ESSAYS IN PRODUCTIVITY AND EFFICIENCY ANALYSIS IN THE PRESENCE OF UNDESIRABLE OUTPUTS

by

Mykhaylo Salnykov
MA in Economics, Economics Education and Research Consortium, Kyiv, Ukraine, 2004
MSc in Environmental Sciences and Policy, Central European University, Budapest, Hungary, 2002
BSc in Ecology, National University of 'Kyiv-Mohyla Academy', Kyiv, Ukraine, 2001

THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

In the
Department of Economics

SIMON FRASER UNIVERSITY

Spring, 2008

APPROVAL

Name: **Mykhaylo Salnykov**

Degree: **Doctor of Philosophy**

Title of Thesis: **Essays in Productivity and Efficiency Analysis in the Presence of Undesirable Outputs**

Examining Committee:

Chair: David Andolfatto
Professor, Department of Economics

Krishna Pendakur
Senior Supervisor
Associate Professor, Department of Economics

Ken Kasa
Supervisor
Professor, Department of Economics

Nancy Olewiler
Supervisor
Professor, Department of Economics

Brian Krauth
Internal Examiner
Associate Professor, Department of Economics

Timo Kuosmanen
External Examiner
Professor, Economic Research Unit
MTT Agrifood Research Finland

Date Defended/Approved: April 1, 2008

Declaration of Partial Copyright Licence

ABSTRACTS

Essay 1 *On the commensurability of directional distance functions*

In the paper we study commensurability property of the directional distance function (DDF). We find that a popular DDF with a fixed directional vector is neither absolute nor ranking commensurable. Nevertheless, this function can be commensurated if the directional vector is commensurated along with the data. We identify a necessary and sufficient condition for a vector that ensures commensurability of the DDF, which helps somewhat narrowing down the key issue for this function in practice—the choice of direction of measurement.

Keywords production theory; directional distance functions; commensurability; efficiency measurement

Subject Terms Production (Economic theory); Industrial efficiency – Measurement; Input-output analysis

JEL Classification Numbers D2, D24

Essay 2 *Smooth homogeneous bootstrap bias correction in frontier models: a Monte Carlo assessment when some outputs are undesirable*

In this paper I test the performance of smooth homogeneous bootstrap bias-correction in multi-output frontier models using hyperbolic efficiency function. I propose an approximation procedure that substantially reduces the nonparametric estimation time while sacrificing little precision compared to the most precise nonparametric alternative. The performance of the uncorrected and bias-corrected estimates is tested in samples of different sizes via Monte Carlo simulation. All techniques perform well in large samples even without correction. Both parametric and nonparametric estimators benefit from the correction regardless of the sample size. Uncorrected nonparametric estimators perform well in large and require bias-correction in smaller samples. In the small samples bias correction shows marginally better results when applied to the parametric estimator.

Keywords production theory; hyperbolic efficiency function; efficiency measurement; smooth homogeneous bootstrap; Monte Carlo simulation

Subject Terms Production (Economic theory); Industrial efficiency – Measurement; Input-output analysis; Bootstrap (Statistics); Monte Carlo method

JEL Classification Numbers C15, D2, D24

Essay 3 *Does emission permit trade hamper development? Capital and output dynamics under an international transferable emission quota trade system*

The paper studies the effects of nontradeable emission quota and transferable emission quota systems on the accumulation of capital and output growth in small open economies. Both types of regulation impede the growth. The transferable emission quota system has different effects on the development of quota buyers and quota sellers. While quota buyers enjoy faster growth in the both capital stock and output as compared to the nontransferable quota system, quota sellers face slower capital accumulation and economic growth.

The simulation using advances in frontier modelling confirms the theoretical findings and reveals that developmental consequences for quota sellers range from a slower capital accumulation to capital stock shrinkage. It also suggests that quota sellers substitute economic production for quota revenues and economic output falls over time.

Keywords production theory; input distance function; emission trade; bootstrap application

Subject Terms Environmental policy – Economic aspects; Emissions trading; Production (Economic theory); Input-output analysis; Bootstrap (Statistics)

JEL Classification Numbers C15, D2, D24, Q25, Q28, Q56

EXECUTIVE SUMMARY

In this dissertation I focus on various issues in productivity and efficiency analysis when the technology produces some undesirable outputs.

In the first essay I study commensurability (independence of units of measurement) property of the directional distance functions -- a common instrument that allows to treat outputs asymmetrically. I discover that directional distance functions with fixed directional vectors, in general, are not commensurable. I then identify necessary and sufficient condition the function must satisfy to be commensurable.

The second essay is a Monte-Carlo study of the smooth homogeneous bootstrap bias correction technique performance when applied to parametric and nonparametric estimators of the hyperbolic efficiency function. Not surprisingly, in the large samples the bias of the frontier estimates is small regardless of the estimator used. In small samples bootstrap allows fixing some bias and performs comparably well when applied to various estimators. I further propose a linear approximation technique for hyperbolic efficiency estimation that performs better than previously used linear approximation based on Taylor expansion.

The final essay develops a theoretical model of the international emission trading scheme similar to Kyoto protocol to study capital and output dynamics of the economies under such an agreement. Advancements in productivity and efficiency analysis are used to simulate the model and assess the magnitude of the effects. I discover that emission buyers will develop faster than under simple cap systems, while emission sellers will suffer slower (or even negative) growth in terms of capital and economic output.

Єдиному

to the One

ACKNOWLEDGEMENTS

I want to voice my wholeheartedly gratitude to

- My PhD supervisory committee members for their valuable comments and support and especially to my senior supervisor Dr. Krishna Pendakur for being extremely helpful and supportive to my studies, research and career.
- Dr. Valentin Zelenyuk for introducing me to the field of productivity and efficiency analysis and Dr. Rolf Färe for introducing Valentin to this field before.
- My external examiner Dr. Timo Kuosmanen and my internal examiner Dr. Brian Krauth for their valuable comments and suggestions and for tolerating my bad English.
- Those at SFU and other places who read through various portions of this work and provided their feedback. My special thanks to Dr. David Andolfatto, Dr. Olena Hankivsky, Dr. David Jacks, Dr. Peter Kennedy, Dr. Nicolas Schmitt, Scott Skjei (all SFU), Dr. Robin Sickles (Rice U), Dr. Leopold Simar (UCL), Dr. Paul Wilson (Clemson U), Dr. Jose Zofio (U Autonoma de Madrid) and anonymous referees.
- I owe gratitude to the staff at the SFU Department of Economics, especially Tim Corram, Lorraine Sam, Tracey Sherwood and Gwen Wild who provided me with necessary technical assistance during my studies. My special thanks to the Graduate Secretary Laura Nielson: if it was not her, I would not graduate as fast and as painlessly
- My thanks to my friends who cheered me up and prayed for me: Andriy Baranskyy, Hillary and Joan Carre, Olena Chekmezova, Dr. Aleh Cherp, Pavlo and Natalya Demchuk, Ian, Ruth, Emily, Mackenzie and Derek Fenton, Yvette and Ross Gorsline, James, Sue and Matt Groening, Galyna Gulenko, Chritopher Henderson, Sofiya Huzenko, Becky Jean, Olga Kesearchuk, Maryana Kolinchak, Vira Kovalyova, Tetyana and Iryna Kuchma, Vera Lantinova, Lesya Lazhevska, Olena Maslyukivska, Bryan McMaster, Vladlena Martsynkevych, Cristin and Colleen McCambridge, Volodymyr Nesterenko, Olena Podolyan, Lance, Kathy, Amy and Bethany Rogers, Yuriy Spivak, Edward and Banani Stanek, Inna Viryasova, Anya Vorobyova, Don and Heather Warren, Ivan Yuryk, Natalya Zelenyuk and Vitaliy Zheka.
- Special thanks are owed to Andrew and Esther Scott and to their sons Mark, Matt and Steven for their endless love and continuous support. I just cannot thank you enough
- From the bottom of my heart I thank my mother Olena Demkovych, my sister Oleksandra and my grandparents Ivan and Antonina Salnykov for many years of support and unconditional acceptance. I also thank my mother Nina Marchenko, my father Serhiy Biriukov and my sister Polina for giving me a gift of their daughter and sister.
- I owe an enormous debt of gratitude to my best friend and wife Nastia Salnykova for living through these challenging but interesting years with me. Thank you for pushing me forward and loving me.
- No litany of thanks would be complete without my thanking The Great Lord, Father of our Lord Jesus Christ, and source of all grace, wisdom, and compassion, Who has given me peace in my heart now and new life in the future.

TABLE OF CONTENTS

LIST OF FIGURES

Essay 1

Essay 2

Essay 3

LIST OF TABLES

Essay 1

Essay 2

Essay 3

Overview

The concept of technology in general and production functions specifically is one of the cornerstones of contemporary microeconomic theory. The traditional microeconomic paradigm is based on the premise that all decision making units are producing the maximum possible output or are collecting the highest possible profit given the scarce input. Yet, even the biggest adherents of the rationality paradigm admit there are cases in the real world where some businesses exploit more lucrative market niches than the others or some financial institutions yield higher returns on their investments. That is: in reality some firms produce more with less.

While the paradigm taken by the young and underexplored field of productivity and efficiency analysis (PEA) may be seen somewhat unusual to mainstream economists, the notions of best practice and benchmarking are common in the business world despite seemingly contradicting the full-rationality approach taken by most economic theory.

PEA takes as a given that there is a best-practice frontier, but that not all units attain it. That is, some agents are, on the face of it, inefficient.

In classical economic theory, one can find a number of examples of failing to reach the frontier, which still can be explained from the rationality premises. Let me provide a short illustrative example.

Suppose a researcher observes a number of players playing a *Prisoners Dilemma* (PD) game. She will most likely observe most of the players getting identical payoffs in the game's Nash equilibrium. Yet, she may also occasionally observe some people collaborating and getting higher payoffs, especially if the pair is playing multi-period PD. Moreover, she may also observe some players defecting while others collaborate resulting in very high payoffs for defectors and loss for collaborators[1]. If the

[1] Interestingly, my experience shows that this outcome is very common in the student environment: I offered students to play Prisoners Dilemma in 6 separate classes I taught. Only twice the Nash equilibrium was a result. Four times one person collaborated and the other defected.

researcher observes the payoffs without being informed about the environment, she may question the rationality of many players in these games. Indeed, given equal resources, some players are more efficient (those defecting while the other collaborates) and some are less efficient. Notably, the rational Nash equilibrium will appear as an inefficient outcome in the researcher's dataset. The reason for this inefficiency is that in this environment both parties cannot achieve the maximum outcome in one game. Yet, this inefficiency is easy to reconcile with the rationality if one *knows* about the PD setup.

It is often suggested that inefficiency is observed because some parameters of the environment are unobserved. Indeed, observed inefficiency in the example above is a result of the researcher's ignorance of the institutional design of the problem.

In many cases difference in observed efficiency is a result of the unobserved managerial abilities of the firms' management. It has been also suggested that inefficiency may be a strategic decision of the incumbent firms to signal low profitability of the market and defer entry of other firms. Inefficiency may also result from rational decisions of the risk averse agents in uncertain environments when one prefers a lower certain payoff to a higher (in expected terms), but less certain one.

One thing is sure though: whenever one observes heterogeneity in the economic performance of different businesses, there is an underlying reason for it.

The literature often distinguishes between two types of inefficiency: technical and allocative. A firm is considered technically efficient if it produces the greatest technically feasible outputs given its inputs. Alternatively, one may view technical efficiency from an input perspective: using the lowest technically feasible input level to obtain given outputs.

A firm is considered allocatively efficient if it produces "the right mix" of outputs or uses "the right mix" of inputs given the prices. Obviously, measuring allocative efficiency requires input and output markets to exist. This is not always the case, however. Many technologies produce so called non-market outputs. Very often these outputs are undesirable, such as air or water pollution in polluting technologies or bad debts in banking.

Non-existence of the markets is not the only problem when dealing with undesirable outputs. These outputs cannot be treated like desirable outputs. When technologies produce desirable outputs only, the studies often depart from a presumption that a firm becomes more efficient when all outputs are scaled up proportionally, while keeping inputs unchanged. By doing so, they rely on so called output distance function (ODF), which find the greatest technically feasible radial expansion of outputs. Obviously, proportional increase of all outputs does not necessarily mean increase of efficiency when one of the outputs is undesirable. A number of solutions have been proposed to deal with this problem.

The earliest studies propose to treat outputs asymmetrically by finding the greatest proportional contraction of the undesirable *and* the greatest proportional expansion of undesirable outputs. The unit moves to the best practice frontier along a hyperbolic curve, which gives the name to the efficiency measure – the hyperbolic efficiency function (HEF).

Later an alternative to HEF was proposed. Directional distance functions (DDFs) in general and output directional distance functions (ODDFs) specifically allow movement to the best practice frontier along an exogenously set directional vector. DDFs are often criticized for the lack of economic intuition as the choice of the directional vector is arbitrary.

An alternative to the HEF and ODDFs that did not receive much popularity in the literature is treating undesirable outputs similarly to inputs and contract inputs and undesirable outputs proportionally using a measure similar to input distance function (IDF).

This dissertation contains three essays dealing with each of the three approaches to technologies when some outputs are undesirable.

The first essay studies the independence of the DDFs to scaling data up or down. This property is often known as the *commensurability* property of the distance functions. In this essay I discover that DDFs with *fixed* directional vector are neither absolutely nor ranking commensurable (i.e., the ranking of the firms change when the data is transformed). I further find that to ensure commensurability the directional vector must transform automatically when the data changes. This allows to limit the scope of the directional vectors suitable for empirical studies.

The second paper studies performance of the smooth homogeneous bootstrap (SHB) bias-correction in the frontier models of technologies producing undesirable outputs. Efficiency estimates are inherently downward bias in most of the frontier models and require bias correction. Statistical properties of the efficiency estimates rarely have analytical formulation justifying use of nonparametric bootstrap to assess statistical inferences. Yet, traditional bootstrap ignores the fact that efficiency estimates are bounded from either above or below. To address this issue SHB was proposed for a use in nonparametric frontier models. Yet, SHB's performance in parametric frontier regression models was never tested. The second paper is a Monte Carlo study of the SHB bias correction performance in nonparametric and parametric models when the technology produces undesirable outputs. I use HEF as a measure of efficiency and propose an alternative nonparametric estimation procedure that is much less computationally demanding and does not sacrifice much of the preciseness as compared to the traditional nonlinear optimization technique.

In the third paper I use the estimated parametric Translog formulation of the IDF to asses the magnitude of the effect of an emission trading scheme on capital and output dynamics of participating countries. I first develop a theoretical model that shows that emission trading promotes faster development in permit buying countries and slower development in quota selling countries as compared to a simple cap system. I then use the parametric IDF estimate to assess the magnitude of this effect, which turns out to be quite substantial.

Mykhaylo Salnykov · Valentin Zelenyuk

ESSAY 1

ON THE COMMENSURABILITY OF DIRECTIONAL DISTANCE FUNCTIONS[1]

M. Salnykov
Department of Economics, Simon Fraser University
8888 University Drive, Burnaby, BC, Canada, V5A 1S6
E-mail: salnykov@sfu.ca
Tel.: +1 604 291 35 08; fax: +1 604 291 59 44

V. Zelenyuk
Kyiv School of Economics and Kyiv Economic Institute
Voloska 10, Office 406, 04070, Kiev, Ukraine;
E-mail: vzelenyuk@eerc.kiev.ua
Tel. +38 044 492 80 12; fax: +38 044 492 80 11

Abstract

Recently, a popular tool for characterizing and estimating production technology and efficiency, known as Shephard's distance functions, was generalized by Luenberger shortage function or the Chambers-Chung-Färe directional distance function. This function possesses many desirable properties; however, it was not known whether it satisfies the *commensurability* property (independence of units of measurement up to a scalar transformation). We address this question in our study and discover interesting results: both positive and negative. We find that a popular directional distance function with a fixed directional vector is not commensurable. Moreover, when the units of measurement change, decision making units may change their ranking if the fixed vector is used. Nevertheless, this function can be commensurated if one is willing to commensurate the directional vector along with the data. We identify a necessary and sufficient condition for a directional vector that ensures commensurability of the function. These results also help somewhat narrowing down the key issue for this function in practice—the choice of direction of measurement.

Keywords production theory; directional distance functions; commensurability; efficiency measurement

JEL Classification Numbers D2, D24

---

[1] An earlier version of this paper was submitted to the Journal of Productivity Analysis. The Journal has requested a revision. We would like to thank Rolf Färe, Peter Kennedy, Robin Sickles, Scott Skjei and anonymous referees for their valuable comments and suggestions on the earlier versions of this paper.

1. INTRODUCTION

Since Shephard (1953), the production theory in neo-classical economics, and especially its duality issues, has been dominated by what his followers called Shephard's distance functions (DFs). Recently, Luenberger (1992, 1994) and Chambers, Chung and Färe (1996, 1998) have introduced and explored what is now widely known as the directional distance function (DDF), or the shortage and benefit function in the terminology of Luenberger. Under quite weak regularity conditions, DDF gives a complete characterization of a technology set and is a generalization of the input and output oriented Shephard's DFs. In a recent paper Färe and Primont (2006) outlined a comprehensive system of duality relationships related to the DDF. Similarly to the traditional DFs, the DDF gives a complete characterization of regular technologies, yet has a somewhat more powerful duality relationship—not only to the revenue or cost functions, but also to the profit function. The DDF also satisfies many desirable properties (e.g., see Luenberger (1992, 1994) and Chambers, Chung and Färe (1996, 1998)). Yet, one very important for an economic measure property has not being investigated for the DDF—the *commensurability* property—introduced to the efficiency analysis by Russell (1987), on the analogy of property introduced by Eichhorn and Voeller (1976). Intuitively, commensurability of an efficiency measure is a property related to the issue of independence of this measure from units of measurement (usually, up to scalar transformation) of the data.

One of the major empirical critiques of the DDF is that it is often not clear what directional vector must be chosen for each empirical study. Indeed, researchers often argue whether input or output orientation must be chosen in a particular research involving Shephard's DFs. For the DDF, however, the choice is even more complicated—there is a continuum of possibilities. A natural way to reduce the set of the feasible directional vectors would be to postulate a list of desirable properties that the DDF must satisfy. Taking this route would be in the fashion of axiomatic approach to efficiency analysis (where DF and DDFs are extensively used) proposed by Färe and Lovell (1978) and elaborated by Bol (1986), Russell (1987, 1990) and others to justify the use of some measures and warn about using others for various cases. This question, however, has not been resolved yet for the DDF.

Our study contributes to the theory of DDF from two perspectives. First, we explore the DDF for the commensurability property and discover quite interesting results: some positive and some negative. Second, our results on commensurability of DDF help narrowing down the key empirical issue for this function—the choice of direction of measurement.

We show that DDFs with a fixed directional vector result in different efficiency scores when the dataset's units of measurement are changed. Moreover, the ranking of the decision making units may change when a fixed directional vector is used. Nevertheless, this shortcoming can be avoided if the directional vector changes when the data is commensurated. We also identify a necessary and sufficient condition for the directional vector that ensures that the DDF is commensurable.

The rest of the paper is organized as follows. We first postulate basic assumptions and definitions. Then we provide and prove sufficient and necessary condition for the absolute commensurability of the directional efficiency measure on an arbitrary technology set. Next we study, the absolute-commensurability and the ranking-commensurability properties for the DDF with the *unit* directional vector. To our surprise, we find that none of these properties is satisfied in the general context. We then also examine whether such DDF can be 'de-commensurated' both ex *post* and ex *ante*. Finally, we find a particular type of DDF that does possess commensurability property.

2. BASIC DEFINITIONS

Let $\mathbf{x} \in \Re_+^N$ denote a vector of inputs, while $\mathbf{y} \in \Re_+^M$ denote a vector of outputs and assume technology can be characterized by a *technology set* T, defined in general terms as

$$T \equiv \{(\mathbf{x}, \mathbf{y}) : \mathbf{x} \; can \; produce \; \mathbf{y}\}.$$

We assume T satisfies the standard regularity conditions of neo-classical production economics.

A1. T is closed and non-empty.

A2. Inputs and outputs are freely disposable:
$(\mathbf{x},\mathbf{y}) \in T \Rightarrow (\mathbf{x}',\mathbf{y}') \in T, \quad \forall \mathbf{x}' \geqq \mathbf{x},\ \mathbf{y}' \leqq \mathbf{y}$.

A3. There is no free lunch, *i.e.* $(\mathbf{0}_N, \mathbf{y}) \in T \Rightarrow \mathbf{y} = \mathbf{0}_M$.

A4. Doing nothing is possible, *i.e.* $(\mathbf{x}, \mathbf{0}_M) \in T,\ \forall \mathbf{x} \in \Re_+^N$.

A5. $P(\mathbf{x}) \equiv \{\mathbf{y} : (\mathbf{x},\mathbf{y}) \in T\}$ is bounded $\forall \mathbf{x} \in \Re_+^N$.

A6. Technology is productive, i.e. $\exists \mathbf{x} \in \Re_+^N : P(\mathbf{x}) \neq \{\mathbf{0}_M\}$.

Assume an arbitrary nonzero directional vector $(-\mathbf{d}_\mathbf{x}, \mathbf{d}_\mathbf{y}) \in \Re_-^N \times \Re_+^M$; $(-\mathbf{d}_\mathbf{x}, \mathbf{d}_\mathbf{y}) \neq \mathbf{0}_N \times \mathbf{0}_M$ (further we assume these conditions without explicitly stating them). Given the regularity conditions (A1-6) and the directional vector, the directional distance function (DDF) defined on a regular technology set T as

$$\vec{D}(\mathbf{x},\mathbf{y} \mid -\mathbf{d}_\mathbf{x}, \mathbf{d}_\mathbf{y}, T) \equiv \sup\{\theta \geq 0 : ((\mathbf{x} - \theta \mathbf{d}_\mathbf{x}), (\mathbf{y} + \theta \mathbf{d}_\mathbf{y})) \in T\}, \tag{1}$$

gives a complete characterization of technology set T. Luenberger (1992, 1994), Chambers, Chung and Färe (1996, 1998), Färe and Grosskopf (2000) and Färe and Primont (2006) derived other properties of the DDF. However, the *commensurability* property of the DDF—the issue we address next—has not been studied yet.

3. COMMENSURABILITY OF DISTANCE FUNCTIONS

When Russell (1987) introduced the *commensurability axiom*, he convincingly argued that commensurability (independence of units of measurement up to scalar transformation) is a very desirable property of any efficiency measure. Indeed, an efficiency measure not satisfying commensurability may cause different researchers using the same data and methodology to arrive at different results—just because one used, for example, kilograms and the other one used pounds to measure inputs or outputs. In the next definition, we will extend the definition of commensurability proposed by Russell, to cover the measures of the type of DDF.

Definition 1 *(Absolute Commensurability)*

Let $E(\mathbf{x}, \mathbf{y}|\mathbf{p}(\mathbf{x}, \mathbf{y}), T) \in \Re_+ \quad \forall \mathbf{x} \in \Re_+^N, \mathbf{y} \in \Re_+^M$ be an efficiency measure on a regular technology set *T*, where $\mathbf{p}$ is a $\Re_+^N \times \Re_+^M \to \Re_+^Z$ mapping resulting a Z-dimensional vector of exogenous parameters of the efficiency measure (e.g., directional vector coordinates in the case of DDF). Let $\tilde{\mathbf{x}} = \Omega_{\mathbf{x}}\mathbf{x}$ and $\tilde{\mathbf{y}} = \Omega_{\mathbf{y}}\mathbf{y}$, where $\Omega_{\mathbf{x}}$ and $\Omega_{\mathbf{y}}$ are (any) diagonal matrices (further called *commensuration* matrices) of dimensions *N*x*N* and *M*x*M*, respectively, with all diagonal elements being strictly positive constants. The efficiency measure $E(\mathbf{x}, \mathbf{y}|\mathbf{p}(\mathbf{x}, \mathbf{y}), T)$ is absolutely commensurable in inputs and outputs *if and only if*

$$E(\mathbf{x}, \mathbf{y}|\mathbf{p}(\mathbf{x}, \mathbf{y}), T) = E(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}|\mathbf{p}(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}), \tilde{T}) \quad \forall \mathbf{x}, \tilde{\mathbf{x}} \in \Re_+^N; \mathbf{y}, \tilde{\mathbf{y}} \in \Re_+^M \quad \forall T$$

where

$$\tilde{T} \equiv \{(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}) : (\mathbf{x}, \mathbf{y}) \in T\} = \{(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}) : (\Omega_{\mathbf{x}}^{-1}\tilde{\mathbf{x}}, \Omega_{\mathbf{y}}^{-1}\tilde{\mathbf{y}}) \in T\} \tag{2}$$

and where $\mathbf{p}(\mathbf{x}, \mathbf{y})$ may or may not be dependent on the data point [2].

Intuitively, the absolute commensurability can be understood as a property of independence (of the efficiency score) from the scale of any inputs and any outputs. For example, the efficiency score obtained for any observation with the inputs expressed in tons and outputs in Watt-hours should be *identical* to the efficiency score of the same observation when inputs are expressed in kilograms and outputs in tons of oil equivalent.

Obviously, commensurability is a very desirable property for any efficiency measure and before any measure is used it is crucial to test the robustness of it to a

[2] As we will see later, allowing for such dependency is one of the simplest ways to ensure absolute commensurability.

change of units of measurement. This test is easy to implement for DDFs. Theorem 1 gives necessary and sufficient condition for absolute commensurability of a DDF.

Theorem 1 For any regular technology set T

$$\vec{D}\left(\mathbf{x},\mathbf{y}\middle|-\mathbf{d}_{\mathrm{x}},\mathbf{d}_{\mathrm{y}},T\right)=\vec{D}\left(\tilde{\mathbf{x}},\tilde{\mathbf{y}}\middle|-\mathbf{d}'_{\mathrm{x}},\mathbf{d}'_{\mathrm{y}},\tilde{T}\right),\ \forall(\mathbf{x},\mathbf{y})\in\Re_{+}^{N}\times\Re_{+}^{M} \quad (3)$$

if and only if

$$\mathbf{d}'_{\mathrm{x}}=\Omega_{\mathrm{x}}\mathbf{d}_{\mathrm{x}} \text{ and } \mathbf{d}'_{\mathrm{y}}=\Omega_{\mathrm{y}}\mathbf{d}_{\mathrm{y}}, \quad (4)$$

where, $\mathbf{d}_{\mathrm{x}},\mathbf{d}'_{\mathrm{x}}\in\Re_{+}^{N},\mathbf{d}_{\mathrm{y}},\mathbf{d}'_{\mathrm{y}}\in\Re_{+}^{M},(\mathbf{d}_{\mathrm{x}},\mathbf{d}_{\mathrm{y}})\neq\mathbf{0},(\mathbf{d}'_{\mathrm{x}},\mathbf{d}'_{\mathrm{y}})\neq\mathbf{0}$.

The proof of this result is given in the Appendix, while here it might be worth emphasizing its meaning. The theorem says that the directional distance function is absolutely commensurable *if and only if* the directional vector is 'commensurated' along with the data. In other words, if one does not commensurate the direction in the same fashion as the data, then the scores obtained from DDF might differ for the same data expressed in different units. Remarkably, such commensuration of the directional vector is sufficient for DDF to be absolute commensurable.

4. COMMENSURABILITY PROPERTIES OF THE DDF WITH A UNIT DIRECTIONAL VECTOR

Consider the DDF with a *unit* directional vector (henceforth UDDF), i.e.

$$\vec{D}(\mathbf{x},\mathbf{y}\,|-\mathbf{1}_{N},\mathbf{1}_{M},T)\equiv\sup\{\theta\geq 0:((\mathbf{x}-\theta\mathbf{1}_{N}),(\mathbf{y}+\theta\mathbf{1}_{M}))\in T\}.$$

Notably, this directional vector is among those recommended in a new textbook on productivity and efficiency analysis (Färe, Grosskopf and Margaritis, 2007). This direction is also a very common choice in many empirical works[3] perhaps due to its simplicity, normalizing nature and, as a consequence, convenience in explaining the

[3] E.g., see Chambers and Färe (1998), who used this direction applied to the social welfare theory; Chambers, Färe and Grosskopf (1996) who applied UDDF to measure productivity growth in Asian-Pacific countries; and Färe, Grosskopf and Zelenyuk (2004), where this direction is used in the context of aggregation issue.

results of measurement. Specifically, an efficiency measure based on such a direction gives *one* number indicating (regardless of the units of measurement) how many units of each input must be deducted and how many units of each output must be added to any particular point in the technology set to reach the (upper) frontier of T. Despite its appealing nature, the UDDF is *not* absolute-commensurable. We show this in the next corollary.

<u>Corollary 1</u> UDDF is not absolute-commensurable for all technologies.

□ The proof follows directly from Theorem 1 by noting that the UDDF does not satisfy the necessary condition for absolute commensurability. ■

The practical implication of this result is that different researchers using the same data and methodology may arrive at different estimates—just because the researchers used different units of measurement. One may wonder whether the results would be qualitatively the same: i.e., if, under some units of measurement, firm A was more efficient than firm B then this ranking would, hopefully, remain unaffected under any other units of measurement (different by a scalar transformation). We thus call this concept *ranking-commensurability*, and formally define and apply it to UDDF below.

<u>Definition 2</u> *(Ranking-Commensurability)*

An efficiency measure $E(\mathbf{x},\mathbf{y}|\mathbf{p}(\mathbf{x},\mathbf{y}),T)$ is said to be ranking-commensurable if and only if

$$E(\mathbf{x}_k,\mathbf{y}_k|\mathbf{p}(\mathbf{x}_k,\mathbf{y}_k),T) > E(\mathbf{x}_j,\mathbf{y}_j|\mathbf{p}(\mathbf{x}_j,\mathbf{y}_j),T) \Leftrightarrow$$

$$E(\tilde{\mathbf{x}}_k,\tilde{\mathbf{y}}_k|\mathbf{p}(\tilde{\mathbf{x}}_k,\tilde{\mathbf{y}}_k),\tilde{T}) > E(\tilde{\mathbf{x}}_j,\tilde{\mathbf{y}}_j|\mathbf{p}(\tilde{\mathbf{x}}_j,\tilde{\mathbf{y}}_j),\tilde{T}) \qquad (5)$$

$$\forall\, \mathbf{x}_k,\mathbf{x}_j \in \Re_+^N;\mathbf{y}_k,\mathbf{y}_j \in \Re_+^M,$$

where $\tilde{\mathbf{x}}_k = \Omega_{\mathbf{x}}\mathbf{x}_k$, $\tilde{\mathbf{x}}_j = \Omega_{\mathbf{x}}\mathbf{x}_j$, $\tilde{\mathbf{y}}_k = \Omega_{\mathbf{y}}\mathbf{y}_k$ and $\tilde{\mathbf{y}}_j = \Omega_{\mathbf{y}}\mathbf{y}_j$ are as in definition 1.

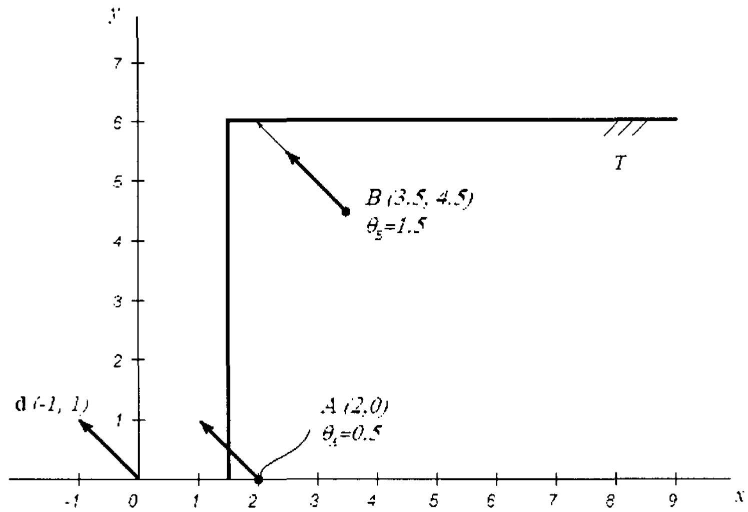

Figure 1-1. Graphical illustration of technology (6), observations A and B and their UDDF scores

The intuition behind (5) is that if an efficiency measure is ranking-commensurable, then changing the units of measurement of any input or/and output by a scalar transformation should not affect the *ranking* of the efficiency scores,

although may change the scores per se. It turns out that UDDF is also *not* ranking commensurable, as we show in the next theorem.

<u>Theorem 2</u> UDDF is not ranking-commensurable for all technologies.

□ To prove this statement, consider a simple single-input-single-output technology,

$$T = \{ (x, y) : y \le 6 \text{ if } x \ge 1.5,\ y = 0 \text{ if } 0 \le x < 1.5 \} \quad x, y \in \Re_+, \tag{6}$$

where the numbers are provided for the sake of illustration. Measuring the UDDF scores of two observations: A at $(x_A = 2, y_A = 0)$ and B at $(x_B = 3.5, y_B = 4.5)$, we would conclude that observation A is more efficient than B, since $\vec{D}(2,0 \mid -1,1) = \theta_A = 0.5$ and $\vec{D}(3.5,4.5 \mid -1,1) = \theta_B = 1.5$. A graphical illustration of this problem is provided in Figure 1-1.

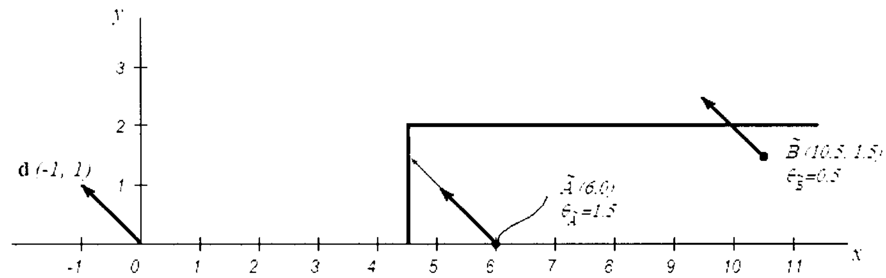

Figure 1-2. Graphical illustration of technology (7), observations $\widetilde{\mathbf{A}}$ and $\widetilde{\mathbf{B}}$ and their UDDF scores

Now, suppose we decide to transform inputs and outputs in such a way that we measure inputs in $\tilde{x} = 3x$ and outputs in $\tilde{y} = y/3$. Combining (2) and (6) implies

$$\widetilde{T} = \{(\tilde{x}, \tilde{y}) : \tilde{y} \le 2 \text{ if } \tilde{x} \ge 4.5, \tilde{y} = 0 \text{ if } 0 \le \tilde{x} < 4.5\} \qquad \tilde{x}, \tilde{y} \in \Re_+ . \qquad (7)$$

Changing the units of measurement of observations would transform observation A into $\widetilde{A}(\tilde{x}_{\widetilde{A}} = 6, \tilde{y}_{\widetilde{A}} = 0)$ and observation B into $\widetilde{B}(\tilde{x}_{\widetilde{B}} = 10.5, \tilde{y}_{\widetilde{B}} = 1.5)$ respectively. This would result in $\vec{D}(6,0 \mid -1,1) = \theta_{\widetilde{A}} = 1.5$ and $\vec{D}(10.5,1.5 \mid -1,1) = \theta_{\widetilde{B}} = 0.5$, *i.e.*, observation B is now concluded to be more efficient than A (graphical illustration is provided in Figure 1-2).

So, the change in units of measurement has led to different ranking. Therefore, the UDDF measure is not ranking commensurable for all technologies. ■

The practical implication of this result is that even the *qualitative* difference in efficiency measurement with UDDF may occur due to changes in the units of measurement, which in turn may lead to different policy implications. It must be also clear that the type of technology we used in our argument was simple for illustration purposes. The same argument would also hold for many other technologies, but just one is enough to prove our claim.

It should be noted, however, that for any regular technology, commensuration will neither affect UDDF scores of the efficient units (i.e. units located on the frontier with the UDDF score equal to zero) nor make any previously inefficient units efficient. In other words, commensuration does not affect identification of efficient units by the UDDF (as a matter of fact, the same is true for *any* DDF). The proof of this fact follows directly from (1) and *definition 1.*

5. ILLUSTRATIVE EXAMPLE

We used data from Kumar and Russell (2002) on capital and labor as inputs and GDP as an output for 57 countries in 1990. Although the study used output oriented Shephard's distance function to estimate Farrell efficiency scores, which are commensurable, we showed that if UDDF was used instead, the estimates would be non-commensurable. First we calculated UDDF efficiency scores based on the original data using data envelopment analysis technique under variable returns to scale; then we calculated UDDF scores based on the commensurated data, where capital and GDP were measured in thousands of the US dollars instead of the US dollars. Commensuration did not affect efficiency scores of the eleven efficient units (UDDF being equal to zero), but did affect all 46 inefficient units. As expected theoretically, not only the absolute values of the scores have changed, but also the ranking altered. Remarkably, 43 out of 46 inefficient observations changed their rankings. The estimates are provided in Table 1-1.

6. DECOMMENSURATION

A natural question now is whether we could remedy the situation with UDDF—with, for example, what we call here as *ex post* and *ex ante* 'de-commensurations'.

<u>Definition 3</u> *(Ex post de-commensuration)*

An efficiency measure $E(\mathbf{x},\mathbf{y}|\mathbf{p}(\mathbf{x},\mathbf{y}),T)$ is de-commensurable *ex post* if and only if

$$\exists G:\Re^1_+ \to \Re^1_+ :\ G\left(E\left(\tilde{\mathbf{x}},\tilde{\mathbf{y}}|\mathbf{p}(\tilde{\mathbf{x}},\tilde{\mathbf{y}}),\tilde{T}\right)\right) = E(\mathbf{x},\mathbf{y}|\mathbf{p}(\mathbf{x},\mathbf{y}),T), \qquad (8)$$

$$\forall T,\ \ \forall \mathbf{x}\in\Re^N_+;\mathbf{y}\in\Re^M_+$$

and where $\tilde{\mathbf{x}} = \Omega_{\mathbf{x}}\mathbf{x}$ and $\tilde{\mathbf{y}} = \Omega_{\mathbf{y}}\mathbf{y}$ are as in definition 1.

Table 1-1. UDDF scores for original and commensurated data

ID	Efficiency before commensuration		Efficiency after commensuration		ID	Efficiency before commensuration		Efficiency after commensuration	
	Score	Rank	Score	Rank		Score	Rank	Score	Rank
1	6502.22	39	5863.1	36	30	4815.73	36	3915.71	30
2	1241.72	18	5926.91	38	31	2705.98	28	3460.46	27
3	926.24	17	982.51	15	32	0	*	0	*
4	508.65	8	559.5	7	33	862.84	14	245.58	1
5	1900.6	24	1754.93	22	34	0	*	0	*
6	484.05	7	843.05	11	35	901.89	16	917.47	13
7	2898.45	29	2613.55	25	36	446.28	6	463.02	5
8	7572.16	41	7448.71	42	37	0	*	0	*
9	859.17	13	899.45	12	38	337.98	4	435.33	4
10	1724.75	23	1555.22	21	39	623.13	11	669.93	10
11	2375.77	27	2367.54	24	40	0	*	0	*
12	534.72	10	644.11	8	41	5564.08	38	4269.24	32
13	3622.59	31	4416.73	34	42	18893.18	45	7826.45	43
14	0.15	2	5912	37	43	1668.94	22	1504.9	20
15	1947.32	25	1955.62	23	44	0	*	0	*
16	1388.65	20	958.87	14	45	3929.33	33	3906.31	29
17	1341.54	19	1181.87	17	46	5189.11	37	5100.62	35
18	0	*	0	*	47	879.35	15	992.45	16
19	0	*	0	*	48	0.01	1	358.68	3
20	0	*	0	*	49	1652.12	21	1495.88	19
21	373.49	5	336.78	2	50	4268.4	34	4338.42	33
22	533.87	9	481.39	6	51	22102.54	46	8194.33	44
23	3651.3	32	3702.19	28	52	17998.61	44	15857.12	45
24	3179.37	30	1357.15	18	53	0	*	0	*
25	857.68	12	663.14	9	54	0	*	0	*
26	4.52	3	29312.57	46	55	7439.29	40	6708.06	40
27	10397.14	43	7434.55	41	56	2375.73	26	2900.18	26
28	9951.74	42	6621.24	39	57	4433.43	35	4036.82	31
29	0	*	0	*					

Notes: * marks efficient units in the Rank column.

Source of original data: Kumar and Russell (2002);

Original data is expressed in US dollars (for GDP and capital) and workers (for employment); commensurated data is expressed in thousands US dollars (for GDP and capital) and workers (for employment).

Efficiency scores are estimated in MATLAB with optimization toolbox.

In words, this is a situation when it is possible to transform, *ex post* or after computation, a UDDF score for 'commensurated' data in such a way that it becomes identical to the UDDF score for the original data. Note that here we require that a de-commensurating transformation G works for all technologies. Therefore, transformation can depend on observed inputs and outputs as well as commensuration matrices, but should be independent of the parameters of the technology. We will now see that, in general, UDDF cannot be remedied by the *ex post* de-commensuration.

<u>Theorem 3</u> UDDF is not de-commensurable *ex post*, independently of technology.

□ Consider a single-input-single-output CRS technology

$$T = \{(x, y) : \lambda x \geq y\} \quad x, y \in \Re_+ . \tag{9}$$

Then,

$$\vec{D}(x, y \mid -1, 1) \equiv \sup\{\theta \geq 0 : \lambda(x - \theta) \geq y + \theta\} = \frac{\lambda x - y}{\lambda + 1} . \tag{10}$$

Let $\Omega_x = \gamma > 1$ and $\Omega_y = 1$, thus $\tilde{x} = \gamma x$ and $\tilde{y} = y$. Combining (2) and (9) implies

$$\tilde{T} = \{(\tilde{x}, \tilde{y}) : \lambda \gamma \tilde{x} \geq \tilde{y}\}, \qquad x, y \in \Re_+$$

and therefore,

$$\vec{D}(\tilde{x}, \tilde{y} \mid -1, 1) = \frac{\lambda \gamma x - \gamma y}{\lambda \gamma + 1} = \frac{\lambda x - y}{\lambda + 1/\gamma} . \tag{11}$$

Finally, combining (10) and (11) implies

$$\vec{D}(x, y \mid -1, 1) = \frac{\lambda + 1/\gamma}{\lambda + 1} \vec{D}(\tilde{x}, \tilde{y} \mid -1, 1) . \tag{12}$$

Thus, (12) proposes a (unique) way to transform (11) to obtain (10) under single-input-single-output CRS technology when output is scaled up. However, this transformation *depends* on parameter of the technology λ, while our definition required such independence (since true technology sets are typically unobserved in practice by researchers). Therefore, (8) does not hold and therefore UDDF is not de-commensurable *ex post*. ■

What must be clear from the intuition of the proofs is that not only the UDDF but *any* DDF with a non-base collinear[4] *fixed* directional vector will not satisfy even ranking-commensurability for *all* technologies (although perhaps for some it might). The intuition for this can be explained by the *additive* nature of such DDFs, which become irreconcilable with the *multiplicative* nature of the commensurability property. This intuition raises another natural question: Is it possible to remedy the UDDF measure in some way to ensure that an efficiency measure *after* commensuration is equal to its value *before* commensuration? One possible remedy might be applied to the (fixed) directional vector. We call this concept '*ex ante* absolute de-commensuration,' and formally define it below.

Definition 4 *(Ex ante de-commensuration)*

An efficiency measure $E(\mathbf{x},\mathbf{y}|\mathbf{p}(\mathbf{x},\mathbf{y}),T)$ is de-commensurable *ex ante* if and only if

$$\exists F:\Re^z \to \Re^z \ : E\left(\tilde{\mathbf{x}},\tilde{\mathbf{y}}\middle|F(\mathbf{p}(\tilde{\mathbf{x}},\tilde{\mathbf{y}})),\tilde{T}\right) = E(\mathbf{x},\mathbf{y}|\mathbf{p}(\mathbf{x},\mathbf{y}),T), \qquad (13)$$

$$\forall T,\ \ \forall\, \mathbf{x}\in\Re_+^N;\mathbf{y}\in\Re_+^M$$

and where $\tilde{\mathbf{x}} = \Omega_{\mathbf{x}}\mathbf{x}$ and $\tilde{\mathbf{y}} = \Omega_{\mathbf{y}}\mathbf{y}$ are as in definition 1.

In words, an efficiency measure is de-commensurable *ex ante* if and only if along with the change of units of measurement of inputs and outputs one shall also modify the parameter vector $\mathbf{p}$ (*e.g.*, the directional vector in the case of DDF) to obtain the same numerical value of efficiency as for the observation in the original units of measurements.

---

[4] By non-base collinear directional vector we mean a vector which is not collinear to any of the base vectors. It is quite straightforward to see that if the directional vector is degenerate, then the DDF is still not commensurable in absolute terms, but is ranking-commensurable for any regular technology.

One may notice a similarity between the *definitions 1* and *4* if $F(\mathbf{p}(\widetilde{\mathbf{x}},\widetilde{\mathbf{y}}))$ is substituted for $\mathbf{p}(\widetilde{\mathbf{x}},\widetilde{\mathbf{y}})$ in the left hand side of (13). Indeed, *definition 4* states that to ensure *ex ante* decommensuration, one provide a mapping F for the directional vector that modifies the vector when the data is commensurated. This vector transformation should make the function absolutely commensurable for any technology.

Trying this on, for our UDDF, we finally get a positive result, which we formally state next.

Corollary 2. UDDF is absolute de-commensurable *ex ante* if its directional vector is 'commensurated' along with the inputs and outputs by pre-multiplying it by the respective commensuration matrices, i.e.,

$$\vec{D}(\widetilde{\mathbf{x}},\widetilde{\mathbf{y}} \mid \Omega_{\mathbf{x}}(-\mathbf{1}_N),\Omega_{\mathbf{y}}\mathbf{1}_M,\widetilde{T}) = \vec{D}(\mathbf{x},\mathbf{y} \mid -\mathbf{1}_N,\mathbf{1}_M,T). \tag{14}$$

To prove this result, one just has to notice that (14) is just a corollary of the Theorem 1. It is worth noting, however, that once the directional vector of UDDF's has been pre-multiplied by the (non-identity) commensuration matrices, the DDF is no longer a UDDF. Therefore, once commensuration takes place and a researcher wants to compare her results to the results of the other study with the 'non-commensurated' data, she cannot use UDDF, but should rather use a directional distance function with a 'commensurated' directional vector. In other words, only the 'lucky first' gets to use DDF with a *unit* directional vector and thus sets the standard for units of measurement. The others have to comply with this standard by commensurating their unit directional vectors with the same matrices that relate their units of measurement to that of this standard. As standard scale is rather a question of tastes, traditions and practices, which may differ from school to school and from society to society, it might be difficult to achieve a consensus over this issue, as it is hard to convince Europeans, for example, to switch to British weights and measures (e.g., pounds) or Americans to metric system (e.g., kilograms). A compromise might be to use a directional distance function with a directional vector that would 'commensurate' itself according to (4) when the data is commensurated.

7. CONCLUDING REMARKS

An interesting special, but still quite broad, case that ensures (4), which actually would not require commensuration of the directional vector explicitly, but would be obtained automatically, is when $(-\mathbf{d}_x, \mathbf{d}_y) = (-\Xi_x \mathbf{x}, \Xi_y \mathbf{y})$, where Ξ_x and Ξ_y are any diagonal matrices of dimensions NxN and MxM, respectively, where the elements are constants that define the direction or, intuitively speaking, assign weights of each input (output) relative to other inputs (outputs) in measuring the distance to the frontier. (The diagonal elements of Ξ_x and Ξ_y can be zero or even negative). Special cases of this function have appeared in the past publications. For example, Chung Färe and Grosskopf (1997) used it for

$$\Xi_x = \mathbf{0}_{N \times N}, \quad \Xi_y = \left(\begin{array}{c|c} \mathbf{I}_{G \times G} & \mathbf{0}_{G \times B} \\ \hline \mathbf{0}_{B \times G} & -\mathbf{I}_{B \times B} \end{array} \right),$$

where $\mathbf{I}_{G \times G}$ is a GxG identity matrix assigning equal weights to G good outputs, $\mathbf{I}_{B \times B}$ is a BxB identity matrix assigning equal weights to B bad outputs and the rest are zero matrices of dimensions indicated in subscripts. Also, Zelenyuk (2002) used it when Ξ_x and Ξ_y were scalars multiplied by corresponding identity matrices. Even more special cases are the popular directional vectors $(-\mathbf{x}, \mathbf{y})$, as well as $(\mathbf{0}, \mathbf{y})$ and $(-\mathbf{x}, \mathbf{0})$ under which one-to-one closed-form relationships with the Shephard's distance functions are known.

It is worth noting that the test for satisfying absolute (and ranking) commensurability helps reducing the problem of choosing the direction of measurement for DDF considerably. In particular, it might discourage one from using DDF with any *fixed* vector, since such DDFs are not ranking-commensurable in addition to not being absolute-commensurable. Alternatively, if for some reason researchers must use a DDF with a fixed vector, then researchers seeking to replicate this study in the future should pay special consideration to the issues of potential non-commensurability associated with such DDFs. Specifically, researchers should try to decommensurate this DDF ex-ante or transform their data into the system of measurement used by the original study. Moreover, since the DDF in (3)-(4) is a special case of the general DDF, it thus not only passes the tests for commensurability, but (under the regularity conditions) also satisfies all the general properties derived by Luenberger (1992, 1994), Chambers, Chung and Färe (1996, 1998), Färe and Grosskopf (2000) and Färe and Primont (2006). Finally, our findings give a simple necessary condition test for verifying computer codes for estimation of DDF with vectors satisfying (4)—if estimates change after multiplication of some input/output vectors by scalars, the code contains a mistake. Finally, a natural extension to our work would be an exploration of other properties for DDF that are generally desirable for theoretical or empirical characterizations of technologies and efficiency measurement, and the ones that can help reducing the choice of direction for measurement even further.

BIBLIOGRAPHY

Bol, George. 1986. On Technical Efficiency Measures: a Remark. *Journal of Economic Theory* 38, 380-385.

Chambers, Robert G., Yangho Chung, and Rolf Färe. 1996. Benefit and Distance Function. *Journal of Economic Theory* 70, 407-419.

Chambers, Robert G., Yangho Chung, and Rolf Färe. 1996. Profit, Directional Distance Function and Nerlovian Efficiency. *Journal of Optimization Theory and Applications* 98, 351-364.

Chambers, Robert G., and Rolf Färe. 1998. Translation Homotheticity. *Economic Theory* 11, 629-641.

Chambers, Robert G., Rolf Färe, and Shawna Grosskopf. 1996. Productivity Growth in APEC Countries. *Pacific Economic Review* 1, 181-190.

Chung, Yangho, Rolf Färe, and Shawna Grosskopf. 1997. Productivity and Undesirable Outputs: a Directional Distance Function Approach. *Journal of Environmental Management* 51, 229-240.

Eichhorn, Wolfgang, and Joachim Voeller. 1976. *Theory of the Price Index, Fisher's Test Approach and Generalizations*. Lectures notes in economics and mathematical systems, Vol. 140. Berlin: Springer-Verlag.

Färe, Rolf, and C.A. Knox Lovell. 1978. Measuring the Technical Efficiency of Production. *Journal of Economic Theory* 19, 150-162.

Färe, Rolf, and Shawna Grosskopf. 2000. Theory and Application of Directional Distance Functions. *Journal of Productivity Analysis* 13, 93-103.

Färe, Rolf, Shawna Grosskopf., and Dimitri Margaritis. 2007. Efficiency and Productivity: Malmquist and More. In *The Measurement of Productive Efficiency and Productivity Growth.* New York: Oxford University Press.

Färe, Rolf, Shawna Grosskopf., and Valentin Zelenyuk. 2004. Aggregation of Cost Efficiency: Indicators and Indexes across Firms. *Academia Economic Papers* 32, 395-412.

Färe, Rolf, and Daniel Primont. 2006. Directional Duality Theory. *Economic Theory* 29, 239-247.

Kumar, Subodh, and R. Robert Russell. 2002. Technological Change, Technological Catch-Up, and Capital Deepening: Relative Contributions to Growth and Convergence. *American Economic Review* 92, 527-548.

Luenberger, David G. 1992. New Optimality Principles for Economic Efficiency and Equilibrium. *Journal of Optimization Theory and Applications* 75, 221-264.

Luenberger, David G. 1994. Optimality and the Theory of Value. *Journal of Economic Theory* 63, 147-169.

Russell, R. Robert. 1987. On the Axiomatic Approach to the Measurement of Technical Efficiency. In: Eichhorn, Wolfgang (ed.), Measurement in Economics: Theory and Application of Economic Indices. Heidelberg: Physica-Verlag.

Russell, R. Robert. 1990. Continuity of Measures of Technical Efficiency. *Journal of Economic Theory* 51, 255-267.

Shephard, Ronald W. 1953. Cost and Production Functions. Princeton, NJ: Princeton University Press.

Zelenyuk, Valentin. 2002. Essays in Efficiency and Productivity Analysis of Economic systems. PhD thesis. Corvallis: Oregon State University.

APPENDIX

Proof of Theorem 1

Sufficiency

□ Suppose $\mathbf{d}'_x = \Omega_x \mathbf{d}_x$ and $\mathbf{d}'_y = \Omega_y \mathbf{d}_y$, then

$$\begin{aligned}\vec{D}\left(\tilde{\mathbf{x}},\tilde{\mathbf{y}}\middle|-\mathbf{d}'_x,\mathbf{d}'_y,\tilde{T}\right) &= \sup\left\{\theta \geq 0 \middle| \left(\left(\tilde{\mathbf{x}}-\theta\mathbf{d}'_x\right),\left(\tilde{\mathbf{y}}+\theta\mathbf{d}'_y\right)\right)\in\tilde{T}\right\} \\ &= \sup\left\{\theta \geq 0 \middle| \left(\Omega_x\left(\mathbf{x}-\theta\mathbf{d}_x\right),\Omega_y\left(\mathbf{y}+\theta\mathbf{d}_y\right)\right)\in\tilde{T}\right\} \\ &= \sup\left\{\theta \geq 0 \middle| \left(\left(\mathbf{x}-\theta\mathbf{d}_x\right),\left(\mathbf{y}+\theta\mathbf{d}_y\right)\right)\in T\right\} \\ &= \vec{D}\left(\mathbf{x},\mathbf{y}\middle|-\mathbf{d}_x,\mathbf{d}_y\right).\end{aligned}$$ ■

Necessity

□ Suppose that (3) holds, i.e.

$$\vec{D}\left(\mathbf{x},\mathbf{y}\middle|-\mathbf{d}_x,\mathbf{d}_y,T\right)=\vec{D}\left(\tilde{\mathbf{x}},\tilde{\mathbf{y}}\middle|-\mathbf{d}'_x,\mathbf{d}'_y,\tilde{T}\right),\quad \forall\left(\mathbf{x},\mathbf{y}\right)\in\Re_+^N\times\Re_+^M$$

Also note that,

$$\begin{aligned}\vec{D}\left(\tilde{\mathbf{x}},\tilde{\mathbf{y}}\middle|-\mathbf{d}'_x,\mathbf{d}'_y,\tilde{T}\right) &= \sup\left\{\theta \geq 0 \middle| \left(\left(\tilde{\mathbf{x}}-\theta\mathbf{d}'_x\right),\left(\tilde{\mathbf{y}}+\theta\mathbf{d}'_y\right)\right)\in\tilde{T}\right\} \\ &= \sup\left\{\theta \geq 0 \middle| \left(\Omega_x\left(\mathbf{x}-\theta\Omega_x^{-1}\mathbf{d}'_x\right),\Omega_y\left(\mathbf{y}+\theta\Omega_y^{-1}\mathbf{d}'_y\right)\right)\in\tilde{T}\right\} \\ &= \sup\left\{\theta \geq 0 \middle| \left(\left(\mathbf{x}-\theta\Omega_x^{-1}\mathbf{d}'_x\right),\left(\mathbf{y}+\theta\Omega_y^{-1}\mathbf{d}'_y\right)\right)\in T\right\} \\ &= \vec{D}\left(\mathbf{x},\mathbf{y}\middle|-\Omega_x^{-1}\mathbf{d}'_x,\Omega_y^{-1}\mathbf{d}'_y\right)\end{aligned}$$

Combining this result with (3), gives us

$$\vec{D}\left(\mathbf{x},\mathbf{y}\middle|-\mathbf{d}_x,\mathbf{d}_y,T\right)=\vec{D}\left(\mathbf{x},\mathbf{y}\middle|-\Omega_x^{-1}\mathbf{d}'_x,\Omega_y^{-1}\mathbf{d}'_y,T\right),\forall\left(\mathbf{x},\mathbf{y}\right)\in\Re_+^N\times\Re_+^M. \qquad (*)$$

Clearly, the statement (*) can hold only if $\mathbf{d}_x = \Omega_x^{-1}\mathbf{d}'_x$; $\mathbf{d}_y = \Omega_y^{-1}\mathbf{d}'_y$, or equivalently, $\mathbf{d}'_x = \Omega_x\mathbf{d}_x$; $\mathbf{d}'_y = \Omega_y\mathbf{d}_y$, which is what is needed to prove the necessity. ■

Mykhaylo Salnykov

ESSAY 2

SMOOTH HOMOGENEOUS BOOTSTRAP BIAS CORRECTION IN FRONTIER MODELS: A MONTE CARLO ASSESSMENT WHEN SOME OUTPUTS ARE UNDESIRABLE

M. Salnykov
Department of Economics, Simon Fraser University
8888 University Drive, Burnaby, BC, Canada, V5A 1S6
E-mail: salnykov@sfu.ca
Tel.: +1 604 291 35 08; fax: +1 604 291 59 44

Abstract

Hyperbolic efficiency measures are popular tools in the assessment of the efficiency of multi-output technologies when the outputs need to be treated asymmetrically. In this paper I test the performance of bias-correction based on the smooth homogeneous bootstrap in multi-output frontier models. In addition, I propose an approximation technique that substantially reduces the nonparametric estimation time while sacrificing little precision compared to the most precise nonparametric alternative.

The performance of the uncorrected and bias-corrected estimates is tested in samples of different sizes via Monte Carlo simulation. Of course, all techniques perform well in large samples even without bias correction. Both parametric and nonparametric estimators benefit from the nonparametric bias correction regardless of the sample size. Uncorrected nonparametric estimators perform well in large samples only and require bias-correction in the medium and small-sized samples. Notably, in the small samples bias correction shows marginally better results when applied to the parametric estimator.

Keywords production theory; hyperbolic efficiency function; efficiency measurement; smooth homogeneous bootstrap; Monte Carlo simulation

JEL Classification Numbers C15, D2, D24

1. INTRODUCTION

Various distance functions are main tools in contemporary efficiency and productivity analysis. Debreu (1951) and Shephard (1953) introduced these instruments for analyzing producer behaviour; Malmquist (1953) proposed an application for analyzing consumer behaviour. The commonly used today distance functions give a complete characterization of technology under a fairly weak set of assumptions about the properties of the technologies. The estimation of these functions requires neither the knowledge of input prices and output prices, nor does it rely on widely disputed assumptions about economic behaviour, such as cost minimization or revenue maximization.

The seminal paper of Farrell (1957) commenced the use of the input and output oriented distance functions in the efficiency analysis. Meeusen and van den Broeck (1977), Aigner, Lovell and Schmidt (1977) and Charnes, Cooper and Rhodes (1978) were among those who popularized the Farrell's overall efficiency measure in the econometric and operations research literature.

The early studies in the field traditionally defined efficiency measures in *radial* terms. Boles (1966) proposed the radial *output* orientation (*i.e.* finding the greatest technologically feasible radial *expansion* of the output bundle) while Shepherd (1953) suggested the radial *input* orientation (which is the smallest possible technologically feasible radial *contraction* of the input bundle). Färe, Grosskopf and Lovell (1985) combined these two approaches by formulating a hyperbolic efficiency function (HEF), which does a simultaneous proportional contraction of the input bundle *and* an expansion of the output bundle.

It has been widely agreed that HEF may be a convenient instrument for analyzing technologies where outputs should be approached asymmetrically. Polluting technologies are a common example of such a case. As a decision making unit (DMU) becomes more efficient, it either decreases its pollution (undesirable output) or increases its economic (desirable) output or does both[1].

An alternative to HEF, the output directional distance functions (ODDF), was proposed a decade later by Chambers, Chung and Färe (1996) and applied to polluting technologies by Chung, Färe and Grosskopf (1997). ODDF suggests movement towards the best practice frontier along the exogenously set directional vector. A major argument for using ODDF instead of HEF is that the traditional (nonparametric) approach towards ODDF estimation is executed using linear optimization. It often consumes substantially less computing resources as compared to the nonlinear optimization used in the nonparametric HEF estimation, especially when numerous iterations are involved (such as in the bootstrapping). An alternative, parametric technique towards HEF estimation recently developed by Cuesta and Zofio (2005), relies on the almost homogeneity property of HEF and is as easy to estimate as the parametrically specified ODDF. Yet, many researchers still prefer nonparametric specification as it requires no ex-ante functional form specification. Arguably, the main advantage of HEF against ODDF, however, is that in many studies ODDF lacks theoretical justification for the choice of the directional vector.

Besides a choice of instruments, a researcher often faces choices regarding estimation techniques. Nonparametric Data Envelopment Analysis (DEA) and the parametric linear programming Translog-specified flexible functional form estimators (hereafter Translog estimators) are the ones most often used when HEF is being estimated. Both estimators are inherently downward biased, yet are consistent[2], as is common in the frontier estimation problems. The bootstrap bias-correction procedure proposed by Simar and Wilson (1998) aims at reducing (and presumably eliminating) the downward bias of the frontier estimates. However, the relative performance of this procedure in parametric versus nonparametric frameworks is underinvestigated.

---

[1] Obviously, shrinking inputs is an alternative approach to the efficiency analysis.

[2] Given absence of the specification bias for the Translog specification.

In this paper I will focus on comparing the relative performance of the smooth homogenous bootstrap bias correction in the alternative HEF estimation frameworks in very small samples when stochastic noise can be neglected. My study is similar to those of, for example, Guilkey, Lovell and Sickles (1983) or Färe, Martins-Filho and Vardanyan (2006, FMV)[3] as it also sets up a Monte Carlo experiment to construct a sample of input-output combinations for each DMU. The crucial difference, however, is that my approach involves analyzing the performance of different techniques based on the bootstrap bias-corrected estimates of the efficiency scores as opposed to the raw scores. This includes comparing the power of the bootstrap-based bias correction in parametric versus nonparametric frameworks depending on the sample size.

Further, my paper shows that by reciprocating some of the data it is possible to linearly approximate HEF. This approach gives more precise results as compared to previously used linear approximation based on Taylor expansion and allows to substantially reduce computational time as compared to the nonlinear programming approach.

The rest of the paper proceeds as follows. The following section provides a brief introduction to the fundamentals of efficiency and productivity analysis and discusses one of the tools used in the field – the hyperbolic distance function. The third section reviews some of the most commonly used techniques for HEF estimation. The fourth section shows how the data set can be modified to overcome some of the shortcomings of the traditional estimation techniques. The fifth section explains the smooth homogeneous bootstrap technique along with the procedures for obtaining bootstrap statistical inferences (namely, bias-corrected estimates of the efficiency scores and the confidence intervals). The sixth section describes the data generating process and outlines the frontier comparison criteria. The seventh section discusses the results and the last section is a conclusion of the paper.

[3] Gulkey et al. (1983) examined the behaviour of three functions used to model cost functions; Färe et al. (2006) studied goodness of approximation for parametric quadratic versus Translog specification of the distance function.

2. HYPERBOLIC EFFICIENCY MEASURE

Suppose the following dataset of i=(1,..., I) DMUs consisting of the triplets of vectors ($\mathbf{x}_i$, $\mathbf{y}_i$, $\mathbf{z}_i$), where $\mathbf{x}_i \in \Re_+^N$ is a vector of inputs, $\mathbf{y}_i \in \Re_+^M$ is a vector of economic (desirable) outputs and $\mathbf{z}_i \in \Re_+^K$ is a vector of undesirable outputs.

Further define a *technology set* as

$$T \equiv \{(\mathbf{x},\mathbf{y},\mathbf{z}) : \mathbf{x} \text{ can produce } (\mathbf{y},\mathbf{z})\}. \tag{1}$$

When using output oriented efficiency measures, it is often more convenient to operate on *output set* defined as

$$P(\mathbf{x}) \equiv \{(\mathbf{y},\mathbf{z}) : (\mathbf{x},\mathbf{y},\mathbf{z}) \in T\}. \tag{2}$$

Both technology and output sets are equivalent representations of the technological process.

In the spirit of the axiomatic approach proposed by Shephard (1970) and then followed widely in the literature on productivity and efficiency analysis (e.g., Färe and Primont 1995 and Färe, Grosskopf, Noh and Weber 2005), below I postulate a set axioms the technology must satisfy to be a valid model of production when some outputs are undesirable.

A1. *Doing nothing is possible.* $(\mathbf{0}_M, \mathbf{0}_K) \in P(\mathbf{x})$ for all $\mathbf{x}$ in $\Re_+^N$.

A2. *No free lunch.* $P(\mathbf{0}_N) = (\mathbf{0}_M, \mathbf{0}_K)$.

A3. *Strong disposability of inputs.* If $(\mathbf{y},\mathbf{z}) \in P(\mathbf{x})$ and $\mathbf{x}' \geqq \mathbf{x}$ then $(\mathbf{y},\mathbf{z}) \in P(\mathbf{x}')$.

A4. *Strong disposability of desirable outputs.* If $(\mathbf{y},\mathbf{z}) \in P(\mathbf{x})$ and $\mathbf{y} \geqq \mathbf{y}'$ then $(\mathbf{y}',\mathbf{z}) \in P(\mathbf{x})$.

A5. *Weak disposability of outputs.* If $(\mathbf{y},\mathbf{z}) \in P(\mathbf{x})$ then $(\theta\mathbf{y}, \theta\mathbf{z}) \in P(\mathbf{x})$ for all $\theta \leq 1$.

A6. *Scarcity.* For all $\mathbf{x}$ in $\Re_+^N$, $P(\mathbf{x})$ is a bounded set.

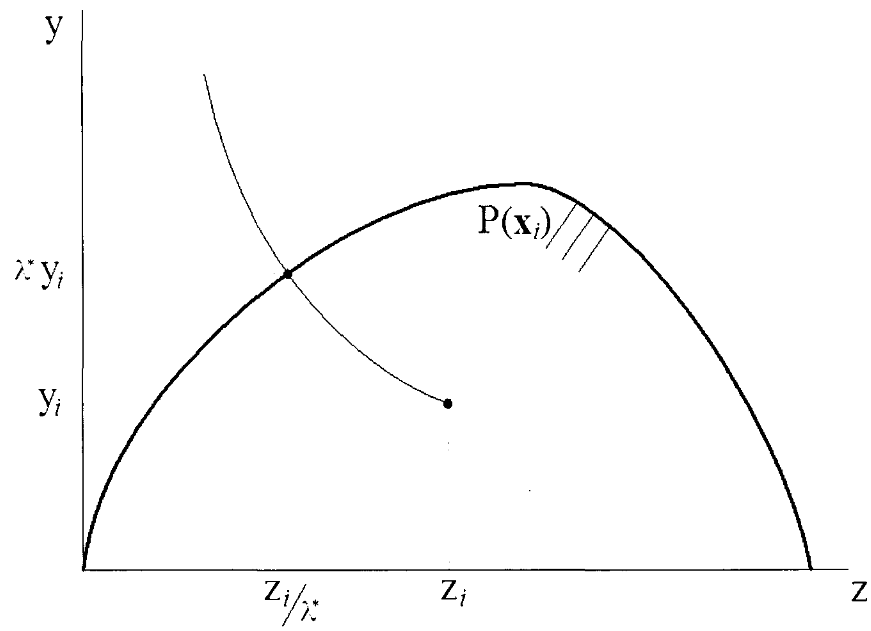

Figure 2-1. Typical shape of the output set when technology produces both desirable (y) and undesirable (z) outputs and the illustration of the hyperbolical projection to the efficient frontier.

A7. *Output closedness.* For all $\mathbf{x}$ in $\Re_+^N$, $P(\mathbf{x})$ is a closed set.

A8. *Input convexity.* P is quasiconcave on $\Re_+^N$, i.e. for all $\mathbf{x}, \mathbf{x}' \in \Re_+^N$ and $0 \le \lambda \le 1$, $P(\mathbf{x}) \cap P(\mathbf{x}') \subseteq P(\lambda \mathbf{x} + (1-\lambda)\mathbf{x}')$.

A9. *Output convexity.* For all $\mathbf{x}$ in $\Re_+^N$, $P(\mathbf{x})$ is a convex set.

A typical shape of the output set $P(\mathbf{x})$ is illustrated on Figure 2-1.

Define an *output hyperbolic efficiency function* on $P(\mathbf{x})$ as

$$H_O(\mathbf{x}, \mathbf{y}, \mathbf{z}) \equiv \sup\{\lambda : (\lambda \mathbf{y}, \lambda^{-1}\mathbf{z}) \in P(\mathbf{x})\} \tag{3}$$

In Figure 2-1 H_O maximizes λ hyperbolically by comparing the ith DMU's observed input-output combination ($\mathbf{x}_i$, $\mathbf{y}_i$, $\mathbf{z}_i$) to the frontier combination $(\mathbf{x},\ \lambda^* \mathbf{y},\ \mathbf{z}/\lambda^*)$, where λ^* is the maximized value of λ.

Similarly to the traditional radial distance functions, HEF gives complete characterization of a technology. Unlike the traditional distance functions, however, HEF allows for asymmetrical treatment of different outputs.

It has been shown by Färe, Grosskopf and Lovell (1985) that HEF possesses the following four properties.

P1. *Representation.* $H_O(\mathbf{x},\mathbf{y},\mathbf{z}) \geq 1$ if and only if $(\mathbf{y},\mathbf{z}) \in P(\mathbf{x})$.

P2. *Almost homogeneity*[4] *of degrees 0,-1,1 and 1.* $H_O(\mathbf{x},\mu^{-1}\mathbf{y},\mu\mathbf{z}) = \mu H_O(\mathbf{x},\mathbf{y},\mathbf{z}) \forall \mu > 0$.

P3. *Monotonicity in* $\mathbf{y}$. $H_O(\mathbf{x},\mu\mathbf{y},\mathbf{z}) \geq H_O(\mathbf{x},\mathbf{y},\mathbf{z}) \quad 0 \leq \mu \leq 1$.

P4. *Monotonicity in* $\mathbf{z}$. $H_O(\mathbf{x},\mathbf{y},\mu\mathbf{z}) \geq H_O(\mathbf{x},\mathbf{y},\mathbf{z}) \quad \mu \geq 1$.

Strong disposability of inputs also implies monotonicity in inputs, which is formulated as following.

P5. *Monotonicity in* $\mathbf{x}$. $H_O(\mu\mathbf{x},\mathbf{y},\mathbf{z}) \geq H_O(\mathbf{x},\mathbf{y},\mathbf{z}) \quad \mu \geq 1$.

3. TRADITIONAL APPROACHES TO HEF ESTIMATION

Empirical estimation of the efficiency measures can be accomplished using either nonparametric programming techniques, semi-parametric methods or parametric frontier regressions.

The most common approach to the estimation of the hyperbolic efficiency measures has been by nonparametric mathematical programming, namely Data Envelopment Analysis (DEA) (see, for example, Färe, Grosskopf, Lovell and Pasurka (FGLP), 1989 and Zofio and Lovell, 2001). An alternative nonparametric approach, the Free Disposable Hull (FDH) technique, can also be used in HEF estimation, yet, it did not gain much popularity as compared to the DEA technique.

---

[4] A function $F(x,y,z)$ is almost homogeneous of degrees k_1, k_2, k_3 and k_4 if and only if $F(\mu^{k_1}x, \mu^{k_2}y, \mu^{k_3}z) = \mu^{k_4}F(x,y)$ (Aczel, 1966).

Semi-parametric and parametric frontier models necessitate an ex-ante knowledge on the distribution of composite error terms or functional specification of the technology, which should satisfy certain regularity conditions (e.g. Chambers, 1988). Ideally, these forms should be flexible, amendable to imposition of the efficiency function properties (such as translation property or homogeneity) and relative easiness of estimation (Cuesta and Zofio, 2005). Two by far most popular parametric specifications are the Translog specification, first introduced by Christensen, Jorgensen and Lau (1971, 1973), and quadratic specification proposed by Färe and Sung (1986). While the literature on the nonparametric techniques of HEF estimation is rather small, it used to be virtually nonexistent on the parametric and semi-parametric side until recently. Yet, in the latest study Cuesta and Zofio (2005) showed the way of incorporating almost homogeneous property of HEF into the Translog flexible specification.

Below, I will review the most common traditional nonparametric and parametric approaches to HEF estimation, namely the DEA and the Translog parametric frontier regression methods.

3.1. DEA SPECIFICATIONS

Data envelopment analysis (DEA) approach to the frontier estimation gained increasing popularity after the papers of Afriat (1972) and Charnes, Cooper and Rhodes (1978). DEA is based on the activity analysis approach to compute efficiency going back to von Neumann (1945), Karlin (1959) and Shephard (1970).

DEA constructs the smallest possible convex hull enveloping all of the data points subject to the assumed returns to scale. As discussed above, each data point i is characterized by a triplet of vectors ($\mathbf{x}_i$, $\mathbf{y}_i$, $\mathbf{z}_i$), where $\mathbf{x}_i \in \Re_+^N$, $\mathbf{y}_i \in \Re_+^M$, $\mathbf{z}_i \in \Re_+^K$ are vector of inputs, desirable and undesirable outputs respectively. Denote the NxI matrix of observed inputs for *all* DMUs by X, the MxI matrix of desirable outputs by Y and the KxI matrix of undesirable outputs by Z. The matrices X, Y and Z are nonnegative having strictly positive row sums and column sums (Karlin, 1959 and

Shepherd, 1974). Further, let $\Psi \in \Re^{I}$ denote an *Ix1* vector of intensity variables for each of the DMUs. Then technology (1) satisfying regularity conditions A1-A9 can be defined as

$$T^{CRS} = \{(\mathbf{x}, \mathbf{y}, \mathbf{z}) : \Psi' X \leq \mathbf{x},\ \Psi' Y \geqq \mathbf{y} \geqq \mathbf{0}_M, \Psi' Z \leqq \mathbf{z}, \Psi \geqq \mathbf{0}_I\}. \quad (4)$$

Technology set (4) defines the smallest convex cone enveloping all of the data points. Modelling set (4) imposes globally constant returns to scale (CRS) assumption.

Different returns to scale can be imposed by modifying constraints placed on the intensity variables. Specifically, non-increasing returns to scale (NIRS) technology set can be defined as

$$T^{NIRS} = \{(\mathbf{x}, \mathbf{y}, \mathbf{z}) : \Psi' X \leq \mathbf{x},\ \Psi' Y \geqq \mathbf{y} \geqq \mathbf{0}_M, \Psi' Z \leqq \mathbf{z}, \Psi' \geqq \mathbf{0}_I, \mathbf{i}'\Psi \leq 1\}, \quad (5)$$

where **i** is an *Ix1* vector consisting of ones.

Technology set (5) places an additional restriction $\mathbf{i}'\Psi \leq 1$ on set (4) and defines the smallest convex weakly disposable hull enveloping all of the data points.

Variable returns to scale (VRS) technology set is defined as

$$T^{VRS} = \{(\mathbf{x}, \mathbf{y}, \mathbf{z}) : \Psi' X \leq \mathbf{x},\ \Psi' Y \geqq \mathbf{y} \geqq \mathbf{0}_M, \Psi' Z \leqq \mathbf{z}, \Psi \geqq \mathbf{0}_I, \mathbf{i}'\Psi = 1\}, \quad (6)$$

Technology set (6) places an additional restriction $\mathbf{i}'\Psi = 1$ on set (4) and defines the smallest convex hull enveloping all of the data points.

It can be easily seen that VRS set satisfies NIRS constraints and NIRS constraints satisfy CRS ones. Not surprisingly, $T^{VRS} \subset T^{NIRS} \subset T^{CRS}$. For a detailed discussion of the returns to scale see Färe and Primont (1995).

CRS nonparametric frontier is both biased and inconsistent unless the true technology is CRS. The same problem applies to NIRS specification. VRS nonparametric frontier is also biased, but consistent even when true technology is CRS or NIRS. Therefore, in this study I focus primarily on the VRS specification of the DEA frontier.

The consistency and the convergence rate of the DEA scores in multi-input, multi-output setup were established analytically (see Kneip, Park and Simar, 1998). Gijbels, Mammen, Park and Simar (1999) obtained limit distribution of the DEA scores in 1-input, 1-output case depending on the curvature of the frontier and the density at the boundary. These results were extended to higher dimensional cases by Kneip, Simar and Wilson (2003).

Specification DEA T (HEF estimation via VRS DEA: traditional approach)

Similarly to FGLP (1989), if VRS is assumed, DEA HEF score for DMU *i* can be estimated by solving a nonlinear optimization problem

$$H_O^{DEA}\left(\mathbf{x}_i, \mathbf{y}_i, \mathbf{z}_i\right) = \max \lambda \qquad (7)$$

s.t.

(i)	$\Psi' X \leq \mathbf{x}_i$	(iv)	$\Psi \geqq \mathbf{0}_I$
(ii)	$\Psi' Y \geqq \lambda \mathbf{y}_i$	(v)	$\mathbf{i}'\Psi = 1$
(iii)	$\Psi' Z \leqq \lambda^{-1} \mathbf{z}_i$		

Whereas nonlinear optimization is more computationally complicated than linear programming, FGLP (1989) also proposed to linearly approximate a nonlinear constrain (iii) of equation (7) around λ=1 resulting in the following specification.

Specification DEA LA (HEF estimation via VRS DEA: linear approximation)

First order linear Taylor approximation of constraint (iii) of equation (7) around λ=1 allows to transform (7) into a linear programming problem

$$H_O^{DEA\,LA}\left(\mathbf{x}_i, \mathbf{y}_i, \mathbf{z}_i\right) = \max \lambda \qquad (8)$$

s.t.

(i)	$\Psi' X \leq \mathbf{x}_i$	(iv)	$\Psi' \geqq \mathbf{0}_I$
(ii)	$\Psi' Y \geqq \lambda \mathbf{y}_i$	(v)	$\mathbf{i}'\Psi = 1$
(iii)	$\Psi' Z \leqq 2\mathbf{z}_i - \lambda \mathbf{z}_i$		

3.2. PARAMETRIC SPECIFICATION

Parametric frontier regressions were pioneered by Aigner and Chu (1968). Their study followed Farrell's idea of describing "an industry envelope isoquant" and described a method of the frontier model estimation that constrained all residuals to be nonpositive. For simplicity, Aigner and Chu estimated one-output, two-input production function

$$y = Ax_1^{\alpha} x_2^{\beta} u \,, \tag{9}$$

where y is an output, x's denote two inputs, u is a random shock (inefficiency) and A, α and β are parameters. (9) can be rewritten in logarithms to obtain

$$e = -\ln A + \ln y + \alpha \ln x_1 + \beta \ln x_2 , \tag{10}$$

where $e = \ln u$. Aigner and Chu further assume that the shocks lie only on one side of the production frontier, thus (10) can be solved by minimizing the linear loss function as opposed to the sum of squared residuals as in, for example, ordinary least squares regressions. Parameters of (10) can be estimated in the framework of linear programming by finding A, α and β that solve

$$\max \sum_{i=1}^{I} \left[\ln y_i - \ln A - \alpha \ln x_{1i} - \beta \ln x_{2i} \right] \tag{11}$$

s.t.

(i) $A, \alpha, \beta \geq 0$

(ii) $\ln y_i - \ln A - \alpha \ln x_{1i} - \beta \ln x_{2i} \leq 0$.

Christensen, Jorgenson and Lau (1971) showed that a homogeneous Translog aggregator function defined as

$$f(x) \equiv \alpha_0 + \sum_{j=1}^{J} \alpha_j \ln x_j + \frac{1}{2} \sum_{j=1}^{J} \sum_{j'=1}^{J} \alpha_{jj'} \ln x_j \ln x_{j'} \tag{12}$$

where $\sum_{j=1}^{J} \alpha_j = 1$, $\alpha_{jj'} = \alpha_{j'j}, \sum_{j=1}^{J} \alpha_{j'j} = 0$ for $j = 1,\ldots,J$ can give a second order approximation to any twice continuously differentiable linear homogenous function.

Diewert (1976) used Aigner and Chu approach and Christensen, Jorgenson and Lau finding to parameterize the Shephard's distance functions to what now is known as *Translog* specification of the distance function by imposing additional conditions on the parameters to account for the properties of the distance functions.

Traditionally, Translog specification was applied to estimate homogeneous distance functions (see, for example, Färe and Grosskopf, 1990). HEF, however, was purposefully avoided as it is not homogeneous of any degree. Cuesta and Zofio (2005) extended Diewert's application to estimate a Translog-specified HEF by using its almost homogeneity property.

In terms of this study, they defined HEF through a Translog function

$$\ln H_O^{TL}(\mathbf{x}_i,\mathbf{y}_i,\mathbf{z}_i) = \alpha_0 + \sum_{n=1}^{N}\beta_n \ln x_i^n + \sum_{m=1}^{M}\gamma_m \ln y_i^m + \sum_{k=1}^{K}\delta_k \ln z_i^k \tag{13}$$

$$+\frac{1}{2}\sum_{n=1}^{N}\sum_{n'=1}^{N}\zeta_{nn'}\ln x_i^n \ln x_i^{n'} + \frac{1}{2}\sum_{m=1}^{M}\sum_{m'=1}^{M}\eta_{mm'}\ln y_i^m \ln y_i^{m'} + \frac{1}{2}\sum_{k=1}^{K}\sum_{k'=1}^{K}\theta_{kk'}\ln z_i^k \ln z_i^{k'}$$

$$+\sum_{n=1}^{N}\sum_{m=1}^{M}\rho_{nm}\ln x_i^n \ln y_i^m + \sum_{n=1}^{N}\sum_{k=1}^{K}\tau_{nk}\ln x_i^n \ln z_i^k + \sum_{m=1}^{M}\sum_{k=1}^{K}\chi_{mk}\ln y_i^m \ln z_i^k$$

where x_i^j, y_i^j and z_i^j denote *j*th elements of vectors $\mathbf{x}_i$, $\mathbf{y}_i$ and $\mathbf{z}_i$ respectively.

Since HEF is almost homogeneous of degrees 0, -1, 1 and 1, the modified Euler theorem introduced by Lau (1972) implies

$$0\cdot\sum_{n=1}^{N}\frac{\partial H_O^{TL}(\cdot)}{\partial x_i^n}x_i^n - 1\cdot\sum_{m=1}^{M}\frac{\partial H_O^{TL}(\cdot)}{\partial y_i^m}y_i^m + 1\cdot\sum_{k=1}^{K}\frac{\partial H_O^{TL}(\cdot)}{\partial z_i^k}z_i^k = 1\cdot H_O^{TL}(\cdot) \tag{14}$$

where $H_O^{TL}(\cdot) = H_O^{TL}(\mathbf{x}_i,\mathbf{y}_i,\mathbf{z}_i)$.

Or alternatively,

$$\sum_{k=1}^{K}\frac{\partial H_O^{TL}(\cdot)}{\partial z_i^k}\cdot\frac{z_i^k}{H_O^{TL}(\cdot)} - \sum_{m=1}^{M}\frac{\partial H_O^{TL}(\cdot)}{\partial y_i^m}\cdot\frac{y_i^m}{H_O^{TL}(\cdot)} \tag{15}$$

$$= \sum_{k=1}^{K}\frac{\partial \ln H_O^{TL}(\cdot)}{\partial \ln z_i^k} - \sum_{m=1}^{M}\frac{\partial \ln H_O^{TL}(\cdot)}{\partial \ln y_i^m} = 1.$$

Combining (13) and (15) results

$$\sum_{k=1}^{K}\left[\delta_k+\sum_{k'=1}^{K}\theta_{kk'}\ln z_i^{k'}+\sum_{n=1}^{N}\tau_{nk}\ln x_i^n+\sum_{m=1}^{M}\chi_{mk}\ln y_i^m\right] \tag{16}$$

$$-\sum_{m=1}^{M}\left[\gamma_m+\sum_{m'=1}^{M}\eta_{mm'}\ln y_i^{m'}+\sum_{n=1}^{N}\rho_{nm}\ln x_i^n+\sum_{k=1}^{K}\chi_{mk}\ln z_i^k\right]=1,$$

which holds if and only if

$$\sum_{k=1}^{K}\delta_k-\sum_{m=1}^{M}\gamma_m=1 \tag{17}$$

$$\sum_{k'=1}^{K}\theta_{kk'}-\sum_{m=1}^{M}\chi_{mk}=0 \qquad k=1..K \tag{18}$$

$$\sum_{k=1}^{K}\tau_{nk}-\sum_{m=1}^{M}\rho_{nm}=0 \qquad n=1..N \tag{19}$$

$$\sum_{k=1}^{K}\chi_{mk}-\sum_{m'=1}^{M}\eta_{mm'}=0 \qquad m=1..M. \tag{20}$$

Specification TL T *(HEF estimation via Translog: traditional approach)*

If HEF is parameterized as in (13), then parameters of the Translog specification can be estimated by solving the following linear programming problem

$$\min\sum_{i=1}^{I}\ln H_O^{TL}(\mathbf{x}_i,\mathbf{y}_i,\mathbf{z}_i) \tag{21}$$

s.t.

(i) $\ln H_O^{TL}(\mathbf{x}_i,\mathbf{y}_i,\mathbf{z}_i)\geq 0 \qquad i=1..I$

(ii) $\dfrac{\partial \ln H_O^{TL}(\mathbf{x}_i,\mathbf{y}_i,\mathbf{z}_i)}{\partial \ln x_i^n}\geq 0 \qquad n=1..N, i=1..I$

$\dfrac{\partial \ln H_O^{TL}(\mathbf{x}_i,\mathbf{y}_i,\mathbf{z}_i)}{\partial \ln y_i^m}\leq 0 \qquad m=1..M, i=1..I$

$\dfrac{\partial \ln H_O^{TL}(\mathbf{x}_i,\mathbf{y}_i,\mathbf{z}_i)}{\partial \ln z_i^k}\geq 0 \qquad k=1..K, i=1..I$

(iii) $\sum_{k=1}^{K}\delta_k - \sum_{m=1}^{M}\gamma_m = 1$ $\sum_{k'=1}^{K}\theta_{kk'} - \sum_{m=1}^{M}\chi_{mk} = 0 \quad k = 1..K$

$\sum_{k=1}^{K}\tau_{nk} - \sum_{m=1}^{M}\rho_{nm} = 0 \quad n = 1..N$ $\sum_{k=1}^{K}\chi_{mk} - \sum_{m'=1}^{M}\eta_{mm'} = 0 \quad m = 1..M$

(iv) $\zeta_{nn'} = \zeta_{n'n} \quad n,n' = 1..N$ $\eta_{mm'} = \eta_{m'm} \quad m,m' = 1..M$

$\theta_{kk'} = \theta_{k'k} \quad k,k' = 1..K$.

(21) minimizes the linear loss function subject to the constraints that reflect the properties of HEF, namely representation (i), monotonicity in arguments (ii), almost homogeneity (iii) as well as ensure that the Young's theorem holds (iv).

Translog estimates of HEF scores for each data point is then estimated by substituting the solution to (21) along with the input/output data into (13).

4. HEF ESTIMATION ON A MODIFIED TECHNOLOGY SET

In this section I will demonstrate how it is possible to modify the dataset to enable HEF estimation using linear approximation as opposed to nonlinear programming. Furthermore, HEF turns out to be homogenous in arguments on the modified set.

Define a partially *reciprocated technology set* as

$$T^R \equiv \{(\mathbf{x}, \tilde{\mathbf{y}}, \mathbf{z}) : \mathbf{x} \text{ can produce } (\tilde{\mathbf{y}}, \mathbf{z})\}, \tag{22}$$

where $\tilde{\mathbf{y}} \in \Re_+^M$ is a vector consisting of the reciprocal values of $\mathbf{y}$.

The output set is defined similarly to (2)

$$P^R(\mathbf{x}) \equiv \{(\tilde{\mathbf{y}}, \mathbf{z}) : (\mathbf{x}, \tilde{\mathbf{y}}, \mathbf{z}) \in T^R\}. \tag{23}$$

If original technology set T satisfies regularity assumptions A1-A9, then the reciprocated technology satisfies the following.

A[R]3. *Strong disposability of inputs.* If $(\tilde{\mathbf{y}}, \mathbf{z}) \in P^R(\mathbf{x})$ and $\mathbf{x}' \geqq \mathbf{x}$ then $(\tilde{\mathbf{y}}, \mathbf{z}) \in P^R(\mathbf{x}')$.

A^{R}4. *Strong reverse disposability of reciprocated desirable outputs.* If $(\tilde{\mathbf{y}},\mathbf{z}) \in P^R(\mathbf{x})$ and $\tilde{\mathbf{y}}' \geqq \tilde{\mathbf{y}}$ then $(\tilde{\mathbf{y}}',\mathbf{z}) \in P^R(\mathbf{x})$.

A^{R}5. *Hyperbolic disposability of outputs.* If $(\tilde{\mathbf{y}},\mathbf{z}) \in P^R(\mathbf{x})$ then $(\theta^{-1}\tilde{\mathbf{y}},\theta\mathbf{z}) \in P^R(\mathbf{x})$ for all $\theta \leq 1$.

A^{R}7. *Output closedness.* For all $\mathbf{x}$ in $\Re_+^N$, $P^R(\mathbf{x})$ is a closed set.

A^{R}8. *Input convexity.* P^R is quasiconcave on $\Re_+^N$, i.e. for all $\mathbf{x},\mathbf{x}' \in \Re_+^N$ and $0 \leq \lambda \leq 1$, $P^R(\mathbf{x}) \cap P^R(\mathbf{x}') \subseteqq P^R(\lambda\mathbf{x} + (1-\lambda)\mathbf{x}')$.

A^{R}9. *Output convexity.* For all $\mathbf{x}$ in $\Re_+^N$, $P^R(\mathbf{x})$ is a convex set.

Define an *output hyperbolic efficiency function* on $P^R(\mathbf{x})$ as

$$H_O^R(\mathbf{x},\tilde{\mathbf{y}},\mathbf{z}) \equiv \sup\{\lambda : (\lambda^{-1}\tilde{\mathbf{y}},\lambda^{-1}\mathbf{z}) \in P^R(\mathbf{x})\}. \qquad (24)$$

Note that H_O^R looks similar to the classical input distance function (see Färe and Primont, 1995). While the traditional input distance function finds the maximum possible contraction of the input vector in the input space, H_O^R searches for the maximum possible contraction of the modified output vectors in the $(\tilde{\mathbf{y}},\mathbf{z})$ space. The latter is equivalent to multiplication of $\mathbf{y}$ and division of $\mathbf{z}$ by the greatest number possible while still staying inside the technology set, which is identical to the traditional HEF. This identity is postulated in the following proposition.

Proposition 1. $H_O(\mathbf{x},\mathbf{y},\mathbf{z}) = H_O^R(\mathbf{x},\tilde{\mathbf{y}},\mathbf{z})$ for all $\mathbf{x}$ in $\Re_+^N$, $\mathbf{y},\tilde{\mathbf{y}}$ in $\Re_+^M$ and $\mathbf{z}$ in $\Re_+^K$, where $\mathbf{y}$ and $\tilde{\mathbf{y}}$ are as in (22).

Proof: $H_O(\mathbf{x},\mathbf{y},\mathbf{z}) \equiv \sup\{\lambda : (\lambda\mathbf{y},\lambda^{-1}\mathbf{z}) \in P(\mathbf{x})\} = \sup\{\lambda : (\mathbf{x},\lambda\mathbf{y},\lambda^{-1}\mathbf{z}) \in T\}$

$$= \sup\{\lambda : (\mathbf{x},\lambda^{-1}\tilde{\mathbf{y}},\lambda^{-1}\mathbf{z}) \in T^R\} = \sup\{\lambda : (\lambda^{-1}\tilde{\mathbf{y}},\lambda^{-1}\mathbf{z}) \in P^R(\mathbf{x})\}$$

$$= H_O^R(\mathbf{x},\tilde{\mathbf{y}},\mathbf{z})$$

Q.E.D.

Having established the identity between $H_O(\mathbf{x},\mathbf{y},\mathbf{z})$ and $H_O^R(\mathbf{x},\tilde{\mathbf{y}},\mathbf{z})$, it is worthwhile to discuss the properties of $H_O^R(\mathbf{x},\tilde{\mathbf{y}},\mathbf{z})$.

P^{R}1. *Representation.* $H_O^R(\mathbf{x},\tilde{\mathbf{y}},\mathbf{z}) \geq 1$ if and only if $(\tilde{\mathbf{y}},\mathbf{z}) \in P^R(\mathbf{x})$.

P^{R}2. *Homogeneity of degree 1 in* $(\tilde{\mathbf{y}},\mathbf{z})$. $H_O^R(\mathbf{x},\mu\tilde{\mathbf{y}},\mu\mathbf{z}) = \mu H_O^R(\mathbf{x},\tilde{\mathbf{y}},\mathbf{z}) \forall \mu > 0$.

P^{R}3. *Monotonicity in* $\tilde{\mathbf{y}}$. $H_O^R(\mathbf{x},\mu\tilde{\mathbf{y}},\mathbf{z}) \geq H_O^R(\mathbf{x},\tilde{\mathbf{y}},\mathbf{z}) \quad \mu \geq 1$.

P^{R}4. *Monotonicity in* $\mathbf{z}$. $H_O^R(\mathbf{x},\tilde{\mathbf{y}},\mu\mathbf{z}) \geq H_O^R(\mathbf{x},\tilde{\mathbf{y}},\mathbf{z}) \quad \mu \geq 1$.

P^{R}5. *Monotonicity in* $\mathbf{x}$. $H_O^R(\mu\mathbf{x},\tilde{\mathbf{y}},\mathbf{z}) \geq H_O^R(\mathbf{x},\tilde{\mathbf{y}},\mathbf{z}) \quad \mu \geq 1$.

While P^{R}1, P^{R}3-5 follow directly from Proposition 1 and the respective properties of the $H_O(\mathbf{x},\mathbf{y},\mathbf{z})$, P^{R}2 deserves a brief proof.

Proposition 2. (*Homogeneity of degree 1 in* $(\tilde{\mathbf{y}},\mathbf{z})$)

$H_O^R(\mathbf{x},\mu\tilde{\mathbf{y}},\mu\mathbf{z}) = \mu H_O^R(\mathbf{x},\tilde{\mathbf{y}},\mathbf{z}) \forall \mu > 0$.

Proof: $$H_O^R(\mathbf{x},\mu\tilde{\mathbf{y}},\mu\mathbf{z}) \equiv \sup_{\lambda}\left\{\lambda : \left(\frac{\mu\tilde{\mathbf{y}}}{\lambda},\frac{\mu\mathbf{z}}{\lambda}\right) \in P^R(\mathbf{x})\right\} = \sup_{\lambda}\left\{\frac{\mu\lambda}{\mu} : \left(\frac{\mu\tilde{\mathbf{y}}}{\lambda},\frac{\mu\mathbf{z}}{\lambda}\right) \in P^R(\mathbf{x})\right\}$$

$$= \mu \sup_{\lambda/\mu}\left\{\frac{\lambda}{\mu} : \left(\frac{\mu\tilde{\mathbf{y}}}{\lambda},\frac{\mu\mathbf{z}}{\lambda}\right) \in P^R(\mathbf{x})\right\} = \mu H_O^R(\mathbf{x},\tilde{\mathbf{y}},\mathbf{z}). \quad \textit{Q.E.D.}$$

Given these properties, estimation of HEF on $P^R(\mathbf{x})$ is similar in its methodology to the estimation of the input distance function that requires simultaneous contraction of all inputs. Below, I will describe the methodologies for HEF estimation on $P^R(\mathbf{x})$ using both DEA and Translog frontier regression methods.

4.1. DEA SPECIFICATION

The VRS estimator of the reciprocated technology set is defined as

$$T^{R\ VRS} = \{(\mathbf{x},\mathbf{y},\mathbf{z}) : \Psi' X \leq \mathbf{x},\ \Psi'\tilde{Y} \leqq \tilde{\mathbf{y}}, \Psi' Z \leqq \mathbf{z}, \Psi' \geqq \mathbf{0}_I, \mathbf{i}'\Psi = 1\}, \quad (25)$$

where $\tilde{Y}$ is an MxI matrix of the reciprocated values of the desirable outputs.

<u>Specification DEA R</u> (HEF linear approximation via VRS DEA on the reciprocated set)

VRS HEF scores on $P^R(\mathbf{x})$ for DMU *i* can be approximated by solving a linear optimization problem

$$H_O^{R\,DEA}(\mathbf{x}_i,\tilde{\mathbf{y}}_i,\mathbf{z}_i)^{-1} = \min \lambda \tag{26}$$

s.t.

(i) $\Psi' X \le \mathbf{x}_i$ (iv) $\Psi \geq \mathbf{0}_I$

(ii) $\Psi' \tilde{Y} \leq \lambda \tilde{\mathbf{y}}_i$ (v) $\mathbf{i}'\Psi = 1$

(iii) $\Psi' Z \leq \lambda \mathbf{z}_i$

Note that similarly to the estimation of the input distance function (see Färe and Primont, 1995), equation (26) estimates a reciprocated measure of the efficiency score.

4.2. PARAMETRIC SPECIFICATION

Similarly to section 3.2, define HEF on $P^R(\mathbf{x})$ through a Translog function

$$\ln H_O^{RTL}(\mathbf{x}_i,\mathbf{y}_i,\mathbf{z}_i) = \alpha_0 + \sum_{n=1}^{N}\beta_n \ln x_i^n + \sum_{m=1}^{M}\gamma_m \ln \tilde{y}_i^m + \sum_{k=1}^{K}\delta_k \ln z_i^k \tag{27}$$

$$+\frac{1}{2}\sum_{n=1}^{N}\sum_{n'=1}^{N}\zeta_{nn'}\ln x_i^n \ln x_i^{n'} + \frac{1}{2}\sum_{m=1}^{M}\sum_{m'=1}^{M}\eta_{mm'}\ln \tilde{y}_i^m \ln \tilde{y}_i^{m'} + \frac{1}{2}\sum_{k=1}^{K}\sum_{k'=1}^{K}\theta_{kk'}\ln z_i^k \ln z_i^{k'}$$

$$+\sum_{n=1}^{N}\sum_{m=1}^{M}\rho_{nm}\ln x_i^n \ln \tilde{y}_i^m + \sum_{n=1}^{N}\sum_{k=1}^{K}\tau_{nk}\ln x_i^n \ln z_i^k + \sum_{m=1}^{M}\sum_{k=1}^{K}\chi_{mk}\ln \tilde{y}_i^m \ln z_i^k .$$

Homogeneity of degree 1 in $(\tilde{\mathbf{y}},\mathbf{z})$ by Euler equation implies that

$$\sum_{m=1}^{M}\frac{\partial H_O^{RTL}(\cdot)}{\partial \tilde{y}_i^m} y_i^m + \sum_{k=1}^{K}\frac{\partial H_O^{RTL}(\cdot)}{\partial z_i^k} z_i^k = H_O^{TL}(\cdot) \tag{28}$$

and consequently

$$\sum_{m=1}^{M}\frac{\partial \ln H_O^{RTL}(\cdot)}{\partial \ln \tilde{y}_i^m} + \sum_{k=1}^{K}\frac{\partial \ln H_O^{RTL}(\cdot)}{\partial \ln z_i^k} = 1, \tag{29}$$

which after some simple algebra results that

$$\sum_{m=1}^{M}\left[\gamma_m+\sum_{m'=1}^{M}\eta_{mm'}\ln\tilde{y}_i^{m'}+\sum_{n=1}^{N}\rho_{nm}\ln x_i^n+\sum_{k=1}^{K}\chi_{mk}\ln z_i^k\right] \tag{30}$$

$$+\sum_{k=1}^{K}\left[\delta_k+\sum_{k'=1}^{K}\theta_{kk'}\ln z_i^{k'}+\sum_{n=1}^{N}\tau_{nk}\ln x_i^n+\sum_{m=1}^{M}\chi_{mk}\ln\tilde{y}_i^m\right]=1.$$

Equation (30) holds if and only if for every possible data point

$$\sum_{m=1}^{M}\gamma_m+\sum_{k=1}^{K}\delta_k=1 \tag{31}$$

$$\sum_{m=1}^{M}\chi_{mk}+\sum_{k'=1}^{K}\theta_{kk'}=0 \quad k=1..K \tag{32}$$

$$\sum_{m=1}^{M}\rho_{nm}+\sum_{k=1}^{K}\tau_{nk}=0 \quad n=1..N \tag{33}$$

$$\sum_{m'=1}^{M}\eta_{mm'}+\sum_{k=1}^{K}\chi_{mk}=0 \quad m=1..M. \tag{34}$$

<u>Specification TL R</u> *(HEF estimation via Translog on the reciprocated set)*

If HEF is parameterized as in (27), then parameters of the Translog specification can be estimated by solving the following linear programming problem

$$\min\sum_{i=1}^{I}\ln H_O^{RTL}(\mathbf{x}_i,\tilde{\mathbf{y}}_i,\mathbf{z}_i) \tag{35}$$

s.t.

(i) $\ln H_O^{RTL}(\mathbf{x}_i,\tilde{\mathbf{y}}_i,\mathbf{z}_i)\geq 0 \qquad i=1..I$

(ii) $\dfrac{\partial\ln H_O^{RTL}(\mathbf{x}_i,\tilde{\mathbf{y}}_i,\mathbf{z}_i)}{\partial\ln x_i^n}\geq 0 \qquad n=1..N, i=1..I$

$\dfrac{\partial\ln H_O^{RTL}(\mathbf{x}_i,\tilde{\mathbf{y}}_i,\mathbf{z}_i)}{\partial\ln \tilde{y}_i^m}\geq 0 \qquad m=1..M, i=1..I$

$\dfrac{\partial\ln H_O^{RTL}(\mathbf{x}_i,\tilde{\mathbf{y}}_i,\mathbf{z}_i)}{\partial\ln z_i^k}\geq 0 \qquad k=1..K, i=1..I$

(iii) $\sum_{m=1}^{M}\gamma_m + \sum_{k=1}^{K}\delta_k = 1$ $\quad\sum_{m=1}^{M}\chi_{mk} + \sum_{k'=1}^{K}\theta_{kk'} = 0 \quad k = 1..K$

$\sum_{m=1}^{M}\rho_{nm} + \sum_{k=1}^{K}\tau_{nk} = 0 \quad n = 1..N$ $\quad\sum_{m'=1}^{M}\eta_{mm'} + \sum_{k=1}^{K}\chi_{mk} = 0 \quad m = 1..M$

(iv) $\zeta_{nn'} = \zeta_{n'n} \quad n, n' = 1..N$ $\quad\eta_{mm'} = \eta_{m'm} \quad m, m' = 1..M$

$\theta_{kk'} = \theta_{k'k} \quad k, k' = 1..K$.

(35) minimizes the linear loss function subject to the constraints that reflect the properties of HEF on $P^R(\mathbf{x})$, namely representation (i), monotonicity in arguments (ii), homogeneity in outputs (iii) as well as ensure that the Young's theorem holds (iv).

Translog estimates of HEF scores for each data point is then computed by substituting the solution to (35) along with the input/output data into (28).

It is easy to see that since $\ln\tilde{y}_i^m = -\ln y_i^m$ for all *i* and *m*, optimal parameters found by (21) are identical to the solution to (35) except of δ's, ρ's and χ's (i.e. parameters of the terms containing $\ln\tilde{y}_i^m$, but not $(\ln\tilde{y}_i^m)^2$, which differ by the sign only. Consequently, the efficiency score estimates found by (21) and (35) are identical.

5. STATISTICAL INFERENCES

Statistical properties of the frontier estimates do not have an analytic formulation in most of the cases (notable exceptions include, for example, early attempts to estimate production frontiers by OLS regressions). As in other cases, when analytic results are not comforting, bootstrap (Efron, 1979 and Efron and Tibsharani, 1993) is an attractive alternative for making inferences.

Bootstrapping relies on the repeated simulation of the data generating process (usually by resampling) and applying the original estimator to every simulated sample so that resulting estimates mimic the sampling distribution of the original estimator. I use the smooth homogeneous bootstrap (SHB) methodology proposed by Simar and Wilson (1998) to obtain statistical properties of the frontier estimates. SHB was proposed as an alternative to the naïve bootstrap, which was shown to give inconsistent results in the frontier problems (Simar and Wilson, 1998). While the naïve bootstrap draws a bootstrap sample from a discrete distribution, SHB uses smoothing based on the kernel estimated densities of the obtained efficiencies to draw a sample from a continuous distribution. In other words, SHB perturbs DMUs' data around their observed values to create a bootstrap sample. Smooth resampling is being repeated B times to obtain bootstrap samples' estimates of the original estimators. I use B=1,000 as Hall (1986) suggests to ensure the adequate coverage of the confidence intervals. The bootstrap estimates are used to obtain bias-corrected estimates of the efficiency scores, their standard errors and confidence intervals.

5.1. SHB PROCEDURE

Simar and Wilson (1998, 2007) multistep procedure for obtaining statistical inference using SHB can be adapted for the aims of the current studies as following.

Along with the starting data on inputs and outputs for each of the DMUs, estimation of each of the specifications provides one a vector of the estimated efficiency scores, say, $\hat{\Lambda} = \{\hat{\lambda}_1, ..., \hat{\lambda}_I\}$. Since a regular kernel estimate does not take into account that $\hat{\Lambda}$ is bounded at 1, with any nonzero bandwidth, a regular kernel suffers from bias in the neighbourhood of unity. Silverman (1986) proposed to solve this problem by reflecting the values of $\hat{\Lambda}$ by constructing a reflected matrix $L = \{\hat{\Lambda}, 2\mathbf{i} - \hat{\Lambda}\}$, which consists of the original vector $\hat{\Lambda}$ and its values reflected around the unity. Simar and Wilson (2007) note that $\hat{\Lambda}$ contains some spurious values equal to 1, which provides with the spurious mass greater than $1/I$ at the boundary value in the discrete density to be smoothed. These values are merely an artefact of the deterministic efficiency analysis and may be excluded for the purpose of selecting a bandwidth.

Step 1. Calculate bandwidth, h, according to the Silverman's adaptive rule

$$h = 1.06 \min\left\{\sigma_L, \frac{iqr(L)}{1.349}\right\} N^{-0.2}, \tag{36}$$

where σ_L is standard error of L; and $iqr(L)$ is its interquartile range.

Step 2. Draw a random sample $\mathrm{B}^* = \{\beta_1^*, ..., \beta_I^*\}$ with replacements from $\hat{\Lambda}$.

Step 3. Calculate $\tilde{\Lambda}^* = \{\tilde{\lambda}_1^*, ..., \tilde{\lambda}_I^*\}$ as

$$\tilde{\lambda}_i^* = \begin{cases} \beta_i^* + h\varepsilon_i^* & \text{if } \beta_i^* + h\varepsilon_i^* \geq 1 \\ 2 - \left(\beta_i^* + h\varepsilon_i^*\right) & \text{otherwise} \end{cases} \tag{37}$$

where ε_i^* is a random deviate drawn from a standard normal distribution, i.e. $\varepsilon_i^* \sim N(0,1)$.

Step 4. As typical when kernel estimators are used, the variance of the bootstrap generated sequence must be corrected by calculating $\Lambda^* = \{\lambda_1^*, ..., \lambda_I^*\}$

$$\lambda_i^* = \overline{\beta}^* + \left(1 + \frac{h^2}{\hat{\sigma}_\lambda^2}\right)^{-\frac{1}{2}} \left(\tilde{\lambda}_i^* - \overline{\beta}^*\right), \tag{38}$$

where $\hat{\sigma}_\lambda^2$ is a sample deviation of $\hat{\Lambda}$; $\overline{\beta}^*$ is a sample mean of B^*.

Step 5. Perturb the original data $\{X, Y, Z\}$ to create a bootstrap sample $\{X_b^*, Y_b^*, Z_b^*\}$ as

$$X_b^* = X, \tag{39}$$

$$Y_b^* = \left\{\frac{\hat{\lambda}_1 \mathbf{y}_1}{\lambda_1^*}, ..., \frac{\hat{\lambda}_I \mathbf{y}_I}{\lambda_I^*}\right\}, \tag{40}$$

$$Z_b^* = \left\{\frac{\lambda_1^* \mathbf{z}_1}{\hat{\lambda}_1}, ..., \frac{\lambda_I^* \mathbf{z}_I}{\hat{\lambda}_I}\right\}. \tag{41}$$

(39)-(41) projects each observation to its estimated efficient peer using the estimate of the efficiency and then projects it off the frontier using a random efficiency score drawn from the smooth kernel density estimate of the score distribution.

Step 6. Using the estimator defined by the specification being analyzed and the bootstrap sample $\{X_b^*, Y_b^*, Z_b^*\}$ as a reference set, obtain bootstrap estimates of the efficiency scores $\hat{\Lambda}_b^* = \{\hat{\lambda}_{b1}^*, ..., \hat{\lambda}_{bI}^*\}$ for each of the original points $\{X_i, Y_i, Z_i\}$.

5.2. SHB STATISTICAL INFERENCE

The SHB procedure generates a set of B bootstrap estimates of the efficiency scores $\{\hat{\Lambda}^*_1,...,\hat{\Lambda}^*_B\}$. Then bootstrap estimated bias of $\hat{\Lambda}^*$, $Bias(\hat{\Lambda}^*)$ is

$$Bias(\hat{\Lambda}^*) = \frac{1}{B}\sum_{b=1}^{B}\left[\hat{\Lambda} - \hat{\Lambda}^*_b\right] = \hat{\Lambda} - \frac{1}{B}\sum_{b=1}^{B}\hat{\Lambda}^*_b . \tag{42}$$

As usual in the bootstrap literature, I assume that the relationship between the original sample (pseudopopulation) and the bootstrap sample mimics the relationship between the true population and the original sample. Therefore, $Bias(\hat{\Lambda}^*) = Bias(\hat{\Lambda})$, which results bias corrected values of $\hat{\Lambda}$, $\hat{\Lambda}^{BC} = \{\hat{\lambda}_i^{BC},...,\hat{\lambda}_I^{BC}\}$

$$\hat{\Lambda}^{BC} = \max\{\hat{\Lambda} + Bias(\hat{\Lambda}), \quad 1\} = \max\left\{2\hat{\Lambda} - \frac{1}{B}\sum_{b=1}^{B}\hat{\Lambda}^*_b, \quad 1\right\}. \tag{43}$$

The *max* operator in (43) ensures that $\hat{\Lambda}^{BC}$ is bounded by unity from below.

Bias corrected estimates of the efficiency scores are unbiased, i.e.

$$E\left[\hat{\lambda}_i^{BC}\right] = \lambda_i \tag{44}$$

Upper bound, $UB(\hat{\Lambda}^{BC})$, and lower bound, $LB(\hat{\Lambda}^{BC})$, of the biased corrected estimates' 95% confidence intervals are computed by finding the respective bounds of $\hat{\Lambda}^*$. Let $\hat{\Lambda}^*_G$ be a (BxI) matrix consisting of all bootstrap estimated $\hat{\Lambda}^*_b$, $b = 1,...,B$. Further, let $\hat{\Lambda}^*_R$ be a (BxI) matrix obtained from $\hat{\Lambda}^*_G$ by ranking elements in each column from the highest to the lowest. Then, the upper bound, $\hat{\Lambda}^*_{UB}$ is the (0.025B)th row of $\hat{\Lambda}^*_R$, while the lower bound, $\hat{\Lambda}^*_{LB}$ is the (0.975B)th row of $\hat{\Lambda}^*_R$. Therefore,

$$UB(\hat{\Lambda}^{BC}) = \hat{\Lambda}^{BC} + \left(\hat{\Lambda}^*_{UB} - \frac{1}{B}\sum_{b=1}^{B}\hat{\Lambda}^*_b\right) \tag{45}$$

$$LB(\hat{\Lambda}^{BC}) = \max\left\{\hat{\Lambda}^{BC} + \left(\hat{\Lambda}^*_{LB} - \frac{1}{B}\sum_{b=1}^{B}\hat{\Lambda}^*_b\right), \quad 1\right\}. \tag{46}$$

6. DATA GENERATING PROCESS AND FRONTIER COMPARISON BENCHMARKS

In this study I assume that the true technology is known and produces one desirable output and one undesirable output using single input. The variables along the best practice frontier (e.g., efficient values) are related to one another by production function

$$y^* = x^{1/2} z^{*1/2}, \tag{47}$$

where y^* is an efficient quantity of desirable output, z^* – efficient quantity of undesirable output and x – input.

I generate the quantities of x and z^* by drawing three separate samples with sizes 20, 50 and 100 respectively from a uniform distribution $x_i \sim U[0,100]$, $z_i^* \sim U[0,100]$. Efficient quantities of the desirable outputs are then obtained by (47).

I assume that the "true' hyperbolic efficiency score λ is

$$\lambda_i = \exp(u_i), \tag{48}$$

where $u_i \sim Exp\left(1/3\right)$, so that $E[\lambda] = 4/3$, a reasonable scenario in the productivity analysis literature (Gijbels, Mammen, Park and Simar, 1999).

The inefficient levels of desirable and undesirable outputs are then obtained as

$$y_i = \lambda_i^{-1} y_i^* \tag{49}$$

$$z_i = \lambda_i z_i^* \tag{50}$$

where y and z are inefficient quantities of the desirable and undesirable outputs.

A good frontier estimator should be able to recover the values of the "true" inefficiency terms $\{\lambda_1,\ldots,\lambda_I\}$ and, consequently, the values of the efficient peers for each of the DMUs $\{(x_1, y_1^*, z_1^*),\ldots,(x_I, y_I^*, z_I^*)\}$ from the off-frontier data points $\{(x_1, y_1, z_1),\ldots,(x_I, y_I, z_I)\}$.

Define simulated efficient peer's triplet for DMU i $(\hat{x}_i, \hat{y}_i^*, \hat{z}_i^*)$ as

$$\hat{x}_i = x_i, \tag{51}$$

$$\hat{y}_i^* = \hat{\lambda}_i^{BC} y_i, \tag{52}$$

$$\hat{z}_i^* = z_i \Big/ \hat{\lambda}_i^{BC}. \tag{53}$$

Similarly to FMV (2006), the first benchmark is *the average Euclidian distance* between the true efficient peers $\{(x_1, y_1^*, z_1^*), ..., (x_I, y_I^*, z_I^*)\}$ and the simulated peers $\{(\hat{x}_1, \hat{y}_1^*, \hat{z}_1^*), ..., (\hat{x}_I, \hat{y}_I^*, \hat{z}_I^*)\}$, $\overline{\Delta}$, defined as

$$\overline{\Delta} = \frac{1}{I}\sum_{i=1}^{I}\sqrt{(y_i^* - \hat{y}_i^*)^2 + (z_i^* - \hat{z}_i^*)^2}. \tag{54}$$

Unlike FMV (2006), who relied on the average shadow price discrepancy and the mean Euclidean distance between the true and estimated Morishima elasticities of substitution (Morishima, 1967), I do not rely on the derivative-based benchmarks. Instead, I rely on the criteria of unbiasedness and efficiency of the estimators to define the following two benchmarks.

Unbiasedness of an estimator is evaluated by the average discrepancy between the true values of efficiencies and the estimated bias-corrected values, $\overline{\Omega}(\hat{\Lambda}^{BC})$

$$\overline{\Omega}(\hat{\Lambda}^{BC}) = \frac{1}{I}\sum_{i=1}^{I}(\hat{\lambda}_i^{BC} - E[\hat{\lambda}_i^{BC}]) = \frac{1}{I}\sum_{i=1}^{I}(\hat{\lambda}_i^{BC} - \lambda_i). \tag{55}$$

If $\overline{\Omega}(\hat{\Lambda}^{BC}) < 0$, the estimator, on average, underestimates true efficiencies and, as a result, the estimated frontier, on average, lies below the true frontier. If $\overline{\Omega}(\hat{\Lambda}^{BC}) > 0$, the estimator, on average, overestimates true efficiencies and the estimated frontier, on average, is above the true frontier. The least biased estimator should have the smallest (in absolute terms) $\overline{\Omega}(\hat{\Lambda}^{BC})$.

Table 2-1. Descriptive statistics of the simulated datasets

	X	Y	Z	λ	DEA T	DEA LA	DEA R	TL T	TL R
N=100									
mean	44.2592	34.0720	76.4650	1.3659	1.2843	1.2161	1.2518	1.3406	1.3406
st. dev	29.7302	20.4642	49.2036	0.3600	0.3420	0.2293	0.3269	0.3605	0.3605
max	99.5908	86.4618	238.3247	2.4294	2.2374	1.7642	2.2353	2.3915	2.3915
min	0.2054	1.5196	0.5488	1.0020	1.0000	1.0000	1.0000	1.0000	1.0000
time					1222.7	9.3	10.1	6.8	6.8
N=50									
mean	46.8064	32.5538	70.8071	1.3772	1.2699	1.2079	1.2067	1.3481	1.3481
st. dev	29.5946	18.5904	44.646	0.3668	0.3352	0.2282	0.2969	0.3623	0.3623
max	94.1532	75.8103	167.9646	2.4519	2.2654	1.7026	2.1296	2.4077	2.4077
min	1.7318	2.534	2.4549	1.0121	1.0000	1.0000	1.0000	1.0000	1.0000
time					103.9	1.9	2.0	1.4	1.4
N=20									
mean	54.8163	42.6058	84.7530	1.3461	1.1628	1.1277	1.1319	1.2015	1.2015
st. dev	30.7573	20.9900	46.4900	0.3418	0.2419	0.1777	0.2057	0.2516	0.2516
max	99.5423	85.7896	170.3358	2.0879	1.7449	1.4781	1.6155	1.7629	1.7629
min	5.6178	5.4295	0.6870	1.0289	1.0000	1.0000	1.0000	1.0000	1.0000
time					5.4	0.5	0.7	0.3	0.3

Efficiency of an estimator is evaluated by the mean squared deviation of the estimated and the true efficiencies

$$\overline{\Phi}\left(\hat{\Lambda}^{BC}\right)=\frac{1}{I}\sum_{i=1}^{I}\left(\hat{\lambda}_i^{BC}-E\left[\hat{\lambda}_i^{BC}\right]\right)^2=\frac{1}{I}\sum_{i=1}^{I}\left(\hat{\lambda}_i^{BC}-\lambda_i\right)^2. \qquad (56)$$

The smallest $\overline{\Phi}\left(\hat{\Lambda}^{BC}\right)$ will be demonstrated with the most efficient estimator.

7. MONTE CARLO SIMULATION RESULTS

Simulation of the data provides with series of input (X), desirable output (Y) and undesirable output (Z) data along with the data on the true efficiency scores (λ). Estimation involving five techniques described above gives five series of the efficiency estimates for each of the samples. Table 2-1 provides the descriptive statistics of the generated samples and the efficiency estimates. In addition, it documents the time taken to estimate the efficiencies using each of the techniques[5].

Note that although efficient values of Zs are generated from a uniform distribution between 0 and 100, the reference set (inefficient) values are obtained by

[5] Estimation time is given for the code executed in MatLab 6.0 using AMD Athlon 64 processor 1.2 GHz, 1.12Gb RAM.

Table 2-2. Benchmarks for the bias uncorrected estimates (U) and bias corrected (BC) estimates

Benchmark	DEA T		DEA LA		DEA R		TL T/R	
	U	BC	U	BC	U	BC	U	BC
N=100								
$\overline{\Delta}$	0.4949	0.2719	1.0541	0.8712	0.6614	0.4169	0.1192	0.0921
$\overline{\Omega}$	-0.0815	-0.0044	-0.1497	-0.0758	-0.1141	-0.0002	-0.0253	-0.0062
$\overline{\Phi}$	0.0141	0.0048	0.0470	0.0313	0.0276	0.0121	0.0024	0.0012
$\sqrt{\overline{\Phi}}$	0.1188	0.0693	0.2168	0.1770	0.1662	0.1105	0.0491	0.0347
N=50								
$\overline{\Delta}$	1.3688	0.7764	1.7354	1.1831	1.8082	1.0344	0.2497	0.1980
$\overline{\Omega}$	-0.1073	-0.0066	-0.1694	-0.0765	-0.1705	-0.0265	-0.0291	0.0209
$\overline{\Phi}$	0.0203	0.0072	0.0556	0.0354	0.0474	0.0174	0.0016	0.0011
$\sqrt{\overline{\Phi}}$	0.1424	0.0846	0.2357	0.1881	0.2176	0.1321	0.0394	0.0334
N=20								
$\overline{\Delta}$	2.9525	2.3756	3.3736	2.6273	3.2219	2.4236	2.4351	1.8360
$\overline{\Omega}$	-0.1833	-0.0894	-0.2183	-0.1228	-0.2142	-0.0918	-0.1445	-0.0823
$\overline{\Phi}$	0.1086	0.0739	0.1244	0.0768	0.1193	0.0776	0.0897	0.0588
$\sqrt{\overline{\Phi}}$	0.3296	0.2719	0.3527	0.2771	0.3453	0.2786	0.2995	0.2426

multiplying Zs by the efficiency scores. Therefore, the reference set values of Zs may be greater than 100.

As expected, Translog estimates do not depend on the reference set (unreciprocated versus reciprocated). Therefore, I will provide common statistics for these two sets of estimates from now on. Both Translog and nonparametric estimates are subject to downward bias in the samples of any size. In addition, one can note that the traditional DEA estimate on the unreciprocated reference set takes 10 to 120 times longer than either its linear approximation or estimation on the reciprocated reference set[6].

To demonstrate the improvement in the estimates resulting from the smooth homogeneous bootstrap bias correction, I present the performance benchmarks of both the uncorrected and bias corrected estimates. These are provided in Table 2-2.

[6] It's logical to expect that a 1,000 iteration bootstrap of the traditional DEA approach should take just under 17 hours for, say, N=50. Yet, due to computational complexity of the nonlinear optimization problems resulting in dumping of memory and the CPU overload, this bootstrap took over 30 hours.

As the table suggests, without correction for bias Translog specification performs very well in the large and medium-size samples, while the DEA specification performs well in the large samples and shows relatively poor performance in the medium-size and small samples. Nonlinear DEA specification on the unreciprocated reference set is the most precise out of the nonparametric alternatives before correction for bias; approximation of the DEA scores on the reciprocated set provides with the estimates that are slightly less accurate than the traditional DEA ones (but require less computing resources) and are more precise than the linearly approximated ones.

Notably, nonparametric bootstrap methods developed for the nonparametric DEA estimators demonstrate good performance when applied to the parametric Translog estimator. This is largely due to the fact that the SHB's difference from the traditional bootstrap techniques is in accounting for the existence of the frontier—a fact independent of the estimation technique.

Figures 2-2 through 2-13 depict HEF (uncorrected and bias-corrected) estimates against true efficiency scores for various sample sizes. The 45 degree line indicates the exact estimation. If a diamond of a particular DMU is located above the 45 degree line then the efficiency score for that DMU is overestimated by the estimator and is underestimated otherwise. In addition, they also depict UB and LB of the bias-corrected estimates against true HEF scores. An ideal estimator should provide with the upper bound values above the 45 degrees line and the lower bound values below the line.

The diagrams suggest that all DEA estimators as well as the Translog parametric estimator underestimate the true efficiencies in samples of any size. In addition, it can be clearly seen that the DEA LA estimator, while accurate enough in estimating efficiencies of relatively more efficient DMUs, does not perform well in estimating efficiencies of the DMUs located further from the frontier. This finding is not surprising as the DEA LA estimator linearly approximates HEF around unity – i.e. the best practice frontier.

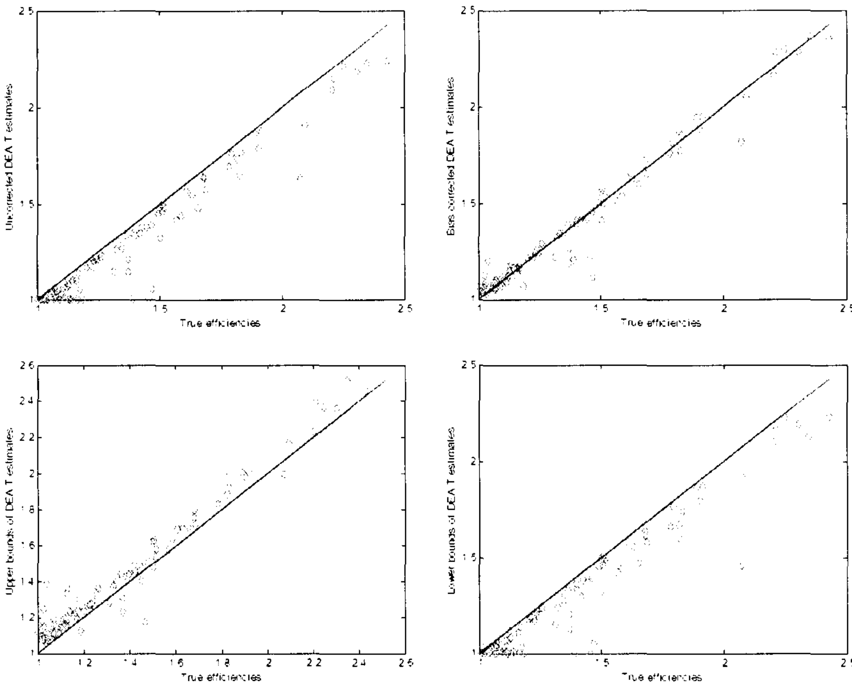

Figure 2-2. Goodness of fit of the DEA T HEF estimates and the their confidence intervals, N=100

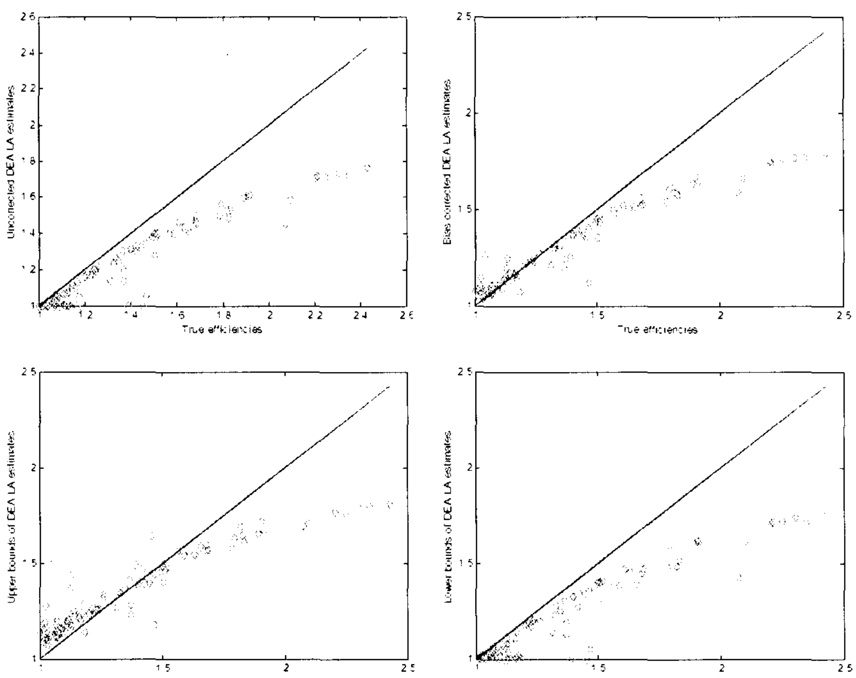

Figure 2-3. Goodness of fit of the DEA LA HEF estimates and the their confidence intervals, N=100

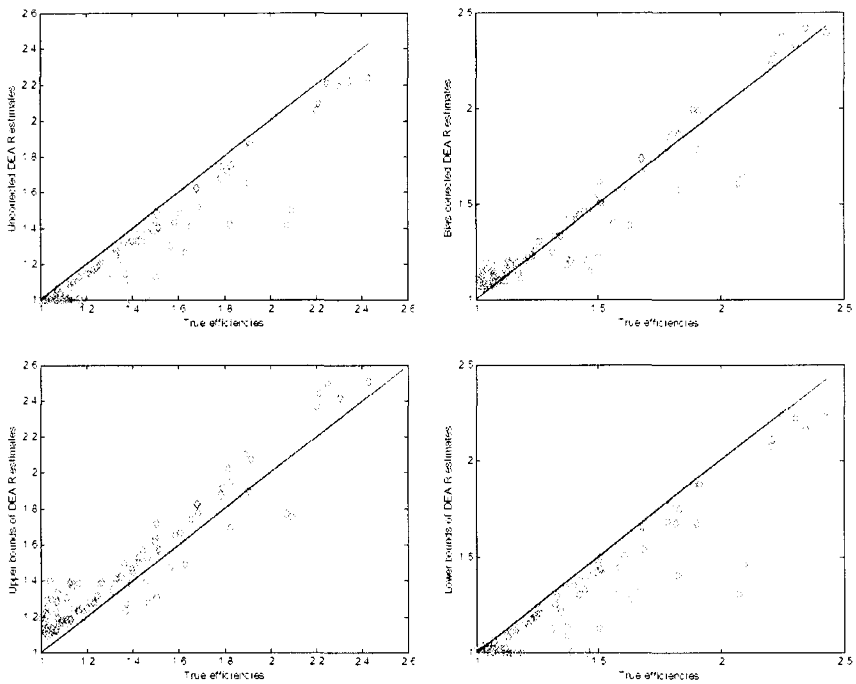

Figure 2-4. Goodness of fit of the DEA R HEF estimates and the their confidence intervals, N=100

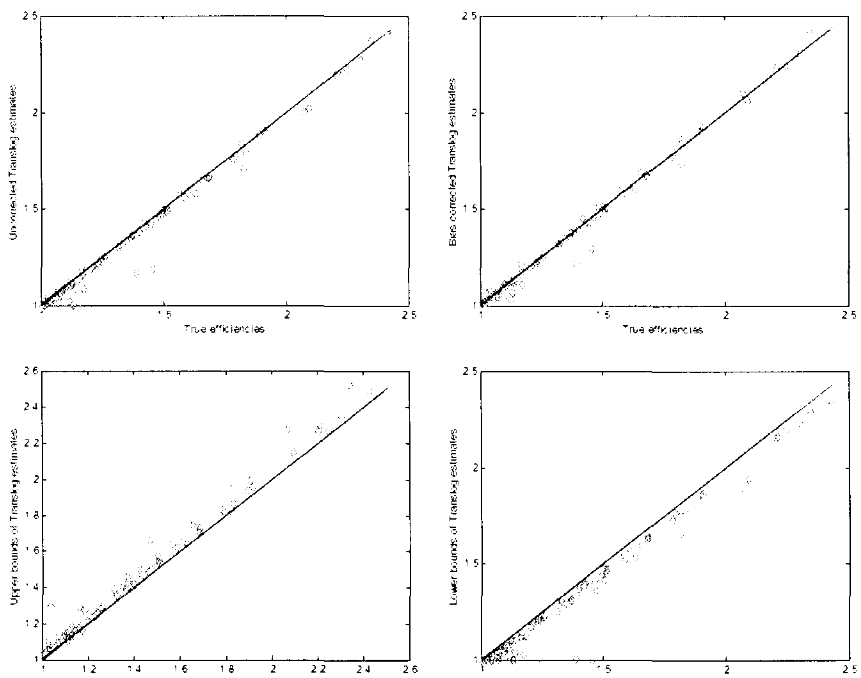

Figure 2-5. Goodness of fit of the Translog HEF estimates and the their confidence intervals, N=100

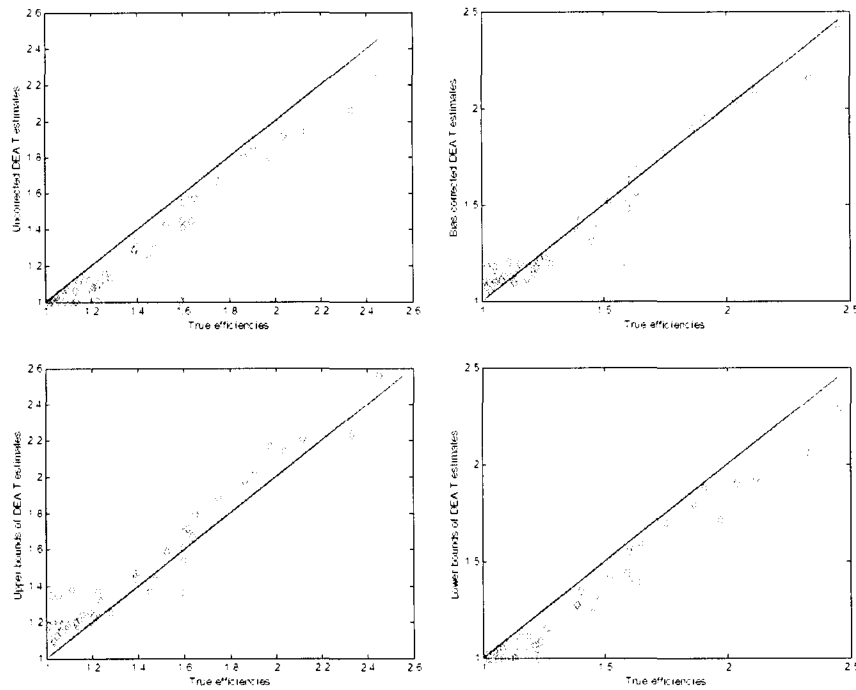

Figure 2-6. Goodness of fit of the DEA T HEF estimates and the their confidence intervals, N=50

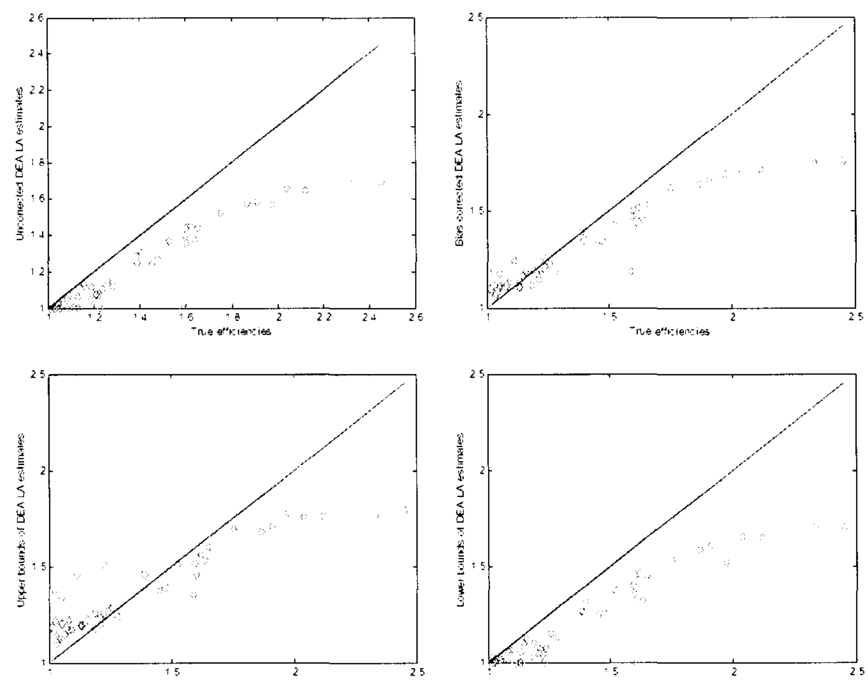

Figure 2-7. Goodness of fit of the DEA LA HEF estimates and the their confidence intervals, N=50

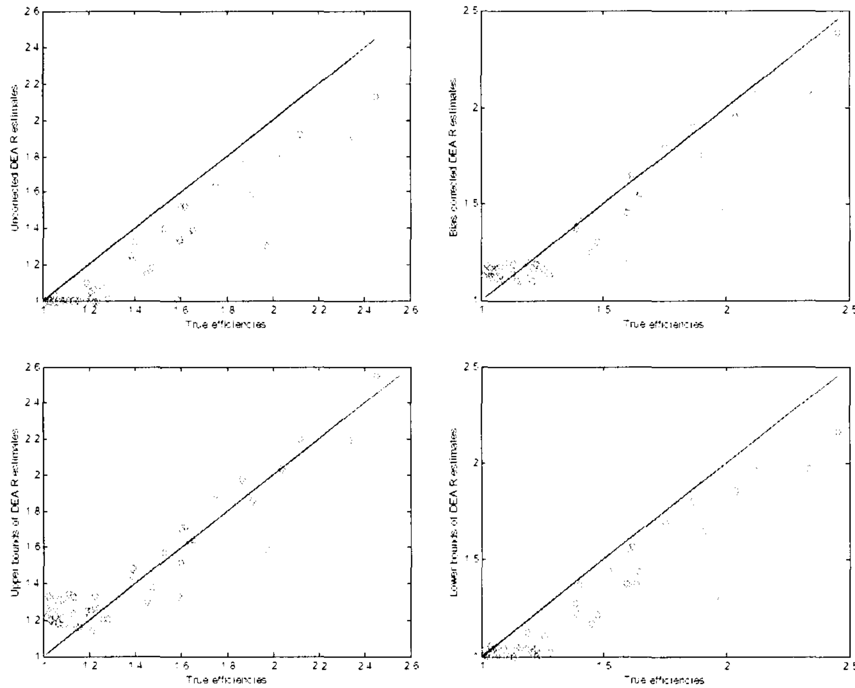

Figure 2-8. Goodness of fit of the DEA R HEF estimates and the their confidence intervals, N=50

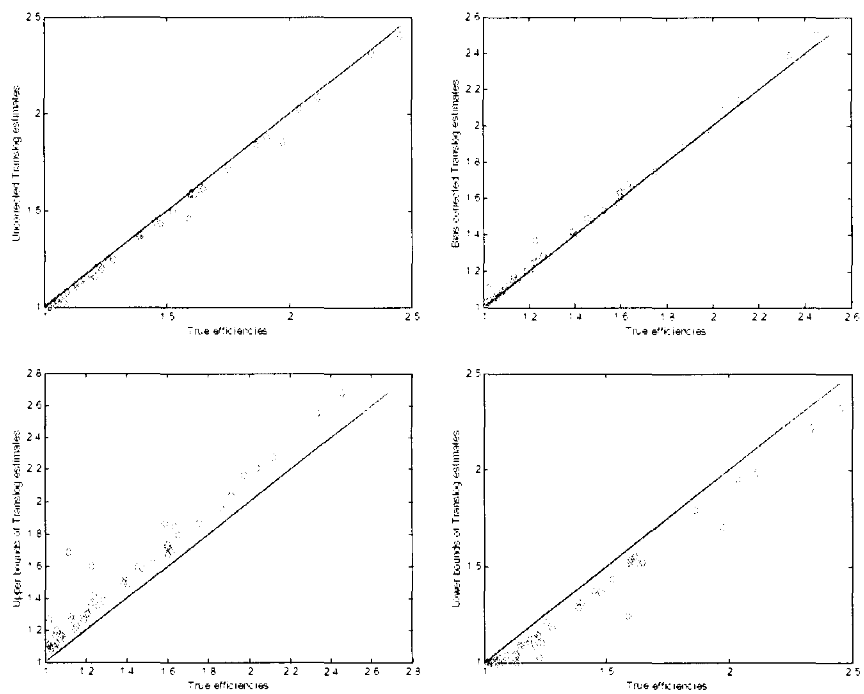

Figure 2-9. Goodness of fit of the Translog HEF estimates and the their confidence intervals, N=50

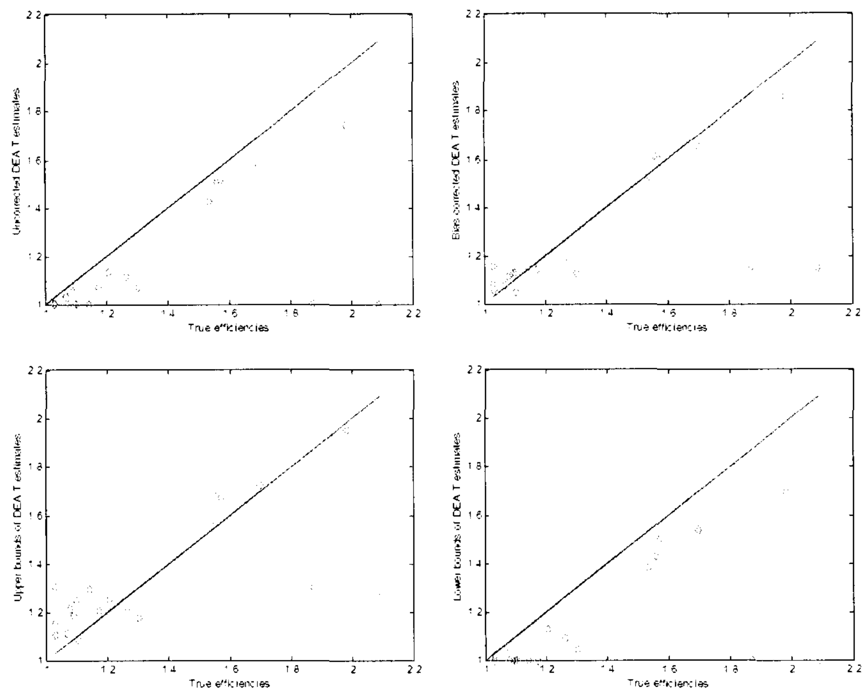

Figure 2-10. Goodness of fit of the DEA T HEF estimates and the their confidence intervals, N=20

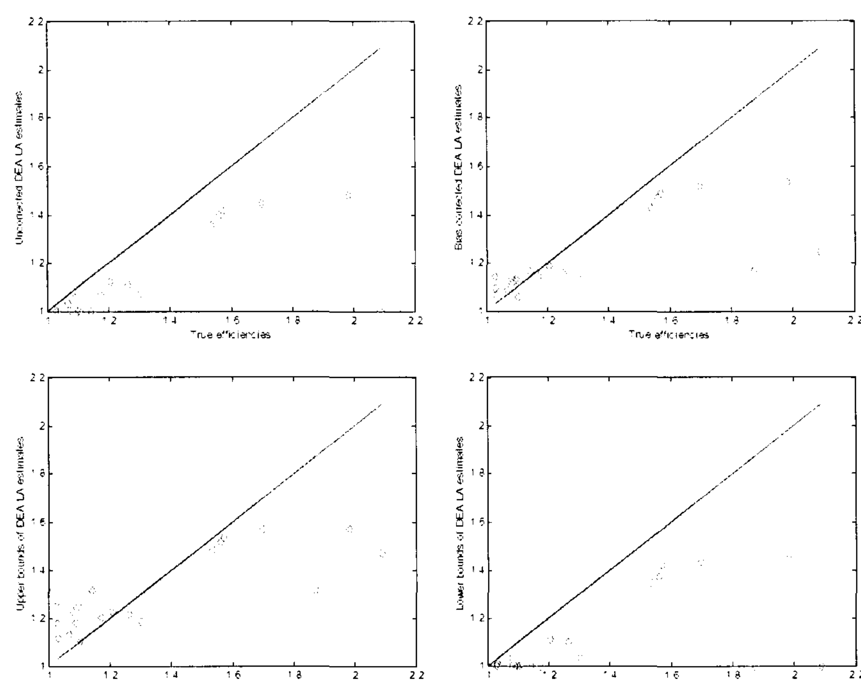

Figure 2-11. Goodness of fit of the DEA LA HEF estimates and the their confidence intervals, N=20

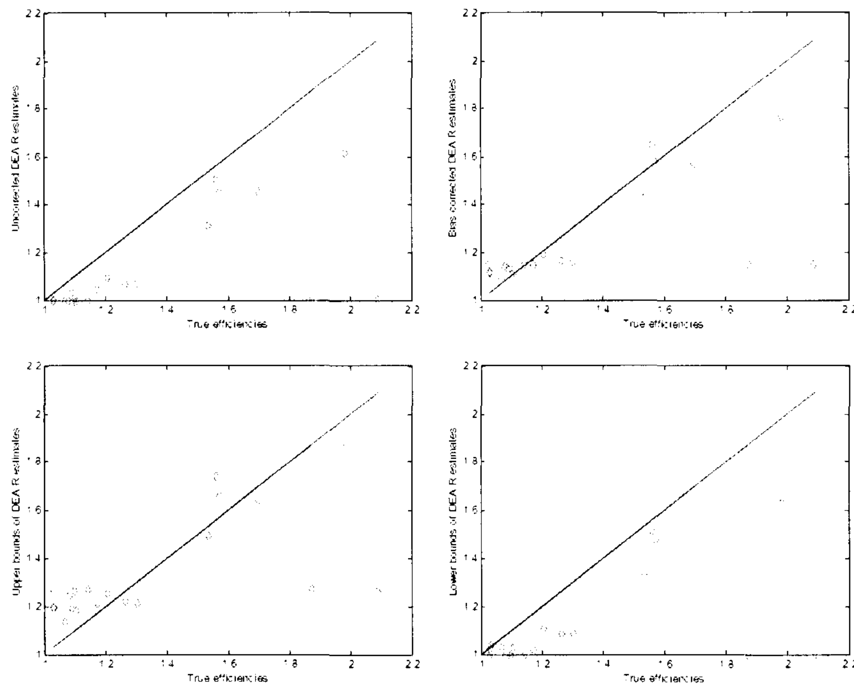

Figure 2-12. Goodness of fit of the DEA R HEF estimates and the their confidence intervals, N=20

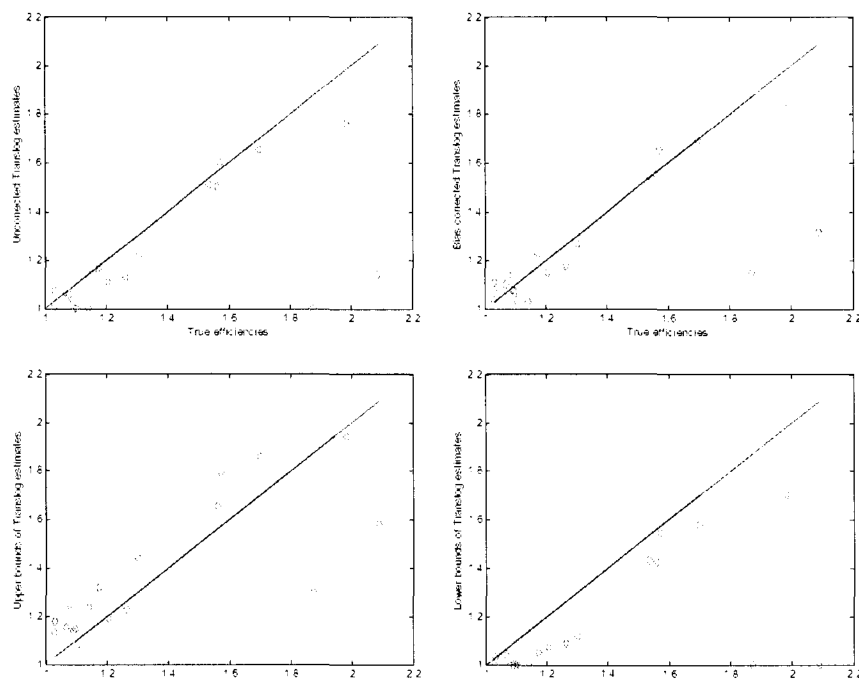

Figure 2-13. Goodness of fit of the Translog HEF estimates and the their confidence intervals, N=20

The diagrams demonstrate that the DEA estimator on the unreciprocated reference set is the most precise of the nonparametric ones. Yet, it may be more suitable to use a linear approximation of DEA estimator on the reciprocated reference set, which is marginally less precise but is substantially less computationally demanding. Depending on the sample size, this estimator requires up to 120 times less computational time than the traditional nonlinear DEA HEF estimator.

The Translog parametric estimator clearly outperforms the nonparametric ones in terms of both the Euclidian distance benchmark and deviations of the estimates from the true efficiency scores. The Translog specification shows marginally worse results in terms of the average bias than the traditional DEA specification.

As it follows from the diagrams, most of the accuracy loss in very small samples occurs because of the relatively bigger part of the sample being on the estimated best practice frontier.

The smooth homogeneous bootstrap allows for obtaining a set of bias-corrected (BC) efficiency scores, along with their confidence intervals represented by the upper and lower bounds (UB and LB respectively) for each DMU and each estimation technique. Even in small samples all estimators correctly identify confidence intervals for almost all DMUs. Notably, LBs are below the true efficiency scores for all estimators and for each DMU regardless of the sample size. UB is often incorrectly identified for very inefficient DMUs. Specifically, DEA LA estimator particularly suffers with underestimating UBs for very inefficient DMUs.

Unsurprisingly, the only approach that gives realistic results in terms of the upper and lower bounds is the Translog parametric. On the nonparametric side, DEA technique on the unmodified reference set gives the most realistic confidence intervals. For the medium size sample, 44 out of 50 estimated confidence intervals include the true efficiency scores and the rest of the confidence intervals somewhat underestimate the true scores. The modification of the reference set leads to some loss in the goodness of the fit of the DEA approach with 38 out of 50 confidence intervals including the true efficiency scores, 2 overestimating the true scores and 10 underestimating them. The traditional linear approximation technique provides realistic confidence intervals when the true efficiency scores are low, yet suffers from a substantial downward bias when the true scores are high.

8. Conclusions

In this paper I have reviewed the popular frameworks for estimating the production frontiers when some outputs are undesirable via the Hyperbolic Efficiency Function – a hybrid of traditional efficiency measures that allows for treating desirable and undesirable outputs asymmetrically. The inherent downward bias of the frontier estimators was addressed by using the smooth homogeneous bootstrap technique – a variation of the conventional bootstrap that takes into account the existence of the frontier. Bootstrap bias-corrected estimates of the efficiency scores and the respective confidence intervals were tested against three benchmarks and the true efficiency scores. Notably, I find that the smooth homogeneous bootstrap originally designed for the nonparametric framework can be applied to parametric problems to substantially improve their estimates' fit.

Unsurprisingly, both parametric and nonparametric techniques give similarly accurate results in large samples both before and after bias correction. Reducing the sample size causes accuracy loss for all estimators. Notably, I find that in medium size samples DEA estimators lose much accuracy both before and after bias correction.

In very small samples both parametric and nonparametric estimators do not perform accurately, primarily due to the larger proportion of DMUs being on the estimated best practice frontier. It turns out that while bias correction procedure can, with relative accuracy, cure the off-frontier DMUs' bias, it demonstrates poor performance when dealing with the incorrectly estimated efficient DMUs. Interestingly, I find that in very small samples the smooth homogeneous bootstrap bias correction performs equally well (if not better) when applied to the parametric estimator as opposed to the DEA estimators.

Despite being relatively inaccurate in small samples, all estimators provide correct confidence intervals for relatively efficient DMUs, but often underestimate the upper bounds of the confidence intervals for relatively inefficient DMUs.

I propose an alternative approach to nonparametric DEA estimation of the linearly approximated hyperbolic efficiency. This is a way to avoid computationally demanding nonlinear programming which may be especially crucial when a large sample and/or multiple iterations are involved. The proposed approach performs substantially better than the previously proposed linear approximation estimator based on the first order linear approximation about the frontier. Most of the performance gain is achieved for the least efficient decision making units.

I conclude that the bias-corrected Translog parametric estimator should be preferred to the uncorrected Translog estimator due to the smaller inward bias of the former. An additional advantage of this estimator – its differentiability – makes it particularly attractive for studies involving shadow prices estimation.

In studies which require nonparametric estimators, traditional bias-corrected DEA estimates provide a relatively good fit, but require substantially more computing resources. The computational time may be significantly reduced by modifying the reference set and sacrificing the precision marginally.

Finally, I would like to point out that the results of this paper were based on the deterministic data generating process and can be extrapolated only to the environments when statistical noise is either absent or can be neglected. Performance of the SHB in a stochastic environment may differ from the one in deterministic environment and can be a subject of future studies.

BIBLIOGRAPHY

Aczél, Janos. 1966. *Lectures on Functional Equations and Their Applications*. New York: Academic Press.

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

Aigner, Dennis J., Shih-fan Chu. 1968. On Estimating the Industry Production Function. *The American Economic Review* 58(4): 826-839.

Aigner, Dennis J., Lovell, C. A. Knox, and Peter Schmidt. 1977. Formulation and Estimation of Stochastic Frontier Production Function Models. *Journal of Econometrics* 6(1): 21-37.

Boles, J.N. 1966. Efficiency Squared-Efficient Computation of Efficiency Indexes. *Proceedings of the Thirty Ninth Annual Meeting of the Western Farm Economics Association* 137-142.

Chambers, Robert G. 1988. Applied Production Analysis: a Dual Approach. New York: Cambridge University Press.

Chambers, Robert G., Yangho H. Chung, and Rolf Färe. 1996. Benefit and Distance Functions. *Journal of Economic Theory* 70(2): 407-419.

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

Christensen, Laurits R., Dale W. Jorgenson, and Lawrence J. Lau. 1973. Transcendental Logarithmic Production Frontiers. *The review of economics and statistics* 55(1): 28-45.

Christensen, Laurits R., Dale W. Jorgenson, and Lawrence J. Lau. 1971. Conjugate Duality and the Transcendental Logarithmic Production Function. *Econometrica* 39(4): 255-256.

Chung, Yangho H., Rolf Färe, and Shawna Grosskopf. 1997. Productivity and Undesirable Outputs: A Directional Distance Function Approach. *Journal of Environmental Management* 51(3): 229-240.

Cuesta, Rafael A., Jose L. Zofio. 2005. Hyperbolic Efficiency and Parametric Distance Functions: With Application to Spanish Savings Banks. *Journal of Productivity Analysis* 24(1): 31-48.

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

Diewert, W. E. 1976. Exact and Superlative Index Numbers. *Journal of Econometrics* 4(2): 115-145.

Efron, Bradley. 1979. Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics* 7(1): 1-26.

Efron, Bradley, and Robert J. Tibshirani. 1993. *An Introduction to the Bootstrap.* Monographs on Statistics and Applied Probability, vol. 57; New York and London: Chapman and Hall.

Färe, Rolf, Shawna Grosskopf. 1990. A Distance Function Approach to Price Efficiency. *Journal of Public Economics* 43(1): 123-126.

Färe, Rolf, Shawna Grosskopf, and Lovell, C. A. Knox. 1985. *The Measurement of Efficiency of Production.* Boston: Kluwer Nijhoff Publishing.

Färe, Rolf, Shawna Grosskopf, C. A. Knox Lovell, and Carl Pasurka. 1989. Multilateral Productivity Comparisons When Some Outputs are Undesirable: A Nonparametric Approach. *The Review of Economics and Statistics* 71(1): 90-98.

Färe, Rolf, Shawna Grosskopf, Dong-Woon Noh, and William Weber. 2005. Characteristics of a Polluting Technology: Theory and Practice. *Journal of Econometrics* 126(2): 469-492.

Färe, Rolf, Carlos Martins-Filho, and Michael Vardanyan. 2006. On Functional Form Representation of Multi-output Production Technologies. Working Paper. Oregon State University, Department of Economics.

Färe, Rolf, and Daniel Primont. 1995. *Multi-output Production and Duality: Theory and Applications.* Boston: Kluwer Academic Publishers.

Färe, Rolf, Keuk Sung. 1986. On Second-order Taylor's-series Approximation and Linear Homogeneity. *Aequationes Mathematicae* 30(1): 180-186.

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

Gijbels, Irene, Enno Mammen, Byeong U. Park, and Leopold Simar. 1999. On Estimation of Monotone and Concave Frontier Functions. *Journal of the American Statistical Association* 94(445): 220-228.

Guilkey, David K., Lovell, C. A. Knox, and Robin C. Sickles. 1983. A Comparison of the Performance of Three Flexible Functional Forms. *International Economic Review* 24(3): 591-616.

Hall, P., 1986. On the Number of Bootstrap Simulations Required to Construct a Confidence Interval. Annals of Statistics, 14: 1453-1462

Karlin, Samuel. 1959. *Mathematical Methods and Theory in Games, Programming and Economics.* Reading, MA: Addison-Wesley.

Kneip, Alois, Byeong U. Park, and Leopold Simar. 1998. A Note on the Convergence of Nonparametric DEA Estimators for Production Efficiency Scores. *Econometric Theory* 14(6): 783-793.

Kneip, Alois, Leopold Simar, and Paul W. Wilson. 2003. *Asymptotics for DEA Estimators in Nonparametric Frontier Models.* Discussion Paper No. 0317.Louvain-la-Neuve, Belgium: Institut de Statistique, Universite Catholique de Louvain.

Lau, Lawrence J. 1972. Profit Functions of Technologies with Multiple Inputs and Outputs. *Review of Economics and Statistics* 54(3): 281-289.

Malmquist, Sten. 1953. Index Numbers and Indifference Surfaces. *Trabajos de Estadistica* 4(1): 209-242.

Meeusen, Wim, van Den Broeck, Julien. 1977. Efficiency Estimation from Cobb-Douglas Production Functions with Composed Error. *International Economic Review* 18(2): 435-444.

Morishima, Michio. 1967. A Few Suggestions on the Theory of Elasticity. *Keizai Hyoron (Economic Review)* 16: 144-150.

Shephard, Ronald W. 1974. *Indirect Production Functions.* Meisenheim Am Glan: Verlag Anton Hain.

Shephard, Ronald W. 1970. *Theory of Cost and Production Functions.* Princeton: Princeton University Press.

Shephard, Ronald W. 1953. *Cost and Production Functions.* Princeton: Princeton University Press.

Silverman, Bernard W. 1986. *Density Estimation for Statistics and Data Analysis.* London: Chapman and Hall.

Simar, Leopold, Paul W. Wilson. 1998. Sensitivity Analysis of Efficiency Scores: How to Bootstrap in Nonparametric Frontier Models. *Management Science* 44(1): 49-61.

Simar, Leopold, Paul W. Wilson. 2007. Statistical Inference in Nonparametric Frontier Models. In *The Measurement of Productive Efficiency and Productivity Growth.* New York: Oxford University Press.

von Neumann, John 1945. A Model of General Economic Equilibrium. *The Review of Economic Studies* 13(1): 1-9.

Zofio, Jose L., Lovell, C. A. Knox. 2001. Graph Efficiency and Productivity Measures: An Application to US Agriculture. *Applied Economics* 33(11): 1433-1442.

Mykhaylo Salnykov

ESSAY 3

DOES EMISSION PERMIT TRADE HAMPER DEVELOPMENT?

CAPITAL AND OUTPUT DYNAMICS UNDER AN INTERNATIONAL TRANSFERABLE EMISSION QUOTA TRADE SYSTEM

M. Salnykov
Department of Economics, Simon Fraser University
8888 University Drive, Burnaby, BC, Canada, V5A 1S6
E-mail: salnykov@sfu.ca
Tel.: +1 604 291 35 08; fax: +1 604 291 59 44

Abstract

I study the effects of nontradeable emission quota and transferable emission quota systems on the accumulation of capital and output growth in small open economies. As in other papers, I find that both types of regulation impede the growth. However, I further find that the transferable emission quota system has different effects on the development of quota buyers and quota sellers. While quota buyers enjoy faster growth in the both capital stock and economic output as compared to the nontransferable quota system, quota sellers face slower capital accumulation and economic growth.

I estimate the input distance function yielding the production frontier and inefficiency parameters using the cross-sectional data on output per capita, CO_2 per capita and capital per capita. I then implement a dynamic simulation using this information to characterise the magnitudes of the effects in the theoretical model. The simulation reveals that developmental consequences for quota sellers range from a slower capital accumulation to capital stock shrinkage. It also suggests that quota sellers will substitute economic production for quota trade revenues and that economic output will fall over time.

Keywords production theory; input distance functions; emission trade; bootstrap application

JEL Classification Numbers C15, D2, D24, Q25, Q28, Q56

1. INTRODUCTION

In recent years, it has become increasingly apparent that the problem of global climate change and the related issue of greenhouse gases (GHG) emissions may be one of the biggest challenges humankind has ever faced. The Kyoto Protocol (KP) is the largest scale attempt to take on this challenge. It is an amendment to the United Nations Framework Convention on Climate Change (FCCC) adopted at the third FCCC conference of the parties (COP) in 1997. The Kyoto Protocol assigns voluntary emission targets (also known as assigned amounts) as a fraction of the emission level in the base year (1990 for most of the KP parties). The assigned amounts are imposed on the Protocol's Annex B[1] countries comprised of 39 industrialized OECD economies as well transitional post-communist economies of Central and Eastern Europe. To help the parties to achieve their commitments, the Protocol defines so called flexibility mechanisms, a set of three clauses that allow the parties to reach their target by using means other than directly via cuts of their emissions, thus reducing the overall costs of meeting the targets. The mechanisms consist of Emission Trading, Joint Implementation and Clean Development Mechanisms[2].

This paper concerns mostly the Emission Trading mechanism that allows Annex B[3] countries that emit at the level below their assigned amounts to sell the balance of their emission quotas (KP, Article 7). The trade system setup was finalized by the Marrakech Accord, the outcome of the COP-7 in 2001. The Accord frames the emission trade in terms of the assigned amount units (AAUs), which is an equivalent of physical reduction of the GHG emissions by one ton of CO_2 equivalent[4]. The

[1] The terms FCCC Annex I and KP Annex B are often used interchangeably. These lists are identical except of Belarus and Turkey, who were not parties to the Convention when the Protocol was adopted.

[2] For a detailed discussion of the KP framework and the flexibility mechanisms, see Barrett (1998).

[3] The notable exceptions include US and Australia, which are Annex B countries that declined to ratify the Protocol. Therefore they are not allowed to trade their credits under the treaty conditions.

[4] The KP sets assigned amounts for six GHGs that differ in their global warming potential. To allow for adequate calculation of total reductions, all gases are commensurated in terms of the global warming potential of CO_2, the most common GHG. One tonne of CO_2equivalent is an amount of a GHG that has the same global warming potential as one tonne of CO_2.

Accord discourages the use of the emission trade revenues for purposes other than the further abatement of greenhouse gases pollution, but does not prohibit it.

As trade is allowed between Annex B countries only, it is expected that the transitional countries of Central and Eastern Europe will sell permits in the first commitment period (2008-2012) as most of them are still below their base year pollution levels due to the economic crisis in the region during the 1990s. It is also expected that in the second commitment period, when the targets will become more restrictive for these countries, they will still sell pollution permits.

This paper analyzes an international pollution permit trade model similar to the Kyoto Protocol and simulates the trade process based on real-life data. Two types of environmental treaties are studied: a simple emission cap system and a cap-and-trade system. Both the theoretical model and the simulation exercise show that in the regulated environments, capital will be accumulated slower in all countries, resulting in slower economic output growth. However, for pollution permit buyers, a cap-and-trade system will result in faster capital accumulation and output growth if compared to a simple cap system. In contrast, pollution permit sellers will face a slower (and occasionally even negative) capital accumulation and output growth if the transferable emission quota (TEQ) trade system is in place.

The model is inspired by so-called green growth models on one hand (see Forster, 1973; Stockey, 1998 and Brock and Taylor, 2004) and general equilibrium models of environment and trade on the other hand (see Pethig, 1976; Copeland and Taylor, 2004 and 2005). As in Copeland and Taylor (2005) I analyze small open economies under an international TEQ trade system. My study differs in a number of ways, however. Firstly, my main focus is on the effect of this system on the development of the economies, namely economic output and capital accumulation dynamics. Secondly, I approach the problem from a dynamic perspective by allowing for technological progress. Finally, I simulate the model using real-life data to estimate the magnitude of the TEQ trade effect on economic output and capital accumulation.

The model will be formulated in Section 2. Section 3 will contain the solution. Discussion of the simulation procedure and results will be provided in Section 4. Section 5 will conclude.

2. MODEL SETUP

I assume a world consisting of small open economies. In time period t each economy i produces economic output, Y, and environmental pollution, Z, using capital, K. Although pollution is an undesirable by- product of economic output, it is often treated as one of the factors of production (see Copeland and Taylor, 2005 and Ishikawa, 2006 for examples). The intuition behind it is that the abatement technologies consume economic resources (Copeland and Taylor, 2005). There exists output augmenting technical change common for all economies, A, normalized to the initial period. Capital stock consists of the undepreciated portion of the previous period capital and the current period investments, I.

$$Y_t^i = A_t \phi^i \left(K_t^i, Z_t^i\right) \tag{1}$$

$$A_t = (1+\eta)A_{t-1} = (1+\eta)^t \tag{2}$$

$$K_t^i = (1-\delta)K_{t-1}^i + I_t^i \tag{3}$$

where technological progress rate $\eta>0$, and depreciation rate $\delta>0$. From now on, I will suppress subscripts and superscripts unless they differ within a single equation or a system of equations.

ϕ is a $\Re_+^2 \to \Re_+$ monotonically increasing in capital, strictly concave and continuously twice differentiable in both arguments function that may or may not be different across economies. I depict projections of on (Y, K) and (Y, Z) spaces on Figure 3-1 (for illustration purposes A is normalized to 1). Concavity in Z may be interpreted as an increasing marginal abatement cost of pollution phenomenon, a common assumption in many environmental economics texts.

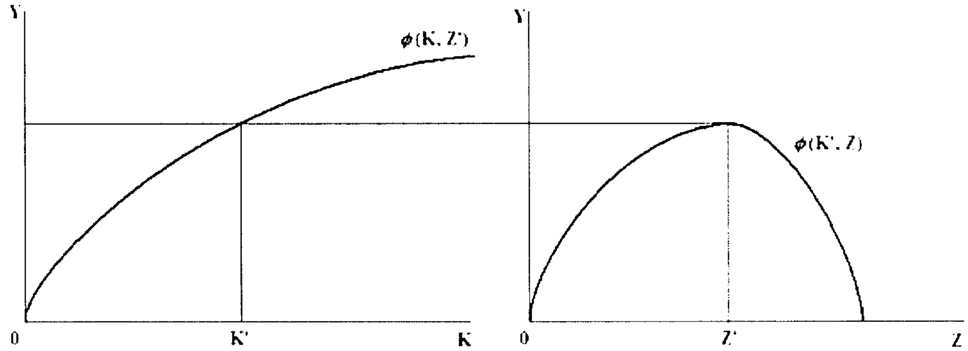

Figure 3-1. Projections of ϕ on a) (Y,K) and b) (Y,Z) spaces. A is normalized to unity.

I impose a standard assumption from the productivity analysis of polluting technologies literature (e.g. Färe, Grosskopf, Noh, and Weber, 2005). Specifically,

$$\forall K' > 0 \ \exists Z' > 0,: \phi_Z(K', Z') = 0. \tag{4}$$

Note that concavity implies $\phi_Z(K', Z) > 0$ if $Z < Z'$ and $\phi_Z(K', Z) < 0$ if $Z > Z'$. Naturally, a rational decision making unit (DMU) will not be located on the negatively sloped portion of the boundary, since the same amount of economic output may be produced with smaller level of pollution (see Figure 3-1 for illustration).

ϕ_{ZK} is positive on the domain[5], i.e. capital and pollution are assumed to be complements in the production process. If this is not the case, no caps are needed to limit the pollution level down: as capital stock increases, optimal pollution level will decrease, which decreases importance of environmental regulation. While assuming a negative ϕ_{ZK} would not be unrealistic in general, it would be quite strange to expect that a supranational authority would encourage such a country to be involved in an environmental treaty, where it would be able to earn emission trade revenues without any abatement efforts, i.e. by selling so called "hot air".

[5] This condition is satisfied for standard production functions including Cobb-Douglass and Constant Elasticity of Substitution production functions.

ϕ satisfies the conventional "no free lunch" condition and limit conditions:

$$\phi(K,0) = \phi(0,Z) = 0 \quad \forall K, Z \geq 0, \tag{5}$$

$$\lim_{K\to 0} \phi_K(K,Z) = \lim_{Z\to 0} \phi_Z(K,Z) = \infty; \quad \lim_{K\to +\infty} \phi_K(K,Z) = 0. \tag{6}$$

I assume that consumption (*C*), global pollution ($\breve{Z}$) and locally produced pollution are perfectly substitutable[6] in the preferences of societies:

$$U_i(C^i, Z^i) = C^i - \theta^i \breve{Z} - \sigma^i Z^i, \tag{7}$$

where σ^i is the shadow price of emissions for country *i*, which can be interpreted as the internal valuation of a pollutant by a society (Salnykov and Zelenyuk, 2005) or as a revealed social preference of a given society with respect to pollution, θ^i - shadow price of the global pollution. These shadow prices is time invariant for each country, but vary across countries. Undoubtedly, in reality shadow prices will depend on many factors, such as the level of pollution or the wealth of the country and may change over time. Nevertheless, the factual observations suggest that these changes occur very slowly and may require a change of generations.

Note that by allowing separate valuation of global and local levels of pollution I separate disutilities resulted by the global scale effect of pollution (such as global climate change) and the local scale effect (e.g., health effect or even feeling bad that one's country pollutes much).

Each country produces a small share of the global pollution. Therefore, $\breve{Z}$ is taken by each country as given.

Two[7] small open economies B and S enter an agreement that is intended to limit their joint emissions. At the time of signing the treaty the countries are characterized by $\{Z_0^B, K_0^B, \sigma^B\}$ and $\{Z_0^S, K_0^S, \sigma^S\}$. The countries agree to allocate

---

[6] Weak seperability of tastes across the set of consumption and pollution is a common assumption in the public economics literature (see Copeland and Taylor, 2005). Assuming perfect substitutability is more restrictive, however. I will elaborate on it below.

[7] Later we will extend the model to an arbitrary number of the treaty participants.

TEQ allowances ($\overline{Z}^B$ and $\overline{Z}^S$) using the grandfathering principle, specifically $\overline{Z}^B = Z_0^B$ and $\overline{Z}^S = Z_0^S$. Any amount of quotas q may be sold by one country to the other at a price p established by the competitive market. The seller cannot emit more than $\overline{Z}^i - q$ and the buyer cannot emit more than $\overline{Z}^i + q$.

The countries devote their economic output to consumption, investment in the capital stock and quota purchases:

$$A\phi(K, Z) \geq C + I + px, \tag{8}$$

where x is a demand for quotas by a given country.

Note that (8) implies that if country is a quota seller, then its income (a total of economic output and trade revenues) would be distributed between consuming and reinvesting into capital.

The decision maker is taken to be a periodically elected governing structure. Therefore, its choices in each period are aimed on maximizing the current period social utility only.

Finally, I assume that capital stock and emission levels are at their long-run equilibrium values in the initial period.

Although some of my assumptions seem untraditional or too restrictive (e.g. myopic central planner, restrictions on the production function, etc.), in these cases I sacrificed generality or conventionality to make the model more congruous with reality.

3. SOLUTIONS TO THE MODEL

I solve the model in two benchmark frameworks: business-as-usual (BAU) and non-transferable quota (NTQ) environment. Then I compare the competitive market equilibrium TEQ solution to the benchmark outcomes.

3.1. BUSINESS AS USUAL

Each country solves

$$\max_{K_t, Z_t}\left\{C_t - \theta\breve{Z}_t - \sigma^i Z_t^i\right\} \tag{9}$$

s.t.

(i) $A_t\phi(K_t, Z_t) \geq C_t + I_t$ (ii) $A_t = (1+\eta)^t$

(iii) $K_t = (1-\delta)K_{t-1} + I_t$ (iv) $C_t, K_t, Z_t, I_t \geq 0$.

with no constraints imposed on Z_t. Solutions $\left\{K_t^{BAU}, Z_t^{BAU}\right\}$ satisfy

$$\begin{cases}\phi_K\left(K_t^{BAU}, Z_t^{BAU}\right) = (1+\eta)^{-t} \\ \phi_Z\left(K_t^{BAU}, Z_t^{BAU}\right) = \sigma\end{cases} \tag{10}$$

Here, (10) implies that both K_t^{BAU} and Z_t^{BAU} are increasing over time. Naturally, economic output Y_t^{BAU} is increasing as well.

3.2. NON-TRANSFERABLE QUOTAS

If the quotas are assigned using the grandfathering rule, but cannot be traded, then each country solves

$$\max_{K_t, Z_t}\left\{C_t - \theta\breve{Z}_t - \sigma^i Z_t^i\right\} \tag{11}$$

s.t.

(i) $A_t\phi(K_t, Z_t) \geq C_t + I_t$ (ii) $A_t = (1+\eta)^t$

(iii) $K_t = (1-\delta)K_{t-1} + I_t$ (iv) $Z_t = \overline{Z}$

(v) $C_t, K_t, I_t \geq 0$.

with solutions $\left\{K_t^{NTQ}, Z_t^{NTQ}\right\}$ satisfying

$$\begin{cases}\phi_K\left(K_t^{NTQ}, Z_t^{NTQ}\right) = (1+\eta)^{-t} \\ Z_t^{NTQ} = \overline{Z}\end{cases}. \tag{12}$$

Here, (12) implies that K_t^{NTQ} is increasing over time while Z_t^{NTQ} is time invariant. Naturally, economic output Y_t^{NTQ} is increasing as well.

3.3. TRANSFERABLE EMISSION QUOTAS

If quotas can change hands, each country solves

$$\max_{K_t, Z_t, x_t} \left\{C_t - \theta \breve{Z}_t - \sigma^i Z_t^i\right\} \tag{13}$$

s.t.

(i) $A_t \phi(K_t, Z_t) \geq C_t + I_t + p_t x_t$ (ii) $A_t = (1+\eta)^t$

(iii) $K_t = (1-\delta)K_{t-1} + I_t$ (iv) $Z_t = \overline{Z} + x_t$

(v) $C_t, K_t, Z_t, I_t \geq 0$.

Optimal demand for quotas x^* by a given country satisfies

$$A\phi_Z\left(K, \overline{Z} + x^*\right) - \sigma = p \tag{14}$$

Market clearing implies $x^{*B} = -x^{*S} = q$ and consequently

$$A\phi_Z^B\left(K^B, \overline{Z}^B + q\right) - \sigma^B = A\phi_Z^S\left(K^S, \overline{Z}^S - q\right) - \sigma^S \tag{15}$$

Note that (15) is a modification of the factor price equalization theorem by Samuelson (1949), where the factor has intrinsic value to the society.

Without loss of generality, I assume that $x^{*B} \geq 0$, i.e. B will buy TEQs while S will sell them. The number of TEQs that change hands will be q.

(15) is an implicit function of q, A, K^B and K^S, where

$$\frac{\partial q}{\partial A} > 0; \qquad \frac{\partial q}{\partial K^B} > 0; \qquad \frac{\partial q}{\partial K^S} < 0 \tag{16}$$

Solutions to (13) $\left\{K_t^{TEQ}, Z_t^{TEQ}\right\}$ satisfy

$$\begin{cases} \phi_K\left(K_t^{TEQ}, Z_t^{TEQ}\right) = (1+\eta)^{-t} \\ Z_t^{TEQ} = \overline{Z} + x_t \end{cases}. \tag{17}$$

Proposition 1. For the TEQ *buyer*, quota trade will imply accumulation of capital and output growth *above* the NTQ level, but below BAU level.

Proof:

See Appendix 1.

Proposition 2. For the TEQ *seller*, quota trade will imply accumulation of capital and output growth *below* the NTQ level.

Proof:

See Appendix 1.

Corollary (the Kurse). Introduction of the tradable emission quotas will lead to faster development of quota buyers and slower development of quota sellers as compared to the nontradable quota system.

The main conclusion of the model is that transferable emission quotas trade will hamper economic development, specifically capital accumulation, in the quota selling countries, while buyers will enjoy a higher rate of economic development as compared to a simple cap regulation. I call this effect the Kyoto[8] curse or the Kurse for short[9].

It should be emphasized, however, that although the model uses a two party pollution permit trade treaty, the conclusions are valid for any multilateral treaty. The competitive market equilibrium will then satisfy

$$p = A\phi_Z^i\left(K^i, \overline{Z}^i + x^i\right) - \sigma^i \qquad \forall i = 1..N \tag{18}$$

$$\sum_{i=1}^{N} x^i = 0.$$

where N is the number of parties in the treaty.

In this case, for any country with $x^i > 0$, Proposition 1 will apply, and for any country with $x^i < 0$, Proposition 2 will be valid. In other words, any party to the treaty selling permits will experience slower development than in the case of a simple cap system, while any buyer will enjoy faster development.

---

[8] The Kyoto Protocol sets the international permit trade system with the initial quota levels (i.e. assigned amounts) established using the gradfathering principle. This setup may lead to the underdevelopment of the quota selling countries as the model suggests.

[9] One can also think of this name as the capital (K) underaccumulation curse, or the K-curse. The original idea of the name belongs to Dr. Nancy Olewiler.

It should be also noted that my study did not analyze welfare effect of the emission trade. As a matter of fact, decisions on the amounts of the quotas bought or sold are based on the utility maximization problems, so countries are becoming better off (at least within a short time span). However, such behavior may harm the future generations as they inherit low capital stock after the treaty comes to an end.

All the way through my analysis I purposefully avoided many (quite realistic) complications. Specifically, contrary to state-of-art papers I analyzed a single sector economy rather than a 1-clean-1-dirty-good economy to enable simulation using real-life data. I also did not explicitly assume existence of an abatement technology. I do recognize, however, that a potential shift in specialization to cleaner goods and more intensive abatement processes will definitely play an important role in the countries' commitment to meet their emission targets. Yet, this specialization change is more likely to occur in the long run rather than in the short run. Complication of the theoretical model to allow for these modifications would be a logical extension of the current paper.

To assess the magnitude of the Kurse for quota buyers and sellers, I simulate my model using real life data.

4. SIMULATION

The model involves two exogenous variables – δ and η – common for all countries. In addition, country specific $\phi^i(\cdot)$ and σ^i are assumed to be exogenous.

Values of shadow prices are estimated based on the estimated functions $\hat{\phi}^i(\cdot)$ as

$$\hat{\sigma}^i = \hat{\phi}^i_Z\left(K^i_0, Z^i_0\right) \tag{19}$$

To estimate the production function, $\phi^i(\cdot)$ I rely on tools from the literature on productivity and efficiency analysis. A short exposure into the fundamentals of this field is given in Appendix 2.

The productivity analysis literature shares an understanding that all DMUs (firms, countries, etc.) have an access to the same technology. However, the level of access differs across the units, which results in a different efficiency, i.e. distance of a given DMU to the best practice frontier. In terms of the present study it can be interpreted as all countries sharing the same best practice frontier, but some are less efficient than the others in producing economic output with the least amount of capital and pollution, which is translated into a different distance to the frontier across countries.

Efficiency is traditionally measured by Shephard-class efficiency measures (Shephard, 1970), such as the input distance function (IDF) and the output distance function (ODF). It has been shown that under relatively weak assumptions about the technology, both IDF and ODF are complete characterizations of technology (Färe and Primont, 1995)

Specifically, Färe and Grosskopf (1990) and Färe, Grosskopf, Lovell and Yaisawarng (1993) demonstrated how the approach of Aigner and Chu (1968) can be used to estimate the IDF and ODF. In this paper, by estimating the IDF, I obtain a parametric estimate of an implicit function relating economic output, capital and emissions.

The choice of IDF over ODF requires some elaboration. The ODF measure expands an observation radially as much as technologically possible in the output space, while IDF contracts the observation radially in the input space (Färe, Grosskopf and Lovell, 1994). While IDF will always project the observation on the portion of the best practice frontier, where economic output is produced with the smallest capital and emissions, ODF may expand the observation to the point on the frontier that is economically irrational to be at. Figure 3-2 provides an illustration of the argument. If A is an observed combination of economic output and pollution $\{\tilde{Y},\tilde{Z}\}$, the ODF scales it up in the economic output space to the efficient peer $\{\tilde{Y}/\theta^*,\tilde{Z}\}$, where θ^* is the ODF score for observation A. The ODF score will seek for

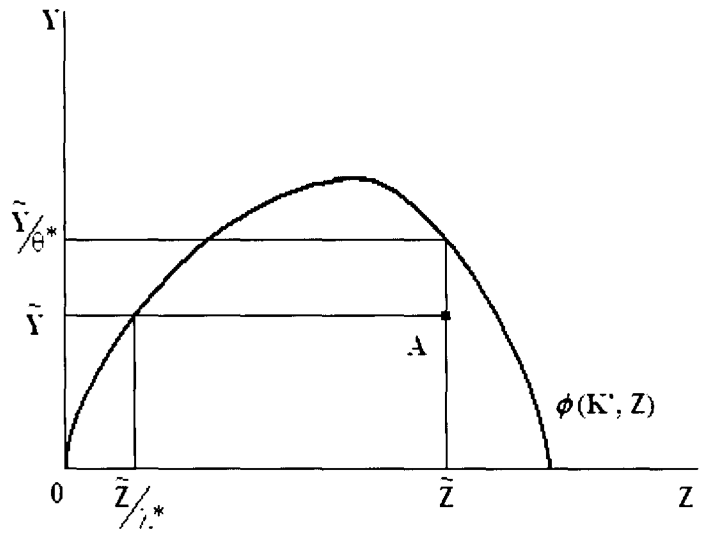

Figure 3-2. IDF projection vs ODF projection: λ^* denotes the IDF measure for observation *A*; θ^*- the ODF measure.

the largest technologically feasible expansion of the economic output given the levels of *K* and *Z*. The IDF contracts the observation down in the (*K*, *Z*) space[10] to the efficient peer $\left\{\tilde{Y}, \tilde{Z}/\lambda^*\right\}$, where λ^* is the IDF score for observation *A*. The IDF score will search for the largest technologically feasible proportional contraction of *K* and *Z* given the level of *Y*. As the diagram suggests, when the technology is polluting, the ODF may yield an efficient peer that is located on the technologically efficient, but economically irrational part of the best practice frontier, while the IDF will always result an efficient peer to be both technologically efficient and economically rational.

Therefore, given assumption of the rationality of DMUs, I conclude that IDF is a better choice of modelling a polluting technology.

Shephard IDF is defined on the technology set *T*, which is in my case

$$T \equiv \left\{(Y,K,Z): Y \in \Re_+ \text{ can be produced given } K \in \Re_+ \text{ and } Z \in \Re_+\right\}. \qquad (20)$$

[10] For the purpose of illustration I depict $\phi(K',Z) = \phi\left(K'/\lambda^*, Z\right)$, but in general, contraction of *K* will yield a best practice frontier not above an old one, so the justification will still hold.

Then the IDF is defined as

$$D_I(Y,K,Z) \equiv \sup\left\{\lambda > 0 : \left(Y, \frac{K}{\lambda}, \frac{Z}{\lambda}\right) \in T\right\}. \tag{21}$$

The IDF is a measure of the distance of a given observation towards the best practice frontier in the (*K, Z*) space. It was shown by Färe and Primont (1995) that if the technology set, *T*, is regular, then IDF is nonincreasing in outputs and nondecreasing in inputs, homogeneous of degree 1 in inputs, is equal to unity if the DMU is on the frontier and is commensurable (independent of units of measurement up to a scalar transformation) as defined by Russell (1987).

The econometric estimation of frontier functions was pioneered by Aigner and Chu (AC) (1968). By following Farrell's (1957) idea of describing "an industry envelope isoquant", they described a method of estimating the frontier model that constrained all residuals to be negative, a full frontier model. For simplicity, AC estimated one-output, two-input Cobb-Douglas production function

$$y = Ax_1^{\alpha} x_2^{\beta} u, \tag{22}$$

where y is an output, x_1 and x_2 are two inputs, u is a random shock and A, α and β are parameters. The authors rewrite (22) in logarithms to obtain

$$e = -\ln A + \ln y - \alpha \ln x_1 - \beta \ln x_2, \tag{23}$$

where $e = \ln u$.

AC argue further that since the shocks lie only on one side of the production frontier (i.e. $e \le 0$), (23) is easily solved by minimizing the linear loss function (rather than sum of squared residuals as was done previously by OLS studies) within the framework of linear programming by finding A, α and β that solve

$$\max \sum_{n=1}^{N} \left[\ln y^n - \ln A - \alpha \ln x_1^n - \beta \ln x_2^n \right] \tag{24}$$

s.t.

(i) $A, \alpha, \beta \geq 0$

(ii) $\ln y^n - \ln A - \alpha \ln x_1^n - \beta \ln x_2^n \leq 0 \quad \forall n = 1..N$.

Later Christensen, Jorgenson and Lau (CJL) (1971) showed that if a homogeneous Translog aggregator function of vector $\mathbf{x} \in \Re^K$ is defined as

$$f(\mathbf{x}) \equiv \alpha_0 + \sum_{k=1}^{K} \alpha_k \ln x_k + \frac{1}{2} \sum_{i=1}^{K} \sum_{j=1}^{K} \beta_{ij} \ln x_i \ln x_j \,, \tag{25}$$

where $\sum_{k=1}^{K} \alpha_k = 1$, $\sum_{k=1}^{K} \beta_{jk} = 0$ and $\beta_{jk} = \beta_{kj}$ for all $j,k = 1..K$, then it can give a second order approximation to any twice continuously differentiable linear homogeneous function.

Diewert (1976) used AC approach and CJL finding to parameterize the Shephard's distance functions to what now is known as *Translog* specification of the distance functions by imposing additional conditions on the parameters to account for the properties of the distance functions. This approach was later used in the aforementioned studies of Färe and Grosskopf (1990) and Färe, Grosskopf, Lovell and Yaisawarng (1993) to estimate IDF and ODF respectively.

In the spirit of the previous works, if the IDF is parameterized by a flexible functional form

$$\begin{aligned} \ln D_I(Y,K,Z) = \quad & \alpha_0 + \alpha_1 \ln Y + \alpha_2 \ln K + \alpha_3 \ln Z \\ & + \beta_1 (\ln Y)^2 + \beta_2 (\ln K)^2 + \beta_3 (\ln Z)^2 \\ & + \gamma_1 \ln Y \ln K + \gamma_2 \ln Y \ln Z + \gamma_3 \ln K \ln Z \,, \end{aligned} \tag{26}$$

then the parameters of (26) can be estimated by solving

$$\min \sum_{n=1}^{N} \ln D_I\left(Y^n, K^n, Z^n\right) \qquad (27)$$

s.t.

(i) $\ln D_I\left(Y^n, K^n, Z^n\right) \geq 0 \qquad \forall n = 1..N$

(ii) $\dfrac{\partial \ln D_I\left(Y^n, K^n, Z^n\right)}{\partial \ln Y^n} \leq 0; \ \dfrac{\partial \ln D_I\left(Y^n, K^n, Z^n\right)}{\partial \ln K^n} \geq 0; \ \dfrac{\partial \ln D_I\left(Y^n, K^n, Z^n\right)}{\partial \ln Z^n} \geq 0$

(iii) $\alpha_2 + \alpha_3 = 1; \quad \beta_2 + \beta_3 + \gamma_3 = 0; \qquad \gamma_1 + \gamma_2 = 1.$

where n denotes the DMU number and N is the total number of the DMUs being analyzed. The first constraint requires observations to be technologically feasible; second – imposes regularity conditions on T; third – enforces homogeneity of degree +1 in inputs. Note that $\sum_{n=1}^{N} \ln D_I(\cdot) = \sum_{n=1}^{N} \left[\ln D_I(\cdot) - \ln 1\right]$, i.e. (27) is equivalent to finding parameters that minimize the sum of all deviations from the frontier. (26) represents a production function shared by all DMUs scaled down in the (K, Z) space by a country-specific efficiency score $D_I\left(Y^n, K^n, Z^n\right) \geq 1$.

It is worth pointing out that the estimates of (26) obtained by (27) are ML if the disturbance is assumed to be half-normally distributed (Greene, 1980). Unfortunately, statistical properties of the estimates of (26) do not have an analytic formulation. As in other cases, when analytic results are not comforting, bootstrap (Efron, 1979 and Efron and Tibsharani, 1993) is an attractive alternative for making inferences. Bootstrapping relies on the repeated simulation of the data generating process and applying the original estimator to every simulated sample so that resulting estimates mimic the sampling distribution of the original estimator.

I use the <u>s</u>mooth <u>h</u>omogeneous <u>b</u>ootstrap (SHB) methodology proposed by Simar and Wilson (1998) to obtain statistical properties of the frontier estimates. SHB was proposed as an alternative to the naive bootstrap, which was shown to give inconsistent results in the frontier problems (<u>S</u>imar and <u>W</u>ilson (SW), 1998). While the naïve bootstrap draws a bootstrap sample from a discrete distribution, SHB uses a smoothing module based on the kernel estimated densities of the obtained efficiency scores to draw a sample from a continuous distribution. In other words, SHB perturbs

DMUs' data around their observed data to create a bootstrap sample. Smooth resampling is being repeated many times (in my case, 1,000 as Hall (1986) suggests to ensure the adequate coverage of the confidence intervals) to obtain samples' estimates of the original estimator (27). The bootstrap estimates are then used to obtain biased corrected estimates of parameters in (26), their standard errors and confidence intervals.

I identify that if a traditional SW smooth homogeneous bootstrap is used in the parametric framework, some estimated variances of some parameters seem to be unrealistically low. Therefore, SHB is executed in two variations: a traditional SW bootstrap and, what I call, a cautious SHB bootstrap, which provide with more realistically looking parameters' variances. Appendix 3 provides discussion of the SHB algorithm, the bias-correction and the confidence interval estimation procedure, as well as explains the difference between the SW and the cautious alternative.

Once one uses (27) to obtain the estimates of the parameters in (26) $\hat{\Xi} = \{\hat{\alpha}_0, \hat{\alpha}_1, \hat{\alpha}_2, \hat{\alpha}_3, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \hat{\gamma}_1, \hat{\gamma}_2, \hat{\gamma}_3\}$, they are used to calculate country-specific estimates of the IDF scores. Therefore, an IDF estimate for country i, $\hat{\lambda}^{*i}$ is

$$\begin{aligned}\hat{\lambda}^{*i} = \quad & \exp\big[\hat{\alpha}_0 + \hat{\alpha}_1 \ln Y^i + \hat{\alpha}_2 \ln K^i + \hat{\alpha}_3 \ln Z^i \\ & + \hat{\beta}_1 (\ln Y^i)^2 + \hat{\beta}_2 (\ln K^i)^2 + \hat{\beta}_3 (\ln Z^i)^2 \\ & + \hat{\gamma}_1 \ln Y^i \ln K^i + \hat{\gamma}_2 \ln Y^i \ln Z^i + \hat{\gamma}_3 \ln K^i \ln Z^i \big],\end{aligned} \tag{28}$$

The IDF score estimates are then used to estimate the individual countries' production functions, $\hat{\phi}^i(\cdot)$.It is given by the implicit function:

$$\begin{aligned}& \big[\hat{\alpha}_0 - \ln \hat{\lambda}^{*i}\big] + \hat{\alpha}_1 \ln Y + \hat{\alpha}_2 \ln K + \hat{\alpha}_3 \ln Z \\ & + \hat{\beta}_1 (\ln Y)^2 + \hat{\beta}_2 (\ln K)^2 + \hat{\beta}_3 (\ln Z)^2 \\ & + \hat{\gamma}_1 \ln Y \ln K + \hat{\gamma}_2 \ln Y \ln Z + \hat{\gamma}_3 \ln K \ln Z = 0.\end{aligned} \tag{29}$$

(29) can be interpreted as follows: all countries share the same best practice frontier, but the country-specific distance to this frontier determines their individual time invariant production function.

The final parameter of the model estimated using data is the shadow price of *Z*. Graphically, shadow prices can be interpreted as a slope of the production function in the (*Y*, *Z*) space. Given parameterization of the production function (26), shadow price of pollution for individual DMUs are estimated according to Färe and Primont (1995) by applying the implicit function theorem.

$$\hat{\sigma}^i = -\frac{\partial\hat{\lambda}^{*i}/\partial Z}{\partial\hat{\lambda}^{*i}/\partial Y} = -\frac{\hat{\alpha}_3 + 2\hat{\beta}_3 \ln Z^i + \hat{\gamma}_2 \ln Y^i + \hat{\gamma}_3 \ln K^i}{\hat{\alpha}_1 + 2\hat{\beta}_1 \ln Y^i + \hat{\gamma}_1 \ln K^i + \hat{\gamma}_2 \ln Z^i} \cdot \frac{Y^i}{Z^i} \tag{30}$$

I use data in per capita terms on GDP (*Y*), capital stock (*K*) and CO_2 pollution (*Z*) on 81 countries in 1995[11]. Table 3-1 provides descriptive statistics on the data set.

Table 3-2 presents the estimated values of the parameters, bootstrap bias corrected (BC) parameter estimates of (26)[12] obtained using the standard SW approach (SW, the first line for each parameter) and a cautious alternative (CA, the second line) as well as the bootstrap estimated standard errors. Figure 3-3 provides graphical representation of the findings. Each pair of candles provides bootstrap estimated statistics for the SW approach (filled bodied candles) and the alternate approach (hollow bodied candles). The candles are centered around the bias corrected estimates of the parameters (marked with diamonds, ♦); the bodies denote 95% confidence intervals and the tails the extremes of the deviations from the bias corrected values. The uncorrected values of the parameters are drawn as solid lines.

As the table and the figure suggest, the standard SW approach results estimated variations of parameters related to *Y* as well as all cross terms seem to be unrealistically low. The reason for that is that Simar and Wilson designed the approach having a nonparametric framework in mind, where the estimates of the efficiency scores rather than the underlying parameters were the major interest. Therefore, the traditional approach creates bootstrap samples, where (*K*, *Z*) duplets are stochastic, but *Y*'s and *K/Z* ratios are deterministic. In the cautious alternative

[11] The set was collected and discussed by Salnykov and Zelenyuk (2005).

[12] The estimated frontier IDF satisfied convexity in *Y* in all observations and quasiconcavity in (*K*,*Z*) in 55 of 81 observations. Although it was not modeled for explicitly, $\phi_{KZ} > 0$ for all 81 observations.

Table 3-1. Descriptive statistics of the dataset and estimates of the IDF and shadow prices

	Units	Mean	Max	Min	St.dev.
Y	US$1,000	16.7945	80.2777	0.2249	22.0875
K	US$1,000	2.0615	12.2794	0.0120	3.0042
Z	metric tons	10.8655	59.0921	0.0149	10.3351
$\hat{\lambda}^*$		1.7387	3.2042	1.0000	0.4752
$\hat{\sigma}$	US$1,000/ton	0.5443	2.7552	3.7 e-11	0.5330

approach I sample both *K/Z* ratios and *Y*'s. As a result, the variances of the estimates increase. The alternative bootstrap standard errors are much larger and I take them as cautious. Moreover, as the diagram suggests, the SW bias-corrected estimates of the parameters lie within the 95% confidence intervals of the alternate approach estimates for almost all parameters.

Confidence intervals are asymmetric around the bias corrected values because SHB draws a bootstrap sample from the asymmetric distribution of efficiency scores.

In the simulation exercise I use the cautious approach bias-corrected estimates; however both sets of estimates were tried and I discovered that the simulation results do not differ qualitatively.

The parameters are then used to obtain estimates of the IDF and shadow prices of CO_2 emissions for every observation. The descriptive statistics for these estimates (the IDF is unlogged) are given in Table 3-1. The statistics show that, on average, countries have a technical ability to decrease their CO_2 pollution and capital stock by 44% without decreasing their GDP. The most inefficient country in the sample (Azerbaijan) can potentially decrease emissions and capital by 69% and remain on the same economic output level. Social willingness to pay for a reduction of CO_2 emissions varies from $2,755 per ton (Sweden) to virtually nothing (Turkmenistan). An average country is willing to forfeit $544 of consumption to abate a ton of CO_2 emissions. The results generally support the findings of Salnykov and Zelenyuk (2005) that developed countries generally have higher valuation of environmental pollution.

Table 3-2. Parameter estimates, bias corrected parameter estimates and their standard errors

	Value	SHB type	BC value	St. err.
$\hat{\alpha}_0$	1.0724	SW	1.3209	0.1834
		CA	0.3056	0.4077
$\hat{\alpha}_1$	-1.0316	SW	-1.0368	3.3e-10
		CA	-0.8699	0.1385
$\hat{\alpha}_2$	0.6529	SW	0.7006	0.0339
		CA	0.3378	0.2152
$\hat{\alpha}_3$	0.3471	SW	0.2994	
		CA	0.6622	
$\hat{\beta}_1$	0.0127	SW	0.0146	5.8e-11
		CA	-0.0047	0.0157

	Value	SHB type	BC value	St. err.
$\hat{\beta}_2$	0.1000	SW	0.1000	4.6e-11
		CA	0.0780	0.0255
$\hat{\beta}_3$	-0.0850	SW	-0.0847	5.9e-11
		CA	-0.1139	0.0250
$\hat{\gamma}_1$	-0.1516	SW	-0.1513	9.6e-11
		CA	-0.1127	0.0464
$\hat{\gamma}_2$	0.1516	SW	0.1513	
		CA	0.1127	
$\hat{\gamma}_3$	-0.0147	SW	-0.0153	
		CA	0.0411	

Note: Value – an estimate from the original sample before bias correction; BC value – bias corrected estimates of the parameters; St.err. – standard errors. For each parameter the first line indicates the value for the traditional SW approach, the second – for the cautious alternative (CA).

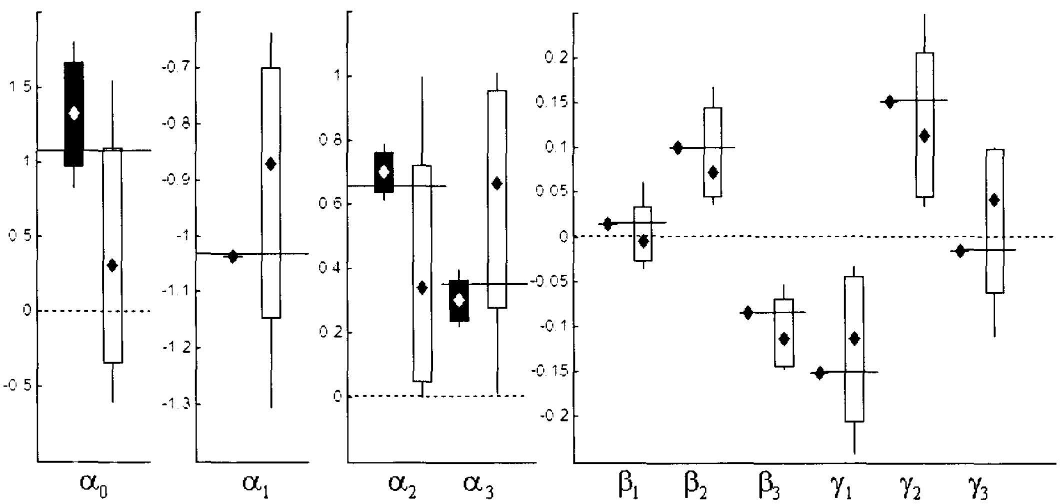

Figure 3-3. Box plots of the parameters' statistics. The body of each candle depicts 95% bootstrap estimated confidence intervals; tails show the highest and the lowest bootstrap estimated deviation from the parameters' bias corrected values marked with diamonds (♦). Initial estimates of the parameters (before bootstrap bias correction) are drawn as solid horizontal lines. Solid bodied candles denote original SW approach bootstrap; hollow bodied candles – cautious alternative.

I simulate international CO_2 permit trade among 31 Kyoto Protocol Annex B countries. Pollution caps are set to the initial period and I assume relatively low economic progress (η=0.01) and depreciation (δ=0.05)[13]. In every period each country simultaneously solves (13); market for pollution permits clears, i.e. (18) is satisfied at any t. Results are then compared to the outcomes of the BAU and NTQ scenarios. I run a simulation over a span of 11 periods.

As expected a priori, the simulation results show that Central and Eastern European post-communist economies will be permit sellers, while developed Western European and non-European Annex B countries will act as permit buyers[14]. The simulation reveals that initially 21 of 31 countries will buy permits; later on, when price for the permits increases, 2 of these countries start selling permits instead.

Pollution permit market size and price dynamics is demonstrated on Figure 3-4. The total amount of pollution quotas allocated to the countries, which chose to be permit sellers is 186.65 units. Market size in the last period simulated is 118.96, i.e. permit sellers cut down their emissions by about 65% total. As quota sellers decrease their emissions, they face increasing marginal abatement cost (which may be also interpreted as opportunity cost of pollution abatement on the margin), which is captured by the increasing pollution permits price.

As the simulation results suggest, the price for permits will increase from $8 per ton to $67 per ton of CO_2 within the first 5 periods of the treaty with the average of $35 per ton. Although my model was not designed to estimate the prices for permits, it is interesting to note that my estimates resemble closely what most experts agree that realistic estimates of the permits in the first 5-year commitment period of the Kyoto Protocol: between $25 and $55 per ton of CO_2 (ICF, 2005).

---

[13] Later I test the sensitivity of the results to the change in parameters.

[14] Notable exceptions include Belgium, which sells permits; Greece and Slovenia, which buy permits initially when prices are low, but then switch to selling permits in the late periods.

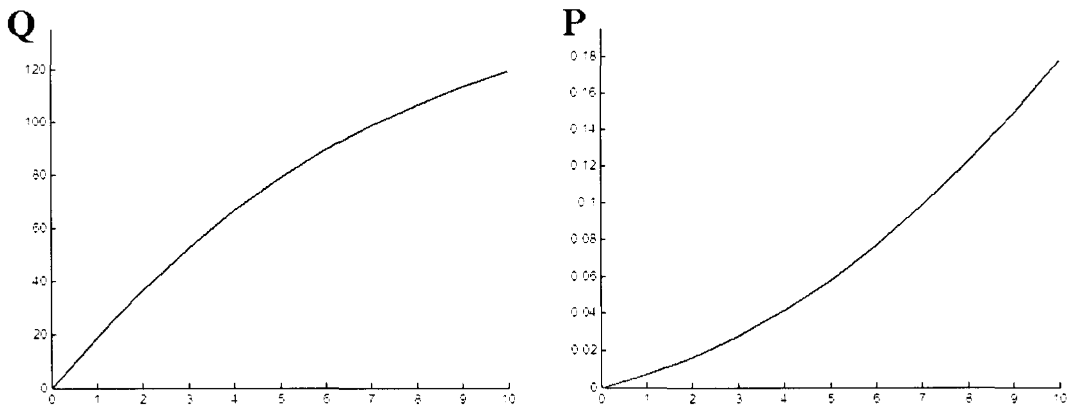

Figure 3-4. Pollution permit market size (a) and price (b) dynamics

Figure 3-5 in Appendix 4 displays simulated model dynamics for economic output, capital stock and emissions along with the corresponding growth rates for a typical permit buying country using Germany as an example. All other quota buyers reveal similar patterns and Germany was used as a typical representative of the dynamic patterns. As the theoretical model predicts, for permit buyers capital accumulation and output growth will be faster as compared to a simple cap regulation outcome, but slower than in an unregulated environment. An interesting finding revealed by the simulation exercise is that capital accumulation rate and output growth rate are faster in the initial periods and slow down later on, when quotas become more expensive, thus increasing cost of output growth for the quota buying economies.

There's a greater variation in behavior for sellers, however. Half of all sellers (6 out of 12)[15] exhibit an initial growth of capital stock with decumulation of capital starting at a particular instance (Latvia depicted on Figure 3-6 in Appendix 4 is a typical representative of this pattern). At the same time, economic output may or may not display an initial growth, but is always following a downward sloping trend in the late periods.

[15] Belgium, Czech Republic, Latvia, Lithuania, Poland, Slovakia.

A third of the sellers (4 out of 12)[16] show a special case of the pattern described above without any initial capital accumulation: both capital stock and economic output start falling from the initial period.

A final category of the sellers[17] exhibits another special case when capital and output grow throughout the entire simulation span (if the simulation span is extended beyond realistic lifetime of the treaty, those countries' growth slows down and eventually goes negative).

As predicted by the theoretical model, both capital and output growth rates stay below the ones under a simple cap regulation. Another interesting feature of the model revealed by the simulation is that there is a certain rigidity in the quota selling behavior: sellers increase the amount of quotas sold over time, primarily due to the increasing quota prices.

The investments in the quota selling countries are still positive despite possible capital stock shrinkage. Capital decumulation rate for all sellers does not go below 5% depreciation rate indicating that positive investments take place. If the simulation is repeated for a high depreciation rate (δ=0.10), no change in the dynamics is being observed. However, this depreciation rate neutrality may be a characteristics of this specific dataset.

A final group of the parties consists of the countries[18] that buy permits in the initial periods, but start selling them in the late periods when price of permits increases. This case is illustrated on Figure 3-7 in Appendix 4. The diagram reveals that the Kurse effect (capital growth below the NTQ level) starts at the instance when emission level starts falling, i.e. when country starts acting as a seller in relative terms with respect to the previous period emission level, not in absolute terms with respect to its initial permit endowment.

---

[16] Bulgaria, Estonia, Romania and Ukraine.

[17] Namely Croatia and Hungary.

[18] Greece and Slovenia.

Running simulation with higher values of the technological progress parameter does not yield qualitatively different results: the sets of quota selling and quota buying countries remain unchanged; quota market volumes grow faster as well as the permits price does; capital decumulation in quota selling countries start earlier, but still stays above the pure depreciation rate.

As simulation suggests, quota buyers are characterized by higher values of shadow prices of pollution. Shadow price of pollution at the equilibrium is equal to the marginal abatement cost of pollution, i.e. an opportunity cost of pollution abatement. By now, it should be clear that countries with higher shadow prices of pollution buy permits, because an opportunity cost of abatement for them is considerably higher than for the quota sellers.

When a quota seller faces favorable prices for emission permits, it shrinks the amount of capital used in the economic production (in the framework of Copeland and Taylor, 2005 that would be an equivalent of shifting capital to abatement from economic production). As a result, capital accumulation rate slows down. As the market price rises, incentives to abate increase and capital stock in the economic production starts shrinking.

5. CONCLUSIONS

This paper has shown that an international tradable emission quota system may hamper the development of a quota selling country. I test and simulate the model against two benchmarks: business as usual and simple cap system. I discover that while quota buyers' economic output is increasing above the output growth rate under a simple cap system, but below business-as-usual output, an economic output of the quota seller will decrease over time. I also identify that capital accumulation in the quota buying economy is more intensive as compared to a simple cap system environment. At the same time, capital accumulation shows a substantial slowdown (and may even be reversed) in the quota selling economies. Such behaviour may threaten economic welfare of future generations when the treaty comes to an end.

The phenomenon I called the Kurse is surprisingly similar to the Dutch Disease[19], a term coined by The Economist (1997). Despite being similar in terms of the effect on the national economy, the Kurse has a different mechanism: permit sellers substitute revenues from economic production for the revenues from the permit trade.

It should be apparent that the described international pollution permit trade systems are not purely illustrative. The model itself as well as the simulation exercise closely mimics conditions of the Kyoto Protocol, specifically the emission trading between Annex B countries. Transitional of Central and Eastern Europe are expected to be sellers of pollution permits under the conditions of the Protocol. It is a common perception that by allowing these countries to sell permits under Kyoto Protocol, economic development in them will be promoted. My study shows, however, that international pollution permit trade may hamper the development of permit selling economies.

On a positive note, I must emphasize that my study modeled a shortsighted behaviour when pollution permit revenues are allowed to be consumed. If the treaty participants are forced to spend their revenues on improving their abatement technology efficiencies, it may be the case that the detrimental effect of the quota trades is avoided. Alternatively, the national governments may be required to bank the trade revenues for the benefits of the future generations.

[19] Dutch Disease is a phenomenon which happens when a country rich in natural resources experiences an appreciation of its currency leading to a declining demand for its exports and, consequently, a fall of competitive sector production.

BIBLIOGRAPHY

Aigner, D., and Chu, S., 1968. On Estimating the Industry Production Function. American Economic Review, 58: 826-39.

Barrett, S., 1998. Political Economy of the Kyoto Protocol. Oxford Review of Economic Policy, 14: 20-39.

Brock, W.A., and Taylor, M.S. 2004. Economic Growth and the Environment: a Review of Theory and Empirics. NBER working paper 10854.

Christensen, L.R., Jorgenson, D.W., and Lau, L.J., 1971. Conjugate Duality and the Transcedental Logarithmic Production Function. Econometrica, 39:255-256.

Copeland, B.R., and Taylor, M.S. 1994. North-South Trade and the Environment. Quaterly Journal of Economics, 109: 755-787.

Copeland, B.R., and Taylor, M.S. 2005. Free Trade and the Global Warming: a Trade Theory View of the Kyoto Protocol. Journal of Environmental Economics and Management, 49: 205-234.

Diewert, E., 1976. Exact and Superlative Index Numbers. Journal of Econometrics, 4: 115-145.

Efron, B., 1979. Bootstrap Methods: Another Look at the Jackknife. Annals of Statistics, 7:1-26.

Efron, B., and Tibsharani, R.J., 1993. An Introduction to the Bootstrap. London: Chapman and Hall.

Färe, R., and Grosskopf, S., 1990. A Distance Function Approach to Price Efficiency. Journal of Public Economics, 43: 123-126.

Färe, R., and Primont, D., 1995. Multi-output Production and Duality: Theory and Applications. Boston: Kluwer.

Färe, R., Grosskopf, S., and Lovell, L., 1994. Production Frontiers. Cambridge: Cambridge University Press.

Färe, R., Grosskopf, S., Lovell, L.,and Yaisawarng, S., 1993. Derivation of Shadow Prices for Undesirable Outputs: a Distance Function Approach, The Review of Economics and Statistics, 75: 347-380.

Färe, R., Grosskopf, S., Noh, D.-W., and Weber,W., 2005. Characteristics of a Polluting Technology: Theory and Practice. Journal of Econometrics, 126: 469-492.

Farrell, M.J., 1957. The Measurement of Productive Efficiency. Journal of Royal Statistical Society, Ser. A,120: 253-281.

Forster, B.A., 1973. Optimal Capital Accumulation in a Polluted Environment. Southern Economic Journal, 39: 544-547.

Greene, W., 1980. Maximum Likelihood Estimation of Econometric Frontier. Journal of econometrics,13: 27-56.

Hall, P., 1986. On the Number of Bootstrap Simulations Required to Construct a Confidence Interval. Annals of Statistics, 14: 1453-1462.

ICF International, 2005. Analyzing the Price of Carbon in 2008-2012. Perspectives, Spring/Summer.

Ishikawa, J, and Kiyono, K., 2006. Greenhouse-Gas Emission Controls in an Open Economy International Economic Review, 47: 431-450.

Pethig, R., 1976. Pollution, Welfare, and Environmental Policy in the Theory of Comparative Advantage. Journal of Environmental Economics and Management, 2: 160-169.

Russell, R. R., 1987. On the Axiomatic Approach to the Measurement of Technical Efficiency. In: Eichhorn, W (ed.), Measurement in Economics: Theory and Application of Economic Indices. Heidelberg: Physica-Verlag.

Salnykov, M., and Zelenyuk, V., 2005. Estimation of Environmental Efficiencies of Economies and Shadow Prices of Pollutants in Countries in Transition. EERC Working Paper Series 05-06e, EERC Research Network, Russia and CIS.

Samuelson, P.A., 1949. International Factor Price Equalization Once Again. Economic Journal, 59: 181-197.

Silverman,B.W., 1986. Density Estimation for Statistics and Data Analysis. London: Chapman and Hall.

Simar, L., and Wilson, P., 1998. Sensitivity Analysis of Efficiency Scores: How to Bootstrap in Nonparametric Frontier Models. Management science, 44: 49-61.

Simar, Leopold, Paul W. Wilson. 2007. Statistical Inference in Nonparametric Frontier Models. In *The Measurement of Productive Efficiency and Productivity Growth.* New York: Oxford University Press.

Shephard, R. W., 1970. Theory of Cost and Production Functions. Princeton: Princeton University Press.

Stokey, N. L., 1998. Are there Limits to Growth? International Economic Review, 39: 1-31.

The Economist, 1977. The Dutch Disease. November 26, 1977, pp. 82-83.

APPENDICES

APPENDIX 1

Proofs

Proposition 1. For the TEQ buyer, quota trade will imply accumulation of capital and output growth above the NTQ level, but below BAU level.

Proof: (17) implies that for any quota buyer

$Z^{TEQ} = \overline{Z} + q > Z^{NTQ}$ and $\phi_K(K^{NTQ}, Z^{NTQ}) = \phi_K(K^{TEQ}, Z^{TEQ})$.

Since $Z^{TEQ} > Z^{NTQ}$ and $\phi_{ZK} > 0$

$\phi_K(K^{TEQ}, Z^{TEQ}) > \phi_K(K^{TEQ}, Z^{NTQ})$ and corollary $\phi_K(K^{NTQ}, Z^{NTQ}) > \phi_K(K^{TEQ}, Z^{NTQ})$,

which implies that

$K_t^{BTEQ} > K_t^{BNTQ}$ and consequently

$Y_t^{BTEQ} > Y_t^{BNTQ}$.

The second part of the statement ($K_t^{BTEQ} < K_t^{BBAU}, Y_t^{BTEQ} < Y_t^{BBAU}$) is easy to prove by noting that $K_1^{BTEQ} < K_1^{BBAU}$ and $Y_1^{BTEQ} < Y_1^{BBAU}$ and using mathematical induction to prove that $K_t^{BTEQ} < K_t^{BBAU}$ and $Y_t^{BTEQ} < Y_t^{BBAU}$ *Q.E.D.*

Proposition 2. For the TEQ seller, quota trade will imply accumulation of capital and output growth below the NTQ level.

Proof: (17) implies that for any quota seller

$Z^{TEQ} = \overline{Z} - q < Z^{NTQ}$ and $\phi_K(K^{NTQ}, Z^{NTQ}) = \phi_K(K^{TEQ}, Z^{TEQ})$.

Since $Z^{TEQ} < Z^{NTQ}$ and $\phi_{ZK} > 0$

$\phi_K(K^{TEQ}, Z^{TEQ}) < \phi_K(K^{TEQ}, Z^{NTQ})$ and corollary $\phi_K(K^{NTQ}, Z^{NTQ}) < \phi_K(K^{TEQ}, Z^{NTQ})$,

which implies that

$K_t^{STEQ} < K_t^{SNTQ}$ and consequently

$Y_t^{STEQ} < Y_t^{SNTQ}$. *Q.E.D.*

APPENDIX 2

Fundamentals of the efficiency and productivity analysis of polluting technologies (following Färe and Primont, 1995 and Färe, Grosskopf, Noh, and Weber, 2005)

Let a vector of N inputs be denoted by $\mathbf{x} = (x_1,...,x_N)$; a vector of M desirable outputs by $\mathbf{y} = (y_1,...,y_M)$ and a vector of K undesirable outputs by $\mathbf{z} = (z_1,...,z_K)$. The technology set is then defined as

$$T \equiv \{(\mathbf{x},\mathbf{y},\mathbf{z}) : \mathbf{y} \in \Re_+^M \text{ can be produced given } \mathbf{x} \in \Re_+^N \text{ and } \mathbf{z} \in \Re_+^K\}.$$

For each input vector $\mathbf{x}$, let $P(\mathbf{x})$ be a set of the feasible (producible) combinations of desirable and undesirable outputs

$$P(\mathbf{x}) = \{(\mathbf{y},\mathbf{z}) : (\mathbf{x},\mathbf{y},\mathbf{z}) \in T\}.$$

This defines an output correspondence, which maps each $\mathbf{x}$ in $\Re_+^N$ to an output set, $P(\mathbf{x}) \subseteqq \Re_+^{M+K}$. Obviously, $(\mathbf{x},\mathbf{y},\mathbf{z}) \in T$ if and only if $(\mathbf{y},\mathbf{z}) \in P(\mathbf{x})$.

I say that the polluting technology set T is regular if the following regularity assumptions are satisfied:

A1. Doing nothing is possible. $(\mathbf{0}_M,\mathbf{0}_K) \in P(\mathbf{x})$ for all $\mathbf{x}$ in $\Re_+^N$.

A2. Strong disposability of desirable outputs. For all $(\mathbf{x},\mathbf{y},\mathbf{z})$ in $\Re_+^{N+M+K}$, if $(\mathbf{y},\mathbf{z}) \in P(\mathbf{x})$ and $\mathbf{y}' \leqq \mathbf{y}$ then $(\mathbf{y}',\mathbf{z}) \in P(\mathbf{x})$.

A3. Strong disposability of inputs. For all $(\mathbf{x},\mathbf{y},\mathbf{z})$ in $\Re_+^{N+M+K}$, if $(\mathbf{y},\mathbf{z}) \in P(\mathbf{x})$ and $\mathbf{x} \leqq \mathbf{x}'$ then $(\mathbf{y},\mathbf{z}) \in P(\mathbf{x}')$.

A4. Scarcity. For all $\mathbf{x}$ in $\Re_+^N$, $P(\mathbf{x})$ is a bounded set.

A5. Output closedness. For all $\mathbf{x}$ in $\Re_+^N$, $P(\mathbf{x})$ is a closed set.

A6. No free lunch. If $(\mathbf{y},\mathbf{z}) \in P(\mathbf{0}_N)$ then $(\mathbf{y},\mathbf{z}) = (\mathbf{0}_M,\mathbf{0}_K)$.

A7. Output convexity. $P(\mathbf{x})$ is convex for all $\mathbf{x}$ in $\Re_+^N$.

While technically undesirable products are technological outputs, their properties more closely resemble those of the inputs. A rational decision making unit (DMU) would want to produce the highest possible economic output while keeping $\mathbf{x}$ and $\mathbf{z}$ constant. Alternatively, a DMU would want to produce the smallest possible undesirable output keeping $\mathbf{y}$ and $\mathbf{x}$ unchanged or produce a given amount of $\mathbf{y}$ and $\mathbf{z}$ with the smallest possible investment of capital.

In spirit of this, the output distance function (ODF) is defined as

$$D_O(\mathbf{x},\mathbf{y},\mathbf{z}) = \inf\left\{\theta : \left(\mathbf{x},\frac{\mathbf{y}}{\theta},\mathbf{z}\right) \in T\right\}.$$

In other words, the ODF score indicates the biggest technologically feasible radial *expansion* of the economic output given the levels of inputs and undesirable outputs.

Similarly, the input distance function (IDF) is defined as

$$D_I(\mathbf{x},\mathbf{y},\mathbf{z}) = \sup\left\{\lambda : \left(\frac{\mathbf{x}}{\lambda},\mathbf{y},\frac{\mathbf{z}}{\lambda}\right) \in T\right\}$$

The IDF score displays the biggest technologically feasible proportional radial *contraction* of inputs and undesirable outputs keeping the economic outputs unchanged.

Both ODF and IDF are complete characterizations of regular technologies and are the most common instruments for estimating production frontiers in the efficiency and productivity analysis literature.

APPENDIX 3

Smooth homogeneous bootstrap and bootstrap statistical inference (adapted from Simar and Wilson, 1998, 2007)

Smooth homogeneous bootstrap algorithm

Let us denote a set of estimated parameters in (26) $\hat{\Xi} = \{\hat{\alpha}_0, \hat{\alpha}_1, \hat{\alpha}_2, \hat{\alpha}_3, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \hat{\gamma}_1, \hat{\gamma}_2, \hat{\gamma}_3\}$. These parameters along with the observed vectors of economic output, *Y*, capital, *K*, and emissions, *Z*, determine a set of estimates of technical efficiencies (IDF scores) for I individual DMUs $\hat{\Lambda} = \{\hat{\lambda}_1, ..., \hat{\lambda}_I\}$. Since a regular kernel estimate does not take into account the condition that $\hat{\Lambda}$ is bounded at 1, with any nonzero bandwidth, a regular kernel estimator suffers from bias in the neighborhood of λ=1. Silverman (1986) proposed to solve this problem by reflecting the values of $\hat{\Lambda}$ by constructing a reflected matrix $L = \{\hat{\Lambda}, 2\mathbf{i} - \hat{\Lambda}\}$, which consists of the original vector $\hat{\Lambda}$ and its values reflected around the unity. Note that as suggested by Simar and Wilson (2007), I ignore spurious values equal to 1, which provides with the spurious mass greater than $1/I$ at the boundary value in the discrete density to be smoothed. These values are merely an artefact of the deterministic efficiency analysis and may be excluded for the purpose of selecting a bandwidth.

Step 1. Calculate bandwidth, *h*, according to the Silverman's adaptive rule $h = 1.06 \min\left\{\sigma_L, \frac{iqr(L)}{1.349}\right\} N^{-0.2}$,

where σ_L is standard error of *L*; and $iqr(L)$ is its interquartile range.

Step 2. Draw a random sample $B^* = \{\beta_1^*, ..., \beta_I^*\}$ with replacements from $\hat{\Lambda}$.

Step 3. Calculate $\tilde{\Lambda}^* = \{\tilde{\lambda}_1^*, ..., \tilde{\lambda}_I^*\}$ as

$$\tilde{\lambda}_i^* = \begin{cases} \beta_i^* + h\varepsilon_i^* & \text{if } \beta_i^* + h\varepsilon_i^* \geq 1 \\ 2 - (\beta_i^* + h\varepsilon_i^*) & \text{otherwise} \end{cases}$$

where ε_i^* is a random deviate drawn from a standard normal distribution, i.e. $\varepsilon_i^* \sim N(0,1)$.

Step 4. As typical when kernel estimators are used, the variance of the bootstrap generated sequence must be corrected by calculating $\Lambda^* = \{\lambda_1^*,...,\lambda_I^*\}$

$$\lambda_i^* = \overline{\beta}^* + \left(1+\frac{h^2}{\hat{\sigma}_\lambda^2}\right)^{-\frac{1}{2}}\left(\widetilde{\lambda}_i^* - \overline{\beta}^*\right),$$

where $\hat{\sigma}_\lambda^2$ is a sample deviation of $\hat{\Lambda}$; $\overline{\beta}^*$ is a sample mean of B^*.

Step 5a. (SW approach) Perturb the original data $\{Y,Z,K\}$ to create a bootstrap sample $\{Y_b^*,Z_b^*,K_b^*\}$ as

$$Y_b^* = Y, \qquad Z_b^* = \left\{\frac{\lambda_1^* Z_1}{\hat{\lambda}_1},...,\frac{\lambda_I^* Z_I}{\hat{\lambda}_I}\right\}, \qquad K_b^* = \left\{\frac{\lambda_1^* K_1}{\hat{\lambda}_1},...,\frac{\lambda_I^* K_I}{\hat{\lambda}_I}\right\}.$$

Step 5a projects each observation to its estimated efficient peer using the estimate of the efficiency measure and then projects it off the frontier using a random efficiency score drawn from the smooth kernel density estimate of the score distribution.

Step 5b. (Cautious alternative) Sample the original data $\{Y,Z,K\}$ with replacements to create a nonsmoothed bootstrap sample $\{\widetilde{Y},\widetilde{Z},\widetilde{K}\}$; use it to create a smoothed bootstrap sample as $\{Y_b^*,Z_b^*,K_b^*\}$

$$Y_b^* = \widetilde{Y}, \qquad Z_b^* = \left\{\frac{\lambda_1^* \widetilde{Z}_1}{\hat{\lambda}_1},...,\frac{\lambda_I^* \widetilde{Z}_I}{\hat{\lambda}_I}\right\}, \qquad K_b^* = \left\{\frac{\lambda_1^* \widetilde{K}_1}{\hat{\lambda}_1},...,\frac{\lambda_I^* \widetilde{K}_I}{\hat{\lambda}_I}\right\}.$$

Step 5b projects each observation in the nonsmoothed bootstrap sample to its estimated efficient peer using the estimate of the efficiency measure for this data point and then projects it off the frontier using a random efficiency score drawn from the smooth kernel density estimate of the score distribution.

Step 6. Run estimator (27) using the perturbed sample $\{Y_b^*, Z_b^*, K_b^*\}$ to obtain bootstrap estimators of the parameters $\hat{\Xi}^*$ of (26).

SHB statistical inference

The SHB procedure generates a set of B bootstrap estimates of the parameters of (26) $\hat{\Xi}^{BS} = \{\hat{\Xi}_1^*, \ldots, \hat{\Xi}_B^*\}$. Then bootstrap estimated bias of $\hat{\Xi}^*$, $Bias(\hat{\Xi}^*)$ is

$$Bias(\hat{\Xi}^*) = \frac{1}{B}\sum_{b=1}^{B}\left[\hat{\Xi} - \hat{\Xi}_b^*\right] = \hat{\Xi} - \frac{1}{B}\sum_{b=1}^{B}\hat{\Xi}_b^* .$$

As usual in the bootstrap literature, I assume that the relationship between the original sample (pseudopopulation) and the bootstrap sample mimics the relationship between the true population and the original sample. Therefore, $Bias(\hat{\Xi}^*) = Bias(\hat{\Xi})$, which results bias corrected values of $\hat{\Xi}$, $\hat{\Xi}^{BC}$

$$\hat{\Xi}^{BC} = \hat{\Xi} + Bias(\hat{\Xi}) = 2\hat{\Xi} - \frac{1}{B}\sum_{b=1}^{B}\hat{\Xi}_b^* .$$

Upper bound, $UB(\hat{\Xi}^{BC})$, and lower bound, $LB(\hat{\Xi}^{BC})$, of the biased corrected estimates' 95% confidence intervals are computed by finding the respective bounds of $\hat{\Xi}^*$. Let $\hat{\Xi}_G^*$ be a (BxI) matrix consisting of all bootstrap estimated $\hat{\Xi}_b^*$, $b = 1, \ldots, B$. Further, let $\hat{\Xi}_R^*$ be a (BxI) matrix obtained from $\hat{\Xi}_G^*$ by ranking elements in each column from the highest to the lowest. Then, the upper bound, $\hat{\Xi}_{UB}^*$ is the (0.025B)th row of $\hat{\Xi}_R^*$, while the lower bound, $\hat{\Xi}_{LB}^*$ is the (0.975B)th row of $\hat{\Xi}_R^*$. Therefore,

$$UB(\hat{\Xi}^{BC}) = \hat{\Xi}^{BC} + \left(\hat{\Xi}_{UB}^* - \frac{1}{B}\sum_{b=1}^{B}\hat{\Xi}_b^*\right)$$

and

$$LB(\hat{\Xi}^{BC}) = \hat{\Xi}^{BC} + \left(\hat{\Xi}_{LB}^* - \frac{1}{B}\sum_{b=1}^{B}\hat{\Xi}_b^*\right).$$

APPENDIX 4

Dynamics of economic output and capital stock for TEQ buyers, sellers and early buyers/late sellers

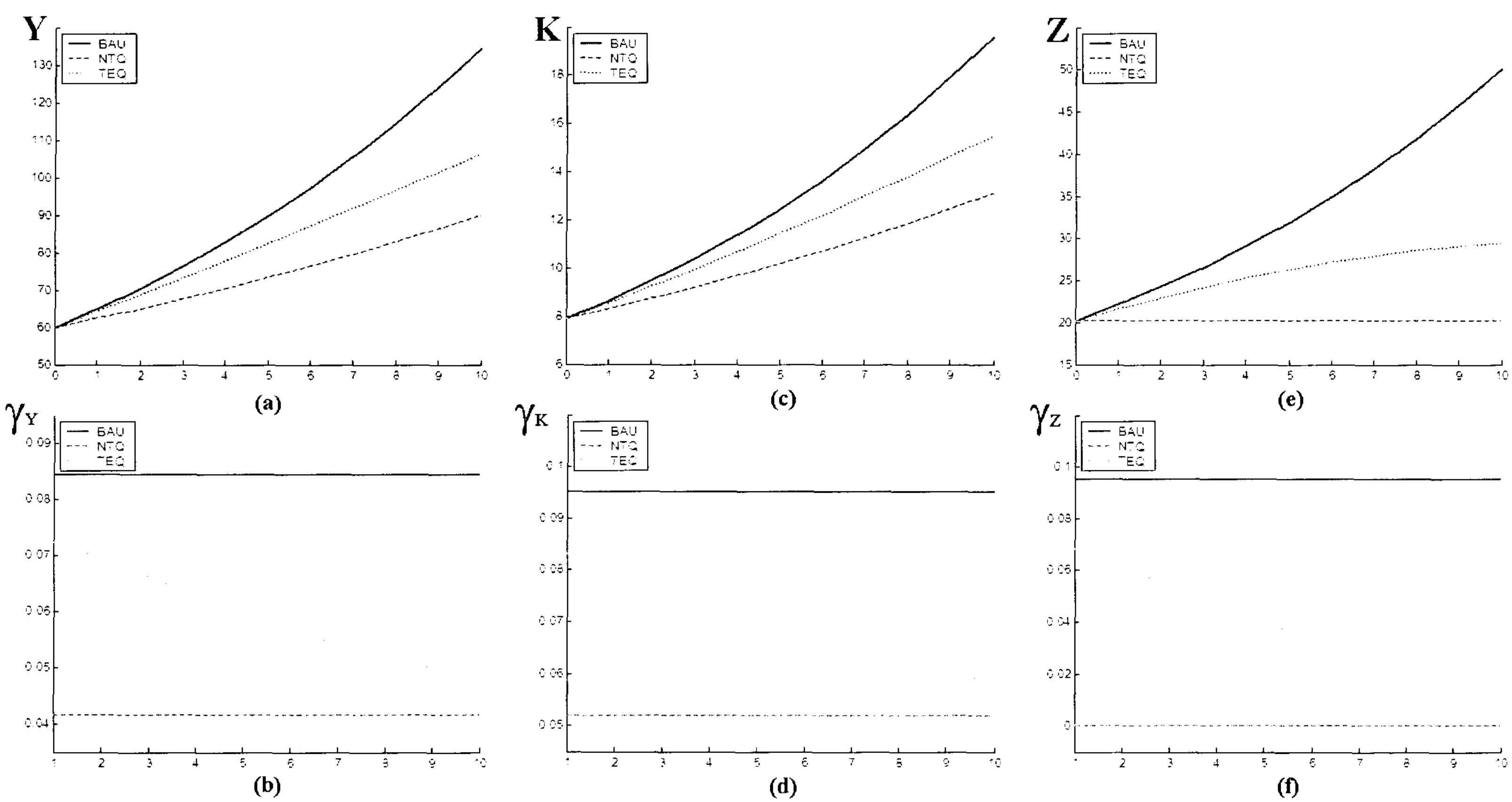

Figure 3-5. Buyer's (Germany) output (a), capital (c) and emission (e) dynamics and the corresponding growth rates (b, d and f respectively).

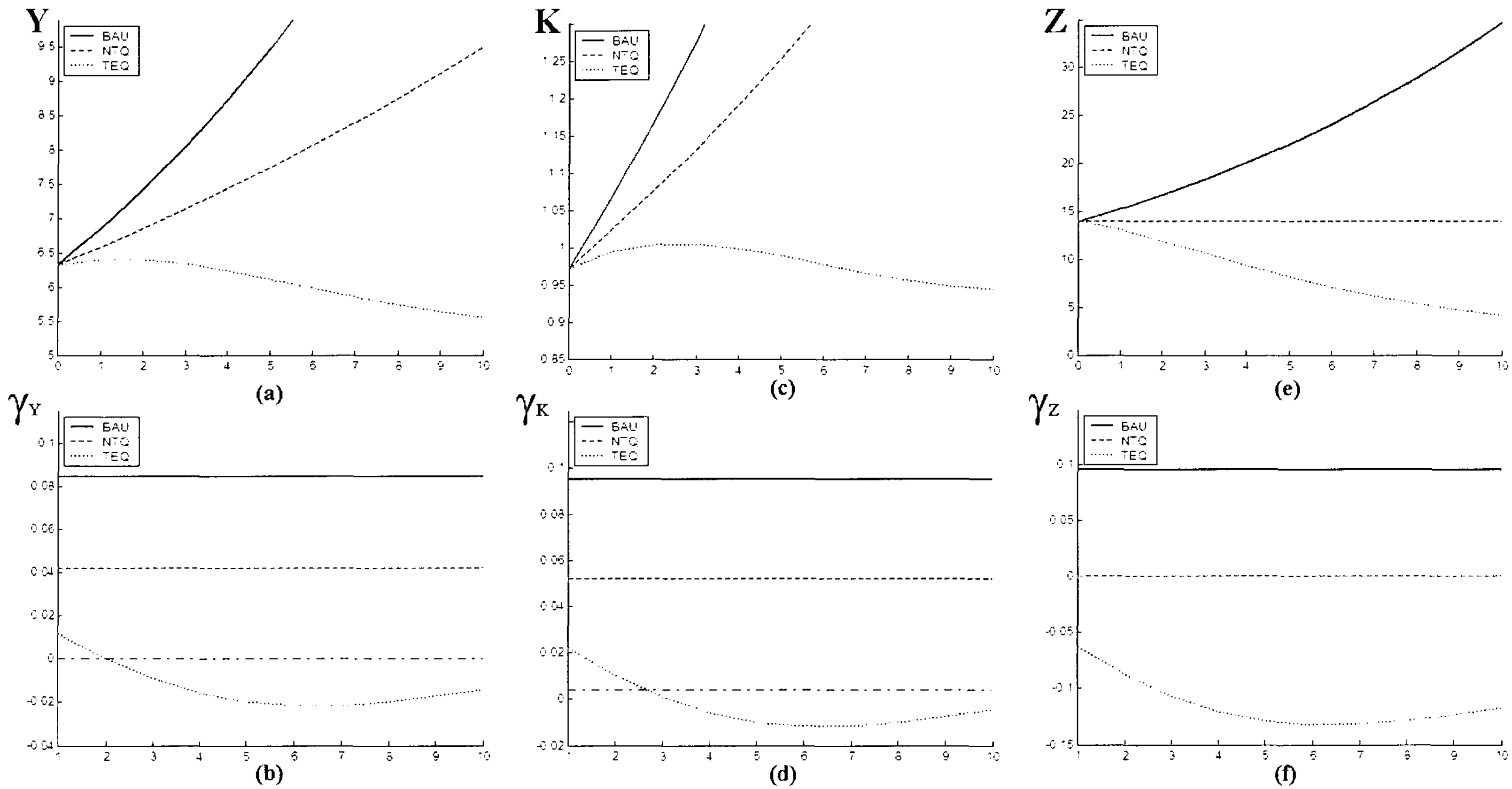

Figure 3-6. Seller's (Latvia) output (a), capital (c) and emission (e) dynamics and the corresponding growth rates (b, d and f respectively).

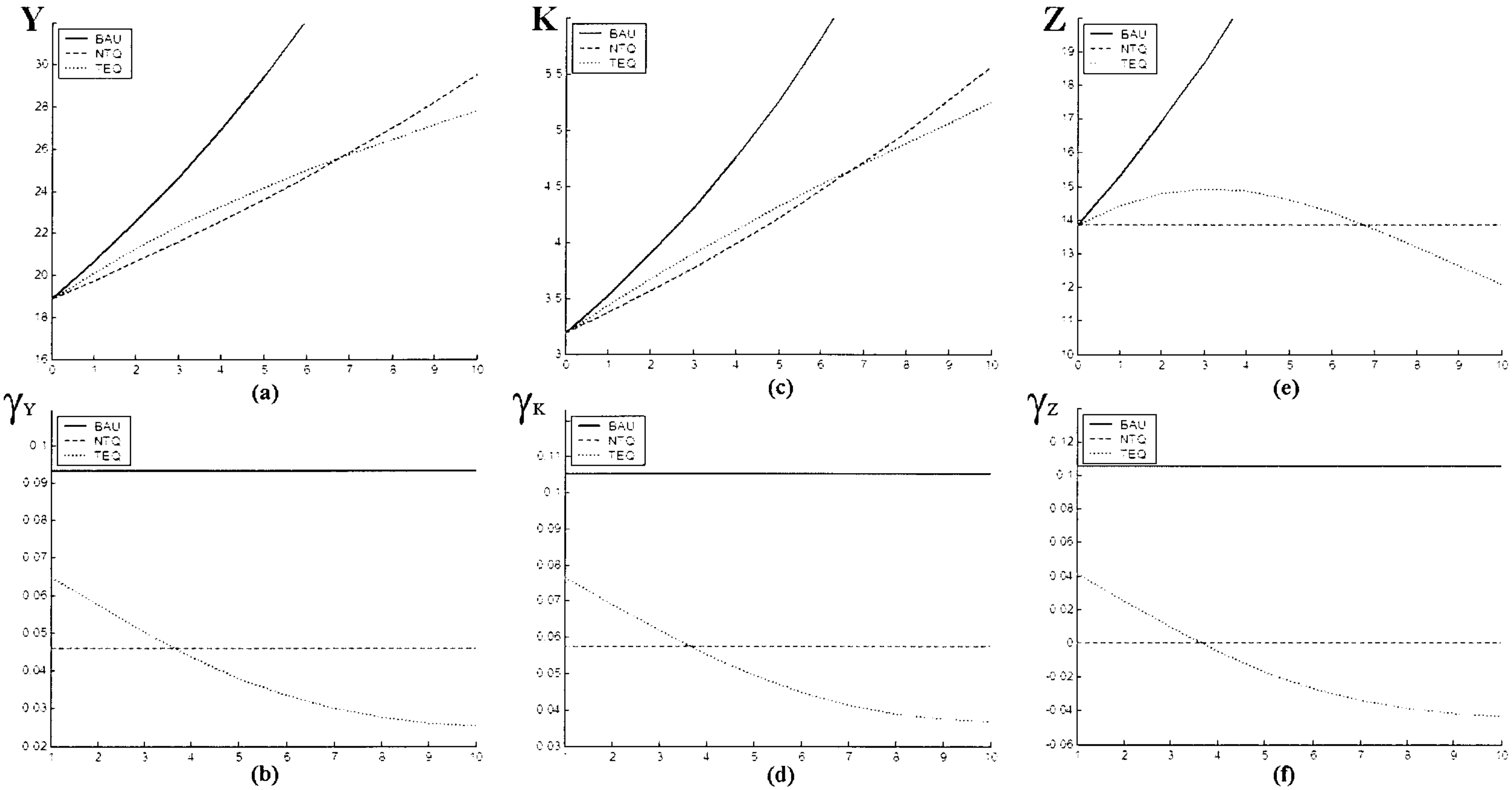

Figure 3-7. Early buyer/late seller's (Slovenia) output (a), capital (c) and emission (e) dynamics and the corresponding growth rates (b, d and f respectively).