These experiences helped me to realize that it was my most urgent task to learn to write papers of publishable quality. At least for me, this was a process of trial and error. Consequently, I didn't give much though to the dissertation, but studied all interesting research problems I knew without following any systematic agenda. My strategy was to take the advantage of the learning curve by producing a large number of papers and to spread the risk by investigating a wide variety of different problems and by targeting my papers to different journals. This was a time of very intensive work and rapid progress.

I have been privileged to share my efforts with two very ambitious, creative, clever, knowledgeable, work-oriented young scholars, Laurens Cherchye and Thierry Post, my non-convex brothers in arms. Indeed, teaming up with two highly productive co-authors has greatly enhanced my productivity. I have learned a lot from the intensive collaboration with Laurens and Thierry, and I hope the feeling is mutual. Perhaps most importantly, I learned to love the writing process, even the associated critical examination, which felt quite discouraging at first. Brainstorming sessions, anticipating the reactions of referees, and actually responding to the presented critique are actually very rewarding. And sharing this all with good friends makes it even more fun. Cheers mates!

For roughly two years, I didn't give much though to this project and felt a bit uneasy to even think about it. I could only hope that my hard work would eventually be rewarded by papers that would form the dissertation some day. I took up the dissertation project again back to my drawing board on Spring 2000. Article I, the paper that took the greatest effort, had just been accepted for publication. Articles II and III had been accepted somewhat earlier. After the discouraging initial record of 8 rejections, I suddenly was in the happy situation where I had the required number of accepted publications at hand, and I merely had to figure out a consistent story to bind the separate papers together. This was not the easiest puzzle to solve, but I think I have managed to fit the pieces together surprisingly well. Looking back, I see that I've followed the correct strategy.

It is worth to mention here that I have done my best to preserve the five articles that constitute this dissertation as strictly as possible in their original form as intended for publication, including any errors that I have become aware of afterwards. Hence, the articles also quite aptly reveal the development of my argumentation skills in the last years, from Article V which was written roughly two years ago in 1999 to the introduction that represent my present condition. I have revised some of our statements in the Introduction section where I review these articles to highlight the contribution to the theme of this dissertation. The errors of more technical nature are addressed in Appendix 1.

I am grateful to my pre-examiners, Professor Rolf Färe and Professor Niels-Christian Petersen, for their encouraging comments and constructive critique. Their fundamental work on the very topics of this dissertation has formed a source of inspiration for me through these years, and I think the influence of both these magnificent authors should be quite evident from this thesis. To be honest, I could not name any more appropriate person to evaluate the contribution of this dissertation than Professors Färe and Petersen. I consider it a great honor to have them both as the pre-examiners and the opponents of this dissertation.

Undoubtedly, the creative, internationally oriented, interdisciplinary atmosphere of the Quantitative Methods unit has also left its mark on my work. I have many pleasant memories from the conference trips, summer schools and suchlike events with my colleagues Tarja Joro, Veli-Pekka Heikkinen, and Pirja Heiskanen. I would also like to thank Mikko Syrjänen, Heikki Siltanen, and Sari Rytökoski for lunch-hour discussions and bringing in some fresh new innovative sprit in the Methods group. My appreciation also goes to Tuula Paukkunen, the secretary of the unit, who has always kindly assisted me with the practicalities.

I wish to acknowledge the inspiration from my parents Pertti and Eeva, whom I truly respect and admire. Since my parents did not have too good possibilities to study in their youth, I have always tried do my best and not to waste the excellent opportunities I have been blessed with. My parents never put much pressure or high expectations on me, but have given me the freedom to pursue my own routes. Their solid confidence in me has been a major resource.

Last but not least, I am indebted to my wife Hanna who has managed our household with great care and skill, giving me the opportunity to focus my attention on this work. My son Teemu deserves a special acknowledgement of the ability to cheer me up and bring me back down on the earth whenever needed. You are the greatest joy of my life. I realize that my long working hours, numerous overseas trips, and observing my intensive thought experiments at home have often made you question is it really worth the effort. I am not sure yet, but I sincerely thank you for all the patience and love you have given me. As far as I can decide, my work is not completed yet, but this is just the beginning. Still, I hope this dissertation proves my efforts have not been totally in vain.

Koivukylä, Vantaa, March 9, 2001.

Timo Kuosmanen

Contents

Introduction

The notion '*Data Envelopment Analysis* '(DEA) was introduced by Charnes, Cooper, and Rhodes (1978). Today, DEA has become an extensively studied, application driven 'multipurpose tool' for analyzing Decision Making Units (DMUs) that consume multiple inputs to produce multiple outputs,[1] where the meaning of the words "DMU", "input", and "output" can be quite freely interpreted. We shall confine attention to production activities that can be characterized as a deterministic process of transforming quantifiable and homogenous inputs to quantifiable and homogenous outputs. That is, any stochastic variations in the process as well as any quality differences and non-measurable factors are assumed non-existing, negligible for the purposes of the analysis, or 'correctable' by means of some kind of data pre-processing. Although we do not impose any other qualifications on the character of the potential activities or the numbers of inputs and outputs, these preconditions can still render some potentially interesting economic activities beyond the scope of our methods. For example, in the service sector the quality of 'production' is of primary importance, but it is extremely difficult - if not impossible - to measure the quality aspect by the degree and accuracy required for efficiency analysis. Still, the focus on the quantifiable activities is the traditional stance of DEA, and we shall not deviate from it here.

DEA is studied and applied in a number of disciplines, including Economics and Econometrics, Operations Research and Management Science, Applied Mathematics, and Statistics. Most traditional application areas of DEA include such categories as agricultural production, health care, education, financial institutions, and public sector production. More 'unusual' applications of DEA include e.g. efficiency assessments with historical production data from Domesday England (McDonald, 1997), evaluation of alternative sites for the super-conducting super-collider in Texas (Thompson et al., 1986), and assessing the multi-dimensional performance of baseball players (Mazur, 1994)[2]. There exists a lot of evidence that demonstrates the overwhelming popularity of DEA. See e.g. Seiford (1996) for a survey article that lists over 600 published articles in refereed journals. We are not aware of the current bibliometric statistics, but in our experience, the volume of DEA papers has recently increased rather than declined.

As noted in the first paragraph of Article I, DEA has traditionally been used for measuring various notions of technical and economic efficiency. However, in contrast to earlier accounts on DEA which typically present DEA as a synonym to efficiency analysis, we here de-emphasize the efficiency assessment aspects of DEA. On one hand, there also exist considerable alternative approaches to efficiency analysis, e.g. the parametric approach known as *Stochastic Frontier Analysis* (SFA: see e.g. Bauer, 1990, for a survey). In fact, the most significant contributions to the actual efficiency gauges date way back to the 'pre-DEA' era (e.g. Debreu, 1951; Shephard, 1951; Farrell, 1957) or have been presented at general level without any direct association with DEA (e.g. Färe and Lovell, 1978; Zieschang, 1984; Chambers *et al.*, 1998; among others). On the other hand, DEA is by no means limited to the context of efficiency analysis either. In fact, DEA is also employed for approximating production functions, recovering unobserved shadow prices of inputs and outputs, providing best-practice benchmarks, revealing shirking in principal-agent problems, and for decision

[1] In the single-input single-output case, the original DEA model coincides with the traditional ratios of (partial) productivity indices, and the engineering definition of efficiency. In this sense, DEA could be viewed as an extension of the traditional output/input ratios to multi-variable settings.
[2] An application to the Finnish version of baseball, Pesäpallo, is reported in Kuosmanen (1999).

support in complicated choice situations, among other uses (see Article I, Section 1, for references).

By the above arguments, we see DEA essentially as a frontier estimation technique, which approximates the production frontier by a piece-wise linear envelopment supported by the observed data points. In other words, DEA constructs a subset of the input-output space that contains all observed DMUs, i.e. DEA literally *envelopes the data.* In economic terminology, such an envelopment set can be viewed as an empirical approximation of the production possibility set characterizing the technology, i.e. the hypograph of the production function. In general, our objective is to reveal some interesting information of the (unknown) underlying technology for the above-discussed purposes by studying the DEA frontiers.

Of course, one could easily imagine many alternative ways of enveloping any given data. In DEA, the specification of the envelopment set is uniquely determined by the apparently universally accepted *minimum extrapolation principle* (Banker, Charnes, and Cooper, 1984). According to this principle, the empirical DEA production set should be *the minimal set* that satisfies some imposed *production assumptions*.[3] The production assumptions of DEA conventionally include *monotonicity* (i.e. free disposability of inputs and outputs), *convexity*, and various notions of *returns to scale* (see e.g. Article I, Section 2, for formal definitions). Thinking in terms of the production frontier, monotonicity can be viewed as a first-order curvature condition, convexity can be seen as a second-order curvature condition, while returns to scale properties can be thought of as homogeneity conditions. This dissertation primarily focuses on the curvature conditions: monotonicity and convexity. The returns to scale properties will be devoted less attention, because 1) the treatment of returns to scale in DEA has already been thoroughly investigated in the earlier literature (see Article I, Section 2, for references), and 2) we are not aware of any acute problems in the standard treatment at the moment.

From operational point of view, the *minimum* extrapolation principle connects DEA in a natural way to optimization. Indeed, DEA typically employs modern Mathematical Programming, especially Linear Programming (LP) techniques. Much of the contemporary research on DEA focuses on the operational issues of DEA. For example, how to formulate a LP problem in such a way that the constraints characterize a particular set of production assumptions is a purely operational issue.

By contrast, some of the most profound methodological aspects of DEA apparently rest on a quite shallow foundation. For example, which assumptions to impose, and by what criteria, are issues which - to a large extent - determine the results one will obtain by DEA. It is worth emphasizing that the specification of production assumptions is not merely a theoretical issue for academic debate. It is a question of vast practical importance for all application of DEA. It seems not exaggerated to claim the success or the failure of a DEA application essentially depends on its assumptions. Unfortunately, the DEA literature only offers minimal guidance on the specification of the production assumption. Many valuable

[3] By the term "production assumption" we explicitly refer to the maintained assumptions (axioms, postulates) concerning the properties of the underlying *production set*. Of course, we also need additional (explicit or implicit) assumptions in DEA. For example, we earlier assumed that the production process does not involve stochastic variations and the data are error-free. Some authors have also imposed additional assumptions on the distribution of inefficiency (see e.g. Banker, 1993). We here exclusively focus on the production assumptions and the discussion on the role of the remaining premises falls beyond the scope of this dissertation. Interested readers can e.g. consult our recent working papers Kuosmanen and Post (1999) and Cherchye, Kuosmanen, and Post (2000, 2001a) for further discussions on the stochastic settings.

discussions with knowledgeable colleagues working in this field, as well as the feedback we have received to our papers, give us an impression that quite many researchers and practitioners feel puzzled about the production assumptions involved in DEA. We therefore believe the role of production assumptions forms a generally interesting as well as challenging subject for this dissertation.

This dissertation consists of 5 distinct articles, which all contribute to the subject from slightly different perspectives. We next take a brief outlook on these papers. Our purpose is merely to clarify the connections between the independent papers, and to underline the most interesting implications to the main theme of this thesis. Moreover, we use this opportunity to briefly discuss some additional points of interest that did not find their way to the original papers. All technical errors we have identified after completing the papers are addressed in Appendix 1.

Article I: Kuosmanen, T. (2001): 'DEA with Efficiency Classification Preserving Conditional Convexity', *European Journal of Operational Research* 132(3), 83-99

The first article provides an elaborate introduction to the subject. First and foremost, this paper emphasizes that there is no valid justification for assuming production sets to be generally monotonous or convex. For example, most commodities are more or less indivisible, but indivisibility of inputs and/or outputs immediately violates both monotonicity and convexity. Furthermore, even if inputs and outputs were divisible, congestion of production factors (Färe and Svensson, 1980) violates monotonicity, while economies of scale and of specialization violate convexity. A classic reference to the economic significance of non-convexities is Farrell (1959), who aptly summarized the economic importance of both these qualifications:

> "A glance at the world about us should be enough to convince us that most commodities are to some extent indivisible and that many have very large indivisibilities. Similarly, whenever one refers to "economies of scale" and of "specialization", one is pointing to concavities in production functions. There is thus no need to argue the importance of either indivisibilities or concavities in production functions – the former are an obvious feature of the real world, and the latter have constituted a central topic in economics since the time of Adam Smith.
>
> (Farrell (1959), p. 378-379) [4]

In his seminal 1957 paper, which laid the foundation for efficiency measurement in general and DEA in particular, Farrell mostly focused on convex technologies in case of constant returns to scale. Nevertheless, Farrell also discussed the possibilities of economies and diseconomies of scale, and he openly admitted the difficulties his method faces in case of economies of scale.[5] Interestingly, Farrell proposed to remedy this problem by the simple "grouping method", which means applying his method of enveloping data to subsets of DMUs operating at the sufficiently similar scale size. Farrell and Fieldhouse (1962) further elaborated on this possibility. Moreover, Farrell and Fieldhouse proposed yet another approach ("overall method") of applying the original Farrell method to data that is

[4] A careful reader notes that (in standard terminology of mathematics) the hypograph of a concave production function is a convex production set, and hence Farrell's statement might appear somewhat confusing. However, Farrell deviated (for convenience) from the mathematical terminology and simply associated "departures from convexity" with "concavity". (Farrell, 1959; footnote 3)

[5] Of course, diseconomies of scale are easily handled in the standard (VRS/BCC) framework. Perhaps the difficulties Farrell saw with the treatment of the economies of scale might explain why Farrell never extended his method towards the so-called 'variable returns to scale' (VRS) technology.

transformed in such a way that constant returns to scale hold in the transformed input-output variables. These subsequent research efforts (following the seminal Farrell (1957) article) suggest that Farrell was genuinely concerned of the restrictive character of convexity in the present context.

The problem with economies of scale has received considerable attention in the more recent DEA literature. Most notably, Petersen (1990) and Bogetoft (1996), and just recently Bogetoft, Tama, and Tind (2000) and Post (2001b) have proposed DEA formulations that relax convexity of the production set, but maintain convexity of the input and/or output correspondences. This specification allows for economies of scale, but (unfortunately) excludes potential economies of specialization in inputs and/or outputs. Indeed, should the input and output correspondences always be convex, there would not be any incentive to specialize in production of some particular inputs or outputs. The notion "economies of specialization" is by definition equivalent to non-convex input/output sets (e.g. Farrell, 1959; Arrow, Ng, and Yang, 1998). As noted already by Adam Smith in his famous needle factory example (to which Farrell refers in the quotation above), specialization can improve productivity to a considerable extent. In Economic theory, increased specialization and division of labor are in fact widely regarded as the main sources of long-term economic growth. Moreover, in many areas of Economics, non-convexities in technology have been systematically raised up as the driving force behind specialization (see e.g. Yang, 1994; Yang and Ng, 1993, 1995; Borland and Yang, 1995; Yang and Rice, 1994; Shi and Yang, 1995).

Petersen (1990) originally motivated the adherence to convexity in input and output space by referring to the '*law of diminishing marginal rates of substitution*'. This notion was also cited by Bogetoft, Bogetoft *et al.*, and Post as the motivation of their DEA formulations. However, we cannot find other references to such law in the context of production in any economic text we are aware of. The fact that neither Petersen nor the other authors citing this law presented any exact reference lead us to suspect an unintentional and unfortunate confusion to the classic 'law of diminishing returns' (discussed in more detail in Article II). That law, however, has nothing to do with convexity of input or output correspondences.

In Article I we also put the content of this suspicious 'law' into question by demonstrating that marginal rates of substitution cannot be guaranteed to diminish in case of increasing marginal products, i.e. the very situation for which these 'relaxed' DEA models cited above were designed. Our example in Article I demonstrates how economies of scale can actually give rise to economies of specialization under differentiability, even if the inputs or outputs would seem to be technical complements by their nature. To illustrate this point more specifically, consider the common textbook claim that input correspondences are convex if production activities can be operated side by side without interfering with each other (see e.g. the quotation of McFadden (1978) in Article I, p. 28). The following example demonstrates that this need not be true in case of economies of scale. Suppose the output y can be produced in two independent production processes involving either input 1 or 2, i.e. $x_1, x_2 \in \Re_+$. Specifically, let the first production processes be characterized by the production function $y = x_1^2$, and let the second (alternative) process be $y = 2x_2^2$. Note that both processes exhibit genuinely increasing returns to scale, with increasing marginal productivity. Suppose further that we can operate these production processes side by side without interfering with each other, so that the total amount of output is simply $y = x_1^2 + 2x_2^2$. Interestingly, one can easily verify that the input correspondences associated with this technology are non-convex for all $y>0$.[6] This simple example suggests that when we allow for economies of scale, it is difficult to ascertain convexity of the input isoquants by reasoning how inputs might interfere with each other. Rather, this example and the discussion in Section 3 of Article I demonstrate that economies of scale can lead to non-convex input correspondences when there is no interaction whatsoever between the inputs, and even when inputs are mutually complementary. By these somewhat discouraging findings we call for extra caution with postulating convexity properties *a priori*.

Besides drawing attention to some unrecognized problems with the standard DEA assumptions, Article I proposes a new less restrictive technology property so as to facilitate a more general (and hence 'safer') DEA approach. That approach is based on the novel *conditional convexity* concept introduced in Article I. The heuristics behind this notion are very similar to that of the Farrell's "grouping method", i.e. we apply the standard DEA to some subsets of DMUs for which convexity can be assumed to hold.[7] The problem is, of course, how to identify those subsets.

In Article I we confine special attention to the variant we call '*efficiency preserving conditional convexity*', or shortly *c-convexity*. The starting point of that special case is the observation that many important functions of DEA could actually be performed without imposing any production assumptions whatsoever, resorting only to the observed facts, i.e. to the discrete set of observed DMUs. For example, we can classify which DMUs are clearly inefficient –being dominated by at least one other DMU in each input and output- and which DMUs seem efficient in the light of the given evidence. We refer to this efficiency classification obtained without any production assumptions whatsoever by spontaneous Pareto-Koopmans efficiency classification. Of course, we can further identify seemingly efficient DMUs as target points or benchmarks so as to facilitate learning from the 'best practices'. In fact, the adjacent literature of nonparametric production analysis (Afriat, 1972; Varian, 1984) does resort to the discrete set of observed DMUs when testing for the consistency of the observed production behavior with hypotheses of profit maximization or cost minimization.[8]

Still, for other purposes of DEA it is desirable to go beyond the observations and construct a continuous empirical frontier. The heuristics behind the c-convexity property are the following: Analogous to the standard DEA, we interpolate the data by constructing a piece-wise linear frontier consisting of hyperplane segments, where the empirically efficient DMUs constitute the vertices of the frontier. We deviate from the basic DEA in that we do not allow any arbitrary ad hoc assumptions to violate the natural spontaneous Pareto-Koopmans efficiency classification. This implies that the resulting production set can be non-convex. We show that in theory, conditional convexity (and hence c-convexity as its special case) implies both monotonicity and convexity, but not conversely. Therefore, conditional convexity is a more general technology property than either convexity or monotonicity.

---

[6] The input sets are convex if the Hessian matrix of the production function is negative semidefinite. However, the Hessian matrix associated with this technology is a diagonal matrix (since $\partial^2 y/\partial x_1 \partial x_2 = \partial^2 y/\partial x_2 \partial x_1 = 0$), the diagonal consisting of positive numbers ($\partial^2 y/\partial x_1^2 = 2$ and $\partial^2 y/\partial x_2^2 = 4$). Hence this matrix is positive definite, implying concave input isoquants and non-convex input sets.

[7] Interestingly, the modern equivalent of the Farrell and Fieldhouse "overall method" is the "Transconcave DEA" approach studied by Post (2000a,b). Of course, the main difficulty with that approach is the specification of appropriate transformations. Post has investigated this problem in more detail, and has obtained quite promising results in simulated experiments by using empirical specification tests.

[8] Most notably, Banker and Maindiratta (1988) and Färe and Grosskopf (1995) have stressed the kinship between the distinct traditions of nonparametric production analysis and DEA.

As discussed in Article I, a potential problem of the proposed approach is the computational complexity. While the basic DEA models involve solving a single LP problem for each DMU we desire to evaluate, the c-convex DEA model presented in Article I boils down to disjunctive programming. Article I illustrated how one could characterize the reference peer groups, and subsequently compute efficiency measures by solving a finitte sequence of LP problems. Unfortunately, the total number of LP problems one needs to solve in this primitive algorithm can be very large depending on the data set. It is therefore of considerable interest to note here the encouraging result that more efficient computational procedures do exist. Very recently, we have found a way of reformulating the disjunctive programming problem associated with the c-convex DEA model equivalently as a binary (0-1) Mixed-Integer Linear Programming Problem (MILP).[9] Attractively, computing efficiency measures relative to the c-convex technologies can be effectively conducted by solving a single MILP problem per each DMU to be evaluated. Today, highly efficient algorithms for solving MILP problems are generally available, and hence (in sharp contrast to the conclusions of Article I) we no longer need to worry about the computational burden of the c-convex DEA model.

From statistical point of view, we can think of the input-output data of DMUs as a sample drawn from the production set, and DEA efficiency measures as estimates for the true efficiency measure. Banker (1993) first demonstrated that these DEA estimates are consistent under some relatively general distribution assumptions (see Simar and Wilson, 2000, for an up-to-date discussion). In Article I we express the conjecture that under the standard distribution assumptions considered by Simar and Wilson (2000), a DEA model based on c-convexity should be statistically consistent for a more general class of technologies than the standard DEA (and Free Disposable Hull (FDH: Deprins, Simar, and Tulkens, 1984)) models that build on the conventional monotonicity and/or convexity properties.

In finite samples, DEA efficiency estimator that builds on c-convexity is exposed to sampling error like all DEA estimators. In Article I we have focused on a variant of c-convexity that preserves efficiency classification of each observed DMU, but not the classifications of the unobserved production vectors. Consequently, in contrast to the standard DEA models, adherence to the minimum extrapolation principle does not suffice to guarantee that the empirical production set is contained within the underlying production set.[10] In other words, inefficiency of some DMUs could be exaggerated in a small sample. Still, in larger samples, under the standard assumptions on the data generating process, the c-convex DEA efficiency estimates should converge for a more general class of technologies than efficiency estimates of any earlier DEA model. Moreover, the c-convex DEA model can never yield smaller efficiency estimates than the convex DEA models. Hence, in light of the alternatives, it would seem hypocritical to immediately object the potential negative small sample errors associated with the c-convex DEA efficiency estimates. Specifically, if the c-convex DEA efficiency estimates happen to exaggerate inefficiency, the convex DEA alternative will always do even worse because of its inherent specification error.

[9] The MILP formulation is presented in the paper in the forthcoming working paper Kuosmanen (2001), which is available from the author upon request. Since the contents of the dissertation have already been formally examined and approved, we refrain from drastic revisions at this stage, following the advise of the honorable custos.

[10] One might attribute this result to the shortcomings of the conditional convexity property. Conversely, this result can be viewed as a rather serious limitation of the fundamental minimum extrapolation principle, i.e. adherence to this principle implies containment of the empirical production set within the true production set only under some particular production assumptions, but not necessarily in general.

Finally, it is always possible to apply the c-convexity condition, in addition to the observed DMUs, to all imaginable input-output vectors as well. It would be straightforward to demonstrate that this second variant of c-convex DEA would simply give the intersection of $MH(S)$ and $CH(S)$, where $MH(S)$ is the monotone hull of observed DMUs and $CH(S)$ is the convex hull of observed DMUs (using the notation and terminology of Article I). This set would of course be contained within the underlying production set for all technologies that satisfy the c-convexity property. But unfortunately, by this more conservative approach we lose the attractive 'smooth' frontier associated with the first variant, and the efficiency measures no longer are continuous and subdifferentiable. There are no free lunches. Nevertheless, this second variant could prove an interesting route to accommodate 'congestion' in the Free Disposable Hull (FDH) framework. That issue that has attracted some interest among dedicated researchers (see e.g. Kerstens and Vanden Eeckout, 1999). The final remark smoothly moves us to Article II.

Article II: Cherchye, L., T. Kuosmanen, and G.T. Post (2001): 'Alternative Treatments of Congestion in DEA: A Rejoinder to Cooper, Gu, and Li', *European Journal of Operational Research* 133(1), 69-74

This paper grew out from two referee reports to the article by Cooper, Gu, and Li (2001), who advocated use of their new approach to deal with congestion in DEA by presenting some numerical counterexamples against the established Färe, Grosskopf, and Lovell (FGL: 1983, 1985) approach. Since many distinguished authors of the field apparently have faced difficulties with the somewhat confusing terminology, we first discussed the meaning and the relationships of the notions of congestion, free disposability, and structural efficiency. By this clarification we hope to direct the on-going debate on this subject towards more substantial issues.

We next refuted the claims of Cooper *et al.* by modifying their counterexamples. Specifically, we imposed some standard technology properties (explicitly assumed by FGL) to demonstrate that the alternative approach advocated by Cooper *et al.* can 'fail' in exactly the same way the FGL model was claimed to fail. To enable readers to follow the counter-arguments, we have placed some necessary information on the examples available in Appendix 2.

But in addition to the obvious, we tried to use this opportunity to emphasize two points of far greater importance in the present context. Firstly, we expressed our skepticism for empirical identification of congestion in a finite sample by any two-stage DEA model (including the two specific models subject to this debate). Our concern essentially arises from the fact that we typically only observe a sample of feasible input-output combinations, rather than the full (super)population. If we observe a particular input-output combination, we believe it must be feasible. But if we do not observe a particular input-output vector, it does not yet imply this production plan would be infeasible. Unfortunately, the two-stage DEA approaches do not take into account the possibility that some feasible production plans may not be observed. In Article II we refer to this phenomenon by the statistical term 'sampling error'. The problem is, the proposed congestion indicators can reflect violations of disposability as they are expected to do, but they can equally well reflect occurrences where our finite sample of DMUs simply does not include observations from all regions of the efficient frontier. Unfortunately, we cannot immediately distinguish the former desirable effect from the latter unwanted influence by these techniques.[11] We therefore suggested establishing a stronger connection with the rapidly developing DEA literature that explicitly deals with the issue of sampling error (e.g. Simar and Wilson, 2000).

Secondly, we politely called for more rigor and transparency with the assumptions. In this respect, the treatment by Cooper *et al.* is an example that falls below any standards. Firstly, Cooper et al. ignored the production assumptions on which the FGL approach was built on – the assumptions that FGL had quite clearly stated in their texts. Worse still, Cooper et al. did not mention what kind of assumptions their own method would necessitate. Yet, these authors made the claim that their own approach is more meaningful than the FGL alternative.

Looking at the bigger picture, these recent discussions are directly related to the old debate between the advocators of the 'loose paradigm' (which Cooper et al. (2000) call the "application oriented" approach) and the more rigorous 'axiomatic approach' to DEA (represented most notably by Färe and Grosskopf). The present discussion has aptly revealed the key advantages of the axiomatic approach. Our firm conviction is that when one becomes concerned of the meaning of the numbers generated by sophisticated DEA formulations (like Cooper *et al.* did), it is impossible to present any sensible argumentation if the maintained premises remain concealed. Yet, we would stress that the axioms are merely useful instruments for the logical reasoning. As Article III next aptly demonstrates, it is possible to reason logically even without mechanically resorting to axiomatics. The ultimate limitation of any axiomatic system is that the axioms themselves cannot be proved correct by the system's internal logic, but must be justified by external arguments.[12] This leaves room for a more holistic reasoning (see Conclusions below for further discussion), which has been absent in the contemporary DEA literature.

Article III: Cherchye, L., T. Kuosmanen, and G.T. Post (2000): 'What is the Economic Meaning of FDH? A Reply to Thrall', *Journal of Productivity Analysis* 13(3), 259-263

Article III continues along the lines of the previous paper, but returns back to the convexity property, now from the perspective of economic efficiency. This paper was written as a response to Thrall (1999) who claimed that the non-convex FDH efficiency analysis has no economic meaning whatsoever. Thrall based his argument on his principal theorem, which essentially demonstrated that a DMU diagnosed technically efficient in the FDH model can be highly inefficient in terms of profit maximization at exogeneously given and fully certain prices.

In our opinion this result is not surprising at all, and certainly does not give rise to the strong negative conclusions drawn by Thrall. Most importantly, a similar result can be easily demonstrated to hold for the standard convex DEA models as well, i.e. a technically efficient DMU according to a DEA model can equally well be highly inefficient in terms of profit

[11] Actually, a similar identification problem pertains to the identification and measurement of *scale inefficiency* by analogous two-stage DEA approaches. For example, we might be tempted to explain observed productivity differences by technical and scale inefficiencies. Unfortunately, DEA based finite sample inference may be seriously biased because the sampling error can only increase the technical efficiency component while it tends to systematically decrease the scale efficiency component (essentially because the CRS technology is less vulnerable to the sampling error). That is, we expect the two-stage DEA approach to exaggerate the importance of scale efficiency relative to technical efficiency.

[12] This point is manifested in the Gödel's famous impossibility theorem.

maximization. This is because technical efficiency is a necessary, but not sufficient condition for economic efficiency. Unlike Thrall, we do not interpret that as a failure of DEA.

Interestingly, when we take the prices explicitly into account to measure *profit efficiency* instead of *technical efficiency*, the convexity assumption becomes redundant. That is, profit efficiency measured relative to the convex monotone hull of DMUs always equals profit efficiency relative to the monotone hull (=FDH) of observations, which in turn equals profit efficiency relative to the discrete set of observed DMUs. This last observation follows directly from the fundamental theorems of nonparametric production analysis by Afriat (1972) and Varian (1984). Therefore, FDH is an equally meaningful technology approximation as the convex DEA (VRS) model in the economic circumstances considered by Thrall. These arguments *ipso facto* refute the claim of Thrall. But again, we used the opportunity to take this line of reasoning a little bit further.

We argue that in many empirically relevant situations, the DMUs do not take prices as exogenously given and fully certain. If we consider profit-maximizing behavior under endogenous prices or price uncertainty, however, the redundancy result of convexity ceases to hold in general. We illustrate the endogeneity argument by a simple numerical example adopted from Thrall.[13] That example demonstrates that under imperfect competition where prices generally depend on the supplied/demanded quantities, a DMU can maximize profit, but still become classified as inefficient in a convex DEA model. As for the price uncertainty, we refer to Article V included in this thesis. The discussion of Article III hence provides an illustrative verbal introduction to the perspective of economic efficiency investigated in the last two papers.[14]

Article IV: Kuosmanen, T., and G.T. Post (2001): 'Measuring Economic Efficiency with Incomplete Price Information', *European Journal of Operational Research*, to appear.

Article IV first rephrases the foundations of the economic efficiency measurement. In this perspective, we motivate the Farrell technical efficiency measure merely as a 'proxy' for economic efficiency. Interestingly, economic efficiency notions require very detailed price (or cost) information, while the technical efficiency proxy cannot immediately utilize any price information whatsoever. In our experience, however, most research situations are characterized as the intermediate cases where some important price data are unavailable or poorly measured, but still, some bounds on the relative magnitudes of prices can be presented based on theoretical knowledge or practical experience of the application area. Hence, in this paper we investigate the possibility of improving the Farrell's proxy by utilizing all price information at hand. For sake of generality (and convenience), we assume the price domain takes a form of a convex pointed cone, which allows for a gradual transition from the perfect price information to the opposite case of no price information. We show how one can derive an interval estimate for cost efficiency by applying the Farrell input measure to 'convexified' input correspondences which are augmented by the polar cones associated with the price domain.

[13] It is worth to add that (like Thrall) we did not impose any production assumptions, but studied meaningfulness of monotonicity from the point of view of the economic objectives.
[14] For further discussion on price endogeneity and uncertainty, we refer to our recent working paper Cherchye *et al.* (2001b).

From operational point of view, these interval estimates can be conveniently computed using the standard weight-restriction, assurance region, and cone-ratio techniques, which have been extensively studied in the recent DEA literature (see Article IV for references). In fact, the incorporation of price information to DEA is frequently mentioned as the motivation of using these techniques. However, the interpretation of the resulting 'efficiency scores' appears to be an acute problem of those techniques, which seems to cause a lot of confusion among the practitioners. Hence, we saw it worth to elaborate on the economic foundation by this paper.

In addition to this theoretical elaboration, our treatment also suggests some distinct material advances in comparison to the earlier proposals. Firstly, we proposed to consider the entire interval of outcomes instead of solely focusing on the 'optimistic' upper bound. This gives a more complete picture of the magnitude of the economic losses due to inefficiency on one hand, and the gains obtainable by improving the performance on the other.[15] Secondly, we stressed that not all price information is always desirable – a point that is typically ignored in the weight-restriction literature. In essence, the value of the additional price information depends on the notion of economic efficiency that is seen appropriate for the situation at hand. For example, if we are primarily interested of cost efficiency, the output prices are completely useless. Thirdly, we again emphasized the restrictive nature of production assumptions, especially convexity. Thus far, the weight restriction approaches have been predominantly employed in the convex DEA framework.[16] For example, one can motivate convex input sets by the focus on cost efficiency, but that is not a valid excuse for assuming convex output sets. Consequently, we showed how to incorporate additional price information in the less restrictive FDH technology. Interestingly, the cost efficiency estimates can in this case be computed by the standard LP codes, complemented with a very simple enumerative principle. An application to European commercial banks demonstrates that even some slight restrictions on the feasible price domain can result in substantial improvement in the efficiency estimates (provided that the restrictions are correct).

Article V: Kuosmanen, T., and G.T. Post (1999): 'Nonparametric Efficiency Analysis under Uncertainty: A First-Order Stochastic Dominance Approach' Helsinki School of Economics and Business Administration, Working paper W-230

Finally, in Article V we extend the economic efficiency analysis circumstances characterized by price uncertainty. More often than not, the producers must commit themselves to certain production plans before they have completely certain information on the input or output prices – consider capital inputs for example. Therefore, the *ex post* performance evaluation undertaken with DEA may not be appropriate when producers face *ex ante* price uncertainty. Of course, evaluation of *ex ante* efficiency of production plans is a complicated task, and to the best of our knowledge, this paper is the first serious attempt to shift the perspective of DEA towards this direction.

To take the risk preferences of DMUs explicitly into account, we formulated the production objective to be maximization of the expected utility of profit. Interestingly, we

[15] Unfortunately, there is an annoying technical error associated with the lower bound, which may cause some confusion. The error is corrected in the Appendix.

[16] For example, Seiford (1996) explicitly stressed that the weight restriction techniques are available for the *convex* DEA models, which can give an impression that weight restrictions cannot be incorporated into the FDH framework. As far as we know, Article IV makes the first proposal to introduce the weight restrictions to the FDH analysis.

have later discovered that this objective function has established as the standard in the theory of firm under uncertainty (the classic references are McCall, 1967; and Sandmo, 1971). Still, it should be emphasized that the nonparametric efficiency tests proposed in this paper do not essentially depend on the expected utility hypothesis. In contrast to our statement on p. 91 (Section 3, the last paragraph), a wide range of alternative non-expected utility theories are consistent with the First-order Stochastic Dominance (FSD) criterion, and can hence be immediately accommodated within our framework.

Our objective is to assess whether the evaluated input-output vector could be viewed as a rational *ex ante* choice that complies with some unknown well-behaved decision rule (i.e. is consistent with the FSD criterion) under some unknown price distribution. Using a similar structure of the price domain as in Article IV, we derive the necessary and sufficient conditions for a DMU to be consistent with the FSD condition. Interestingly, we note that these conditions include some standard conditions of economic and technical efficiency as the limiting special cases. This indicates the possibility to extend the traditional *ex post* efficiency conditions in a relatively straightforward way towards *ex ante* efficiency evaluation. Moreover, this proves that the proposed FSD test approach is easily implementable by the standard computational techniques.

Of course, we might also think of a degree measure of inefficiency in case our test indicates inefficiencies. However, such measure should almost inevitably be based on the 'utility' scale. But unfortunately, measuring the utility loss of inefficiency seems highly problematic, given our nonparametric orientation. Moreover, the complexity of the current setting involving price uncertainty appears to render the dual approach of Debreu (1951) ineffective. Due to these difficulties, we abstract from measuring the degree of inefficiency and merely focus on identification of inefficient DMUs.

As for the production assumptions, we elaborate on our earlier argument (see Article III) that under price uncertainty the convexity of production possibility set is generally not a harmless condition, which is in a sharp contrast to the standard result under price certainty. In contrast to convexity, we show that monotonicity does remain as a harmless condition in the new setting as far as the utility function is increasing in profit.

The restrictive nature of convexity essentially arises from the allowance for risk preferences, i.e. risk aversion. Interestingly, it can be easily demonstrated that if producers are assumed to be *risk neutral*, the standard result of harmless convexity is restored. This is because risk neutrality implies that the expected utility of profit always equals the utility of the expected profit, or equivalently, the utility of profit at the expected prices, i.e. (using the obvious notation)

$$EU(\boldsymbol{p}) = U(E\boldsymbol{p}) = U(E(p)y - E(w)x).$$

To prove the harmless character of convexity in this case, one can simply rephrase the Afriat's Theorem (e.g. Varian (1984)) in terms of expected prices (instead of the fully certain prices).

It remains an interesting challenge for future research to investigate how the power of the proposed nonparametric stochastic dominance test could be improved by imposing higher order restrictions on the risk preferences, e.g. by using the Arrow-Pratt measure of risk aversion or the certainty equivalent as instruments. Conversely, it would be interesting to find a way to recover how large the degree of risk aversion or the certainty equivalent would have to be in order to be able to ‘rationalize’ a given non-dominated input-output vector.

Conclusions and Discussion

We have argued that DEA is essentially a nonparametric approach to frontier estimation, in which the imposed production assumptions play a central and determinant role. We discussed the meaning and the interpretation of the production assumption, which typically include such properties as monotonicity, convexity, and returns to scale. We introduced some new aspects such as economies of scale and of specialization as well as incomplete or uncertain price information into this discussion. Our remarks especially focused on the various convexity properties, which were debated from both technological and economical perspectives. Moreover, some difficulties in the treatment of congestion using the two-stage DEA techniques were pointed out. In addition to the theoretical discussion, we present some readily implementable operational DEA techniques, most notably the c-convex and the price-augmented production sets, which build directly on the observed data.

Although the mathematical definition of production assumptions does not leave any room for speculation, there are considerable differences in the interpretation of the assumptions between different authors. We therefore find it illustrative to categorize alternative perspectives to production assumptions according to their source and their function. The production assumptions can originate from 1) rational arguments and intuition or 2) empirical considerations and evidence. Secondly, we can categorize the production assumptions according to their intended purpose as i) descriptive statements or ii) instrumental hypotheses. This classification gives us four alternative views of the nature of production assumptions.

Table 1: Alternative perspectives to production assumptions: example arguments and representative references

	Rational	**Empirical**
Descriptive	Production possibilities are convex, because the manufacturing process is both additive and divisible. Arrow and Hahn (1971)	Evidence from empirical studies supports the assumption that production possibilities are convex. Tulkens (1993)
Instrumental	Convexity is a 'harmless' regularity condition under profit maximization. Varian (1984)	"Decision-Makers find the results of the convex DEA model more convincing than the results from the non-convex model."

Table 1 illustrates these four different perspectives in case of the convexity assumption. To illustrate each category, we formulated some arbitrarily chosen example arguments to justify the convexity assumption. Moreover, we included a representative references to papers where the perspective is present. Note that our examples and the references are not meant to be directly related. For example, Tulkens (1993) takes exactly the opposite stance to the example argument presented in Table 1. Moreover, we are not aware of any explicit arguments where empirical evidence had been used to back up instrumental assumptions, and we see that this position is virtually non-existing in the literature.

In the articles that constitute the main content of this dissertation, we have criticized production assumptions from all these perspectives, focusing on various convexity assumptions. Firstly, we have presented rational arguments for why standard production assumptions like convexity need not hold in reality, drawing attention to such aspects as economies of scale and of specialization. Moreover, we have pointed out some important limitations in the line of reasoning presented in the earlier literature. We refer especially to Article I for further discussion.

Secondly, we have investigated the existing empirical evidence, which cannot be described exhausting at the moment, and found some indication of violations of the common assumptions (see e.g. Article I, Section 3, p. 33). However, we consider the existing empirical findings regarding the production properties tentative at best. In our opinion, a serious empirical program would necessitate more developed statistical techniques, which could account for both the sampling error and the data problems. Although some considerable steps have recently been taken in that research direction (see e.g. Article II, Section 6, for references), a number of problems remain unsolved. Furthermore, we face rather fundamental identification problems whenever we investigate a spontaneously generated real-world economic data (rather than fully controlled and isolated production experiments, like e.g. field trials in agriculture). For example, the fact that we do not currently observe any DMU producing in the congested or non-convex region of the production frontier does not generally imply the production technology would be monotonous and convex. It may simply be the case that the economic motives (e.g. profit maximization) drive the DMUs to produce at the economically advantageous regions of the frontier even though the underlying technology actually is non-monotonous and non-convex. That is, an economic selection effect occurs, which makes the data look like it had come from a "nicely behaved" convex monotonous technology (analogous to e.g. Afriat, 1972; and Varian, 1984). This identification problem is an important (yet largely ignored) consideration in many functions of DEA, including the planning of the future activities, in the (re)allocation of resources to the DMUs, as well as in the forecasting of the firm behavior.

Thirdly, from the purely instrumentalist perspective, we do not see there any material value whatsoever in such production assumptions that are harmless, i.e. do not interfere with the results. As discussed in Article III, the discrete set of observations would then give the identical results. That is, we object using the duality relationships between the economic objective functions and the technology representations as a justification for production assumptions in DEA. On the other hand, the instrumental statistical argument of imposing production assumptions to reduce the sampling error in efficiency estimates has the obvious cost of increased risk of specification error, without any associated gain in the asymptotic properties. As noted above, we are not aware of systematic instrumentalist accounts based on empirical findings.

By the above arguments, we call for displacing the production assumptions from the leading role they now play the contemporary methodology. Instead, we would like to see there a genuinely data driven DEA, which truly lets the data 'speak for itself' rather than force it to the idiom of some arbitrary ad hoc postulates. In our opinion, the first important step in the correct direction is taking the discrete set of empirical observations as the basic starting point of investigations. Note that focusing on the discrete set of observations is perfectly in line with the minimal extrapolation principle of DEA when no production assumptions are enforced. The five articles below all share this data focused perspective. It is evident from these papers that our point is not to abandon envelopment sets or DEA. We have merely observed that the strongest cases for enveloping the data are typically the motivated from the perspective of economic conditions, objectives or efficiency criteria, rather than form technology properties. Under such circumstances, the discrete set of observations is at least an equally good technology representation as any envelopment set. We do recognize the possibility of eliciting technical information (e.g. engineering consideration, see e.g. Chenery, 1949) that can be incorporated into the analysis by means of global or local production assumptions. However, we find this a quite difficult route, limited to very simple production activities.

Finally, we think the fully blown methodology would also more explicitly recognize the serious data limitations, incomplete information, and uncertainties involved in production. This would call for more transparent model specification and justification of the methodological choices, reflected also in the conclusions and implications. Most certainly, this means 'weaker' results in the sense that greater emphasis should be based on investigation of alternative scenarios, speculative 'what if…' thought experiments, examination of the extreme cases, and more systematic sensitivity and robustness analyses. However, results of such comprehensive analyses would almost certainly be 'stronger' in terms of their empirical content.

REFERENCES

Afriat, S. (1972): Efficiency Estimation of Production Functions, *International Economic Review* 13(3), 568-598

Arrow, K.J., and F.H. Hahn (1971): *General Competitive Analysis*, Holden-Day Inc., San Francisco. Oliver & Boyd, Edinburgh.

Arrow, K.J., Y.-K. Ng, and X. Yang (Eds.) (1998): *Increasing Returns and Economic Analysis*, New York: St. Martin's Press; London: Macmillan Press

Banker, R.D. (1993): Maximum Likelihood, Consistency and Data Envelopment Analysis: A Statistical Foundation, *Management Science* 39(10), 1265 – 1273

Banker, R.D., A. Charnes, W. W. Cooper (1984): Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis, *Management Science*, 30(9), 1078 – 1092

Banker, R.D., and A. Maindiratta (1988): Nonparametric Analysis of Technical and Allocative Efficiencies in Production, *Econometrica* 56(6), 1315-1332.

Bauer, P.W. (1990): Recent Developments in the Econometric Estimation of Frontiers, *Journal of Econometrics* 46(1/2), 39-56

Bogetoft, P. (1996): DEA on A Relaxed Convexity Assumption, *Management Science* 42(3), 457-465

Bogetoft, P., J.M. Tama, and J. Tind (2000): Convex Input and Output Projections of Nonconvex Production Possibility Sets, *Management Science* 46(6), 858-869

Borland, J., and X. Yang (1995): Specialization, Product Development, Evolution of the Institution of the Firm, and Economic Growth, *Journal of Evolutionary Economics* 5(1), 19-42

Chambers, R. G., Y. Chung, and R. Färe (1998): Profit, Directional Distance Functions, and Nerlovian Efficiency, *Journal of Optimization Theory and Applications* 98(2), 351-364

Charnes, A., W.W. Cooper, and E. Rhodes (1978): Measuring the Efficiency of Decision Making Units, *European Journal of Operational Research*, 2(6), 429 – 444

Chenery, H. (1949): Engineering Production Functions, *Quarterly Journal of Economics* 63(4), 507-531

Cherchye, L., T. Kuosmanen, G.T. Post (2000a): New Tools for Dealing with Errors-In-Variables in DEA, Catholic University of Leuven, Center for Economic Studies, Discussion paper Series DPS 00.06.

Cherchye, L., T. Kuosmanen, G.T. Post (2001a): Nonparametric Efficiency Estimation in Stochastic Environments, *Operations Research* (to appear)

Cherchye, L., T. Kuosmanen, G.T. Post (2001b): Nonparametric Production Analysis under Alternative Price Conditions, Catholic University of Leuven, Center for Economic Studies, Discussion paper Series 01.05.

Cooper, W.W., B. Gu, and S. Li (2001): Comparisons and Evaluations of Alternative Approaches to the Treatment of Congestion in DEA, *European Journal of Operational Research* 133(1)

Debreu, G. (1951): The Coefficient of the Resource Utilization, *Econometrica* 19(3), 273 - 292

Deprins, D., D. Simar, H. Tulkens (1984): Measuring Labor Efficiency in Post Offices, p. 243 - 267 in M. Marchand, P. Pestieau and H. Tulkens, *The Performance of Public Enterprises: Concepts and Measurements*, North Holland

Farrell, M. J. (1957): The Measurement of Productive Efficiency, *Journal of Royal Statistical Society*, Series A, 120(3), 253 - 290

Farrell, M.J. (1959): Convexity assumption in theory of competitive markets, *Journal of Political Economy* 67, 377-391.

Farrell, M.J., and M. Fieldhouse (1962): Estimating Efficient Production Functions under Increasing Returns to Scale, *Journal of Royal Statistical Society*, Series A 125, 252-267

Färe, R., and S. Grosskopf (1995): Non-Parametric Tests of Regularity, Farrell Efficiency, and Goodness of Fit, *Journal of Econometrics* 69, 415-425.

Färe, R., S. Grosskopf, and C. A. K. Lovell (1983): The Structure of Technical Efficiency, *Scandinavian Journal of Economics* 85(2), 181-190

Färe, R., S. Grosskopf, and C.A.K. Lovell (1985): *The Measurement of Efficiency of Production*, Kluwer Academic Publishers, Boston

Färe, R., and C. A. K. Lovell (1978): Measuring the Technical Efficiency of Production, *Journal of Economic Theory* 19, 150-162

Färe, R., and L. Svensson (1980): Congestion of Production Factors, *Econometrica* 48, 1743-1753

Hanoch, G., and M. Rothschild (1972): Testing Assumptions of Production Theory: A Nonparametric Approach, *Journal of Political Economy* 80, 256-275

Mazur, M. J. (1994): Evaluating the Relative Efficiency of Baseball Players, in Charnes, A., W. Cooper, A. Y. Lewin, and L. M. Seiford (eds.): *Data Envelopment Analysis: Theory, Methodology and Applications*, Kluwer Academic Publishers

McCall, J.J. (1969): Competitive Production for Constant Risk Utility Functions, *Review of Economic Studies* 34, 417-420.

McDonald, J. (1997): Manorial Efficiency in Domesday England, *Journal of Productivity Analysis* 8(2), 199-213

McFadden, D. (1978): Cost, Revenue, and Profit Functions, in M. Fuss and D. McFadden (Eds.): *Production Economics: A Dual Approach to Theory and Applications*, North-Holland, Amsterdam

Kerstens, K., and P. Vanden Eeckout (1999): Measuring Congestion with Convex and Non-Convex Technologies: Definition, Measures, Tests and Empirical Illustration, paper presented at the 6th European Workshop on Efficiency and Productivity Analysis, Copenhagen, Denmark, October 29-31 (on-line copy currently available at http://www.flec.kvl.dk/6EWEPA/abstracts.htm)

Kuosmanen, T. (1999): Data Envelopment Analysis of Non-Convex Technologies: With an Application to Finnish Super League Pesis Players, Helsinki School of Economics and Business Administration, Working Papers W-224

Kuosmanen, T. (2001): An Efficient MILP Algorithm for C-Convex DEA Models, an unpublished manuscript

Kuosmanen, T., and G.T. Post (1999): Robust Efficiency Measurement: Dealing with Outliers in Data Envelopment Analysis, Rotterdam Institute for Business Economic Studies, Report 9911

Petersen, N. C., (1990): Data Envelopment Analysis on a Relaxed Set of Assumptions, *Management Science*, 36(3), 305-314

Post, G. T. (2001a): Transconcave Data Envelopment Analysis, *European Journal of Operational Research* 132(2), 131-146

Post, G.T. (2001b): Estimating Non-convex Production Sets Using Transconcave DEA, *European Journal of Operational Research* (to appear)

Sandmo, A. (1971): On the Theory of the Competitive Firm under Price Uncertainty, *American Economic Review* 61, 65-73.

Seiford, L. M. (1996): The Most Influential DEA Publications: The Evolution and the State of the Art (1978-1995), *Journal of Productivity Analysis* 7, 99-137

Shephard. R.W. (1953): *Cost and Production Functions*, Princeton University Press, Princeton.

Shi, H., and X. Yang (1995): A New Theory of Industrialization, *Journal of Comparative Economics* 20(2), 171-189

Simar, L. and P. Wilson (2000): Statistical Inference in Nonparametric Frontier Models: The State of the Art, *Journal of Productivity Analysis* 13(1), pp. 49-78

Thompson, R.G., F.D. Singleton Jr., R.M. Thrall, and B.A. Smith (1986): Comparative Site Evaluations for Locating a High-Energy Physics Lab in Texas, *Interfaces* 16, 35-49

Thrall, R.M. (1999): What is the Economic Meaning of FDH? *Journal of Productivity Analysis* 11(3), 243-250

Varian, H (1984): The Nonparametric Approach to Production Analysis, *Econometrica* 52(3), 579-597

Yang, X. (1994): Endogenous vs. Exogenous Comparative Advantage and Economies of Specialization vs. Economies of Scale, *Journal of Economics* (*Zeitschrift für Nationalökonomie*) 60(1), 29-54

Yang, X., Y.-K. Ng (1993): *Specialization and Economic Organization*, Elsevier Science, Amsterdam

Yang, X., Y.-K. Ng (1995): Theory of the Firm and Structure of Residual Rights, *Journal of Economic Behavior and Organization* 26(1), 107-128

Yang, X., and R. Rice (1994): An Equilibrium Model Endogenizing the Emergence of a Dual Structure between the Urban and Rural Sectors, *Journal of Urban Economics* 35(3), 346-368

Zieschang, K. D. (1984): An Extended Farrell Technical Efficiency Measure, *Journal of Economic Theory* 33, 387-396

DEA WITH EFFICIENCY CLASSIFICATION PRESERVING CONDITIONAL CONVEXITY

Timo Kuosmanen*
Helsinki School of Economics and Business Administration,
Department of Economics and Management Science,
P.O. Box 1210, 00101 Helsinki, FINLAND
Tel. +358-9-431 38537
Fax. +358-9-431 38535
E-mail: kuosmane@hkkk.fi

A preprint of the article to be published in
European Journal of Operational Research,
Vol. 132, Iss. 3, 2001, pp. 83-99

Abstract: We propose to relax the standard convexity property used in Data Envelopment Analysis (DEA) by imposing additional qualifications for feasibility of convex combinations. We specifically focus on a condition that preserves the Koopmans efficiency classification. This yields an efficiency classification preserving conditional convexity property, which is implied by both monotonicity and convexity, but not conversely. Substituting convexity by conditional convexity, we construct various empirical DEA approximations as the minimal sets that contain all DMUs and are consistent with the imposed production assumptions. Imposing an additional disjunctive constraint to standard convex DEA formulations can enforce conditional convexity. Computation of efficiency measures relative to conditionally convex production set can be performed through Disjunctive Programming.

Key words: *Data envelopment analysis, Nonparametric efficiency analysis, Conditional convexity, Disjunctive Programming*

* This paper develops the line of argumentation sketched in the working paper Kuosmanen (1999), available from the author upon request. Knowledgeable comments from L. Cherchye, T. Post, L. Simar, J. Tind, and the two anonymous referees are gratefully acknowledged. This study was supported by the Graduate School in Systems Analysis, Decision making and Risk Management; the Foundation of Helsinki School of Economics and Business Administration; Yrjö Jahnsson's Foundation; and the Foundation for Economic Education.

1. INTRODUCTION

Data Envelopment Analysis (DEA), originating from Farrell's (1957) seminal work and popularized by Charnes, Cooper and Rhodes (1978), provides a flexible nonparametric doctrine for empirical production analysis. In recent decades, DEA has rapidly expanded towards new application areas (see e.g. Seiford (1996) for a survey). In addition to its original use in efficiency measurement, DEA is also employed for approximating production possibility sets or input/output correspondences (see Färe and Grosskopf (1995)), recovering shadow prices (see Färe, Grosskopf and Nelson (1990)), providing best-practice benchmarks (e.g. Bogetoft and Hougaard (1999)), as monitoring tool in agency problems (Bogetoft (1994)), and as a "lazy man's decision support tool" (Doyle (1995)).

The key contribution of DEA to efficiency analysis, and empirical production analysis in general, is the possibility to approximate unobservable production technologies from empirical input-output data of Decision Making Units (henceforth DMUs) without imposing overly restrictive parametric assumptions. See e.g. Bauer (1990) for a discussion of parametric frontier models. Still, DEA models usually impose *monotonicity* (i.e. free disposability of *all* inputs and outputs) and *convexity* assumptions, sometimes complemented with assumptions concerning returns to scale properties (see e.g. Seiford and Thrall, 1990).

It is to be noted that the basic determinism postulate of DEA, which states the production set should contain all observed Decision Making Units (DMUs), already suffices for many important functions of DEA. Additional properties such as monotonicity and convexity become useful when we wish to extend the scope of the analysis to assessments concerning the degree of inefficiency (using e.g. the radial Debreu-Farrell measures), and the shape of the "best-practice" production frontier. Unfortunately, these assumptions can be viewed overly restrictive, too. In fact, these properties can be violated in many important situations well recognized in economic theory. *Congestion* of production factors (see Färe and Svensson, 1980) can violate monotonicity. Many other features such as increasing marginal product of inputs, or indivisibility of inputs and outputs can violate convexity. To quote McFadden (1978, p. 8): "*Convexity… holds if the technology is such that substitution of one input combination for a second, keeping output constant, results in a diminishing marginal reduction in the second input combination, or if production activities can be operated side by side (or sequentially) without interfering with each other. However, the importance of* [monotonicity and convexity] *in traditional production analysis lies in their analytic convenience rather than in their economic realism.*"

Restrictiveness of the standard monotonicity and convexity conditions provides a substantial motivation to look for more general properties, on which DEA models could be more firmly based. In this paper we propose to relax the convexity property by imposing additional qualifications on technical feasibility of convex combinations of observed DMUs. The focus on convexity instead of monotonicity is justified by the fact that unlike monotonicity, convexity does not interfere with the static taxonomy of efficiency measures presented by Farrell (1957) and extended by Färe, Grosskopf and Lovell (1983, 1985). We specifically focus on developing a condition, which preserves the standard efficiency classification based on the classic Koopmans (1951) definition. This condition yields a very general conditional convexity property, which is implied by both monotonicity and convexity, but not conversely. Substituting convexity by conditional convexity, we construct various empirical DEA approximations as the minimal sets that contain all DMUs and are consistent with the imposed production assumptions. Imposing an additional disjunctive constraint to the convex DEA models can enforce conditional convexity. Computation of efficiency measures relative to this production set can be performed through Disjunctive Programming.

The rest of the paper unfolds as follows. Section 2 presents the necessary notation and terminology by reviewing the static taxonomy of efficiency measures. Section 3 discusses the role of convexity and monotonicity properties in the DEA reference technologies. Section 4 presents the central concept of this paper: viz. conditional convexity. In Section 4 we also discuss how an empirical production set can be constructed based on efficiency preserving conditional convexity. In Section 5 we formulate the Disjunctive Programming problems for characterizing various reference technologies and computing efficiency measures relative to them. Section 6 illustrates the approach by a simple numerical example. Finally, Section 7 draws our conclusive remarks.

2. STATIC TAXONOMY OF EFFICIENCY MEASURES

DEA models assess various performance dimensions of decision-making units (DMUs) that allocate inputs $x = (x_1 \ldots x_q)^T \in \Re_+^q$ to produce outputs $y = (y_1 \ldots y_p)^T \in \Re_+^p$. Inputs and outputs are assumed to be observable and to completely characterize the production process. In this paper we focus on the static or cross-sectional efficiency analysis. For dynamic performance evaluation, see e.g. Färe and Grosskopf (1996). This section presents the necessary conceptual apparatus for the subsequent sections by discussing the static taxonomy of technical efficiency. For sake of simplicity, we mostly focus on input side, but the framework laid down in this section extends to output and full input-output space as well, see e.g. Färe, Grosskopf and Lovell (1985) for further discussion.

In this paper we characterize the production technology in terms of the *production possibility set*

$$T = \left\{ (x, y) \in \Re_+^{q+p} \,\middle|\, \text{input } x \text{ can produce output } y \right\}.$$

The production set T is generally assumed to be closed and nonempty. Additional properties discussed in this paper include:

***Monotonicity*:** *A production set T is said to be monotonous if for all* (x,y): $(x, y) \in T \Rightarrow (x + u, y - v) \in T \; \forall u \in \Re_+^q, v \in \Re_+^p$.

Convexity: *Let X and Y denote a* $(q \times n)$ *input matrix and a* $(p \times n)$ *output matrix respectively, and let* X_j (Y_j) *denote the column j of X* (Y). *A production set T is said to be convex if for all X,Y:* $(X_j, Y_j) \in T \; \forall j = 1,...,n \Rightarrow (X\mathit{l}, Y\mathit{l}) \in T \; \forall \mathit{l} \in \Re_+^n, e\mathit{l} = 1$.

Constant Returns to Scale (CRS): *A production set T is said to exhibit constant returns to scale if* $T = aT, a > 0$.

Four remarks of these properties are worth noting. First, monotonicity is equivalent to free disposability of *all* inputs and outputs. We think of monotonicity as a property of a production set, while disposability is a property associated with inputs and outputs. Second, monotonicity and convexity properties can be straightforwardly associated with subsets of T (i.e. input or output correspondences) as well. Monotonicity (convexity) of production set naturally implies monotonicity (convexity) of these subsets, but the converse need not hold. Third, returns to scale properties are commonly associated with a production set, but also with production vectors, as discussed below in more detail. Finally, the basic DEA toolbox includes alternative returns to scale properties as well, see e.g. Seiford and Thrall (1990). In this paper we mainly focus on the constant returns to scale (CRS) property. We briefly note in Section 5 that the basic alternative properties are obtained by simple modifications to the models presented here.

In what follows we shall also employ an equivalent representation of production possibilities: the *input distance function* by Shephard (1953) defined as

$$D_T(x,y) = Sup\{q \in \Re_+ \,|\, (x/q, y) \in T\}.$$

If inputs are weakly disposable, i.e. $(x,y) \in T \Rightarrow (x/q, y) \in T \forall q \in (0,1]$, then $D_T(x,y) \geq 1$ is equivalent to $(x,y) \in T$, and $D_T(x,y) = 1$ characterizes the *isoquant* of the input correspondence (e.g. Färe (1988)).

Following Koopmans (1951) (see also Färe, 1988), the efficient subset of the production set *T* can be defined as

$$Eff.T = \{(x,y) \in T \,|\, x' \leq x, y' \geq y, (x',y') \neq (x,y) \Rightarrow (x',y') \notin T\}.$$

In the subsequent sections we will also employ another relevant subset, the weak efficient subset defined as (e.g. Färe, 1988):

$$WEff.T = \{(x,y) \in T \,|\, x' < x, y' > y \Rightarrow (x',y') \notin T\}.$$

Clearly, $WEff.T \supseteq Eff.T$. That is, efficiency implies weak efficiency, but the reverse relationship should not necessarily hold.

There are a number of alternative measures for gauging the degree of inefficiency (see e.g. De Borger et al. (1998)). In this paper we confine attention to the standard Debreu-Farrell input measure, which is simply the inverse of the input distance function, i.e.

$$DF_T(x,y) = D_T(x,y)^{-1}. \quad (1)$$

As noted by Färe and Lovell (1978), the Debreu-Farrell input measure can fail to fully account for all inefficiency in the sense of Koopmans. Most notably, the Debreu-Farrell measure may equal unity for an inefficient DMU, and the reference point $(DF_L(x,y)x,y)$ does not necessarily belong to the efficient subset. Nevertheless, the remaining nonradial 'slacks' are in the original Farrell's framework captured by the allocative efficiency component, which depends on the objective function of the producer. In this paper we leave the explicit producer's objective unspecified, and hence focus on technical efficiency.

Färe, Grosskopf and Lovell (1983, 1985) extended the static taxonomy to include additional notions of *structural efficiency* and *scale efficiency*. For these purposes, define the *monotone hull* (*MH*) of *T* as the smallest monotone set that contains *T*, i.e.

$$MH(T) = \{(x,y) \,|\, x = x'+u; y = y'-v; (x',y') \in T; (u,v) \in \Re_+^{q+p}\}.$$

Next, define the smallest monotone set that exhibits CRS and contains T as the *ray-unbounded monotone hull* (RMH), i.e.

$$RMH(T) = \{(x, y) | (x, y) \in lMH(T), l > 0\}.$$

Structural efficiency (STR) and scale efficiency (SCA) can now be defined as

$$STR_T(x, y) = \frac{DF_{MH(T)}(x, y)}{DF_T(x, y)} \text{ and}$$

$$SCA_T(x, y) = \frac{DF_{RMH(T)}(x, y)}{DF_{MH(T)}(x, y)}$$

respectively. Production is diagnosed structurally efficient if and only if it occurs in a non-congested or "economic" region of production, where marginal product of every input is non-negative. Moreover, production is diagnosed scale efficient if and only if production takes place at *the most productive scale size* (see Banker et al. (1984) for discussion). Note that producing on the most productive scale need not always comply with the primary objectives of the producer (e.g. profit maximization, cost minimization), and consequently, scale efficiency should not be automatically included in overall efficiency criteria. For further details of structural and scale efficiency measures, see Färe, Grosskopf and Lovell (1983, 1985).

3. DEA APPROACH

The previous section phrased in terms of a theoretical production set T. Too often, the true production technology cannot be characterized precisely enough by engineering blueprints or other theoretical knowledge of the production process. The key feature of DEA is that it allows us to approximating the production technology directly from the observed production data. Let our observed sample consist of n DMUs, and denote the matrix of output vectors by $Y = (y^1 \cdots y^n)$ and the matrix of input vectors by $X = (x^1 \cdots x^n)$. We use $l = (l^1 \cdots l^n)^T$ to denote the vector of intensity variables, and $e = (1 \cdots 1)$. Finally, $S = \{1,...,n\}$ denotes an index set of cardinality n.

In deterministic DEA it is generally assumed that all observed production vectors are technically feasible, i.e. $(X_j, Y_j) \in T \; \forall j \in S$. This seems reasonable if the relevant input-output variables can be measured accurately enough; after all, these input-output combinations are observed. As already suggested by Tulkens and Vanden Eeckaut (1999)**,** this basic assumption already allows one to distinguish between efficient, weak efficient, and inefficient DMUs respectively, i.e.

$$Eff.(X, Y) = \{j \in S | \nexists i \in S : x_i \leq x_j, y_i \geq y_j, (x_i, y_i) \neq (x_j, y_j)\}, \tag{2}$$

$$WEff.(X, Y) = \{j \in S | \nexists i \in S : x_i < x_j, y_i > y_j\}, \tag{3}$$

$$Ineff.(X,Y)=\{j\in S \mid j\notin WEff.(X,Y)\}. \tag{4}$$

Distinguishing these subsets does not require any assumptions other than the determinism condition discussed above. Consequently, we will refer to these subsets as the 'spontaneous' efficiency classification. These sets are sufficient as such e.g. for benchmarking purposes (see Bogetoft and Hougaard (1999)). Also ordinal ranking of DMUs based on the number of dominating/dominated DMUs is possible (Tulkens and Vanden Eeckaut (1999)).

Further assumptions are typically necessary for measuring the (cardinal) degree of inefficiency in terms of the standard taxonomy. The original DEA models formulated by Farrell (1957) and Charnes et al. (1978) were based on the maintained assumption that the production set T satisfies all three additional properties given in Section 2: monotonicity, convexity, and CRS. This assumption allows us to approximate the production set by the *ray-unbounded convex monotone hull* (henceforth *RCMH*) of the observed production vectors, i.e.

$$RCMH(X,Y)=\left\{(x,y)\in\Re_{+}^{q+p}\left|\begin{pmatrix} y \\ -x\end{pmatrix}\leq\begin{pmatrix} Yl \\ -Xl\end{pmatrix}; l\in\Re_{+}^{n}\right.\right\}.$$

Consistent with the *minimum extrapolation principle* (see Banker et al. (1984) and Bogetoft (1994)), *RCMH*(*X,Y*) is the minimal set that contains all observations and complies with the imposed properties. Although *RCMH*(*X,Y*) does not impose any parametric structure on production possibilities, postulating T to be monotonous and convex and to exhibit CRS is quite restrictive. Nevertheless, for all convex technologies, *RCMH*(*X,Y*) is contained in the congestion and scale adjusted reference set *RMH*(*T*) used in the static decomposition.

In the last two decades, a considerable development towards more general properties has taken place in DEA. Perhaps the most standard DEA reference technology, which relaxes the assumption of CRS but maintains assumptions of monotonicity and convexity, is the *convex monotone hull* (henceforth *CMH*) (see Afriat (1972), Färe, Grosskopf, and Logan (1983), and Banker et al. (1984)) defined as

$$CMH(X,Y)=\left\{(x,y)\in\Re_{+}^{q+p}\left|\begin{pmatrix} y \\ -x\end{pmatrix}\leq\begin{pmatrix} Yl \\ -Xl\end{pmatrix}; el=1; l\in\Re_{+}^{n}\right.\right\}.$$

Obviously, $CMH(X,Y)\subseteq RCMH(X,Y)\ \forall(X,Y)$. Moreover, for convex T, *CMH*(*X,Y*) is contained within *MH*(*T*). Hence, *CMH*(*X,Y*) can be viewed as a proxy for the congestion adjusted reference set *MH*(*T*) in the static taxonomy.

We can proceed by relaxing monotonicity from CMH, which yields the *convex hull* (*CH*) (Charnes et al. (1985)), i.e.

$$CH(X,Y)=\left\{(x,y)\in\Re_{+}^{q+p}\left|\begin{pmatrix} y \\ -x\end{pmatrix}=\begin{pmatrix} Yl \\ -Xl\end{pmatrix}; el=1; l\in\Re_{+}^{n}\right.\right\}.$$

Now, $CH(X,Y)\subseteq CMH(X,Y)\ \forall(X,Y)$. Moreover, for convex T, *CH*(*X,Y*) is contained in T, which justifies its use as a reference technology. Unfortunately, the assumption of convex T is almost void of economic realism. Like noted by McFadden (1978), it is mainly used for its analytical convenience. See Guesnerie (1975) and Cherchye et al. (2000), for further discussion.

In application oriented DEA, empirical realism should outweigh analytical convenience. Perhaps the most important drawback of this property lies in the fact that convexity typically interferes with the spontaneous efficiency classification. Hence, an erroneously imposed convexity assumption can have a substantial impact on policy recommendations. Dramatic effects of the convexity postulate have been reported in empirical studies. See e.g. Deprins et al. (1984), Tulkens (1993), Dekker and Post (1999), and Kuosmanen (1999). For example, in the retail banking application reported by Tulkens (1993, p. 192-197), 74.6% of public bank branches were found efficient assuming monotonicity only, but only 5.2% of all bank branches remained efficient when both monotonicity and convexity were imposed. For private bank branches, the same study reports 57.8% average efficiency without convexity, and 5.5% efficiency with convexity. We do not know which of these results are more correct, but it obviously takes a lot of faith to maintain convexity hypothesis without any suspicion.

Consequently, an alternative strategy has been considered in the literature, see e.g. Deprins et al. (1984) and Tulkens (1993). In this approach it is the convexity property that is relaxed, and monotonicity that is maintained. This gives the so-called Free Disposable Hull (FDH), which we here dub for sake of uniformity as *monotone hull* (henceforth *MH*), i.e.

$$MH(X,Y) = \left\{ (x,y) \in \Re_{+}^{q+p} \left| \begin{pmatrix} y \\ -x \end{pmatrix} \leq \begin{pmatrix} Yl \\ -Xl \end{pmatrix}; el = 1;\ l^{j} \in \{0,1\};\ j \in S \right. \right\}.$$

A particularly attractive feature of monotonicity is its redundancy in the standard Koopmans notion of technical efficiency. That is, in contrast to convexity, monotonicity does not interfere with the efficient, weak efficient, or inefficient subsets (2)-(4).

If T is monotonous, it contains $MH(X,Y)$. Nevertheless, monotonicity of T can be viewed equally unrealistic as convexity. See e.g. Färe and Svensson (1980) for discussion on congestion. According to Färe, Grosskopf and Lovell (1983), examples of efficiency losses due to congestion include traffic congestion in the production of transportation, reduced grain yield due to excessive fertilization in agriculture, and output losses due to featherbedding and other union work rules.

From the applied point of view, *MH* has also other serious limitations. One frequently cited shortcoming is that after accounting for potential radial inefficiency measured by the Debreu-Farrell measure (1), considerable non-radial slacks typically remain (see e.g. De Borger et al. (1998) for discussion). Secondly, the marginal properties of *MH* are rather bizarre: Consider for example the *scale elasticity* measure that can be defined in terms of the distance function and its gradient (see e.g. Färe, Gosskopf, and Lovell (1988)) as

$$e(x,y) = -\frac{D_T(x,y)}{\nabla D_T(x,y) \cdot (0,y)}. \qquad (5)$$

Scale elasticity is an important indicator of returns to scale: increasing, constant or decreasing returns to scale are said to prevail when $e > 1$, $e = 1$ or $e < 1$. In case of *MH*, however, practically useful approximation of scale elasticity cannot be obtained due to the discrete nature of *MH*. The same problem equally concerns the elasticities of substitution and transformation. Thirdly, *MH* is subject to a considerable small sample error, which raises a need for large data sets. Although statistical procedures to deal with the third limitation have been developed (see e.g. Park et al. (forthcoming), Gijbels et al. (1998), Simar and Wilson (1998)), at least the first two shortcomings remain.

Finally, a number of intermediate models that typically intend to weaken the general convexity assumption underlying *CMH*(*X,Y*) have been presented. In the literature on nonparametric production analysis, Hanoch and Rothschild (1972) and Varian (1984) have discussed the possibility of imposing convexity on the input correspondence, which does not necessarily imply convexity of the production set. In the DEA literature, Petersen (1990), Bogetoft (1996), Bogetoft et al. (forthcoming), and Post (forthcoming) have further developed the idea of convex input/output correspondences of a nonconvex production set. Furthermore, Kerstens and Vanden Eeckaut (1998, 1999) have presented models that impose monotonicity and CRS, but not convexity.

Consider, for example, a production set that starts from *MH* and additionally imposes convexity of input correspondences, i.e.

$$CIMH\,(X,Y) = \left\{ (x,y) \in \Re_{+}^{q+p} \middle| x \geq Xl\,;\, l^{j} y \leq l^{j} y^{j} \;\; \forall j \in S; el = 1; l \in \Re_{+}^{n} \right\} \;\; .$$

There are some special occasions where convexity of input correspondence can be a reasonable and harmless requirement, e.g. when the managerial objective function is linear in inputs or outputs. See for example Hanoch and Rothschild (1972) or Kuosmanen and Post (1999a) for discussion on the economic interpretation of *CIMH*(*X,Y*) in terms of cost efficiency. Unfortunately, the general justification of these partially convex models seems problematic, especially as the analysis of pure technical efficiency is concerned. These models clearly combine the negative features such as ignorance of congestion and ad hoc specification of convexity, but the concrete benefits of these alternatives are unclear.

Yet, it has been suggested that convexity of input and/or output correspondences would be a standard property of technology. Specifically, Petersen (1990, p. 307) refers to the microeconomic *"law of diminishing marginal rates of substitutions"* that would imply such property. Also Bogetoft (1996) and Bogetoft et al. (forthcoming) cite this law as the motivation for their production assumptions. However, to the best of our knowledge, such law does not exist in economic theory. There are such classic laws of *diminishing returns* and *variable proportions* (see e.g. Schumpeter, 1966), but (as pointed out by both referees) these laws are related to monotonicity or lack of it (see Färe and Svensson, 1980) rather than convexity.

To demonstrate the difficulties with a maintained hypothesis of convex input correspondence, consider the following simple exercise in differential calculus. As noted above, the production frontier can be represented using the input distance function, i.e. $D_T(x,y) = 1$, akin to a *production function* in an implicit form. For analytical convenience, assume D_T is twice continuously differentiable. The following general expression can be derived by differentiating the implicit production relation $D_T = 1$ with respect to inputs *i* and *j*:

$$\left. \frac{dx_i}{dx_j} \right|_{D(x,y)=1} = - \frac{\partial D_T(x,y)/\partial x_j}{\partial D_T(x,y)/\partial x_i} \,. \tag{6}$$

This expression defines the marginal rate of substitution between inputs *i* and *j* in the neighborhood of the frontier point (*x*, *y*), i.e. the slope of the input isoquant. When (*x*, *y*) is non-congested, the distance function increases in *x*, and the partial derivatives in (6) are positive, which gives the usual decreasing input isoquants. Differentiating (6) again with respect to x_j (omitting the arguments (*x*,*y*) in D_T) gives

$$\left.\frac{d^2 x_i}{dx_j^2}\right|_{D(x,y)=1} = \frac{\left(\partial^2 D_T / \partial x_i \partial x_j\right)\cdot\left(\partial D_T / \partial x_j\right) - \left(\partial^2 D_T / \partial x_j^2\right)\cdot\left(\partial D_T / \partial x_i\right)}{\left(\partial^2 D_T / \partial x_i^2\right)^2}.$$

This second derivative reflects the curvature of the isoquant. Convexity of the input isoquant requires nonnegative second derivative for all *x*. Clearly, the sign of the nominator is determinant in this respect.

Now what happens in case of increasing returns if the marginal product of input *j* is increasing, i.e. $\partial^2 D_T / \partial x_j^2 > 0$? Consider a noncongested point (*x*,*y*) where the both first-order partial derivatives are positive. Convexity of the input isoquant is preserved if

$$\partial^2 D_T / \partial x_i \partial x_j \geq \frac{\partial D_T / \partial x_i}{\partial D_T / \partial x_j} \partial^2 D_T / \partial x_j^2 .$$

In case the marginal product is increasing, the right hand side of this inequality is strictly positive. The left-hand side of the inequality represents the pure substitution effect in production, which could be positive or negative depending on the nature of inputs. Inputs *i* and *j* can be called technical complements *if* $\partial^2 D_T / \partial x_i \partial x_j \geq 0$, and technical substitutes if $\partial^2 D_T / \partial x_i \partial x_j \leq 0$. Clearly, *all* inputs must be sufficiently complementary in order to preserve convexity of the input isoquant.

In verbal terms, increasing marginal products can provide substantial incentives for utilizing scale economies by specialization. In case of substitute inputs, the specialization and substitution effects are mutually enhancing, which results as a concave input isoquant (non-convex input correspondence). In case of complementary inputs, however, there is a conflict between substitution and specialization effects. When convexity of the whole production set is relaxed, maintaining convexity of the input correspondence requires extreme complementarity in all inputs that always outweighs any scale effect. This seems a difficult assumption, as the relative magnitudes of the scale and substitution effects are difficult (if not impossible) to ascertain a priori.

4. CONDITIONAL CONVEXITY

As sufficiently general empirical production sets apparently cannot be built on the standard monotonicity and convexity properties, it is worth to consider alternative weaker technology properties. Since both monotonicity and CRS properties have already been given an explicit interpretation in terms of the static taxonomy, it seems fruitful to focus on the remaining convexity property. This section proposes a relaxation of the convexity property, which we call *conditional convexity.*

Definition *(General conditional convexity)**:** Let C denote an arbitrary well-defined logical condition. Production set T is convex conditional upon C, if for all X,Y* and $l \in \Re_{+}^{n}, el = 1$: $(X,Y) \in T \wedge C \Rightarrow (Xl, Yl) \in T$.

Clearly, if production set is convex, it is also conditionally convex for any well-defined *C*, but the converse need not be true.

To operationalize the general conditional convexity property, we need to specify the condition *C* explicitly. One particularly interesting possibility is to enforce the spontaneous technical efficiency classification by putting convexity conditional upon its preservation. This condition is motivated by the argument that the appeal of the frequently imposed monotonicity property mostly underlies in the similar feature, rather than in its economic realism. Hence, in this paper we focus on exploring convexity conditional specifically upon preservation of efficiency classification. In terms of the above definition, the condition C is defined as

Definition (*Condition* C^{EP}): $Xl \not< x, Yl \not> y \;\; \forall (x,y) \in WEff.T$

Substituting *C* by Condition C^{EP} in the definition of general conditional convexity gives the *efficiency classification preserving conditional convexity* property. For sake of brevity, we will henceforth refer to this property by the abbreviated term *c-convexity*. Specifically, this property states that if no weak efficient production plan is dominated by a convex combination of feasible production plans contained in the production set, then the convex combination is feasible.

As noted above, convexity implies c-convexity, but the converse need not be true. Interestingly, the same holds for monotonicity, as the following proposition demonstrates:

PROPOSITION: If $T \subset \Re_{+}^{p+q}$ represents a closed, nonempty, monotone production set, then *T* is c-convex.

Proof: Let *X'* and *Y'* denote input and output matrices of an arbitrary subset of *m* production vectors $(X'_j, Y'_j) \in T \; \forall j = 1,...,m$, and let $(x(l), y(l))$ denote an arbitrary convex combination $x(l) = X'l; y(l) = Y'l; el = 1; l \in \Re_{+}^{n}$. The production set *T* is c-convex if for all $l, (x,y) \in WEff.T$: $x(l) \not< x, y(l) \not> y \Rightarrow (x(l), y(l)) \in T$. By monotonicity of *T* and the definition of the weak efficient subset, $(\tilde{x}, \tilde{y}) \notin T$ is equivalent to $\exists (x,y) \in WEff.T : \tilde{x} < x, \tilde{y} > y$. That is, an infeasible production vector necessarily dominates at least one weak efficient production vector. If this is not true for $(x(l), y(l))$, it immediately follows that $(x(l), y(l)) \in T$. *Q.E.D.*

As c-convexity does not generally imply monotonicity (see e.g. Figure 1), c-convexity can be viewed by the previous proposition as a more general property of production sets than monotonicity or convexity.

Now, how could we use c-convexity in DEA? In fact, empirical DEA production sets can be constructed on c-convexity in a similar fashion as on standard convexity. Note that the c-convexity depends on the weak efficient subset of technology. As the theoretical weak efficient subset is unknown, we resort to its empirical approximation *WEff*.(*X*, *Y*). The minimal set containing all DMUs and satisfying c-convexity is the *c-convex hull* (*CCH*) defined as

$$CCH(X,Y)=\left\{(x,y)\in\Re_{+}^{q+p}\left|\begin{pmatrix}y\\-x\end{pmatrix}=\begin{pmatrix}Yl\\-Xl\end{pmatrix}\right.;el=1;\ l\in\Re_{+}^{n};C^{EP}\right\}. \tag{7}$$

Imposing monotonicity in addition to c-convexity gives the *c-convex monotone hull* (*CCMH*)

$$CCH(X,Y)=\left\{(x,y)\in\Re_{+}^{q+p}\left|\begin{pmatrix}y\\-x\end{pmatrix}\leq\begin{pmatrix}Yl\\-Xl\end{pmatrix}\right.;el=1;\ l\in\Re_{+}^{n};C^{EP}\right\}. \tag{8}$$

Finally, imposing CRS in addition to c-convexity and monotonicity yields the *ray-unbounded c-convex monotone hull* (*RCCMH*), i.e.

$$RCCMH(X,Y)=\left\{(x,y)\in\Re_{+}^{q+p}\left|\begin{pmatrix}y\\-x\end{pmatrix}\leq a\begin{pmatrix}Yl\\-Xl\end{pmatrix}\right.;a>0;el=1;\ l\in\Re_{+}^{n};C^{EP}\right\}. \tag{9}$$

Figure 1 illustrates *CCH, CCMH* and *RCCMH* sets in a single input-output case. Black and white dots represent DMUs, the black ones are weak efficient, the white ones are inefficient. The dashed line with short gaps represents the boundary of *CCH*, the solid line represents the boundary of *CCMH*, and the straight line with long dash represents the *RCCMH* frontier.

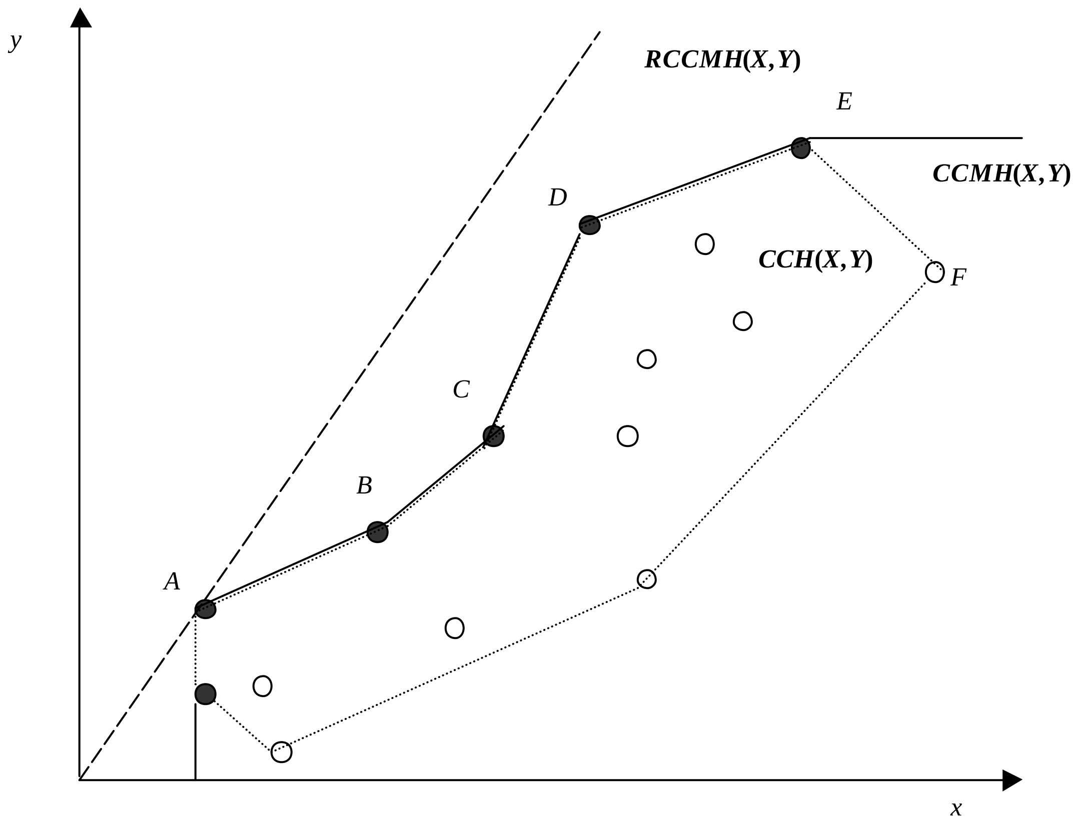


Figure 1: Illustration of CCH, CCMH and RCCMH sets

Three points are worth noting: First, both c-convexity and monotonicity preserve efficiency status of all weak efficient DMUs, but that feature need not carry over to the CRS property, like Figure 1 illustrates. Note that preserving efficiency classification under CRS is unnecessary. Note also that *RCCMH* differs from the CRS model by Charnes et al. (1978) in that it need not be convex in input and output spaces. Second, *CCH* and *CCMH* production sets do not restrict returns to scale to behave monotonically as *CMH* does: From A to C via B these technologies exhibit decreasing returns to scale, from C to D increasing returns prevail, while from D to E again decreasing returns prevail. In *CMH* there can only be a single increasing part followed by a decreasing part, which is a strong restriction. Third, *CCH* set indicates DMU *F* could suffer from congestion, which we would overlook by resorting to *MH* or other monotone approximations. However, one should take the inherent small sample error into account before drawing conclusions based on these methods.

Consider properties of these c-convex approximations. Note first that unlike monotonicity or convexity, c-convexity of *T* does not guarantee $CCH(X,Y)$ would be contained within *T*. It is somewhat unfortunate that the attractive general character of the c-convexity property does not fully carry over to finite sample approximations. Still, c-convexity remains a more general property than convexity even in small samples. Clearly, all c-convex empirical sets are contained within their corresponding convex counterparts, i.e. $CCH(X,Y) \subseteq CH(X,Y)$, $CCMH(X,Y) \subseteq CMH(X,Y)$, and $RCCMH(X,Y) \subseteq RCMH(X,Y)$ for all (X,Y). This further implies that the c-convex approximations yield higher efficiency measures than their convex counterparts, e.g. $DF_{CCMH}(x,y) \geq DF_{CMH}(x,y)$ for all (x,y). Naturally, the c-convex empirical sets coincide with their convex counterparts if no convex combination interferes with the efficiency classification.

As for monotonicity, *CCH* need not contain *MH*. Conversely, *MH* need not contain *CCH* either. Consequently, although c-convexity is a more general property of theoretical production sets than monotonicity, this need not be the case in empirical approximations. Nevertheless, as *CCMH* is monotonous, it necessarily contains *MH*. Note that although monotonicity of *T* implies conditional convexity, *MH* need not contain *CCMH*, as a set of discrete DMUs presented by (X,Y) does not satisfy monotonicity. Still, we conjecture that if *MH* converges to a true monotone production set *T* as the sample size increases (see Park et al. (forthcoming) for sufficient conditions), then so does *CCMH*. Note that *MH* and *CCMH* coincide if every vertex of the *MH* frontier represents an observed DMU. The relationship between *MH* and *CCMH* can be further clarified by the fact that *MH* preserves the efficiency classification of both the observed DMUs and all non-observed production plans of the input-output space, whereas *CCMH* preserves the efficiency classification of the observed DMUs only.

It should be already evident that the standard static taxonomy of efficiency is available for the c-convex production sets along the lines of Section 2. Now how would imposing c-convexity instead of convexity influence the decomposition? As for structural efficiency, there is no difference whatsoever. This is simply because the condition C^{EP} cannot influence the convex combinations of DMUs lying on the congested or uneconomical part of the isoquant. That is, the backward bending boundaries of a convex set and a c-convex set are always exactly the same. However, for measuring technical and scale inefficiencies, the convexity assumption can have a dramatic effect. Obviously, an erroneously imposed convexity postulate can only decrease the measured overall "technical" inefficiency (OTE), i.e. $OTE=DF*STR*SCA$. Thus, erroneously specified convexity can both underrate overall efficiency, and distort the decomposition by overvaluing the technical and scale inefficiencies relative to the structural component. These effects of convexity should be taken into account in interpretation of results. The more general c-convexity property could help to remedy some of the potential specification errors.

Finally, note that it is possible to further relax our assumptions by applying c-convexity on the input or output isoquants, but not for the whole production set *T*. Following Bogetoft (1996), we could assume monotonicity of the production set, and apply c-convexity to inputs only. However, for sake of brevity we abstract here from these obvious alternative formulations.

5. EFFICIENCY MEASUREMENT

In this section we show that efficiency measures relative to c-convex production sets can be computed by Disjunctive Programming (DP). We also briefly discuss how the optimal solution to a DP problem can be inferred from optimal solutions to a series of Linear Programming (LP) problems akin to the standard DEA LP formulations.

In principal, *CCH* is obtained from *CH* by simply excluding all efficiency classification violating convex combinations. We can test whether any convex combination violates efficiency status of a weak efficient DMU j by solving the following Linear Programming test problem:

$$\begin{aligned} &\max_{q,l} \; q \\ &s.t.\; (1-q)x^j = Xl \\ &(1+q)y^j = Yl \\ &el = 1 \\ &l \in \Re_+^w \end{aligned} \tag{10}$$

This test problem uses the graph measure by Briec (1997) as an instrument. Let q^*, l^* represent the optimal solution to problem (10). $q^* > 0$ indicates that the a convex combination of the reference DMUs $h = \{i \in S \mid l^{i*} > 0\}$ violates the efficiency classification. Consequently, that combination should be eliminated as infeasible. This elimination can be enforced by imposing to (10) an additional constraint

$$l \in \bigcup_{i \in h} \{l \mid l^i = 0\}. \tag{11}$$

That is, at least one of the current reference DMUs is excluded from the reference set by constraining the lambda weight equal to zero.

We can solve the test problem (10) again together with the constraint (11), and subsequently impose additional constraints until the optimal solution to the test problem equals zero. Hence, potential non-uniqueness of any reference set h does not matter. This procedure is repeated for all weak efficient DMUs. Let h_k denote the resulting set of indices $j \in S$ of DMUs that form an efficiency classification violating combination k, $k = 1,\ldots,r$. Furthermore, denote the set of efficiency classification violating combinations by $H = \{h_1, \ldots, h_r\}$. Elimination of combinations $h \in H$ can be enforced by invoking a so-called *disjunctive constraint* (in the conjunctive normal form) written as

$$l \in \bigcap_{h_k \in H} \left[\bigcup_{j \in h_k} \{ l \,|\, l^j = 0 \} \right].$$

Incorporating this constraint to *CH* gives *CCH* an equivalent formulation to (7), i.e.

$$CCH(X,Y) = \left\{ (x,y) \in \Re_+^{q+p} \,\middle|\, \begin{pmatrix} y \\ -x \end{pmatrix} = \begin{pmatrix} Yl \\ -Xl \end{pmatrix}; el = 1; \; l \in \bigcap_{h_k \in H} \left[\bigcup_{j \in h_k} \{ l \,|\, l^j = 0 \} \right]; \; l \in \Re_+^n \right\}.$$

Similarly, we can write *CCMH* equivalent to (8) as

$$CCMH(X,Y) = \left\{ (x,y) \in \Re_+^{q+p} \,\middle|\, \begin{pmatrix} y \\ -x \end{pmatrix} \leq \begin{pmatrix} Yl \\ -Xl \end{pmatrix}; el = 1; \; l \in \bigcap_{h_k \in H} \left[\bigcup_{j \in h_k} \{ l \,|\, l^j = 0 \} \right]; \; l \in \Re_+^n \right\}.$$

Finally, *RCCMH* can be written equivalent to (9) as

$$RCCMH(X,Y) = \left\{ (x,y) \in \Re_+^{q+p} \,\middle|\, \begin{pmatrix} y \\ -x \end{pmatrix} \leq \begin{pmatrix} Yl \\ -Xl \end{pmatrix}; \; l \in \bigcap_{h_k \in H} \left[\bigcup_{j \in h_k} \{ l \,|\, l^j = 0 \} \right]; \; l \in \Re_+^n \right\}.$$

Computation of the standard efficiency measures relative to *CCH*, *CCMH* or *RCCMH* involves solving a DP problem for each DMU. For example, the Debreu-Farrell input measure for DMU j can be computed relative to *CCMH* as the optimal solution to the problem

$$\begin{aligned} &\min_{q,l} \; q \\ &s.t. \; qx^j \geq Xl \\ &y^j \leq Yl \qquad (12) \\ &el = 1 \\ &l \in \Re_+^w \\ &l \in \bigcap_{h_k \in H} \left[\bigcup_{j \in h_k} \{ l \,|\, l^j = 0 \} \right] \qquad (13) \end{aligned}$$

The analogous *CCH* measure is simply obtained by substituting inequalities in (12) by equalities, while the *RCCMH* measure is obtained by deleting the constraint $el = 1$. Note that increasing (decreasing) returns to scale can be enforced by substituting this constraint by $el \geq 1$ ($el \leq 1$). Also alternative orientations or non-radial gauges can be computed relative to *RCCMH*, *CCMH* or *CCH* in a straightforward fashion.

Optimization theory knows a number of alternative finitely converging algorithms for solving Linear Programming problems with disjunctive constraints. See e.g. Sherali and Shetty (1980) for details. Let us suffice here to point some brief remarks on a basic relaxation principle. Note first that relaxing the disjunctive constraint (13) yields an ordinary DEA LP problem (12). The disjunctive constraint can be viewed to represent alternative sets of constraints, of which at least one must be satisfied. These constraints preserve the linear structure, enforcing lambda weights of the particular subsets of DMUs equal to zero. In fact, this is equivalent to excluding the corresponding subsets from the data matrices *X* and *Y*. These alternative data sets span a number of convex sets in the input-output space, which could be viewed as some kind of sub-technologies. For example, the *CCMH* set of Figure 1 can be presented as a union of three CMHs spanned by subsets (A,B), (B,C), and (C,D,E) respectively. A straightforward, albeit computationally hard way is simply to compute efficiency measures relative to all alternative sub-technologies i.e. by solving a relaxed version of LP problem (12). All these sub-technologies are feasible, so the minimum of the resulting optimal solutions is the optimal solution to the DP problem (12) with constraint (13). As a conclusion, the optimal solution to the DP problem can be inferred from optimal solutions to a series of LP problems.

The explicit characterization of c-convex technologies can be convenient for many 'predictive' purposes of DEA. However, it also entails a substantial computational burden. Computationally superior algorithms could be developed if we ignore the explicit characterization, and focus on solving efficiency measures as usual in DEA (see Kuosmanen (1999) for some heuristics). While this indeed seems an interesting avenue to explore further, we will abstract from this issue in the current paper and leave it for future research.

As for recovering shadow prices or approximating the elasticities of the frontier, c-convex reference sets provide equally fit frameworks as their convex counterparts, although the results can differ considerably. For simplicity, we have only considered the envelopment side (primal) problems here, but it is trivial to infer the optimal multiplier (dual) weights using the information of the optimal l vector, i.e. find the minimal dimensional hyperplane that supports all reference DMUs and the projected reference point of the evaluated DMU. In practice, one can solve the multiplier (dual) DEA formulation relative to the optimal reference set. Alternatively, it is possible to use the multiplier formulation throughout instead of (12), by modifying the data matrices instead of imposing restrictions on lambda weights, recalling that restricting a lambda weight equal to zero corresponds to excluding the corresponding DMU from the data. The multiplier weights can be used for example for computing elasticity of scale (5), or for inferring the qualitative (local) returns to scale properties. Moreover, the multiplier weights can be given a shadow price interpretation, which can be motivated by price endogeneity or uncertainty (see Kuosmanen and Post (1999b) for further discussion). Although recovering shadow prices is one promising by-product of the proposed model, its specific treatment falls beyond the scope of this paper, and so we leave it also for future research.

6. ILLUSTRATIVE EXAMPLE

We illustrate the practice of efficiency measurement relative to the c-convex reference sets by a simple example. For sake of brevity we focus on the *CCMH* set solely. For a more extensive application of the approach to a real-world data of 194 Finnish pesis batters (a game akin to baseball) with a single-input-three-output technology, see Kuosmanen (1999).

Consider the single-input single-output data of 5 DMUs given in Table 1, and illustrated by Figure 2. It is easy to see that DMUs #1, #2, #3, and #4 are efficient, and hence also weak efficient. The only inefficient one is the DMU #5. Now what would be the degree of inefficiency of this unit? In terms of the Debreu-Farrell input measure, *MH* efficiency amounts to 81.8%, while *CMH* efficiency is only 45.5%. Clearly, the convexity assumption makes a big difference. Also DMUs #2 and #3 would appear inefficient if convexity were imposed. Let us now compute efficiency measure of DMU #5 for the *CCMH* model.

Table 1: Example data set

	DMU 1	DMU 2	DMU 3	DMU 4	DMU 5
X	2	6	9	11	11
Y	2	3	6	11	5

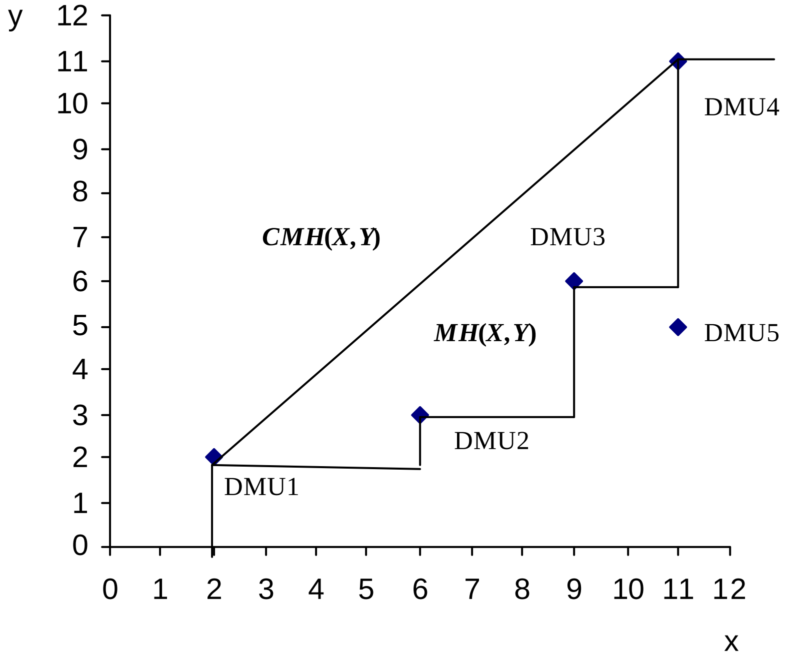


Figure 2: CMH and MH frontiers in the example data set

We start by characterizing the feasible convex subsets as explained in the previous section. We solve the test problem (10) for all efficient DMUs, excluding the ineficient DMU #5 from the reference data. For DMU j, j =1,2,3,4, the test problem reads:

Problem 6.1

$$\underset{q,l}{Max}\ q$$

$$s.t.\ (1-q)x_j \geq 2l_1 + 6l_2 + 9l_3 + 11l_4$$

$$(1+q)y_j \leq 2l_1 + 3l_2 + 6l_3 + 11l_4$$

$$l_1 + l_2 + l_3 + l_4 = 1$$

$$l_1, l_2, l_3, l_4 \geq 0$$

For DMUs #1 and #4 the optimal solution to this problem is $q = 0$. For DMUs #2 and #3 the optimal solutions are: DMU #2:$q = 1/3$; $l = \left(7/9, 0, 0, 2/9\right)$; DMU #3: $q = 1/5$; $l = \left(37/45, 0, 0, 8/45\right)$. As for these two DMUs $q > 0$, the convex combination of DMUs #1 and #4, which have lambda weights greater than zero in both solutions, must be eliminated by constraining $l_1 = 0$ or $l_4 = 0$.

We proceed by solving two new problems with restriction $l_1 = 0$ imposed in Problem 6.2, and restriction $l_4 = 0$ imposed in Problem 6.3. For DMU *j*, *j* = 2,3, the two test problems read:

Problem 6.2

$$\underset{q,l}{Max}\ q$$
$$s.t.\ (1-q)x_j \geq 2l_1 + 6l_2 + 9l_3 + 11l_4$$
$$(1+q)y_j \leq 2l_1 + 3l_2 + 6l_3 + 11l_4$$
$$l_1 + l_2 + l_3 + l_4 = 1$$
$$l_1, l_2, l_3, l_4 \geq 0$$
$$l_1 = 0$$

Problem 6.3

$$\underset{q,l}{Max}\ q$$
$$s.t.\ (1-q)x_j \geq 2l_1 + 6l_2 + 9l_3 + 11l_4$$
$$(1+q)y_j \leq 2l_1 + 3l_2 + 6l_3 + 11l_4$$
$$l_1 + l_2 + l_3 + l_4 = 1$$
$$l_1, l_2, l_3, l_4 \geq 0$$
$$l_4 = 0$$

The optimal solutions are the following. Problem 6.2: DMU #2: $q = 0$; DMU #3: $q \cong 0.0882$; $l = \left(0, 57/102, 0, 45/102\right)$; Problem 6.3: DMU #2: $q = 12/17$; $l = \left(9/17, 0, 8/17, 0\right)$; DMU #3: $q = 0$. Elimination of the infeasible convex combinations is carried out by the following restrictions: $l_1, l_2 = 0$ or $l_1, l_4 = 0$ or $l_3, l_4 = 0$.

We proceed by formulating and solving the following three new problems corresponding to the three restrictions above:

Problem 6.4

$$\underset{q,l}{Max}\ q$$
$$s.t.\ (1-q)x_j \geq 2l_1 + 6l_2 + 9l_3 + 11l_4$$
$$(1+q)y_j \leq 2l_1 + 3l_2 + 6l_3 + 11l_4$$
$$l_1 + l_2 + l_3 + l_4 = 1$$
$$l_1, l_2, l_3, l_4 \geq 0$$
$$l_1, l_2 = 0$$

Problem 6.5

$$\underset{q,l}{Max}\ q$$
$$s.t.\ (1-q)x_j \ge 2l_1 + 6l_2 + 9l_3 + 11l_4$$
$$(1+q)y_j \le 2l_1 + 3l_2 + 6l_3 + 11l_4$$
$$l_1 + l_2 + l_3 + l_4 = 1$$
$$l_1, l_2, l_3, l_4 \ge 0$$
$$l_1, l_4 = 0$$

Problem 6.6

$$\underset{q,l}{Max}\ q$$
$$s.t.\ (1-q)x_j \ge 2l_1 + 6l_2 + 9l_3 + 11l_4$$
$$(1+q)y_j \le 2l_1 + 3l_2 + 6l_3 + 11l_4$$
$$l_1 + l_2 + l_3 + l_4 = 1$$
$$l_1, l_2, l_3, l_4 \ge 0$$
$$l_3, l_4 = 0$$

The optimal solutions to Problems 6.4 - 6.6 for DMUs #1 - #4 are all less than or equal to zero. Hence, the reference sets of Problems 6.4 – 6.6 characterize three convex monotone hulls that form CCMH as their union. Figure 3 illustrates the so-found convex sets. In that figure, CMH 1 corresponds to Problem 6.4, CMH 2 corresponds to Problem 6.5, and CMH 3 corresponds to Problem 6.6.

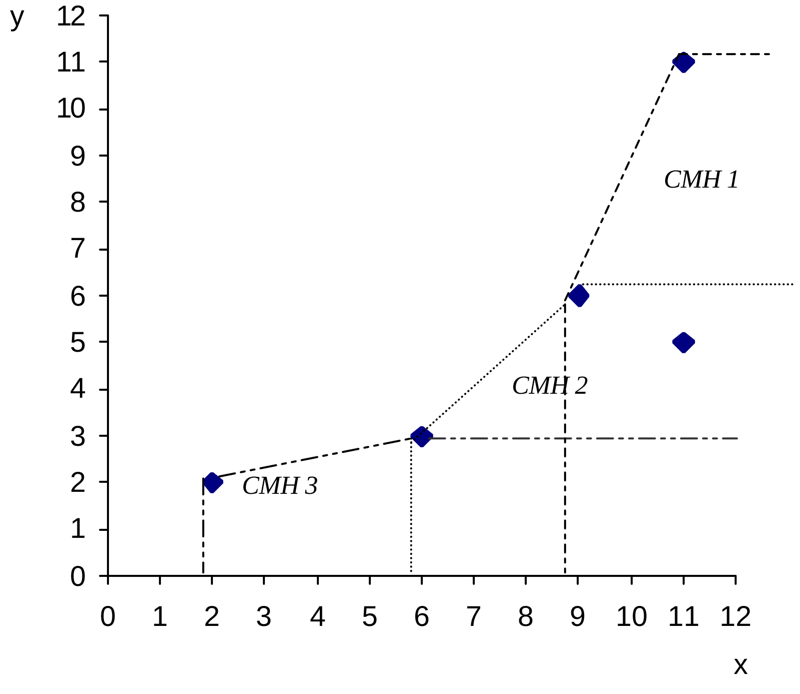


Figure 3: CCMH as the union on three convex subsets

Next, we can compute efficiency measures for inefficient units, say DMU #5, by solving the Debreu-Farrell measures relative to *CMH 1*, *CMH 2*, and *CMH 3* respectively, and consequently selecting the minimum of the feasible optimal solutions. For example, the efficiency score relative to *CMH 1* can be solved as the optimal solution to the LP problem:

Problem 6.7

$$\underset{q,l}{Min}\ q$$

$$s.t.\ qx_j \geq 2l_1 + 6l_2 + 9l_3 + 11l_4$$

$$y_j \leq 2l_1 + 3l_2 + 6l_3 + 11l_4$$

$$l_1 + l_2 + l_3 + l_4 = 1$$

$$l_1, l_2, l_3, l_4 \geq 0$$

$$l_1, l_2 = 0$$

Note that Problem 6.7 differs from Problem 6.4 only in that we have modified the objective function and the first two constraints so as to compute the Debreu-Farrell measure instead of the Briec graph measure. The corresponding DEA problems for computing Debreu-Farrell measures relative to *CMH 2* and *CMH 3* are modified accordingly from Problems 6.5 and 6.6 respectively, and are omitted here as obvious. For DMU #5, the optimal solutions are: *CMH 1*: $q = 9/11$; $l = (0,0,1,0)$; *CMH 2*: $q = 8/11$; $l = \left(0, \frac{1}{3}, \frac{2}{3}, 0\right)$; *CMH 3*: infeasible. The Debreu-Farrell efficiency measure for DMU #5 is then *Min*(9/11, 8/11) = 8/11. Like expected, this measure falls between the corresponding *CMH* (5/11) and *MH* (9/11) scores.

We solved the total of 17 LP problems (Problem 6.1 for 4 DMUs, Problems 6.2 and 6.3 for 2 DMUs, Problems 6.4 - 6.6 for 2 DMUs, and finally 3 LPs akin to Problem 6.7 for a single DMU). This small example gives a somewhat discouraging impression of the computational burden associated with the approach. Note that most of the computational cost was associated with characterizing the sub-technologies, i.e. *CMH*s *1, 2*, and *3*. Once this has been done, computing efficiency scores for inefficient units has the marginal cost of 3 LPs only. Finally, this simple and straightforward relaxation strategy is not necessarily the most efficient algorithm for computing efficiency scores. If it is unnecessary to explicitly characterize the sub-technologies, an iterative branch and bound type of approach could reduce computational cost considerably. However, as noted in the previous section, the development of efficient computation codes is left for future research

7. CONCLUDING REMARKS

We proposed to relax the standard convexity property by invoking additional qualifications or conditions for feasibility of convex combinations. We especially focused on a condition that preserves the spontaneous Koopmans efficiency classification. We abbreviated the efficiency classification preserving conditional convexity as c-convexity. As both monotonicity and convexity imply c-convexity, but the converse does not generally hold, c-convexity can be viewed as the most general property of these three.

Substituting convexity by c-convexity, we constructed empirical DEA production sets as the smallest polyhedrons containing all DMUs, consistent with the imposed production assumptions. These c-convex production sets combine the attractive marginal properties of convex DEA technologies, allowing e.g. the measurement of scale elasticity, with preservation of the efficiency classification thus far only associated with the extremely non-convex *MH* model. Unfortunately, the general theoretical character of the c-convexity property does not fully carry over to empirical DEA approximations. Still, c-convexity provides a more general approximation than standard convexity even in small samples.

The distinct feature of the DEA formulations based on c-convexity is the disjunctive constraint that eliminates the infeasible, efficiency classification violating convex combinations. Consequently, solving the DEA model generally involves Disjunctive Programming. We briefly outlined a relaxation strategy for inferring the optimal solution from optimal solutions to a series of ordinary Linear Programming problems, using the fact that CCMH can be viewed as a union of convex monotone hulls of subsets of DMUs. This approach provides an explicit characterization of the production set, but is often computationally heavy. Development of a more efficient algorithm tailored for DEA provides an interesting challenge for future research.

Another promising avenue for the future research is to try to give the shadow prices of the c-convex frontiers a rigorous economic meaning. Price endogeneity and uncertainty, which have been almost ignored in this literature so far, provide a possible source of motivation for such an investigation.

We illustrated the computational formula by a simple single-input single-output example with 5 DMUs. Although the computational burden associated with this model is considerably higher than in the basic DEA models, the solutions are still obtained by a finite number of Linear Programming problems. We are quite convinced that the computational capacity currently available to researchers suffices to compute the proposed model even in relatively large-scale problems. It is also generally expected that the computational capacity can rapidly improve in the future. Therefore, development of models based on less restrictive maintained hypotheses appear affordable if the additional computational burden constitutes the only cost.

REFERENCES

Afriat, S. (1972): Efficiency Estimation of Production Functions, *International Economic Review* 13, 568-598.

Banker, R.D., A. Charnes, W. W. Cooper (1984): Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis, *Management Science* 30(9), 1078 - 1092

Bauer, P.W. (1990): Recent Developments in the Econometric Estimation of Frontiers, *Journal of Econometrics* 46(1/2), 39-56

Bogetoft, P. (1994): Incentive Efficient Production Frontiers: An Agency perspective on DEA, *Management Science* 40(8), 959-968

Bogetoft, P. (1996): DEA on A Relaxed Convexity Assumption, *Management Science* 42(3), 457-465

Bogetoft, P., and J.L. Hougaard (1999): Efficiency Evaluations Based on Potential (Non-Proportional) Improvements, *Journal of Productivity Analysis* 12, 233-247

Bogetoft, P., J.M. Tama, and J. Tind (forthcoming): Convex Input and Output Projections of Nonconvex Production Possibility Sets, *Management Science*

Briec, W. (1997): A Graph-Type Extension of Farrell Technical Efficiency Measure, *Journal of Productivity Analysis* 8, 95-110

Charnes, A., W. Cooper, and E. Rhodes (1978): Measuring the Efficiency of Decision Making Units, *European Journal of Operational Research* 2(6), 429 - 444

Charnes, A., W.W. Cooper, B. Golany, L. Seiford, and J. Stutz (1985): Foundations of Data Envelopment Analysis for Pareto-Koopmans Efficient Empirical Production Frontiers, *Journal of Econometrics* 30(1-2), 91-107

Cherchye, L., T. Kuosmanen, and G. T. Post (2000): What is the Economic Meaning of FDH? A Reply to Thrall, *Journal of Productivity Analysis* 13(3), 259-263

De Borger, B., G. D. Ferrier, and K. Kerstens (1998): The Choice of a Technical Efficiency Measure on the Free Disposal Hull Reference Technology: A Comparison Using US Banking Data, *European Journal of Operational Research* 105, 427-446

Dekker, D. J. and G. T. Post. (1999): *A Quasi-Concave DEA Model with an Application for Bank Branch Performance Evaluation*, Rotterdam Institute for Business Economic Studies (RIBES), Report no. 9906, under revision for European Journal of Operational Research

Deprins, D., D. Simar, H. Tulkens (1984): Measuring Labor Efficiency in Post Offices, p. 243 - 267 in M. Marchand, P. Pestieau and H. Tulkens, *The Performance of Public Enterprises: Concepts and Measurements*, North Holland

Doyle, R. (1995): Multiattribute Choice for the Lazy Decision Maker: Let the Alternatives Decide!, *Organizational Behavior & Human Decision Processes* 62(1), 87-100

Farrell, M. J. (1957): The Measurement of Productive Efficiency, *Journal of Royal Statistical Society*, Series A, 120(3), 253 - 290

Färe, R. (1988): *Fundamentals of Production Theory*, Springer-Verlag, Berlin

Färe, R., and S. Grosskopf (1995): Non-Parametric Tests of Regularity, Farrell Efficiency, and Goodness of Fit, *Journal of Econometrics* 69, 415-425.

Färe, R., and S. Grosskopf (1996): *Intertemporal Production Frontiers: With Dynamic DEA*, Kluwer Academic Publishers, Boston

Färe, R., S. Grosskopf, and J. Logan (1983): The Relative Efficiency of Illinois Electric Utilities, *Resources and Energy* 5, 349-367

Färe, R., S. Grosskopf, and C.A.K. Lovell (1983): The Structure of Technical Efficiency, *Scandinavian Journal of Economics* 85(2), 181-190

Färe, R., S. Grosskopf, and C.A.K. Lovell (1985): *The Measurement of Efficiency of Production*, Boston, Kluwer

Färe, R., S. Grosskopf, and C.A.K. Lovell (1988): Scale Elasticity and Scale Efficiency, *Journal of Institutional and Theoretical Economics* (JITE) 144, 721-729

Färe, R., S. Grosskopf, and J. Nelson (1990): On Price Efficiency, *International Economic Review* 31(3), 709-720

Färe, R., and C.A.K. Lovell (1978): Measuring Technical Efficiency of Production, *Journal of Economic Theory* 19(1), 150-162

Färe, R., L. Svensson (1980): Congestion of Production Factors, *Econometrica* 48(7), 1745-1753

Gijbels, I., E. Mammen, B.U. Park, and L. Simar (1999): On estimation of monotone and concave frontier functions, *Journal of the American Statistical Association* 94, 220-228

Guesnerie R. (1975): Pareto Optimality in Non-Convex Economies, *Econometrica* 43, 1-29.

Hanoch, G., and M. Rothschild (1972): Testing Assumptions of Production Theory: A Nonparametric Approach, *Journal of Political Economy* 80, 256-275

Kerstens, K., and P. Vanden Eeckaut (1998): *Distinguishing Technical and Scale Efficiency on Non-Convex and Convex technologies: Theoretical Analysis and Empirical Illustrations*, Louvain-la-Neuve, UCL (CORE Discussion Paper 9855)

Kerstens, K., and P. Vanden Eeckaut (1999): Estimating Returns to Scale Using Non-Parametric Deterministic Technologies: A New Method Based on Goodness-Of-Fit, *European Journal of Operational Research* 113(1), 206-214

Koopmans, T.C. (1951): Analysis of Production as an Efficient Combination of Activities, in T.C. Koopmans (Ed.), *Activity Analysis of Production and Allocation*, Wiley, New York.

Kuosmanen, T. (1999): *Data Envelopment Analysis of Non-convex Technology: With an Application to Finnish Super League Pesis Players*, Helsinki School of Economics and Business Administration, Working Papers, W-224

Kuosmanen, T., and G. T. Post (1999a): *Measuring Economic Efficiency with Incomplete Price Information*, Rotterdam Institute for Business Economic Studies (RIBES), Report no. 9934

Kuosmanen, T., and G. T. Post (1999b): *Nonparametric Efficiency Analysis under Uncertainty: A First-Order Stochastic Dominance Approach*, Helsinki School of Economics and Business Administration, Working Papers, W-230

McFadden, D. (1978): Cost, Revenue, and Profit Functions, in M. Fuss and D. McFadden, (Eds.): *Production Economics: A Dual Approach to Theory and Applications*, North-Holland, Amsterdam.

Park, B.U., L. Simar and C. Weiner (forthcoming): FDH efficiency scores from a stochastic point of view, *Econometric Theory*.

Petersen, N. C., (1990): Data Envelopment Analysis on a Relaxed Set of Assumptions, *Management Science* 36(3), 305-314

Post, G.T. (forthcoming): Estimating Non-convex Production Sets Using Transconcave DEA, *European Journal of Operational Research*

Seiford, L.M. (1996): Data Envelopment Analysis: The Evolution of the State of the Art (1978-1995), *Journal of Productivity Analysis* 7, 99-137

Seiford, L.M., and R. Thrall (1990): Recent Developments in DEA: The Mathematical Programming Approach to Frontier Analysis, *Journal of Econometrics* 46(1-2), 7-38.

Schumpeter, J.A. (1966): *History of Economic Analysis*, Oxford University Press, Oxford

Shephard, R.W. (1953): *Cost and Production Functions*, Princeton University Press, Princeton

Sherali, H.D., and C.M. Shetty (1980): *Optimization with Disjunctive Constraints*, Springer-Verlag, Berlin

Simar, L., and P. Wilson (1998): Sensitivity Analysis of Efficiency Scores: How to Bootstrap in Non-parametric Frontier Models, *Management Science* 44(11), 49-61

Tulkens, H. (1993): On FDH Analysis: Some Methodological Issues and Applications to Retail Banking, Courts and Urban Transit, *Journal of Productivity Analysis* 4(1-2), 183-210

Tulkens, H., and P. Vanden Eeckaut (1999): Mesurer l'efficaté: avec ou sans frontières, p. 75-100 in P.Y. Badillo and J.C. Paradi (Eds.), La Méthode DEA: analyse de performances, Hemès science publication, Paris

Varian, H.R. (1984): The Non-parametric Approach to Production Analysis, *Econometrica* 52, 279-297.

ALTERNATIVE TREATMENTS OF CONGESTION IN DEA:

A REJOINDER TO COOPER, GU, AND LI*

LAURENS CHERCHYE
Catholic University of Leuven, Belgium
E-mail: Laurens.Cherchye@econ.kuleuven.ac.be

TIMO KUOSMANEN
Helsinki School of Economics and Business Administration, Finland
E-mail: Kuosmane@hkkk.fi

THIERRY POST
Erasmus University Rotterdam, The Netherlands
E-mail: GTPost@few.eur.nl

A preprint of the article to be published in
European Journal of Operational Research,
Vol. 133, Iss. 1, 2001, pp. 69-74

1. Introduction

The treatment of congestion within the DEA framework has received considerable attention in the recent literature, see e.g. Cooper, Seiford and Zhu (2000a,b) and Färe and Grosskopf (2000). Cooper, Gu, and Li [this issue], claim that the standard procedure to deal with congestion in DEA by Färe, Grosskopf, and Lovell (1985), henceforth FGL, can i) fail to identify congestion when it is present, and ii) identify congestion when it is not present. Cooper *et al.* therefore advocate an alternative Cooper, Thompson, and Thrall (1996) procedure (hereafter CTT), extended by Brockett, Cooper, Shin and Wang (1998), which according to these authors does not suffer from similar problems.

We have two main objections against the reasoning by Cooper *et al.*. First, these authors use the FGL procedure for a purpose -detecting congestion- for which the procedure was not originally intended. In addition, Cooper *et al.* apply FGL to technologies for which it was never intended. Second, even if we adhere to the purpose and the technologies used by Cooper *et al.*, it is logically incorrect to conclude from one or two examples that CTT works better. FGL indeed fails to detect congestion for the two example technologies, and CTT indeed succeeds for these technologies. However, simple counterexamples demonstrate CTT

* This rejoinder was proposed by Professor Jyrki Wallenius to whom we are grateful. We develop and elaborate the argumentation initiated in the two referee reports by the second author.

can fail for the purpose of detecting congestion as well. Before presenting the counterexamples, brief clarifications of the notion of congestion and the related concepts of structural efficiency and free disposability are in order.

2. Definition of Congestion

The key to the debate concerning treatment of congestion in DEA lies in the meaning and definition of congestion. The notion of input congestion originates from Färe and Svensson (1980), who motivated their concept by referring to Turgot's classic 'law of variable proportions', also known as 'law of diminishing returns'. See e.g. Brue (1993) for a historical review of this classic economic concept, and Färe (1980) for an elaborate formal treatment. Cooper *et al.* criticize the FGL model for adhering to this 'law', but they do not present any arguments to back up their negative position. Moreover, it seems that these authors somewhat confuse the intuitive meaning underlying this 'law'. Therefore, some conceptual clarification is in order.

The 'law of diminishing returns' is generally known as a stylized fact, which states that if a single input is increased while other inputs are held constant, the marginal product of the variable input diminishes. In other words, product mix diseconomies are expected to emerge due to unbalanced input proportions. In everyday English language, congestion generally means overcrowding, or more specifically, concentration of some material objects in a small or narrow space. It should be noted that in the classic treatments of the law of diminishing returns (e.g. Turgot, Ricardo, Malthus), the space (i.e. the cultivated land area) is viewed as an essential production factor, which represents the fixed input factor on which a variable input (e.g. labor) congests. For a modern illustration, consider the following stylized example from transportation industry. Suppose we use two inputs, vehicles and road-network, to produce hauling service. Traffic congestion typically occurs when we increase the number of vehicles (input 1) in a fixed road network (input 2). As traffic increases, an additional vehicle produces only a marginal transportation output, like the law of variable proportions suggests. In colloquial language we typically use traffic congestion to describe a situation where the marginal transportation output of an additional car is substantially decreased due to interference with other units of the same input (vehicles). In this sense, the law of diminishing returns and congestion are intimately related, as discussed in Färe and Svensson (1980).

However, we would make a clear distinction between the everyday meaning of congestion, and the congestion property currently debated in the DEA literature. Cooper *et al.* refer to the following definition of congestion adopted from Färe *et al.* (1985), using $L(y) = \{x \mid \text{input } x \in \Re_{+}^{l} \text{ can produce output } y \in \Re_{+}^{m}\}$ for the input correspondence:

Definition 1: The production technology is input congested if for some $y \geq 0$ and $x \in L(y)$, there exists an $x' \geq x$ such that $x' \in L(qy), 0 \leq q < 1, x' \notin L(y)$.

Interestingly, this definition differs from the original definitions given in Färe and Svensson (1980) in three respects. In Definition 1, congestion is a *qualitative* property associated with the production technology, while Färe and Svensson (1980) associated congestion with the input vector that is congested. In Definition 1, congestion can be a property associated with a technology even if all observed DMUs (Decision Making Units) would produce in the congestion-free region of the production set. The second difference is that Färe and Svensson phrased their definition in terms of (single-input) production functions while Definition 1 uses input correspondence. Finally, Färe and Svensson explicitly distinguish between three degrees of congestion, i.e. output-limitational congestion, monotone output-limitational congestion, and output-prohibitive congestion. These more specific terms already suggest that Färe and Svensson, followed by Färe *et al.*, are not talking about congestion in the everyday sense of the word. Rather, they are interested in particularly severe forms of congestion that can limit or completely prohibit production. Note that product mix diseconomies suggested by the law of variable proportions do not yet imply input congestion in the sense of Definition 1. Specifically, this law only suggests that the marginal product of input becomes small, but not necessarily negative (as required in Definition 1).

The relationship between congestion and scale diseconomies is also interesting. The law of variable proportions does not say anything about what should happen in case all outputs are simultaneously augmented, maintaining the original proportions. Return to our stylized transportation example. If both the number of vehicles and the road-network capacity are increased proportionately so that the car-density remains constant, traffic congestion should not emerge. However, it is still possible that marginal transportation output decreases or even becomes negative, but for a different reason such as 'managerial' difficulties due to increased complexity of the road-network. In standard microeconomic reasoning, output losses caused by an equiproportionate scale expansion would not be associated with the law of diminishing returns, congestion, or product mix diseconomies. Rather, such losses would be associated with scale diseconomies.

It is important to note that Definition 1 can associate congestion both with product mix diseconomies and scale diseconomies. Still, it appears that Färe *et al.* are thinking of congestion primarily as a serious form of product mix diseconomies. Note that the assumption of *weak disposability of inputs*[17] which Färe *et al.* (1985) explicitly introduce and maintain throughout their analysis, suffices to eliminate the effects of severe scale diseconomies and hence focuses attention on the product mix diseconomies in the spirit of the classical authors. Therefore, the assumption of weak disposability is an integral part of the congestion analysis by Färe *et al.* (1985), like Färe and Grosskopf (2000) also emphasize. By contrast, Cooper *et al.* seem to have the scale diseconomies in mind since they explicitly criticize the law of variable proportions and the weak disposability assumption. In our opinion, both product mix and scale perspectives are interesting. However, Definition 1 is a potential source of confusion because it doesn't distinguish between the two.

3. Congestion versus Structural Efficiency

Recall that Definition 1 to which Cooper *et al.* refer describes congestion as a qualitative property of technology. This definition does not suggest any *quantitative* measure such as the *"amount of congestion"* or *"congesting amount"* that Cooper *et al.* associate with DMUs. As these concepts are not properly defined, our first objection against the comparison presented by Cooper *et al.* is that the ultimate purpose of the CTT procedure is not clearly stated. From the treatment of the examples, we infer that the purpose is to detect whether or not the technology is congested.

[17] Inputs are weakly disposable iff $x \in L(y)$ => $qx \in L(y)$ for all $q > 1, y \in \Re^m_+$.

Interestingly, detecting occurrence of congestion was not the original purpose of the FGL procedure. Rather, that procedure was intended for measuring *structural efficiency* as a component of overall efficiency. Specifically, structural efficiency (STR) is defined as:

(1) $STR(x,y) = D_L(x,y)/D_{FDHL}(x,y)$.

In this expression, $D_L(x,y) = Inf\{q \mid qx \in L(y)\}$ denotes the Debreu-Farrell input measure -or alternatively interpreted, the inverse of Shephard's input distance function- relative to *L*, and *FDHL* denotes the free disposable hull of *L*, i.e. the smallest set that contains *L*(*y*) and satisfies free (=strong) disposability of inputs.[18]

Färe *et al.* presented this concept of structural efficiency to extend the well-known Farrell (1957) decomposition of overall efficiency (Färe *et al.* (1983, 1985: chapters 3, 8 and 9)). Specifically, they used the following decomposition:

(2) *OE = FTE * STR * SCA * AE,*

where *OE* = Overall Efficiency, *FTE* = Farrell ("pure") Technical Efficiency, *STR* = Structural Efficiency, *SCA* = Scale Efficiency, and *AE* = Allocative Efficiency. See Färe *et al.* (1985, pp. 187-191) for further details.

Färe *et al.* presented DEA models for approximating overall efficiency and its different components. Specifically, the FGL model for approximating structural efficiency is the model that Cooper *et al.* criticize for failure to identify congestion (see Cooper *et al.*, model (3)). However, it is clear from the Färe *et al.* (1985, chapter 3) that the original purpose of the STR measure did not include identifying whether or not congestion occurs in a particular technology. Rather, Färe *et al.* were interested in assessing the extent to which congestion affects a DMU's overall inefficiency.

Cooper *et al.* have apparently confused the notions of congestion and structural efficiency, which explains some of their negative conclusions. The two examples do not provide any evidence that the FGL model would fail to measure what it is proposed to measure, viz. structural efficiency. For example, the pyramid shaped production set in the second example is indeed congested according to the definition given by Färe *et al.* (consider e.g. DMU G = (7.5,7.5) that uses higher amounts of both input than DMU R = (5,5) but cannot produce the output of R, i.e. 10). Still, G is structurally efficient because congestion does not affect its Farrell efficiency (i.e. the radial input contraction).

Of course, the second example does show that FGL can fail for the purpose of detecting congestion. This failure occurs because FGL restricts attention to the observed DMUs and the Farrell efficiency measure (which suffices for measuring structural efficiency). However, as discussed above, FGL was not originally intended for the purpose of detecting congestion. One could equally well argue that CTT fails to measure structural efficiency.

---

[18] Inputs are freely disposable iff $L(y) = L(y) + \Re_+^l$ for all $y \in \Re_+^m$. It is important to distinguish *free* (or strong) disposability from *weak* disposability (see Footnote 16). FGL examines congestion by dropping the strong disposability assumption, but it still maintains weak disposability.

4. Congestion versus Free Disposability

FGL assesses the influence of the free disposability assumption on efficiency. If the assumptions that Färe *et al.* impose on the technology are satisfied, lack of free disposability of inputs is equivalent to input congestion (see Färe *et al.* (1985, pp. 68), and Färe and Grosskopf (1983) for a formal proof). This implies that FGL cannot find structurally inefficient DMUs (detect congestion) if the technology is not congested.

Still, Cooper *et al.* claim that this failure can occur, and use their first example as proof for this claim. In that example, the production set is assumed to take a form of a plateau. Clearly, congestion does not occur simply because output cannot increase or decrease by any feasible input adjustments. Still, FGL does classify DMUs 6 and 7 as structurally inefficient, which seems to wrongly suggest congestion. Hence, the Cooper *et al.* claim seems correct for this technology. However, Cooper *et al.* forgot to mention that a plateau technology does not satisfy the FGL technology assumptions!

Specifically, Färe *et al.* (1985, pp. 23) assume that $L(0) = \Re^l_+$. We might call this the 'idling assumption', since it essentially says that any combination of nonnegative inputs can be held idle. If the technology does not satisfy this assumption, then lack of free input disposability and congestion (in the sense of Definition 1) no longer are equivalent. Specifically, lack of free input disposability becomes a necessary but not sufficient condition for congestion. Hence, FGL may detect lack of disposability for non-congested technologies. This is exactly what happens in the first example, where the plateau technology violates $L(0) = \Re^l_+$.

It should be obvious that any DEA model (including both FGL and CTT) can fail for a technology that does not satisfy the maintained production assumption. Apart from the idling assumption, also other frequently imposed postulates including envelopment of the observed DMUs and convexity provide potential sources of specification error. It therefore seems more meaningful to compare alternative DEA models on the level of the assumptions they impose on the technology. Unfortunately, this is difficult in this particular case because the exact assumptions underlying the CTT model are unclear. Of course, we can discuss about the validity of the idling assumption in specific applications. However, Cooper *et al.* do not discuss validity of the FGL assumptions, but rather ignore them completely.

5. Counterexamples

Even if we adhere to the purpose of detecting congestion, it is not correct to claim that CTT works better than FGL. Cooper *et al.* correctly show that FGL identifies congestion while it is not present in the first example, and that FGL fails to identify congestion while it is present in the second example, even though both examples violate the FGL assumptions. However, it is logically incorrect to conclude from these two examples that CTT is fit for detecting congestion, or that it is better than the FGL procedure. In fact, two simple counterexamples suffice to show that CTT can fail in cases in which the FGL procedure succeeds.

To demonstrate this, consider the first example Cooper *et al.* adopted from Färe *et al.* (1985, pp. 76-77)**.** Let the input correspondence $L(2)$ be the plateau assumed by Cooper *et al.*, but following FGL assume further that $L(0)=\Re_{+}^{l}$. Clearly, the output of DMU 7 decreases from 2 to 0, if input 1 is even slightly increased. Hence, congestion occurs according to Definition 1. The FGL structural efficiency measure correctly identifies that DMU 7 produces at the backward bending (congested) part of the frontier, whereas the slack based CTT measure fails to identify congestion under these assumptions.

In the second example adopted from Brockett *et al.* (1998) we might augment the pyramid shaped production set by further assuming either weak disposability of inputs, or constant returns to scale (i.e. $x \in L(y) \Rightarrow qx \in L(qy)$ for all $q > 0, y \in \Re_{+}^{m}$). Note that the original FGL model imposes both assumptions. Under these alternative assumptions, DMU G does not lie on the congested part of the frontier but rather in the interior of the production set, as the FGL structural efficiency measure correctly suggests. See Fig. 1 in Färe and Grosskopf (2000) for illustration. By contrast, the slack based CTT measure spuriously identifies G congested when this is not the case (congestion still occurs elsewhere but DMU G does not produce on the congested part of the frontier). These remarks should suffice to show that congestion diagnoses essentially depend on what assumptions we are willing to impose on the production technology.

6. Conclusions

The FGL procedure fails to identify congestion in the Cooper *et al.* examples. However, FGL was originally proposed for measuring structural efficiency rather than detecting congestion. In addition, the example technologies do not satisfy the basic assumptions underlying FGL. In this respect, the critique by Cooper *et al.* is somewhat misplaced. Still, it would be interesting to develop a procedure for detecting congestion for a broad class of technologies. Unfortunately, we cannot conclude from the Cooper *et al.* examples that CTT is fit for that purpose, or that it is better than FGL. In fact, simple counterexamples show that CTT can fail in cases in which FGL succeeds. We therefore call for a more systematic treatment of this issue. A first step into that direction would be to formalize the purpose of CTT and its underlying assumptions.

A point that has been ignored so far is the effect of sampling error in two-stage DEA models such as FGL and BTT. In practical applications, the production set is estimated from a finite sample of empirical data. Finite samples generally do not contain sufficient observations to fully represent the production possibilities. It is well known that DEA models which build on more relaxed sets of postulates are generally exposed to greater small sample error. For example, the stage one of the FGL model (see Cooper *et al.*, Formula 1) that imposes weak disposability cannot yield a smaller efficiency score than the stage two of the FGL model (Cooper *et al.*, Formula 3) that builds on free disposability. As such, differences between efficiency scores in the two stages of the FGL model (or slack values in the two stages of CTT) can be due to congestion, but also to the weaker discriminatory power of the stage one model, which arises from the small sample error. Consequently, these DEA models as such cannot provide empirical evidence of occurrence of congestion in a finite sample of DMUs.

In addition, empirical data are almost always contaminated by errors-in-variables. For example, much empirical research uses accounting data that can give a flawed representation of the underlying economic values, e.g. because of debatable valuation and depreciation schemes. In case of sampling error and errors-in-variables, structurally inefficient observations do not generally constitute convincing evidence of occurrence of congestion. Therefore, future research attention should be directed at integrating the measures for structural efficiency and the tests for congestion with the rapidly growing DEA literature on dealing with sampling error (see e.g. Grosskopf (1996) for a survey, and Simar and Wilson (2000) for recent advances on that area) and errors-in-variables (see Post (1999, Section 4.2) for a survey).

References

Brockett, P.L., W.W. Cooper, H.C. Shin, and Y. Wang (1998): Inefficiency and Congestion in Chinese Production Before and After the 1978 Economic Reforms, *Socio-Economic Planning Sciences* 32, 1-20

Brue, S.L. (1993): Retrospectives: The Law of Diminishing Returns, *Journal of Economic Perspectives* 7(3), 185-192

Cooper, W.W., B. Gu, and S. Li (2001): Comparisons and Evaluations of Alternative Approaches to the Treatment of Congestion in DEA, *European Journal of Operational Research*, this issue

Cooper, W.W., L.M. Seiford, and J. Zhu (2000a): A Unified Additive Model Approach for Evaluating Inefficiency and Congestion, *Socio-Economic Planning Sciences* 34, 1-26

Cooper, W.W., L.M. Seiford, and J. Zhu (2000b): Slacks and Congestion: A Response to the Comments of R. Färe and S. Grosskopf, *Socio-Economic Planning Sciences* (to appear)

Cooper, W.W., R.G. Thompson, and R.M. Thrall (1996): Introduction: Extensions and New Developments in DEA, *Annals of Operations Research* 66, 3-45

Farrell, M. J. (1957): The Measurement of Productive Efficiency, *Journal of Royal Statistical Society*, Series A, 253 - 290

Färe, R. (1980): *Laws of Diminishing Returns*, Springer-Verlag, Berlin

Färe, R., and S. Grosskopf (1983): Measuring Congestion in Production, *Zeitschrift für Nationalökonomie* 43, 257-271

Färe, R., and S. Grosskopf (2000): Slacks and Congestion: A Response, *Socio-Economic Planning Sciences* 34, 35-50

Färe, R., S. Grosskopf, and C.A.K. Lovell (1983): The Structure of Technical Efficiency, *Scandinavian Journal of Economics* 85(2), 181-190

Färe, R., S. Grosskopf, and C.A.K. Lovell (1985): *The Measurement of Efficiency of Production*, Kluwer-Nijhoff Publishing, Boston

Grosskopf (1996): Statistical Inference and Nonparametric Efficiency: A Selective Survey, *Journal of Productivity Analysis 7*, 161-176.

Post, G.T. (1999): *Finding the Frontier: Methodological Advances in DEA*, Tinbergen Institute Research Series 211, Rotterdam, The Netherlands.

Simar, L., and P. Wilson (2000): Statistical Inference in Nonparametric Frontier Models: The State of the Art, *Journal of Productivity Analysis* 13(1), 49-78

WHAT IS THE ECONOMIC MEANING OF FDH?

A REPLY TO THRALL*

LAURENS CHERCHYE
Catholic University of Leuven, Belgium
E-mail: Laurens.Cherchye@econ.kuleuven.ac.be

TIMO KUOSMANEN
Helsinki School of Economics and Business Administration, Finland
E-mail: Kuosmane@hkkk.fi

THIERRY POST
Erasmus University Rotterdam, The Netherlands
E-mail: GTPost@few.eur.nl

A preprint of the article originally published in
Journal of Productivity Analysis,
Vol. 13, Iss. 3, 2000, pp. 259-263

1. Introduction

In a recent issue of the *Journal of Productivity Analysis*, Thrall (1999) called for abandoning the Free Disposable Hull (FDH, Deprins et al., 1984) approximation of production possibilities as economically meaningless in comparison to the Convex Monotone Hull (CMH; Banker et al., 1984) approximation. This strong conclusion was solely based on Thrall's Principal Theorem, which essentially demonstrates that FDH can give a technically efficient classification to output-input vectors that are inefficient in terms of profit maximisation, i.e. at all non-negative price vectors there exists an alternative output-input vector that yields higher profit. In this short communication, we argue that the economic meaning of the competing empirical production sets cannot be inferred from this theorem. Specifically, we demonstrate that both empirical production sets are economically equally meaningful under the economic conditions that underlie Thrall's theorem. In addition, we demonstrate that FDH can be economically more meaningful than CMH under non-trivial alternative economic conditions.

* We thank C.A. Knox Lovell for helpful comments.

2. Technical efficiency versus profit efficiency

FDH relies on the sole assumption that production possibilities satisfy free disposability. The key difference between FDH and CMH is that CMH imposes the additional assumption that production possibilities are convex[1]. Free disposability is a generally accepted assumption in production economics. Färe and Grosskopf (1983) show that imposing free disposability in fact implies a congestion-adjusted approximation. Note that this congestion adjustment is harmless as far as the usual (Pareto-Koopmans or Debreu-Farrell) notions of technical efficiency are concerned. By contrast, to the best of our knowledge, there appear to be no valid theoretical arguments for assuming a priori that production possibilities are truly convex (see also McFadden, 1978, pp. 8-10). In addition, various empirical studies of non-trivial industries suggest violations of the convexity hypothesis (see e.g. Hasenkamp, 1976; Kuosmanen, 1999; Dekker and Post, 1999). Therefore, FDH seems to have a comparative advantage for analysing technical efficiency.

Ironically, Thrall's negative conclusion towards FDH originates from confusing technical and economic efficiency criteria. Thrall considered the evaluation of efficiency in terms of profit maximisation at given prices, but he used the Pareto-Koopmans notion of technical efficiency as the only efficiency criterion. Theoretically, technical efficiency is a *necessary* but generally not a *sufficient* condition for overall economic efficiency (Farrell, 1957). Therefore, FDH technical efficiency cannot guarantee profit maximisation, as Thrall correctly points out, but it should be added that CMH technical efficiency does not guarantee it either.

When an appropriate efficiency criterion that complies with the assumed profit maximisation objective is adopted (i.e. *profit efficiency*, as originally proposed by Nerlove, 1965)[2], both approximations appear equally meaningful. The Nerlovian profit efficiency criterion compares the obtained profit level to the highest profit level attainable given technology and output and input prices: an output-input combination is labelled efficient if, for given prices, no higher profit level is shown to be attainable. A standard result in production analysis is that production sets can be 'monotonised' and 'convexified' without harm for the purpose of measuring profit efficiency, because monotonisation and convexification do not interfere with the maximum profit level (see e.g. Varian, 1984). In fact, from production economics we know that cost, revenue and profit functions associated with a particular technology equal those corresponding to its appropriately convexified counterpart, a point which is nicely illustrated by Diewert (1982; p. 538-540). We conclude that maximum profit levels associated with the FDH and CMH approximations are identical. Therefore, we do not see any reason to discriminate between the two approximations as regards economic efficiency.

3. Endogenous prices and price uncertainty

The equivalence result mentioned at the end of the previous section only applies to Nerlovian profit efficiency in Farrell's framework, which is economically relevant if producers take exogenously fixed prices as given. That is a valid assumption under economic conditions of perfect competition or price rationing under perfect certainty. In real-life industries, however, these conditions are mostly not satisfied even by approximation[3]. In many cases, prices and quantities of inputs and outputs are mutually dependent and the price-taking assumption is no longer valid. Also, decision-makers frequently face uncertainty about ex post prices when ex ante allocating resources and producing output (McCall, 1967;

Sandmo, 1971). In these instances, the economic justification of convexification breaks down, and FDH generally becomes economically more meaningful than CMH.

First, if input and output prices vary with quantities demanded or supplied, the evaluated firm can in general not be compared with a convex combination of firms with a different output-input structure, which are confronted with different output and input prices. Indeed, firms of which convex combinations dominate the evaluated firm may actually be associated with strictly *lower* profit levels. This point can be illustrated using the example considered by Thrall (1999). That example involved three firms producing two outputs, y_1 and y_2, from a single input x. The observed vectors (y_1, y_2, x) are (1, 10, 1), (2,2,1) and (10,1, 1). The output-input vector (2,2,1) is FDH efficient, as it is not dominated by any of the two other observations, but CMH inefficient, because convex combinations of (1, 10, 1) and (10,1, 1) do dominate it. However, this does not necessarily imply that (2,2,1) is not profit-maximising. Consider, for example, the following linear inverse demand functions for the two outputs (with p_i the price of y_i (i=1,2)): $p_1=10 - y_1 + ½ y_2$; $p_2=10 - y_2 + ½ y_1$. The output bundle (2,2) is associated with revenue of 36, while output bundles (1,10) and (10,1) both yield revenue of 19 only. Since all three firms use the same input amount, the profit associated with (2,2,1) is necessarily higher than that corresponding to (1,10,1) and (10,1,1). Hence, the output-input vector (2,2,1) cannot be demonstrated to be economically inefficient. While the argument for convexification breaks down if prices are endogenous, free disposability can be maintained without harm in the economically meaningful region where marginal costs of inputs and marginal revenues of outputs are nonnegative.

Second, convexification requires that producers have perfect information about input and output prices when fixing their production plans. When prices are uncertain, however, it can be unfair and misleading to measure economic efficiency *at the ex post prices*. Rather, profit realisations at all possible ex post price scenarios have to be taken into account. FDH is economically appealing under uncertainty due to the fact that for every interior point there exists an actually observed point that yields higher profit at *all* possible (non-negative) price vectors. This has a natural interpretation in terms of *first-order stochastic dominance*, a well-known decision criterion for choice under uncertainty. In particular, under FDH the profit distribution of an interior point is always stochastically dominated by that of an actually observed point (see Kuosmanen and Post, 1999, for a more elaborate discussion). This needs not be the case with CMH. Although for each interior point there exists at *every single* price vector an observed DMU that yields higher profit, CMH reference units need not yield higher profits at *all* price vectors. Hence, CMH does not have the same stochastic dominance interpretation as FDH.

4. The interpretation of substitution rates

A final point concerns Thrall's claim that CMH efficiency analysis provides the researcher with additional information on the terms of substitution. We do not believe this provides a valid argument in favour of CMH and against FDH. Firstly, one has to distinguish between the (supply-side) technical substitution properties on the one hand and the (demand-side) preference-related substitution effects on the other. In contrast to Thrall's claims, CMH can only approximate the former type of substitution rates, and only so for a truly convex production set. Secondly, although FDH does not directly provide estimates of the technical substitution properties, it may be complemented by various nonparametric or parametric approaches (see Kuosmanen (1999) for a detailed discussion and references). For example, Thiry and Tulkens (1992) used a two-stage technique that first filters out FDH inefficient DMUs and subsequently estimates the frontier using parametric regression techniques.

5. Conclusions

We pointed out that FDH remains an economically meaningful empirical production set. Specifically, we showed that Thrall's main argument against FDH, concerning the potential disparity between technical and profit efficiency classifications, builds on an inappropriate efficiency criterion. We subsequently demonstrated that FDH and CMH approximations actually lead to exactly the same profit efficiency results when a suitable criterion is applied. In addition, we discussed non-trivial economic conditions (imperfect competition and price uncertainty) under which FDH may even become economically more meaningful than CMH. We want to underline that there are no a priori economic grounds to believe that the true production set is convex. Therefore, FDH has a comparative advantage over CMH as regards technical efficiency analysis and the related decomposition of economic efficiency into allocative and technical components.

Nevertheless, some non-economic arguments in favour of CMH can be acknowledged. Firstly, simple linear programming techniques can compute e.g. Farrell (1957) technical efficiency measures relative to CMH (see Banker et al., 1984). However, we would add that FDH Farrell estimates are equally simply obtained from integer programming, which is no problem to modern-day solvers, or from enumeration, which does not require mathematical programming at all (see Tulkens, 1993). Second, if the technology would be truly convex - by some strike of luck - the CMH approximation may yield better technical efficiency estimates than the FDH approximation. This, however, would in principle require some empirical verification of the convexity property, which to the best of our knowledge is not easily obtained. Moreover, it is possible to reduce the small sample error associated with FDH using information from the asymptotic distributions of efficiency estimates (Park et al., 1997), or alternatively from simulated empirical distributions generated with bootstrapping techniques (e.g. Simar and Wilson, 1998). Including additional local production information (see e.g. Thannassoulis and Allen, 1998) could also help to remedy this problem. Finally, if the production set would indeed happen to be convex, possible advantages of CMH over FDH generally vanish for large samples.

Notes

1. Free disposability and convexity are assumptions about the true production possibilities. Both FDH and BCC impose the additional (determinism) assumption that observed input-output vectors are a sample from the true production possibilities.
2. See also Banker and Maindiratta (1988) and Chambers et al. (1998).
3. Nerlove (1965, pp. 90) already criticised Farrell's (1957) treatment, which essentially remains embedded in modern DEA, for the fact that it is not generally applicable under imperfect competition.

References

Banker, R. D., A. Charnes, and W. W. Cooper. (1984). “Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis.” *Management Science* 30, 1078 - 1092

Banker, R.D. and A. Maindiratta. (1988). “Nonparametric Analysis of Technical and Allocative Efficiencies in Production.” *Econometrica* 56, 1315-1332.

Chambers, R.G., Y. Chung and R. Färe. (1998). “Profit, Efficiency and Nermovian Efficiency.” *Jounal of Optimization Theory and Applications* 98, 351-364.

Dekker, D.J. and G.T. Post. (1999). "A Quasi-Concave DEA Model with an Application for Bank Branch Performance Evaluation.” Rotterdam Institute for Business Economics Studies (RIBES) Working Paper 9906, ISBN 90-5086-319-1, under revision for *European Journal of Operational Research*

Deprins, D., L. Simar and H. Tulkens. (1984). “Measuring Labor Efficiency in Post Offices.” In M. Marchand, P. Pestieau and H. Tulkens (eds.), *The Performance of Public Enterprises: Concepts and Measurements.* Amsterdam: North Holland.

Diewert, E. (1978). “Duality Approaches to Microeconomic Theory.” In K.J. Arrow and M.D. Intriligator (eds.), *Handbook of Mathemetical Economics Volume 2*. Amsterdam: North-Holland.

Farrell, M. J. (1957). “The Measurement of Productive Efficiency.” *Journal of the Royal Statistical Society* Series A 120, 253 - 281

Färe, R. and S. Grosskopf. (1983). “Measuring Congestion in Production.” *Zeischrift fur Nationalökonomie* 43, 257-271.

Hasenkamp, G. (1976). “A Study of Multiple-Output Production Functions.” *Journal of Econometrics* 4, 253-262.

Kuosmanen, T. (1999). “Data Envelopment Analysis of Non-Convex Technology: With an Application to Finnish Super League Pesis Players.” Helsinki School of Economics and Business Administration Working Papers W-224, under revision for *European Journal of Operational Research.*

Kuosmanen, T. and G. T. Post. (1999). “Nonparametric Efficiency Analysis under Uncertainty: A First-Order Stochastic Dominance Approach.” Helsinki School of Economics and Business Administration Working Paper W-230, under revision for *Journal of Productivity Analysis*.

McCall, J. J. (1967): "Competitive Production for Constant Risk Utility Functions", *Review of Economic Studies* 34, 417-420

McFadden, D. (1978). “Cost, Revenue and Profit Functions.” In M. Fuss and D. McFadden (eds.), *Production Economics: A Dual Approach to Theory and Applications*. Amsterdam: North-Holland.

Nerlove, M. (1965). "*Estimation and Identification of Cobb-Douglas Production Functions."* Chicago: Rand McNally.

Park, B., L. Simar, and C. Weiner. (1997). “FDH Efficiency Scores from a Stochastic Point of View.” *Econometric Theory*, forthcoming.

Sandmo, A. (1971). "On the Theory of the Competitive Firm under Price Uncertainty." *American Economic Review* 61, 65-73

Simar, L. and P. Wilson (1998). “Sensitivity Analysis of Efficiency Scores: How to Bootstrap in Non-parametric Frontier Models.” *Management Science* 44, 49-61

Thanassoulis, E. and R. Allen. (1998). “Simulating Weight Restrictions in Data Envelopment Analysis.” *Management Science* 44, 586-594.

Thrall, R.M. (1999). “What is the Economic Meaning of FDH?” *Journal of Productivity Analysis* 11, 243-250.

Thiry, B. and H. Tulkens. (1992). “Allowing for technical Inefficiency in Parametric estimation of Production Functions for Urban Transit Firms.” *Journal of Productivity Analysis* 3, 45-65.

Tulkens, H. (1993). “On FDH Analysis: Some Methodological Issues and Applications to Retail Banking, Courts and Urban Transit.” *Journal of Productivity Analysis* 4,183-210.

Varian, H. (1984). “The Nonparametric Approach to Production Analysis.” *Econometrica* 50, 945-974.

MEASURING ECONOMIC EFFICIENCY WITH INCOMPLETE PRICE INFORMATION

WITH AN APPLICATION TO EUROPEAN COMMERCIAL BANKS

TIMO KUOSMANEN
Helsinki School of Economics and Business Administration, Finland
E-mail: Kuosmane@hkkk.fi

THIERRY POST
Erasmus University Rotterdam, The Netherlands
E-mail: GTPost@few.eur.nl

A preprint of the article to be published in
European Journal of Operational Research[*]

Abstract

Measuring economic efficiency requires complete price information, while resorting to technical efficiency exclusively does not allow one to utilise any price information. In most studies, at least some information on the prices is available from theory or practical knowledge of the industry under evaluation. In this paper we extend the theory of efficiency measurement to accommodate incomplete price information by deriving upper and lower bounds for Farrell's overall economic efficiency. The bounds typically give a better approximation for economic efficiency than technical efficiency measures that use no price data whatsoever. From an operational point of view, we derive new Data Envelopment Analysis (DEA) models for computing these bounds using standard linear programming. The practical application of these estimators is illustrated with an empirical application to large European Union commercial banks.

Key words: Economic efficiency measurement, imperfect price information, Data Envelopment Analysis (DEA), weight restricted models, Free Disposable Hull (FDH)

1. Introduction

In the literature of productivity and efficiency analysis, much discussion has focused on economic efficiency measures, as well as on the economic justification of technical efficiency measures. Following Farrell’s (1957) seminal paper, economic efficiency can be decomposed into two components: allocative efficiency and technical efficiency. The technical component requires quantitative volume data of inputs and outputs only, while associated prices or cost shares are also necessary for measuring allocative efficiency. Already Debreu (1951) and Farrell (1957) expressed their concern about the ability to measure prices accurately enough to make good use of economic efficiency measurement. For example, accounting data can give a poor approximation for economic prices (i.e. marginal opportunity costs), because of debatable valuation and depreciation schemes. Several authors, including Charnes and Cooper (1985) cite this concern as a motivation for emphasizing technical efficiency measurement. Consequently, many studies in the more application oriented side of Operations Research, including the seminal articles on Data Envelopment Analysis (DEA) by Charnes, Cooper and Rhodes (1978), Banker, Charnes and Cooper (1984), and Charnes et al. (1985), assess efficiency solely in terms of technical efficiency.

Interestingly, the radial (i.e. Debreu-Farrell) technical efficiency measure provides a theoretical upper bound for economic efficiency, as Debreu (1951) and Shephard (1953) already noted. However, this measure need not be a particularly good approximation for economic efficiency, as it does not utilize any price information whatsoever. In numerous empirical studies, at least some rough information on the economic prices is available from theory or practical knowledge of the industry under evaluation. One source of price information is prior knowledge on the quality or risk of the different inputs and outputs. For example, primary inputs are typically more expensive than secondary inputs. Therefore, the unit of labour input of key personnel (e.g. detectives, dentists, university professors, surgeons, teachers) is more expensive than that of assisting staff (e.g. secretaries, research assistants, cleaners, janitors). As capital inputs are concerned, the unit cost of equity capital exceeds that of debt, because equity involves more risk for the capital suppliers than debt does. Consequently, there exist both need and opportunities to include incomplete price information in efficiency analysis, so as to improve approximation of economic efficiency concepts.

In this paper, we extend the classic Farrell's theoretical framework to include incomplete price information. Assuming that a convex polyhedral cone can represent the price domain, we derive both upper and lower bounds for economic efficiency. The bounds typically give a better approximation for true economic efficiency than technical efficiency measures that use no price data whatsoever. For sake of generality, we only assume production technologies are closed. In particular, we do not impose convexity or monotonicity of the production set. Therefore, the general theory developed in this paper is not limited to some restrictive classes of technologies, or particular approaches of approximating them, but are ready to put in practice both in parametric and nonparametric streams of efficiency analysis.

In addition to the theoretical discussion, we show how standard linear programming techniques can obtain DEA estimators for these bounds from empirical data. These estimators have some technical similarity to the well-known 'weight-restricted' DEA models, as initiated by Thompson et al. (1986).[19] Those models enrich original DEA models with managerial judgement by restricting the dual multipliers to convex polyhedral cones. The dual multipliers represent the marginal properties (shadow prices) of the production set, and restricting them can help to obtain a better approximation for the production set. Therefore, weight restrictions are a useful approach for including production information.

[19] For a detailed survey of the weight-restricted DEA models, see Allen et al. (1997).

The inclusion of price information for the purpose of economic efficiency estimation (i.e the subject of this study) is also frequently cited as a rationale for using weight-restricted models (e.g. Thompson et al., 1990). However, there are at least three important objections against using weight restrictions for including price information. First, it conceals that different economic efficiency measures require different kinds of price information. For example, information on input prices can improve the estimation of cost efficiency. By contrast, output prices are not required, and can even seriously distort the analysis. Still, the weight-restricted models give the impression that all price information is desirable. Second, by restricting the dual multipliers to price cones, the model looses flexibility to obtain information on the marginal properties of the production set. Third, the use of weight restrictions is restricted to DEA models that give a locally linear approximation for the production set, and hence assume that the production set is convex. However, there are few valid theoretical or empirical arguments in favour of convexity, and, in many cases, it is desirable to relax convexity, e.g. to allow for economies of scale and of specialisation (see e.g. Kuosmanen (2001) and Cherchye, Kuosmanen, and Post (2000)). For these reasons, we think weight-restricted DEA models have limited use for measuring economic efficiency.

To remedy the above problems, the empirical estimators derived below explicitly distinguish between the marginal properties of the production set on one hand, and the economic prices on the other. In addition, we derive both upper and lower bounds for economic efficiency, whereas weight-restricted DEA models usually consider an upper bound only. Finally, our estimators build on the assumption of free disposability of inputs and outputs only (like the Free Disposable Hull (FDH) model by Deprins, Simar, and Tulkens (1984)), which makes them consistent estimators for both convex and non-convex technologies.

The remainder of this paper is organised as follows. In Section 2 we present the Farrell decomposition along with some definitions of various efficiency measures. Section 3 discusses how incomplete price information can be used for deriving an interval for economic efficiency. Section 4 discusses standard mathematical programming techniques can obtain nonparametric estimators for these bounds from empirical data. Section 5 illustrates the approach with an empirical application for large European Union commercial banks, and show that even a single restriction on prices can substantially improve the economic efficiency estimation. Section 6 gives our conclusive remarks.

2. Farrell decomposition

The Farrell decomposition is a fundamental cornerstone of the theory of efficiency measurement. Farrell (1957) explicitly decomposed overall *economic efficiency* into components of *technical efficiency* and *allocative efficiency*.[20] Farrell's general framework applies to a number of different ways to define and measure economic efficiency. In efficiency analysis, the organisations under study differ both in terms of their objectives and their environment (compare e.g. business enterprises versus non-for-profit organisations). Therefore, alternative measures are needed for a meaningful assessment of economic efficiency for different organisations. Following Farrell, we focus on cost minimising behaviour of producers. In this perspective, the appropriate measure for economic efficiency is *cost efficiency*, associated with the *input oriented* technical efficiency measure (i.e. output

[20] In fact, Farrell originally used the term "price efficiency" instead of allocative efficiency, but we agree with Kopp (1981) who pointed out that the assumed price-taking nature of production units that is better captured by the term allocative efficiency.

is held constant). However, the proposed approach is directly applicable to analogous *revenue efficiency* and *profit efficiency* (Nerlove, 1965) measures, and the associated output oriented and non-oriented technical efficiency measures respectively (see e.g. Chambers et al. (1996, 1998); and Briec and Lesourd (1999)).

It is first necessary to introduce some notation. Throughout the paper, production inputs are represented by the input vector $x = (x_1 \ \cdots \ x_q) \in \Re_+^q$, whereas outputs are denoted by $y = (y_1 \ \cdots \ y_q) \in \Re_+^p$. Input prices in terms of marginal opportunity cost are represented by the price vector $w = (w_1 \ \cdots \ w_q)^T \in \Re_+^q$. The empirical production data of the production units are represented by the output matrix $Y = (y_1 ... y_n)^T$, with $y_j = (y_{1j} \cdots y_{pj})$, and the input matrix $X = (x_1 ... x_n)^T$, with $x_j = (x_{1j} \cdots x_{qj})$. In addition, we use the index set $S = \{1,...,n\}$.

There are many alternative ways to characterise the production technology (see e.g. Färe (1988) for an elaborate discussion). The most general representation is the production possibilities set *T*:

(2.1) $T := \{(x, y) \in \Re_+^{q+p} \,|\, x \text{ can produce } y\}$

For the purposes of Sections 2 and 3, however, a more convenient representation of the technology is the *input correspondence*

(2.2) $L(y) := \{x \,|\, (x, y) \in T\},\ y \in \Re_+^p,$

which maps outputs *y* into subsets *L*(*y*) of inputs. In other words, the input set *L*(*y*) denotes all input vectors *x* that yield output *y*. For sake of generality, in Sections 2 and 3 we only postulate *T* and *L* to be nonempty and closed. Specifically, we do not impose any convexity or monotonicity assumptions whatsoever. We think this is an important point of deviation from the earlier theoretical developments, which are heavily built on these two simplifying assumptions. We stress that there are no valid theoretical or empirical arguments for claiming *T* or *L* to be convex *a priori* (see e.g. Farrell (1959); and more recent papers Cherchye et al. (2000) and Kuosmanen (2001) on this point). In addition, as demonstrated below, convexity assumptions can be unnecessary for employing duality theory of optimization, and for deriving tractable model formulations.

Cost efficiency is conventionally defined as the ratio of the minimum cost of producing the output of the evaluated production unit to the actual cost at the given input prices and technology.[21] Using the standard notion of *cost function*, i.e.

(2.3) $C^L(y; w) := \underset{x}{Min}\{xw \,|\, x \in L(y)\}$

cost efficiency can be formally defined as

[21] It is worth to note that since we assume away any stochastic disturbances, we measure efficiency of technically feasible production vectors $(x,y) \in T$ only, so we use *minimum* and *maximum* rather than *infimum* and *supremum*.

(2.4) $CE^{L}(x,y;w):=\dfrac{C^{L}(y;w)}{xw}$.

In the subsequent section, we shall utilise an equivalent formulation of cost efficiency that views cost efficiency as the cost function at normalised prices, i.e.

(2.5) $CE^{L}(x,y;w)=\underset{x'\in L(y)}{Min}\left\{x'w' \middle| w'=aw, a\in\Re_{+}; xw'=1\right\}$.

In this measure, prices are normalised such that the normalised cost of the evaluated production vector equals unity, i.e. $xw'=1$. It is worth to note at least the following properties of this measure: Firstly, cost efficiency is homogeneous of degree one in input prices. Furthermore, the measure is restricted by construction to the closed interval [0,1], the value of unity representing total efficiency.

Number of alternative measures for *technical efficiency* have been proposed in the literature (see De Borger et al., 1998, for an up-to-date presentation). Nevertheless, it turns out that the only measure with a tractable dual formulation is the radial measure proposed by Debreu (1951) (see also Shephard (1953)), and adopted by Farrell. The *Debreu-Farrell input efficiency measure* defines inefficiency as the maximum equiproportionate reduction of inputs that is attainable without reducing any of the outputs, formally

(2.6) $DF^{L}(x,y):=\underset{q}{Min}\left\{q \middle| qx\in L(y)\right\}$.

Analogous output and graph measures are also available (see e.g. Farrell, 1957; and Chambers et al., 1998)

Allocative efficiency can be seen as the cost efficiency measure (or the overall economic efficiency in general) applied to the technically efficient reference production plan, i.e. $(DF^{L}(x,y)x,y)$. In case of cost savings, allocative efficiency is formally defined as

(2.7) $AE^{L}(x,y;w):=\dfrac{C^{L}(y;w)}{DF^{L}(x,y)xw}$.

Although the above definition associates allocative efficiency intimately with cost efficiency, it is worth to note that allocative efficiency can be equivalently defined without any such reference by using the radial Debreu-Farrell input gauge and the minimum isocost hyperplane $H^{L}(y;w)=\left\{x' \middle| x'w=C^{L}(y;w)\right\}$, i.e.

(2.8) $AE^{L}(x,y;w)=DF^{H}(DF^{L}(x,y)x,y)$.

It is immediate from (2.6) that the allocative efficiency measure has by definition the homogeneity property in input prices. By construction, allocative efficiency is restricted to the interval [0,1]. Moreover, it is obvious from (2.6) that the product of allocative efficiency and technical efficiency always equals cost efficiency, i.e.

(2.9) $CE^{L}(x,y;w)=AE^{L}(x,y;w)\times DF^{L}(x,y)\quad\forall x\in L(y), w\in\Re_{+}^{q}$.

Equation (2.9) is the fundamental Farrell decomposition. Among other things, this decomposition implies that technical efficiency is a necessary, but not sufficient condition for economic efficiency. This result is also known as the *Mahler inequality* (see e.g. Färe and Primont, 1995), i.e.

(2.10) $CE^L(x,y;w) \leq DF^L(x,y) \quad \forall x \in L(y), w \in \Re^q_+$.

Figure 1 illustrates the Farrell decomposition graphically with two inputs, at a fixed level of output *y*. The production unit *k* lies in the interior of the (non-convex) input set *L* so it is inefficient both in technical and economic sense. The Debreu-Farrell technical efficiency measure is the ratio OR_{DF}/Ok. Allocative efficiency is the ratio OR'_{CE}/OR_{DF}. Finally, cost efficiency is OR'_{CE}/Ok, which we obtain by multiplying the technical efficiency and allocative efficiency indices.

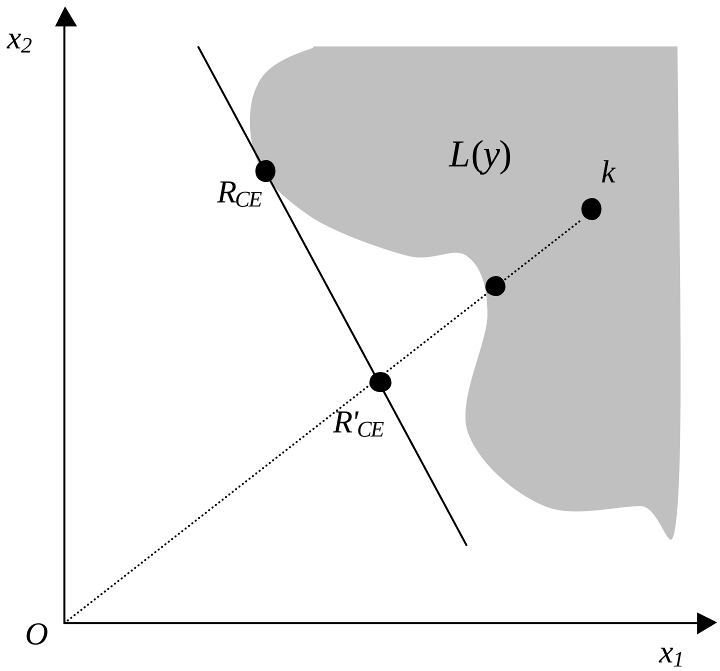


Figure 1: Farrell decomposition of cost efficiency (OR'_{CE}/Ok) as a product of technical efficiency (OR_{DF}/Ok) and allocative efficiency (OR'_{CE}/OR_{DF}) components.

3. Farrell framework under incomplete price information

The economic and technical efficiency measures are extreme cases with respect to the requirement and the utilisation of price information. The economic efficiency measures require complete and accurate information. On the other hand, no prices appear in the definitions of technical efficiency measures, cf. e.g. (2.6). However, we believe many research situations are best described by the intermediate case where some incomplete price information is available. In that case, the economic approach can not be used, because the price information is incomplete. However, resorting to the technical approach leaves potentially valuable price information unutilised. Below, we demonstrate how incomplete price information can be accommodated into the Farrell framework to obtain better approximations for economic efficiency than available from technical efficiency measures.

In this paper we relax the information requirement of prices by only assuming that the factor prices belong to domain $W \subseteq \Re_{+}^{q}$. For simplicity, we use a polyhedral convex cone to represent the price domain. More specifically, we assume the following structure for the price domain:

(3.1) $$W = \left\{ w \in \Re_{+}^{q} \middle| Aw \geq 0 \right\}.$$

This cone represents the price domain in terms of l linear inequalities. A is a $l \times q$ matrix composed of l row vectors $A_1,\ldots,A_l$. Denote the index set $L=\{1,\ldots,l\}$. Heuristically, each row vector A_j, $j \in L$ can be considered as an input vector. Interpreted in this way, price domain W is a collection of all nonnegative prices vectors at which the cost of each input vector A_j, $j \in L$ is greater than or equal to zero. As the limiting special cases, we have $W = \Re_{+}^{q}$ if A is void. By contrast, the price cone reduces to a specific price vector unique up to a multiplication by any positive scalar if $Aw \geq 0$ can be expressed in terms of q linearly independent equality restrictions $Bw = 0$, where B is a $q \times q$ matrix of full rank.

The price domain cone W involves two sets that are interesting from an economic perspective, and that will prove helpful for approximating economic efficiency. The first set is the negative polar cone of W, which has the economic interpretation of the set of input vectors associated with a nonnegative cost at *all* prices in W:

(3.2) $$W^{U} = \left\{ x \in \Re^{q} \middle| xw \geq 0 \ \ \forall w \in W \right\} = \left\{ x \in \Re^{q} \middle| x \geq bA; b \in \Re_{+}^{l} \right\} .$$

The second set contains all input vectors associated with a nonnegative cost at *some* prices in W:

(3.3) $$W^{V} = \left\{ x \in \Re^{q} \middle| \exists w \in W : xw \geq 0 \right\} = \left\{ x \in \Re^{q} \middle| x \geq aA_{i}; i \in L; a \geq 0 \right\} .$$

It should be noted that the assumed form of the price domain W could turn out impractical or irrelevant in some empirical situations. However, in many cases, price information does take the form of linear inequalities, or linear inequalities can give a good approximation for more complicated price structures. For example, the information that equity capital is more expensive than debt capital can be represented by a single linear inequality. In addition, the limiting cases of exact and complete price information on one hand, and no price information whatsoever on the other, are special cases of this more general price domain, as shown in more detail below. Finally, polyhedral convex cones are computationally convenient, because they can be included in Linear Programming models.

Figure 2 illustrates a graphical example of a price domain as represented by the convex cone W. This price domain is obtained by specifying $A = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}$, which effectively restricts the relative price $W1/W2$ of the two inputs to the closed interval [0.5, 2].

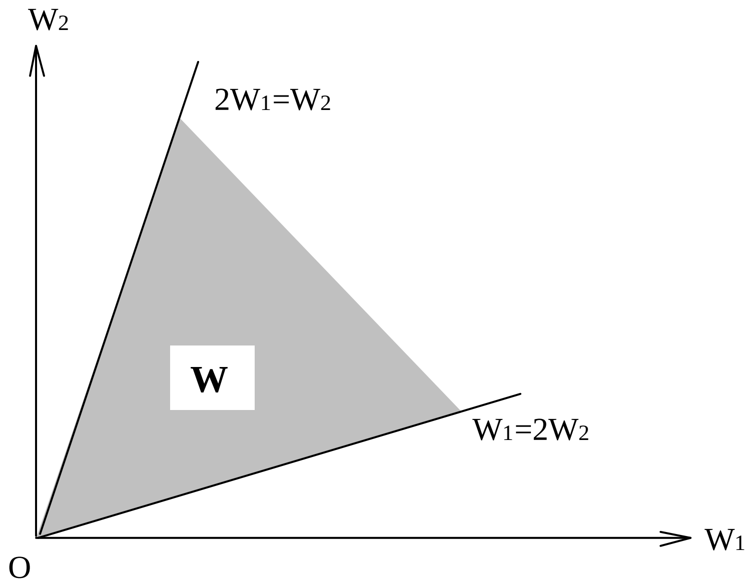


Figure 2: Example of the price domain as represented by the convex cone W

To obtain an upper bound for cost efficiency, we propose to use the *maximum* value of cost efficiency over the price domain, i.e.:

(3.4) $$\overline{CE}^{L}(x,y;W)=\max_{w\in W} CE^{L}(x,y;w)=\max_{w\in W}\left(\min_{x'\in L(y)}\left\{x'w \mid xw=1\right\}\right).$$

We will refer to this maximum value as the *upper bound* of cost efficiency, using analogy to nonparametric production analysis (e.g. Varian, 1984), especially Banker and Maindiratta (1988) who derived upper and lower bounds for profit efficiency using their subset-rationalization condition.

Employing duality theory of optimization, we can relate this upper bound to the Debreu-Farrell measure (2.6).

THEOREM 3.1: For all closed nonempty input sets $L(y)$ and price domains W, the upper bound of cost efficiency $\overline{CE}^{L}(x,y;W)$ is equal to the Debreu-Farrell input measure $DF^{U}(x,y)$ relative to the augmented input set $U(y)=co(L(y))+W^{U}$, where $co(L(y))$ denotes the convex hull of the input set.

PROOF 3.1 In problem (3.4), the inputs set $L(y)$ can be harmlessly substituted by it's convex hull, i.e. $co(L(y))$, because the objective function is linear in inputs (see Varian (1984) for further details). Hence, problem (3.4) can equivalently be defined as:

(i) $$\overline{CE}^{L}(x,y;W)=\max_{w\in W}\min_{x'\in co(L(y))}\left\{x'w \mid xw=1\right\}=\min_{x'\in co(L(y))}\max_{w\in W}\left\{x'w \mid xw=1\right\}.$$

This problem embeds the following Linear Programming problem:

(ii) $$\max_{w\in W}\left\{x'w \mid xw=1\right\}=\max_{w\in\Re_{+}^{q}}\left\{x'w \mid xw=1;Aw\geq 0\right\}.$$

The dual associated with this problem is:

(iii) $$\min_{q,b\in\Re_{+}^{l}}\left\{q \mid x'+bA\leq qx\right\}.$$

Substituting (iii) for (ii) in (i) gives:

(iv) $$\overline{CE}^L(x,y;W) = \min_{x' \in co(L(y))} \left(\min_{q,b \in \Re^l_+} \{ q \mid x' + b A \le q x \} \right) = \min_q \{ q \mid q x \in co(L(y)) + W^U \} \quad .$$

$$= DF^U(x,y) \, . \qquad \textit{Q.E.D.}$$

The above theorem demonstrates that it makes good sense to use radial Debreu-Farrell measure against the augmented input set to obtain a proxy for economic efficiency, as is sometimes done in the DEA literature. In addition, this theorem underlines the interpretation of such efficiency score as the *upper bound* of cost efficiency. Heuristics behind the equivalence are as follows. The augmented reference set $U(y)$ represents all input vectors x that involve a minimum cost of xw at most, for *all* prices $w \in W$. As $L(y)$ is always contained within $U(y)$, the upper bound $\overline{CE}^L(x,y;W)$ can never exceed the value of $DF^L(x,y)$, but depending on the price domain W, it can be substantially lower. Thus, this upper bound can give a substantially better approximation for economic efficiency than the Debreu-Farrell measure alone. Note that the elements of $U(y)$ need not be included in $L(y)$, and hence they need not suffice to produce output y. For example, $U(y)$ has convex isoquants even if the isoquants of the true production set are non-convex. However, the non-feasible elements can be harmlessly used for cost comparisons, because $L(y)$ necessarily contains at least one input vector that can produce y at a cost of xw at most. As $DF^U(x,y)$ is computationally convenient, we will elaborate on how to obtain empirical approximations for $U(y)$ in Section 4 below.

Figure 3 illustrates graphically the upper bound of cost efficiency relative to $L(y)$, which equals the Debreu-Farrell measure relative to set $U(y)$. The slopes of the two isocost lines drawn in Figure 3 represent the extreme scenarios of relative prices included in a hypothetical price domain W. Notice that the radial reference point R'_{CE} is an infeasible production plan, as it lies outside the input correspondence $L(y)$. However, $L(y)$ contains two production plans R_{CE1} and R_{CE2} that produce the output y at the same cost as the reference point R'_{CE}. Note that while $L(y)$ is non-monotonous and non-convex, $U(y)$ is both monotonous and convex. Therefore, for the purpose of evaluating cost efficiency, the input set can be 'monotonised' and 'convexified' without harm. However, this does neither imply that the entire production set T could be convexified, nor that the convexification would be allowed for measuring technical or allocative efficiency.

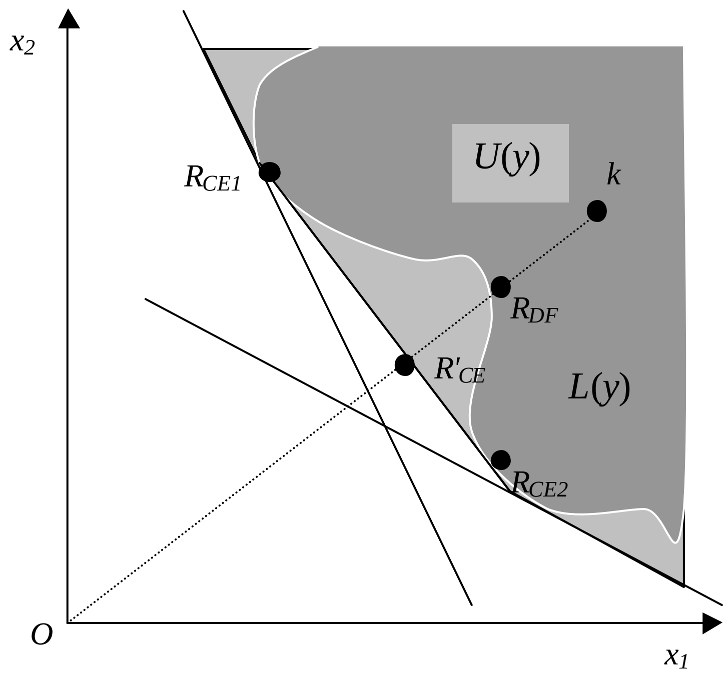


Figure 3: Illustration of the upper bound of cost efficiency, which is the Debreu-Farrell measure relative to set the augmented set U rather than L(y).

Interestingly, if price information is complete, i.e. the price domain W only contains multiples of the true price vector, the upper bound (3.4) always equals the cost efficiency measure (2.4). On the other hand, if no price information is available, sets $U(y)$ and $L(y)$ coincide, and (3.4) equals the Debreu-Farrell measure (2.6) in the special case of monotonous and convex input set. This, however, need not be the case in general. Therefore, the upper bound (3.4) can give a better approximation for cost efficiency in case of non-convex technologies even when no price information is available, i.e. price domain W contains the whole non-negative orthant of real valued vector space. Finally, the properties of measure (3.4) effectively induce a gradual transition from Debreu-Farrell efficiency (associated with no price information) to economic efficiency (associated with full complete price information).

A lower bound for cost efficiency can be obtained in a similar way. More specifically:

(3.5) $$\underline{CE}^L(x,y;W) = \min_{w \in W} \min_{x' \in L(y)} \{x'w \mid xw = 1\} \ .$$

Using duality theory, we can also relate this lower bound to the Debreu-Farrell measure (2.6).

THEOREM 3.2: For all closed nonempty input sets $L(y)$ and price domains W, the lower bound of cost efficiency $\underline{CE}^L(x,y;W)$ is equal to Debreu-Farrell input measure $DF^V(x,y)$ defined with respect to the augmented input set $V(y) = co(L(y)) + W^V$.

PROOF 3.2 In problem (3.5), the objective function is linear in inputs and prices, and hence $L(y)$ can be replaced by $co(L(y))$, and moreover the order of the two minimization problems can be safely reversed:

(i) $\underline{CE}^{L}(x, y; W) = \min_{x \in co(L(y))} \left(\min_{w \in W} \{ x' w | x w = 1 \} \right).$

This problem embeds the following Linear Programming problem:

(ii) $\min_{w \in W} \{ x' w | x w = 1 \} = \min_{w \in \Re^{q}_{+}} \{ x' w | x w = 1; Aw \geq 0 \}$.

The dual associated with this problem is:

(iii) $\max_{q, b \in \Re^{l}_{+}} \{ q \,| x' - bA \geq q x \} = \min_{q, a \geq 0} \{ q \,| x' + a A_i \leq q x; \; i \in L \}$.

Substituting (ii) by (iii) in (i) gives

(vi) $\underline{CE}^{L}(x, y; W) = \min_{x' \in co(L(y))} \left(\min_{q, a \geq 0} \{ q \,| x' + a A_i \leq q x; \; i \in L \} \right)$

$= \min_{q} \{ q \,| q x \in co(L(y)) + W^{V} \} = DF^{V}(x, y)$. *Q.E.D.*

Combined with the upper bound, the lower bound can be used to construct an interval approximation for the true cost efficiency. Theorem 3.2 implies that the radial technical efficiency measure relative to an augmented input set can in fact be used for computing the entire interval for economic efficiency. We think that also the worst case scenario, as reflected by the augmented reference set *V*(*y*) gives valuable information in many applications. The set *V*(*y*) represents all input vectors that produce output *y* at a minimum cost of xw at most, for *some* prices $w \in W$. Obviously, the reference set *U*(*y*) is always contained within *V*(*y*). Again, the elements of *V*(*y*) need not be contained within *L*(*y*). However, *L*(*y*) contains input vectors that can produce *y* at a cost of xw for some prices. As $DF^{V}(x, y)$ is computationally convenient, we will elaborate on how to obtain empirical approximations for *V*(*y*) in Section 4 below.

Figure 4 illustrates graphically the lower bound relative to set *V*(*y*). Notice that the radial reference point in the boundary of *V*(*y*), denoted by R'_{CEL}, is associated with the nonradial reference point in the boundary of *L*(*y*), denoted by R_{CEL}. In fact, R_{CEL} represents the optimal *x* to the minimization problem (3.5). Interestingly, the upper and lower bounds of cost efficiency provide some economic justification for the general idea of non-radial projections to the technical frontier. Unfortunately, any well-known non-radial measure does not generally coincide with these reference points. As shown by Figure 4, achieving this reference may actually require increase in some inputs.

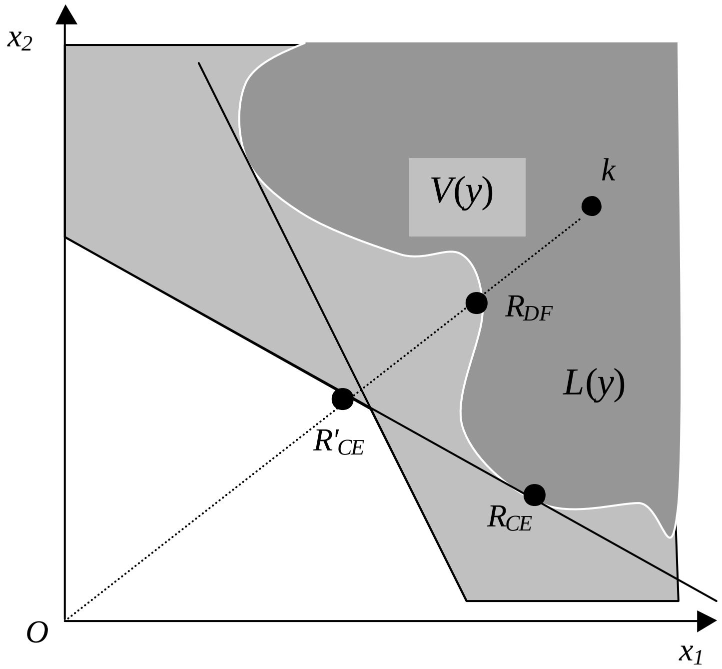


Figure 4: Illustration of the lower bound of cost efficiency, which is the Debreu-Farrell measure relative to set V

Consider the special case of full information. As could be expected the sets $U(y)$ and $V(y)$ coincide, and the lower bound equals the upper bound. On the other hand, in case of no prior price information, the lower bound can fall considerably short of the upper bound. Still, the lower bound typically exceeds zero, which suggests that the theoretical lower bound can give new insight even when prices are totally unavailable.

Finally, using the Farrell decomposition (2.9) and the cost efficiency bounds derived above, we can construct a theoretical interval for allocative efficiency, i.e.

(3.6) $$AE^L \in \left[\underline{AE}^L, \overline{AE}^L\right],$$

where the lower bound is given by

(3.7) $$\underline{AE}^L = \underline{CE}^L / DF^L,$$

and the upper bound is respectively

(3.8) $$\overline{AE}^L = \overline{CE}^L / DF^L.$$

It is immediate from (3.6) and (3.7) that the interval for allocative efficiency is simply cost efficiency interval normalised by technical efficiency. Nevertheless, it can be highly informative to decompose the estimated cost efficiency measures into its two components so as to gain insight of the underlying sources of inefficiency, i.e. to what extent the observed overall inefficiency is due to inefficient planning versus inefficient operation. In this respect, even a rough interval for allocative efficiency can provide valuable information in comparison to the level of technical efficiency. Moreover, the interval generally diminishes by a more accurate price information.

4. Empirical estimators

Above we extended the general theory of efficiency measurement to bridge the gap between the opposite cases of perfect price information on one hand and no price information whatsoever on the other. Although the above measure can deal with incomplete price information, complete information on the input set $L(y)$ is still required. Unfortunately, one typically faces imperfect information on the production possibilities in empirical analysis. However, empirical inference is possible by approximating the true input correspondence by particular empirical input sets constructed from empirical data.

Various approximations can be employed, depending on the assumptions imposed on production possibilities and the data used for approximating it. In the nonparametric DEA literature, the most popular approximation of production possibilities set T is the *convex monotone hull* of observations used, among others, by Banker *et al.* (1984) for estimating Debreu-Farrell input efficiency and Banker and Maindiratta (1988) for estimating overall economic efficiency. In case of these standard convex DEA technologies, the upper bound for cost efficiency can be easily computed by standard weight restriction techniques, e.g. by incorporating the price cone *W* in the cone-ratio DEA model by Charnes et al. (1990).

However, the standard DEA models build on the assumption of convex production set (i.e. T=co(T)), which is more restrictive than convexity of the input correspondences (i.e. $L(y)$=co($L(y)$), $y \in \Re^p_+$). As emphasised already by Farrell (1959) (and more recently Kuosmanen (2001) and Cherchye et al. (2000)), there is no reason to expect production technologies to be convex, and economic phenomena such as economies of scale and of specialization significantly violate convexity. Because of the restrictive nature of the convexity assumption, we choose to derive our empirical estimators from a convexity-free DEA model.

In the previous sections of this study we only assumed sets $L(y)$ and T to be closed and nonempty. However, we need to impose some additional structure for the production possibilities in order to obtain any meaningful empirical reference set. Therefore, we henceforth resort to the *free disposable hull* (FDH) technology (Deprins *et al.* (1984), and Tulkens (1993)). The FDH approximation of the input correspondence L can be formally represented by:

(4.1) $$\hat{L}_{FDH}(y) = \left\{ x \in \Re^q_+ \middle| y \le l^T Y; x \ge l^T X; l^T e = 1; l_j \in \{0,1\} \quad \forall j \in S \right\} .$$

If the production set T is monotonous, i.e. inputs and outputs are freely (=strongly) disposable, and all the observations included in the data set are technically feasible, then $\hat{L}_{FDH}(y) \subseteq L(y)$. Note that these two assumptions are also imposed in standard DEA, the only difference between the free disposable hull and convex monotone hull being the (debatable) convexity postulate. Under the two assumptions imposed here, measuring technical efficiency relative to the FDH approximation gives an upper bound for the true Debreu-Farrell efficiency, i.e.

(4.2) $\overline{DF}^{L}(x,y) = DF^{\hat{L}_{FDH}}(x,y) = \min_{q,l}\left\{q \,\middle|\, l^T X \le qx;\ l^T Y \ge y; l^T e = 1;\ l_j \in \{0,1\}\ \forall j \in S\right\}$.

Solving a Mixed Integer-Linear Programming problem for each DMU can obtain these estimates. Tulkens (1993) has also provided a specially tailored enumeration algorithm for this problem.

Unfortunately, a lower bound for technical efficiency cannot be obtained when L is unknown, except for the value of zero, which provides a theoretical infimum when no free lunch is allowed, i.e. production of positive output bundles is impossible with non-positive input levels.

The FDH approximation also forms a good starting point for cost efficiency evaluation. For this purpose, we construct an FDH approximation for the price-augmented reference set $U(y)$ as

(4.3) $\hat{U}_{FDH}(y) = \left\{x \in \Re^q \,\middle|\, x \ge l^T c(y) + bA;\ b \in \Re_+^l;\ l \in \Re_+^{cardc(y)};\ l^T e = 1\right\}$,

where $c(y)$ denotes the input matrix of such observed production plans that produce higher output than y, i.e.

(4.4) $c(y) = \left\{x_j \in X \,\middle|\, y_j \ge y\right\}$.

Using Theorem 3.1 we obtain the following upper bound for cost efficiency

(4.5) $\overline{CE}^{\hat{L}_{FDH}}(x,y) = DF^{\hat{U}_{FDH}}(x,y)$

$= \min_{q,l,b}\left\{q \,\middle|\, l^T c(y) + bA \le qx;\ b \in \Re_+^l;\ l \in \Re_+^{cardc(y)};\ l^T e = 1\right\}$.

Note that problem (4.5) can be solved by standard Linear Programming algorithms (after first enumerating the matrix $c(y)$), despite the fact that we built on a non-convex FDH technology.

Similarly, for computing the empirical lower bound for cost efficiency we first construct the empirical reference set $\hat{V}_{FDH}(y)$:

(4.6) $\hat{V}_{FDH}(y) = \left\{x \in \Re^l \,\middle|\, x \ge l^T c(y) + aA_i,\ i \in L;\ a \ge 0;\ l \in \Re_+^{cardc(y)};\ l^T e = 1\right\}$.

Using Theorem 3.2, the empirical lower bound for cost efficiency can be computed as

(4.7) $\underline{CE}^{\hat{L}_{FDH}}(x,y) = DF^{\hat{V}_{FDH}}(x,y)$

$= \min_{q,l,a}\left\{q \,\middle|\, qx \ge l^T c(y) + aA_i;\ i \in L;\ a \ge 0;\ l \in \Re_+^{cardc(y)};\ l^T e = 1\right\}$.

Here we can easily enumerate the solution by solving the embedded problem

(4.8) $\min_{q,l,a}\left\{q \,\middle|\, qx \geq l^T c(y) + aA_i; a \geq 0; l \in \Re_+^{cardc(y)}; l^T e = 1\right\}$

for each $i=1,\ldots,l$ and subsequently select the minimum of the obtained solutions. This is an effective strategy especially when l is small. For larger problems, more efficient algorithms cold be developed, but we leave that issue for further research.

As pointed out in the previous section, the convexification of input correspondence is harmless due to the linearity of the cost function. Therefore, the price-augmented FDH input set $\hat{U}_{FDH}(y)$ can be equivalently obtained as the input correspondence of empirical production sets that impose convexity in input space only (see e.g. Bogetoft (1996), and Post (2001)). In the above FDH formulas, potential non-convexity in the graph or output space is taken into account by first filtering out such observations that produce smaller amount of any output than the evaluated production plan.

The empirical bounds of cost efficiency derived above have following properties. If FDH input set is contained within the true input correspondence, then the empirical upper bound is always greater than or equal to the theoretical upper bound, i.e. $\overline{CE}^{\hat{L}^{FDH}} \geq \overline{CE}^{L}$. Moreover, as the theoretical upper bound is greater than or equal to the true cost efficiency, i.e. $\overline{CE}^{L} \geq CE^{L}$, it follows that the true cost efficiency can never exceed the empirical upper bound, i.e. $\overline{CE}^{\hat{L}^{FDH}} \geq CE^{L}$. In addition, under some generally accepted assumptions of the distribution of inefficiency terms, the FDH converges to the true production set as the sample size is increased (see Park, Simar and Weiner, 1997). Therefore, the empirical upper bound asymptotically converges to the theoretical upper bound, i.e. $\overline{CE}^{\hat{L}^{FDH}} \to \overline{CE}^{L}$.

Similarly, the empirical lower bound is always greater than or equal to the theoretical lower bound, i.e. $\underline{CE}^{\hat{L}^{FDH}} \geq \underline{CE}^{L}$. However, as $\underline{CE}^{L} \leq CE^{L}$, the empirical lower bound can generally be greater or less than the true cost efficiency. Thus, the lower bound is essentially a sample property. We would underline that the true economic efficiency can in some cases fall short of the empirical lower bound i.e. due to the finite sample error. However, the empirical lower bound asymptotically converges to the corresponding theoretical value, i.e. $\underline{CE}^{\hat{L}^{FDH}} \to \underline{CE}^{L}$. Furthermore, the lower bound estimate relative to a sample can still be a valuable information for practical purposes, as there is no evidence of DMUs operating with lower cost than the reference cost on the lower bound.

Unfortunately, a more precise theoretical interval than the natural [0,1] range cannot be obtained for allocative efficiency, because we can only characterize upper bounds for both cost and technical efficiency. Still, it can be interesting to consider allocative efficiency interval observed in the *sample* relative to the empirical technology, i.e.

(4.9) $AE^{\hat{L}_{FDH}} \in \left[\underline{AE}^{\hat{L}_{FDH}}, \overline{AE}^{\hat{L}_{FDH}}\right].$

The upper and lower bound of allocative efficiency can be computed from (3.6) and (3.7) by substituting L by its empirical approximation. In a large enough data set the sample interval can give a good approximation due to the asymptotic properties of FDH approximation.

5. Empirical illustration

The nonparametric approach to efficiency measurement has seen extensive application for studying the financial industry. For example, Berger and Humphrey (1997) find that 69 out of 122 frontier efficiency studies for financial institutions use the nonparametric approach. To illustrate the use of incomplete price information in economic efficiency analysis, we performed an empirical application for commercial banks. Specifically, we used a data set with 1997 financial statement data of the 453 largest commercial banks in the European Union[22]. For convenience, we use a simplified representation of the bank technology, which involves a single output, *total earning assets*, and a three inputs, *equity capital*, *debt capital* and *operational costs* (which aggregates all inputs apart from equity and debt). All variables are measured in millions of Euro. Figure 1 illustrates graphically the employed technology. Furthermore, Table 1 presents some descriptive statistics for the data set.

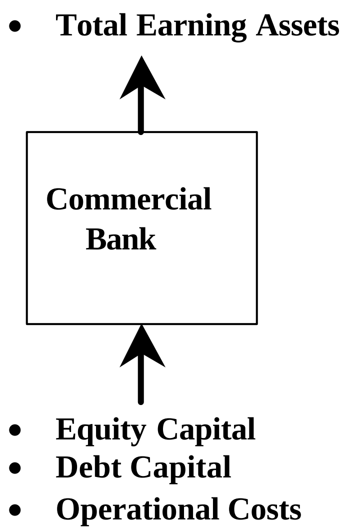


Figure 5: Graphical representation of the bank technology.

Table 1: Descriptive statistics for the bank sample.

	Tot. Earn. Ass.	Eq. Cap.	Debt cap.	Op. Costs
Mean	1027.88	468.49	23389.64	24283.46
Median	222.25	64.34	4005.60	3964.36
Maximum	30808.54	10433.01	568501.80	574533.80
Minimum	8.35	1.23	500.70	533.43
St. Dev.	2547.49	1286.28	60211.62	62961.11
Skewness	5.82	4.83	4.84	4.77
Kurtosis	51.52	29.90	31.24	29.91

To measure cost efficiency and allocative efficiency, information is required on the prices of the inputs, say $w = (w_1 \quad w_2 \quad w_3)$, where w_1 denotes the price of equity capital, w_2

[22] In this article, we use BankScope data provided by Bureau van Dijk Nederland.

the price of debt capital and w_3 the price of other inputs. Our data set includes operational cost, which is computed using the price of other inputs, w_3. However, the prices of capital inputs are not available. The price of debt capital is not available, because interest revenues and interest expenditures are aggregated as *net interest income*. In addition, the cost of capital for equity relates to the *ex ante* required rate of return from future dividend payments and capital gains. This rate of return can't be inferred from *ex post* financial statements. In addition, the required rate of return on equity generally depends on the debt/equity ratio, and therefore can not be assumed to apply to reference units that have a different debt/equity ratio than the evaluated DMU. These complications exclude the possibility of direct measurement of cost efficiency and allocative efficiency.

As discussed above, the input-oriented Debreu-Farrell measure for technical efficiency can be used as an upper bound for cost efficiency. However, prior price information can derive better bounds for cost efficiency. For example, economic theory assumes that the cost of equity capital exceeds that of debt capital. The rationale behind this assumption is that equity involves more risk for the capital supplier than debt does, because debt claims are senior to equity claims. This assumption can be enforced by specifying the matrix A as:

(5.1) $A = (1 \quad -1 \quad 0)$.

Additional information on the prices could be found, e.g. using empirical estimates from capital market data or subjective assessments by bank managers. However, for the purpose of illustration, we resort to this single assumption on the relative price of equity and debt capital. Our purpose is to demonstrate by this example that valuable information can be gained by this simple restriction on input prices.

For these purposes we computed the Debreu-Farrell measure (4.2), and the empirical upper and lower bounds for cost efficiency associated with A, i.e. (4.4) and (4.7). We subsequently computed the empirical bounds on allocative efficiency, i.e. (3.6) and (3.7). Table 2 reports the summary statistics of the cost efficiency bounds and Debreu-Farrell technical efficiency.[23]

Table 2: Statistics on the cost efficiency bounds and Debreu-Farrell measures

	Cost efficiency Lower bound	**Cost efficiency Upper bound**	**Debreu-Farrell**
Mean	0.114	0.978	0.988
Median	0.048	1	1
Maximum	1	1	1
Minimum	0.007	0.312	0.369
St. Dev.	0.176	0.048	0.041
Skewness	3.374	-6.962	-9.141
Kurtosis	12.574	83.382	123.774

[23] More detailed estimation results can be obtained from the authors upon request.

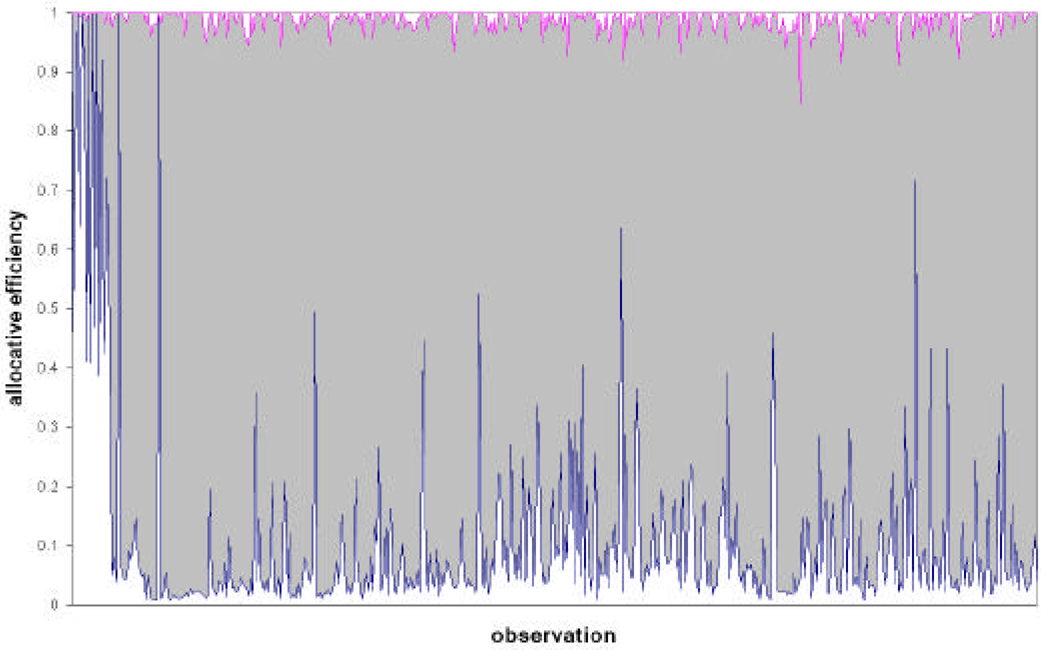


Figure 6: White area represents the improvement in cost efficiency bounds associated with the information that the cost of capital of equity exceeds that of debt.

Figure 6 graphically illustrates the relative improvement in the cost efficiency bounds achieved by incorporating price restriction (5.1), in other words, the computed cost efficiency interval is presented at the normalised scale [0, $DF^{\hat{L}_{FDH}}(x,y)$]=[0,1]. Alternatively, The grey area between the upper line and the lower line in Figure 6 can be viewed as the allocative efficiency interval $\left[\underline{AE}^{\hat{L}^{FDH}}, \overline{AE}^{\hat{L}^{FDH}}\right]$ for the observations. As allocative and cost efficiency intervals generally improve at the same rate, in what follows we refer to intervals and bounds in both meanings. Table 3 further summarises the improvements from using prior price information relative to interval from [0, $DF^{\hat{L}_{FDH}}(x,y)$], i.e. the interval associated with no prior information.

Table 3: Statistics on the improvements due to price information

	Lower Bound	**Upper Bound**
Mean	0.114	0.989
Median	0.049	1.000
Maximum	1.000	1.000
Minimum	0.007	0.846
St. Dev.	0.176	0.018
Skewness	3.371	-2.210
Kurtosis	12.554	11.122

It is clear both from Figure 6 and Table 3 that for a number of banks, incorporating price information substantially narrow intervals from [0,1], i.e the interval associated with no prior information. We see that the reduction in the intervals achieved by the restriction (5.1) solely, as represented by the white area, is quite substantial relative to the additional information content in (5.1). Obviously, for some banks, the improvements are more substantial than for others. For example, for a number of banks the lower bound equals unity. These banks are demonstrably cost efficient in this sample at all prices within the price cone. Note that in the classic framework, when full price data was unavailable, the cost efficiency interval for a Debreu-Farrell efficient bank was the maximal [0,1]. For some of these banks, we were able to provide evidence of their efficiency that was previously unavailable.

In addition, for many banks the upper bound is substantially lower than the Debreu-Farrell measure. For these banks, (1) the radial projection point is on a non-convex part of the isoquant, or alternatively (2) the marginal rates of technical substitution at the projection point differ from the input prices contained within the price cone.

For the majority of banks, however, the improvement in the interval was due to the introduction of the lower bounds. There certainly is a correlation between upper and lower bounds, so that when upper bounds are relatively close to unity, like with most banks in this data set, the lower bounds capture a greater proportion of the narrowing of the bounds, and vice versa. However, notice also that the above results may be subject to small sample bias. The empirical measures are biased above the theoretical bounds if the empirical production set is only a subset of the true production set, as is true in small samples. As discussed above, this positive bias preserves the status of the empirical upper bound as an upper bound for true cost efficiency. However, the empirical lower bound is not necessarily a lower for true cost efficiency, and should be interpreted as a sample property rather than a population property. This needs to be taken into consideration in the interpretation of the empirical bounds. Still, the empirical and theoretical bounds converge as the sample size increases.

We conclude by pointing some interesting opportunities to extend this type of empirical analysis. Naturally, collecting more accurate price restrictions can help to further improve the efficiency intervals. Moreover, when more accurate price data is not available, it can be worth to consider alternative price scenarios, speculate what would happen if the price domain would be specified in a different way, or how particular production assumptions (i.e. convexity, constant returns to scale, etc.) would affect the results. However, these speculations are beyond the scope of the current application. Our purpose was to show that the theoretical framework proposed in this paper provides an operationally tractable basis that requires little additional information, and most remarkably, rests of minimal assumptions on the unknown production process.

6. Concluding remarks

Economic efficiency measures require complete price data, while technical efficiency measures do not utilise any price information whatsoever. In many empirical studies, price information is considered highly relevant, but price data available is considered incomplete for using economic efficiency concepts. We derived upper and lower bounds for overall economic efficiency and allocative efficiency assuming incomplete price data in the form of a convex polyhedral cone. Standard linear programming techniques can be used for computing nonparametric estimates for these bounds from empirical data. The application for European commercial banks presented in Section 5 shows that these bounds can give a substantially better approximation for economic efficiency than technical efficiency measures; a single theoretically sound assumption about the relative prices of equity and debt capital generated a substantial improvement.

In the Introduction, we noted that our approach has a strong similarity with so-called weight restricted DEA models. Now, how do the empirical bounds we derived from Farrell's theoretical model compare to the weight restricted DEA models? In contrast to the weight-restricted models, we distinguish between the marginal properties (shadow prices) of the production set on one hand, and the economic prices on the other. This distinction has at least three important implications. First, it allows one to incorporate information in both envelopment and multiplier problems. Hence, including economic prices does not interfere with the elicitation of information on the shadow prices. Second, the distinction reveals that different notions of economic efficiency require different kinds of price information. For example, it is immediately clear that cost efficiency measures require information on the input prices, but not on the output prices. Third, the distinction facilitates the analysis of non-convex production technologies, which we consider desirable e.g. to account for economies of scale and of specialisation. In addition to distinguishing between technical and economic information, we compute both upper and lower bounds so as to present the whole interval of possible efficiency values, whereas weight restricted DEA models usually consider an upper bound at 'most favourable' weights only.

Regardless of the limitations of weight restricted DEA models for measuring economic efficiency, proper restrictions of technology parameters can help to improve the approximation of the production set. However, it should be understood that weight restrictions reflect technical information on marginal products or marginal substitution or transformation rates, rather than information on economic prices. In fact, weight restrictions have been utilized to reflect the former information in DEA right from the beginning, for example to impose particular assumptions of returns-to-scale (see e.g. Banker et al, 1984). Therefore, it is useful to distinguish between technology related information on one hand, and price or value related information on the other, both of which can be included in the model by parameter restrictions. We conclude by calling for more attention to the interpretation of weight restricted DEA models.

References

Allen, R., A. D. Athanassopoulos, R.G. Dyson and E. Thanassoulis (1997): Weight Restrictions and Value Judgements in DEA: Evolution, Development and Future Directions. *Annals of Operations Research* 73, 13-34.

Banker, R.D., Charnes, A., and W. W. Cooper (1984): Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis, *Management Science* 30(9), 1078 – 1092.

Banker, R.D. and A. Maindiratta (1988): Nonparametric analysis of technical and allocative efficiencies in production, *Econometrica* 56(6), 1315-1332.

Berger, A. N. and D. B. Humphrey (1997): Efficiency of Financial Institutions: International Survey and Directions for Future Research, *European Journal of Operational Research* 98, 175-212.

Bogetoft, P. (1996): DEA on a Relaxed Convexity Assumption, *Management Science* 42(3), 457-465

Briec, W., and Lesourd (1999): Metric Distance Function and Profit: Some Duality Results, *Journal of Optimization Theory and Applications* 101, 15-33

Chambers, R.G., Y. Chung, and R. Färe (1998): Benefit and Distance Functions, *Journal of Economic Theory* 70, 407-419

Chambers, R.G., Y. Chung, and R. Färe (1998): Profit, Directional Distance Functions and Nerlovian Efficiency, *Journal of Optimization Theory and Applications* 98(2), 351-364.

Charnes, A. and W. W. Cooper (1985): Preface to Topics in Data Envelopment Analysis, *Annals of Operations Research* 2, 59-94.

Charnes, A., W. Cooper, B. Golany, L. Seiford, J. Stutz (1985): Foundations of Data Envelopment Analysis for Pareto-Koopmans Efficient Empirical Production Functions, *Journal of Econometrics,* 30(1), 91 – 107

Charnes, A., W. Cooper, Z. M. Huang and D. B. Sun (1990): Polyhedral cone-ratio DEA models with an illustrative application to large commercial banks, *Journal of Econometrics* 46, 73-91.

Charnes, A., W. Cooper, and E. Rhodes (1978): Measuring the Efficiency of Decision Making Units, *European Journal of Operational Research* 2, 429 – 444.

Cherchye, L., T. Kuosmanen, and G.T. Post (2000): What is the Economic Meaning of FDH? A Reply to Thrall, *Journal of Productivity Analysis* 13(3), 259-263

De Borger, B., G. D. Ferrier, and K. Kerstens (1998): The Choice of a Technical Efficiency Measure on the Free Disposal Hull Reference Technology: A Comparison Using US Banking Data, *European Journal of Operational Research* 105, 427-446

Debreu, G. (1951): The Coefficient of the Resource Utilization, *Econometrica* 19(3), 273 – 292.

Deprins, D., D. Simar, H. Tulkens (1984): Measuring Labor Efficiency in Post Offices, p. 243 - 267 in M. Marchand, P. Pestieau and H. Tulkens (Eds.), *The Performance of Public Enterprises: Concepts and Measurements*, North Holland.

Farrell, M. J. (1957): The Measurement of Productive Efficiency, *Journal of the Royal Statistical Society* Series A 120(3), 253 – 281.

Farrell, M.J. (1959): Convexity assumption in theory of competitive markets, *Journal of Political Economy* 67, 377-391.

Färe, R. (1988): *Fundamentals of Production Theory*, Springer-Verlag, Berlin

Färe R., and D. Primont (1995): *Multi-Output Production and Duality: Theory and Applications*, Kluwer Academic Publishers, Boston.

Kopp, R. J. (1981): The Measurement of Productive Efficiency: A Reconsideration, The Quarterly Journal of Economics, 96(3), 477-503

Kuosmanen, T. (2001): DEA with Efficiency Classification Preserving Conditional Convexity, forthcoming in *European Journal of Operations Research*

Nerlove, M. (1965): *Estimation and Identification of Cobb-Douglas Production Functions*, Rand McNally, Chicago, Illinois

Park, B.U., L. Simar and C. Weiner (1997): FDH Efficiency Scores from a Stochastic Point of View, Discussion Paper 9715, Institut de Statistique, Universtite Catholique de Louvain, under revision for *Econometric Theory*

Post, G.T. (2001): Estimating Non-Convex Production Sets: Imposing Convex Input Sets and Output Sets in Data Envelopment Analysis, forthcoming in *European Journal of Operations Research.*

Shephard, R.W. (1953): *Cost and Production Functions*, Princeton University Press, Princeton

Thompson, R.G., F.D. Singleton Jr., R.M. Thrall, and B.A. Smith (1986): Comparative Site Evaluations for Locating High Energy Lab in Texas, *TIMS Interfaces*, 1380-1395

Thompson, R.G., L.N. Langemeier, C.-T. Lee, E. Lee, and R.M. Thrall (1990): The Role of Multiplier Bounds in Efficiency Analysis with Application to Kansas Farming, Journal of Econometrics 46, 93-108

Tulkens, H. (1993): On FDH Analysis: Some Methodological Issues and Applications to Retail Banking, Courts and Urban Transit, *Journal of Productivity Analysis* 4(1-2), 183-210

Varian, H.R. (1984): The Non-parametric Approach to Production Analysis, *Econometrica* 52, 279-297

NONPARAMETRIC EFFICIENCY ANALYSIS UNDER UNCERTAINTY:

A FIRST-ORDER STOCHASTIC DOMINANCE APPROACH

TIMO KUOSMANEN
Helsinki School of Economics and Business Administration, Finland
E-mail: Kuosmane@hkkk.fi

THIERRY POST
Erasmus University Rotterdam, The Netherlands
E-mail: GTPost@few.eur.nl

Helsinki School of Economics and Business Administration
Working Papers W-230*
August 1999

Abstract

This paper extends the nonparametric approach to efficiency analysis to include uncertainty for input-output prices. We introduce a novel concept for efficiency under uncertainty together with necessary and sufficient first-order stochastic dominance (FSD) conditions. These FSD conditions require minimal assumptions with respect to the preferences of the decision-maker and the distribution of the prices. In addition, the FSD conditions include the well-established efficiency conditions by Koopmans and Nerlove as limiting cases. Furthermore, the FSD conditions can be tested empirically using standard mathematical programming techniques.

Key words: *nonparametric efficiency analysis, Data Envelopment Analysis, performance evaluation under uncertainty, stochastic dominance.*

1. Introduction

Following Koopmans (1951), most studies in the field of nonparametric efficiency analysis, including the classic DEA papers by Charnes, Cooper and Rhodes (1978), Banker, Charnes and Cooper (1984), and Charnes et al. (1985), assess efficiency in terms of *technical efficiency*. Technical efficiency is an important organisational objective. However, it generally is not the sole objective. Therefore, technical efficiency generally is a necessary,

ISSN 1235-5674, ISBN 951-791-413-X

but not sufficient, condition for *effectiveness*, i.e. the extent to which DMUs succeed in achieving their organisational objectives. Consequently, there is the need to include additional preference information in efficiency analysis, so as to discriminate between technically efficient production plans.

One attractive approach to include additional preference information is to use behavioural assumptions from microeconomic theory. Most notably, Nerlove (1965) derived the concept of profit efficiency from the assumption that DMUs maximise profit given input and output prices. In the nonparametric literature, Varian (1984), Banker and Maindiratta (1988), and Färe and Grosskopf (1997), among others, have proposed empirical tests and efficiency measures related to Nerlovian profit efficiency[24]. A problem not recognised thus far is that the input-output prices typically involve uncertainty. If this is the case, evaluating efficiency of DMUs at the realised prices observed *ex post* by the analyst is unfair towards the units who actually did not have all that information at hand when they fixed their production plans.

This paper extends the nonparametric approach to efficiency analysis to include uncertainty. Using the theory of first-order stochastic dominance (Hadar and Russell 1969), we derive necessary and sufficient conditions for efficiency under uncertainty that require minimal assumptions about the statistical distribution of the prices and the organisational objectives of the DMUs. These conditions include the condition for Koopmans' technical efficiency as the limiting case where prices are unrestricted strictly positive random variables. Similarly, the condition for Nerlovian profit efficiency is the limiting case where prices are known with complete certainty. In addition, we derive empirical tests for the necessary and sufficient conditions that can be implemented using standard mathematical programming techniques.

The remainder of the text is organised as follows. Section 2 introduces the established efficiency concepts by Koopmans and Nerlove. Section 3 discusses production analysis under price uncertainty, and introduces the first-order stochastic dominance conditions. Section 4 illustrates how the FSD conditions can be tested empirically. Finally, Section 5 draws conclusions and gives suggestions for future research.

2. Efficiency definitions

We intend to evaluate the efficiency of DMUs that operate under a particular production technology, as represented by the production possibilities set:

(1) $$T = \{(y, x) \in \Re_{+}^{q+m} \mid \text{input } x \text{ can produce output } y\} \ .$$

Throughout the text, the production data of the DMUs are represented by the output matrix $Y = (y_1 ... y_n)^T$, with $y_j = (y_{1j} \cdots y_{qj})$, and the input matrix $X = (x_1 ... x_n)^T$, with $x_j = (x_{1j} \cdots x_{mj})$. In addition, we use $S = \{1,...,n\}$.

[24] Neither Varian, nor Banker and Maindiratta explicitly refer to Nerlove, but the concept of profit efficiency is fundamentally embedded into both studies.

Koopmans efficiency[25] for the evaluated (say k-th) unit can be formally defined as follows: DMU k is technically efficient in Koopmans sense if and only if, given the production technology, none of its inputs or outputs could be improved without worsening some of its other inputs or outputs, i.e.:

(2) $$\max_{x,y,s',s}\left\{ \begin{pmatrix} s' & s \end{pmatrix} e \,\middle|\, \begin{pmatrix} y - s' \\ -x - s \end{pmatrix} \geq \begin{pmatrix} y_k \\ -x_k \end{pmatrix}; (y,x) \in T; (s',s) \in \Re_{+}^{q+m} \right\} = 0 .$$

In the nonparametric literature, Charnes, Cooper and Rhodes (1978), Banker, Charnes and Cooper (1984), and Charnes et al. (1985), among others, have proposed empirical tests and efficiency measures related this efficiency concept.

The recognised problem of technical efficiency is that it typically is a necessary, but not sufficient condition for optimal economic behaviour. Technical efficiency considers means for achieving given ends, but the specification of the ends themselves is ignored. In other words, potential inefficiencies in the *allocation* of resources to produce the optimal output bundle are not taken into account. Theoretically, this point was already manifested in the seminal Farrell decomposition (Farrell, 1957).

One possibility to enrich efficiency analysis is to change the perspective towards economic efficiency concepts. Most notably, Nerlove (1965) introduced the concept of profit efficiency. This efficiency concept uses the microeconomic assumption of profit maximisation subject to technological constraints, as represented by the production set, and environmental factors, as represented by output prices $p_k = (p_{1k} \cdots p_{qk})^T \in \Re_{+}^{q}$ and input prices $w_k = (w_{1k} \cdots w_{mk})^T \in \Re_{+}^{m}$.

A DMU k is profit efficient if and only if it is not possible to achieve higher profit, given the technology and the prices, i.e.:

(3) $$\max_{y,x}\left\{ (yp_k - xw_k) - (y_k p_k - x_k w_k) \,\middle|\, (y,x) \in T \right\} = 0$$

In the nonparametric literature, Varian (1984), Banker and Maindiratta (1988), and Färe and Grosskopf (1997), among others, have proposed empirical tests and efficiency measures related to Nerlovian profit efficiency.

A problem not recognised thus far is that the input-output prices typically involve uncertainty for the DMUs. Prices are typically determined by demand and supply forces that are (partly) beyond the control of the DMUs, and in addition that are (partly) unpredictable for the DMUs. For example, in the classic Cournot (1838) duopoly model, DMUs control their production plans, but prices are determined on the market. Under such circumstances, DMUs require perfect information on demand responses and production plans of competitors, to perfectly anticipate (the Nash equilibrium) prices. Moreover, even if the nominal prices

[25] Also known as Pareto-Koopmans efficiency. This should not be confused to another well-established technical efficiency concept proposed by Debreu (1951) and especially Farrell (1957) that exclusively focuses on radial (equiproportional) input reduction and/or output augmentation.

would be fixed i.e. by long-term contracts or regulatory acts, the unpredictable fluctuations in uncontrollable environmental factors such as business cycles, interest rates, and inflation guarantee that the real prices in terms of opportunity costs typically involve uncertainties beyond hedging. Thus, evaluating efficiency of DMUs at the realised prices observed *ex post* by the analyst is unfair towards the units who actually did not have all that information at hand when they fixed their production plans.

3. Efficiency analysis under uncertainty

In contrast to the standard model, we assume the DMU under evaluation faces uncertainty for the price vectors. More specifically, we assume that the prices are random variables with domain $D \subseteq \Re_{+}^{s+m}$ and joint distribution function $F : D \rightarrow [0,1]$. Both the price domain D and the distribution function F can be harmlessly assumed DMU specific.

For simplicity, we use a polyhedral convex cone to represent the price domain. More specifically, we assume the following structure for the price domain:

(4) $$D = \left\{ (p,w) \in \Re_{+}^{q+m} \middle| Ap + Bw \geq 0; (p,w) \geq e \right\} .$$

This cone represents the price domain in terms of l linear inequalities. A and B are $l \times q$ and $l \times m$ matrices respectively and c is a $l \times 1$ vector. In addition, the prices are assumed to be strictly positive. Restricting the prices to exceed an arbitrary positive constant e imposes that assumption. It should be noted that the assumed form of price domain may turn out impractical or irrelevant in some empirical cases. However, in many cases, price information does take the form of linear inequalities, or linear inequalities can give a good approximation for more complicated price structures. Especially, the limiting cases of exact and complete price information on one hand, and no price information whatsoever on the other, are special cases of this more general price domain, as shown in more detail below. Finally, polyhedral convex cones are computationally convenient, because they can be included in Linear Programming models.

In DEA literature, Charnes et al. (1990), among others, have used polyhedral convex cones in the so-called weight-restricted DEA models (henceforth WR models). For a detailed survey of WR models, see Allen et al. (1997). Despite the technical similarity, the WR models should not be confused with the approach adopted in this paper. Firstly, WR models use polyhedral convex cones to impose restrictions on the marginal substitution rates of the production set (2.1). By contrast, we deal with the assumptions about the input prices required for measuring economic efficiency, rather than the structure of the production possibilities. Secondly, WR models only focus on the most favourable substitution rates, whereas we consider the least favourable prices (the lower bounds below) in addition to the most favourable ones (the upper bounds below). Finally, WR models relate specifically to the empirical convex hull approximation by Banker et al. (1984). By contrast, we define bounds relative to the true technology T, and show how empirical estimates can be obtained for various empirical approximations, including non-convex ones.

To include uncertainty, we formulate production analysis in terms of the Von Neumann-Morgenstern (1967) Expected Utility Theory. More specifically, we assume that the evaluated DMU selects its production plan to maximise the expected value of a Von Neumann-Morgenstern utility function $U : \Re^1 \to \Re^1$ that is defined over profit, subject to the technological constraints, and subject to the distribution of prices[26]. The following constrained maximisation problem represents the decision problem of the evaluated DMU:

(5) $$\max_{y,x} \{E(U(yp - xw)) | (y,x) \in T\}.$$

The purpose of efficiency analysis is to evaluate whether or not the choice of the production variables was rational, maximising the expected utility function of the DMU. First, notice that it is consistent with the Von Neumann-Morgenstern axioms to express expected utility as follows:

(6) $$E(U(yp - xw)) = \int_D U(yp - xw) \partial F(p,w).$$

Thus, analogous to (3), the DMU k maximises expected utility if and only if:

(7) $$\max_{y,x} \left\{ \int_D U(yp - xw) \partial F(p,w) - \int_D U(y_k p, x_k w) \partial F(p,w) \middle| (y,x) \in T \right\} = 0.$$

Unfortunately, this can not be solved without imposing further assumptions concerning the preference structure (U) and the distribution function (F). However, we intend to impose minimal distribution and preference assumptions, so as to preserve the nonparametric nature of the standard framework. Therefore, we propose to use decision criteria from the theory of stochastic dominance, which is a theory of choice under uncertainty that has seen considerable theoretical development and empirical application in economics in the last decades[27]. The most general form of stochastic dominance makes no assumptions with respect to the form of the disturbance distribution. Furthermore, no specific functional form for decision-maker preferences is assumed. Rather, stochastic dominance relies on assumptions about the general characteristics of decision-maker preferences. In this paper, we use the most general first-order stochastic dominance (henceforth FSD) criterion (Hadar and Russell, 1969). This criterion assumes that decision-makers maximise expected utility and, in addition, that decision-makers are non-satiable, i.e. they prefer more to less.

In general, a distribution function $G_1(z)$ stochastically dominates another distribution function $G_2(z)$ by first-order, if and only if the former involves a probability of exceeding any value for z greater than or equal to that of the latter (with inequality for at least one value for z), i.e.:

---

26 In this paper, we assume that input-output quantities are fully controllable. However, sometimes it is necessary to include uncertainty for the quantities as well. For example, in the Bertrand (1883) oligopoly model, DMUs set prices, and quantities are determined on the market. However, if DMUs can freely choose between Bertrand and Cournot strategies, the resulting Nash equilibrium tends to be the Cournot equilibrium (Qin and Stuart, 1997).

27 For a detailed survey and analysis of the expected utility and stochastic dominance literature, see Levy (1992).

(8) $G_2(z) \succ G_1(z) \quad \forall z \in \Re^1$.

Here '$\succ$' is used to denote the condition that '$\geq$' holds for the entire domain with '>' holding for at least some values of the domain.

In the production problem discussed above, prices are considered random variables. Therefore, profit is a random variable with the following probability distribution:

(9) $$P(y_k p - x_k w \leq p) = \int_{\{(p,w)\in D | y_k p - x_k w \leq p\}} \partial F(p,w).$$

Applying the FSD rule (8) to this distribution function, we find that the evaluated DMU is stochastically dominated by first-order by some feasible production plan, if and only if:

(10) $$\exists (y,x) \in T : \int_{\{(p,w)\in D | y_k p - x_k w \leq p\}} \partial F(p,w) \succ \int_{\{(p,w)\in D | yp - xw \leq p\}} \partial F(p,w) \quad \forall p \in \Re^1.$$

Consequently, a *necessary* condition for optimal production is:

(11) $$\max_{y,x}\left\{ \min_{p}\left\{ \int_{\{(p,w)\in D | y_k p - x_k w \leq p\}} \partial F(p,w) - \int_{\{(p,w)\in D | yp - xw \leq p\}} \partial F(p,w) \right\} \middle| (y,x) \in T \right\} = 0.$$

Similarly, the evaluated DMU stochastically dominates all feasible production plans if and only if:

(12) $$\int_{\{(p,w)\in D | yp - xw \leq p\}} \partial F(p,w) \succ \int_{\{(p,w)\in D | y_k p - x_k w \leq p\}} \partial F(p,w) \quad \forall p \in \Re^1; (y,x) \in T.$$

Consequently, a *sufficient* condition for optimal production is:

(13) $$\max_{y,x}\left\{ \max_{p}\left\{ \int_{\{(p,w)\in D | y_k p - x_k w \leq p\}} \partial F(p,w) - \int_{\{(p,w)\in D | yp - xw \leq p\}} \partial F(p,w) \right\} \middle| (y,x) \in T \right\} = 0.$$

These rules do not require preference assumptions in addition to non-satiability. However, they do require information on the distribution function F. However, the following *necessary* condition for (11) can be obtained for an arbitrary distribution:

(14) $$\max_{y,x}\left\{ \min_{p,w}\left\{ (yp - xw) - (y_k p - x_k w) \middle| (p,w) \in D \right\} \middle| (y,x) \in T \right\} = 0.$$

In other words, optimal behaviour requires that input-output vectors that yield higher profit than the evaluated DMU at all possible price vectors are not attainable.

Similarly, the following *sufficient* condition for (13) can be obtained for an arbitrary distribution:

$$\textbf{(15)} \qquad \max_{y,x,p,w} \left\{ (yp - xw) - (y_k p - x_k w) \middle| (p,w) \in D; (y,x) \in T \right\} = 0 .$$

In other words, if the evaluated DMU yields the maximum profit at all possible price vectors, it necessarily maximises expected utility.

Using duality theory, the following equivalent condition can be obtained for problem (14):

$$\textbf{(16)} \qquad \max_{y,x,s',s} \left\{ e(s' \quad s)e \middle| (y - s', x + s) \in U ; (y,x) \in T ; (s', s) \in \Re_{+}^{q+m} \right\} = 0 .$$

Here U represents the following polyhedral convex cone:

$$\textbf{(17)} \qquad U = \left\{ (y,x) \in \Re^{q+m} \middle| \begin{pmatrix} y \\ -x \end{pmatrix} \geq \begin{pmatrix} y_k + zA \\ -x_k + zB \end{pmatrix} ; z \in \Re_{+}^{l} \right\} .$$

This cone represents all input-output values that yield more profit than the evaluated DMU at all prices from D. It can be constructed from the restrictions that A and B impose on the relative prices.

Proof Problem (14) embeds the following Linear Programming problem, using y^* and x^* for the optimal solutions to the maximisation problem:

$$\min_{p,w} \left\{ (y^* p - x^* w) - (y_k p - x_k w) \middle| (p,w) \in D \right\} =$$
$$\min_{p,w} \left\{ (y^* p - x^* w) - (y_k p - x_k w) \middle| Ap + Bw \geq 0; (p,w) \geq e \right\} .$$

The dual formulation of this problem is:

$$\max_{s',s,,z} \left\{ e(s' \quad s)e \middle| \begin{pmatrix} y^* - s' \\ -x^* - s \end{pmatrix} \geq \begin{pmatrix} y_k + zA \\ -x_k + zB \end{pmatrix} ; (s', s) \in \Re_{+}^{q+m} ; z \in \Re_{+}^{l} \right\} =$$
$$\max_{s',s} \left\{ e(s' \quad s)e \middle| (y^* - s', x^* + s) \in U \right\} .$$

Substituting this problem in (14), we find (16).?

Similarly, duality gives the following equivalent condition for (15):

$$\textbf{(18)} \qquad \max_{y,x} \left\{ \min_{s',s} \left\{ e(s' \quad s)e \middle| (y - s', x + s) \in V ; (y,x) \in T \right\} \middle| (s', s) \in \Re_{+}^{q+m} \right\} = 0 .$$

Here V represents the following polyhedral convex cone:

$$\textbf{(19)} \qquad V = \left\{ (y,x) \in \Re^{q+m} \middle| \begin{pmatrix} y \\ -x \end{pmatrix} \leq \begin{pmatrix} y_k - zA \\ -x_k - zB \end{pmatrix} ; z \in \Re_{+}^{l} \right\} .$$

This cone represents all input-output vectors that yield *less* profit than DMU k at all prices in *D*. Like U, it can be constructed from the restrictions that A and B impose on the relative prices.

Proof Problem (15) embeds the following Linear Programming problem, using $y*$ and $x*$ for the optimal solutions:

$$\max_{p,w}\left\{(y*p - x*w) - (y_k p - x_k w) \middle| (p,w) \in D\right\} =$$
$$\max_{p,w}\left\{(y*p - x*w) - (y_k p - x_k w) \middle| Ap + Bw \geq 0; (p,w) \geq e\right\} .$$

The dual formulation of this problem is:

$$\min_{s',s,z}\left\{e\left(s' \quad s\right)e \middle| \begin{pmatrix} y*-s' \\ -x*-s \end{pmatrix} \leq \begin{pmatrix} y_k - zA \\ -x_k - zB \end{pmatrix}; (s',s) \in \Re_+^{q+m}; z \in \Re_+^l \right\} =$$
$$\max_{s',s}\left\{e\left(s' \quad s\right)e \middle| \left(y*-s', x*+s\right) \in V\right\} .$$

Substituting this problem in (15), we find (18). ?

Figures 1, 2 and 3 graphically illustrate the FSD conditions for a single input - single output technology. Figure 1 illustrates the price domain used in this example as represented by the polyhedral convex cone *D*. Figure 2 displays the production set *T* and the polyhedral convex set U related to *D*. The shaded area represents the intersection of T and U. This intersection represents the attainable input-output vectors that yield higher profit than DMU k at all prices included in *D*. Since this intersection is non-empty, DMU k is stochastically dominated by first-order. Figure 3 displays the production set *T* and the polyhedral convex set *V*. The shaded area represents the set of attainable input-output vectors that do not yield less profit than DMU k at all prices included in *D*. Since this set is non-empty, DMU k does not satisfy the sufficient condition for optimal production.

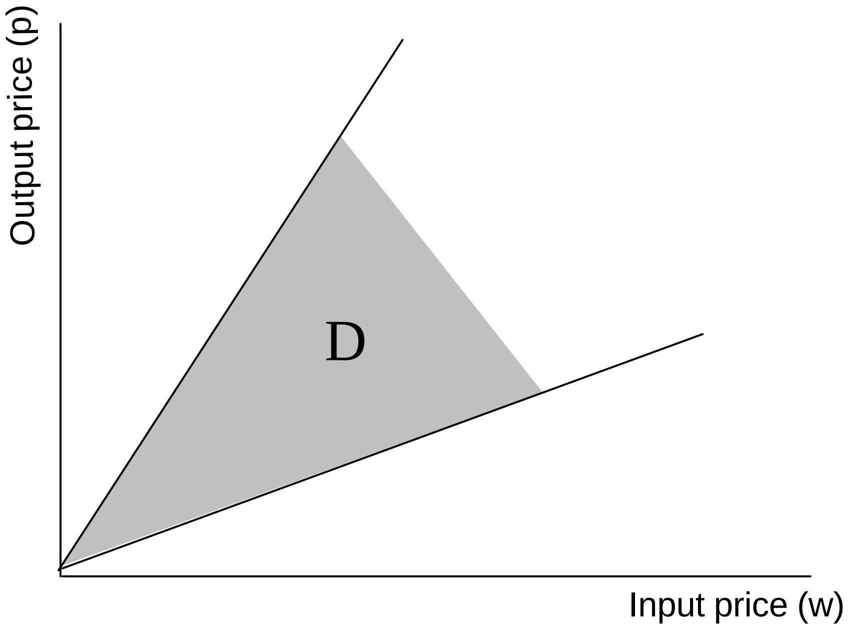


Figure 1: Example of price domain as represented by polyhedral convex cone D.

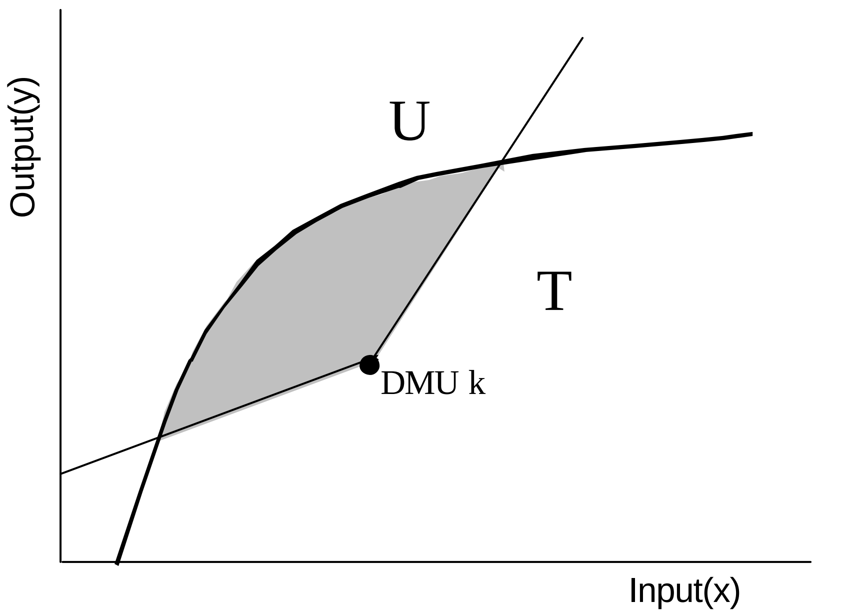


Figure 2: The shaded area represents the intersection of the production set T and the polyhedral cone U. This intersection represents the feasible production plans that dominate DMU k in FSD sense.

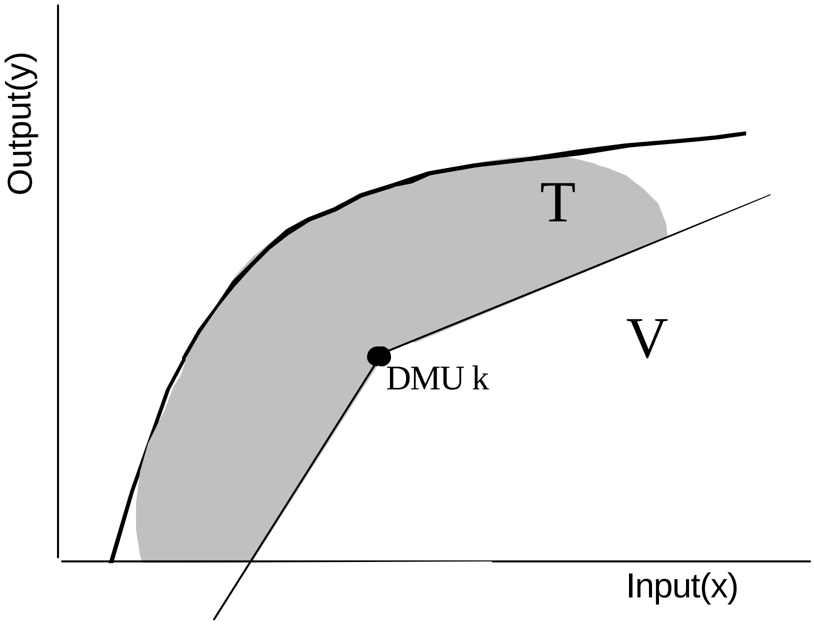


Figure 3: The shaded area represents the set of attainable input-output vectors that do not yield less profit than DMU k at all prices included in D.

These efficiency conditions have some important limiting cases. Obviously, if the price domain is restricted to a unique price vector, i.e. $D = (p_k, w_k)$, both condition (14) and condition (15) are equivalent to the condition for Nerlovian profit efficiency (3). By contrast, if the relative prices are completely unrestricted, i.e. $A = 0$ and $B = 0$, we find the following condition for (16):

(20) $$\max_{x,y,s',s}\left\{ e\left(s' \quad s\right)e \left| \begin{pmatrix} y - s' \\ -x - s \end{pmatrix} \geq \begin{pmatrix} y_k \\ -x_k \end{pmatrix}; (y,x) \in T; (s',s) \in \Re_+^{q+m} \right. \right\} = 0 .$$

Apart from the constant e, this condition is equivalent to the Koopmans condition for technical efficiency (2). Therefore, the efficiency conditions by Koopmans and Nerlove represent the limiting cases of the necessary condition for FSD efficiency, as illustrated by Figure 4.

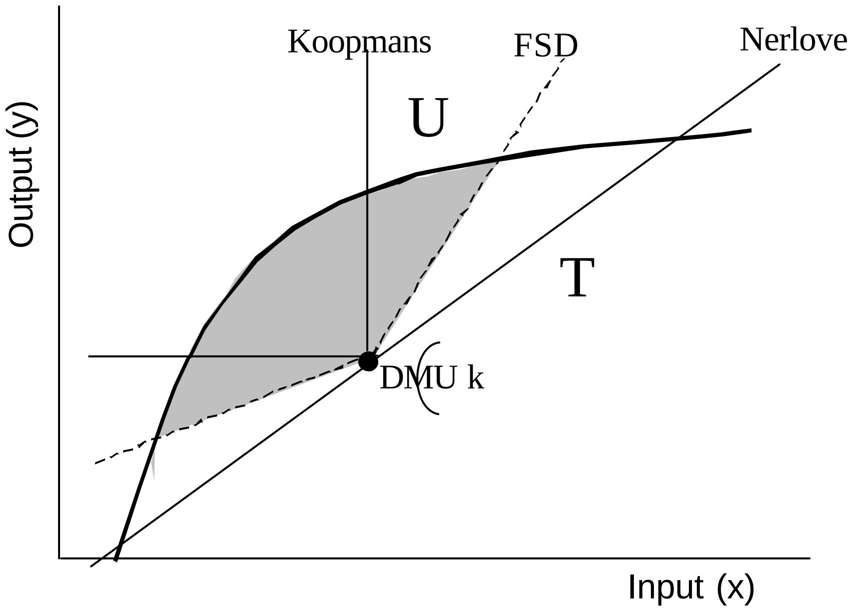


Figure 4: Koopmans technical efficiency and Nerlovian profit efficiency as limiting cases of the necessary condition for FSD efficiency.

4. Empirical tests

Resorting to the above FSD conditions can reduce the information requirement for the preference structure and the price distribution. However, information on the production set is still required. Unfortunately, one typically faces imperfect information on the production possibilities in the empirical analysis of production data. However, empirical tests can be obtained approximating the true production set by particular empirical production set constructed from empirical data.

Various approximations can be employed, depending on the assumptions imposed on the production set and the data used for approximating it. In Banker, Charnes and Cooper (1984) three key assumptions were imposed. Firstly, the production set was assumed to contain all observations, i.e.:

(21) $(X, Y) \in T.$

Furthermore, the production set was assumed to satisfy free disposability, i.e.:

(22) $(y_1, x_1) \in T \Rightarrow (y_2, x_2) \in T \;\forall (y_2, x_2) \in \Re_+^{s+m} : y_2 \leq y_1; x_2 \geq x_1.$

Finally, the production set was assumed convex, i.e.:

(23) $(Y,X) \in T \Rightarrow (l^T Y, l^T X) \in T \;\; \forall l : e^T l = 1, l \in \Re_+^n$ **.**

Banker *et al.* (1984) used the *convex hull* of observations as an empirical production set. The convex hull satisfies the so-called *minimum extrapolation* principle, being the smallest set in $\Re_+^{q+m}$ consistent with the above assumptions. More formally, the following empirical production set was used to approximate the true production set:

(24) $$\hat{T}_{BCC} = \left\{ (y,x) \in \Re_+^{q+m} \left| \begin{pmatrix} y \\ -x \end{pmatrix} \leq \begin{pmatrix} l^T Y \\ -l^T X \end{pmatrix} ; l^T e = 1; l \in \Re_+^n \right. \right\} \; .$$

This empirical production set has become the standard approximation for T in the nonparametric literature. For example, it was used by Banker *et al.* (1984) and Charnes *et al.* (1985), among others, for testing and measuring Koopmans' technical efficiency. In addition, Varian (1984), Banker and Maindiratta (1988), and Färe and Grosskopf (1997), among others, used it for testing and measuring Nerlovian profit efficiency.

An attractive feature of the convex hull is that it can be used for assessing profit efficiency even if the true production set does not satisfy free disposability and convexity. If the true production set violates free disposability or convexity, the convex hull can contain input-output vectors outside the production set. However, profit is concave and monotonic in output and input. Therefore, the data set necessarily contains an observation that produces at least as much profit as an arbitrary vector from the convex hull, i.e.:

(25) $$\exists j \in S : y_j p_k - x_j w_k \geq y p_k - x w_k \;\; \forall (y,x) \in \Re_+^{q+m}; l \in \Re_+^{p+m} :$$
$$\begin{pmatrix} y \\ -x \end{pmatrix} \leq \begin{pmatrix} l^T Y \\ -l^T X \end{pmatrix} ; l^T e = 1 \, .$$

Therefore, the convex hull implicitly represents feasible profit levels. Consequently, the production set may harmlessly violate free disposability and convexity, provided all observations are contained within the same production set (21).

However, under uncertainty, the convex hull can approximate the production set only if the latter truly is convex. This is because (25) does not imply that the convex hull contains an observation that produces at least as much profit as an arbitrary vector at all prices from the price domain. Therefore, the goodness of the convex hull depends on whether the production set truly is convex or not. Convexity requires the production technology to exhibit particular theoretical characteristics such as non-increasing marginal products and non-increasing marginal rates of substitution and transformation (Madden, 1986). From empirical point of view, little evidence on these theoretical characteristics has been presented (Brue, 1993). In case of violations, an alternative empirical production set is required. A number of alternative sets that do not require convexity have been proposed in the DEA literature, including Deprins, Simar, and Tulkens (1984), Petersen (1990), and Bogetoft (1996).

Figure 5 illustrates this point graphically. The shaded area represents the intersection of the convex hull of the observations and the polyhedral cone U. This intersection is nonempty, so the necessary condition (16) implies DMU k has to be inefficient. However, the intersection of U and the true, non-convex production set T is empty. Thus, DMU k can actually maximise expected utility. Therefore, if T is non-convex, the convex hull approximation may lead to erroneous conclusions.

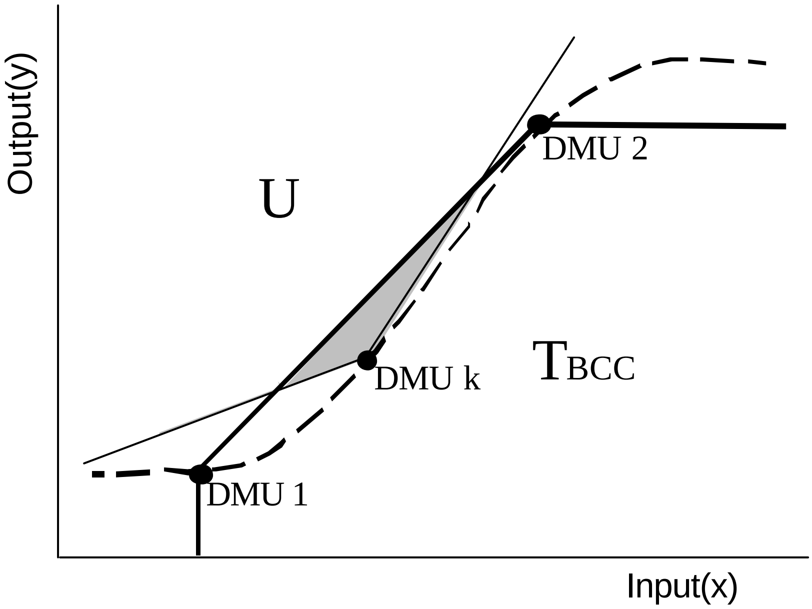


Figure 5: The shaded area represents the intersection of the convex hull of observations and the polyhedral cone U. This intersection is non-empty, falsely implying inefficiency of DMU k.

In contrast to convexity, free disposability can be imposed without harm. This is because the data set necessarily contains an observation that yields at least as much profit, at all possible prices, as an arbitrary feasible input-output vector that produces less output and uses more input than the observation, i.e.:

(26) $$\exists j \in S : y_j p - x_j w \geq yp - xw \, \forall (p,w) \in D; (y,x) \in T : \begin{pmatrix} y \\ -x \end{pmatrix} \leq \begin{pmatrix} y_j \\ -x_j \end{pmatrix}.$$

The smallest subset of $\Re_+^{q+m}$ consistent with the assumptions of containment of observations (21) and free disposability (22) is the so-called Free Disposable Hull (henceforth FDH) set by Deprins *et al.* (1984) (see also Tulkens, 1993). More formally, the FDH set can be represented by:

(27) $$\hat{T}_{FDH} = \left\{ (y,x) \in \Re_+^{q+m} \middle| \begin{pmatrix} y \\ -x \end{pmatrix} \leq \begin{pmatrix} l^T Y \\ -l^T X \end{pmatrix}; l^T e = 1; l_j \in \{0,1\}; j \in S \right\} .$$

Substituting that set for the true production set in the necessary FSD condition (16) gives the following empirical condition:

$$\textbf{(28)} \quad \max_{y,x,s',s}\left\{e(s' \quad s)e \,\middle|\, (y-s', x+s)\in U; (y,x)\in \hat{T}_{FDH}; (s',s)\in\Re_{+}^{q+m}\right\} = 0 \Leftrightarrow$$

$$\max_{l,s',s,z}\left\{e(s' \quad s)e \,\middle|\, \begin{pmatrix} l^T Y - s' \\ -l^T X - s \end{pmatrix} \geq \begin{pmatrix} y_k + zA \\ -x_k + zB \end{pmatrix}; l^T e = 1; l_j \in \{0,1\}\ j\in S; (s',s)\in \Re_{+}^{q+m}; z\in \Re_{+}^{l}\right\} = 0$$

If the observations are contained within the production set, this measure gives a necessary condition for optimal production under uncertainty.

Similarly, substituting the FDH set for the true production set in the sufficient FSD condition (15) gives the following empirical condition:

$$\textbf{(29)} \quad \max_{y,x,p,w}\left\{(yp - xw) - (y_k p - x_k w) \,\middle|\, (p,w)\in D; (y,x)\in \hat{T}_{FDH}\right\} = 0 \Leftrightarrow$$

$$\max_{l,s',s,z}\left\{(l^T Yp - l^T Xw) - (y_k p - x_k w) \,\middle|\, Ap - Bw \geq 0; p,w \geq e; l^T e = 1; l_j \in \{0,1\}\ j\in S\right\} = 0 .$$

Since the FDH set typically is smaller than the true production set, this condition does not give a sufficient condition for optimal production. However, it does give a sufficient condition for dominance over all observations in the sample. Solving a Mixed Integer-Linear Programming problem for each DMU can check these conditions.

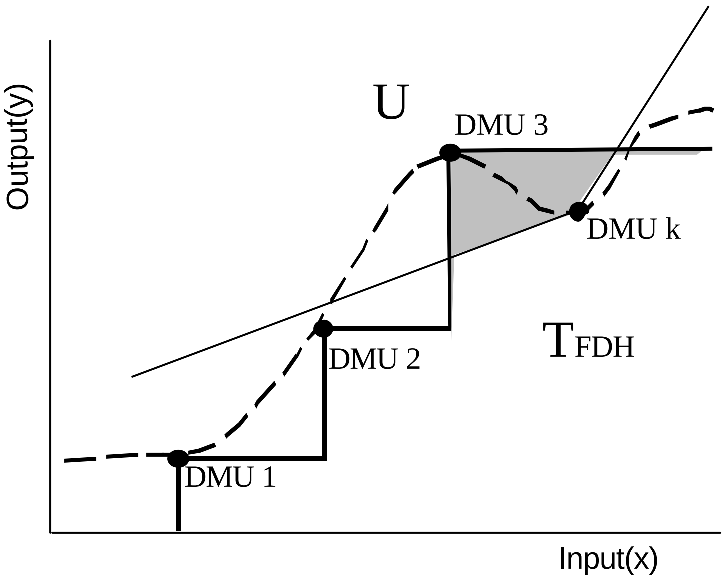


Figure 6: The shaded area represents the intersection of the FDH production set and the polyhedral cone U. Although DMU k produces on the frontier of the true technology, it does not maximize expected utility as there exist observed points that yield higher profit at all possible prices.

Figure 6 illustrates the application of the FDH set to test the empirical FSD condition (28). The shaded area represents the intersection of the FDH production set and the polyhedral cone *U*. Although DMU *k* produces on the frontier of the true technology, and hence is Koopmans efficient, it does not maximize expected utility as there exist observed points that yield higher profit at all possible prices

5. Concluding remarks

We have presented an approach based on first-order stochastic dominance that extends the nonparametric efficiency analysis to deal with uncertainty related to input-output prices. We have derived necessary and sufficient first-order stochastic dominance conditions that production vectors must satisfy in order to maximise expected utility. Interestingly, the necessary FSD efficiency condition contains the well-established conditions for Koopmans technical efficiency and Nerlovian profit efficiency as its limiting cases.

In addition, we have discussed empirical conditions that can test whether observed behaviour is consistent with expected utility maximisation. These conditions can be checked using standard mathematical programming techniques. A complication for testing FSD empirically is that convexity for the production set cannot be assumed without harm. Therefore, if violations of convexity are anticipated, the FDH set is preferred to the convex hull, which is typically used for testing profit efficiency conditions. From a methodological point of view, the approach adopted in this paper provides a powerful economic justification for the FDH approximation of technology set. We believe a substantial proportion of the appeal related to the convex monotone hull is due to the well-known economic underpinnings, i.e. the property (25). One important implication of this paper is that the convenient microeconomic properties of the convex monotone hull can be seriously inflated in a case of price uncertainty.

Nevertheless, it should be understood that the FSD approach is not limited to particular empirical production sets. Alternative empirical approximations or estimations could be used, depending on the purposes of the application and the available information on the production relationships.

In this paper, we assumed that input-output quantities are controllable, and we only considered uncertainty related to the input-output prices. However, in some cases, including Bertrand competition, it is necessary to include uncertainty for the quantities as well. Therefore, the inclusion of uncertainty for the input-output variables is an interesting route for future research. Furthermore, we only discussed testing efficiency of individual DMUs, given a specific theoretical or empirical production set. However, the problem can be easily turned around to test properties of the production set, given the set of efficient DMUs, as in Varian (1984).

To preserve the applicability related to the established nonparametric tools, we deliberately avoided imposing assumptions that could be considered restrictive in any sense. Resorting to first order stochastic dominance allowed us to effectively reduce the information requirement on the utility function and price distributions. However, in principle nothing prevents one to apply higher-order stochastic dominance rules. Higher order criteria involve more discriminating power than lower order ones, because they induce a larger reduction of the set of not-dominated choice alternatives. Nevertheless, that power has to be balanced against the stringency of the preference imposed assumptions (e.g. risk aversion for second-order stochastic dominance and decreasing absolute risk aversion for third-order stochastic dominance). Still, extending the FSD approach towards higher-order criteria is another interesting challenge for future research.

References

Allen, R., A. D. Athanassopoulos, R.G. Dyson and E. Thanassoulis (1997): Weight Restrictions and Value Judgements in DEA: Evolution, Development and Future Directions. *Annals of Operations Research* 73, 13-34.

Banker, R.D., Charnes, A., and W. W. Cooper (1984): Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis, *Management Science* 30, No. 9, 1078 - 1092

Banker, R.D. and A. Maindiratta (1988): Nonparametric Analysis of Technical and Allocative Efficiencies in Production, *Econometrica* 56 (6), 1315-1332.

Bertrand, J. (1883): Théorie Mathématique de la Richesse Sociale, *Journal des Savants*, 499 - 508

Bogetoft, P. (1996): DEA on Relaxed Convexity Assumptions, *Management Science* 42, 457-465.

Brue, S.L. (1993): Retrospectives: The Law of Diminishing Returns, *Journal of Economic Perspectives*, 7 (3), 185-192.

Charnes, A., W. Cooper, Z. M. Huang and D. B. Sun (1990): Polyhedral Cone-ratio DEA Models with an Illustrative Application to Large Commercial Banks, *Journal of Econometrics* 46, 73-91.

Charnes, A., W. Cooper, and E. Rhodes (1978): Measuring the Efficiency of Decision Making Units, *European Journal of Operational Research* 2, 429 - 444

Charnes, A., W. Cooper, B. Golany, L. Seiford, J. Stutz (1985): Foundations of Data Envelopment Analysis for Pareto-Koopmans Efficient Empirical Production Functions, *Journal of Econometrics*, 30, No. 1, 91 – 107

Cournot, A. (1838): *Recherches sur les Principes Mathématiques de la Théorie des Richesses*. English edition: *Researches into the Mathematical Principles of the Theory of Wealth*, Edited by N. Bacon, New York, Macmillan

Debreu, G. (1951): The Coefficient of the Resource Utilization, *Econometrica* 19 (3), 273 - 292

Deprins, D., D. Simar, H. Tulkens (1984): Measuring Labor Efficiency in Post Offices, p. 243 - 267 in M. Marchand, P. Pestieau and H. Tulkens (Eds.),, *The Performance of Public Enterprises: Concepts and Measurements*, North Holland

Farrell, M. J. (1957): The Measurement of Productive Efficiency, *Journal of the Royal Statistical Society* Series A 120 (3), 253 - 281

Färe, R., and S. Grosskopf (1997): Profit efficiency, Farrell decompositions and the Mahler inequality, *Economics Letters* 57, 283-287

Hadar, J. and W. R. Russell (1969): Rules for Ordering Uncertain Prospects, *American Economic Review* 59, 25-34.

Levy, H. (1992): Stochastic Dominance and Expected Utility: Survey and Analysis, *Management Science* 38 (4), 555-593.

Koopmans, T. (1951): *Activity Analysis of Production and Allocation*, John Wiley & Sons Inc., New York

Madden, P. (1986) *'Concavity and Optimisation in Microeconomics'*, ISBN 0-631-14192-8.

Neumann, J. Von, and O. Morgenstern (1967): *Theory of games and economic behaviour*, Wiley, New York.

Nerlove, M. (1965): *Estimation and Identification of Cobb-Douglas Production Functions*, Rand McNally, Chicago, Illinois.

Petersen, N. C. (1990): Data Envelopment Analysis on a Relaxed Set of Assumptions. *Management Science*, vol. 36 (3), 305-314.

Qin, C.-Z., and Stuart, C. (1997): Bertrand versus Cournot revisited, *Economic Theory*, vol. 10 (3), 497-507

Tulkens, H. (1993): On FDH Efficiency Analysis: Some Methodological Issues and Application to Retail Banking, Courts, and Urban Transit, *The Journal of Productivity Analysis*, 4, No. 1, 183-210

Varian, H (1984): The Nonparametric Approach to Production Analysis, *Econometrica* 52 (3), 579-597.

Appendix 1: Errata

We have identified the following errors in the original articles.

Article II:

Some typing errors are noted:

On p. 53, Section 2, 1st paragraph. "The second difference is that Färe and Svensson phrased their definition in terms of (single-input) production functions…". This should read [(single-output) production functions].

p. 55, Section 4, 3rd paragraph. "It therefore seems more meaningful to compare alternative DEA models on the level of the assumptions…" [at the level of the assumptions].

Article III:

P. 60, Section 2, 1st paragraph. "Note that this congestion adjustment is harmless as far as the usual (Pareto-Koopmans or Debreu-Farrell) notions of technical efficiency are concerned." This statement is true as for the Pareto-Koopmans notion of technical efficiency, but not so in the case of Debreu-Farrell notion. See Article II for further details.

P. 63, References: The paper by Diewert was published in 1982 as stated in the text, not in 1978 as claimed in the list of references.

Article IV:

Firstly, an annoying technical error. We are grateful to Professor Niels-Christian Petersen for pointing out this technical error in his pre-examination report of this dissertation.

From Article IV (p. 71):

> "The price domain cone W involves two sets that are interesting from an economic perspective, and that will prove helpful for approximating economic efficiency. […] The second set contains all input vectors associated with a nonnegative cost at some prices in *W*:
>
> (3.3) $W^V = \left\{ x \in \Re^q \,\middle|\, \exists w \in W : xw \geq 0 \right\}$."

In Article IV we assumed that this set is equivalent to the following set:

$$\tilde{W}^V = \left\{ x \in \Re^q \,\middle|\, x \geq a\, A_i ; i \in L ; a \geq 0 \right\} ,$$

and proceeded to derive Proof 3.2 and empirical estimator (4.7), formulated in terms of this set. Unfortunately, the two sets are not equivalent. Example: Consider a case with two inputs and suppose A=0, which implies that $\tilde{W}^V = \Re^2_+$. Consider the input vector $x = (-1 \quad 1)$. Clearly, x does not belong to the set $\tilde{W}^V = \Re^q_+$. Yet, $x \in W^V$ since there exist feasible prices at which the cost is positive, e.g. $x = (0 \quad 1)$. It follows that Proof 3.2 (associated with Theorem 3.2 which still is correct) and the empirical estimator (4.7) are incorrect.

To correct this problem, we can introduce the following alternative set:

$$W^V = \left\{ x \in \Re^q \left| x_i \geq \boldsymbol{a} A_i \quad \forall \boldsymbol{a} \in \Re^l_-; i \in L \right. \right\}.$$

In Kuosmanen and Post (2001) we present a revised proof of the Lower Bound Theorem 3.2., which implies that this set is equivalent to (3.3).

To facilitate empirical application, we proposed an enumerative Linear Programming procedure for computing the empirical estimator $\underline{CE}^{L_{FDH}}(x, y; W)$ relative to the Free Disposable Hull (FDH) input set $\hat{L}_{FDH}(y) = \left\{ x \in \Re^q_+ \left| y \leq y_j; x \geq x_j; j \in S \right. \right\}$. We next briefly outline a revised algorithm to correct the error identified above.

To characterize that reference technology, the original article first identified (by enumeration) the subset of observed production plans that yield output vectors greater than or equal to y, i.e. $c(y) = \left\{ x_j \in X \left| y_j \geq y \right. \right\}$. (For notational convenience, in Article IV $c(y)$ was used interchangeably for this discrete set of input vectors, and the matrix consisting of these vectors.) Since the set $c(y)$ necessarily includes all the extreme points of $\hat{L}_{FDH}(y)$, the lower bound estimator $\underline{CE}^{L_{FDH}}(x, y; W)$ can equivalently be written in terms of this set as

$$\underline{CE}^{L_{FDH}}(x, y; W) = \min_{w \in W} \left(\min_{x_j \in c(y)} \left\{ x_j w \left| xw = 1 \right. \right\} \right).$$

We can next safely reserve the order of the two minimization problems:

$$\underline{CE}^{L_{FDH}}(x, y; W) = \min_{x_j \in c(y)} \left(\min_{w \in W} \left\{ x_j w \left| xw = 1 \right. \right\} \right).$$

Thus, the lower bound estimator can be easily computed by the following simple algorithm: 1) Identify the input set $c(y)$ by enumeration. 2) Solve the embedded LP problems $\min_{w \in W} \left\{ x_j w \left| xw = 1 \right. \right\}$ for each $x_j \in c(y)$. 3) Select the minimum of the optimal solutions to those problems.

Note: In large-scale applications, the computation burden can be reduced by using additional information (e.g. from technical efficiency classifications, or cost efficiency upper bounds) to decrease the size of $c(y)$ by eliminating all 'inefficient' observations that cannot solve the subsequent minimization problem in Step 3.

To conclude, we happily noticed that actually the correct algorithm outlined above was originally used for computing the lower bound estimates in the empirical application presented in Article IV. The error with the input set was introduced in one of the subsequent revisions.

Finally, there is an error in the references concerning the forthcoming papers Kuosmanen (2000) and Post (2000). Both are scheduled for publication in 2001.

Article V:

p. 101, Section 5, 4th paragraph: "In this paper, we assumed that input-output quantities are controllable, and we only considered uncertainty related to the input-output prices. However, in some cases, including Bertrand competition, it is necessary to include uncertainty for the quantities as well." This remark on Bertrand competition takes a shortcut, which may appear confusing. Actually, the case of Bertrand competition does not directly relate to the quantity uncertainty at all. It merely points to the fact that in some situations it may be more appropriate to model the prices (rather than quantities) as decision variables, while the demanded/supplied quantities would be determined by the market forces. In such a setting, it might be reasonable to associate the uncertainty to the quantities.

References

Kuosmanen, T., and G.T. Post (2001): Notes on: Measuring Economic Efficiency with Incomplete Price Information, an unpublished manuscript

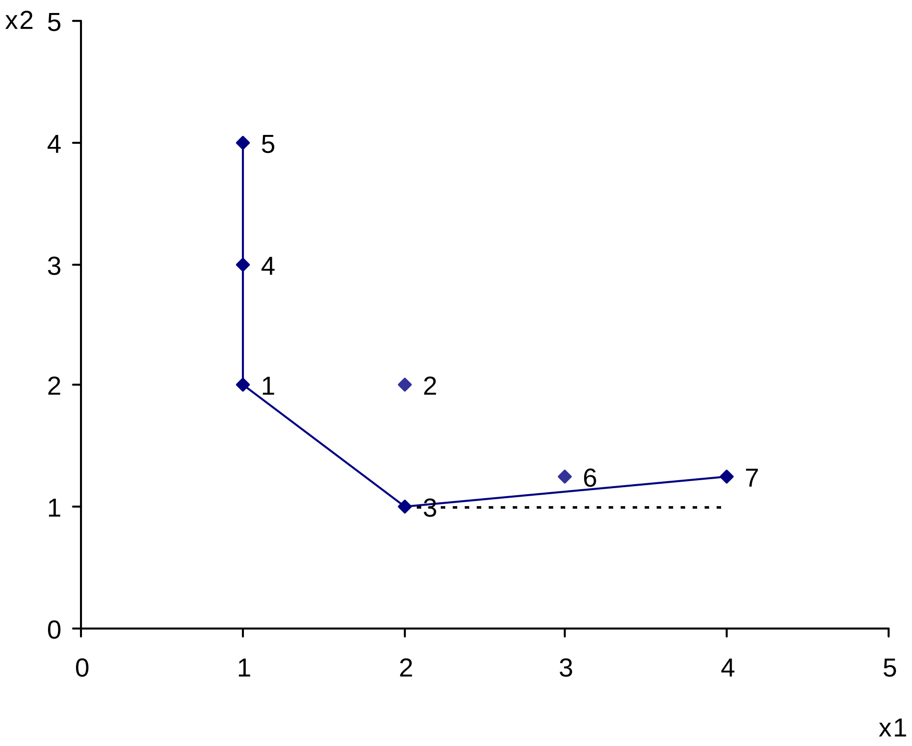
Starmer, C. (2000): Developments in Non-Expected Utility Theory: The Hunt for a Descriptive Theory of Choice under Risk, *Journal of Economic Literature* 37, 332-382

Appendix 2: Examples

Details of the two examples discussed in Article II.

Example 1:

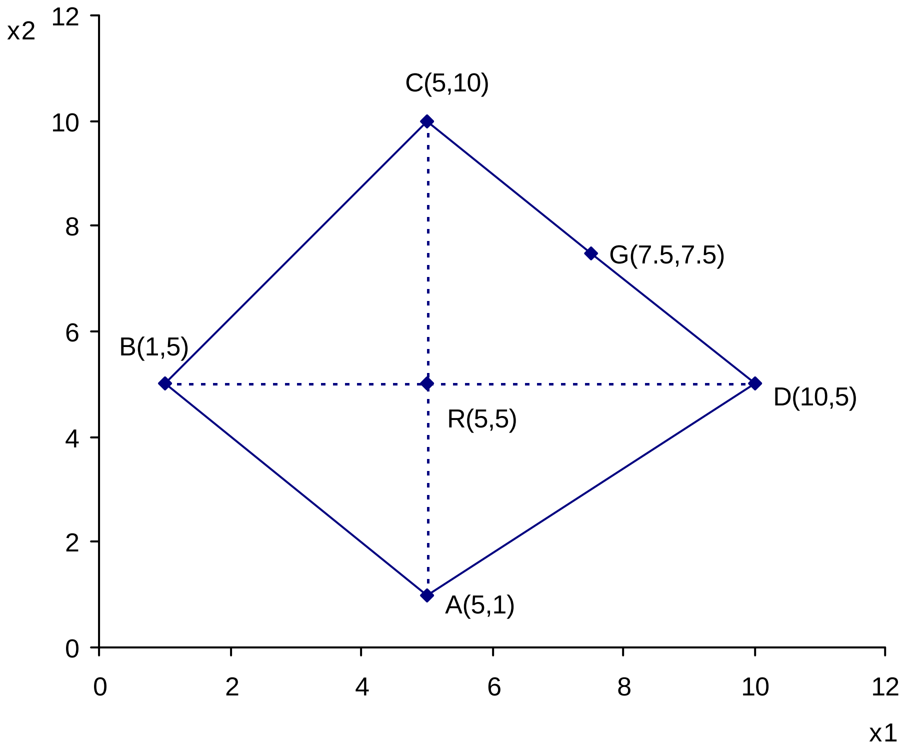
DMU	Output y	Input x_1	Input x_2
1	2	1	2
2	2	2	2
3	2	2	1
4	2	1	3
5	2	1	4
6	2	3	1.25
7	2	4	1.25



Source: Färe, R., S. Grosskopf, and C.A.K. Lovell (1985): *The Measurement of Efficiency of Production*, Kluwer Academic Publishers, Boston

Example 2:

DMU	Output y	Input x_1	Input x_2
A	1	5	1
B	1	1	5
C	1	5	10
D	1	10	5
G	1	7.5	7.5
R	10	5	5



Source: Brockett, P.L., W.W. Cooper, H.C. Shin, and Y. Wang (1998): Inefficiency and Congestion in Chinese Production Before and After the 1978 Economic Reforms, *Socio-Economic Planning Sciences* 32, 1-20