FACULTEIT ECONOMIE EN
BEDRIJFSWETENSCHAPPEN

Beyond black box modeling in production: methodological advances in nonparametric methods

Proefschrift voorgedragen tot
het behalen van de graad van
Doctor in de
Economische Wetenschappen

door

Pieter Jan Kerstens

Number 574 2017

Doctoral committee

Advisor:	Prof. dr. Laurens Cherchye (KU Leuven)
Co-advisor:	Prof. dr. Bram De Rock (Université Libre de Bruxelles & KU Leuven)
Members:	Prof. dr. Jaap Bos (Maastricht University) Prof. dr. em. Knox Lovell (University of Queensland) Prof. dr. Tom Van Puyenbroeck (KU Leuven)
Chairman:	Prof. dr. Frits Spieksma (KU Leuven)

Acknowledgements

"A mind without instruction can no more bear fruit than can a field, however fertile, without cultivation."

— Marcus Tullius Cicero

The road to a PhD contains many bumps, obstacles, side roads and hills to climb. Yet, it also takes one past beautiful scenery and along its path one meets many great people. Although I still had to do the walking myself, I had many great people accompanying me. Some of them from the start to the end; others only later on. Without them this PhD would not be possible.

I was fortunate to have two good coaches supporting me over these four years. First of all, prof. dr. Laurens Cherchye, my main advisor, who stood by me with sound advice, a sharp mind and a lot of patience. He gave me every freedom to wander off of the main road to explore its many side roads when pursuing my own research interests. Yet, he also ensured I kept on track to complete this PhD. Among many things, I learned from him that a good idea alone is not enough to make a good paper: how you tell it is equally important. Second, I am indebted to prof. dr. Bram De Rock, my co-advisor, for encouraging me from time to time to take my eyes of the road and to take a closer look at the surroundings and the view. I recall many discussions where – in his no-nonsense style – he encouraged me to try to look at the bigger picture of what we are doing and what we want to do. Together, Laurens and Bram were great coaches.

My sincerest thanks goes to prof. dr. Jaap Bos, prof. dr. em. Knox Lovell and prof. dr. Tom Van Puyenbroeck for their time and effort invested in reading my work as members of my doctoral committee and for their constructive comments that helped to improve this thesis. I enjoyed our stimulating discussions where you offered new ideas and perspectives. I also thank prof. dr. Frits Spieksma for kindly agreeing to be chairman.

Along the way I met a variety of people that had an impact on me one way or the other. Members of the doctoral reading group on household economics at ULB come to mind and in particular, Alex, Denni and Tom. Thank you for the lively discussions both inside and outside the reading group. I am also indebted to Frederic: I learned a lot from our many inspiring discussions and appreciate his unrelenting enthusiasm.

At KULAK I had the good fortune to have nice and supportive colleagues. I appreciate the many lunches we shared, conversations and the occasional trip to the "Blauwe Frituur". Special thanks go to Marijn and my officemates Veerle, Sam, Khushboo, Jenke and Dieter. I would also like to thank the administrative staff at KULAK and FEB Leuven for their support throughout these years. In particular, I thank Caroline, Whitney, Annie and Heidi.

Whenever I was in Leuven I enjoyed a cup of coffee, lunch or the occasional beer with my friends and colleagues of the "MASE". A warm thank you goes to Joris, Geert and Marten. I further thank my "WIT" friends Atul, Bruno, Korneel, Ward and Yves for many great moments. In Antwerpen I could always rely on my "Sint-Ludgardis" friends. Our walks, board games, dinner parties, bar nights and weekends are a welcome distraction to the daily hassles.

I am incredibly fortunate to be surrounded by a loving and supportive family. I would like to thank some of them in particular. First of all, I thank my parents for their continued support, their love, patience and advice. Second, I thank my grandparents who had, and still have, an influence in shaping the person I am today. Third, I thank my sister Carolien for her support and a listening ear. Finally, I am immensely grateful to have met my girlfriend Sanne along the road. Your unconditional love, comfort and understanding are refreshing. I look forward to the rest of our journey together.

Finally, a word of thanks to you, dear reader, for your interest in my PhD thesis. I hope you find the read stimulating and hope that it may inspire you!

Pieter Jan Kerstens
Antwerpen, August 2017.

Contents

General introduction

Nonparametric production analysis supplements observed input-output combinations with a number of assumptions to model the production process. Notwithstanding its simple basic premises, this black box analysis is quite powerful and has low data requirements. However, it can be made even more powerful by opening up this black box. The common thread in this thesis is to go beyond this black box modeling. "Beyond black box modeling" here refers to two interpretations. The first and conventional interpretation refers to more realistic models of production processes by, for example, explicitly modeling the different subprocesses and their links (cfr. Chapter 2) or by modeling intertemporal links between processes over time (cfr. Chapter 4). Apart from this conventional interpretation, it also refers to the idea of looking beyond mere efficiency scores or productivity measures produced by these black box models. An equally important analysis is tracing the underlying factors of these results (cfr. Chapters 2 and 3) and learning from dominating peers (cfr. Chapter 5). This thesis contributes to the existing literature on nonparametric productivity and efficiency analysis in a number of ways.

We first start with a brief crash course in nonparametric production analysis in **Chapter 1**. This chapter is designed to be self-contained and to acquaint the reader with all the necessary tools to understand the following chapters.

It has long been known that the average cost of production, defined as total cost divided by quantity produced, can be reduced by increasing the scale of production (i.e., economies of scale) or by synergies in production (i.e., economies of scope). Either of both choices has its advantages and drawbacks. In agriculture there is a tendency to specialize which is actively promoted by governments. The majority of the existing studies in the efficiency and productivity literature treat agricultural production as a black box where the subprocesses are overlooked. This hampers comparison of mixed and specialized farms to assess whether some reallocation among the different farming activities can be fruitful and how this reallocation could improve productivity over time. **Chapter 2** contributes to the literature by introducing a nonparametric measure of "coordination productivity growth" where the subprocesses are explicitly modeled in the production technology. Put differently: we open up the black box of farming activities by modeling both the crop and livestock activities separately and how both interact. Coordination productivity growth then measures the additional potential gain in productivity by reallocating resources between crop and livestock activities. The coordination productivity indicator is decomposed into a coordination technical inefficiency change component and a coordination technical change component. This decomposition allows assessment of reallocation impacts on the different sources of productivity growth. The empirical application focuses on a large panel of English and Welsh farms over the period $2007-2013$. The results show that coordination inefficiency significantly increases with the proportion of resources allocated to livestock production in economic and statistical terms. Coordination inefficient farms should generally allocate more land to crop production.

The decomposition of total factor productivity is usually concerned with the underlying drivers. These underlying drivers tell us to what extent, for example, productivity growth is due to improved technology, efficiency improvement, or changes in returns-to-scale. However, in general, total factor productivity indicators (indexes) cannot be disentangled into components of aggregate output change and aggregate input change. Thus, one cannot measure to what extent total factor productivity growth (decline) is due to expansion (decline) in aggregate output and decline (expansion) in aggregate input. Productivity indicators (indexes) which do posses this property are said to be “additively (multiplicatively) complete”. **Chapter 3** contributes to the literature by introducing a decomposition of the additively complete Luenberger-Hicks-Moorsteen total factor productivity indicator into the usual components: technical change, technical inefficiency change and scale inefficiency change. Our approach is general in that it does not require differentiability or convexity of the production technology. Therefore, it is applicable to a wide variety of production technologies. Using a nonparametric framework, the empirical application focuses on the agricultural sector at the state-level in the U.S. over the period $1960-2004$. The results show that Luenberger-Hicks-Moorsteen total factor productivity increased substantially in the considered period. This productivity growth is due to output growth rather than input decline, although the extent depends on the convexity assumption of the technology. Technical change is the main driver, while the role of technical inefficiency change and scale inefficiency change also depends on the convexity assumption of the technology.

Production at one point in time is not independent of earlier production. Many inputs influence production over many future time periods. This has long been recognized and studied in the literature. However, much of this literature has focused on dynamically efficient production behavior from a technical perspective (i.e., through explicit modeling of the production process). By contrast, far less work has tackled the issue from an economic perspective (i.e., through behavioral modeling of firms' operations). **Chapter 4** proposes a nonparametric methodology for intertemporal production analysis that accounts for durable as well as storable inputs. Durable inputs contribute to the production outputs in multiple consecutive periods. Storable inputs are non-durable and can be stored in inventories for use in future periods. We explicitly model the possibility that firms use several vintages of the durable inputs, i.e., they invest in new durables and scrap older durables over time. Furthermore, we allow for production delays of durable inputs. We characterize production behavior that is dynamically cost efficient, which allows us to evaluate the efficiency of observed production decisions. For cost inefficient behavior, we propose a measure to quantify the degree of inefficiency. An attractive feature of this measure is that it can be decomposed in period-specific cost inefficiencies. We demonstrate the usefulness of our methodology through an application on Swiss railway companies.

In empirical applications of efficiency and productivity analysis an often disregarded aspect is the direction in which to project onto the production frontier. Although some literature started focusing on choosing this direction in some optimal way, perhaps the most common choice is the radial direction. However, little attention is paid to the managerial interpretations of such a choice. Furthermore, empirical applications often focus exclusively on modeling the production process and computing efficiency/productivity. The analysis often stops there even though the next step is equally – perhaps even more – important: what are the important peers? From which should a specific DMU learn? **Chapter 5** attempts to tackle both issues. As a by product of efficiency computations we have a matrix of intensity variables. This matrix represents a graph showing the dominating peers that make up the (artificial) benchmark DMU for every DMU. We propose to identify the key DMUs in the DMU network by varying the direction vectors and aggregating these matrices of intensity variables. The key DMUs are then those DMUs with the largest eigenvector centrality which is computed from the eigenvalue decomposition of this aggregate matrix of intensity variables. Through a comparative analysis we identify key objectives from the key DMUs characteristics. These key objectives are used as direction vectors in subsequent analysis.

Once efficiency scores are computed and dominating peers are identified for a given DMU, we need to analyze these individual dominating peers to determine what one can learn from them. Visualizing these dominating peers in comparison to the benchmarked DMU can simplify the task at hand. We are not aware of any visualization tool in the literature that allows to quickly compare DMUs in terms of their input-output mix and scale. Therefore, we further propose a visualization tool to compare dominating peers with the benchmarked DMU in terms of input-output mix and scale. Our tool is easy to compute and allows to quickly visualize all dominating peers in comparison to a benchmarked DMU and can complement, for example, radar plots. Thus, the tool enriches the toolbox of efficiency analysis and helps in moving beyond the conventional analysis.

Each of these chapters was originally conceived as separate papers and the chapters are therefore self-contained. This also means that some of the arguments and definitions are repeated in some of the chapters and notation can differ over the chapters. We next turn to a basic introduction of production theory. The focus is on nonparametric techniques to analyze production behavior. This should familiarize the reader with the tools used throughout this thesis.

Chapter **1**

Introduction: A crash course on nonparametric analysis of production

"The neoclassical theory of production postulates that firms maximize profits (and minimize costs) subject to certain technological constraints. ... The conventional analysis of these questions proceeds by first postulating a parametric form for the production function (or some equivalent parametric representation of the technology) and then using standard statistical techniques to estimate the unknown parameters from the observed data. This procedure suffers from the defect that the maintained hypothesis of parametric form can never be directly tested: it must be taken on faith."

— Hal R. Varian[1]

Efficiency analysis and productivity is the main object of interest in this thesis. Before turning to the contributions in this thesis it is useful to provide a "crash course" on the basic concepts that are used in the following chapters. Thus, this chapter should equip the reader with the necessary tools to dive into the following chapters. We introduce basic technical concepts and do a limited literature review so that this thesis is self-contained. But, our goal is not to be exhaustive. To focus our discussion, a selection is made and we refer interested readers to the books of, for example, Fried et al. (1993) and Färe and Primont (1995) for a more in depth discussion.

1.1 Introduction

In a production context, nonparametric analysis refers to the tradition of the axiomatic production theory in line of Farrell (1957), Shephard (1970), Afriat (1972) and Varian (1984). Axiomatic production theory starts from a minimal set of axioms to characterize the production possibilities set and some behavioral assumptions. In contrast, parametric methods of production analysis posit a specific functional form for the technology and an additive composite stochastic error term. This stochastic composite error term

[1]Varian (1984, p.579)

captures both inefficiency and measurement error. They frequently proceed by further assuming a specific distribution for both error terms and then estimating the technology through econometric estimation. Their advantage is the natural way in which various forms of uncertainty are handled while their main disadvantage is clearly the high risk of functional form misspecification. This is largely avoided by nonparametric methods we employ in this thesis: the functional form is not imposed a priori but is determined from the data itself. In this way the "data speaks for itself". An important caveat is that nonparametric methods can still suffer from misspecification e.g., due to outliers, the data generating process and dynamic misspecification (e.g., Kasparis and Phillips (2012)).

Farrell (1957)'s seminal work laid the foundation for modern efficiency analysis. He proposed to envelop the data to provide a conservative estimate of the production possibilities set. Then, he defined "technical efficiency" as the maximal output that could be produced for a given inputs. Similarly, "price efficiency" (or cost efficiency) was defined. Both proposed efficiency measures were ratio based so as to give them an easy interpretation. Shephard (1970) provided a formal axiomatic framework to characterize production. Following the ideas of Farrell and Shephard, Charnes et al. (1978) provided an easy implementable linear programming formulation of what has become known as Data Envelopment Analysis (DEA). Their method also has the advantage of being able to deal straightforwardly with multiple inputs and multiple outputs. Building on work by Afriat (1972), Hanoch and Rothschild (1972) and Diewert and Parkan (1983), Varian (1984) showed how to test cost minimizing behavior by means of the "Weak Axiom of Cost Minimization" (WACM) by assuming free disposability of inputs, nestedness of input requirement sets and cost minimizing behavior. One can then determine how far a firm deviates from cost minimizing behavior by measuring cost efficiency as a ratio of minimal cost over observed cost such as proposed by Farrell. Varian proposed a similar test for profit maximization and showed how to recover information about the underlying technology from cost minimizing or profit maximizing behavior. Finally, Banker and Maindiratta (1988) extend Varian's work and show the link with DEA models.

1.1.1 The firm's objective: the pursuit of profit

In line with the neo-classical framework one frequently assumes that firms are price takers; that one firm cannot influence the market price ("atomism") and that it operates in a market with perfect competition. Let $\mathbf{Z} = (-\mathbf{X}, \mathbf{Y}) \in \mathbb{R}^n_- \times \mathbb{R}^m_+$ be a "netput" vector where inputs are assumed negative and outputs positive. The firm's profit maximization problem can then be stated as (Mas-Colell et al., 1995):

$$\max_{\mathbf{Z}} \mathbf{P} \cdot \mathbf{Z}' \tag{1.1a}$$

$$\text{s.t. } F(\mathbf{Z}) \leq 0, \tag{1.1b}$$

where the transformation function $F(\cdot)$ is differentiable and has the property that $F(\mathbf{Z}^*) = 0$ for all points $\mathbf{Z}^*$ on the boundary. $F(\cdot)$ represents the production technology. The production trade-off at the boundary between any two netputs l and k is given by the marginal rate of transformation (MRT). The MRT shows the marginal change in netput k due to a marginal change in netput l. The MRT is found by totally differentiating $F(\mathbf{Z}^*) = 0$

$$\frac{\partial F(\mathbf{Z}^*)}{\partial Z^{*l}} dZ^{*l} + \frac{\partial F(\mathbf{Z}^*)}{\partial Z^{*k}} dZ^{*k} = 0,$$

and rearranging terms:

$$MRT_{k,l} \equiv \frac{dZ^{*k}}{dZ^{*l}} = -\frac{\frac{\partial F(\mathbf{Z}^*)}{\partial Z^{l*}}}{\frac{\partial F(\mathbf{Z}^*)}{\partial Z^{k*}}}. \tag{1.2}$$

From the first order conditions of optimality for (1.1) we know that an optimal solution $\mathbf{Z}^*$ must satisfy

$$\mathbf{P} = \lambda \cdot \nabla F(\mathbf{Z}^*),$$

for some $\lambda \geq 0$. Then, for any netput l and k we have

$$\frac{P^l}{P^k} = \frac{\frac{\partial F(\mathbf{Z}^*)}{\partial Z^{l*}}}{\frac{\partial F(\mathbf{Z}^*)}{\partial Z^{k*}}} = -MRT_{k,l}. \tag{1.3}$$

Thus, under profit maximizing behavior the ratio of prices equals minus the marginal rate of transformation. This relation forms the fundamental building block to (i) reconstruct the technology set from observed profit maximizing choices and (ii) recover relative "shadow" prices from knowledge of the technology set.

Figure 1.1 illustrates the above graphically for the case of two inputs holding output constant. The dots are the observed input combinations to produce the output $\mathbf{Y}$. The dotted lines are the prices faced in the different situations (e.g., different firms with different prices or single firm over time). Under profit maximization we know that these relative prices are tangent to the technology frontier. Thus, an "outer" approximation of the technology set is then marked by the different relative prices. Conversely, given different input combinations and knowledge of the technology set $\mathcal{I}(\mathbf{Y})$, we can estimate the relative prices a profit maximizing firm would have faced.

The remaining parts of this chapter serve first to introduce an axiomatic approach to reconstruct the technology set in the absence of price information and the measurement of technical efficiency. The second part then discusses a nonparametric way of testing "economic optimizing" (e.g., profit maximizing or cost minimizing) behavior with minimal assumptions regarding the underlying technology. The third section demonstrates how the two previous sections are linked together by duality. Finally, we discuss measures of productivity before concluding in the final section.

1.2 Technical efficiency

The transformation function $F(\cdot)$ in (1.1) can also be represented through a production set. The remaining part of this chapter uses this set representation of technology.

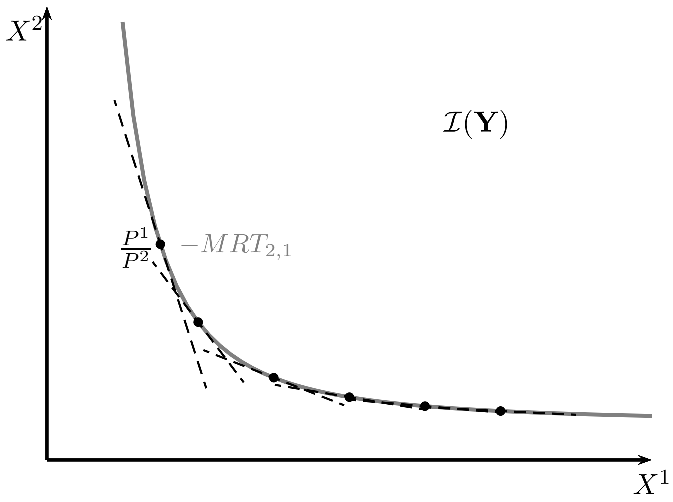


Figure 1.1: Cost minimization: $\frac{P^1}{P^2} = -MRT_{2,1}$.

1.2.1 Set representation of technology

The production literature uses production possibilities sets as alternative representations of technology. Let the technology set $\boldsymbol{\mathcal{Y}} \subset \mathbb{R}^n_+ \times \mathbb{R}^m_+$ be defined as:

$$\boldsymbol{\mathcal{Y}} = \left\{ (\mathbf{X}, \mathbf{Y}) \in \mathbb{R}^{n+m}_+ | \mathbf{X} \text{ can produce } \mathbf{Y} \right\}. \tag{1.4}$$

This technology set simply collects all feasible input-output combinations. One can take slices of this set by fixing the inputs or the outputs yielding the output sets

$$\mathcal{P}(\mathbf{X}) = \left\{ \mathbf{Y} \in \mathbb{R}^m_+ | \mathbf{X} \text{ can produce } \mathbf{Y} \right\}, \tag{1.5}$$

and the input requirements set

$$\mathcal{I}(\mathbf{Y}) = \left\{ \mathbf{X} \in \mathbb{R}^n_+ | \mathbf{X} \text{ can produce } \mathbf{Y} \right\}. \tag{1.6}$$

The technology set is commonly assumed to satisfy the following axioms:

Axiom 1.1 (inaction)**.** $(\mathbf{0}_n, \mathbf{0}_m) \in \boldsymbol{\mathcal{Y}}$.

Axiom 1.2 (no free lunch)**.** *if* $(\mathbf{X}, \mathbf{Y}) \in \boldsymbol{\mathcal{Y}}$ *and* $\mathbf{X} = \mathbf{0}_n$*, then* $\mathbf{Y} = \mathbf{0}_m$.

Axiom 1.3 (closedness)**.** $\boldsymbol{\mathcal{Y}}$ *is closed.*

Axiom 1.4 (free disposability of inputs and outputs)**.** $(\mathbf{X}, \mathbf{Y}) \in \boldsymbol{\mathcal{Y}}$ *and* $(\mathbf{X}', -\mathbf{Y}') \geq (\mathbf{X}, -\mathbf{Y}) \Rightarrow (\mathbf{X}', \mathbf{Y}') \in \boldsymbol{\mathcal{Y}}$.

Axiom 1.1 states that producing nothing is possible, while Axiom 1.2 states that nothing must come out of nothing, i.e., there is no such thing as a free lunch. Axiom 1.3

is a mathematical regularity property. Axiom 1.4 then states that disposal of inputs or outputs is costless.[2] In addition, often convexity of the technology set is assumed:

Axiom 1.5 (convexity)**.** *$\mathcal{Y}$ is a convex set:* $(\mathbf{X},\mathbf{Y}) \in \mathcal{Y}$ *and* $(\mathbf{X}',\mathbf{Y}') \in \mathcal{Y} \Rightarrow \forall\lambda \in [0,1]:$ $\lambda(\mathbf{X},\mathbf{Y}) + (1-\lambda)(\mathbf{X}',\mathbf{Y})' \in \mathcal{Y}$.

Convexity implies time or space divisibilities of inputs and outputs. It is often perceived as innocuous, while in fact it is a quite stringent assumption on the underlying production technology. In contrast, for economic efficiency convexity of the technology set (input requirement set) does not alter profit efficiency (cost efficiency) evaluations (cfr. Section 1.3). Examples of a hypothetical input requirement set and output set satisfying the above axioms are depicted in Figure 1.2 and Figure 1.3.

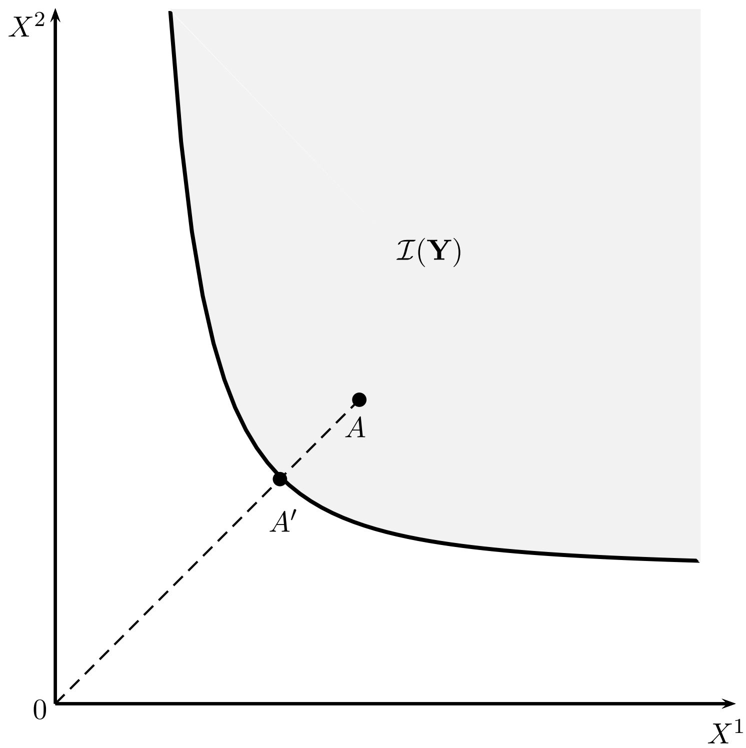


Figure 1.2: Input requirement set with an inefficient observation A and its weakly efficient projection A' on the frontier.

We remark that these production possibilities sets can also be defined over subsets of the inputs and outputs. One example is Chapter 2 where we use subprocess production possibilities sets to model different farming activities. Another example is the multi-output production models of Cherchye et al. (2013, 2014) and Cherchye et al. (2016)

[2]There are many situations where this assumption is questionable: e.g., in the case of congestion or pollution-generating activities. We refer to Dakpo et al. (2016) for an extensive review of different approaches to model pollution-generating activities.

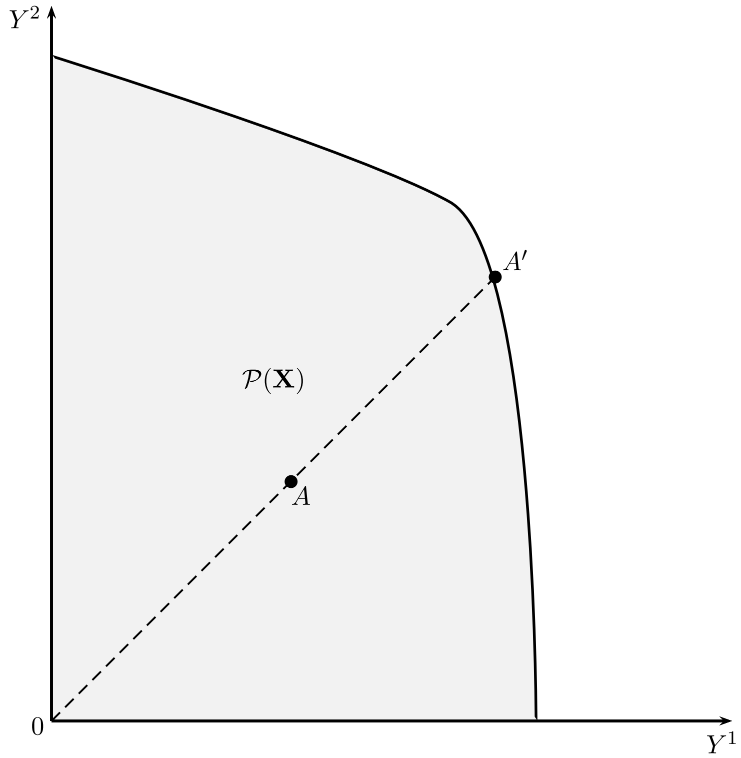


Figure 1.3: Output set with an inefficient observation A and its weakly efficient projection A' on the frontier.

where they model output-specific technologies which have their own output-specific inputs and joint inputs shared by every output. Naturally, these production possibilities sets can have their own different set of axioms.

Intuitively, a firm is "technically efficient" if it produces the maximum amount for all outputs with the minimal amount of all inputs (Koopmans, 1951). More formally, an observation $(\mathbf{X}_0, \mathbf{Y}_0) \in \boldsymbol{\mathcal{Y}}$ is "Pareto-Koopmans efficient" if

$$\{(\mathbf{X}, \mathbf{Y}) \in \boldsymbol{\mathcal{Y}} | (\mathbf{X}, -\mathbf{Y}) \leq (\mathbf{X}_0, -\mathbf{Y}_0)\} = \emptyset.$$

The Pareto-Koopmans efficient frontier is characterized by

$$Eff\, \boldsymbol{\mathcal{Y}} \equiv \{(\mathbf{X}, \mathbf{Y}) \in \boldsymbol{\mathcal{Y}} |\ \nexists (\mathbf{X}', \mathbf{Y}') \in \boldsymbol{\mathcal{Y}} : (\mathbf{X}', -\mathbf{Y}') \leq (\mathbf{X}, -\mathbf{Y})\}.$$

One can straightforwardly define similar Pareto-Koopmans efficient input requirement sets (output sets) by fixing outputs (inputs). Tone (2001)'s Slacks-Based Measure (SBM) model is one example among many implementing a Pareto-Koopmans based efficiency measure. Besides this "strong efficiency" notion, a firm can be "weakly efficient" if it is not possible to simultaneously reduce all its inputs and expand all its outputs proportionally by a common factor:

$$Isoq\, \boldsymbol{\mathcal{Y}} \equiv \left\{(\mathbf{X}, \mathbf{Y}) \in \boldsymbol{\mathcal{Y}} | \forall \theta \in [0, 1) : \left(\theta \mathbf{X}, \frac{\mathbf{Y}}{\theta}\right) \notin \boldsymbol{\mathcal{Y}}\right\}. \tag{1.7}$$

This technology set isoquant contains all weakly efficient input-output combinations. The hyperbolic graph efficiency measure of Färe et al. (1985) implements this "weakly efficient" notion by computing the radial distance to $Isoq\,\boldsymbol{\mathcal{Y}}$. Clearly, $Eff\,\boldsymbol{\mathcal{Y}} \subseteq Isoq\,\boldsymbol{\mathcal{Y}}$. Similarly, isoquants for the input requirement set and the output set can be defined:

$$Isoq\,\mathcal{I}(\mathbf{Y}) \equiv \{\mathbf{X} \in \mathcal{I}(\mathbf{Y}) | \forall \theta \in [0,1) : \theta\mathbf{X} \notin \mathcal{I}(\mathbf{Y})\}\,, \tag{1.8}$$

$$Isoq\,\mathcal{P}(\mathbf{X}) \equiv \{\mathbf{Y} \in \mathcal{P}(\mathbf{X}) | \forall \theta \in (1,\infty) : \theta\mathbf{Y} \notin \mathcal{P}(\mathbf{X})\}\,. \tag{1.9}$$

These isoquants serve as the basis for Shephard (1970)'s distance functions to determine technical efficiency and quantify the deviation from technical efficiency. Graphically, these isoquants correspond to contraction (expansion) of all inputs (outputs) along a ray originating in the origin. This is illustrated in Figure 1.2 as follows: the inefficient observation A is contracted towards the origin until it reaches the frontier at point A'. Figure 1.3 shows how the inefficient observation A is expanded along a ray through the origin until it reaches the frontier at A'.

1.2.2 Measurement of efficiency

If a firm is not technically efficient then it is useful to measure the degree of inefficiency and determine the technical efficient input-output combination. There are a wide variety of different measurements of (in)efficiency. We only discuss those that straightforwardly follow from our previous discussion.

The hyperbolic graph efficiency measure of Färe et al. (1985) measures the distance of a given observation to the weakly efficient frontier (1.7) from the origin:

$$\gamma(\mathbf{X},\mathbf{Y}) \equiv \inf\left\{\gamma | \left(\gamma\mathbf{X}, \frac{\mathbf{Y}}{\gamma}\right) \in Isoq\,\boldsymbol{\mathcal{Y}}\right\}. \tag{1.10}$$

Similarly, Shephard (1970)'s input distance function computes the distance along a ray from the origin to $Isoq\,\mathcal{I}(\mathbf{Y})$ while holding outputs constant:

$$\theta(\mathbf{X},\mathbf{Y}) \equiv \sup\left\{\theta | \left(\frac{\mathbf{X}}{\theta}, \mathbf{Y}\right) \in Isoq\,\mathcal{I}(\mathbf{Y})\right\}. \tag{1.11}$$

This measure is equal to or larger than one. Analogously, Shephard (1970)'s output distance function computes the distance along a ray from the origin to $Isoq\,\mathcal{P}(\mathbf{X})$ while holding inputs constant:

$$\delta(\mathbf{X},\mathbf{Y}) \equiv \inf\left\{\delta | \left(\mathbf{X}, \frac{\mathbf{Y}}{\delta}\right) \in Isoq\,\mathcal{P}(\mathbf{X})\right\}. \tag{1.12}$$

This output distance function is situated between zero and one. These distance measures have in common that they compute the distance to the technology set along a ray from the origin. However, nothing prevents us from choosing a different "direction" of measurement. A more general distance function measuring (in)efficiency in a chosen direction is based on Luenberger (1992)'s shortage function. First introduced in a consumption setting by Luenberger, it was then introduced by Chambers et al. (1996b) in a production setting as the directional distance function:

$$D(\mathbf{X}, \mathbf{Y}; \mathbf{g}) = \sup \left\{ \beta \in \mathbb{R} : (\mathbf{X} - \beta \mathbf{g}_x, \mathbf{Y} + \beta \mathbf{g}_y) \in \boldsymbol{\mathcal{Y}} \right\}, \tag{1.13}$$

where $\mathbf{g} = (\mathbf{g}_x, \mathbf{g}_y) \in \mathbb{R}^{n+m}$ is the direction vector. Intuitively, $D(\mathbf{X}, \mathbf{Y}; \mathbf{g})$ then projects $(\mathbf{X}, \mathbf{Y})$ onto the boundary of the technology set by simultaneously scaling the inputs and outputs in the chosen direction $\mathbf{g}$ by a common factor β.

Shephard's distance functions (1.11) and (1.12) arise as special cases of (1.13). Specifically, $D(\mathbf{X}, \mathbf{Y}; (\mathbf{X}, \mathbf{0}_m)) = 1 - \theta(\mathbf{X}, \mathbf{Y})^{-1}$ and $D(\mathbf{X}, \mathbf{Y}; (\mathbf{0}_n, \mathbf{Y})) = \delta(\mathbf{X}, \mathbf{Y})^{-1} - 1$. Furthermore, the hyperbolic graph efficiency measure $\log \gamma^{-1}(\mathbf{X}^*, \mathbf{Y}^*)$ over the transformed production set

$$\boldsymbol{\mathcal{Y}}^* = \left\{ (\mathbf{X}^*, \mathbf{Y}^*) = (\exp(\mathbf{X}./\mathbf{g}_x), \exp(\mathbf{Y}./\mathbf{g}_y)) \in \mathbb{R}_+^{n+m} | \mathbf{X} \text{ can produce } \mathbf{Y} \right\}$$

can be shown to equal (1.13) (Simar and Vanhems, 2012). Here, ./ denotes element-wise division.

Chambers et al. (1998, Lemma 2.2) prove the following properties of the directional distance function:

Proposition 1.1. *If $\boldsymbol{\mathcal{Y}}$ satisfies Axiom 1.1–1.4, then $D(\mathbf{X}, \mathbf{Y}; \mathbf{g})$ has the following properties:*

(a) $D(\mathbf{X} - \alpha \mathbf{g}_x, \mathbf{Y} + \alpha \mathbf{g}_y; \mathbf{g}) = D(\mathbf{X}, \mathbf{Y}; \mathbf{g}) - \alpha$*;*

(b) $D(\mathbf{X}, \mathbf{Y}; \mathbf{g})$ *is upper semicontinuous in* $\mathbf{X}$ *and* $\mathbf{Y}$ *(jointly);*

(c) $D(\mathbf{X}, \mathbf{Y}; \alpha \mathbf{g}) = \frac{1}{\alpha} D(\mathbf{X}, \mathbf{Y}; \mathbf{g}), \alpha > 0$*;*

(d) $\mathbf{Y}' \geq \mathbf{Y} \Rightarrow D(\mathbf{X}, \mathbf{Y}'; \mathbf{g}) \leq D(\mathbf{X}, \mathbf{Y}; \mathbf{g})$*;*

(e) $\mathbf{X}' \geq \mathbf{X} \Rightarrow D(\mathbf{X}', \mathbf{Y}; \mathbf{g}) \geq D(\mathbf{X}, \mathbf{Y}; \mathbf{g})$*;*

(f) if $\boldsymbol{\mathcal{Y}}$ *is convex (i.e., Axiom 1.5),* $D(\mathbf{X}, \mathbf{Y}; \mathbf{g})$ *is concave in* $(\mathbf{X}, \mathbf{Y})$*.*

Property (a) is the translation invariance property showing that translating an observation in the direction of measurement results in a translation by the same factor of the directional distance function evaluated on the original observation. In essence, Property (b) means that the value of the directional distance function for arguments near $(\mathbf{X}, \mathbf{Y})$ is equal to or lower than $D(\mathbf{X}, \mathbf{Y}; \mathbf{g})$. This technical property is weaker than continuity. Scaling the direction vector is equivalent to inversely scaling the original directional distance function (Property (c)). The directional distance function is monotone decreasing in outputs and monotone increasing in inputs as denoted by Properties (d) and (e). Finally, convexity of the technology set implies a concave directional distance function (Property (f)).

1.2.3 Empirical approximation of production sets

In practice, the production technology is not known to the empirical analyst and has to be estimated from some data $S = \{(\mathbf{X}_k, \mathbf{Y}_k)\}_{k=1,\dots,K}$. Farrell (1957)'s seminal work laid the foundation for modern efficiency analysis. He proposed to envelop the data to provide a conservative estimate of the production possibilities set. Then, he defined "technical efficiency" as the maximal output that could be produced for given inputs. Shephard (1970) provided a formal axiomatic framework to characterize production and his input (1.11) and output (1.12) distance functions, directly borrowed from Farell, operationalize (1.8) and (1.9). Building on the ideas of Farrell and Shephard, Charnes et al. (1978) provided an easy implementable linear programming (LP) formulation of what has become known as DEA. Their method also has the advantage of being able to easily deal with multiple inputs and multiple outputs.

In order to arrive at an empirical approximation of the technology set, one additional axiom is required:

Axiom 1.6 (observability means feasibility)**.** $\forall(\mathbf{X}_k, \mathbf{Y}_k) \in S \Rightarrow (\mathbf{X}_k, \mathbf{Y}_k) \in \boldsymbol{\mathcal{Y}}$.

Hence, we assume that all observations are measured without noise. The above axioms can be supplemented with particular assumptions on returns-to-scale.

Axiom 1.7 (returns-to-scale)**.** $\boldsymbol{\mathcal{Y}}$ *exhibits*

- *constant returns-to-scale (CRS):* $(\mathbf{X}, \mathbf{Y}) \in \boldsymbol{\mathcal{Y}} \Rightarrow (\delta\mathbf{X}, \delta\mathbf{Y}) \in \boldsymbol{\mathcal{Y}},\ \forall\delta > 0$*;*
- *non-increasing returns-to-scale (NIRS):* $(\mathbf{X}, \mathbf{Y}) \in \boldsymbol{\mathcal{Y}} \Rightarrow (\delta\mathbf{X}, \delta\mathbf{Y}) \in \boldsymbol{\mathcal{Y}},\ \forall\delta \in [0, 1]$*;*
- *non-decreasing returns-to-scale (NDRS):* $(\mathbf{X}, \mathbf{Y}) \in \boldsymbol{\mathcal{Y}} \Rightarrow (\delta\mathbf{X}, \delta\mathbf{Y}) \in \boldsymbol{\mathcal{Y}},\ \forall\delta \geq 1$*;*
- *variable returns-to-scale (VRS): otherwise.*

CRS means that an observation can be scaled up or down by any positive factor. NIRS means that an observation can only be scaled down, while NDRS implies that an observation can only be scaled up. Finally, VRS satisfies NIRS and NDRS in different regions of the production possibilities set (Färe et al., 1994a). These different returns-to-scale assumptions can be imposed to further strengthen the empirical bite of the technology approximation, but imposing them can lead to misspecification error.

Extending Bogetoft (1996), Briec et al. (2004) provided a unified algebraic representation for the minimum extrapolating technology consistent with the data S (Axiom 1.6), Axioms 1.3-1.4, the different returns to scale assumptions (Axiom 1.7) for both convex (i.e., Axiom 1.5) and non-convex technologies:

$$\boldsymbol{\mathcal{Y}}^{\Lambda,\Gamma} = \left\{ (\mathbf{X}_0, \mathbf{Y}_0) | \sum_{k=1}^{K} \delta\lambda_k \mathbf{X}_k \leq \mathbf{X}_0, \sum_{k=1}^{K} \delta\lambda_k \mathbf{Y}_k \geq \mathbf{Y}_0, \lambda_k \in \Lambda, \delta \in \Gamma \right\}, \tag{1.14a}$$

where

$$\Lambda \in \begin{cases} NC & = \left\{ \boldsymbol{\lambda} \in \mathbb{R}_{+}^{K} | \sum_{k=1}^{K} \lambda_k = 1, \lambda_k \in \{0,1\} \right\}; \\ C & = \left\{ \boldsymbol{\lambda} \in \mathbb{R}_{+}^{K} | \sum_{k=1}^{K} \lambda_k = 1, \lambda_k \geq 0 \right\} \end{cases} \tag{1.14b}$$

and

$$\Gamma \in \begin{cases} VRS & = \{\delta \in \mathbb{R}_{+} | \delta = 1\}; \\ CRS & = \{\delta \in \mathbb{R}_{+} | \delta \geq 0\}; \\ NIRS & = \{\delta \in \mathbb{R}_{+} | 0 \leq \delta \leq 1\}; \\ NDRS & = \{\delta \in \mathbb{R}_{+} | \delta \geq 1\}. \end{cases} \tag{1.14c}$$

In contrast to other representations, this formulation has the advantage that the returns to scale parameter δ is separate from the usual intensity variable $\boldsymbol{\lambda}$. The returns to scale parameters in Γ allow to scale the observations in line with the different returns to scale assumptions of Axiom 1.7. The intensity variables λ_k serve as weights (i.e., their sum is one) to construct a benchmark observation $\left(\sum_{k=1}^{K} \delta\lambda_k \mathbf{X}_k, \sum_{k=1}^{K} \delta\lambda_k \mathbf{Y}_k\right)$ as a scaled, weighted sum of observed inputs and outputs. Convexity of the technology set can be imposed by selecting C from Λ and allows to construct the (artificial) benchmark observation as a convex combination of observed input-output combinations. Selecting NC from Λ then boils down to selecting only one observed input-output combination which is scaled by δ as the benchmark.

It should be noted that (1.14) only satisfies Axiom 1.1 when $(\mathbf{0}_n, \mathbf{0}_m) \in S$ and Axiom 1.2 when, in addition, $(\mathbf{0}_n, \mathbf{Y}_k) \notin S$ with $\mathbf{Y}_k > 0$. Figure 1.4 shows inner approximations for both the input requirements set and the output set. The dots represent observations. The shaded area represents $\boldsymbol{\mathcal{Y}}^{C,VRS}$ and imposes convexity while the dashed boundary represents $\boldsymbol{\mathcal{Y}}^{NC,VRS}$. We have $\boldsymbol{\mathcal{Y}}^{NC,\Gamma} \subseteq \boldsymbol{\mathcal{Y}}^{C,\Gamma}$.

The directional distance function is then simply computed by plugging $\boldsymbol{\mathcal{Y}}^{\Lambda,\Gamma}$ in (1.13):

$$D(\mathbf{X}_0, \mathbf{Y}_0; \mathbf{g}) = \left\{ \max_{\beta \in \mathbb{R}, \boldsymbol{\lambda} \in \mathbb{R}_{+}^{K}} \beta \text{ s.t. } (\mathbf{X}_0 - \beta \mathbf{g}_x, \mathbf{Y}_0 + \beta \mathbf{g}_y) \in \boldsymbol{\mathcal{Y}}^{\Lambda,\Gamma} \right\}. \tag{1.15}$$

This is a linear program when imposing convexity (i.e., $\Lambda = C$) and therefore can be easily solved by standard LP solvers. To be concrete, the linear program one solves to compute (1.13) on $\boldsymbol{\mathcal{Y}}^{C,\Gamma}$ is

$$D(\mathbf{X}_0, \mathbf{Y}_0; \mathbf{g}) = \left\{ \max_{\beta \in \mathbb{R}, \boldsymbol{\lambda} \in \mathbb{R}_{+}^{K}} \beta \text{ s.t. } \sum_{k=1}^{K} \lambda_k \mathbf{X}_k \leq \mathbf{X}_0 - \beta \mathbf{g}_x, \sum_{k=1}^{K} \lambda_k \mathbf{Y}_k \geq \mathbf{Y}_0 + \beta \mathbf{g}_y, \lambda_k \in \Gamma \right\}, \tag{1.16a}$$

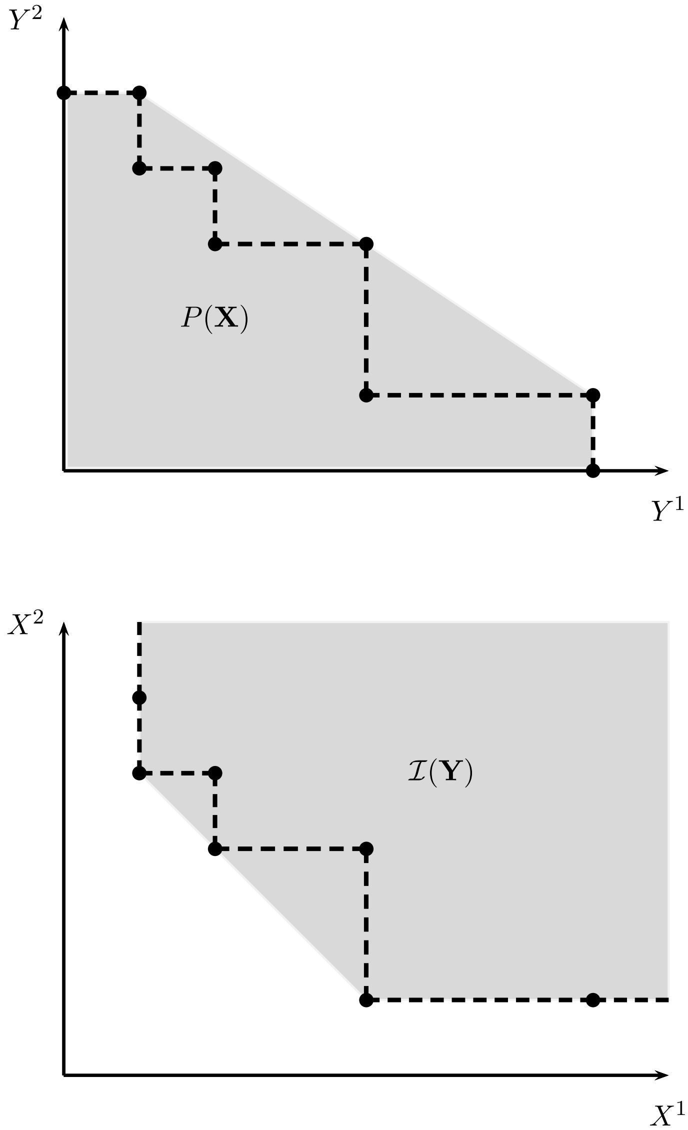


Figure 1.4: Convex (gray area) or free disposal hull (dashed lines) approximation of output and input requirements set.

with

$$\Gamma \in \begin{cases} VRS & = \left\{ \lambda_k \geq 0 | \sum_{k=1}^{K} \lambda_k = 1 \right\}; \\ CRS & = \{ \lambda_k \geq 0 \}; \\ NIRS & = \left\{ \lambda_k \geq 0 | \sum_{k=1}^{K} \lambda_k \leq 1 \right\}; \\ NDRS & = \left\{ \lambda_k \geq 0 | \sum_{k=1}^{K} \lambda_k \geq 1 \right\}. \end{cases} \tag{1.16b}$$

In this linear program δ and its constraints are merged into $\boldsymbol{\lambda}$.

When not imposing convexity (i.e., $\Lambda = NC$), (1.15) is a mixed-integer programming (MIP) problem. Leleu (2006) shows how this can be transformed into an LP formulation. Conveniently, as first pointed out by Tulkens (1993) for Shephard distance functions, an equivalent formulation exists which is solvable using simple enumeration methods. First, one can use Simar and Vanhems (2012)'s transformation of the technology set:

$$\boldsymbol{\mathcal{Y}}^{*NC,\Gamma} = \left\{ (\mathbf{X}^*, \mathbf{Y}^*) \in \mathbb{R}_+^{n+m} | \exists (\mathbf{X}, \mathbf{Y}) \in \boldsymbol{\mathcal{Y}}^{NC,\Gamma} : \mathbf{X}^* = \exp(\mathbf{X}./\mathbf{g}_x), \mathbf{Y}^* = \exp(\mathbf{Y}./\mathbf{g}_y) \right\},$$

before computing the hyperbolic distance function over $\boldsymbol{\mathcal{Y}}^{*NC,\Gamma}$ by enumeration as described in Briec and Kerstens (2006). For the VRS case (i.e., $\Gamma = VRS$), this enumeration formulation is

$$\gamma(\mathbf{X}^*, \mathbf{Y}^*) = \min_{\substack{\forall k=1,\dots,K: \\ \mathbf{X}_k^* \leq \mathbf{X}^*, \mathbf{Y}_k^* \geq \mathbf{Y}^*}} \left(\max \left\{ \max_{i=1,\dots,n} \left(\frac{X_k^{*,i}}{X^{*,i}} \right), \max_{j=1,\dots,m} \left(\frac{Y^{*,j}}{Y_k^{*,j}} \right) \right\} \right).$$

Finally, $\log \gamma^{-1}(\mathbf{X}^*, \mathbf{Y}^*)$ then yields the directional distance function computed over technology $\boldsymbol{\mathcal{Y}}^{NC,VRS}$. Solving $\gamma(\mathbf{X}^*, \mathbf{Y}^*)$ thus involves one loop over all observations $(\mathbf{X}_k^*, \mathbf{Y}_k^*)$ and checking whether they are dominating $(\mathbf{X}^*, \mathbf{Y}^*)$. If they are dominating then one computes the expression within brackets. These enumeration formulations are easy to implement and considerable faster than an LP.

The above transformation assumes $\mathbf{g} > 0$, but when some $(\mathbf{g}_x^U, \mathbf{g}_y^V) = (\mathbf{0}_{|U|}, \mathbf{0}_{|V|})$ for $U \subseteq \{1, \dots, n\}, V \subseteq \{1, \dots, m\}$ then $\mathbf{X}^{*,U} = \mathbf{X}^U$ and $\mathbf{Y}^{*,V} = \mathbf{Y}^V$ (Daraio and Simar, 2014, Section 3.2). Unfortunately, if $\mathbf{g} < 0$ then the above transformation cannot be applied.[3]

Let us now turn to measuring economic efficiency under some behavioral optimizing assumption while imposing minimal assumptions on the technology.

1.3 Economic efficiency

The previous section showed how one can reconstruct the technology set from a discrete number of observations by imposing a number of plausible assumptions regarding the

[3] Cherchye et al. (2001) provide the only known enumeration algorithm which handles $\mathbf{g} < 0$ for $\boldsymbol{\mathcal{Y}}^{NC,VRS}$ only.

technology. Thus, the goodness-of-fit of the approximation hinges on the validity of these assumptions and the number of observations.

If price information is available to the empirical analyst then a different approach is possible that requires far less assumptions with regard to the technology at the cost of assuming some economic optimizing behavior (e.g., profit maximization, revenue maximization or cost minimization).[4] In this section we turn to the question whether there are any testable nonparametric implications on a discrete dataset $S = \{(\mathbf{W}_k, \mathbf{Y}_k, \mathbf{P}_k, \mathbf{X}_k)\}_{k=1,\dots,K}$ associated with economic (e.g., profit maximizing or cost minimizing) optimizing behavior.

1.3.1 Profit maximization

We start with the hypothesis of profit maximization and further assume that a firm cannot influence the market price (i.e., "atomism"); that it is a price taker and that it operates in a market with perfect competition. Let us retake the formulation of (1.1) and start with defining what profit maximization entails for a discrete dataset.

Definition 1.1 (p-rationality)**.** *A technology set $\boldsymbol{\mathcal{Y}}$ p-rationalizes the observed behavior $\{(\mathbf{W}_k, \mathbf{Y}_k, \mathbf{P}_k, \mathbf{X}_k)\}_{k=1,\dots,K}$ if $\mathbf{W}_k \cdot \mathbf{Y}'_k - \mathbf{P}_k \cdot \mathbf{X}'_k \geq \mathbf{W}_k \cdot \mathbf{Y}' - \mathbf{P}_k \cdot \mathbf{X}'$ for all $(\mathbf{X}, \mathbf{Y}) \in \boldsymbol{\mathcal{Y}}$ for $k = 1, \dots, K$.*

P-rationality simply requires that the observed choice $(\mathbf{X}_k, \mathbf{Y}_k)$ in situation k yields maximum profit over all feasible choices $(\mathbf{X}, \mathbf{Y})$ within the technology set $\boldsymbol{\mathcal{Y}}$. The next step is finding testable conditions implied by p-rationality.

Building on work by Afriat (1972); Hanoch and Rothschild (1972); Diewert and Parkan (1983), Varian (1984) showed the following equivalence result:[5]

Theorem 1.1 (Weak Axiom of Profit Maximization (WAPM))**.** *The following conditions are equivalent:*

(1) There exists a technology set that p-rationalizes the data.

(2) $\mathbf{W}_i \cdot \mathbf{Y}'_i - \mathbf{P}_i \cdot \mathbf{X}'_i \geq \mathbf{W}_i \cdot \mathbf{Y}'_j - \mathbf{P}_i \cdot \mathbf{X}'_j$ for all $i, j = 1, \dots, K$.

(3) There exists a closed, convex, free disposal production set that p-rationalizes the data.

Equivalence here means that testing any of the above statements should always give the same result as testing any other of these statements. Thus, if the test of statement (2)

[4]This economic optimization problem is often multi-dimensional which further complicates performance assessment. Public sector managers, in particular, face multiple objectives set by the state to maximize social welfare. Pestieau and Tulkens (1993) argue: *"Producing too little or employing too many factors as compared to what is technically feasible cannot be justified in terms of any of the other objectives listed above."* Hence, they argue, technical efficiency is the only objective which does not impede the achievement of the other social objectives. But, see Peters (1985) for a different opinion.

[5]Varian (1984) used the term "negative monotonic production set" which he admits is "essentially a free disposal hypothesis" of inputs and outputs.

on the dataset is positive, then this immediately implies that there exists an underlying, unknown, technology consistent with statement (3). The elegance of this result is that, apart from Axiom 1.6, statement (2) does not make any assumptions with regards to the technology, but that a positive outcome of this test implies that we can reconstruct this underlying technology by assuming Axioms 1.3-1.5 (cfr. statement (3)). Although convexity is not assumed to check statement (2), the equivalent statement (3) states that a convex production set exists which is associated with a p-rational dataset. Thus, imposing convexity on the reconstructed underlying technology does not alter the result of p-rationality.

The above can be operationalized as follows:

$$\Pi(\mathbf{W}_k, \mathbf{P}_k) = \left\{ \max_{(\mathbf{X},\mathbf{Y}) \in S} \mathbf{W}_k \cdot \mathbf{Y}' - \mathbf{P}_k \cdot \mathbf{X}' \right\} - \left[\mathbf{W}_k \cdot \mathbf{Y}_k' - \mathbf{P}_k \cdot \mathbf{X}_k' \right] \geq 0.$$

Profit efficiency is then marked by $\Pi(\mathbf{W}_k, \mathbf{P}_k) = 0$ while profit inefficiency $\Pi(\mathbf{W}_k, \mathbf{P}_k) > 0$ measures the additional profit that can be gained on top of the current realized profit. $\Pi(\mathbf{W}_k, \mathbf{P}_k)$ is straightforwardly solved by simple enumeration over S.

Varian (1990) advocates the use of measures of deviations from optimizing behavior in an economic sense, because "… *exactly* optimizing behavior isn't a very plausible hypothesis to begin with" (Varian, 1990, p.129). If $\Pi(\mathbf{W}_k, \mathbf{P}_k)$ is small then a firm is more-or-less optimizing. This contrasts with the usual hypothesis testing where the optimization hypothesis is rejected if the estimated parameters deviate in a statistical sense from the values implied by optimization. One should be careful not to confound significance in a statistical sense with significance in an economic sense: one does not necessarily imply the other. We refer the interested reader to McCloskey and Ziliak (1996) for more discussion on this issue.

1.3.2 Cost minimization

One can derive testable implications in a similar way for cost minimization. In contrast to the profit maximization case, we do not need the assumption of perfect competition, atomism or the price taking assumption for output prices. However, we maintain the assumption that firms are price takers in input prices. Thus, cost minimization is a far less restrictive assumption. Furthermore, we do not need output prices to test for cost minimization. A family of input requirements sets "c-rationalizes" the data $S = \{(\mathbf{P}_k, \mathbf{X}_k, \mathbf{Y}_k)\}_{k=1,\dots,K}$ if $\mathbf{X}_k$ solves

$$\min_{\mathbf{X} \in \mathcal{I}(\mathbf{Y}_k)} \mathbf{P}_k \cdot \mathbf{X}'$$

for all $k = 1, \dots, K$. Note that this does not require output prices. We have the following definition:

Definition 1.2 (c-rationality). *An input requirements set $\mathcal{I}(\mathbf{Y})$ c-rationalizes the observed behavior $\{(\mathbf{P}_k, \mathbf{X}_k, \mathbf{Y}_k)\}_{k=1,\dots,K}$ if $\mathbf{P}_k \cdot \mathbf{X}_k' \leq \mathbf{P}_k \cdot \mathbf{X}'$ for all $\mathbf{X} \in \mathcal{I}(\mathbf{Y}_k)$ for $k = 1, \dots, K$.*

Since $\mathcal{I}(\mathbf{Y})$ contains all input combinations that can produce at least $\mathbf{Y}$, we require them to be nested:

$$\text{if } \mathbf{X} \in \mathcal{I}(\mathbf{Y}) \text{ and } \mathbf{Y} \geq \mathbf{Y}', \text{ then } \mathbf{X} \in \mathcal{I}(\mathbf{Y}').$$

Equivalently, this means that $\mathcal{I}(\mathbf{Y}) \subseteq \mathcal{I}(\mathbf{Y}')$ for $\mathbf{Y} \geq \mathbf{Y}'$ and essentially it is a free disposal of outputs assumption. If one does not assume nested input requirement sets then one can always trivially c-rationalize the data by $\mathcal{I}(\mathbf{Y}_k) = \{\mathbf{X}_k\}$ for $k = 1, \ldots, K$ and $\mathcal{I}(\mathbf{Y}) = \emptyset$ otherwise.

The testable implication, dubbed Weak Axiom of Cost Minimization, is again due to Varian (1984):[6]

Theorem 1.2 (Weak Axiom of Cost Minimization (WACM))**.** *The following conditions are equivalent:*

(1) There exists a family of nested input requirement sets $\{\mathcal{I}(\mathbf{Y})\}$ that c-rationalizes the data.

(2) If $\mathbf{Y}_j \geq \mathbf{Y}_i$, then $\mathbf{P}_i \cdot \mathbf{X}_j' \geq \mathbf{P}_i \cdot \mathbf{X}_i'$ for all $i, j = 1, \ldots, K$.

(3) There exists a family of nontrivial, closed, convex, free disposal input requirement set that c-rationalizes the data.

The same discussion as for profit maximization applies here. One should be careful to note that statement (3) applies to the *input requirement set* only and not to the production possibilities set as a whole. In particular, a positive outcome of statement (2) implies that there exists a closed, convex, free disposal input requirement set (i.e., statement (3)). However, this does not imply that convexifying the entire production possibilities set has no effect on the outcome of the cost minimization test. Thus, convexity is only a harmless assumption for cost minimization when applied to the input requirement set. We can operationalize WACM by defining:

$$C(\mathbf{P}_k) = \frac{\{\min_{\mathbf{X} \in D_k} \mathbf{P}_k \cdot \mathbf{X}'\}}{\mathbf{P}_k \cdot \mathbf{X}_k'} \leq 1,$$

with $D_k = \{(\mathbf{X}, \mathbf{Y}) \in S | \mathbf{Y} \geq \mathbf{Y}_k\}$. Under perfect cost minimization, this relative cost efficiency measure $C(\mathbf{P}_k) = 1$. Otherwise, $1 - C(\mathbf{P}_k)$ represents the fraction of observed costs $\mathbf{P}_k \cdot \mathbf{X}_k'$ that can be saved.

1.4 Duality: two sides of the same coin

In the introduction we briefly mentioned the link between technical and economic efficiency. This section elaborates on how the previous two sections are linked by means of duality. Recall from the beginning of this chapter that the ratio of prices equals minus

[6]Free disposal here is with respect to the inputs only. Varian (1984) talks of input requirement sets that are “positive monotonic”.

the marginal rate of transformation for a profit-maximizing choice (cfr. (1.3)). Thus, when observing profit-maximizing observations we can reconstruct the boundary of a technology set, consistent with profit-maximization, using the ratio of prices at every one of these observations as illustrated in Figure 1.1 in input space. Conversely, knowledge of the technology set – and hence the marginal rate of transformation – allows us to reconstruct a profit function consistent with the technology set. This section formalizes this intuition.

Chambers et al. (1998, Lemma 2.1) show that the directional distance function (1.13) completely characterizes the technology for $\mathbf{g} > 0$: i.e., $D(\mathbf{X}, \mathbf{Y}; \mathbf{g}) \geq 0 \iff (\mathbf{X}, \mathbf{Y}) \in \boldsymbol{\mathcal{Y}}$. Thus, we can rewrite the earlier profit maximization problem (1.1) in terms of the directional distance function:

$$\max_{\mathbf{X},\mathbf{Y}} \mathbf{W} \cdot \mathbf{Y}' - \mathbf{P} \cdot \mathbf{X}' \tag{1.17a}$$

$$\text{s.t. } D(\mathbf{X}, \mathbf{Y}; \mathbf{g}) \geq 0. \tag{1.17b}$$

If $(\mathbf{X}, \mathbf{Y}) \in \boldsymbol{\mathcal{Y}}$, then its technically efficient projection $(\tilde{\mathbf{X}}, \tilde{\mathbf{Y}}) = (\mathbf{X} - D(\mathbf{X}, \mathbf{Y}; \mathbf{g})\mathbf{g}_x, \mathbf{Y} + D(\mathbf{X}, \mathbf{Y}; \mathbf{g})\mathbf{g}_y) \in \boldsymbol{\mathcal{Y}}$, because

$$\begin{aligned}
(\tilde{\mathbf{X}}, \tilde{\mathbf{Y}}) \in \boldsymbol{\mathcal{Y}} &\iff D(\tilde{\mathbf{X}}, \tilde{\mathbf{Y}}, \mathbf{g}) \geq 0 \\
&\iff D(\mathbf{X} - D(\mathbf{X}, \mathbf{Y}; \mathbf{g})\mathbf{g}_x, \mathbf{Y} + D(\mathbf{X}, \mathbf{Y}; \mathbf{g})\mathbf{g}_y, \mathbf{g}) \\
&\quad = D(\mathbf{X}, \mathbf{Y}, \mathbf{g}) - D(\mathbf{X}, \mathbf{Y}, \mathbf{g}) = 0 \geq 0.
\end{aligned}$$

Now, in general, for $(\tilde{\mathbf{X}}, \tilde{\mathbf{Y}}) \in \boldsymbol{\mathcal{Y}}$ we have that

$$\begin{aligned}
\Pi(\mathbf{W}, \mathbf{P}) = &\left\{ \max_{\mathbf{X},\mathbf{Y}} \mathbf{W} \cdot \mathbf{Y}' - \mathbf{P} \cdot \mathbf{X}' \right\} - \left[\mathbf{W} \cdot \tilde{\mathbf{Y}}' - \mathbf{P} \cdot \tilde{\mathbf{X}}' \right] \geq 0 \\
\iff &\left\{ \max_{\mathbf{X},\mathbf{Y}} \mathbf{W} \cdot \mathbf{Y}' - \mathbf{P} \cdot \mathbf{X}' \right\} - \left[\mathbf{W} \cdot \mathbf{Y}' - \mathbf{P} \cdot \mathbf{X}' \right] \geq D(\mathbf{X}, \mathbf{Y}; \mathbf{g}) \left(\mathbf{W}\mathbf{g}_y' + \mathbf{P}\mathbf{g}_x' \right) \\
\iff &\frac{\Pi(\mathbf{W}, \mathbf{P})}{\mathbf{W}\mathbf{g}_y' + \mathbf{P}\mathbf{g}_x'} \geq D(\mathbf{X}, \mathbf{Y}; \mathbf{g}).
\end{aligned}$$

If $\boldsymbol{\mathcal{Y}}$ is convex, then by the supporting hyperplane theorem we can exactly partition the space $\mathbb{R}^{n+m}$ in $\boldsymbol{\mathcal{Y}}$ and $\mathbb{R}^{n+m} \setminus \boldsymbol{\mathcal{Y}}$ using a hyperplane with parameters $(\mathbf{W}, \mathbf{P}) \geq 0$ such that:

$$D(\mathbf{X}, \mathbf{Y}; \mathbf{g}) = \inf_{(\mathbf{W},\mathbf{P}) \geq 0} \left\{ \frac{\Pi(\mathbf{W}, \mathbf{P})}{\mathbf{W}\mathbf{g}_y' + \mathbf{P}\mathbf{g}_x'} \right\}. \tag{1.18}$$

This shows that the technology set can be approximated by means of a rescaled version of the profit function. The following LP program operationalizes this result:

$$\min_{t_k, \mathbf{W}_k, \mathbf{P}_k \geq 0} t_k - [\mathbf{W}_k \cdot \mathbf{Y}_k' - \mathbf{P}_k \cdot \mathbf{X}_k'] \tag{1.19a}$$

$$\text{s.t. } t_k \geq \mathbf{W}_k \cdot \mathbf{Y}_s' - \mathbf{P}_k \cdot \mathbf{X}_s' \qquad \forall s \in S \tag{1.19b}$$

$$\mathbf{W}_k \mathbf{g}_y' + \mathbf{P}_k \mathbf{g}_x' = 1. \tag{1.19c}$$

An important remark is that if the true underlying technology $\boldsymbol{\mathcal{Y}}$ is non-convex then its convex hull is recovered by the profit function (Kuosmanen, 2003). Thus, using the profit function yields only a very crude – convexified – approximation of a non-convex technology. Conversely, the profit function can be approximated by means of the directional distance function:

$$\pi(\mathbf{W}, \mathbf{P}) = \sup_{(\mathbf{X}, \mathbf{Y}) \geq 0} \left\{ [\mathbf{W} \cdot \mathbf{Y}' - \mathbf{P} \cdot \mathbf{X}'] + D(\mathbf{X}, \mathbf{Y}; \mathbf{g}) \left(\mathbf{W} \mathbf{g}_y' + \mathbf{P} \mathbf{g}_x' \right) \right\}. \tag{1.20}$$

Graphically, this boils down to finding the point $(\tilde{\mathbf{X}}, \tilde{\mathbf{Y}})$ lying on a line parallel to $(-\mathbf{g}_x, \mathbf{g}_y)$. This point lies on the boundary of the technology set (i.e., $D(\tilde{\mathbf{X}}, \tilde{\mathbf{Y}}, \mathbf{g}) = 0$) and can also be seen as the projected point of a technically inefficient observation $(\mathbf{X}, \mathbf{Y})$ projected onto the boundary of the technology set in the direction $(\mathbf{g}_x, \mathbf{g}_y)$.

1.5 Intermezzo: the "Tinbergen rule" in production theory

One well-known early result in macroeconomics is the "Tinbergen rule". The Tinbergen rule is named after Dutch economist and Nobel laureate Jan Tinbergen. Loosely stated, this rule says that you need at least as many instruments as there are objectives. In the present production context – in our interpretation –, the inputs can be regarded as instruments while the outputs can be regarded as objectives. Mechanically this then means that the more inputs the firm has at its disposal, the more possibilities it has to produce outputs efficiently (in a technical efficiency or economic efficiency sense): e.g., it can choose from an entire spectrum of input-output combinations to produce efficiently. Economically, this manifests itself in the firm choosing from a variety of strategies to be, e.g., a profit maximizer.

When using dominance-based tests in empirical applications, one often finds many efficient observations when the number of observations is small relative to the number of input and output dimensions. The link to the "curse of dimensionality" is relatively straightforward: when increasing the dimensions of the production space (i.e., the inputs and/or outputs) while keeping the number of observations constant, the volume of this production space increases exponentially. Thus, the probability for a given observation to be close to another observation decreases sharply. Hence, the probability for a given observation of having a dominating peer is even smaller. As these nonparametric tests discussed in this chapter are based on a comparison of the evaluated firm with its (input and/or output) dominating peers, we necessarily find a large number of efficient firms as they have no dominating peers to compare with. Convex DEA models (i.e., with reconstructed production set $\boldsymbol{\mathcal{Y}}^{C,\Gamma}$) are the exception as they combine peer units, which are not necessarily dominating, into a hypothetical dominating peer.

1.6 Productivity

We are not only interested in the performance of firms, but also how this performance changes over time. Using only discrete data $(\mathbf{X}_{k,t}, \mathbf{Y}_{k,t})_{t=1}^{T}$ for firms $k = 1, \ldots, K$ we would like to quantify the change in performance over time or productivity.

1.6.1 Malmquist and Hicks-Moorsteen productivity indexes

Intuitively, productivity change between period t and $t + 1$ in the one input-one output case can be expressed in ratio form as:

$$\frac{Y_{0,t+1}/X_{0,t+1}}{Y_{0,t}/X_{0,t}}. \tag{1.21}$$

An advantage of this ratio-based formulation is that it is invariant to a change in the unit of measurement: e.g., converting the monetary outputs and inputs from US dollars to Euros does not affect the results. A disadvantage is that (1.21) can be undefined (e.g., when $X_{0,t+1} = 0$ or $Y_{0,t} = 0$) or zero (e.g., when $Y_{0,t+1} = 0$). Furthermore, this formulation satisfies some common sense properties. For example: if $(X_{0,t+1}, Y_{0,t+1}) = (2X_{0,t}, 2Y_{0,t})$ then intuitively there is no change in productivity which is indicated by a value of 1. However, if $(X_{0,t+1}, Y_{0,t+1}) = (X_{0,t}, 2Y_{0,t})$ then clearly the firm has become more productive because it is able to produce twice the amount of output holding input constant. Indeed, the above formula yields 2 indicating improvement of productivity by a factor 2. Conversely, if $(X_{0,t+1}, Y_{0,t+1}) = (2X_{0,t}, 1.5Y_{0,t})$ then productivity equals 3/4 and has declined because by doubling the level of input the firm has less than doubled its output.

In a multi input – multi output setting one would like to retain the intuitive interpretation of (1.21). Hence, we need some form of aggregation of outputs and inputs. The productivity literature derives index numbers that quantify productivity and propose decompositions that tell us something about the underlying factors that drive changes in productivity.

Distance function approach

The seminal work of Caves et al. (1982) introduces the Malmquist productivity index, named after Sten Malmquist (1953). The output based Malmquist productivity index is defined as:

$$M_{CCD} = \sqrt{M_t \cdot M_{t+1}}, \tag{1.22a}$$

with the period $s = \{t, t+1\}$ output based Malmquist productivity index

$$M_s = \frac{\delta_s(\mathbf{X}_{0,t+1}, \mathbf{Y}_{0,t+1})}{\delta_s(\mathbf{X}_{0,t}, \mathbf{Y}_{0,t})} \tag{1.22b}$$

where $\delta_s(\cdot)$ is the Shephard output distance function (1.12) defined over the technology of period s. The Shephard distance function characterizes the technology and serves as an aggregator function for the inputs and outputs. The Malmquist output productivity index measures changes in output distance to the efficient frontier from period t to period $t+1$. $M_{CCD} > 1$ indicates productivity improvement while $M_{CCD} < 1$ reflects deterioration in productivity. No change in productivity occurs for $M_{CCD} = 1$. It is common to take a geometric mean of M_t and M_{t+1} to avoid an arbitrary choice of base period. Analogously, the Malmquist input quantity index is defined.

The link between the intuitive notion of productivity (1.21) and the Malmquist index (1.22) is not immediately clear. The former has an average product interpretation while the Malmquist index only has this interpretation when the distance functions are computed with respect to a CRS technology. Grifell-Tatjé and Lovell (1995) show that the Malmquist index is biased under non-constant returns-to-scale: i.e., it systematically under- or overestimates productivity change. The source of this bias is exactly the effect of returns-to-scale.

Different sources affect productivity change. The most important ones are efficiency change, technical change and scale efficiency change. This can answer important questions for individual firms such as: Is the improvement in productivity because *we* became more efficient and/or because *everyone* became more efficient? The former is known as efficiency change while the latter is known as technical change. Much of the literature focused on (correctly) decomposing M_{CCD} into these various sources when not imposing CRS on the underlying technology. We refer to Grosskopf (2003) for a historic overview and additional references.

The Malmquist index is not a measure of total factor productivity (TFP), because *"... [it] measures the displacement of the production frontier at a specific point and neglects scale economies. For this reason the CCD Malmquist can be interpreted as a technology index, i.e., a measure of local technical progress (or regress)"* (Peyrache, 2014, p.435). It neglects scale economies, because it only measures changes in either output or input distances to the efficient frontier. Thus it only tells half of the story, because it ignores any changes in input or output distances to the efficient frontier.

Bjurek (1996) defines the Hicks-Moorsteen TFP index as an output quantity index divided by an input quantity index:

$$HM = \sqrt{HM_t \cdot HM_{t+1}}, \tag{1.23a}$$

with the period $s = \{t, t+1\}$ Hicks-Moorsteen index

$$HM_s = \frac{MO_s(\mathbf{X}_{0,s}, \mathbf{Y}_{0,t}, \mathbf{Y}_{0,t+1})}{MI_s(\mathbf{X}_{0,t}, \mathbf{X}_{0,t+1}, \mathbf{Y}_{0,s})} = \frac{\delta_s(\mathbf{X}_{0,s}, \mathbf{Y}_{0,t+1})/\delta_s(\mathbf{X}_{0,s}, \mathbf{Y}_{0,t})}{\theta_s(\mathbf{X}_{0,t+1}, \mathbf{Y}_{0,s})/\theta_s(\mathbf{X}_{0,t}, \mathbf{Y}_{0,s})}. \tag{1.23b}$$

Furthermore, contrary to the Malmquist index, the Hicks-Moorsteen index satisfies the determinateness property meaning that it is always well-defined for all observations and time periods (Briec and Kerstens, 2011). A decomposition of the Hicks-Moorsteen is provided by Nemoto and Goto (2005) in a stochastic frontier analysis framework, O'Donnell (2012a) and Diewert and Fox (2014, 2017) for free disposal technologies. Finally, Färe et al. (1996) show that the Malmquist index equals the Hicks-Moorsteen index if and only if the technology is inversely homothetic and exhibits CRS.

Economic approach

When prices are available then one can use these as weights to aggregate the respective outputs and inputs. One such index is the Törnqvist (1936) productivity index:[7]

$$\begin{aligned}\ln T &\equiv \ln TO - \ln TI && (1.24)\\ &= \sum_{v=1}^{m}\left(\frac{r_t^v + r_{t+1}^v}{2}\right)\ln\left(\frac{Y_{0,t+1}^v}{Y_{0,t}^v}\right) - \sum_{u=1}^{n}\left(\frac{c_t^u + c_{t+1}^u}{2}\right)\ln\left(\frac{X_{0,t+1}^u}{X_{0,t}^u}\right),\end{aligned}$$

with period $s = \{t, t+1\}$ revenue shares $r_s^v = \frac{W_{0,s}^v Y_{0,s}^v}{\mathbf{W}_{0,s} \cdot \mathbf{Y}_{0,s}'}$ and cost shares $c_s^u = \frac{P_{0,s}^u X_{0,s}^u}{\mathbf{P}_{0,s} \cdot \mathbf{X}_{0,s}'}$. The Törnqvist productivity index is defined as the ratio of a Törnqvist output index to a Törnqvist input index. The attractive feature of the Törnqvist index is that it is easy to compute from price and quantity data. Thus, it does not require the computation of distance functions as in the Malmquist index (1.22) or the Hicks-Moorsteen index (1.23) by means of linear programming or econometric techniques.

Exact and superlative index numbers

In Section 1.1 we found that under economic optimization the ratio of prices equals minus the marginal rate of transformation. Essentially this means that the boundary of the technology set can be approximated by the ratio of prices under profit maximization. Conversely, the boundary of the technology set can recover the ratio of prices that would prevail under profit maximization (i.e., "shadow prices"). Section 1.4 formally showed how the profit function can approximate the directional distance function which completely characterizes the production technology and how the directional distance function can approximate the profit function.

With these duality results in mind it makes sense to wonder whether a similar link exists between the distance function approach and the economic approach to productivity. In other words: can the Malmquist index (1.22) and/or the Hicks-Moorsteen index (1.23) be approximated by the Törnqvist productivity index (1.24) (and vice versa)? It turns out that – under specific conditions – the answer is positive.

In order to relate the Malmquist index and the Törnqvist index two conditions must be satisfied. The first one is a behavioral condition: one must assume that firms are

[7]The Törnqvist price index is historically attributed to Törnqvist (1936), but according to Balk (2008) it was originally introduced in Törnqvist and Törnqvist (1937).

profit maximizers. With Section 1.1 and duality of Section 1.4 in mind the necessity of this condition is easy to understand: (1.22) implicitly and (1.24) explicitly aggregate all inputs and outputs using a weighted sum. The associated distance function $\delta_s(\cdot)$ uses the shadow prices of the efficient projections, while the Törnqvist uses the observed prices. The only way for both weighted sums to be equal is by having the same weights or if the ratios of shadow prices equal the ratios of market prices.

The second condition is a technological one: the technology must have a particular "flexible" functional form so that is sufficiently smooth and differentiable. Diewert (1976, p.115) calls a functional form "flexible" if *"it can provide a second order approximation to an arbitrary twice differentiable linearly homogeneous function"*. An index number is *"said to be 'superlative' if it is exact (i.e., consistent with) for a 'flexible' aggregator functional form"* (Diewert, 1976, p.115). Combining both conditions, Caves et al. (1982, Theorem 3) formally show that the Malmquist index (1.22) is equal to the Törnqvist productivity index (1.24) if (i) the Shephard output distance functions can be represented by a CRS translog functional form with identical second-order coefficients and (ii) the firm is a profit maximizer. Note that the assumption of profit maximization implies that the firm is also technically efficient: i.e., $\delta_s(\mathbf{X}_{0,s}, \mathbf{Y}_{0,s}) = 1$ for $s = \{t, t+1\}$ in (1.22).[8] Finally, Mizobuchi (2016, Corollary 2) shows that the Hicks-Moorsteen index (1.23) also equals the Törnqvist index under the above mentioned conditions.

1.6.2 Indexes vs. indicators

The previous subsection started out with (1.21) as an intuitive ratio-based measure of productivity. However, we can equivalently express a productivity measure in terms of differences:

$$[Y_{0,t+1} - X_{0,t+1}] - [Y_{0,t} - X_{0,t}]. \tag{1.25}$$

This difference-based measure is not units invariant like the ratio-based measure (1.21), but in contrast to (1.21) it is well defined even when some of the variables equal zero and it is translation invariant: adding a constant $\tau \in \mathbb{R}$ to all outputs and inputs has no effect (cfr. (1.21)). Furthermore, it has some intuitive properties. For example: if $(X_{0,t+1}, Y_{0,t+1}) = (X_{0,t} + 2, Y_{0,t} + 2)$ then intuitively there is no change in productivity which is indicated by a value of 0. However, if $(X_{0,t+1}, Y_{0,t+1}) = (X_{0,t}, Y_{0,t} + 2)$ then clearly the firm has become more productive because it is able to produce two more units of output holding input constant. Indeed, the above formula yields 2 indicating improvement of productivity by 2 units. Conversely, if $(X_{0,t+1}, Y_{0,t+1}) = (X_{0,t} + 2, Y_{0,t} + 1.5)$ then productivity equals -0.5 and has declined because by using two more units of input the firm has added less than two units of its output.

We refrain from presenting productivity indicators in a multi input – multi output setting, because this is the subject of Chapter 2 and Chapter 3 which in turn discuss the Luenberger productivity indicator and the Luenberger-Hicks-Moorsteen TFP indicator.

[8] Balk (1998) generalizes this result by not requiring technical efficiency while Diewert and Fox (2010) further relax the assumption of perfect competition and allow for increasing returns-to-scale.

Analogously to the case of productivity indexes, exact and superlative indicators exist for these productivity indicators. We refer the interested reader to Chambers (2002).

It should be clear that both indexes and indicators have their specific merits in specific situations and that it often depends on data characteristics and/or the user's preference.

1.7 Concluding remarks

We presented a brief overview of necessary tools for technical and economic efficiency measurement. We discussed the underlying theory and nonparametric tools to evaluate efficiency and demonstrated how to operationalize them on discrete data. We showed how duality theory acts as the mold between technical and economic efficiency. Finally, we discussed how productivity measures the performance of firms over time and how different approaches to productivity measurement are linked once again by duality. We end this chapter by discussing some of the limitations of the presented tools.

A first issue pertains to the "no noise, no outlier" assumption (Axiom 1.6). This is a strong assumption which is seldom valid in practice. A number of tools exist that relax this assumption such as partial frontiers (order-m and order-α) discussed in Daraio and Simar (2007) and superefficiency models of Banker and Chang (2006) for outlier detection. The presented methods in this thesis can quite easily be modified along those lines. Finally, we did not touch upon the statistical properties of DEA estimators. It can be shown that DEA estimators are consistent under some mild conditions: i.e., the estimation error (bias) goes to zero as the sample size increases (Banker, 1993). We refer the interested reader to the book of Daraio and Simar (2007) for more details.

A second point of discussion is with regard to how measured inefficiency should be interpreted. The presented measure of profit efficiency relies on the assumptions of perfect competition, atomism and price taking behavior in input and output prices while the measure of cost efficiency maintains only the price taking assumption in input prices. The presented economic efficiency tests then constitute a joint test of all these assumptions. Therefore, the measured economic inefficiency could also be regarded as a violation of any of these assumptions. We here mention some examples of papers that address these issues. Cherchye et al. (2002) relax the assumption of exogenous prices by allowing for prices that depend on quantities through a known inverse demand function. Portela and Thanassoulis (2014) propose a modification using only observed prices and quantities and propose a decomposition of cost efficiency. Lee and Johnson (2015) go a step further and identify the Nash market equilibrium for given inverse demand and supply functions. Finally, Carvajal et al. (2013) develop revealed preference tests of the Cournot model while Carvajal et al. (2014) develop nonparametric tests for models of multi-product oligopoly.

Furthermore, the presented technical efficiency measures rely on the assumption that the empirical approximation to the technology is appropriate for the true technology. This empirical approximation in turn depends on the validity of the modeling assumptions. It is then entirely possible that the measured inefficiency can (partially) be explained by some market or technology constraint which is unobserved by the empirical analyst and hence not accounted for in the modeling. For instance, capacity constraints determine maximum output per firm and can therefore (partially) explain measured output technical inefficiency or profit inefficiency.

The following chapters constitute the main body of this thesis. Chapter 2 and chapter 3 contribute to the productivity literature. Chapter 2 proposes a productivity measure that measures potential gains of input reallocation which can be decomposed in underlying factors. It uses a nonparametric empirical approximation of the technology which models and links the individual farm activities. Chapter 3 presents a decomposition of the Luenberger-Hicks-Moorsteen total factor productivity indicator along with an empirical application on US state-level agricultural data. Chapter 4 contributes to the literature on economic efficiency by proposing a framework for intertemporal cost minimization and corresponding nonparametric tests. The empirical illustration compares and highlights differences between the static cost minimization model and our intertemporal cost minimization model. Chapter 5 determines the direction vectors in (1.13) through a comparative analysis of key DMUs and presents a visualization tool to compare DMUs in terms of their input-output mix similarity and scale.

Chapter 2

To mix or specialise? A coordination productivity indicator for English and Welsh farms

"Most traditional DEA models treat their reference technologies as black boxes. Our network models, developed for the Swedish Institute for Health Economics (IHE), allow us to look into these boxes and to evaluate organizational performance and its component performance."

— Rolf Färe and Shawna Grosskopf[1]

2.1 Introduction

The economic choice for a farmer about whether to engage in specialised or mixed agriculture is based on a comparison of the gains from economies of scale versus the gains from diversifying risk and economies of scope.[2] Crop-specific capital (*e.g.*, harvesters and ploughs) and livestock-specific capital (*e.g.*, milking robots) are expensive and benefit considerably from economies of scale. Furthermore, these fixed costs of capital can be spread more over higher production volumes. Thus, a return on these investments can only be achieved by increasing the scale of operation, which in turn leaves little room for other farming activities (Chavas and Aliber, 1993; Fernandez-Cornejo et al.,

[0]This chapter is based on joint work with Frederic Ang (Swedish University of Agricultural Sciences) and a slightly different version was published as Ang and Kerstens (2016) in Journal of Agricultural Economics. The paper was joint winner of the "Prize Essay Competition" for young researchers of the Agricultural Economics Society (AES) in 2016.

[1]Färe and Grosskopf (2000, p.35)

[2]The literature makes a further distinction between "economies of scope" and "economies of diversification". Following Panzar and Willig (1981), the former focuses on measuring the cost of complete specialisation while the latter focuses on measuring the cost of partial specialisation (Chavas and Kim, 2010). The specialisation scheme to use is context dependent and in addition can be due to data limitations. Complete specialisation has more severe data requirements as it requires the observation of completely specialised farms. Since we do observe completely specialised farms in the data, we stick to the economies of scope terminology. We refer to Chavas and Kim (2010) and references therein for a discussion.

1992). European agriculture is nowadays increasingly characterised by specialised production. In the light of the liberalisation of the Common Agricultural Policy, input and output prices are becoming more volatile, increasing the volatility of the farmer's income. Economic intuition suggests that diversifying into more farming activities and mixed farming generates economies of scope resulting from complementarities between different farming activities, which allows for production at lower cost (Chavas, 2008).[3] Moreover, researchers and policy makers are increasingly concerned about the negative environmental impact of nutrient surplus associated with specialisation (Ryschawy et al., 2012).

The efficiency and productivity of a farm play an essential role for its long-term viability, requiring coordination of crop- and livestock-specific inputs. The overwhelming majority of the studies in the efficiency and productivity literature treat agricultural production as a black-box where the subprocesses are overlooked.[4] This implicitly leaves the question unanswered whether more or less specialisation would be needed for efficiency gains. In addition, this hampers the comparability of farms with different subprocesses. The difficulty of modelling these subprocesses may explain why most empirical studies only focus on specialised farms. We introduce a coordination Luenberger productivity indicator that addresses these problems.

Färe and Whittaker (1995) introduce an efficiency framework that takes into account the production of intermediate inputs on the farm. In their model, crop output can also be used as a feed input in the livestock enterprise. Focussing on a sample of cereal farms, Färe et al. (1997) develop an efficiency framework where land use can be reallocated. Cherchye et al. (2013) develop a general framework that opens the black-box of production by explicitly modelling input allocation in a multi-output setting. They distinguish different subdivisions with their own output. Every output uses its own associated output-specific inputs and common joint inputs that are shared by all outputs. They develop a radial input-oriented framework. Using this framework, Cherchye et al. (2017b) address the question of efficient allocation of common output-specific inputs over subdivisions. They develop a coordination efficiency measure that quantifies the possible efficiency gains from reallocating some inputs over the subdivisions.[5]

[3]The distinction between economies of scale and economies of scope is less clear-cut in case of non-separability between outputs. Loosely defined, separability between two outputs implies that a change in production of one output does not affect production of the other output. Chavas and Kim (2010) introduce a diversification measure and show that this measure can be decomposed into four effects: complementarity among outputs, scale, convexity and fixed costs. Under a pattern of complete specialization, Chavas and Kim (2010, Proposition 2) show that this diversification measure only depends on the complementarity among outputs effect and fixed costs effect. Separability implies that this complementarity effect equals zero so that economies of diversification only depend on fixed costs under a pattern of complete specialization.

[4]Färe and Whittaker (1995), Färe et al. (1997), Jaenicke (2000), Skevas et al. (2012) and Chen (2012) are exceptions.

[5]There exists a stream of literature that handles interfirm reallocation of inputs. This social planner perspective to reallocation starts from a fixed industrywide amount of inputs (e.g., quota) that needs to be reallocated among firms. The consequences for individual firms include break-ups, mergers, interfirm reallocation of resources or shutting down their operations entirely. We refer to Peyrache (2013) for a

Using a nonparametric framework, this paper extends the static, radial framework suggested by Cherchye et al. (2017b) to a dynamic context of coordination Luenberger productivity growth. Cherchye et al. (2017b)'s radial framework only identifies inefficiencies in the input direction. We generalise this approach by developing a directional distance function framework where inputs as well as outputs are choice variables, which is consistent with profit-maximising behaviour. The Luenberger productivity indicator builds on contemporaneous and intertemporal directional distance functions. It measures productivity growth by simultaneously assessing changes in the input and output level over time and can be decomposed into components of technical inefficiency change and technical change. Explicitly taking into account the subprocesses of crop production and livestock production, our framework is able to adequately compare the efficiency and productivity of crop farms, mixed farms and livestock farms. It also indicates whether coordination efficiency gains would be associated with specialisation or diversification towards mixed farming. Moreover, we define a coordination productivity indicator that measures productivity growth due to optimal reallocation of process-specific inputs over time which is decomposable in a coordination technical inefficiency change component and a coordination technical change component. This decomposition allows us to assess how reallocation affects the different sources of productivity growth. The empirical application focusses on panel data from mixed and specialised farms in England and Wales over the period $2007 - 2013$.

The remainder of the paper is structured as follows. The next section describes the theoretical framework for measuring the coordination Luenberger productivity indicator and its components. This is followed by the practical implementation and empirical application. The final section concludes.

2.2 Mixed farm model

In this section we describe our mixed farm model. We distinguish two interdependent subprocesses with their own technology. We then propose a coordination Luenberger productivity indicator and its decomposition that identifies how coordination inefficiency affects the different sources of productivity growth.

2.2.1 Model and technology description

We identify 2 processes: the crop subprocess (C) and the livestock subprocess (L). The network structure is shown in Figure 2.1. Following Färe and Whittaker (1995), both processes are interdependent because the livestock process uses unmarketed residue of crops as feed for its livestock in addition to feed bought on the market. Note that

comprehensive framework which introduces an efficiency indicator of industry configuration. Examples of application include reallocation of fishing quota (Andersen and Bogetoft, 2007), sugar beet contracts (Bogetoft et al., 2007) and nitrogen emissions permits (Nielsen et al., 2014). We take the farmer's perspective in the present study to determine how an individual farmer can become more efficient by reallocating inputs among farming activities.

manure could be modelled as a livestock output, which can serve as an input for future crop production. However, we do not include manure in our model due to a lack of availability of manure data.[6]

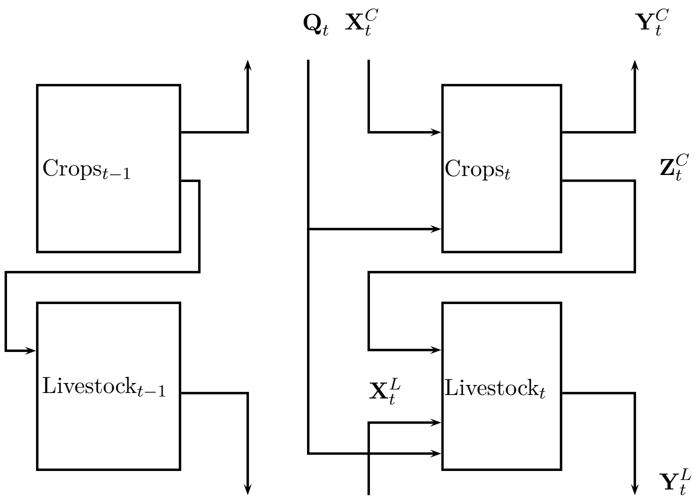


Figure 2.1: Network structure of the model.

The crop subprocess has the following inputs and outputs:

- $\mathbf{X}_t^C \in \mathbb{R}_+^{N_C}$: inputs such as labour, seeds, etc;
- $\mathbf{Y}_t^C \in \mathbb{R}_+^{O_C}$: outputs such as wheat, barley, etc;
- $\mathbf{Z}_t^C \in \mathbb{R}_+^{O_C}$: outputs that are not sold, but used as feed in the same period for the livestock.

The livestock subprocess has the following inputs and outputs:

- $\mathbf{X}_t^L \in \mathbb{R}_+^{N_L}$: inputs such as labour, feed, etc;
- $\mathbf{Z}_t^C \in \mathbb{R}_+^{O_C}$: outputs from the crops used as feed in the same period;
- $\mathbf{Y}_t^L \in \mathbb{R}_+^{O_L}$: outputs such as milk, meat, etc;

Note that $\mathbf{X}_t^L$ and $\mathbf{Z}_t^C$ can have common inputs: the farmer buys additional feed for his livestock on top of the feed he already collected from his crops. Define the index set $H = \{1, \ldots, N_L\} \cap \{1, \ldots, O_C\}$ for these common inputs.

There can be some inputs that are shared by both processes but which are not joint. Land use is such an input: the farmer has to decide how much of his land area to use for crop production and livestock production. Thus, these subprocess-specific inputs

[6] At the time of writing, manure data were only available for the years 2012 and 2013 for a subsample of farms.

have to be allocated among both processes. In line with Cherchye et al. (2017b), this allocation might not be optimal and a better reallocation is possible. Let $\mathbf{X}_t \in \mathbb{R}^C_+$ with $C \subseteq \{1, \ldots, N_C\} \cap \{1, \ldots, N_L\}$ be the process-specific inputs that have to be allocated among both subprocesses such that

$$X_t^{C,m} + X_t^{L,m} = X_t^m \; \forall m \in C. \tag{2.1a}$$

Thus, C is the subset of inputs, common to crop and livestock, that can be reallocated among both subprocesses.[7] In line with Färe et al. (1997), this application regards crop land and livestock land as reallocatable, fixed inputs.

Furthermore, the levels of the other process-specific inputs must also be adjusted for the new reallocation:

$$\sum_{\forall i \in \{1,\ldots,N_C\} \setminus C} p_t^{C,i} X_t^{C,i} + \sum_{\forall j \in \{1,\ldots,N_L\} \setminus C} p_t^{L,j} X_t^{L,j} = PEXP_t, \tag{2.1b}$$

where $\mathbf{p}_t^C \in \mathbb{R}^{N_C}_{++}$ and $\mathbf{p}_t^L \in \mathbb{R}^{N_L}_{++}$ are the prices of crop-specific and livestock-specific inputs and $PEXP_t$ is the total process-specific expenditures.[8] This budget constraint allows the farmer to redistribute the process-specific budget over the crop and livestock activities while staying within his budget. In general, we do not know the individual farmer's credit or budget constraints, but we do observe his total process-specific expenditures. Therefore, process-specific expenditures can be reallocated within the farmer's observed budget.

Finally, the crop and livestock process share a joint input $\mathbf{Q}_t \in \mathbb{R}^M_+$ (*e.g.*, buildings and machinery). Joint inputs are inputs which are shared by the different subprocesses (see Cherchye et al. (2013)). Some of these joint inputs are fixed: let $F \subseteq \{1, \ldots, M\}$ denote the set of fixed joint inputs.

We now define the technology of each subprocess by their production set. The crop subprocess production set is:

$$\boldsymbol{\mathcal{Y}}_t^C = \left\{ (\mathbf{X}_t^C, \mathbf{Q}_t) \text{ produces } (\mathbf{Y}_t^C, \mathbf{Z}_t^C) \right\}. \tag{2.2}$$

Similarly, the livestock subprocess production set is:

$$\boldsymbol{\mathcal{Y}}_t^L = \left\{ (\mathbf{X}_t^L, \mathbf{Z}_t^C, \mathbf{Q}_t) \text{ produces } \mathbf{Y}_t^L \right\}. \tag{2.3}$$

In the remainder of this paper, we assume the following basic axioms for both subprocesses:

Axiom 2.1 (strong disposability of inputs)**.** $(\mathbf{x}, \mathbf{y}) \in \boldsymbol{\mathcal{Y}}$ *and* $\mathbf{x}' \geq \mathbf{x} \Longrightarrow (\mathbf{x}', \mathbf{y}) \in \boldsymbol{\mathcal{Y}}$

[7] This assumes that these common inputs can be freely reallocated between both subprocesses without losing value in a marginal product sense. This means, for example, that one unit of livestock land can be reallocated to crop land and yields equal amounts of crop output as any other unit of crop land.

[8] In the absence of price data, one could equivalently work with expenditures. The quantities are then expenditures and the modified budget constraint would be (2.1b) without prices.

Axiom 2.2 (strong disposability of outputs)**.** $(\mathbf{x}, \mathbf{y}) \in \boldsymbol{\mathcal{Y}}$ *and* $\mathbf{y}' \leq \mathbf{y} \Longrightarrow (\mathbf{x}, \mathbf{y}') \in \boldsymbol{\mathcal{Y}}$

Axiom 2.3 (convexity)**.** *Technology set* $\boldsymbol{\mathcal{Y}}$ *is convex.*

The overall network production set is:

$$\boldsymbol{\mathcal{Y}}_t = \left\{ (\mathbf{X}_t^C, \mathbf{Q}_t, \mathbf{Y}_t^C, \mathbf{Z}_t^C) \in \boldsymbol{\mathcal{Y}}_t^C \text{ and } (\mathbf{X}_t^L, \mathbf{Z}_t^C, \mathbf{Q}_t, \mathbf{Y}_t^L) \in \boldsymbol{\mathcal{Y}}_t^L \right\}, \tag{2.4}$$

and satisfies the above axioms by construction.

2.2.2 The Luenberger productivity indicator and its decomposition

We use Luenberger's directional distance function to measure technical inefficiency by simultaneously contracting inputs and expanding outputs. This is consistent with profit maximisation. Shephard (1970)'s input and output distance functions are special cases of the directional distance function (Chambers et al., 1996a) and are consistent with cost minimisation and revenue maximisation, respectively. Define, for notational convenience, $\mathbf{X}_t = (\mathbf{X}_t^C, \mathbf{X}_t^L, \mathbf{Z}_t^C, \mathbf{Q}_t)$ as the input vector and $\mathbf{Y}_t = (\mathbf{Y}_t^C, \mathbf{Y}_t^L, \mathbf{Z}_t^C)$ as the output vector. The directional distance function proposed by Chambers et al. (1996b) is:

$$D_t(\mathbf{X}_t, \mathbf{Y}_t; \mathbf{g}_t) = \sup \left\{ \beta \in \mathbb{R} : (\mathbf{X}_t - \beta \mathbf{g}_{x,t}, \mathbf{Y}_t + \beta \mathbf{g}_{y,t}) \in \boldsymbol{\mathcal{Y}}_t \right\}, \tag{2.5}$$

if $(\mathbf{X}_t - \beta \mathbf{g}_{x,t}, \mathbf{Y}_t + \beta \mathbf{g}_{y,t}) \in \boldsymbol{\mathcal{Y}}_t$ for some β and $D_t(\mathbf{X}_t, \mathbf{Y}_t; \mathbf{g}_t) = -\infty$ otherwise. Here, $\mathbf{g}_t = (\mathbf{g}_{x,t}, \mathbf{g}_{y,t})$ represents the direction vector. The directional distance function is a special case of Luenberger (1992)'s shortage function.

We denote the time-related directional distance function for $(a, b) \in \{t, t+1\} \times \{t, t+1\}$:

$$D_b(\mathbf{X}_a, \mathbf{Y}_a; \mathbf{g}_a) = \sup \left\{ \beta \in \mathbb{R} : (\mathbf{X}_a - \beta \mathbf{g}_{x,a}, \mathbf{Y}_a + \beta \mathbf{g}_{y,a}) \in \boldsymbol{\mathcal{Y}}_b \right\}.$$

Furthermore, we distinguish between $D_t(\mathbf{X}_t, \mathbf{Y}_t; \mathbf{g}_t | R)$ and $D_t(\mathbf{X}_t, \mathbf{Y}_t; \mathbf{g}_t | NR)$: the former allows for reallocation of the process-specific inputs over the subprocesses (i.e., (2.1)), while the latter keeps these fixed.

In an analogous way, we define "reallocative" and "non-reallocative" Luenberger productivity indicators $L_{t,t+1}(\cdot | R)$ and $L_{t,t+1}(\cdot | NR)$ respectively. The Luenberger productivity indicator proposed by Chambers (2002) is defined as:

$$\begin{aligned} &L_{t,t+1}(\mathbf{X}_t, \mathbf{Y}_t, \mathbf{X}_{t+1}, \mathbf{Y}_{t+1}; \mathbf{g}_t, \mathbf{g}_{t+1} | c) \\ &\quad = \frac{1}{2} \left[(D_t(\mathbf{X}_t, \mathbf{Y}_t; \mathbf{g}_t | c) - D_t(\mathbf{X}_{t+1}, \mathbf{Y}_{t+1}; \mathbf{g}_{t+1} | c)) \right. \\ &\quad \left. + \, (D_{t+1}(\mathbf{X}_t, \mathbf{Y}_t; \mathbf{g}_t | c) - D_{t+1}(\mathbf{X}_{t+1}, \mathbf{Y}_{t+1}; \mathbf{g}_{t+1} | c)) \right], \end{aligned} \tag{2.6}$$

for $c \in \{R, NR\}$. It can be additively decomposed into a technical inefficiency change component and a technical change component:

$$\begin{aligned} L_{t,t+1}(\cdot|c) &= (D_t(\mathbf{X}_t, \mathbf{Y}_t; \mathbf{g}_t|c) - D_{t+1}(\mathbf{X}_{t+1}, \mathbf{Y}_{t+1}; \mathbf{g}_{t+1}|c)) \\ &+ \frac{1}{2}\left[(D_{t+1}(\mathbf{X}_{t+1}, \mathbf{Y}_{t+1}; \mathbf{g}_{t+1}|c) - D_t(\mathbf{X}_{t+1}, \mathbf{Y}_{t+1}; \mathbf{g}_{t+1}|c))\right. \\ &+ \left.(D_{t+1}(\mathbf{X}_t, \mathbf{Y}_t; \mathbf{g}_t|c) - D_t(\mathbf{X}_t, \mathbf{Y}_t; \mathbf{g}_t|c))\right] \\ &\equiv \Delta TEI(c) + \Delta T(c), \end{aligned} \tag{2.7}$$

where the first difference is the technical inefficiency change component $\Delta TEI(c)$ and the arithmetic average of the two last differences captures technical change $\Delta T(c)$ (Chambers et al., 1996b). The technical inefficiency change component quantifies the change in relative position of a given observation to the (shifted) production frontier. The technical change component measures the change in the production frontier itself and is therefore a measure of technical progress or regress. The relevant distance functions are depicted in Figure 2.2.

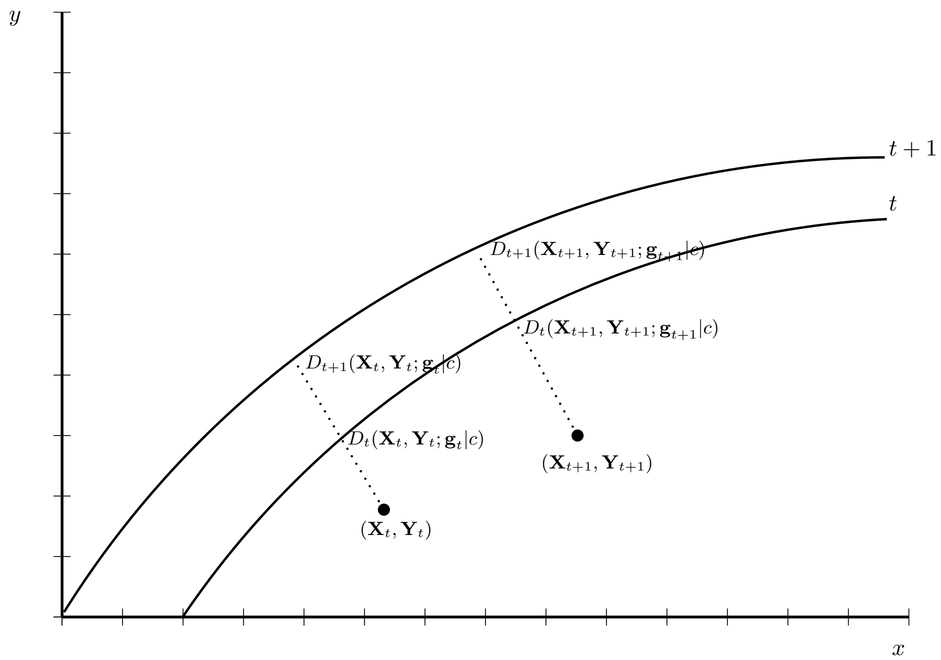


Figure 2.2: Distance functions of the Luenberger productivity indicator.

2.2.3 Coordination inefficiency

Cherchye et al. (2017b) consider a model of a Decision Making Unit with several subdivisions (they give an example of a university with subdivisions in research and teaching). They are interested in measuring whether efficiency gains are possible from reallocating common inputs over the different subdivisions. To this end, they distinguish radial measures of decentralised and centralised efficiency. Decentralised efficiency is the radial measure of efficiency when the current allocation is preserved over the subdivisions. In contrast, centralised efficiency is the radial measure of efficiency when the allocation is free to change over the subdivisions. Then, they define coordination efficiency as the ratio of centralised over decentralised efficiency. We make use of a directional distance function framework, which is more flexible in that it allows for varying input and output levels.

An equivalent difference-based coordination inefficiency measure is:

$$CI = D_t(\mathbf{X}_t, \mathbf{Y}_t; \mathbf{g}_t|R) - D_t(\mathbf{X}_t, \mathbf{Y}_t; \mathbf{g}_t|NR). \tag{2.8}$$

where R, NR denote reallocation and no reallocation from present farm organisation, respectively. Here, one can see that $D_t(\mathbf{X}_t, \mathbf{Y}_t; \mathbf{g}_t|R) \geq D_t(\mathbf{X}_t, \mathbf{Y}_t; \mathbf{g}_t|NR)$, because the *status quo* allocation represented by $D_t(\cdot|NR)$ is always attainable when reallocation is allowed. Positive values for CI indicate that inefficiencies may arise from suboptimal allocation of inputs.

2.2.4 Coordination productivity indicator

We measure how much productivity growth is affected by reallocation of inputs over time by comparing the reallocative Luenberger productivity indicator with the non-reallocative Luenberger productivity indicator. $L_{t,t+1}(R) > (<) L_{t,t+1}(NR)$ indicates that a farmer becomes better (worse) at reallocating over time which leads to improved (worsened) productivity growth. We define a "coordination Luenberger productivity indicator" $CL_{t,t+1}$ as the difference between $L_{t,t+1}(R)$ and $L_{t,t+1}(NR)$:

$$\begin{aligned} CL_{t,t+1} &\equiv L_{t,t+1}(R) - L_{t,t+1}(NR) \\ &= [\Delta TEI(R) - \Delta TEI(NR)] + [\Delta T(R) - \Delta T(NR)] \\ &\equiv \Delta CI + \Delta CT, \end{aligned} \tag{2.9}$$

where ΔCI is coordination inefficiency change and ΔCT is coordination technical change. ΔCI measures the change in coordination inefficiency that can be ascribed to reallocation of process-specific inputs over time. ΔCT measures changes in the production frontier due to reallocation of process-specific inputs over time.

2.3 Practical implementation

The empirical analyst can compute efficiency and productivity measures using either a parametric or nonparametric approach. The parametric approach takes into account stochastic factors and does not treat all deviations from the frontier as inefficiency. However, it requires a specification of a functional form and technical changes cannot be determined at the firm level. We opt for the nonparametric approach which does not require such a specification and allows for determination of firm-specific technical changes (Oude Lansink et al., 2015).

We assume that we have data

$$S = \left\{ \mathbf{p}_{k,t}^C, \mathbf{X}_{k,t}^C, \mathbf{p}_{k,t}^L, \mathbf{X}_{k,t}^L, \mathbf{Z}_{k,t}^C, \mathbf{Q}_{k,t}, \mathbf{Y}_{k,t}^C, \mathbf{Y}_{k,t}^L \right\}_{t=1}^T$$

for Decision-Making Unit (DMU) $k = 1, \ldots, K$. The DMU under evaluation is $k = 0$.

2.3.1 Technology

The crop production set for a variable-returns-to-scale technology can be empirically approximated as:

$$\hat{\mathfrak{Y}}_t^C = \Big\{ (\mathbf{X}_{0,t}^C, \mathbf{Q}_{0,t}, \mathbf{Y}_{0,t}^C, \mathbf{Z}_{0,t}^C) : \sum_{k=1}^K \lambda_{k,t} \mathbf{X}_{k,t}^C \leq \mathbf{X}_{0,t}^C, \tag{2.10a}$$

$$\sum_{k=1}^K \lambda_{k,t} \mathbf{Q}_{k,t} \leq \mathbf{Q}_{0,t}, \tag{2.10b}$$

$$\sum_{k=1}^K \lambda_{k,t} (\mathbf{Y}_{k,t}^C + \mathbf{Z}_{k,t}^C) \geq (\mathbf{Y}_{0,t}^C + \mathbf{Z}_{0,t}^C), \tag{2.10c}$$

$$\sum_{k=1}^K \lambda_{k,t} = 1, \tag{2.10d}$$

$$\lambda_{k,t} \geq 0 \,\Big\}. \tag{2.10e}$$

The livestock production set for a variable-returns-to-scale technology can be empirically approximated as:

$$\hat{\mathfrak{Y}}_t^L = \Big\{ (\mathbf{X}_{0,t}^L, \mathbf{Z}_{0,t}^C, \mathbf{Q}_{0,t}, \mathbf{Y}_{0,t}^L) : \sum_{k=1}^K \gamma_{k,t} X_{k,t}^{L,h} \leq X_{0,t}^{L,h} \; \forall h \notin H, \tag{2.11a}$$

$$\sum_{k=1}^K \gamma_{k,t} (Z_{k,t}^{C,h} + X_{k,t}^{L,h}) \leq Z_{0,t}^{C,h} + X_{0,t}^{L,h} \; \forall h \in H, \tag{2.11b}$$

$$\sum_{k=1}^K \gamma_{k,t} \mathbf{Q}_{k,t} \leq \mathbf{Q}_{0,t}, \tag{2.11c}$$

$$\sum_{k=1}^K \gamma_{k,t} \mathbf{Y}_{k,t}^L \geq \mathbf{Y}_{0,t}^L, \tag{2.11d}$$

$$\sum_{k=1}^K \gamma_{k,t} = 1, \tag{2.11e}$$

$$\gamma_{k,t} \geq 0 \,\Big\}. \tag{2.11f}$$

These approximations are the inner bound approximations of the technology (Varian, 1984). From these subprocesses we obtain the approximation $\hat{\mathfrak{Y}}_t$ of the overall technology by taking the intersection of $\hat{\mathfrak{Y}}_t^C$ and $\hat{\mathfrak{Y}}_t^L$.

2.3.2 Inefficiency measurement

The implementation of the directional distance function (2.5) is:

$$D_t(\mathbf{X}_{0,t}, \mathbf{Y}_{0,t}; \mathbf{g}_t) = \sup\left\{\beta \in \mathbb{R} : (\mathbf{X}_{0,t} - \beta\mathbf{g}_{x,t}, \mathbf{Y}_{0,t} + \beta\mathbf{g}_{y,t}) \in \hat{\mathfrak{Y}}_t\right\}. \tag{2.12}$$

The combination of subprocesses is implemented by solving linear programme (2.13) which essentially combines (2.10), (2.11) and (2.12). The linear programmes are relegated to the Appendix to conserve space. In line with the literature (*e.g.*, Chambers et al. (1996b)), we select $\mathbf{g}_{x,t} = \mathbf{X}_t$ and $\mathbf{g}_{y,t} = \mathbf{Y}_t$ as the directional vectors. This choice ensures that the contemporaneous directional distance function is feasible (Briec and Kerstens, 2009) and can be interpreted as the maximum proportional contraction of variable inputs and simultaneously as the maximum proportional expansion of outputs.

The directional distance function which allows for reallocation of land and the process-specific variable costs is computed by solving the linear programme (2.14) in the Appendix. Compared to (2.13), the crop-specific ($\mathbf{X}^C_{0,t}$) and livestock-specific ($\mathbf{X}^L_{0,t}$) variable inputs are additional choice variables in this linear programme. Furthermore, it has two additional constraints to ensure that (i) the sum of the optimal crop land and livestock land is equal to the total land area; (ii) the process-specific variable costs can be optimally redistributed over the crop and livestock activities. (ii) is in line with Färe and Grosskopf (2012)'s cost-constrained efficiency measure. Consider the following example to see why this redistribution of the process-specific variable costs is necessary: it would make little sense for a fully specialised livestock farm to diversify into crops without this reallocation of the budget, for he would not be able to buy the necessary seeds for his crop land (i.e., $X^{C,m}_{0,t} = 0$ in (2.13a)). Therefore, he must be able to reallocate part of his budget to crop specific inputs (such that $X^{C,m}_{0,t} > 0$). An analogous reasoning holds for a fully specialised crop farm.[9]

Note that our model makes several implicit assumptions about land use. First, we assume that land is immediately reallocatable and costless.[10] Second, we assume that all the farm's utilised land is substitutable between crops and livestock and thus we do not take into account heterogeneity of land quality. In practice, at least some mixed farms are mixed precisely because some of the land is not suitable for crop production. We can drop this assumption by treating reallocated units of land differently in the production technology depending on their land quality. For instance, one could do this by discounting reallocated units of land in $\hat{\mathfrak{Y}}^C_t$ and $\hat{\mathfrak{Y}}^L_t$. This would reflect that, for example, 1 unit of livestock land reallocated to crop production is worth 0.8 unit of crop land. This modification implements the "input specificity" concept described in

[9] In exceptional cases, this budget constraint may lead to $D_a(\mathbf{X}_{0,b}, \mathbf{Y}_{0,b}; \mathbf{g}_b|R) < D_a(\mathbf{X}_{0,b}, \mathbf{Y}_{0,b}; \mathbf{g}_b|NR)$ for $(a, b) \in \{t, t+1\} \times \{t, t+1\}$ and $a \neq b$ if the technically efficient allocations fall outside the budget constraint. One can solve this by using expenditures instead of (implicit) quantities, as prices are effectively equal to unity when using expenditures. However, using expenditures conflates technical and economic efficiency.

[10] This assumption can be weakened by assuming that reallocation leads to temporary reductions in production. We refer to Oude Lansink and Stefanou (2001), Nemoto and Goto (1999, 2003) and Silva and Stefanou (2003, 2007) for specific examples to model these adjustment costs.

Caballero and Hammour (1998). However, this modification requires detailed data on land quality which is unavailable to us at the time of writing.

2.4 Empirical Application

2.4.1 Data description

Our empirical application focusses on a large sample of specialised and mixed farms in England and Wales. We obtain data from the Farm Business Survey (FBS) dataset covering the period $2007 - 2013$. The FBS dataset provides farm-level information on economic and physical characteristics. It is unbalanced but statistically representative. Farms remain in the panel for a maximum of on average $5 - 7$ years. To model the complex production processes on the farm in a detailed way, this paper exploits the rich characterisation of outputs and inputs of the FBS dataset. We distinguish 2 outputs, 12 variable inputs and 6 fixed factors. The outputs are crop production and livestock production. Note that, in general, one is likely to encounter a lot of specialisation in data with few aggregate outputs than in data with more disaggregate outputs. Joint non-reallocatable variable inputs are energy use, water use, hired labour and other inputs (costs on insurance, bank charges, professional fees, vehicle tax and other general farming costs). Crop-specific inputs are seed and young plants, fertilisers, crop protection and other variable crop costs. Livestock-specific inputs are bought feed and fodder, veterinary costs and medicine, and other livestock costs and the non-marketed crop output used as feed. Family labour and joint capital costs are joint non-reallocatable fixed factors. Aggregated crop-specific capital costs (permanent crops, debtors of crop subsidies, off-farm grain storage, crops, cultivations and stores) and livestock-specific capital costs (livestock and forage) are crop- and livestock-specific fixed factors, respectively. Crop land and livestock land are assumed to be fixed factors that are reallocatable among the outputs. This implies that total land use (and thus also farm size) is assumed to be fixed for a given year, but that the farmer can choose how much land to allocate to crop production and livestock production. Except for (hired and family) labour and land, which are measured in annual working hours and hectares, respectively, all inputs and outputs are measured in constant 2007£. We compute implicit quantities of outputs and capital costs by calculating the ratio of value to the respective price index. We aggregate the monetary crop-specific, livestock-specific and joint variable inputs as implicit quantities by computing the ratio of their aggregated value to their corresponding aggregated Törnqvist price index. Price indices vary over the years but not over the farms, implying that differences in the composition or quality of inputs and outputs are reflected by differences in implicit quantity (Cox and Wohlgenant, 1986). The separate price indices are obtained from the Eurostat (2015) database.

Data Envelopment Analysis (DEA) is sensitive to different environmental conditions (*e.g.*, weather conditions), outliers and measurement errors. We address these drawbacks as follows. First, we control for environmental differences by separately running the DEA models per region. East Midlands (EM), East of England (EE), South East (SE), North

East (NE), North West (NW), Yorkshire & the Humber (YH), South West (SW), West Midlands (WM) and Wales (WA) are the considered regions. Second, we remove influential outliers using the approach developed by Banker and Chang (2006). We run DEA model (2.14) for each observation by excluding the observation itself from the reference technology. Outliers are situated well outside the adjusted reference technology and appear 'super-efficient' (Banker and Chang, 2006) with a score substantially below zero. We only include the observations with a $D_t(\cdot|NR)$ between the 5 and 95 percentile.[11] Since we explicitly account for heterogeneous technologies in our specification, we include specialised as well as mixed farms in our analysis. The eventual dataset contains $12,738$ observations for a period of seven years.

Table 2.1 shows the descriptive statistics of the variables used in the analysis. In particular, the last column indicates the presence of completely specialised farms in our sample (i.e., $Y_t^C + Z_t^C = 0$, $Y_t^L = 0$ and $X_t^{C,m}, X_t^{L,m} = 0$).

Variables		Dimensions	Average	Std. Dev.	Min
Crop-specific variable inputs	X_t^C	Constant 2007£	47,731	144,414	0
Livestock-specific variable inputs	X_t^L	Constant 2007£	15,160	28,083	0
Non-labour joint variable inputs	$Q_t^f,\ f \notin F$	Constant 2007£	19,894	36,630	1002
Hired labour		Ann. Working Hours	4,977	15,592	0
Family labour	$Q_t^f,\ f \in F$	Ann. Working Hours	2,631	969	5
Joint capital		£	1,086,669	1,505,083	11,564
Crop-specific capital	$X_t^{C,m},\ m \notin C$	Constant 2007£	85,416	250,381	0
Livestock-specific capital	$X_t^{L,m},\ m \notin C$	Constant 2007£	90,660	104,562	0
Crop land	$X_t^{C,m},\ m \in C$	Hectares	131	334	0
Livestock land	$X_t^{L,m},\ m \in C$	Hectares	224	286	0
Total crop output	$Y_t^C + Z_t^C$	Constant 2007£	94,271	323,051	0
Crop output used as feed	Z_t^C	Constant 2007£	2,449	7,781	0
Livestock output	Y_t^L	Constant 2007£	165,403	376,680	-2,893

Table 2.1: Descriptive statistics of variables.

2.4.2 Static analysis: decomposing technical inefficiency

Table 2.2 presents the results of the static analysis of coordination inefficiency, CI, technical inefficiency when process-specific inputs over crops and livestock are optimally chosen, $D_t(\mathbf{X}_t, \mathbf{Y}_t; \mathbf{g}_t|R)$, and technical inefficiency when reallocation of process-specific inputs is not allowed, $D_t(\mathbf{X}_t, \mathbf{Y}_t; \mathbf{g}_t|NR)$.[12] $D_t(\mathbf{X}_t, \mathbf{Y}_t; \mathbf{g}_t|NR)$ ranges from 0.044 (in NE) to 0.131 (in WA). Considering our specification of directional vectors, this means that farms in NE and WA could simultaneously expand their output levels and contract their input levels by on average 4.4% and 13.1%, respectively if their land use would remain fixed. These regions also provide the lowest (0.092) and highest (0.194)

[11] Infeasibilities may appear when the considered observation has a peer with a projected negative output. Ray (2008) shows that these observations then have a score of lower than -1. We treat these observations as outliers.

[12] We only include the arithmetic averages to conserve space, but yearly results are available from the authors upon request.

corresponding $D_t(\mathbf{X}_t, \mathbf{Y}_t; \mathbf{g}_t|R)$ if land use would be optimally reallocated. The wedge between $D_t(\mathbf{X}_t, \mathbf{Y}_t; \mathbf{g}_t|R)$ and $D_t(\mathbf{X}_t, \mathbf{Y}_t; \mathbf{g}_t|NR)$, CI, is on average small and ranges from 0.036 (in EM) to 0.063 (in WA and WM). Thus, several regions may reduce technical inefficiency by optimally diverting land use to livestock and crops.

This table also analyses the differences in CI, $D_t(\mathbf{X}_t, \mathbf{Y}_t; \mathbf{g}_t|R)$ and $D_t(\mathbf{X}_t, \mathbf{Y}_t; \mathbf{g}_t|NR)$ among livestock farms (crop production covers 0-33% of total utilised land area), mixed farms (livestock production/crop production covers 33-66% of total utilised land area) and crop farms (livestock production covers 0-33% of total utilised land area).

Table 2.2 shows a clear pattern in differences in CI regarding farm types. CI in livestock farms is higher than CI in crop farms. CI in mixed farms is higher (lower) than CI in crop (livestock) farms. Specialisation in crop production thus leads to reduction in coordination inefficiency and better allocation of process-specific inputs. In what follows, we discuss the results that are significant at the 10% level using the Wilcoxon rank test (see Table 2.5 in the Appendix). CI is significantly higher in livestock farms than in crop farms for all regions. Livestock farms have a significantly higher CI than do mixed farms in NW, SW and WA. In EM, CI is significantly higher in mixed farms than in livestock farms, but the difference is very small (0.006). In the majority of regions (EE, EM, NW, SE, SW and WM), CI is higher in mixed farms than in crop farms. In summary, these differences in CI are not only statistically significant, but also economically significant.

Turning to the non-reallocative directional distance function $D_t(\cdot|NR)$, no such pattern is present: in some regions (NW, SE, SW, WA, WM and YH) $D_t(\cdot|NR)$ of livestock farms is higher than for crop farms, while in others (EE, EM and NE) the opposite holds. This result is significant for the majority of the regions. Similar ambiguity holds for comparing mixed farms to specialised farms.

These inefficiencies are generally lower than those found in the efficiency literature on agriculture in the United Kingdom.[13] However, it is difficult to compare our results with those in the literature. All previous results use radial measures focussing solely on input reductions or output expansions. In addition, all (but Hadley et al. (2013)) employ Stochastic Frontier Analysis where inefficiency is separated from random noise. Finally, previous studies use a subsample of our sample by only focussing on one type of farms or using data spanning different periods.

Table 2.3 and Figure 2.3 analyse the land use changes that are associated with eliminating coordination inefficiency. If $CI = 0$, farms do not need to change land use (n.). If $CI > 0$, (2.14) also allows us to compute the optimal land allocations. Farms should then either allocate more land to livestock (-) or crops (+). Interestingly, the required reallocation is considerably skewed towards more crop land use. For almost every region, there is a higher proportion of farms that would need to allocate more land to crops than to livestock. This holds for livestock farms, mixed farms as well as crop farms. This implies gains from specialisation (diversification towards mixed farming) for crop (livestock) farms. This confirms the above finding that livestock farms have more scope to reduce inefficiency by reallocating process-specific inputs.

[13]Similar studies were conducted by Areal et al. (2012); Hadley (2006); Hadley et al. (2013); Karagiannis et al. (2002, 2004); Wilson et al. (2001).

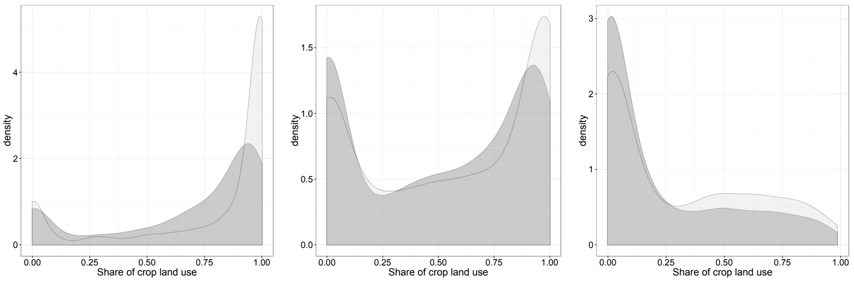
Region		Total sample	Livestock	Mixed	Crops
EE	CI	0.037	0.073	0.057	0.027
	$D_t(\cdot\|R)$	0.139	0.145	0.164	0.138
	$D_t(\cdot\|NR)$	0.102	0.071	0.107	0.110
EM	CI	0.036	0.046	0.052	0.030
	$D_t(\cdot\|R)$	0.095	0.086	0.138	0.099
	$D_t(\cdot\|NR)$	0.059	0.040	0.085	0.070
NE	CI	0.048	0.049	0.050	0.044
	$D_t(\cdot\|R)$	0.092	0.093	0.114	0.089
	$D_t(\cdot\|NR)$	0.044	0.044	0.064	0.045
NW	CI	0.044	0.048	0.044	0.025
	$D_t(\cdot\|R)$	0.112	0.122	0.105	0.069
	$D_t(\cdot\|NR)$	0.069	0.074	0.061	0.044
SE	CI	0.042	0.063	0.050	0.028
	$D_t(\cdot\|R)$	0.112	0.137	0.136	0.095
	$D_t(\cdot\|NR)$	0.070	0.074	0.086	0.067
SW	CI	0.053	0.059	0.048	0.036
	$D_t(\cdot\|R)$	0.175	0.186	0.185	0.145
	$D_t(\cdot\|NR)$	0.122	0.127	0.137	0.109
WA	CI	0.063	0.064	0.038	0.034
	$D_t(\cdot\|R)$	0.194	0.196	0.087	0.069
	$D_t(\cdot\|NR)$	0.131	0.132	0.049	0.035
WM	CI	0.063	0.073	0.069	0.043
	$D_t(\cdot\|R)$	0.170	0.192	0.184	0.131
	$D_t(\cdot\|NR)$	0.108	0.119	0.115	0.087
YH	CI	0.048	0.053	0.039	0.043
	$D_t(\cdot\|R)$	0.100	0.105	0.117	0.093
	$D_t(\cdot\|NR)$	0.052	0.053	0.078	0.051

Table 2.2: Average static coordination inefficiency per region and level of specialisation.

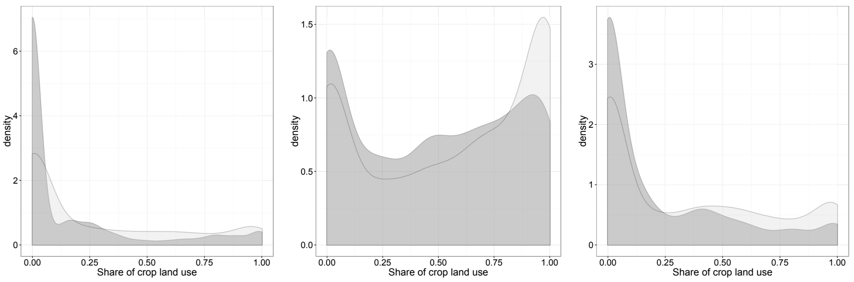
These results hold only to a much lesser extent to WA, where the majority of farms (63.6%) should not change their land allocation although the overwhelming majority of farms are livestock farms.

2.4.3 Dynamic analysis: decomposing Luenberger productivity growth

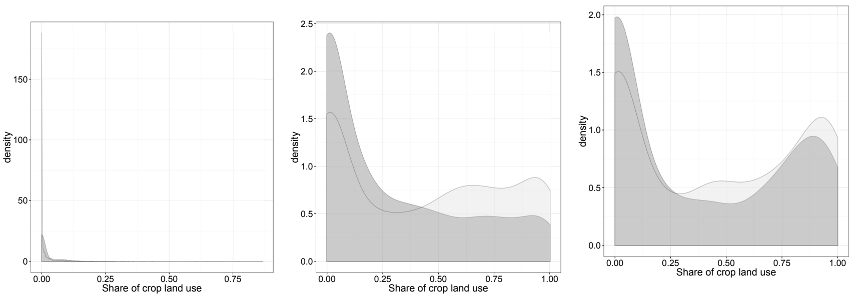
Table 2.4 presents the average coordination Luenberger productivity growth and its decomposition into coordination technical change ΔCT and coordination inefficiency change ΔCI for all regions. The chosen directional vectors ensure that all contemporaneous DEA scores are feasible. Infeasibilities may arise for the components where the time period of the observation differs from the time period of the reference technology.



(a) East of England. (b) East Midlands. (c) North East.



(d) North West. (e) South East. (f) South West.



(g) Wales. (h) West Midlands. (i) Yorkshire & the Humber.

Figure 2.3: Distribution of optimal (□) and actual (■) land allocation in function of the proportion of land allocated to crops.

region	Livestock			Mixed: 50 − 66% Livestock			Mixed: 50 − 66% Crops			Crops		
	-	n.	+	-	n.	+	-	n.	+	-	n.	+
EE	0.022	0.078	0.099	0.014	0.000	0.040	0.015	0.004	0.066	0.118	0.080	0.465
EM	0.036	0.172	0.170	0.039	0.008	0.043	0.037	0.005	0.049	0.106	0.051	0.285
NE	0.132	0.279	0.308	0.023	0.002	0.055	0.028	0.010	0.036	0.051	0.011	0.065
NW	0.123	0.381	0.315	0.018	0.002	0.021	0.010	0.002	0.014	0.026	0.024	0.064
SE	0.068	0.135	0.192	0.049	0.002	0.061	0.041	0.001	0.079	0.093	0.082	0.199
SW	0.114	0.281	0.331	0.034	0.002	0.061	0.019	0.002	0.044	0.022	0.031	0.059
WA	0.169	0.636	0.182	0.005	0.001	0.003	0.002	0.000	0.000	0.000	0.000	0.001
WM	0.062	0.233	0.349	0.021	0.002	0.071	0.027	0.004	0.046	0.044	0.027	0.116
YH	0.072	0.227	0.251	0.040	0.005	0.029	0.025	0.005	0.026	0.115	0.032	0.174

Table 2.3: Share of farms that should allocate more land to livestock (-), crops (+) or which mix should remain unchanged (n.) (averaged over all years).

No easy solutions exist to avoid infeasibilities. Briec and Kerstens (2009) therefore recommend to simply report the number of infeasibilities (Table 2.7 in the Appendix). The share of infeasibilities is very to moderately small, ranging from 4.67% to 23.67%.

Depending on the region, average $L_{t,t+1}(NR)$ per year ranges from -29.5% to 8.8%. Whereas annual average productivity declines in EM (-29.5%), NE (-3.3%), SE (-7.3%), SW (-1.1%) and YH (-4.1%), it increases in EE (4.9%), NW (8.8%), WA (0.8%) and WM (1.7%). The average coordination Luenberger productivity growth ranges from -9.7% to 15.9%, depending on the region. This is driven by ΔCT rather than ΔCI. The ability to reallocate process-specific inputs over time does not change substantially, whereas changes in the technology due to reallocation plays an important role.

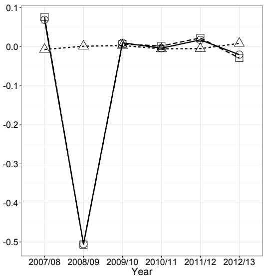
In what follows, we only discuss the results that are statistically significant at the 10% level according to the Kolmogorov-Smirnov test reported in Table 2.6 in the Appendix. Except for WM and YH, there are no significant differences in distribution of $L_{t,t+1}(NR)$ according to farm types. In contrast, the distributions of $CL_{t,t+1}$ differ according to farm types, although the sign of the statistical dominance is unclear in all regions except for NE.

Figure 2.4 shows the average coordination Luenberger productivity growth over time for each region. In every region, $CL_{t,t+1}$ is driven by ΔCT. Several large fluctuations occur for $CL_{t,t+1}$ and ΔCT between some years, which may be caused by weather conditions or by a few frontier farms that drive ΔCT.

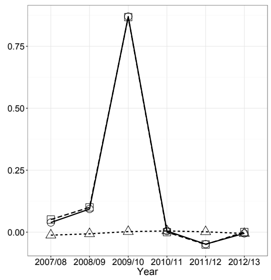
Although the results show a clear pattern, we remain cautious as we have not taken into account heterogeneity of land quality. We cannot rule out the possibility that the results are partly an artifact of the data and the assumption of complete substitutability between crop land use and livestock land use. This may be an issue especially in regions with heterogeneous soils (*e.g.*, NE and WA).

	regions								
	EE	EM	NE	NW	SE	SW	WA	WM	YH
$L_{t,t+1}(NR)$	0.049	-0.295	-0.033	0.088	-0.073	-0.011	0.008	0.017	-0.041
$L_{t,t+1}(NR)$ for livestock farms	-0.009	-0.165	-0.035	0.107	-0.031	-0.017	0.007	0.015	0.044
$L_{t,t+1}(NR)$ for mixed farms	-0.025	-1.364	-0.004	-0.037	-0.191	-0.001	0.083	-0.017	-0.012
$L_{t,t+1}(NR)$ for crop farms	0.061	-0.374	-0.028	-0.015	-0.095	0.005	0.083	0.021	-0.135
$CL_{t,t+1}$	-0.072	0.159	0.019	-0.097	0.059	0.008	-0.006	0.051	0.014
$CL_{t,t+1}$ for livestock farms	-0.042	0.043	0.029	-0.112	0.027	0.014	-0.006	0.086	-0.048
$CL_{t,t+1}$ for mixed farms	-0.003	0.856	-0.030	-0.004	0.183	-0.005	-0.061	0.029	-0.015
$CL_{t,t+1}$ for crop farms	-0.079	0.228	-0.008	-0.013	0.077	-0.007	-0.061	-0.016	0.083
ΔCT	-0.072	0.161	0.015	-0.094	0.058	0.008	-0.002	0.056	0.016
ΔCT for livestock farms	-0.041	0.045	0.027	-0.111	0.026	0.015	-0.002	0.093	-0.045
ΔCT for mixed farms	0.002	0.865	-0.046	0.017	0.179	-0.010	-0.059	0.026	-0.014
ΔCT for crop farms	-0.078	0.231	-0.016	-0.004	0.075	-0.012	-0.059	-0.015	0.084
ΔCI	-0.001	-0.003	0.004	-0.002	0.001	0.000	-0.004	-0.005	-0.002
ΔCI for livestock farms	-0.001	-0.002	0.002	-0.001	0.001	-0.001	-0.004	-0.007	-0.003
ΔCI for mixed farms	-0.004	-0.009	0.016	-0.021	0.004	0.004	-0.001	0.003	-0.001
ΔCI for crop farms	-0.001	-0.003	0.009	-0.009	0.002	0.004	-0.001	-0.001	-0.000

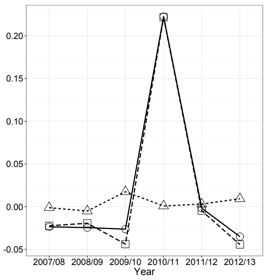
Table 2.4: Average Luenberger productivity change and its components.



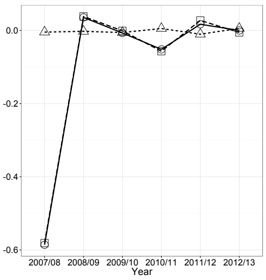
(a) East of England.



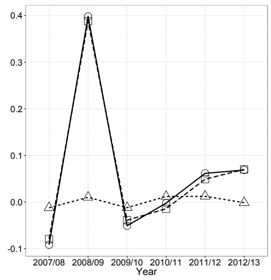
(b) East Midlands.



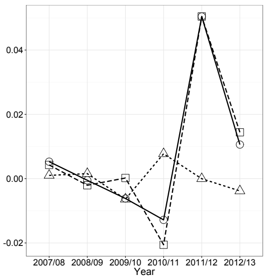
(c) North East.



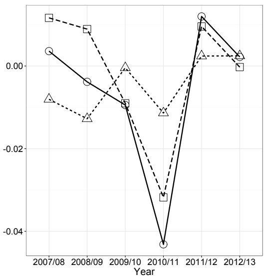
(d) North West.



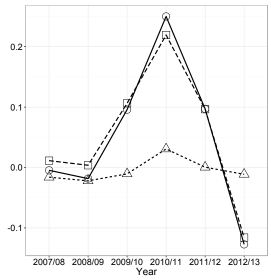
(e) South East.



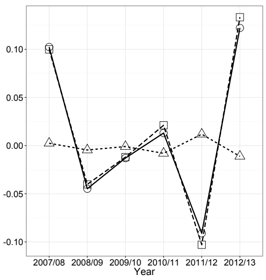
(f) South West.



(g) Wales.



(h) West Midlands.



(i) Yorkshire & the Humber.

Figure 2.4: Decomposition of $CL_{t,t+1}$ (—o—) in ΔCI (-△-) and ΔCT (-□-) per region.

2.5 Conclusions

This paper develops a nonparametric measure of coordination Luenberger productivity growth where the subprocesses are explicitly modelled in the production technology. This indicator allows us to assess the change in the farmers' ability to allocate inputs over crop and livestock outputs over time. Focussing on a large panel of English and Welsh farms over the period $2007-2013$, this paper demonstrates how better coordination of process-specific inputs may increase efficiency and productivity. We decompose coordination Luenberger productivity growth into coordination technical change and coordination inefficiency change. We compute the efficiency and productivity measures separately per region.

The static analysis shows a clear pattern: crop farms have a lower coordination inefficiency than livestock farms, *i.e.*, they allocate their process-specific inputs more adequately. This result is statistically significant across all regions. Furthermore, coordination inefficiency in mixed farms is higher (lower) than coordination inefficiency in crop (livestock) farms. Coordination efficiency gains are associated with allocating more land use to crop production. In contrast, no such pattern exists considering the results for the non-reallocative directional distance function, which is now the standard way of measuring technical inefficiency. This demonstrates that richer modelling of subprocesses uncovers an additional source of inefficiency due to misallocation of resources.

According to the dynamic analysis, average non-reallocative Luenberger productivity growth per year ranges from -29.5% to 8.8%, with considerable differences across regions. The Kolmogorov-Smirnov test finds almost no significant distributional differences in farm types. We further find that average coordination Luenberger productivity growth ranges from -9.7% to 15.9%, depending on the region. This is driven by coordination technical change rather than coordination inefficiency change. The ability to reallocate process-specific inputs over time does not change substantially, whereas changes in the technology due to reallocation plays an important role. The Kolmogorov-Smirnov test shows significant distributional differences in farm types, which contrasts the findings regarding the non-reallocative Luenberger productivity indicator. However, we find inconclusive evidence about which farm type stochastically dominates. Again, modelling subprocesses and allowing for reallocation reveal differences in optimally allocating resources over time. These differences are linked to heterogeneity in production technologies of different farm types.

Although researchers and policy makers identified an interest in stimulating mixed agriculture due to its environmental benefits, our results indicate that caution may be required. Since coordination efficiency gains are generally associated with more crop production for all farm types, one should stimulate mixed farming in livestock farms rather than crop farms. However, this does not necessarily imply that crop farms are more able to optimally allocate resources over time. Despite the clear patterns in the results, we remain prudent about the policy implications as the clear patterns may partly be an artifact of the data and the assumption of complete substitutability between crop land use and livestock land use.

We have several recommendations for future research. First, we recommend opening the black-box of efficiency and productivity by explicitly modelling the subprocesses. This can guide decision makers in coordinating the subprocesses to enhance efficiency and productivity. Second, this framework can be extended by including stochastic factors. Agricultural production is impacted by weather conditions, which cannot be influenced by the farms through choices of inputs and outputs. Efficiency is biased downwards (upwards) under bad (good) weather conditions. We have only partially controlled for this issue by running the nonparametric models per region. This problem can be dealt with in a more structural way by using Stochastic Frontier Analysis or making DEA conditional on environmental variables (de Witte and Kortelainen, 2013; Jeong et al., 2010). Finally, this framework can be augmented by taking into account intertemporal linkages. For instance, manure from livestock enterprises can be modelled as future inputs of crop production. Applied to the context of English and Welsh agriculture, this will be possible if more fertiliser surveys become available.

2.A Additional Tables

		Region								
		EE	EM	NE	NW	SE	SW	WA	WM	YH
CI	Livestock - Mixed	0.227	0.028	0.920	0.057	0.676	0.016	0.001	0.752	0.346
	Livestock - Crops	0.000	0.000	0.069	0.000	0.000	0.000	0.000	0.000	0.003
	Crops - Mixed	0.000	0.000	0.228	0.008	0.000	0.000	0.692	0.000	0.278
$D_t(\cdot\|R)$	Livestock - Mixed	0.117	0.000	0.166	0.097	0.274	0.903	0.000	0.673	0.489
	Livestock - Crops	0.583	0.111	0.627	0.000	0.000	0.000	0.000	0.000	0.299
	Crops - Mixed	0.127	0.001	0.130	0.024	0.000	0.000	0.673	0.001	0.197
$D_t(\cdot\|NR)$	Livestock - Mixed	0.000	0.000	0.089	0.357	0.016	0.263	0.000	0.935	0.030
	Livestock - Crops	0.000	0.000	0.648	0.000	0.728	0.001	0.000	0.001	0.908
	Crops - Mixed	0.550	0.072	0.083	0.060	0.013	0.001	0.762	0.011	0.038

Table 2.5: P-values of Wilcoxon rank test.

		Region								
		EE	EM	NE	NW	SE	SW	WA	WM	YH
$L_{t,t+1}(NR)$	Livestock - Mixed	0.553	0.547	0.547	0.147	0.917	0.835	0.291	0.061	0.640
	Livestock - Crops	0.274	0.254	0.183	0.123	0.957	0.563	0.291	0.148	0.046
	Crops - Mixed	0.993	1.000	1.000	0.722	0.998	1.000	1.000	0.980	0.812
$CL_{t,t+1}$	Livestock - Mixed	0.179	0.410	0.338	0.784	0.073	0.334	0.037	0.102	0.515
	Livestock - Crops	0.000	0.000	0.846	0.090	0.000	0.001	0.037	0.002	0.017
	Crops - Mixed	0.001	0.003	0.995	0.330	0.014	0.091	1.000	0.366	0.799
ΔCT	Livestock - Mixed	0.149	0.878	0.503	0.338	0.202	0.080	0.049	0.501	0.626
	Livestock - Crops	0.000	0.000	0.603	0.005	0.000	0.000	0.049	0.002	0.008
	Crops - Mixed	0.002	0.001	0.885	0.083	0.027	0.107	1.000	0.564	0.202
ΔCI	Livestock - Mixed	0.154	0.437	0.313	0.548	0.985	0.082	0.229	0.648	0.743
	Livestock - Crops	0.000	0.001	0.587	0.020	0.001	0.000	0.229	0.001	0.068
	Crops - Mixed	0.059	0.002	0.894	0.271	0.016	0.067	1.000	0.496	0.962

Table 2.6: P-values of Kolmogorov-Smirnov test for non-reallocative Luenberger productivity growth, coordination Luenberger productivity growth, coordination technical change and coordination inefficiency change.

Region	$D_{t+1}(\mathbf{X}_t, \mathbf{Y}_t; \mathbf{g}_t \vert NR)$	$D_t(\mathbf{X}_{t+1}, \mathbf{Y}_{t+1}; \mathbf{g}_{t+1} \vert NR)$	$D_{t+1}(\mathbf{X}_t, \mathbf{Y}_t; \mathbf{g}_t \vert R)$	$D_t(\mathbf{X}_{t+1}, \mathbf{Y}_{t+1}; \mathbf{g}_{t+1} \vert R)$
EE	14.06	13.75	8.97	8.61
EM	13.67	13.99	10.32	9.36
NE	21.63	16.79	15.75	12.26
NW	14.95	13.52	11.42	9.66
SE	16.92	11.70	12.10	8.09
SW	9.52	9.18	7.84	6.84
WA	6.92	9.49	4.67	7.05
WM	18.49	12.71	14.17	8.55
YH	23.67	17.11	18.35	12.62

Table 2.7: Share of infeasibilities over all years (in %).

2.B Linear Programmes

$D_t(\mathbf{X}_{0,t}, \mathbf{Y}_{0,t}; \mathbf{g}_t = (\mathbf{g}_{x,t}, \mathbf{g}_{y,t})|NR) =$

$$\max_{\substack{\beta,\\ \lambda_{k,t},\gamma_{k,t}\geq 0}} \quad \beta$$

$$\text{s.t.} \quad \sum_{k=1}^{K} \lambda_{k,t} X_{k,t}^{C,m} \leq X_{0,t}^{C,m} - \beta g_{x,t}^{C,m} \qquad \forall m \notin C, \tag{2.13a}$$

$$\sum_{k=1}^{K} \lambda_{k,t} X_{k,t}^{C,m} \leq X_{0,t}^{C,m} \qquad \forall m \in C, \tag{2.13b}$$

$$\sum_{k=1}^{K} \lambda_{k,t} Q_{k,t}^{f} \leq Q_{0,t}^{f} - \beta g_{Q,t}^{f} \qquad \forall f \notin F, \tag{2.13c}$$

$$\sum_{k=1}^{K} \lambda_{k,t} Q_{k,t}^{f} \leq Q_{0,t}^{f} \qquad \forall f \in F, \tag{2.13d}$$

$$\sum_{k=1}^{K} \lambda_{k,t} (\mathbf{Y}_{k,t}^{C} + \mathbf{Z}_{k,t}^{C}) \geq (\mathbf{Y}_{0,t}^{C} + \mathbf{Z}_{0,t}^{C}) + \beta \mathbf{g}_{y,t}^{C}, \tag{2.13e}$$

$$\sum_{k=1}^{K} \lambda_{k,t} = 1, \tag{2.13f}$$

$$\sum_{k=1}^{K} \gamma_{k,t} X_{k,t}^{L,m} \leq X_{0,t}^{L,m} - \beta g_{x,t}^{L,m} \qquad \forall m \notin C, \forall m \notin H, \tag{2.13g}$$

$$\sum_{k=1}^{K} \gamma_{k,t} X_{k,t}^{L,m} \leq X_{0,t}^{L,m} \qquad \forall m \in C, \tag{2.13h}$$

$$\sum_{k=1}^{K} \gamma_{k,t} (Z_{k,t}^{C,h} + X_{k,t}^{L,h}) \leq (Z_{0,t}^{C,h} + X_{0,t}^{L,h}) - \beta g_{x,t}^{h} \qquad \forall h \in H, \tag{2.13i}$$

$$\sum_{k=1}^{K} \gamma_{k,t} Q_{k,t}^{f} \leq Q_{0,t}^{f} - \beta g_{Q,t}^{f} \qquad \forall f \notin F, \tag{2.13j}$$

$$\sum_{k=1}^{K} \gamma_{k,t} Q_{k,t}^{f} \leq Q_{0,t}^{f} \qquad \forall f \in F, \tag{2.13k}$$

$$\sum_{k=1}^{K} \gamma_{k,t} \mathbf{Y}_{k,t}^{L} \geq \mathbf{Y}_{0,t}^{L} + \beta \mathbf{g}_{y,t}^{L}, \tag{2.13l}$$

$$\sum_{k=1}^{K} \gamma_{k,t} = 1. \tag{2.13m}$$

$$D_t(\mathbf{X}_{0,t}, \mathbf{Y}_{0,t}; \mathbf{g}_t = (\mathbf{g}_{x,t}, \mathbf{g}_{y,t})|R) =$$

$$\max_{\substack{\beta,\\ \lambda_{k,t},\gamma_{k,t}\geq 0,\\ \mathbf{X}^C_{0,t},\mathbf{X}^L_{0,t}\geq 0}} \quad \beta \tag{2.14a}$$

$$\text{s.t.} \quad (2.13a) - (2.13m) \tag{2.14b}$$

$$X^{C,m}_{0,t} + X^{L,m}_{0,t} = X^m_{0,t} \qquad \forall m \in C, \tag{2.14c}$$

$$\sum_{\forall i \in \{1,\ldots,N_C\}\setminus C} p^{C,i}_{0,t} X^{C,i}_{0,t} + \sum_{\forall j \in \{1,\ldots,N_L\}\setminus C} p^{L,j}_{0,t} X^{L,j}_{0,t} = PEXP_{0,t} \tag{2.14d}$$

Chapter 3

Decomposing the Luenberger-Hicks-Moorsteen Total Factor Productivity indicator: An application to U.S. agriculture

"By far the largest portion of the literature on total factor productivity is devoted to problems of measurement rather than to problems of explanation. In recognition of this fact changes in total factor productivity have been given such labels as The Residual or The Measure of Our Ignorance."

— Dale W. Jorgenson and Zvi Griliches[1]

3.1 Introduction

Assessing the drivers of productivity growth is important for business and economic policy. Their identification allows monitoring of industries and can guide policymakers in their decisions. Hence, an abundant literature has sought to decompose various measures of productivity growth into components of technical change, efficiency change and scale efficiency change.[2] The literature has largely focused on ratio-based productivity "indexes". Yet, O'Donnell (2012a) recently shows that not all such decomposable indexes are "multiplicatively complete" (*i.e.*, consisting of a *ratio* of an output aggregator to an input aggregator), while all multiplicatively complete indexes are decomposable in this way. He demonstrates that the class of multiplicatively complete productivity indexes includes Laspeyres, Paasche, Fischer, Törnqvist and Bjurek (1996)'s Hicks-Moorsteen indexes, but does not include the popular Malmquist index of Caves et al. (1982).

[0] This chapter is based on joint work with Frederic Ang (Swedish University of Agricultural Sciences) and a slightly different version was published as Ang and Kerstens (2017a) in European Journal of Operational Research.

[1] Jorgenson and Griliches (1967, p.249)

[2] See Färe et al. (1998) and Grosskopf (2003) for historical overviews.

Ratio-based productivity indexes are undefined when one or more of the variables are equal or close to zero (Balk et al., 2003). Difference-based productivity "indicators" do not suffer from this problem and are thus particularly useful in regulatory contexts.

Difference-based indicators were developed to measure Total Factor Productivity (TFP) growth based on Luenberger (1992)'s shortage function. This directional distance function, introduced by Chambers et al. (1996b) in a production context, extends the Shephard input and output distance functions by allowing for simultaneous contraction of inputs and expansion of outputs. Chambers (2002) introduced a general difference-based Luenberger productivity indicator which can be decomposed in a technical change and efficiency change component (Chambers et al., 1996b).[3] Since its introduction, it has frequently been applied in empirical applications (*e.g.*, Nakano and Managi (2008)) and additional decompositions of its technical change component (*e.g.*, Briec and Peypoch (2007)) and efficiency change component (*e.g.*, Epure et al. (2011)) have been proposed in the literature. However, the Luenberger productivity indicator is not "additively complete" (*i.e.*, consisting of a *difference* between an output aggregator and an input aggregator) and thus cannot be disentangled into components of output growth and input growth.

Briec and Kerstens (2004) introduced the Luenberger-Hicks-Moorsteen (LHM) TFP indicator, which is a difference-based, additively complete alternative to the ratio-based, multiplicatively complete Hicks-Moorsteen index.[4] Notwithstanding the attractive properties of the LHM TFP indicator, only few empirical studies can be found in the literature (*e.g.*, Barros et al. (2008) and Managi (2010)). One possible reason for the limited number of applications is the fact that a full decomposition into components of technical change, technical inefficiency change and scale inefficiency change has hitherto not been developed. A first effort was made by Managi (2010) who decomposed the LHM TFP indicator into components of technical change and (in)efficiency change. However, this decomposition lacks a scale inefficiency change component and does not correctly capture technical change and technical inefficiency change (see Appendix 3.A). No full decomposition of a difference-based TFP indicator being additively complete is thus presently known in the literature.

The current paper contributes to the existing literature by introducing a decomposition of the additively complete LHM TFP indicator into components of technical change, technical inefficiency change and scale inefficiency change. Our decomposition is general in that it does not require convexity or differentiability of the technology set. It is simi-

[3]The "economic" approach to productivity measurement requires price information and if in addition (i) some assumptions can be made about firm behavior and (ii) the technology is approximated by a known flexible functional form up to the second order, then one can use a "superlative" index as advocated by Diewert (1976). Chambers (2002) showed that the Bennet-Bowley indicators are exact and superlative approximations of the Luenberger productivity indicator under (i) profit-maximizing behavior and (ii) a quadratic technology directional distance function. Recently, Ang and Kerstens (2017b) show that the Bennet-Bowley profit indicator is also an exact and superlative approximation for the LHM TFP indicator under a suitable price normalization.

[4]See Briec et al. (2012) for exact relations between the Luenberger-Hicks-Moorsteen TFP indicator and the Hicks-Moorsteen TFP index.

lar to Diewert and Fox (2014, 2017)'s decomposition of the ratio-based Hicks-Moorsteen TFP index.

Using a nonparametric framework, we illustrate the decomposition with an empirical application to state-level data of the U.S. agricultural sector over the period $1960-2004$. Since our decomposition is suitable for non-convex as well as convex technologies, we demonstrate its flexibility by using the Free Disposal Hull as well as Data Envelopment Analysis. To the best of our knowledge, no other studies using the same dataset have investigated the issue of potential non-convexities. However, we believe that such an investigation is particularly relevant in the context of the agricultural sector. Inputs such as capital equipment are nondivisible, potentially leading to non-convexities.

This paper is structured as follows. The next section describes Luenberger's directional distance function and the LHM TFP indicator. We then introduce our complete decomposition and apply this to state-level data of the U.S. agricultural sector over the period $1960-2004$. The final section concludes.

3.2 The Luenberger-Hicks-Moorsteen TFP indicator

Let $\mathbf{x}_t \in \mathbb{R}^n_+$ be the nonnegative inputs that are used to produce nonnegative outputs $\mathbf{y}_t \in \mathbb{R}^m_+$. We define the technology set in the usual way:

$$\boldsymbol{\mathcal{Y}}_t = \left\{(\mathbf{x}_t, \mathbf{y}_t) \in \mathbb{R}^{n+m}_+ | \mathbf{x}_t \text{ can produce } \mathbf{y}_t\right\}.$$

Furthermore, we make the following minimal assumptions on the technology set (Chambers, 2002):

Axiom 3.1 (Closedness). *$\boldsymbol{\mathcal{Y}}_t$ is closed.*

Axiom 3.2 (Free disposability of inputs and outputs). *if $(\mathbf{x}'_t, -\mathbf{y}'_t) \geq (\mathbf{x}_t, -\mathbf{y}_t)$ then $(\mathbf{x}_t, \mathbf{y}_t) \in \boldsymbol{\mathcal{Y}}_t \Rightarrow (\mathbf{x}'_t, \mathbf{x}'_t) \in \boldsymbol{\mathcal{Y}}_t$.*

Axiom 3.3 (Inaction). *Inaction is possible: $(\mathbf{0}^n, \mathbf{0}^m) \in \boldsymbol{\mathcal{Y}}_t$.*

Convexity of the technology set is thus not a necessary condition for our decomposition.[5] We illustrate this in our empirical application.

Luenberger's directional distance function is a measure of technical inefficiency as it simultaneously contracts inputs and expands outputs. The directional distance function proposed by Chambers et al. (1996b) is:

$$D_t(\mathbf{x}_t, \mathbf{y}_t; \mathbf{g}_t) = \sup\left\{\beta \in \mathbb{R} : (\mathbf{x}_t - \beta\mathbf{g}^i_t, \mathbf{y}_t + \beta\mathbf{g}^o_t) \in \boldsymbol{\mathcal{Y}}_t\right\}, \tag{3.1}$$

[5]In fact, the LHM TFP indicator and our decomposition are applicable to a wider range of non-convex models that satisfy the above axioms and for which the directional distance function can be defined. Examples of these non-convex models include the Constant-Elasticity-of-Substitution-Constant-Elasticity-of-Transformation model of Färe et al. (1988), relaxed convexity model of Petersen (1990) and Bogetoft (1996), selective convexity model of Podinovski (2005) and B-convexity model of Briec and Liang (2011).

if $(\mathbf{x}_t - \beta\mathbf{g}_t^i, \mathbf{y}_t + \beta\mathbf{g}_t^o) \in \boldsymbol{\mathcal{Y}}_t$ for some β and $D_t(\mathbf{x}_t, \mathbf{y}_t; \mathbf{g}_t) = -\infty$ otherwise. Here, $\mathbf{g}_t = (\mathbf{g}_t^i, \mathbf{g}_t^o)$ represents the direction vector. The directional distance function is a special case of Luenberger (1992)'s shortage function.

We denote the time-related directional distance function for $(a, b) \in \{t, t+1\} \times \{t, t+1\}$:

$$D_b(\mathbf{x}_a, \mathbf{y}_a; \mathbf{g}_a) = \sup\left\{\beta \in \mathbb{R} : (\mathbf{x}_a - \beta\mathbf{g}_a^i, \mathbf{y}_a + \beta\mathbf{g}_a^o) \in \boldsymbol{\mathcal{Y}}_b\right\}.$$

Next, we turn to the Luenberger-Hicks-Moorsteen (LHM) TFP indicator proposed by Briec and Kerstens (2004). This can be seen as the difference-based equivalent of the ratio-based Hicks-Moorsteen (HM) TFP index. They define the LHM TFP indicator with base period t as the difference between a Luenberger output quantity indicator and a Luenberger input quantity indicator:

$$\begin{aligned} &LHM_t(\mathbf{x}_{t+1}, \mathbf{y}_{t+1}, \mathbf{x}_t, \mathbf{y}_t; \mathbf{g}_t, \mathbf{g}_{t+1}) \\ &= [D_t(\mathbf{x}_t, \mathbf{y}_t; (0, \mathbf{g}_t^o)) - D_t(\mathbf{x}_t, \mathbf{y}_{t+1}; (0, \mathbf{g}_{t+1}^o))] \\ &- \left[D_t(\mathbf{x}_{t+1}, \mathbf{y}_t; (\mathbf{g}_{t+1}^i, 0)) - D_t(\mathbf{x}_t, \mathbf{y}_t; (\mathbf{g}_t^i, 0))\right] \\ &\equiv LO_t(\mathbf{x}_t, \mathbf{y}_t, \mathbf{y}_{t+1}; \mathbf{g}_t^o, \mathbf{g}_{t+1}^o) - LI_t(\mathbf{x}_t, \mathbf{x}_{t+1}, \mathbf{y}_t; \mathbf{g}_t^i, \mathbf{g}_{t+1}^i). \end{aligned} \tag{3.2}$$

Similarly, a base period $t+1$ LHM TFP indicator is defined as:

$$\begin{aligned} &LHM_{t+1}(\mathbf{x}_{t+1}, \mathbf{y}_{t+1}, \mathbf{x}_t, \mathbf{y}_t; \mathbf{g}_t, \mathbf{g}_{t+1}) \\ &= [D_{t+1}(\mathbf{x}_{t+1}, \mathbf{y}_t; (0, \mathbf{g}_t^o)) - D_{t+1}(\mathbf{x}_{t+1}, \mathbf{y}_{t+1}; (0, \mathbf{g}_{t+1}^o))] \\ &- \left[D_{t+1}(\mathbf{x}_{t+1}, \mathbf{y}_{t+1}; (\mathbf{g}_{t+1}^i, 0)) - D_{t+1}(\mathbf{x}_t, \mathbf{y}_{t+1}; (\mathbf{g}_t^i, 0))\right] \\ &\equiv LO_{t+1}(\mathbf{x}_{t+1}, \mathbf{y}_{t+1}, \mathbf{y}_t; \mathbf{g}_t^o, \mathbf{g}_{t+1}^o) - LI_{t+1}(\mathbf{x}_t, \mathbf{x}_{t+1}, \mathbf{y}_{t+1}; \mathbf{g}_t^i, \mathbf{g}_{t+1}^i). \end{aligned} \tag{3.3}$$

O'Donnell (2012a, p.258, footnote 5) defines additive completeness as follows:

Definition 3.1 (Additive completeness). *Formally, let $TFPI(x_t, q_t, x_s, q_s)$ denote an index number that compares TFP in period s with TFP in period t using period s as a base. $TFPI(x_t, q_t, x_s, q_s)$ is additively complete if and only if it can be expressed in the form $TFPI(x_t, q_t, x_s, q_s) = Q(q_t) - Q(q_s) - X(x_t) + X(x_s)$ where $Q(\cdot)$ and $X(\cdot)$ are non-negative non-decreasing functions satisfying the translation property $Q(q+\lambda q) = Q(q)+\lambda$ and $X(x + \lambda x) = X(x) + \lambda$ for $\lambda > 0$.*[6]

$LHM_t(\cdot)$ and $LHM_{t+1}(\cdot)$ are "additively complete" in O'Donnell's sense. This can be verified from their definitions above where the directional distance function, along with its corresponding direction vector, serves as the output (using $(0, \mathbf{g}_s^o)$) and input (using $(\mathbf{g}_s^i, 0)$) aggregator functions.[7]

[6] Balk et al. (2003, p.157) show that the input quantity change indicators $LI_t(\cdot)$ and $LI_{t+1}(\cdot)$ with $\mathbf{g}_t^i = \mathbf{g}_{t+1}^i$ satisfy the desirable properties for a quantity change indicator listed in Diewert (2005). This equally applies to the output quantity change indicators $LO_t(\cdot)$ and $LO_{t+1}(\cdot)$ with $\mathbf{g}_t^o = \mathbf{g}_{t+1}^o$.

[7] Luenberger (1992)'s *shortage function* differs from Chambers (2002)' Luenberger *productivity indicator*. The shortage function satisfies the translation property. It is an aggregator function that can be

Finally, one takes an arithmetic average of LHM_t and LHM_{t+1} to avoid an arbitrary choice of base periods:[8]

$$LHM_{t,t+1}(\mathbf{x}_t, \mathbf{y}_t, \mathbf{x}_{t+1}, \mathbf{y}_{t+1}; \mathbf{g}_t, \mathbf{g}_{t+1}) = \frac{1}{2}\left[LHM_t + LHM_{t+1}\right]. \tag{3.4}$$

The HM TFP index is defined as the ratio of an output index to an input index. Similarly, we can show that the LHM TFP indicator equals the difference between an output indicator and an input indicator, which are themselves arithmetic averages of two output and two input indicators:

$$\begin{aligned} LHM_{t,t+1} &= \frac{1}{2}\left[LO_t + LO_{t+1}\right] - \frac{1}{2}\left[LI_t + LI_{t+1}\right] \\ &\equiv LO_{t,t+1} - LI_{t,t+1}. \end{aligned} \tag{3.5}$$

3.3 Decomposition of the Luenberger-Hicks-Moorsteen indicator

This section introduces our LHM decomposition along with illustrative figures in the one input - one output dimension to provide the intuition. We show an example with a non-convex technology (*i.e.*, Free Disposal Hull), as convexity is not a necessary assumption for our decomposition. Note, however, that one can also use our approach for a convex technology.

In line with the decomposition of the HM TFP index, the LHM TFP indicator can be decomposed using the output direction or input direction.[9] We focus on the decomposition using the output direction, but provide a similar decomposition using the input direction in Appendix 3.B. Our LHM decomposition is a specific case (analogous

used to compute components of an additively complete indicator (such as the LHM TFP indicator), but is *not* additively complete. Chambers (2002) defines the Luenberger productivity indicator as follows:

$$\begin{aligned} &L_{t,t+1}(\mathbf{x}_t, \mathbf{y}_t, \mathbf{x}_{t+1}, \mathbf{y}_{t+1}; \mathbf{g}_t, \mathbf{g}_{t+1}) \\ &\quad = \frac{1}{2}\Big[(D_t(\mathbf{x}_t, \mathbf{y}_t; \mathbf{g}_t) - D_t(\mathbf{x}_{t+1}, \mathbf{y}_{t+1}; \mathbf{g}_{t+1})) \\ &\quad + (D_{t+1}(\mathbf{x}_t, \mathbf{y}_t; \mathbf{g}_t) - D_{t+1}(\mathbf{x}_{t+1}, \mathbf{y}_{t+1}; \mathbf{g}_{t+1}))\Big], \end{aligned}$$

All directional vectors are determined in the input direction as well as the output direction, *i.e.*, $\mathbf{g}_a = (\mathbf{g}_a^i, \mathbf{g}_a^o) > 0$. This prevents us from disentangling the indicator into separate output and input aggregator functions.

[8] This average can be harder to interpret in regulatory and managerial contexts in which a clearer target is required. This can easily be accounted for by a different choice of weights for both periods: *i.e.*, we can define $LHM_{t,t+1} = \zeta LHM_t + (1-\zeta)LHM_{t+1}$ with weights $\zeta \in [0,1]$. One can then for example set $\zeta = 0$ or $\zeta = 1$. These weights trickle down in the technical change and scale inefficiency change components of our decomposition in a straightforward way.

[9] The technical change and technical inefficiency change components in particular are completely determined by this choice. The additive completeness property of the LHM TFP indicator can guide this decision by checking whether LHM TFP is mostly driven by $LO_{t,t+1}$ or $LI_{t,t+1}$. This contrasts with the Luenberger productivity indicator where both inputs and outputs contribute to its components.

to the multiplicatively complete case discussed in Section 3.7 of O'Donnell (2012a)) of an additively complete indicator that uses the directional distance function as the aggregator function for both inputs and outputs. Hence, in our case the mix efficiency change components are all 0 and our decomposition consists of three components:

$$LHM_{t,t+1} = \Delta T^o_{t,t+1} + \Delta TEI^o_{t,t+1} + \Delta SEC^o_{t,t+1}, \tag{3.6}$$

representing technical change, technical inefficiency change and scale inefficiency change respectively.[10] Given the close relation to the HM TFP index, it is no surprise that our decomposition is similar to Diewert and Fox (2014, 2017)'s decomposition of the HM TFP index.

The technical change component is

$$\begin{aligned}\Delta T^o_{t,t+1} = &\frac{1}{2}\left\{[D_{t+1}(\mathbf{x}_t, \mathbf{y}_t; (0, \mathbf{g}^o_t)) - D_t(\mathbf{x}_t, \mathbf{y}_t; (0, \mathbf{g}^o_t))]\right.\\ &\left.+ \left[D_{t+1}(\mathbf{x}_{t+1}, \mathbf{y}_{t+1}; (0, \mathbf{g}^o_{t+1})) - D_t(\mathbf{x}_{t+1}, \mathbf{y}_{t+1}; (0, \mathbf{g}^o_{t+1}))\right]\right\}\\ \equiv &\frac{1}{2}\left\{\Delta T^o_t + \Delta T^o_{t+1}\right\}.\end{aligned} \tag{3.7}$$

Technical change $\Delta T^o_{t,t+1}$ is the arithmetic average of ΔT^o_t and ΔT^o_{t+1}. Figure 3.1 depicts these technical change components. The arithmetic average is used to avoid an arbitrary choice of the observation under evaluation. Here, ΔT^o_t measures the difference in efficiency for observation $(\mathbf{x}_t, \mathbf{y}_t)$ evaluated against production frontier $t+1$ and t. An upward (downward) shift of the production frontier between t and $t+1$, indicating technical progress (regress), results in a positive (negative) difference. ΔT^o_{t+1} is similar to ΔT^o_t but evaluated for observation $(\mathbf{x}_{t+1}, \mathbf{y}_{t+1})$. Thus, technical change measures (local) shifts of the production frontier itself.

The technical inefficiency change component is

$$\Delta TEI^o_{t,t+1} = D_t(\mathbf{x}_t, \mathbf{y}_t; (0, \mathbf{g}^o_t)) - D_{t+1}(\mathbf{x}_{t+1}, \mathbf{y}_{t+1}; (0, \mathbf{g}^o_{t+1})), \tag{3.8}$$

and measures the change between period t and period $t+1$ in the relative position to the production frontier. Positive (negative) values of $\Delta TEI^o_{t,t+1}$ indicate efficiency improvement (deterioration) over time: $(\mathbf{x}_{t+1}, \mathbf{y}_{t+1})$ is located closer to (farther from) the $t+1$ frontier than $(\mathbf{x}_t, \mathbf{y}_t)$ was to the t frontier. In Figure 3.2 this means that $D_{t+1}(\mathbf{x}_{t+1}, \mathbf{y}_{t+1})$ is smaller (larger) than $D_t(\mathbf{x}_t, \mathbf{y}_t)$. Note that ΔTEI^o only measures the evolution in technical efficiency of the observation under consideration without taking into account changes of the production frontier over time.

This technical inefficiency change component can be further decomposed in the same way as done by Epure et al. (2011) for the Luenberger indicator into "pure" inefficiency and, for example, congestion changes.

[10] Managi (2010)'s decomposition lacks a scale inefficiency change component. We refer to Appendix 3.A for a discussion.

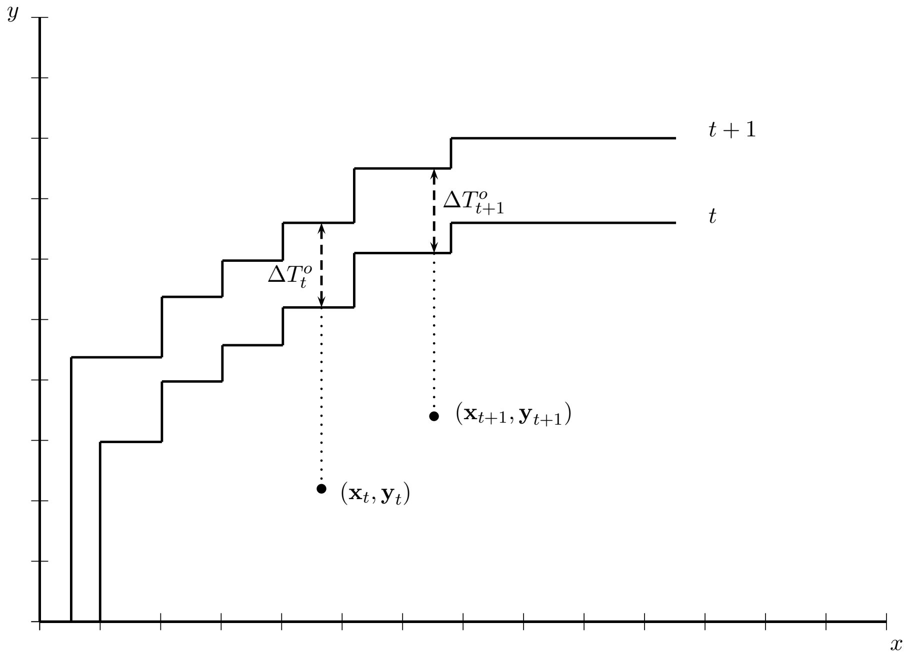


Figure 3.1: Technical change.

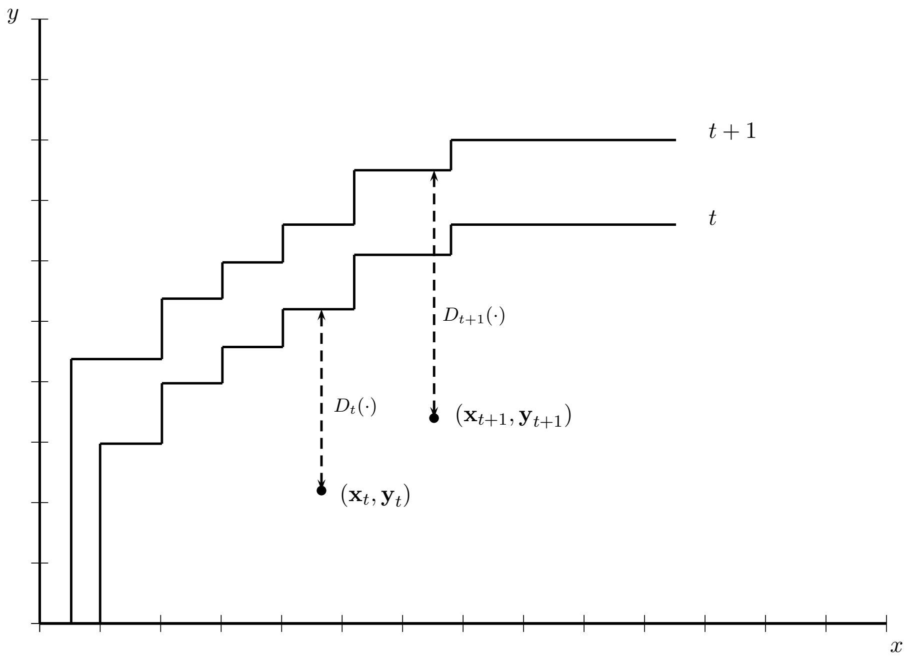


Figure 3.2: Technical inefficiency change.

Finally, from the residual

$$\begin{aligned} LHM_{t,t+1}-&\Delta T^o_{t,t+1} - \Delta TEI^o_{t,t+1} \\ =&\frac{1}{2}\left\{[D_t(\mathbf{x}_{t+1},\mathbf{y}_{t+1};(0,\mathbf{g}^o_{t+1})) - D_t(\mathbf{x}_t,\mathbf{y}_{t+1};(0,\mathbf{g}^o_{t+1}))]\right. \\ &\left.+ [D_{t+1}(\mathbf{x}_{t+1},\mathbf{y}_t;(0,\mathbf{g}^o_t)) - D_{t+1}(\mathbf{x}_t,\mathbf{y}_t;(0,\mathbf{g}^o_t))]\right\} \\ -&\frac{1}{2}\left\{\left[D_t(\mathbf{x}_{t+1},\mathbf{y}_t;(\mathbf{g}^i_{t+1},0)) - D_t(\mathbf{x}_t,\mathbf{y}_t;(\mathbf{g}^i_t,0))\right]\right. \\ &\left.+ \left[D_{t+1}(\mathbf{x}_{t+1},\mathbf{y}_{t+1};(\mathbf{g}^i_{t+1},0)) - D_{t+1}(\mathbf{x}_t,\mathbf{y}_{t+1};(\mathbf{g}^i_t,0))\right]\right\}, \end{aligned} \tag{3.9}$$

we can distill the scale inefficiency change component as follows. First, we define the projections of $\mathbf{y}_t$ and $\mathbf{y}_{t+1}$ on the production frontier at time t using notation of Diewert and Fox (2017):

$$\mathbf{y}^*_t = \mathbf{y}_t + D_t(\mathbf{x}_t,\mathbf{y}_t;(0,\mathbf{g}^o_t))\mathbf{g}^o_t \tag{3.10a}$$

$$\mathbf{y}^{**}_{t+1} = \mathbf{y}_{t+1} + D_t(\mathbf{x}_{t+1},\mathbf{y}_{t+1};(0,\mathbf{g}^o_{t+1}))\mathbf{g}^o_{t+1} \tag{3.10b}$$

Similarly, we define the projections of $\mathbf{y}_t$ and $\mathbf{y}_{t+1}$ on the production frontier at time $t+1$:

$$\mathbf{y}^{**}_t = \mathbf{y}_t + D_{t+1}(\mathbf{x}_t,\mathbf{y}_t;(0,\mathbf{g}^o_t))\mathbf{g}^o_t \tag{3.11a}$$

$$\mathbf{y}^*_{t+1} = \mathbf{y}_{t+1} + D_{t+1}(\mathbf{x}_{t+1},\mathbf{y}_{t+1};(0,\mathbf{g}^o_{t+1}))\mathbf{g}^o_{t+1} \tag{3.11b}$$

Then, respectively adding and subtracting $D_t(\mathbf{x}_t,\mathbf{y}_t;(0,\mathbf{g}^o_t))$ and $D_{t+1}(\mathbf{x}_{t+1},\mathbf{y}_{t+1};(0,\mathbf{g}^o_{t+1}))$ to and from (3.9), and using the translation property of the directional distance function and the definitions of the projections above, we find the scale inefficiency change component:

$$\begin{aligned} \Delta SEC^o_{t,t+1} =& \frac{1}{2}\left\{[D_t(\mathbf{x}_t,\mathbf{y}^*_t;(0,\mathbf{g}^o_t)) - D_t(\mathbf{x}_t,\mathbf{y}^{**}_{t+1};(0,\mathbf{g}^o_{t+1}))]\right. \\ &- \left[D_t(\mathbf{x}_{t+1},\mathbf{y}_t;(\mathbf{g}^i_{t+1},0)) - D_t(\mathbf{x}_t,\mathbf{y}_t;(\mathbf{g}^i_t,0))\right] \\ &+ [D_{t+1}(\mathbf{x}_{t+1},\mathbf{y}^{**}_t;(0,\mathbf{g}^o_t)) - D_{t+1}(\mathbf{x}_{t+1},\mathbf{y}^*_{t+1};(0,\mathbf{g}^o_{t+1}))] \\ &\left.- \left[D_{t+1}(\mathbf{x}_{t+1},\mathbf{y}_{t+1};(\mathbf{g}^i_{t+1},0)) - D_{t+1}(\mathbf{x}_t,\mathbf{y}_{t+1};(\mathbf{g}^i_t,0))\right]\right\} \\ \equiv& \frac{1}{2}\left\{SOC^o_t - SIC^o_t + SOC^o_{t+1} - SIC^o_{t+1}\right\} \\ \equiv& \frac{1}{2}\left\{\Delta SEC^o_t + \Delta SEC^o_{t+1}\right\}, \end{aligned} \tag{3.12}$$

which has the interpretation of measuring changes in "global" returns to scale in line with Diewert and Fox (2017). As a result, our scale inefficiency change component does not require differentiability or convexity of the production technology. Figure 3.3 illustrates the intuition behind (3.12). Again, the arithmetic average of ΔSEC^o_t and ΔSEC^o_{t+1} is used to avoid an arbitrary choice of base period for the technology. Both components have a similar interpretation as a finite difference approximation of the frontier's gradient. ΔSEC_t^o is a finite difference approximation of the frontier t's gradient and measures the change in inputs and outputs *along* the frontier when going from $(\mathbf{x}_t, \mathbf{y}_t)$ to $(\mathbf{x}_{t+1}, \mathbf{y}_{t+1})$. The change in inputs and outputs is measured separately: the SOC_t^o (SIC_t^o) subcomponent of ΔSEC_t^o keeps the inputs (outputs) constant while measuring the change in the level of outputs (inputs).

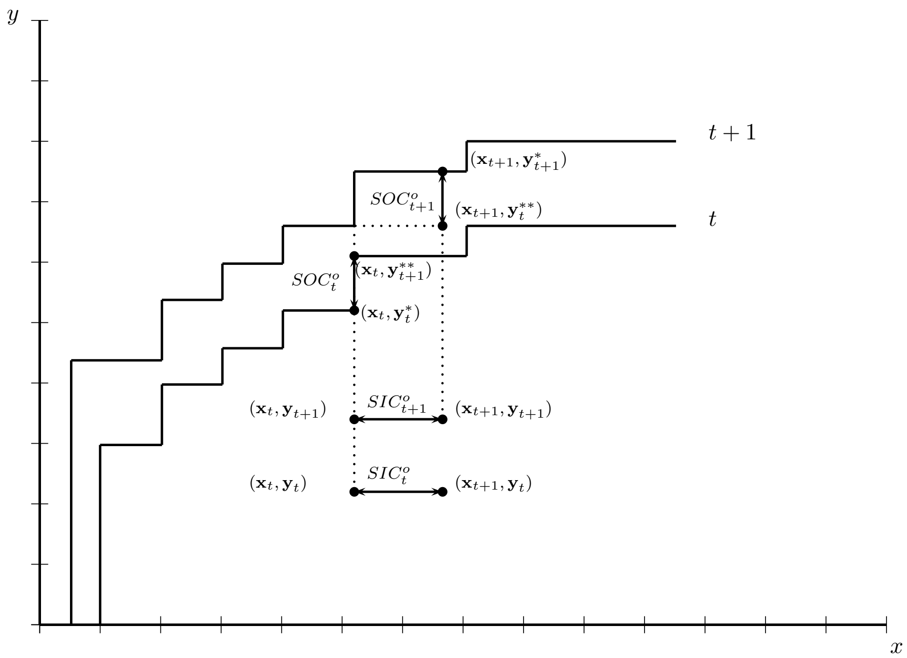


Figure 3.3: Scale inefficiency change.

This "residual" approach of Diewert and Fox (2017) differs from the traditional "Constant-Returns-to-Scale-Variable-Returns-to-Scale" (CRS-VRS) approach of Färe et al. (1994b) for the Malmquist index and Epure et al. (2011) for the Luenberger indicator.[11] The CRS-VRS approach compares the VRS frontier to a (hypothetical) benchmark CRS frontier to detect changes in returns to scale over time. In contrast, the residual approach directly considers changes in the frontier's gradient over time to assess scale inefficiency change. Thus, the main difference is that the Färe et al. (1994b) approach relies on two frontiers (VRS and CRS) to measure scale inefficiency change, while the residual approach of Diewert and Fox (2017) only uses one frontier (VRS in our case).

From a theoretical point of view, CRS is often not a realistic assumption whereby this hypothetical CRS frontier to measure changes in returns-to-scale is not appropriate. In contrast, the main strength of the residual approach is that we do not need to introduce

[11] Grifell-Tatjé and Lovell (1995) show that the Malmquist index allowing for VRS is biased when measuring productivity for a non-constant returns to scale technology. Bjurek (1996) points out that the Hicks-Moorsteen index does not suffer from this shortcoming. The reason is that the Malmquist index only measures changes in technology in either output or input orientation. Therefore, the output (input) Malmquist index does not pick up the effect of input (output) changes due to returns-to-scale which results in a bias under increasing or decreasing returns-to-scale.

a CRS component into the LHM TFP indicator to detect changes in returns-to-scale. If the technology exhibits CRS then this will be automatically reflected in zero values for the $\Delta SEC^o_{t,t+1}$ component even if we use a VRS approximation. Of course, depending on the application at hand and results of a preliminary test on returns-to-scale, the LHM TFP indicator and our decomposition can also be computed under other returns-to-scale assumptions. From a practical point of view, an obvious drawback to the "CRS-VRS" approach is that it is sensitive to outliers, because the CRS frontier can be spanned by a few (extreme) observations. This drawback can be reduced by using appropriate techniques such as order-m (Cazals et al., 2002) or order-α (Aragon et al., 2005).

The accuracy of the residual approach to approximate the gradient of the frontier depends on the "step-size", *i.e.*, the gap SIC^o_t and SIC^o_{t+1} between the frontier projections of $\mathbf{x}_t$ and $\mathbf{x}_{t+1}$ for the decomposition using output directions. The larger the step-size, the cruder the approximation.[12] Thus, a big change in inputs for a DMU from period t to period $t+1$ can give a cruder approximation of the frontier's gradient. This is the major disadvantage of the residual approach. Thus, both approaches have their own distinct advantages and disadvantages from a theoretical and a practical point of view. Future work might compare both approaches to understand in which situations each approach works well.

As a final remark, observe that both $\Delta T^o_{t,t+1}$ and $\Delta SEC^o_{t,t+1}$ are the arithmetic average of a Laspeyres (using base period t) and a Paasche (using base period $t+1$) type indicator.

3.4 Empirical application: U.S. agriculture

There is considerable interest in measuring and decomposing TFP of the economy in general and its different sectors. Measurement of TFP in agriculture, in particular, is important to assess our continuing ability to feed an increasing world population on a fixed amount of natural resources (such as land). Thus, policy makers are interested to learn about the underlying drivers of TFP growth and the contribution of output changes and input changes on TFP. The US department of agriculture (USDA) provides state level panel data on agricultural inputs and outputs alongside its estimates of productivity growth. These data are publicly available which also partially explains its popularity among previous research. Previous studies of TFP growth on the USDA data include Luh and Stefanou (1991), Zofío and Lovell (2001), Ball et al. (2004) and O'Donnell (2012b) among others. We investigate LHM TFP growth of U.S. agriculture across 48 states.[13] We use our newly developed LHM decomposition to determine the main drivers of productivity growth. Specifically, we investigate the extent to which LHM TFP growth

[12] This step-size is analogous to h in the commonly used definition of a derivative of a function f: $f'(x) = \lim_{h\to 0} \frac{f(x+h)-f(x)}{h}$. The more h approaches zero, the better the approximation of the derivative at the evaluated point. Likewise, the smaller SIC^o_t and SIC^o_{t+1}, the better the approximation of the frontier's gradient.

[13] The dataset does not include data from Alaska and Hawaii.

is driven by output growth and input growth, on the one hand, and technical change, technical inefficiency change and scale inefficiency change, on the other hand.

3.4.1 Data description

We use U.S. state-level agricultural panel data compiled by the U.S. Department of Agriculture (USDA). The data ranges from 1960 to 2004 and includes prices and quantities for 3 outputs (crops, livestock and other) and 4 inputs (land, intermediate, capital and labor). Table 3.1 contains mean values and the coefficient of variation per subperiod of 11 years. A full description of the data can be found in USDA (2016). The summary statistics suggest that aggregate production has substantially increased. Aggregate use of land, labor and to a lesser extent capital have decreased, while aggregate intermediate input use has increased. The low coefficient of variation of land use reveals that this production factor cannot be adjusted instantaneously.

	Period		Pacific	Mountain	Northern Plains	Southern Plains	Corn Belt	Southeast	Northeast	Lake States	Appalachian	Delta States
	1960/71	Mean	3300931.169	7255411.046	5840331.457	6292994.186	5004879.143	2215366.354	1768427.738	2444184.002	2356925.337	1520081.600
		CV	0.041	0.039	0.007	0.010	0.007	0.070	0.107	0.043	0.056	0.029
	1971/82	Mean	2893980.439	6720070.751	5687708.224	5781558.555	4767466.397	1773007.009	1394621.216	2199276.592	1956513.091	1343990.332
Land		CV	0.015	0.007	0.011	0.017	0.009	0.027	0.014	0.014	0.021	0.016
	1982/93	Mean	2710142.283	6508197.735	5544756.703	5535859.150	4632340.008	1484494.228	1267039.795	2112296.486	1818028.473	1206858.098
		CV	0.029	0.011	0.004	0.004	0.011	0.067	0.055	0.024	0.030	0.039
	1993/04	Mean	2590402.638	5967026.474	5581123.690	5718432.955	4582192.192	1428024.529	1190460.233	2085066.363	1802684.458	1209153.491
		CV	0.016	0.035	0.005	0.011	0.007	0.020	0.021	0.009	0.013	0.018
	1960/71	Mean	6388062.631	5374177.714	9020391.093	5551339.202	16939475.421	4322275.998	5303636.777	8170759.798	4824174.845	3346989.032
		CV	0.047	0.125	0.092	0.126	0.048	0.119	0.034	0.027	0.064	0.125
	1971/82	Mean	7546312.374	7223653.517	11783582.906	7805959.230	18720865.969	5445467.048	5643208.149	9328635.249	5770043.712	4124055.132
Intermediate		CV	0.111	0.091	0.128	0.094	0.074	0.112	0.090	0.112	0.093	0.086
	1982/93	Mean	8462752.863	6899795.667	12869397.188	7703305.897	16985319.256	5663158.403	5712198.784	9970093.780	5939957.239	4740621.090
		CV	0.058	0.029	0.032	0.070	0.056	0.037	0.026	0.057	0.022	0.130
	1993/04	Mean	11556159.900	8073898.489	14492935.479	8891404.135	17845564.638	6924523.321	6077658.726	11068114.621	7896907.296	6121353.879
		CV	0.082	0.070	0.078	0.064	0.047	0.066	0.054	0.059	0.119	0.031
	1960/71	Mean	2277278.160	1965743.390	3839212.101	2420531.367	7485812.020	1409850.653	2761049.666	4235861.165	2550447.895	1202072.627
		CV	0.020	0.058	0.054	0.056	0.084	0.070	0.021	0.030	0.067	0.115
	1971/82	Mean	2583856.991	2443132.645	4662093.957	3052573.807	9993671.366	1854854.361	3040945.252	4891432.936	3233950.577	1718949.272
Capital		CV	0.079	0.084	0.072	0.085	0.095	0.103	0.068	0.075	0.086	0.106
	1982/93	Mean	2340101.255	2292679.329	4213514.172	2901421.560	8602936.055	1665912.719	2711313.433	4626726.086	2817472.143	1619430.276
		CV	0.128	0.114	0.108	0.108	0.159	0.147	0.124	0.131	0.141	0.144
	1993/04	Mean	1983069.814	1936337.760	3376963.177	2316924.548	6070538.603	1370479.997	1983837.703	3562676.158	2361032.305	1264252.022
		CV	0.029	0.016	0.026	0.032	0.056	0.023	0.052	0.040	0.016	0.020
	1960/71	Mean	11826308.223	7251927.972	11451651.884	10457257.796	26640405.728	8806948.892	12416331.593	17517663.881	16278275.980	8020645.246
		CV	0.124	0.089	0.134	0.141	0.164	0.117	0.182	0.141	0.171	0.192
	1971/82	Mean	10262405.279	6501263.073	10321361.426	7673646.814	20171125.527	6484264.056	9595326.458	13841201.345	9745925.952	4745991.646
Labor		CV	0.059	0.033	0.062	0.096	0.066	0.101	0.040	0.030	0.137	0.149
	1982/93	Mean	9271807.049	6202223.028	9061470.223	6519170.104	16138668.519	4774474.559	8008139.689	12006149.317	7133646.807	3299937.462
		CV	0.064	0.088	0.117	0.051	0.095	0.074	0.136	0.126	0.163	0.090
	1993/04	Mean	10210352.950	5202355.190	7081823.093	7023374.727	11939104.652	4363279.813	6389996.042	7731857.592	6268745.715	2921548.411
		CV	0.083	0.043	0.056	0.047	0.097	0.042	0.052	0.136	0.050	0.059
	1960/71	Mean	9125685.221	4287806.931	7041572.037	4374994.511	14332254.967	4522798.287	4261514.605	5899717.197	5730387.079	3013945.475
		CV	0.080	0.081	0.111	0.061	0.087	0.054	0.036	0.063	0.050	0.093
	1971/82	Mean	13188542.391	5597955.159	10329731.388	5405400.050	20625602.933	6242424.518	4699512.097	8385250.909	6594263.573	3872567.263
Crops		CV	0.157	0.115	0.143	0.169	0.142	0.114	0.084	0.179	0.085	0.134
	1982/93	Mean	17643771.954	6701908.463	12996711.723	5879907.506	22994725.677	7229992.378	5461209.103	10186991.736	6992866.349	4664010.212
		CV	0.084	0.061	0.138	0.089	0.170	0.056	0.048	0.134	0.120	0.131
	1993/04	Mean	23286483.981	7886660.357	16120799.908	6506193.595	27046336.520	8590499.524	5693994.670	12044270.033	7794004.214	5409819.779
		CV	0.068	0.049	0.120	0.090	0.091	0.046	0.041	0.108	0.054	0.110
	1960/71	Mean	5327584.310	4919007.684	7454614.625	5173132.409	17052320.948	4112320.193	6257852.014	9324537.634	4515637.364	2939895.358
		CV	0.057	0.129	0.098	0.101	0.026	0.158	0.014	0.030	0.063	0.165
	1971/82	Mean	6152491.657	6380074.865	9212170.296	7281860.962	15516508.437	5433833.885	6445471.439	9292008.630	5271046.811	3749067.268
Livestock		CV	0.044	0.034	0.043	0.031	0.047	0.054	0.073	0.053	0.069	0.028
	1982/93	Mean	7493949.676	6568411.982	10365093.348	7989950.726	14086364.477	6324043.452	7531591.704	10371877.289	6873033.353	4410247.573
		CV	0.082	0.043	0.050	0.063	0.025	0.071	0.025	0.021	0.077	0.122
	1993/04	Mean	9685007.165	8609192.715	11405247.563	9848376.630	14512438.991	8009618.671	8382637.716	10506374.382	9332273.382	6095279.863
		CV	0.076	0.098	0.036	0.041	0.034	0.060	0.028	0.033	0.052	0.048
	1960/71	Mean	1505214.023	754115.275	763723.724	1373528.695	973630.348	1109452.294	542393.806	574042.980	631465.964	510886.261
		CV	0.072	0.074	0.111	0.261	0.085	0.066	0.202	0.186	0.154	0.123
	1971/82	Mean	1287921.873	565772.514	619433.118	715874.669	779303.889	807393.456	366832.739	430656.260	433338.235	391387.399
Other		CV	0.062	0.101	0.163	0.123	0.104	0.086	0.089	0.095	0.059	0.065
	1982/93	Mean	1666910.125	968193.737	1430545.782	1362035.393	843864.319	869125.637	491187.178	646790.108	588487.411	619379.070
		CV	0.121	0.141	0.136	0.175	0.171	0.193	0.091	0.126	0.295	0.408
	1993/04	Mean	2550514.545	1351626.269	1934947.718	1728703.994	1218992.400	1394982.468	667759.874	904455.293	1049200.724	992821.059
		CV	0.178	0.142	0.134	0.135	0.106	0.230	0.209	0.181	0.216	0.128

Table 3.1: Mean and coefficient of variation (CV) for quantities per subperiod in 1996 US dollars ($\times 10^3$).

The USDA identifies 10 regions of agricultural production in the U.S. An overview is provided in Table 3.2.

Region	States
Pacific	CA, OR, WA
Mountain	AZ, CO, ID, MT, NM, NV, UT, WY
Northern Plains	KS, ND, NE, SD
Southern Plains	OK, TX
Corn Belt	IA, IL, IN, MO, OH
Southeast	AL, FL, GA, SC
Northeast	CT, DE, MA, MD, ME, NH, NJ, NY, PA, RI, VT
Lake States	MI, MN, WI
Appalachian	KY, NC, TN, VA, WV
Delta States	AR, LA, MS

Table 3.2: Regions of agricultural production.

We compute LHM TFP growth and its output-oriented decomposition for every state over the selected time period. We compare across all 48 states when computing the necessary distance functions and thus assume that all states have access to a similar production technology. This is also the approach of Zofío and Lovell (2001) and Ball et al. (2010). Alternatively, we could compare states within the same agricultural region (see Table 3.2). However, this would limit the set of observations to 2 or 3 for some regions, which may be insufficient.[14]

We first conduct the analysis for a non-convex technology (using Free Disposal Hull under a variable-returns-to-scale assumption) and then repeat the analysis for a convex technology (using Data Envelopment Analysis under a variable-returns-to-scale assumption). This shows the applicability of our decomposition for both technologies and highlights potential differences that can arise due to convexity assumptions of the production technology.

3.4.2 Non-convex technology

In practice, $\boldsymbol{\mathcal{Y}}_t$ is unknown and needs to be estimated from the K observations in the dataset. The smallest enveloping non-convex approximation under variable-returns-to-scale (VRS) is given by:

$$\hat{\boldsymbol{\mathcal{Y}}}_t = \left\{ (\mathbf{x}_{0t}, \mathbf{y}_{0t}) | \sum_{k=1}^{K} \lambda_k \mathbf{x}_{kt} \leq \mathbf{x}_{0t}, \sum_{k=1}^{K} \lambda_k \mathbf{y}_{kt} \geq \mathbf{y}_{0t}, \sum_{k=1}^{K} \lambda_k = 1, \lambda_k \in \{0,1\} \right\}, \tag{3.13}$$

and can be plugged in (3.1) to compute the directional distance function in practice. The resulting program is a mixed-integer program and can be computationally harder to solve

[14] O'Donnell (2012b) applies window analysis to circumvent this problem, but uses rather large windows for some regions. This can dampen the estimated rates of technical change.

than the usual linear program. As first pointed out by Tulkens (1993), there exists an equivalent formulation based on enumeration which is considerably easier to solve. The enumeration formulation for directional distance functions with $\mathbf{g}_t > 0$ proposed by Cherchye et al. (2001) is:

$$D_b(\mathbf{x}_{0a}, \mathbf{y}_{0a}; \mathbf{g}_a) = \max_{k \in \{1,\dots,K\}} \left\{ \min_{\substack{j \in \{1,\dots,m\}, \\ v \in \{1,\dots,n\}}} \left\{ \frac{Y^j_{kb} - Y^j_{0a}}{g^{oj}_a}, \frac{X^v_{0a} - X^v_{kb}}{g^{iv}_a} \right\} \right\}, \tag{3.14}$$

with $(a, b) \in \{t, t+1\} \times \{t, t+1\}$. This allows us to compute all distance functions needed for the LHM TFP indicator and its decomposition. In line with the literature, we choose $\mathbf{g}^i_a = \mathbf{x}_{0a}$ and $\mathbf{g}^o_a = \mathbf{y}_{0a}$ such that β can be interpreted as the maximum proportional expansion (contraction) in the output (input) direction.[15] Since we work with aggregate data, all of our chosen directional vectors are nonzero. Moreover, the data set only contains nonnegative outputs $\mathbf{y}_t \in \mathbb{R}^m_+$. As a result, we can use the simplified formula (3.14).[16]

Main findings for the U.S

We first present the results for the U.S. as a whole before presenting individual results for the agricultural regions. We first consider the average LHM TFP change in Figure 3.4. This is computed in a given year by taking the average LHM TFP of all states. This figure shows several considerable LHM TFP changes over time. Until 1979 − 1980, bad years offset good years resulting in only marginal cumulative LHM TFP growth over this period. After this period, positive growth rates dominate negative growth rates resulting in a positive cumulative LHM TFP growth of 78.61% in 2004. This boils down to an average LHM TFP growth of 1.79% per year.

Figure 3.5 also shows the underlying drivers of these trends. Up to 1979 − 1980, cumulative LHM TFP growth is driven by $LI_{t,t+1}$. Subsequently, both input decline and output growth contribute to substantial LHM TFP growth. Cumulative output growth is 44.10%, while cumulative input decline is 34.51%. This means that U.S. agricultural production simultaneously increases output production at an average rate of 1% per year while decreasing input use at an average rate of 0.78% per year.

We now turn to our LHM TFP decomposition. Technical progress is the main driver of LHM TFP growth which is partly offset by scale inefficiency growth. Over the entire period, technical progress increased with 139.57% on average while cumulative scale inefficiency change reached −60.63%. Technical inefficiency change plays virtually no role. Table 3.3 summarizes these results and also lists the minimal and maximal values

[15] This choice of the direction vector takes into account state heterogeneity and projects each observation in a different direction onto the frontier. Recently, more advanced data-driven approaches were developed that determine the direction vectors using the analyzed firm's configuration (see Daraio and Simar (2016) for technical details and Epure (2016) for a management-oriented discussion). Finally, a homogeneous direction vector is more desirable, for example, for regulators in sectors where heterogeneity in input-output configurations is low.

[16] We use Bogetoft and Otto (2015)'s Benchmarking package in R to compute the distance functions.

of the LHM TFP indicator and its components per subperiod of 11 years. It also lists the corresponding states.

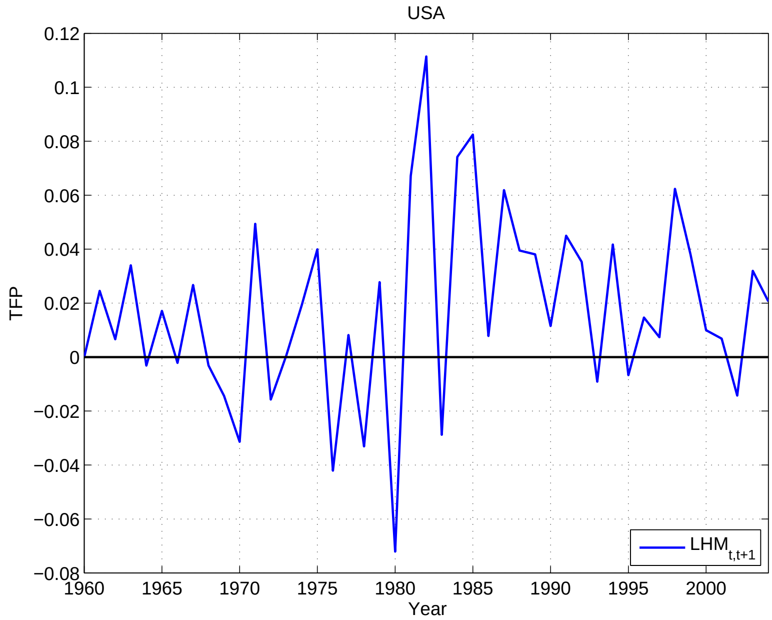


Figure 3.4: Mean TFP change in the U.S. using a non-convex technology.

		$LHM_{t,t+1}$	$LO_{t,t+1}$	$LI_{t,t+1}$	$\Delta T^o_{t,t+1}$	$\Delta TEI^o_{t,t+1}$	$\Delta SEC^o_{t,t+1}$
$\sum_{t=1960}^{2004}$ mean(states)		78.61	44.10	-34.51	139.57	-0.32	-60.63
Avg growth rate		1.79	1.00	-0.78	3.17	-7.30 $\times 10^{-3}$	-1.38
min	1960/71	-5.72 (OK)	-4.70 (OK)	-6.61 (RI)	-8.77 (FL)	-1.45 (OK)	-30.20 (RI)
	1971/82	-1.99 (WY)	-1.18 (IN)	-2.46 (SC)	-13.81 (AZ)	-0.57 (PA)	-7.95 (DE)
	1982/93	0.30 (FL)	-2.02 (SD)	-4.76 (NH)	-1.22 (AR)	-0.28 (MO)	-14.44 (NH)
	1993/04	-1.03 (VT)	-0.60 (WY)	-3.11 (MA)	-3.38 (AL)	-1.40 (OK)	-10.13 (DE)
max	1960/71	3.29 (RI)	2.99 (NV)	2.02 (CO)	33.49 (RI)	0.00 (all but OK)	9.68 (FL)
	1971/82	3.90 (OK)	4.12 (NE)	2.98 (ID)	7.64 (NH)	1.45 (OK)	14.60 (AZ)
	1982/93	7.45 (UT)	4.13 (AR)	1.42 (OK)	17.79 (NH)	0.57 (PA)	6.81 (UT)
	1993/04	5.34 (MA)	5.42 (SC)	1.99 (TN)	13.76 (DE)	0.28 (MO)	7.61 (AL)

Table 3.3: LHM TFP growth and its components in the U.S. over 1960 – 2004 (in %) using a non-convex technology.

Main findings per region

Figure 3.6 depicts the average cumulative LHM TFP and its components for every region over time. The mean is computed with respect to all states in that particular agricultural production region. The highest cumulative TFP growth is achieved by the Northeast, Southeast, Corn Belt and Delta States with 84.47% – 95.62%. They are followed by the Pacific, Northern Plains, Appalachian, Lake States and Mountain

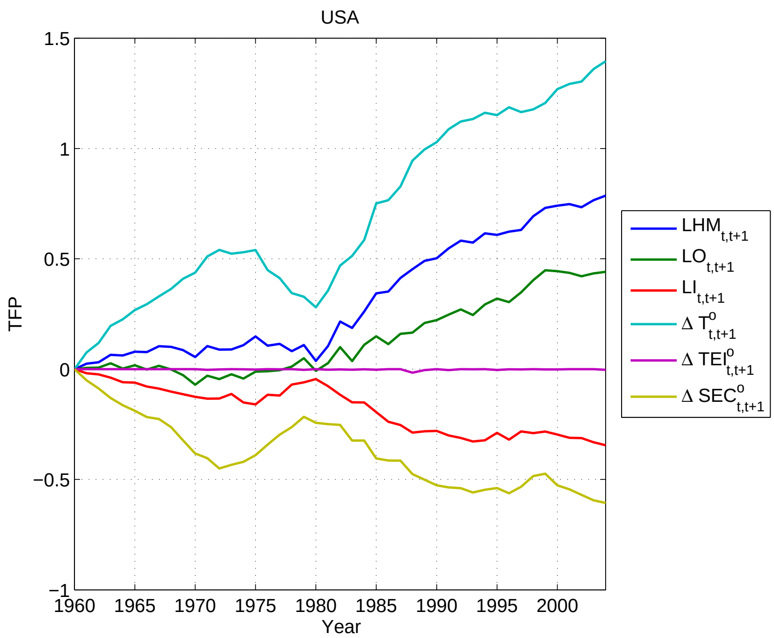


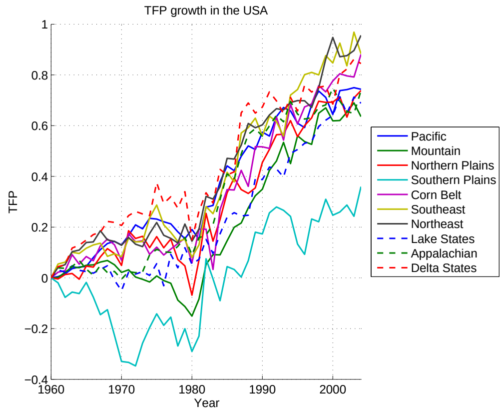
Figure 3.5: Mean cumulative TFP growth in the U.S. and it components using a non-convex technology.

regions with 63.56% − 74.36%. Finally, the Southern Plains region is severely behind the other regions with a cumulative TFP growth of 35.95%.

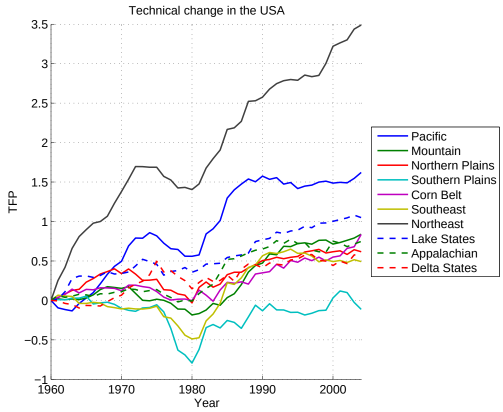
Although almost all regions experience technical progress, there are diverging trends among the different regions. Positive (negative) cumulative technical change over the whole time period indicates progress (regress) in terms of production technology. The Northeast experienced the largest cumulative technical progress (349%). The Pacific region is second with 162.2% and the Lake States are third with 104.9%. The Mountain, Corn Belt, Appalachian, Northern Plains, Delta States and Southeast experience milder technical progress between 49.23% − 83.65%. The Southern Plains is the only region with a cumulative technical regress of 11.42%, mainly due to a severe dip in the period 1975 − 1980 from which it only slowly recovers.

Technical inefficiency change generally plays a minor role. Positive (negative) cumulative technical inefficiency change indicates that the distance to the frontier decreases (increases) over the whole time period. Negative changes in cumulative technical inefficiency change are quickly followed by positive changes. These spikes are visible in the Southern Plains, Northern Plains, Corn Belt, Delta States and Lake States. There is only a negative cumulative technical inefficiency change in the Southern Plains, due to a drop in technical inefficiency by 7.71% in 2004.

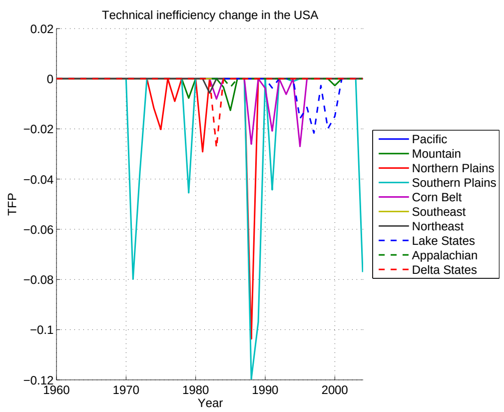
The trend in the scale inefficiency change is the mirror image of the trend in technical change: regions with positive (negative) technical change experience negative (positive) scale inefficiency. Positive (negative) cumulative scale inefficiency change indicates that the region operates at a more (less) optimal scale over the whole time period. The Southern Plains, Southeast, Delta States, Northern Plains, Corn Belt experience the highest positive cumulative scale inefficiency change between $4.48\% - 55.07\%$. Cumulative scale inefficiency change is negative in the Appalachian, Mountain and Lake States (between -1.00% and -36.05%). Cumulative scale inefficiency change is most negative in the Pacific (-87.82%) and Northeast (-253.4%) regions.



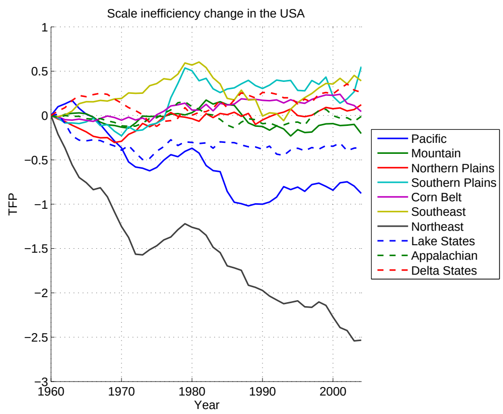
(a) Average cumulative TFP growth per region.



(b) Average cumulative technical change per region.



(c) Average cumulative technical inefficiency change per region.



(d) Average cumulative scale inefficiency change per region.

Figure 3.6: Cumulative TFP growth and its decomposition per U.S. region using a non-convex technology.

3.4.3 Convex technology

Since we only have 48 observations per year, a non-convex technology might provide limited discriminating power resulting in many efficient observations. Therefore, we repeat the analysis for a convex VRS representation of the production technology using

Data Envelopment Analysis (DEA). The smallest enveloping approximation is given by:

$$\hat{\boldsymbol{\mathcal{Y}}}_t = \left\{ (\mathbf{x}_{0t}, \mathbf{y}_{0t}) | \sum_{k=1}^{K} \lambda_k \mathbf{x}_{kt} \leq \mathbf{x}_{0t}, \sum_{k=1}^{K} \lambda_k \mathbf{y}_{kt} \geq \mathbf{y}_{0t}, \sum_{k=1}^{K} \lambda_k = 1, \lambda_k \geq 0 \right\}, \tag{3.15}$$

and can be plugged in (3.1) to compute the directional distance function in practice. The resulting linear program with $(a, b) \in \{t, t+1\} \times \{t, t+1\}$ is:

$$\begin{aligned} D_b(\mathbf{x}_{0a}, \mathbf{y}_{0a}; \mathbf{g}_a) = \max_{\beta, \lambda_k \geq 0} \beta \text{ s.t.} & \sum_{k=1}^{K} \lambda_k \mathbf{x}_{kb} \leq \mathbf{x}_{0a} - \beta \mathbf{g}_a^i, \\ & \sum_{k=1}^{K} \lambda_k \mathbf{y}_{kb} \geq \mathbf{y}_{0a} + \beta \mathbf{g}_a^o, \\ & \sum_{k=1}^{K} \lambda_k = 1. \end{aligned} \tag{3.16}$$

This allows us to compute all necessary distance functions needed for all the components of the LHM TFP indicator. As for the FDH analysis, we choose $\mathbf{g}_a^i = \mathbf{x}_{0a}$ and $\mathbf{g}_a^o = \mathbf{y}_{0a}$.

Main findings for the U.S

We present the results for the U.S. as a whole before presenting individual results for the agricultural regions.[17] We first consider the average annual LHM TFP change in Figure 3.7. This is computed in a given year by taking the average LHM TFP of all states. This figure shows considerable fluctuations in annual LHM TFP changes over time. Overall, years with LHM TFP growth dominate years with LHM TFP decline.

Figure 3.8 shows the cumulative LHM TFP growth and the underlying drivers. Our main finding for the U.S. as a whole is that LHM TFP clearly increases over time. The LHM TFP indicator increases by 70.46% between 1960 and 2004. This boils down to an average LHM TFP growth of 1.60% per year. LHM TFP growth is driven by output growth (+62.98%) rather than input decline (−7.47%). In the period 1977−1982, $LI_{t,t+1}$ contributes to a temporary slowdown in LHM TFP growth. $LI_{t,t+1}$ only plays a minor role in the remaining periods.

We now turn to our LHM decomposition. Our decomposition shows that technical change (+70.55%) is the main driver, while technical inefficiency change (−1.99%) and scale inefficiency change (+0.42%) only play a minor role. Table 3.4 summarizes these results and lists the minimal and maximal values of the LHM TFP indicator and its components per subperiod of 11 years. It also lists the corresponding states.

[17] Infeasibilities may arise for the components where the year of the observation differs from the year of the reference technology. As there is no easy solution to solve this problem, Briec and Kerstens (2009) recommend to report the infeasibilities. There were only infeasibilities in the computation of $\Delta T^o_{t,t+1}$ and $\Delta SEC^o_{t,t+1}$ for Rhode Island.

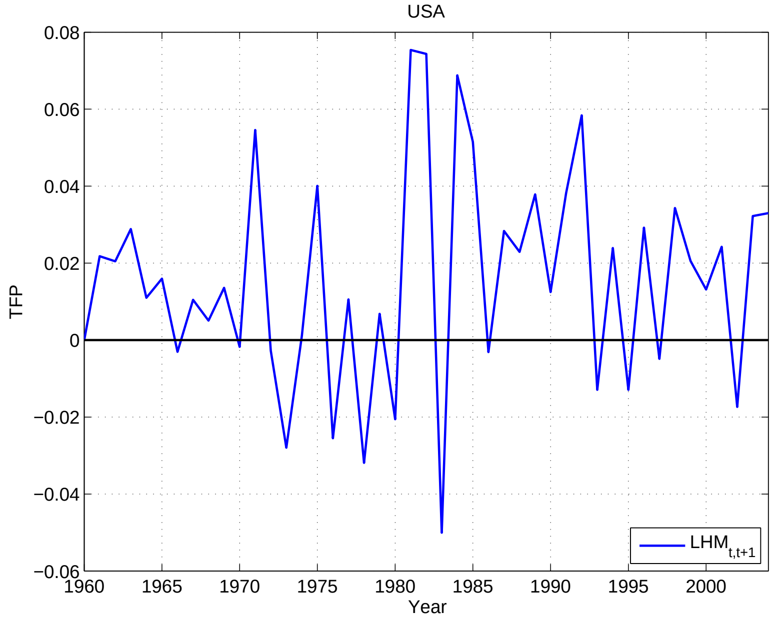


Figure 3.7: Mean TFP change in the U.S. using a convex technology.

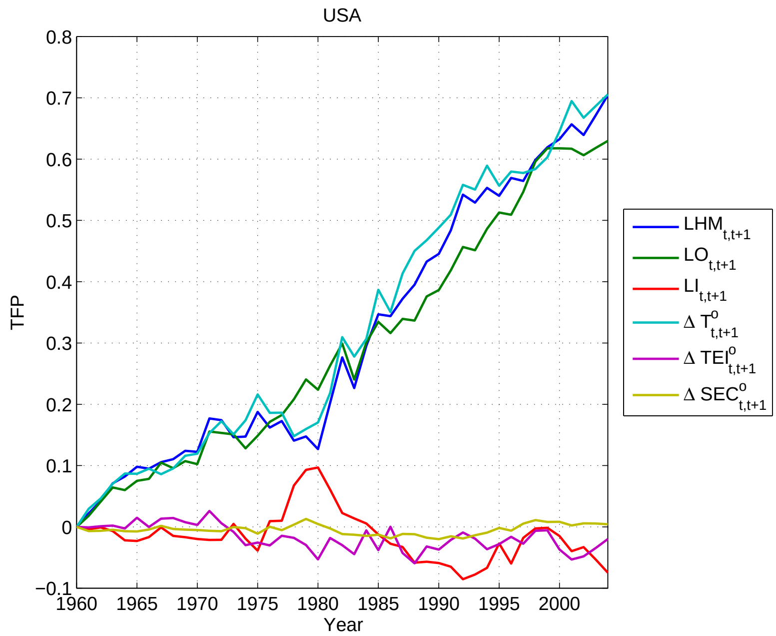


Figure 3.8: Mean cumulative TFP growth and its components in the U.S. using a convex technology.

		$LHM_{t,t+1}$	$LO_{t,t+1}$	$LI_{t,t+1}$	$\Delta T^o_{t,t+1}$	$\Delta TEI^o_{t,t+1}$	$\Delta SEC^o_{t,t+1}$
$\sum_{t=1960}^{2004}$ mean(states)		70.46	62.98	-7.47	70.55	-1.99	0.42
Avg growth rate		1.60	1.43	-0.17	1.60	-0.05	9.50 $\times 10^{-3}$
min	1960/71	-4.37 (OK)	-3.84 (NJ)	-6.61 (RI)	-3.83 (OK)	-1.60 (OK)	-3.32 (NV)
	1971/82	-1.28 (WV)	-1.55 (MO)	-1.84 (RI)	-0.67 (FL)	-2.92 (WY)	-1.65 (DE)
	1982/93	-0.36 (TN)	-1.53 (NH)	-3.52 (KS)	0.64 (FL)	-2.64 (MO)	-1.63 (SD)
	1993/04	-2.41 (WY)	-2.34 (WY)	-2.70 (RI)	-1.34 (KY)	-3.65 (WY)	-0.93 (LA)
max	1960/71	7.16 (ND)	7.21 (AR)	3.18 (AR)	5.42 (NV)	3.72 (ND)	1.76 (LA)
	1971/82	3.77 (IL)	3.51 (WA)	2.63 (DE)	4.05 (ND)	1.59 (OK)	1.33 (OR)
	1982/93	5.64 (DE)	4.83 (WV)	1.99 (OK)	5.67 (DE)	3.62 (MT)	2.03 (IA)
	1993/04	4.24 (AL)	4.11 (SD)	2.53 (KY)	4.38 (MS)	2.89 (MO)	2.54 (TN)

Table 3.4: TFP growth and its components in the U.S. covering the years $1960 - 2004$ (in %) using a convex technology.

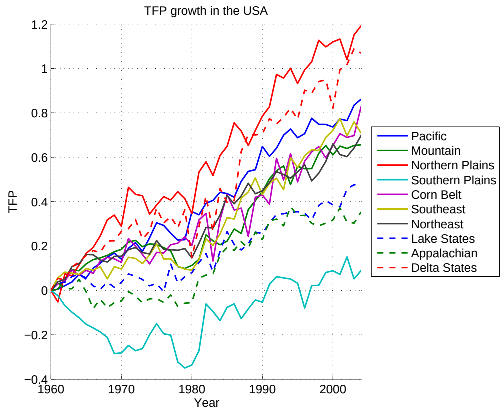
Main findings per region

Figure 3.9 depicts the mean cumulative LHM TFP and its components for every region over time. The mean is computed with respect to all states in that particular agricultural production region. The Northern Plains experienced the highest cumulative LHM TFP growth (119.3%) while the Southern Plains experienced the lowest cumulative LHM TFP growth (8.96%) over the entire period. Between them, Delta States experience the second highest cumulative LHM TFP growth of 107%. Pacific, Corn Belt, Southeast, Northeast and Mountain regions have similar levels of cumulative LHM TFP growth of $65.58\% - 86.18\%$. The cumulative LHM TFP growth of Lake States and Appalachian regions varies in the range $35.36\% - 46.83\%$.

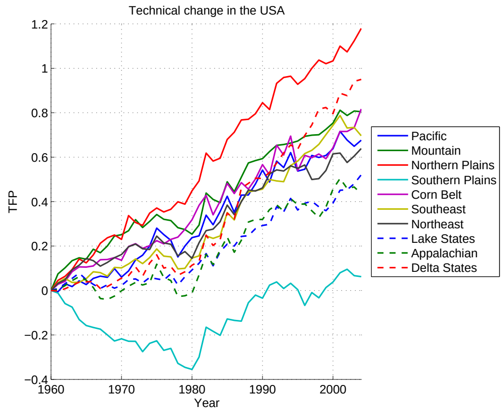
Being the main driver of LHM TFP growth, similar trends occur for technical change. The Northern Plains region has the highest rate of cumulative technical change (117.9%) and the Southern Plains the lowest (6.25%). Again, Delta States experience the second highest rate of technical change of 95%. The other regions can roughly be classified in two clusters. The first cluster consists of the Corn Belt, Mountain, Southeast, Pacific and Northeast regions ($63.89\% - 81.67\%$). The second cluster consists of Lake States and Appalachian ($43.6\% - 52.09\%$).

In terms of cumulative technical inefficiency change, there are diverging trends among the different regions. Pacific, Northern Plains, Delta States and Northeast experience a positive cumulative technical inefficiency change between 4.86% and 10.82%. The six remaining regions experience a negative cumulative technical inefficiency change. Cumulative technical inefficiency change is mildly negative (between -7.01% and -1.58%) in the Southeast, Corn Belt, Lake States and Mountain regions. This is worse in the Southern Plains and Appalachian, where the cumulative technical inefficiency change is -12.96% and -23.61%, respectively.

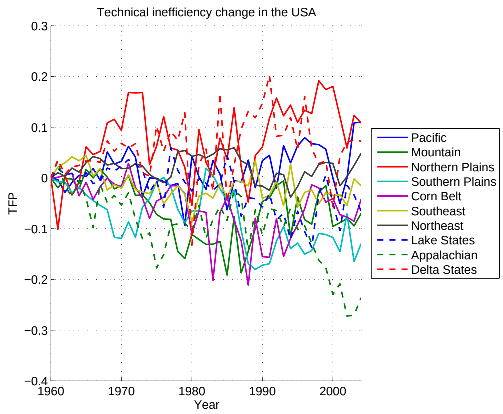
Again, there are diverging trends for cumulative scale inefficiency change. The Southern Plains experience the highest increase in cumulative scale inefficiency change (15.67%) followed closely by the Appalachian region (15.38%). The Northern Plains experience a negative cumulative scale inefficiency change (-9.5%). Between these extremes, the Pacific, Corn Belt, Delta States, Southeast and Lake States have a positive cumulative scale inefficiency change in the range of 1.29% – 7.56%. In contrast, the cumulative scale inefficiency change of the Northeast and Mountain regions is negative (−6.64% and −7.88%, respectively).



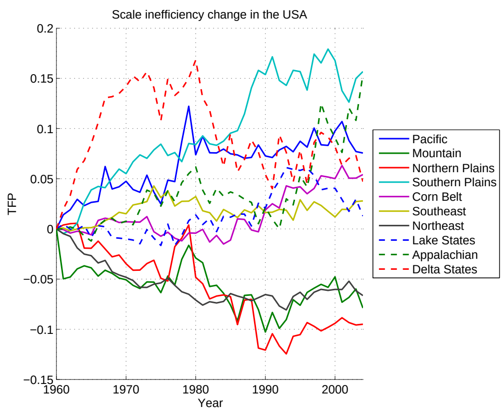
(a) Average cumulative TFP growth per region.



(b) Average cumulative technical change per region.



(c) Average cumulative technical inefficiency change per region.



(d) Average cumulative scale inefficiency change per region.

Figure 3.9: Cumulative TFP growth and its decomposition per U.S. region using a convex technology.

Although all U.S. regions experienced LHM TFP growth in the period 1960 – 2004, this analysis shows that the contribution of the underlying factors varies considerably per region. Technical change is the main driver of LHM TFP growth for all U.S. regions. In addition, several U.S. regions partly increased TFP by becoming more efficient over time and/or operating at a more optimal scale. Other regions mainly relied on technical change to increase LHM TFP.

3.4.4 Discussion

The results depend on the convexity assumption of the technology. We test the hypothesis whether the distributions of the LHM TFP indicator and its components for FDH and DEA are not significantly different using a Kolmogorov-Smirnov test. This non-parametrically tests the hypothesis H_0 whether two samples are drawn from the same underlying distribution. We conduct the test for every year separately, resulting in 44 different test hypotheses for every component. The results at the 10% significance level are presented in Table 3.5. For the majority of years, the distributions of the $LHM_{t,t+1}$ and its components $LO_{t,t+1}$ and $LI_{t,t+1}$ are not statistically different using FDH and DEA. In contrast, the distributions of $\Delta T^o_{t,t+1}$ are statistically significant for a majority of years and the distributions of $\Delta TEI^o_{t,t+1}$ and $\Delta SEC^o_{t,t+1}$ under both technologies are significantly different for all years. These results in conjunction with Table 3.3 and Table 3.4 lead us to the following qualitative conclusions.

	$LHM_{t,t+1}$	$LO_{t,t+1}$	$LI_{t,t+1}$	$\Delta T^o_{t,t+1}$	$\Delta TEI^o_{t,t+1}$	$\Delta SEC^o_{t,t+1}$
Reject H_0 per year at 10%	9/44	12/44	8/44	25/44	44/44	44/44
Reject H_0 at 10%	No	Yes	Yes	Yes	Yes	Yes

Table 3.5: Results of Kolmogorov-Smirnov test for distributions under non-convex and convex technologies.

Both results suggest there is substantial LHM TFP growth over the entire period which is mainly driven by technical progress. Both the DEA and FDH results indicate that output growth dominates input decline, although this finding is much more pronounced for the DEA results. A possible explanation for the smaller contribution of input decline is that some quasi-fixed inputs (*e.g.*, land) are not constantly adjusted over time or that input reduction is not an objective for some inputs such as land and labor.

We analyze LHM TFP growth and technical change across time, farm types and agricultural intensity rates. Table 3.6 shows the results of the Kolmogorov-Smirnov test testing equality of distributions for LHM TFP growth rates and technical changes for consecutive subperiods of eleven years in line with Table 3.3. Regarding FDH, all distributions of consecutive LHM TFP growth rates and technical changes are significantly different at the 10% level. Regarding DEA, the distributions of the LHM TFP growth rates and technical changes between 1982/93 and 1993/04 are not significantly different at the 10% level, while these are significantly different comparing the preceding time periods. This suggests that distributional differences in productivity growth driven by shifts in technology may decrease in importance throughout time.

In line with Ang and Kerstens (2016), we assess whether there are distributional differences in LHM TFP growth rates and technical changes between farm types in Table 3.7. We rank the farm regions by the ratio of crop production to total production considering the whole time period. This leads to a classification of 3 crop regions (Corn Belt, Northern Plains and Pacific), 4 mixed regions (Delta States, Southeast, Appalachian, and Northeast) and 3 livestock regions (Lake States, Mountain area and Southern

	1960/71 – 1971/82	1971/82 – 1982/93	1982/93 – 1993/04
FDH: Reject H_0 at 10% $LHM_{t,t+1}$	Yes	Yes	Yes
FDH: Reject H_0 at 10% $\Delta T_{t,t+1}$	Yes	Yes	Yes
DEA: Reject H_0 at 10% $LHM_{t,t+1}$	Yes	Yes	No
DEA: Reject H_0 at 10% $\Delta T_{t,t+1}$	Yes	Yes	No

Table 3.6: Results of Kolmogorov-Smirnov test for distributions of LHM TFP growth rates and technical changes for consecutive time periods.

Plains). Regarding the FDH results, the distributions of the LHM TFP growth rates of mixed regions and livestock regions are significantly different at the 10% level, while these are not significantly different at the 10% level comparing crop regions to mixed regions and livestock regions. Interestingly, regarding the FDH results, the distributions of the technical changes are significantly different at the 10% level comparing all types of regions. Regarding the DEA results, the distributions of the LHM TFP growth rates and technical changes of crop regions and mixed regions, and mixed regions and livestock regions, are not significant at the 10% level, while these are significantly different at the 10% level comparing crop regions to livestock regions. In summary, there seems to be ambiguity in how regional differences in specialization may drive differences in the distribution of LHM TFP growth and technical changes.

	Crops - Mixed	Mixed - Livestock	Crops - Livestock
FDH: Reject H_0 at 10% $LHM_{t,t+1}$	No	Yes	No
FDH: Reject H_0 at 10% $\Delta T_{t,t+1}$	Yes	Yes	Yes
DEA: Reject H_0 at 10% $LHM_{t,t+1}$	No	No	Yes
DEA: Reject H_0 at 10% $\Delta T_{t,t+1}$	No	No	Yes

Table 3.7: Results of Kolmogorov-Smirnov test for distributions of LHM TFP growth rates and technical changes covering the whole time period among farm types.

We also assess whether there are distributional differences in LHM TFP growth rates and technical changes between agricultural intensity rates in Table 3.8. We rank the farm regions by the Industry Specialization Index (ISI) for agriculture considering 1963 – 2004.[18] The U.S. Bureau of Economic Analysis (BEA) computes the ISI as the agricultural industry's share of the state-level Gross Domestic Product divided by the agricultural industry's share of the U.S. total for the same statistic. The complete dataset can be found in BEA (2016). We rank the regions by ISI, which leads to a classification of 3 low ISI regions (Northeast, Lake States and Southeast), 4 medium ISI regions (Appalachian, Southern Plains, Pacific and Corn Belt) and 3 high ISI regions (Mountain area, Delta States and Northern Plains). With respect to the FDH results, comparing distributions of the LHM TFP growth rates for all groups do not yield any significant difference at the 10% level. The distributions of the technical changes are significantly different at the 10% level comparing low ISI regions to medium and high

[18]Data for 1960 – 1962 are unavailable.

ISI regions. Regarding the DEA results, the distributions of the LHM TFP growth rates and technical changes are significantly different at the 10% level comparing medium ISI regions to high ISI regions. Similar to the preceding section, there thus seems to be ambiguity in how regional differences in agricultural intensity may drive differences in the distribution of LHM TFP growth and technical changes.

	Low - Medium	Medium - High	Low - High
FDH: Reject H_0 at 10% $LHM_{t,t+1}$	No	No	No
FDH: Reject H_0 at 10% $\Delta T_{t,t+1}$	Yes	No	Yes
DEA: Reject H_0 at 10% $LHM_{t,t+1}$	No	Yes	No
DEA: Reject H_0 at 10% $\Delta T_{t,t+1}$	No	Yes	No

Table 3.8: Results of Kolmogorov-Smirnov test for distributions of LHM TFP growth rates and technical changes covering the whole time period among agricultural intensity rates.

The contribution of technical inefficiency change to LHM TFP growth is less clear-cut. Using FDH, technical inefficiency change is virtually nonexistent. Further inspection reveals that most (contemporaneous) technical inefficiency scores are zero using FDH. This drives the extremely low technical inefficiency change. Therefore, these remarkable results may be due to lower discriminatory power of FDH in this case since there are relatively few observations per year compared to the number of inputs and outputs. Using DEA, there is a small cumulative increase in technical inefficiency change.

The results differ more for the scale inefficiency change component. There is a substantial increase in cumulative scale inefficiency change using FDH, whereas there is almost no cumulative scale inefficiency change using DEA. Again, this may be due to the higher discriminatory power of DEA.

Our DEA results are in line with other empirical studies that analyze the TFP growth in the U.S. agricultural sector using the same data source. Zofío and Lovell (2001), Ball et al. (2010), O'Donnell (2012b) and Ball et al. (2016) also find substantial TFP growth.[19] It is driven by technical progress rather than efficiency change in line with Zofío and Lovell (2001) and Ball et al. (2016). Following Ball et al. (2016), TFP growth is also due to output growth rather than changes in the input level.

3.5 Conclusions

This paper decomposes the additively complete LHM TFP indicator into components of technical change, technical inefficiency change and scale inefficiency change. Our approach is general in that it does not require differentiability or convexity of the production technology. Using a nonparametric framework, the empirical application focuses on state-level data of the U.S. agricultural sector over the period $1960 - 2004$. We compute

[19] Zofío and Lovell (2001) only analyze TFP growth over the period $1960 - 1990$.

the scores using FDH and DEA to show the flexibility of our decomposition and to investigate the potential issue of non-convexities in the agricultural sector. Furthermore, we analyze LHM TFP growth and technical change across time, farm types and agricultural intensity rates.

The FDH results show that LHM TFP has increased by 78.61% in the considered period. This is due to output growth (+44.10%) as well as input decline (−34.51%). Technical change (+130.57%) and scale inefficiency change (−60.63%) are the main drivers, while technical inefficiency change (−0.32%) only plays a minor role.

Following the DEA results, LHM TFP has increased by 70.46% for the considered period. This productivity growth is due to output growth (+62.98%) rather than changes in the input level (−7.47%). Technical change is the main driver (+70.55%), while technical inefficiency change (−1.99%) and scale inefficiency change (+0.42%) only play a minor role.

The results thus depend on whether we use FDH or DEA. Although this may partly be driven by the underlying true production technology, we note that FDH may result in too low discriminatory power to compute the distance functions given the relatively low number of observations for the number of variables in this application.

Following the Kolmogorov-Smirnov tests, there seem to be differences in the distributions of LHM TFP growth and technical change across time, farm types and agricultural intensity rates. We suspect that policy instruments and factor endowments (*e.g.*, soil and weather conditions) may drive differences across time, farm types and agricultural intensity rates, potentially resulting in differing distributions in LHM TFP growth and technical change. For instance, agricultural support payments with restrictions on land use (Just and Kropp, 2013) and ethanol subsidies (Motamed et al., 2016) likely have an impact on geographical specialization. This information would be relevant for policy makers. Such an empirical investigation is left for future research.

There are a number of limitations to our work. First, we did not account for possible intertemporal linkages in the modeling of technology in the empirical application. Capital and land are prime examples of "durable" inputs (see Chapter 4) which link production in subsequent periods. Accounting for this in modeling of the technology likely affects the TFP results and its decomposition. Second, the residual approach in defining the scale inefficiency change component differs from the conventional CRS-VRS approach. The accuracy of this residual approach depends on the "step-size", while the CRS-VRS approach uses a (hypothetical) CRS benchmark susceptible to a few (extreme) observations. The first question pertains to the definition of the scale inefficiency change component: can one use the CRS-VRS approach instead of the residual approach and, if so, what is the interpretation of the remaining component? Next, it is of practical interest to determine the conditions under which one should prefer one approach over the other via simulations and empirical applications.

3.A Managi's (2010) decomposition

Managi (2010) decomposes the Luenberger-Hicks-Moorsteen indicator into technical change (TC):

$$\begin{aligned}TC = &\left[D_{t+1}(\mathbf{x}_{t+1}, \mathbf{y}_{t+1}; (\mathbf{g}^i_{t+1}, 0)) - D_t(\mathbf{x}_{t+1}, \mathbf{y}_{t+1}; (\mathbf{g}^i_{t+1}, 0))\right] \\ &- \left[D_{t+1}(\mathbf{x}_{t+1}, \mathbf{y}_{t+1}; (0, \mathbf{g}^o_{t+1})) - D_t(\mathbf{x}_{t+1}, \mathbf{y}_{t+1}; (0, \mathbf{g}^o_{t+1}))\right],\end{aligned}$$

and the residual being efficiency change (EC):

$$\begin{aligned}EC = &\frac{1}{2}\{D_t(\mathbf{x}_t, \mathbf{y}_t; (0, \mathbf{g}^o_t)) - D_t(\mathbf{x}_t, \mathbf{y}_{t+1}; (0, \mathbf{g}^o_{t+1})) \\ &- D_t(\mathbf{x}_{t+1}, \mathbf{y}_t; (\mathbf{g}^i_{t+1}, 0)) + D_t(\mathbf{x}_t, \mathbf{y}_t; (\mathbf{g}^i_t, 0)) \\ &+ D_{t+1}(\mathbf{x}_{t+1}, \mathbf{y}_t; (0, \mathbf{g}^o_t)) + D_{t+1}(\mathbf{x}_{t+1}, \mathbf{y}_{t+1}; (0, \mathbf{g}^o_{t+1})) \\ &- D_{t+1}(\mathbf{x}_{t+1}, \mathbf{y}_{t+1}; (\mathbf{g}^i_{t+1}, 0)) + D_{t+1}(\mathbf{x}_t, \mathbf{y}_{t+1}; (\mathbf{g}^i_t, 0))\Big\} \\ &- D_t(\mathbf{x}_{t+1}, \mathbf{y}_{t+1}; (0, \mathbf{g}^o_{t+1})) - D_{t+1}(\mathbf{x}_{t+1}, \mathbf{y}_{t+1}; (\mathbf{g}^i_{t+1}, 0)) + D_t(\mathbf{x}_{t+1}, \mathbf{y}_{t+1}; (\mathbf{g}^i_{t+1}, 0))\end{aligned}$$

However, this decomposition is incomplete. First, it lacks a scale (in)efficiency change component. Second, there is no reason for TC to be defined as a difference between an output-oriented technical change component and an input-oriented technical change component. Furthermore, TC is only defined with respect to observations in period $t+1$, although there is no clear reason to favor those to observations in period t. Finally, the EC component does not capture technical (in)efficiency change.

3.B Decomposition using the input direction

The decomposition using the input direction is:

$$LHM_{t,t+1} = \Delta T^i + \Delta TEI^i + \Delta SEC^i, \tag{3.17}$$

representing technical change, technical inefficiency change and scale inefficiency change, respectively.

The technical change component is defined as:

$$\begin{aligned}\Delta T^i = &\frac{1}{2}\Big\{\left[D_{t+1}(\mathbf{x}_t, \mathbf{y}_t; (\mathbf{g}^i_t, 0)) - D_t(\mathbf{x}_t, \mathbf{y}_t; (\mathbf{g}^i_t, 0))\right] \\ &+ \left[D_{t+1}(\mathbf{x}_{t+1}, \mathbf{y}_{t+1}; (\mathbf{g}^i_{t+1}, 0)) - D_t(\mathbf{x}_{t+1}, \mathbf{y}_{t+1}; (\mathbf{g}^i_{t+1}, 0))\right]\Big\},\end{aligned} \tag{3.18}$$

and the same interpretation as before. Technical inefficiency change is:

$$\Delta TEI^i = D_t(\mathbf{x}_t, \mathbf{y}_t; (\mathbf{g}^i_t, 0)) - D_{t+1}(\mathbf{x}_{t+1}, \mathbf{y}_{t+1}; (\mathbf{g}^i_{t+1}, 0)). \tag{3.19}$$

From the residual

$$
\begin{aligned}
LHM_{t,t+1} - \Delta T^i - \Delta TEI^i = \quad & & (3.20)\\
& \frac{1}{2}\left\{\left[D_t(\mathbf{x}_t,\mathbf{y}_t;(0,\mathbf{g}^o_t)) - D_t(\mathbf{x}_t,\mathbf{y}_{t+1};(0,\mathbf{g}^o_{t+1}))\right]\right.\\
& \left. + \left[D_{t+1}(\mathbf{x}_{t+1},\mathbf{y}_t;(0,\mathbf{g}^o_t)) - D_{t+1}(\mathbf{x}_{t+1},\mathbf{y}_{t+1};(0,\mathbf{g}^o_{t+1}))\right]\right\}\\
& -\frac{1}{2}\left\{\left[D_t(\mathbf{x}_{t+1},\mathbf{y}_t;(\mathbf{g}^i_{t+1},0)) - D_t(\mathbf{x}_{t+1},\mathbf{y}_{t+1};(\mathbf{g}^i_{t+1},0))\right]\right.\\
& \left. + \left[D_{t+1}(\mathbf{x}_t,\mathbf{y}_t;(\mathbf{g}^i_t,0)) - D_{t+1}(\mathbf{x}_t,\mathbf{y}_{t+1};(\mathbf{g}^i_t,0))\right]\right\},
\end{aligned}
$$

we recover the scale inefficiency change component in a similar way as before. Define the projections of $\mathbf{x}_t$ and $\mathbf{x}_{t+1}$ on the production frontier at time t:

$$\mathbf{x}^*_t = \mathbf{x}_t - D_t(\mathbf{x}_t,\mathbf{y}_t;(\mathbf{g}^i_t,0))\mathbf{g}^i_t \qquad (3.21a)$$

$$\mathbf{x}^{**}_{t+1} = \mathbf{x}_{t+1} - D_t(\mathbf{x}_{t+1},\mathbf{y}_{t+1};(\mathbf{g}^i_{t+1},0))\mathbf{g}^i_{t+1}, \qquad (3.21b)$$

and the projections of $\mathbf{x}_t$ and $\mathbf{x}_{t+1}$ on the production frontier at time $t+1$:

$$\mathbf{x}^{**}_t = \mathbf{x}_t - D_{t+1}(\mathbf{x}_t,\mathbf{y}_t;(\mathbf{g}^i_t,0))\mathbf{g}^i_t \qquad (3.22a)$$

$$\mathbf{x}^*_{t+1} = \mathbf{x}_{t+1} - D_{t+1}(\mathbf{x}_{t+1},\mathbf{y}_{t+1};(\mathbf{g}^i_{t+1},0))\mathbf{g}^i_{t+1}. \qquad (3.22b)$$

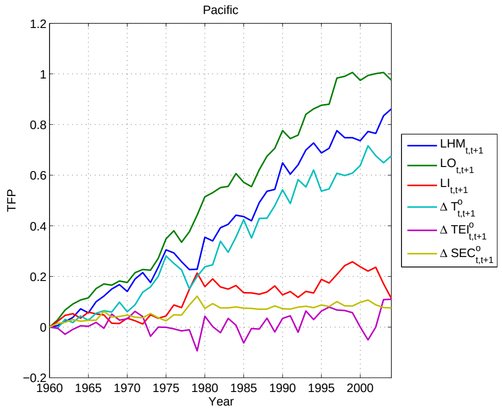
Respectively adding and subtracting $D_t(\mathbf{x}_t,\mathbf{y}_t;(\mathbf{g}^i_t,0))$ and $D_{t+1}(\mathbf{x}_{t+1},\mathbf{y}_{t+1};(\mathbf{g}^i_{t+1},0))$ to and from (3.20), and using the translation property of the directional distance function and the definitions of the projections above, we find the scale inefficiency change component:

$$
\begin{aligned}
\Delta SEC^i = & \frac{1}{2}\left\{\left[D_t(\mathbf{x}_t,\mathbf{y}_t;(0,\mathbf{g}^o_t)) - D_t(\mathbf{x}_t,\mathbf{y}_{t+1};(0,\mathbf{g}^o_{t+1}))\right]\right. & (3.23)\\
& - \left[D_t(\mathbf{x}^{**}_{t+1},\mathbf{y}_t;(\mathbf{g}^i_{t+1},0)) - D_t(\mathbf{x}^*_t,\mathbf{y}_t;(\mathbf{g}^i_t,0))\right]\\
& + \left[D_{t+1}(\mathbf{x}_{t+1},\mathbf{y}_t;(0,\mathbf{g}^o_t)) - D_{t+1}(\mathbf{x}_{t+1},\mathbf{y}_{t+1};(0,\mathbf{g}^o_{t+1}))\right]\\
& \left. - \left[D_{t+1}(\mathbf{x}^*_{t+1},\mathbf{y}_{t+1};(\mathbf{g}^i_{t+1},0)) - D_{t+1}(\mathbf{x}^{**}_t,\mathbf{y}_{t+1};(\mathbf{g}^i_t,0))\right]\right\}\\
& \equiv \frac{1}{2}\left\{\Delta SEC^i_t + \Delta SEC^i_{t+1}\right\}.
\end{aligned}
$$

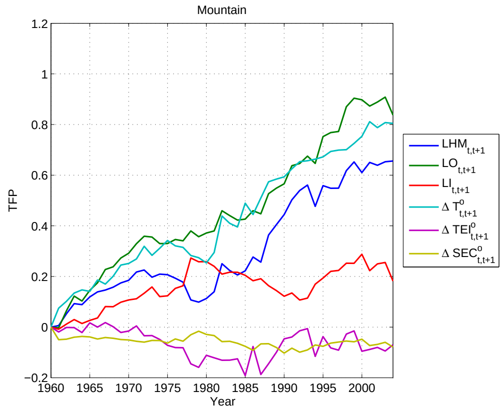
3.C State-level TFP figures

This appendix includes the LHM TFP indicator and its components per agricultural region. Each figure is constructed by averaging over all states in that particular agricultural region in every year.

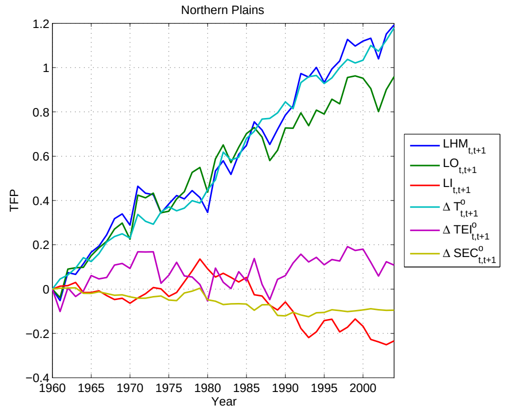
3.C.1 Convex technology



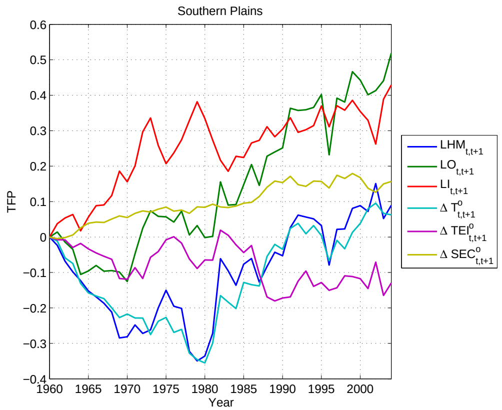
(a) Pacific.



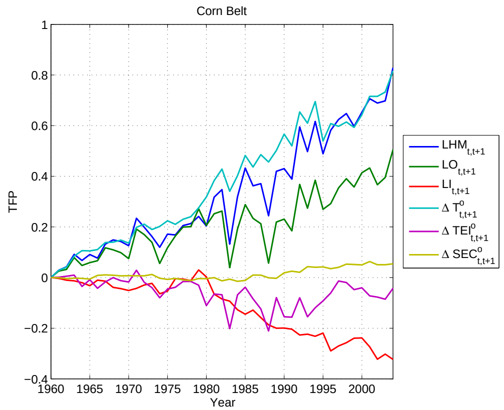
(b) Mountain.



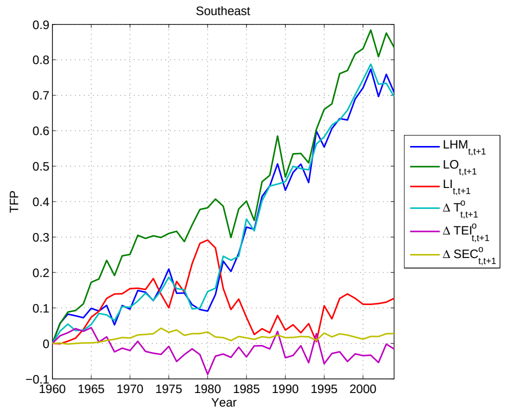
(c) Northern Plains.



(d) Southern Plains.

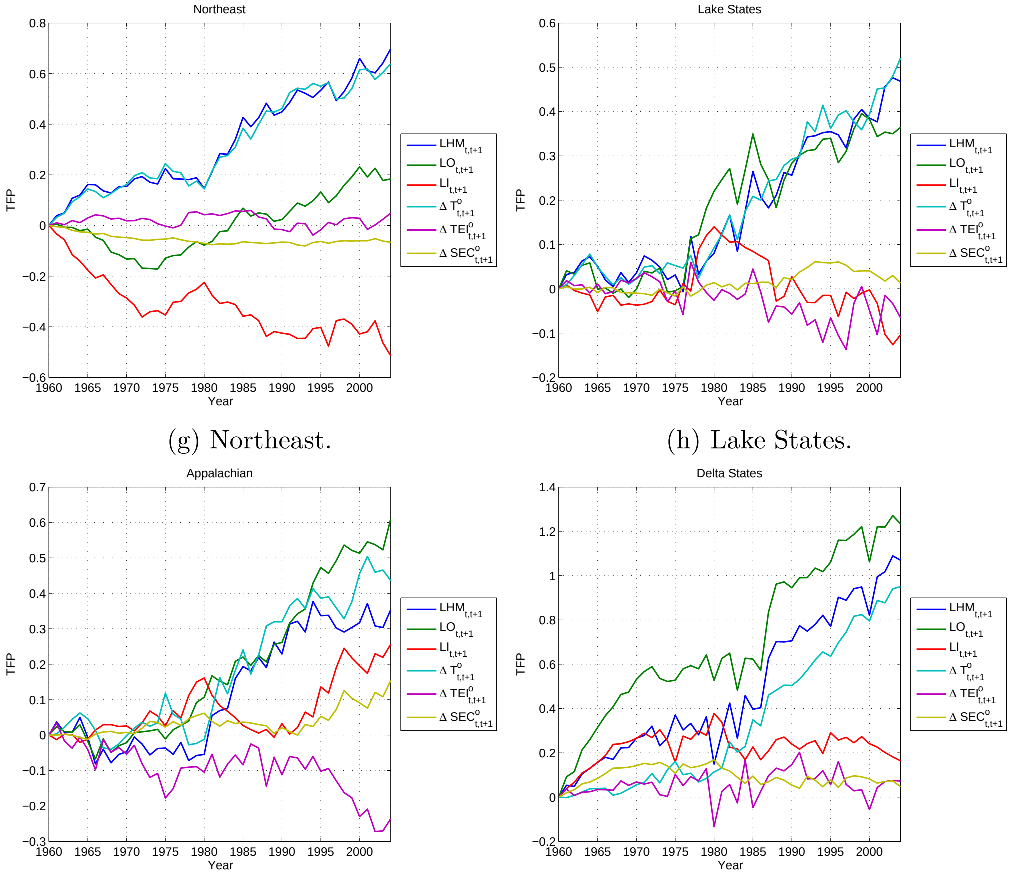


(e) Corn Belt.



(f) Southeast.

Figure 3.10: Cumulative LHM TFP indicator and its components per agricultural region under a convex technology.



(g) Northeast.

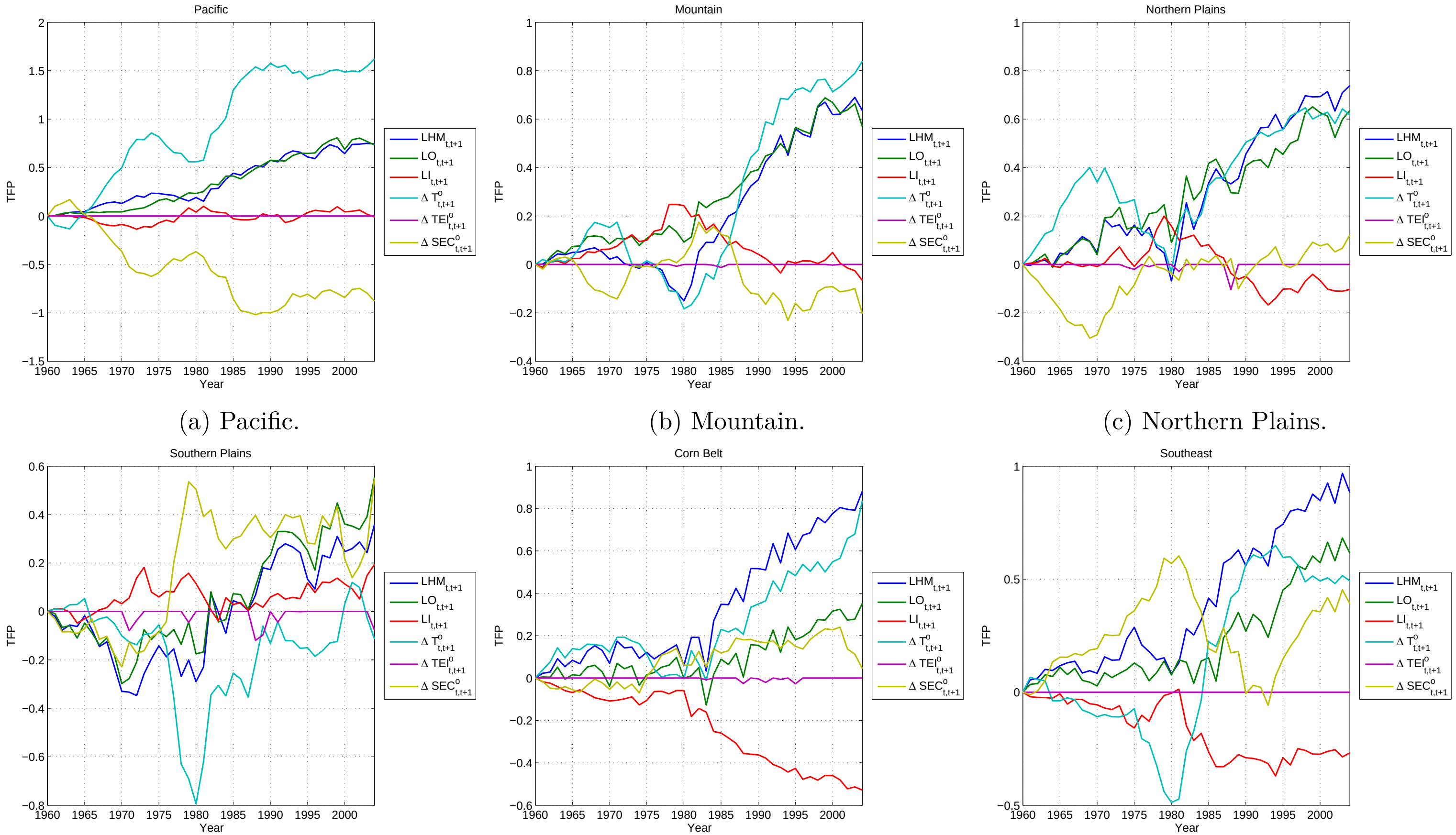
(h) Lake States.

(i) Appalachian.

(j) Delta States.

Figure 3.10: Cumulative LHM TFP indicator and its components per agricultural region under a convex technology.

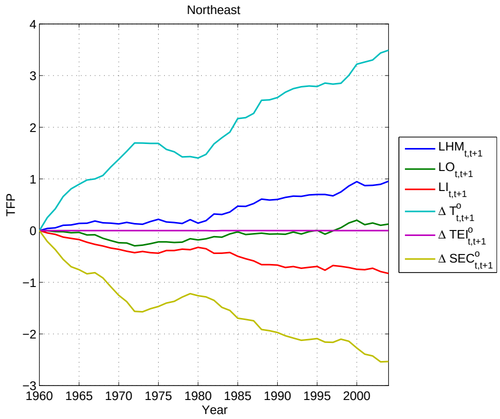
3.C.2 Non-convex technology



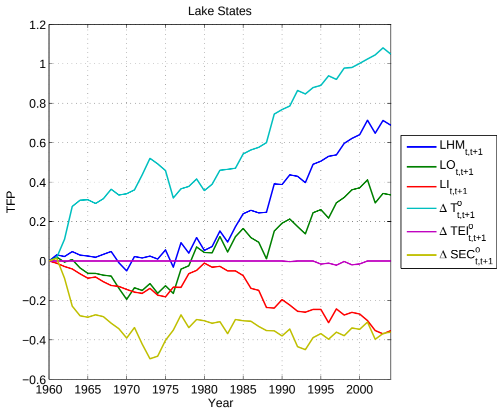
(a) Pacific. (b) Mountain. (c) Northern Plains.

(d) Southern Plains. (e) Corn Belt. (f) Southeast.

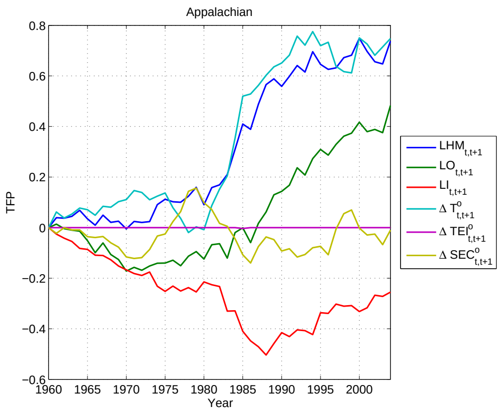
Figure 3.11: Cumulative LHM TFP indicator and its components per agricultural region under a non-convex technology.



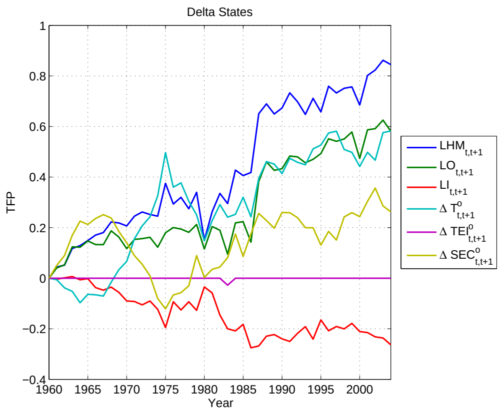
(g) Northeast.



(h) Lake States.



(i) Appalachian.



(j) Delta States.

Figure 3.11: Cumulative LHM TFP indicator and its components per agricultural region under a non-convex technology.

Chapter **4**

Production with storable and durable inputs: nonparametric analysis of intertemporal efficiency

"Imagine that you visit a shipyard. Day by day a tremendous amount of production activity of great variety is carried on, yet no ships are turned out. [...] The shipbuilding production system, like construction, is a dynamically evolving process."

— Ronald W. Shephard[1]

4.1 Introduction

Many production decisions have long term consequences for production and are capital-intensive: should a firm merely buy new machines to replace older machines that have reached (physical) end of life? Or should the firm invest in new machines to expand its production capacity? These capital-intensive investments are "durable" by nature, because they have a long term impact on production. Furthermore, firms often buy far larger quantities of inputs then they currently need. This can be economically rational for a number of reasons: there are discounts on bulk purchases of inputs or firms expect input prices to rise in the near future. These "storable" inputs can be stored in inventories and are used over several time periods. These durable and storable inputs used in production limit the flexibility of a firm in adjusting its input mix. In this paper, we introduce a novel methodology for economic (cost) efficiency analysis that explicitly takes these intertemporal aspects of firms' production behavior into account. This obtains a more realistic modeling of intertemporal relations in production situations where storability and durability of inputs are relevant, which is often the case in real-life settings.

[0] This chapter is based on joint work with Laurens Cherchye (KU Leuven) and Bram De Rock (ULB & KU Leuven).

[1] Preface to Shephard and Färe (1980, p.V)

4.1.1 Intertemporal efficiency and regulation

In regulated industries it is particularly vital that the regulator takes intertemporal dependencies of production into account in the regulation exercise. However, regulators generally do not incorporate these interdependencies in practice. This is sometimes motivated by a lack of panel data, which forces regulators to limit the analysis to cross-sectional data (Pollitt, 2005). However, also the used definition of capital costs (Shuttleworth, 2005) and capitalization policies across firms, countries or industries (Haney and Pollitt, 2013) can be contested.[2] Clearly, not taking these dependencies into account can lead to erroneous cost reduction targets. Shuttleworth (2005) reports a case where Ofgem, a UK electricity distribution regulator, imposed a too strong target for one distributor (Seeboard) while imposing a too loose target on another (Southern). This discrepancy was due to the fact that Ofgem only considered operational expenses, while disregarding capital expenses. And it happened that Southern was characterized by high capital expenses and low operational expenses, while the opposite applied to Seeboard. Balk et al. (2010) also discuss the importance of correctly considering capital expenses in the regulation analysis and note that there is no consensus on how to properly deal with capital expenses in the studies they considered.

The relation between regulatory regime and investment has received a lot of attention (see Guthrie (2006) for a discussion). Focusing on cross-sectional data can lead to penalization of firms that invest while rewarding those that delay investments. Nick and Wetzel (2015) conclude that firms have an incentive to cut investments when the regulator uses a static benchmarking model. Our empirical application to Swiss railway companies will show that the resulting dynamic efficiency conclusions may significantly differ from the ones that are based on a static efficiency analysis. In our opinion, this directly motivates the practical relevance of our methodology, as these differences may substantially affect the regulatory policies that are based on the efficiency assessment.

4.1.2 Efficiency analysis with durable and storable inputs

The existing literature has devoted much attention to the analysis of dynamically efficient production behavior from a technical perspective (see Fallah-Fini et al. (2013) for a recent review). Such technical efficiency analysis then focuses, for example, on the modeling of production delays, inventories, capital (quasi-fixed factors in general), adjustment costs and learning. By contrast, far less work has tackled the issue from an economic perspective.[3] Importantly, however, the distinction between economic and technical efficiency analysis becomes particularly relevant in dynamic decision settings.

For durable inputs, it has long been known that firms do not scrap old (durable) capital equipment the moment new equipment becomes available. The process of replacing capital equipment is rather gradual. Firms deciding on new capital equipment face

[2]There does exist general guidelines on capital measurement. We refer to the manual on capital measurement of the OECD (2009) for an example.

[3]Notable exceptions include Nemoto and Goto (1999, 2003); Ouellette and Yan (2008); Silva and Stefanou (2003). We discuss the relation between our framework and this existing work in Section 4.2.

different substitution possibilities between inputs before (ex ante) and after (ex post) the purchase: once capital equipment is installed, it remains in use until the end of its predetermined lifetime (Forsund and Hjalmarsson, 1974; Johansen, 1959). Thus, while firms might seem inefficient from a technical perspective, they may actually be efficient from an economic perspective.[4]

Similarly, when deciding upon storable inputs, firms typically plan their production in advance for a certain time horizon. They form expectations on prices and demand and then decide on the amount of necessary inputs to acquire. Clearly, if prices of storable inputs vary over time, this can again generate significant discrepancies between technical and economic efficiency analysis.

In this paper, we present a unifying framework to analyze intertemporal cost minimizing behavior with both durable and storable inputs. For durable inputs, our framework explicitly models the possibility that firms use several vintages: they invest in new durables and scrap older durables over time. Furthermore, we allow for production delays of durable inputs. We also show how our framework can incorporate alternative hypotheses such as degressive write-off of durables over time.

A main distinguishing feature of our methodology is that it is intrinsically nonparametric (in the spirit of Afriat (1972), Varian (1984) and Banker and Maindiratta (1988)): it can analyze production behavior without imposing any (usually non-verifiable) functional structure on the production technology. We characterize production behavior that is intertemporal cost efficient, which allows us to evaluate the efficiency of observed production decisions. For cost inefficient behavior, we propose a measure that quantifies the degree of inefficiency. This intertemporal inefficiency measure has the attractive property that it can be decomposed in period-specific cost inefficiencies.

4.1.3 Outline

The remainder of this paper unfolds as follows. In Section 4.2, we discuss the connection between our work and the closely related literature on both intertemporal production models and efficiency analysis. Sections 4.3 to 4.5 formally introduce our methodology. After introducing our general set-up in Section 4.3, we first consider the case where one has full information on allocations of storables and write-offs of durables in Section 4.4, to subsequently present the case where limited or no such information is known in Section 4.5. Section 4.6 presents some extensions to the basic framework. Section 4.7 contains the empirical application of our methodology to Swiss regional railway companies. Specifically, this analysis will demonstrate the relevance of accounting for the intertemporal (durable) nature of capital expenses in a regulated production environment. Finally, Section 4.8 concludes and points out a number of interesting extensions.

[4] Wibe (2008) coined the term “rational inefficiency” to mark this difference.

4.2 Related literature

Our framework for intertemporal production analysis bears close connections with a number of existing studies on the analysis of efficient production behavior. Most of this earlier work appeared under the label Data Envelopment Analysis (DEA), which is often used to refer to the nonparametric analysis of production efficiency. In what follows, we discuss the relation with earlier literature on network DEA, efficiency analysis with quasi-fixed inputs, and DEA with lagged input effects. In turn, this will allow us to articulate the specificities of our own contribution.

First, our work is closely related to the literature on network DEA (Färe and Grosskopf, 2000) and dynamic DEA (Färe and Grosskopf, 1996). In an early contribution to this literature, Färe (1986) showed how to measure output efficiency by allowing for inputs that are allocatable over time, which are similar in nature to what we call storable inputs. He makes a distinction between inputs for which the allocation over time is known and inputs for which (only) the total amount is known but not how this amount is allocated over time. Importantly, however, he does not consider the intermediate case with new inputs in every period that are to be allocated over multiple time periods. In a similar fashion, Färe et al. (1997) model fixed but allocatable inputs over outputs and develop an output efficiency measure that locates potential efficiency gains due to the reallocation of inputs over the outputs. Färe et al. (2010) consider the problem of resource allocation over time distinguishing between the decision of when to start allocation and over how many periods to allocate the resources. They also consider this problem under specific returns-to-scale assumptions, capacity constraints and technical change. Again, he does not consider the problem of allocating new inputs after the start period. In the current setting we consider the time frame fixed, but allow for new inputs in every time period that need to be allocated in subsequent time periods. Finally, inventories are also explicitly modeled in Hackman and Leachman (1989)'s general framework of production.

Similarly to our use of durable inputs, Färe et al. (2007) construct a network DEA model with durable and instantaneous inputs to model technology adoption, where one of the technologies is vintage. Durable inputs are vintage-specific, and the adoption of a new technology is accomplished by diverting instantaneous inputs away from the vintage technology to the new technology. Kao (2013) models a dynamic DEA model where the intertemporal dependence among production processes is modeled by quasi-fixed inputs or intermediate products. The overall system efficiency measure can be decomposed as a weighted sum of per-period efficiency measures. These per-period efficiency measures are not necessarily unique and hence not comparable among different DMUs. In a two-stage DEA model Kao and Hwang (2008) maximize the efficiency of stage 1 while maintaining the overall system efficiency. In this way stage 1 efficiency is maximal for all DMUs and can be compared among DMUs.

All these network DEA models have in common that they measure technical efficiency (without price information) and not economic efficiency (with price information). Such technical efficiency analysis requires specific assumptions regarding the nature of the production technology.[5,6]

Furthermore, our concept of durable inputs is also related to the notion of quasi-fixed inputs. Nemoto and Goto (1999, 2003) model adjustment costs due to quasi-fixed inputs and develop an efficiency measure. They treat quasi-fixed inputs as intermediate outputs which are used as inputs in subsequent periods. Their model was extended by Ouellette and Yan (2008) by weakening the restrictions on capital investment. Similarly, Silva and Stefanou (2003) develop nonparametric tests for investment in quasi-fixed inputs with internal adjustment costs in the spirit of Varian (1984).

Next, Chen and van Dalen (2010) incorporate lagged effects of inputs on outputs in DEA efficiency measurement. The relation between output and delayed inputs is fixed parametrically. Thus, they assume that these productive effects are known a priori and estimate these by a fixed effect panel vector autoregressive model in their empirical application. This makes their efficiency measure highly dependent on their parametric specification of the productive effects.

Basically, our contribution is that we present a unifying framework to nonparametrically analyze economic (cost) efficiency in intertemporal production with both storable and durable inputs. We explicitly model the fact that these two types of inputs are used over several time periods: storable inputs are allocated over multiple periods, and durable "vintage" inputs are not immediately replaced by newer durable inputs (thus following Johansen (1959) and Forsund and Hjalmarsson (1974)). In addition, we also allow for production delays of durable inputs over time. Next, we propose a cost inefficiency measure that can be decomposed in per-period inefficiencies.[7] Finally, as compared to the literature on quasi-fixed inputs, we do not focus on the issue of adjustment costs, but rather consider the replacement of vintages of durables over time from a cost perspective (see also the introduction of Section 4.3).

4.3 Set-up

We assume a balanced panel setting with K firms that are observed T times. For each firm k and time period t, we observe the S-dimensional output $\mathbf{y}_{k,t} \in \mathbb{R}^S_+$, the N-dimensional storable input $\mathbf{q}_{k,t} \in \mathbb{R}^N_+$, the M-dimensional durable input $\mathbf{Q}_{k,t} \in \mathbb{R}^M_+$ and the corresponding discounted input prices $\mathbf{p}_{k,t} \in \mathbb{R}^N_{++}$ and $\mathbf{P}_{k,t} \in \mathbb{R}^M_{++}$ respectively. For every $k = 1, \ldots, K$, this defines the dataset

$$\mathcal{S}_k = \left\{(\mathbf{p}_{k,t}, \mathbf{q}_{k,t}, \mathbf{P}_{k,t}, \mathbf{Q}_{k,t}, \mathbf{y}_{k,t}) | t = 1, \ldots, T\right\}.$$

[5] In our concluding Section 4.8 we will indicate the possibility to conduct a technical efficiency analysis in the intertemporal framework (for economic efficiency analysis) that we develop in the following sections. These technical efficiency formulations could subsequently establish a formal link with the existing network DEA models.

[6] See the early contribution of Shephard and Färe (1980) for a formal treatment in terms of set representation and distance functions.

[7] We note that Kao (2013) proposed a similar decomposition of efficiency scores in per-period efficiencies for a DEA model with quasi-fixed inputs.

To keep our exposition simple, we assume that firms have perfect foresight, i.e., they exactly anticipate the future prices. In fact, it is fairly easy to extend our method to account for predicted prices that deviate from the prices that are realized ex post.[8] But this would only complicate our reasoning without really adding new insights. Also, the fact that we evaluate a firm's cost efficiency in terms of realized prices makes that we may interpret measured inefficiencies as (ex post) prediction errors.

Storable inputs are divisible and we assume they are used over J periods: a fraction is used in each period, while the remaining part is stored for the next periods. Storable inputs can only be used once and are nondurable. Durable inputs are indivisible and usable in multiple periods before reaching end of life status. This is where they differ from storable inputs. Durable inputs are related to quasi-fixed inputs in that they have an effect over multiple periods, but differ from quasi-fixed inputs because they may also be adjusted instantaneously (e.g., one can stop using a laptop or company car immediately). In that sense, we can see quasi-fixed inputs as a subset of durable inputs. In general, durable inputs are seen as investments: a firm intends to use the durable input for a number of periods and writes off the cost of investment over these periods. Examples of durable inputs include machines, equipment, company cars, etc. To keep the exposition simple, we also assume they are used over J periods. We show in Section 4.6 how this assumption can be relaxed.

Our behavioral hypothesis is that firms are intertemporally cost minimizing. To formalize this assumption, we represent firm technologies in terms of input requirement sets $\mathcal{I}_t(\mathbf{y}_{k,t})$ for the output of firm k produced at time period t. These sets are defined in the usual way, i.e.,

$$\mathcal{I}_t(\mathbf{y}_{k,t}) = \left\{(\mathbf{q}, \mathbf{Q}) \in \mathbb{R}_+^{N+M} | (\mathbf{q}, \mathbf{Q}) \text{ can produce } \mathbf{y}_{k,t}\right\}.$$

Next, we make use of quantity allocations $(\mathfrak{q}_t^1, \ldots, \mathfrak{q}_t^J)_{t=1}^T$ of storable inputs and price write-offs $(\mathfrak{P}_t^1, \ldots, \mathfrak{P}_t^J)_{t=1}^T$ of durable inputs. These allocations and write-offs will be used to distribute firm k's input costs over the J relevant time periods, and are subject to the adding-up restrictions $\mathbf{q}_{k,t} = \sum_{j=1}^J \mathfrak{q}_{k,t}^j$ and $\mathbf{P}_{k,t} = \sum_{j=1}^J \mathfrak{P}_{k,t}^j$.[9] Then, we say that firm k minimizes its total production costs over the time horizon $[J, \ldots, T]$ if it chooses the allocation $(\mathfrak{q}_t^1, \ldots, \mathfrak{q}_t^J)_{t=1}^T$ and write-off $(\mathfrak{P}_t^1, \ldots, \mathfrak{P}_t^J)_{t=1}^T$ that solves

$$\min_{\substack{(\mathfrak{q}_t^1, \ldots, \mathfrak{q}_t^J)_{t=1}^T \\ (\mathfrak{P}_t^1, \ldots, \mathfrak{P}_t^J)_{t=1}^T}} \sum_{t=J}^{T} \sum_{j=t-J+1}^{t} \left(\mathbf{p}_{k,j} \mathfrak{q}_j^{t-j+1} + \mathfrak{P}_j^{t-j+1} \mathbf{Q}_{k,j}\right) \tag{4.1a}$$

$$\text{s.t.} \quad \left(\sum_{j=1}^{J} \mathfrak{q}_{t-j+1}^j, \sum_{j=1}^{J} \mathbf{Q}_{k,t-j+1}\right) \in \mathcal{I}_t(\mathbf{y}_{k,t}) \qquad \forall t = J, \ldots, T, \tag{4.1b}$$

[8] For example, a simple solution consists of verifying the cost efficiency conditions that we define below for alternative specifications of (anticipated) prices, as a robustness analysis.

[9] In Appendix 4.A, we explain the economic intuition of the write-offs $(\mathfrak{P}_t^1, \ldots, \mathfrak{P}_t^J)_{t=1}^T$ as representing (in monetary terms) marginal productivities of the durable inputs.

where the last feasibility constraint states that the allocation of storables and durables effectively admits the production of the output $\mathbf{y}_{k,t}$ for the given technology. The fact that the storable and durable input quantities are summed over J present and past periods reveals the intertemporal dependency of firm k's production decisions.

Table 4.1 sharpens the intuition of the above concepts through a simple example that shows a firm's observed costs and production costs over time for $J = 2$. The table illustrates two crucial points. First, observed costs and production costs generally differ. Thus, any efficiency comparison using observed costs instead of production costs is potentially overly pessimistic. Second, the lack of information on allocations and write-offs beyond the observed time frame limits any test of (4.1) to the time period $[2, \ldots, T]$ when $J = 2$ and $[J, \ldots, T]$ in general. This explains why we only consider the period $[J, \ldots, T]$ in (4.1) instead of $[1, \ldots, T]$ in our minimization program.

t	observed cost	production cost	
		storable inputs	durable inputs
1	$\mathbf{p}_1\mathfrak{q}_1$	$\mathbf{p}_1\mathfrak{q}_1^1 + ?$	$\boldsymbol{\mathfrak{P}}_1^1\mathbf{Q}_1 + ?$
2	$\mathbf{p}_2\mathfrak{q}_2$	$\mathbf{p}_2\mathfrak{q}_2^1 + \mathbf{p}_1\mathfrak{q}_1^2$	$\boldsymbol{\mathfrak{P}}_2^1\mathbf{Q}_2 + \boldsymbol{\mathfrak{P}}_1^2\mathbf{Q}_1$
3	$\mathbf{p}_3\mathfrak{q}_3$	$\mathbf{p}_3\mathfrak{q}_3^1 + \mathbf{p}_2\mathfrak{q}_2^2$	$\boldsymbol{\mathfrak{P}}_3^1\mathbf{Q}_3 + \boldsymbol{\mathfrak{P}}_2^2\mathbf{Q}_2$
4	$\mathbf{p}_4\mathfrak{q}_4$	$\mathbf{p}_4\mathfrak{q}_4^1 + \mathbf{p}_3\mathfrak{q}_3^2$	$\boldsymbol{\mathfrak{P}}_4^1\mathbf{Q}_4 + \boldsymbol{\mathfrak{P}}_3^2\mathbf{Q}_3$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
t	$\mathbf{p}_t\mathfrak{q}_t$	$\mathbf{p}_t\mathfrak{q}_t^1 + \mathbf{p}_{t-1}\mathfrak{q}_{t-1}^2$	$\boldsymbol{\mathfrak{P}}_t^1\mathbf{Q}_t + \boldsymbol{\mathfrak{P}}_{t-1}^2\mathbf{Q}_{t-1}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$T-1$	$\mathbf{p}_{T-1}\mathfrak{q}_{T-1}$	$\mathbf{p}_{T-1}\mathfrak{q}_{T-1}^1 + \mathbf{p}_{T-2}\mathfrak{q}_{T-2}^2$	$\boldsymbol{\mathfrak{P}}_{T-1}^1\mathbf{Q}_{T-1} + \boldsymbol{\mathfrak{P}}_{T-2}^2\mathbf{Q}_{T-2}$
T	$\mathbf{p}_T\mathfrak{q}_T$	$\mathbf{p}_T\mathfrak{q}_T^1 + \mathbf{p}_{T-1}\mathfrak{q}_{T-1}^2$	$\boldsymbol{\mathfrak{P}}_T^1\mathbf{Q}_T + \boldsymbol{\mathfrak{P}}_{T-1}^2\mathbf{Q}_{T-1}$
$T+1$	?	$? + \mathbf{p}_T\mathfrak{q}_T^2$	$? + \boldsymbol{\mathfrak{P}}_T^2\mathbf{Q}_T$

Table 4.1: Overview of production costs with storable and durable inputs for $J = 2$.

4.4 Complete information

We next turn to deriving operational conditions for cost minimizing behavior as defined in (4.1). As indicated in the Introduction, we derive nonparametric conditions in the spirit of Afriat (1972); Banker and Maindiratta (1988); Varian (1984), which make minimal assumptions regarding the production technology. To set the stage, we first consider the limiting case that is characterized by full information on the quantity allocations of the storable inputs and the price write-offs of the durable inputs, that is, for each firm k the empirical analyst observes the allocations $(\mathfrak{q}_{k,t}^1, \mathfrak{q}_{k,t}^2, \ldots, \mathfrak{q}_{k,t}^J)$ and the write-offs $(\boldsymbol{\mathfrak{P}}_{k,t}^1, \boldsymbol{\mathfrak{P}}_{k,t}^2, \ldots, \boldsymbol{\mathfrak{P}}_{k,t}^J)$ at every time period $t = 1, ..., T$.

Such complete information greatly simplifies matters. From (4.1), it is easy to verify that, for a given specification of storable allocations and durable write-offs, the production costs for any time period t are defined independently of the production costs for other time periods. As an implication, firm k behaves consistently with (4.1) if and only if it solves, for every $t = J, ..., T$,

$$\min_{\substack{(\mathfrak{q}_j^1,\ldots,\mathfrak{q}_j^J)_{j=t-J+1}^t \\ (\mathfrak{P}_j^1,\ldots,\mathfrak{P}_j^J)_{j=t-J+1}^t}} \sum_{j=t-J+1}^{t} \left(\mathbf{p}_{k,j}\mathfrak{q}_j^{t-j+1} + \mathfrak{P}_j^{t-j+1}\mathbf{Q}_{k,j}\right) \tag{4.2a}$$

$$\text{s.t.} \quad \left(\sum_{j=1}^{J} \mathfrak{q}_{t-j+1}^{j}, \sum_{j=1}^{J} \mathbf{Q}_{k,t-j+1}\right) \in \mathcal{I}_t(\mathbf{y}_{k,t}) \tag{4.2b}$$

Putting it differently, dynamically cost minimizing behavior under complete information can be represented as statically cost minimizing behavior for every period t. Varian (1984) developed the nonparametric characterization of such static cost minimization.[10] Thus, we can obtain our empirical condition for dynamic cost efficiency by translating Varian's reasoning to our particular setting.

Throughout, we will adopt the next two axioms regarding the production technology (given by $\mathcal{I}_t(\mathbf{y}_{k,t})$):

Axiom 4.1 (observability means feasibility)**.** *For all* $t = 1, \ldots, T$ *and* $k = 1, \ldots, K$: $(\mathbf{p}_{k,t}, \mathfrak{q}_{k,t}^1, \ldots, \mathfrak{q}_{k,t}^J, \mathfrak{P}_{k,t}^1, \ldots, \mathfrak{P}_{k,t}^J, \mathbf{Q}_{k,t}, \mathbf{y}_{k,t}) \in \mathcal{S}_k \Rightarrow \left(\sum_{j=1}^J \mathfrak{q}_{k,t-j+1}^j, \sum_{j=1}^J \mathbf{Q}_{k,t-j+1}\right) \in \mathcal{I}_t(\mathbf{y}_{k,t})$.

Axiom 4.2 (nested input sets)**.** *For all* $t = 1, \ldots, T$ *and* $k, s = 1, \ldots, K$: $\mathbf{y}_{s,t} \geq \mathbf{y}_{k,t} \Rightarrow \mathcal{I}_t(\mathbf{y}_{s,t}) \subseteq \mathcal{I}_t(\mathbf{y}_{k,t})$.[11]

In words, Axiom 4.1 says that there are no significant measurement errors in the data.[12] Axiom 4.2 says that, for a given time period t, input requirement sets are nested: if firm s produces at least the same output as firm k (i.e., $\mathbf{y}_{s,t} \geq \mathbf{y}_{k,t}$), then the input set for s must be contained in the set for k (i.e., $\mathcal{I}_t(\mathbf{y}_{s,t}) \subseteq \mathcal{I}_t(\mathbf{y}_{k,t})$).[13] Intuitively, this means that outputs are freely disposable. These are the only two production axioms that we will assume in the sequel of this paper.

[10] Varian (1984) characterized cost minimizing production behavior in terms of the so-called Weak Axiom of Cost Minimization (WACM). Basically, Proposition 4.1 will state this WACM criterion for our intertemporal setting.

[11] Throughout $\mathbf{y}_{s,t} \geq \mathbf{y}_{k,t}$ should be interpreted as vector inequalities, implying that the inequality needs to hold for all components.

[12] Clearly, this axiom may often be problematic in practical situations. In such instances, we can use alternative techniques to explicitly account for errors. For example, one may adjust our methodology by integrating it with the probabilistic method which Cazals, Florens, and Simar (2002) and Daraio and Simar (2005, 2007) originally proposed in a DEA context. To focus our discussion, we do not consider this extension here.

[13] We remark that this assumes that different firms s and k face the same technology in period t. Obviously, we can also use other hypotheses regarding technological homogeneity/heterogeneity across firms and time periods. For example, we may assume homogeneous technologies (only) for subsets of firms (e.g., defined on the basis of observable firm characteristics), or firm-specific technologies that are constant over time. For compactness, we will again not explicitly implement this.

Then, we define

$$c_{k,t} = \min_{s \in D_k^t} \left\{ \sum_{j=t-J+1}^{t} \left(\mathbf{p}_{k,j} \mathfrak{q}_{s,j}^{t-j+1} + \mathfrak{P}_{k,j}^{t-j+1} \mathbf{Q}_{s,j} \right) \right\}. \tag{4.3}$$

for

$$D_k^t = \left\{ s | \mathbf{y}_{s,t} \geq \mathbf{y}_{k,t} \right\}, \tag{4.4}$$

i.e., the set of observed firms s that produce at least the same output as firm k in period t (i.e., $\mathbf{y}_{s,t} \geq \mathbf{y}_{k,t}$). By construction, we have $k \in D_k^t$, so that $D_k^t \neq \emptyset$. In words, $c_{k,t}$ represents the minimal cost over this set D_k^t. Obviously, we can compute $c_{k,t}$ by simply enumerating over all $s \in D_k^t$ if the allocations $(\mathfrak{q}_{k,t}^1, \mathfrak{q}_{k,t}^2, \ldots, \mathfrak{q}_{k,t}^J)$ and write-offs $(\mathfrak{P}_{k,t}^1, \mathfrak{P}_{k,t}^2, \ldots, \mathfrak{P}_{k,t}^J)$ are given.

We can now state the following result.

Proposition 4.1. *Firm k solves* (4.1) *for a production technology that satisfies Axioms* (4.1) *and* (4.2) *if and only if, for all* $t = J, ..., T$,

$$\sum_{j=t-J+1}^{t} \left(\mathbf{p}_{k,j} \mathfrak{q}_{k,j}^{t-j+1} + \mathfrak{P}_{k,j}^{t-j+1} \mathbf{Q}_{k,j} \right) = c_{k,t}. \tag{4.5}$$

Proof. We use the equivalence between (4.1) and (4.2). Then, the result follows from Theorem 1 (statements (1) and (2)) of Varian (1984). □

This results directly suggests the next measure of cost inefficiency for every period t:

$$CE_k^t \equiv \sum_{j=t-J+1}^{t} \left(\mathbf{p}_{k,j} \mathfrak{q}_{k,j}^{t-j+1} + \mathfrak{P}_{k,j}^{t-j+1} \mathbf{Q}_{k,j} \right) - c_{k,t}, \tag{4.6}$$

Obviously, firm k meets the empirical cost minimization criterion (4.5) in Proposition 4.1 if and only if $CE_k^t = 0$. More generally, we have $CE_k^t \geq 0$, and the value of CE_k^t indicates how much firm k deviates from cost minimizing behavior at time t.

When aggregating over all $t = J, \ldots, T$, we can similarly define an overall cost inefficiency measure as

$$CE_k \equiv \sum_{t=J}^{T} \sum_{j=t-J+1}^{t} \left(\mathbf{p}_{k,j} \mathfrak{q}_{k,j}^{t-j+1} + \mathfrak{P}_{k,j}^{t-j+1} \mathbf{Q}_{k,j} \right) - \sum_{t=J}^{T} c_{k,t}. \tag{4.7}$$

By construction, we have

$$CE_k = \sum_{t=J}^{T} \left(\sum_{j=t-J+1}^{t} \left(\mathbf{p}_{k,j} \mathfrak{q}_{k,j}^{t-j+1} + \mathfrak{P}_{k,j}^{t-j+1} \mathbf{Q}_{k,j} \right) - c_{k,t} \right) = \sum_{t=J}^{T} CE_k^t, \tag{4.8}$$

which yields the next result.

Proposition 4.2. $CE_k = 0 \Leftrightarrow CE_k^t = 0 \ \forall t = J, \dots, T$.

Proof. The result follows from (4.7) and the definitional fact that $CE_k^t \geq 0$. □

In words, firm k minimizes its total production costs over the full period $[J, \dots, T]$ if and only if its production costs are minimal in every single period t. Essentially, this result shows that our overall cost inefficiency measure CE_k satisfies the aggregate indication axiom of Blackorby and Russell (1999).

4.5 Incomplete information

The previous section assumed an ideal scenario in which the empirical analyst had full knowledge of the allocations $(\mathfrak{q}_{k,t}^1, \mathfrak{q}_{k,t}^2, \dots, \mathfrak{q}_{k,t}^J)$ and the write-offs $(\mathfrak{P}_{k,t}^1, \mathfrak{P}_{k,t}^2, \dots, \mathfrak{P}_{k,t}^J)$. In practice, however, only very limited information on allocations and write-offs is often available. It may even happen that such information is completely absent. This section shows how to proceed in such (more realistic) instances.

Formally, we will assume that the available information is captured by the polyhedron

$$\Theta(\mathbf{A}, \mathbf{b}) \equiv \left\{ \boldsymbol{\rho} \in \mathbb{R}_+^{TJ(N+M)} : \mathbf{A}\boldsymbol{\rho} \geq \mathbf{b} \right\}, \tag{4.9}$$

which represents L restrictions on the allocations of storable inputs and on the write-offs of durable inputs. Specifically, $\mathbf{A}$ is a $L \times TJ(N+M)$ matrix and $\mathbf{b}$ a $L \times 1$ vector, and $\boldsymbol{\rho}$ represents all vectors that satisfy the constraints imposed by $\mathbf{A}$ and $\mathbf{b}$.

To structure our discussion, we will first consider the limiting case in which we cannot use any information on firms' allocations and write-offs, which corresponds to $\Theta = \mathbb{R}_+^{TJ(N+M)}$. Subsequently, we will discuss the intermediate scenario where some information is available, i.e., $\Theta \subset \mathbb{R}_+^{TJ(N+M)}$.

4.5.1 No information on allocations and write-offs

In the absense of full information on storable allocations and durable write-offs, we can no longer check the condition (4.2) independently for every single time period t. In this case, we verify if there exists at least one possible specification of $(\mathfrak{q}_{k,t}^1, \mathfrak{q}_{k,t}^2, \dots, \mathfrak{q}_{k,t}^J)$ and $(\mathfrak{P}_{k,t}^1, \mathfrak{P}_{k,t}^2, \dots, \mathfrak{P}_{k,t}^J)$ that makes firm k's behavior consistent with the overall cost minimization condition (4.1). More specifically, we define feasible allocations and write-offs that present firm k as efficient as possible. This evaluates firm k in the most favorable light and, thus, gives this firm the benefit-of-the-doubt in the absence of full information.[14]

[14] This benefit-of-the-doubt idea is intrinsic to DEA efficiency evaluations. See, for example, Cherchye et al. (2007) for a detailed discussion of the benefit-of-the-doubt interpretation of DEA models in the specific context of composite indicator construction.

The following linear program operationalizes this idea:

$$\min_{\substack{c_{k,t}\geq 0,\\ (\mathfrak{q}^1_{s,t},\ldots,\mathfrak{q}^J_{s,t})_{t=1}^T\geq 0,\\ (\mathfrak{P}^1_{k,t},\ldots,\mathfrak{P}^J_{k,t})_{t=1}^T\geq 0}} \sum_{t=J}^{T}\left(\sum_{j=t-J+1}^{t}\left(\mathbf{p}_{k,j}\mathfrak{q}^{t-j+1}_{k,j}+\mathfrak{P}^{t-j+1}_{k,j}\mathbf{Q}_{k,j}\right)-c_{k,t}\right) \tag{4.10a}$$

$$\text{s.t. } c_{k,t}\leq \sum_{j=t-J+1}^{t}\mathbf{p}_{k,j}\mathfrak{q}^{t-j+1}_{s,j}+\mathfrak{P}^{t-j+1}_{k,j}\mathbf{Q}_{s,j} \qquad \forall s\in D^t_k,\ \forall t=J,\ldots,T, \tag{4.10b}$$

$$\sum_{j=1}^{J}\mathfrak{q}^j_{s,t}=\mathbf{q}_{s,t} \qquad \forall s\in D^t_k,\ \forall t=1,\ldots,T, \tag{4.10c}$$

$$\sum_{j=1}^{J}\mathfrak{P}^j_{k,t}=\mathbf{P}_{k,t} \qquad \forall t=1,\ldots,T, \tag{4.10d}$$

In this program, the objective minimizes firm k's cost inefficiency (as defined in (4.7)) in terms of the chosen allocation and write-off schemes. The first constraint imposes that $c_{k,t}$ effectively represents the minimal cost to produce the output $\mathbf{y}_{k,t}$ (over the set D^t_k). The second and third constraints impose the adding-up restrictions that apply to feasible specifications of $(\mathfrak{q}^1_{k,t},\mathfrak{q}^2_{k,t},\ldots,\mathfrak{q}^J_{k,t})$ and $(\mathfrak{P}^1_{k,t},\mathfrak{P}^2_{k,t},\ldots,\mathfrak{P}^J_{k,t})$. Intuitively, cost inefficiency occurs as soon as some other firm s is characterized by a lower production cost than firm k no matter what allocations and write-offs are used.[15]

The allocations and write-off schemes of inputs acquired in periods $[T-J+2,\ldots,T]$ deserve some discussion at this point. These inputs are used beyond the time horizon T (until $T+J-1$ for inputs acquired in period T) and for them we can predict the optimal allocation of the LP solver. Since shifted cost allocations to periods $[T+1,\ldots,T+J-1]$ do not enter the objective of (4.10), optimal choices for firm k's allocations and write-offs of $[T-J+2,\ldots,T]$ consist of distributing these costs entirely over the periods beyond T. Thus, firm k is efficient by default for the periods $[T-J+2,\ldots,T]$. This is an inherent feature of the benefit-of-the-doubt idea. Furthermore, this prediction becomes invalid when partial information on allocations and write-offs is added.

Recall from Section 4.3 that we cannot evaluate dynamic cost efficiency for periods $[1,\ldots,J-1]$ due to data limitations. Together with our discussion in the previous paragraph, this implies that we can only effectively discriminate among firms within the

[15] Similarly as in conventional multiplier formulations of DEA, the allocations and write-offs that solve (4.10) are not necessarily unique. As a result, the efficiency decomposition in (4.8) is also not necessarily unique. However, one can use a similar procedure as in Kao and Hwang (2008) to find a unique decomposition by maximizing per-period efficiency while maintaining overall efficiency.

time window $[J, \ldots, T-J+1]$. Note the impact of both J and T on the size of this time window: the larger J (T), the smaller (larger) this time window and the more (fewer) periods of efficient by default. This reveals a clear trade-off between J and T. While T is often determined by the data at hand, one can vary J to check sensitivity of the results.

4.5.2 Partial information on allocations and write-offs

In many practical situations, it is possible to put some additional restrictions on the feasible allocation and write-off schemes. Such partial information can be incorporated by suitably specifying $\Theta(\mathbf{A}, \mathbf{b})$. Correspondingly, we can append the restriction

$$(\mathfrak{q}^1_{s,t}, \ldots, \mathfrak{q}^J_{s,t}, \mathfrak{P}^1_{k,t}, \ldots, \mathfrak{P}^J_{k,t})^T_{t=1} \in \Theta(\mathbf{A}, \mathbf{b}) \tag{4.11}$$

to program (4.10), and solve the resulting (linear) problem. Clearly, by using restriction (4.11) we constrain the solution space, which will generally result in higher values of the computed cost inefficiencies.

To take a specific instance, let $(\mathfrak{q}^{u,A}_{k,v})_{u \in U \subseteq [1,\ldots,J]}$ represent lower bounds on the quantity allocations of the storable inputs for firm k and time period(s) $v \in V \subseteq [J, \ldots, T]$. Similarly, let $(\mathfrak{P}^{w,A}_{k,z})_{w \in W \subseteq [1,\ldots,J]}$ be known lower bounds on the price write-offs of the durable inputs for time period(s) $z \in Z \subseteq [J, \ldots, T]$. We then define

$$\Theta = \Big\{ \mathfrak{q}^u_{k,v} \geq \mathfrak{q}^{u,A}_{k,v},\ \forall u \in U,\ \forall v \in V \\ \mathfrak{P}^w_{k,z} \geq \mathfrak{P}^{w,A}_{k,z},\ \forall w \in W,\ \forall z \in Z \Big\}.$$

As a limiting case, instantaneous input consumption complies with $\mathfrak{q}^{u,A}_{k,v} = (\mathbf{q}_{k,v}, 0, \ldots, 0)$ or, equivalently, $\mathfrak{P}^{w,A}_{k,z} = (\mathbf{P}_{k,z}, 0, \ldots, 0)$.

4.5.3 Write-off hypotheses

By using this approach, we can actually include (and check) alternative hypotheses regarding the allocation of the durable costs to individual time periods (i.e., specific write-off schemes). On the one hand, write-off schemes are often dictated by standard accounting practices for specific durable inputs so that write-off schemes are public information. Imposing these write-off schemes in the LP then allows to check whether these write-off schemes are cost efficient for a firm. On the other hand, durable inputs are often an aggregate of multiple durable inputs (cfr. "capital" in the empirical illustration) so that it is unclear a priori what the exact write-off scheme is. Imposing alternative write-off hypotheses in the LP can then help to clarify this. For example, it might often be reasonable to assume that the firm's valuation of a durable input diminishes over time. In our framework, this corresponds to

$$\mathfrak{P}^1_{k,t} \geq \mathfrak{P}^2_{k,t} \geq \ldots \geq \mathfrak{P}^J_{k,t}, \tag{4.12}$$

which complies with a degressive write-off of investment costs. From our above explanation, it follows that this is also consistent with the assumption of technological improvement, where older machines are scrapped and replaced by newer – technologically improved – ones over time.

Alternatively, a linear write-off of investment corresponds to

$$\mathfrak{P}^1_{k,t} = \mathfrak{P}^2_{k,t} = \ldots = \mathfrak{P}^J_{k,t}, \tag{4.13}$$

implying that $\mathfrak{P}^j_{k,t} = \mathbf{P}_{k,t}/J \ \forall j = 1, \ldots, J$. Both hypotheses can be tested by adding (4.12) or (4.13) for $t = J, \ldots, T$ to $\Theta(\mathbf{A}, \mathbf{b})$.

4.6 Extensions

We next focus on a number of extensions of our basic framework set out in the previous section. These extensions highlight the versatility of our framework and, of course, are not exhaustive. First, we show how to convert our (difference) cost inefficiency measures (4.6) and (4.7) in ratio form. Then, we discuss the extension of our framework to allow for heterogeneous input lifetime and production delays. Finally, we indicate how to proceed in the absence of input price information by applying shadow pricing. Here, we will also explain the decomposition of cost inefficiency as defined above in terms of technical and allocative inefficiency. We will illustrate the different extensions in our empirical application in Section 4.7.

4.6.1 Ratio measures of inefficiency

A downside of our cost inefficiency measure in (4.7) is that it is not invariant to rescaling of prices and inputs. However, one can turn this difference measure into a ratio measure of cost inefficiency by an appropriate normalization. In principle, a multitude of normalizations are possible. A natural choice is to divide by the actual cost, i.e.,

$$RCE_k \equiv \frac{CE_k}{\sum_{t=J}^{T} \sum_{j=t-J+1}^{t} \left(\mathbf{p}_{k,j} \mathfrak{q}^{t-j+1}_{k,j} + \mathfrak{P}^{t-j+1}_{k,j} \mathbf{Q}_{k,j} \right)} \tag{4.14}$$

This relative measure is situated between 0 and 1 and expresses the proportion of total production costs that can be saved by minimizing total production costs over the periods $[J, \ldots, T]$.[16]

Analogously to (4.8), we can decompose this overall ratio measure in terms of per-period measures.[17] In this case, we have that RCE_k equals a weighted sum of per-period cost inefficiencies in ratio form RCE^t_k. Specifically, it uses the period-specific weights

$$w^t_k \equiv \frac{\sum_{j=t-J+1}^{t} \left(\mathbf{p}_{k,j} \mathfrak{q}^{t-j+1}_{k,j} + \mathfrak{P}^{t-j+1}_{k,j} \mathbf{Q}_{k,j} \right)}{\sum_{t=J}^{T} \sum_{j=t-J+1}^{t} \left(\mathbf{p}_{k,j} \mathfrak{q}^{t-j+1}_{k,j} + \mathfrak{P}^{t-j+1}_{k,j} \mathbf{Q}_{k,j} \right)}, \tag{4.15}$$

[16] This normalization mirrors the one used by Chambers et al. (1998) for profit efficiency.

[17] The following decomposition parallels Färe and Zelenyuk (2003)'s decomposition of industry revenue efficiency as a weighted sum of firms' revenue efficiency.

which represent the proportions of total production costs allocated to every period t. This obtains

$$\begin{aligned}
RCE_k &= \sum_{t=J}^{T} w_k^t \frac{CE_k^t}{\sum_{j=t-J+1}^{t}\left(\mathbf{p}_{k,j}\mathfrak{q}_{k,j}^{t-j+1} + \mathfrak{P}_{k,j}^{t-j+1}\mathbf{Q}_{k,j}\right)} \\
&= \sum_{t=J}^{T} w_k^t \left(1 - \frac{c_{k,t}}{\sum_{j=t-J+1}^{t}\left(\mathbf{p}_{k,j}\mathfrak{q}_{k,j}^{t-j+1} + \mathfrak{P}_{k,j}^{t-j+1}\mathbf{Q}_{k,j}\right)}\right) \\
&= \sum_{t=J}^{T} w_k^t RCE_k^t, \qquad (4.16)
\end{aligned}$$

As a final note, we indicate that $1 - RCE_k$ and $1 - RCE_k^t$ give the conventional cost efficiency measures, i.e., minimal cost divided by actual cost.

Alternatively, (4.14) can be decomposed as a combination of a relative measure of static cost saving and a static misallocation measure:

$$\begin{aligned}
RCE_k = &\frac{CE_k}{\sum_{t=1}^{T}\left(\mathbf{p}_{k,t}\mathbf{q}_{k,t} + \mathbf{P}_{k,t}\mathbf{Q}_{k,t}\right)} \times \frac{\sum_{t=1}^{T}\left(\mathbf{p}_{k,t}\mathbf{q}_{k,t} + \mathbf{P}_{k,t}\mathbf{Q}_{k,t}\right)}{\sum_{t=J}^{T}\sum_{j=t-J+1}^{t}\left(\mathbf{p}_{k,j}\mathfrak{q}_{k,j}^{t-j+1} + \mathfrak{P}_{k,j}^{t-j+1}\mathbf{Q}_{k,j}\right)} \\
&= \text{static cost saving} \times \underbrace{\text{static misallocation}}_{\geq 1} \qquad (4.17)
\end{aligned}$$

The first term measures the potential cost saving as a fraction of total observed costs while the second term shows the fraction of observed costs over total production costs. This second term highlights potential mismatch of static and dynamic cost efficiency measures. One can further decompose (4.17) similarly as in (4.16).

4.6.2 Heterogeneous input lifetime and production delays

Until now, we have assumed that all durable inputs have the same lifetime J. Admittedly, this may sometimes be a too strong assumption. In addition, our current specification does not allow for production delays. As we show next, we can solve both issues by making use of the concept of delay matrices.

Specifically, let $\mathbf{D}^D = (\mathbf{d}_1^D, \ldots, \mathbf{d}_J^D) \in \{0,1\}^{M \times J}$ denote a binary delay matrix, where each row represents a durable input. For example, for the durable input m we may use one of the following specifications:

- $(1, \underbrace{0, \ldots, 0}_{J-1})$ if input m is an instantaneous input;
- $(1, \ldots, 1)$ if input m is a durable input with lifetime J;
- $(\underbrace{0, \ldots, 0}_{U}, \underbrace{1, \ldots, 1}_{J-U})$ if input m is a durable input usable after a delay of $U < J$ periods with a lifetime of $J - U$ periods;

- $(\underbrace{1,\dots,1}_{U},\underbrace{0,\dots,0}_{J-U})$ if input m is a durable input usable over $U < J$ periods.

Similarly, let $\mathbf{D}^S = (\mathbf{d}_1^S, \dots, \mathbf{d}_J^S) \in \{0,1\}^{N\times J}$ represent a binary delay matrix for the storable inputs. Then, we can formulate the next modified optimization problem of firm k:

$$\min_{\substack{(\mathfrak{q}_t^1,\dots,\mathfrak{q}_t^J)_{t=1}^T\\ (\mathfrak{P}_t^1,\dots,\mathfrak{P}_t^J)_{t=1}^T}} \sum_{t=J}^{T}\sum_{j=t-J+1}^{t} \mathbf{p}_{k,j}\mathbf{d}_{t-j+1}^S\mathfrak{q}_j^{t-j+1} + \mathfrak{P}_j^{t-j+1}\mathbf{d}_{t-j+1}^D\mathbf{Q}_{k,j} \tag{4.18a}$$

$$\text{s.t.}\quad \left(\sum_{j=1}^{J}\mathbf{d}_j^S\mathfrak{q}_{t-j+1}^j, \sum_{j=1}^{J}\mathbf{d}_j^D\mathbf{Q}_{k,t-j+1}\right) \in \mathcal{I}_t(\mathbf{y}_{k,t}) \qquad \forall t = J,\dots,T \tag{4.18b}$$

Closer inspection reveals that $\mathbf{d}_j^S\mathfrak{q}_{t-j+1}^j$ and $\mathbf{d}_j^D\mathbf{Q}_{k,t-j+1}$ only selects those inputs that are usable in period j. In this case, J stands for the maximum lifetime over all durable and storable inputs. We further note that we can include a resale value for all durable inputs simply by setting J to maximum lifetime + 1. The final write-off of the durable inputs is then the resale value of the durable inputs which can be subtracted from the per-period cost to reflect resale of the durable inputs.

The associated linear program is

$$\min_{\substack{c_{k,t}\geq 0,\\ (\mathfrak{q}_{s,t}^1,\dots,\mathfrak{q}_{s,t}^J)_{t=1}^T\geq 0,\\ (\mathfrak{P}_{k,t}^1,\dots,\mathfrak{P}_{k,t}^J)_{t=1}^T\geq 0}} \sum_{t=J}^{T}\left(\sum_{j=t-J+1}^{t}\left(\mathbf{p}_{k,j}\mathbf{d}_{t-j+1}^S\mathfrak{q}_{k,j}^{t-j+1} + \mathfrak{P}_{k,j}^{t-j+1}\mathbf{d}_{t-j+1}^D\mathbf{Q}_{k,j}\right) - c_{k,t}\right) \tag{4.19a}$$

$$\text{s.t.}\ c_{k,t} \leq \sum_{j=t-J+1}^{t}\mathbf{p}_{k,j}\mathfrak{q}_{s,j}^{t-j+1}\mathbf{d}_{t-j+1}^S + \mathfrak{P}_{k,j}^{t-j+1}\mathbf{d}_{t-j+1}^D\mathbf{Q}_{s,j} \qquad \forall s \in D_k^t,\ \forall t = J,\dots,T, \tag{4.19b}$$

$$\sum_{j=1}^{J}\mathfrak{q}_{s,t}^j\mathbf{d}_j^S = \mathbf{q}_{s,t} \qquad \forall s \in D_k^t,\ \forall t = 1,\dots,T, \tag{4.19c}$$

$$\sum_{j=1}^{J}\mathfrak{P}_{k,t}^j\mathbf{d}_j^D = \mathbf{P}_{k,t} \qquad \forall t = 1,\dots,T, \tag{4.19d}$$

$$(\mathfrak{q}_{s,t}^1,\dots,\mathfrak{q}_{s,t}^J,\mathfrak{P}_{k,t}^1,\dots,\mathfrak{P}_{k,t}^J)_{t=1}^T \in \Theta(\mathbf{A},\mathbf{b}). \tag{4.19e}$$

It is easy to verify that this program reduces to (4.10) for $\mathbf{D}^S = \mathbb{1}_{N\times J}$ and $\mathbf{D}^D = \mathbb{1}_{M\times J}$. Furthermore, any zero values in $\mathbf{D}^S$ and $\mathbf{D}^D$ immediately imply zero values for the corresponding allocations and write-offs. In other words, the use of delay matrices allows us to impose a priori restrictions on the storable allocations and durable write-offs. We also remark that, in principle, we can specify firm-specific delay matrices if this seems desirable.

4.6.3 Shadow prices and technical inefficiency

So far, we have focused on economic (cost) efficiency, which requires price information for the relevant inputs. By contrast, technical efficiency analysis does not require such price information and, thus, can be used if limited price information is available.

Generally, technical efficiency criteria/measures can be characterized as economic efficiency criteria/measures evaluated at so-called "shadow prices".[18] Thus, by establishing the shadow price representation of our dynamic efficiency concepts, we can define technical efficiency notions that explicitly account for the dynamic (storable and durable) nature of the inputs.

It is easy to use shadow pricing if the exact allocation of storable inputs over time periods (i.e., $(\mathfrak{q}^1_{s,t}, \ldots, \mathfrak{q}^J_{s,t})_{t=1}^T$) is known to the empirical analyst. In that case, it suffices to solve (4.19) with the input prices $(\mathbf{P}_{k,t})_{t=1}^T$ and $(\mathbf{p}_{k,t})_{t=1}^T$ as additional free variables that are subject to a non-negativity constraint and the normalization

$$\sum_{t=J}^{T}\left(\sum_{j=t-J+1}^{t} \mathbf{p}_{k,j}\mathfrak{q}_{k,j}^{t-j+1}\mathbf{d}_{t-j+1}^{S} + \mathfrak{P}_{k,j}^{t-j+1}\mathbf{d}_{t-j+1}^{D}\mathbf{Q}_{k,j}\right) = 1.$$

We remark that this price normalization implies $CE_k = RCE_k$. The resulting shadow cost inefficiency measure (or technical inefficiency measure TE_k) measures cost efficiency using shadow prices that give the benefit-of-the-doubt to the firm. Thus, these shadow prices are chosen as to maximize the shadow cost efficiency of the firm.

Let TE_k represent the "technical inefficiency" measure that is obtained as the solution of the resulting linear program. By construction, we have $TE_k \leq RCE_k$. The difference between TE_k and RCE_k gives us a measure AE_k of "allocative inefficiency", i.e.,

$$TE_k + AE_K = RCE_k \Leftrightarrow AE_k = RCE_k - TE_k. \tag{4.20}$$

Allocative inefficiency is the difference between cost inefficiency using market prices and shadow cost inefficiency using optimal shadow prices. This shows that reducing cost inefficiency can be achieved by a combination of two effects: (i) reducing technical

[18] In DEA terminology, this shadow price characterization of technical efficiency corresponds to the "multiplier" formulation of DEA models. Practical applications often make use of DEA models in "envelopment" form, which is dual to this multiplier formulation. In our set-up, the envelopment formulation can be obtained as the dual of the linear program that we define below (to compute TE_k). We refer to Färe and Primont (1995) for a more detailed discussion.

inefficiency by minimizing the level of inputs conditional on outputs and (ii) selecting the cost minimizing input mix out of all technically efficient inputs. Allocative inefficiency can occur when the realized a posteriori prices deviate from the expected a priori input prices. In our context of durable and storable inputs, intertemporal dependencies can exacerbate this allocative inefficiency: buying storable and/or durable inputs right before a persistent price drop can cause a persistent per-period inefficiency over subsequent periods, because the cost of these inputs is distributed over J periods. Thus, one slip up can resonate for many subsequent time periods.

Interestingly, because $RCE_k = CE_k$ under our price normalization and using (4.7), we can also decompose AE_k in period-specific allocative inefficiencies, as follows:

$$AE_k \equiv \sum_{t=J}^{T} \left(RCE_k^t - TE_k^t \right) = \sum_{t=J}^{T} AE_k^t. \tag{4.21}$$

Finally, matters are more complicated when the allocation $(\mathbf{q}_{s,t}^1, \ldots, \mathbf{q}_{s,t}^J)_{t=1}^T$ is unobserved. In that case, the analogue of the programming problem (4.19) becomes nonlinear in unknown prices and quantities. We can restore linearity by making specific assumptions regarding the storable input allocation. For example, if we are willing to assume that all DMUs allocate their storable inputs in the same way over time, then we can use a similar procedure as outlined in Cook et al. (2000) and Cherchye et al. (2013).

4.7 Empirical illustration

We apply our model to a panel dataset of Swiss regional railway companies that was also studied by Farsi et al. (2005).[19] The original panel is unbalanced and contains yearly information on 50 railway companies over the period 1985 – 1997. From this dataset, we constructed a balanced panel that covers the 13-year period for 37 companies, and which contains all input and output information needed to apply our methodology. The constructed balanced panel contains 481 ($= 13 \times 37$) firm observations.

In what follows, we first motivate our selection of outputs and inputs, where we will use capital expenses as a durable input. Then, we present the results of our empirical analysis, which will mainly focus on overall and per-period efficiencies as well as technical and allocative inefficiency. We conclude by a number of sensitivity checks.

4.7.1 Output and input specification

The original dataset contains information on total expenses, labor and energy expenses, as well as the total number of employees, electricity consumption, network length, total number of available seats, total number of train-kilometers, passenger-kilometers and ton-kilometers. Capital expenses are defined as the residual after deducting labor and energy from the total expenses. Prices for labor and energy are found by dividing

[19] The data are available at `http://people.stern.nyu.edu/wgreene/Econometrics/PanelDataSets.htm`, which also contains a detailed description of all variables.

(labor and energy) expenses by total quantity (number of employees and total electricity consumption in kWh, respectively). The price of capital is defined by dividing capital expenses by total number of seats.

Following Farsi et al. (2005), we use total passenger-kilometers and ton-kilometers as our two outputs in our following analysis. Next, we have three inputs: labor, energy and capital expenses. We refer to Farsi et al. (2005) for more details on the data. Table 4.2 presents summary statistics. The most important observation is that labor and capital expenditures are the major costs (52.82% and 43.41% on average), while energy expenditures represents only a small fraction of the total cost (i.e., 3.77% on average).

	mean	std	median	min	max	share (in %)
Passenger output in passenger kilometers (Q2) ($\times 10^8$)	0.2843	0.5192	0.0919	0.0041	3.1100	95.27
Goods output in ton kilometers (Q3) ($\times 10^7$)	0.2135	0.8158	0.0226	0.0000	5.9400	4.73
Length of railway network in km (NETWORK)	39.5340	61.0973	22.8200	3.8980	376.9970	n.a.
Number of stations on the network (STOPS)	20.8274	20.1613	15.0000	4.0000	121.0000	n.a.
Labor price adjusted for inflation (PL) ($\times 10^5$)	0.8550	0.0602	0.8557	0.6093	1.0493	n.a.
Number of employees (STAFF) ($\times 10^3$)	0.1401	0.2517	0.0520	0.0120	1.6410	n.a.
Labor expenditures (LABOREXP) ($\times 10^7$)	1.2189	2.1945	0.4406	0.0985	14.6988	52.82
Price of electricity in CHF per kWh (PE)	0.1574	0.0240	0.1580	0.0763	0.2652	n.a.
Total consumed electricity in kWh (KWH) ($\times 10^7$)	0.5775	1.0317	0.1980	0.0082	6.5849	n.a.
Energy expenditures (ELECEXP) ($\times 10^5$)	8.4849	12.9842	3.0220	0.1400	81.0408	3.77
Capital price per seat (PK) ($\times 10^6$)	0.2182	0.3644	0.0872	0.0212	2.4105	n.a.
Quantity of Capital (CAPITAL)	43.4092	9.4026	41.7978	23.8892	77.3315	n.a.
Capital expenditures ($\times 10^6$)	8.7849	13.4185	3.9922	0.6119	87.9753	43.41
Total costs adjusted for inflation (CT) ($\times 10^8$)	0.2182	0.3644	0.0872	0.0212	2.4105	100

Table 4.2: Summary statistics of the railway data (481 observations).

As indicated above, capital expenditures form a prime example of durable inputs. Therefore, while we consider labor and energy expenditures as instantaneously consumed (i.e., not storable or durable), we will treat capital expenditures as durable. For the general model specification (with capital usable in J years), this obtains the $3 \times J$ delay matrix

$$\mathbf{D}^D = \begin{pmatrix} 1 & 0 \ldots 0 \\ 1 & 0 \ldots 0 \\ 1 & \underbrace{1 \ldots 1}_{J-1} \end{pmatrix},$$

with labor, energy and capital corresponding to rows 1, 2 and 3, respectively.

Table 4.3 reports summary statistics on our durable input. We see that, in nominal terms, capital costs are steadily increasing until 1991. Within individual years, we also observe considerable variation across firms. The magnitude of this variation is fairly stable over time.

It follows from our discussion in the previous sections that treating capital expenses as a durable input requires us to use discounted prices, and to specify the lifetime of capital (i.e., J). In our application all provided prices are adjusted for inflation with respect to 1997 prices. Next, capital costs are related to equipment as well as materials. This makes it hard to specify the exact lifetime of this durable input. For this reason,

year	mean	std	min	max
85	7821.9918	13415.8909	675.5952	79364.5942
86	7947.3623	13476.0897	611.8980	79606.2277
87	8226.9291	13474.9261	614.8458	78815.1313
88	8769.3795	13483.8216	725.2748	79002.7419
89	9293.5348	13890.6071	1104.4175	79672.0210
90	9531.1382	14310.4937	833.1193	82232.3410
91	9683.8843	14759.9689	991.5967	86268.9002
92	9541.3585	14988.2197	949.5515	87975.3422
93	8868.5766	13414.5760	940.9871	78469.9559
94	8709.7330	13665.2525	704.6633	80731.6209
95	8707.0591	12669.9599	691.9920	73890.1362
96	8396.3872	12086.1047	780.2419	70634.0593
97	8706.7027	12561.9671	873.0000	73087.0000

Table 4.3: Capital costs: summary statistics per year (in 1000 CHF).

and to clearly demonstrate the potential impact of intertemporal dependencies between inputs, we will mainly focus on a minimalistic scenario with $J = 2$. As an additional exercise, we will also consider alternative values for J, to check robustness of our main conclusions.

4.7.2 Cost efficiency analysis

In summary, the dynamic nature of our empirical analysis relates to a single durable input, capital expenses. Moreover, this input represents a fairly large fraction of the total cost relative to the instantaneous inputs, labor and energy expenses (see our discussion of Table 4.2). In what follows, we will show that ignoring the intertemporal (durable) aspect of capital can substantially affect the efficiency analysis. In turn, referring to our discussion in the Introduction, this can considerably distort regulatory policies (in our case for Swiss railway companies) that are based on the efficiency results. Obviously, these distortions will generally be more pronounced in production settings where durable inputs form an even more important fraction of total costs, and in settings with storable inputs in addition to durable inputs.

We first consider the differences in cost inefficiencies between the dynamic and static setting. Figure 4.1 shows the differences between the CE_k^t-values for our dynamic model (with $J = 2$) and the static model (which corresponds to $J = 1$). The differences in per-period inefficiencies are quite substantial: in some years (such as 89, 92 and 94) ignoring intertemporal effects leads to an underestimation of productive inefficiency by as much as 4 million CHF, while in other years (e.g., 88, 90 and 93) it leads to an overestimation by no less than 6 million CHF. The differences are statistically significant: comparing the cumulative density functions of CE_k^t using a Kolmogorov-Smirnov test, we reject at the 1% significance level the hypothesis that both densities have the same underlying distribution.[20]

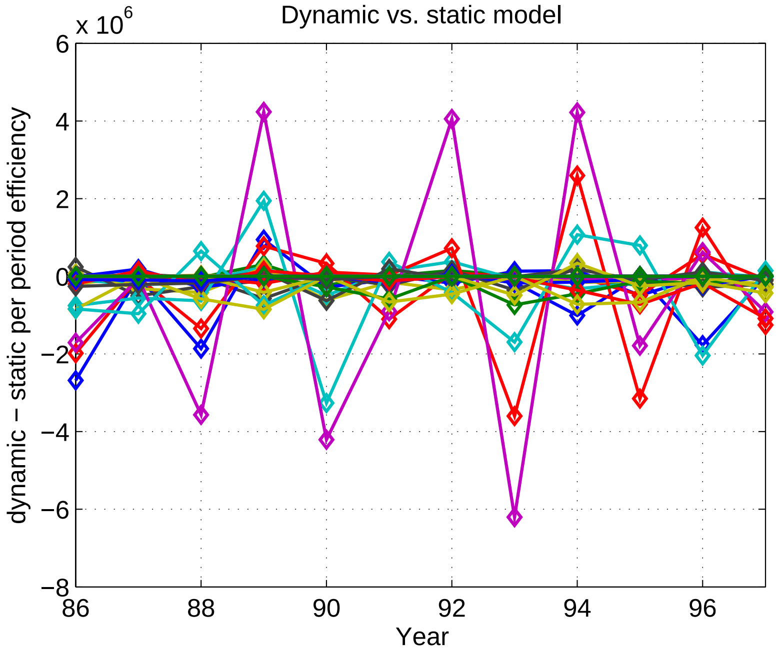


Figure 4.1: Dynamic ($J = 2$) vs. static ($J = 1$) CE_k^t.

Next, we turn to the ratio measure RCE_k that we defined in (4.14). Comparing the results for $J = 1$ with those for $J = 2$ provides further insight into the severity and frequency of disagreement between the dynamic and static inefficiency models. Figure 4.2 depicts a histogram of the differences in RCE_k^t-values. Both models agree in terms of RCE_k^t in 61.94% of all cases. For the cases where they do not agree, this histogram shows that inefficiency is more often overestimated by the static model: in 13.02% of all cases the static model overestimates inefficiency by at least 5%, while inefficiency is underestimated by at least 5% in only 4.14% of the cases.

Figure 4.3 decomposes RCE_k^t in a static cost saving and a static misallocation component (as in (4.17)). The former represents the relative cost saving with regard to total observed costs, while the latter shows the discrepancy between observed and allocated costs. These histograms show that 3.76% – 13.39% of total costs is not allocated over the 12 years under the dynamic model.

Next, we can classify the firms based on the evolution of RCE_k^t over time. We distinguish 4 different categories: (i) persistently efficient firms over time (i.e., $\forall t : RCE_k^t \leq 10^{-8}$); (ii) mostly improving firms (i.e., $\sum_{t=1987}^{1997} \left[RCE_k^{t-1} - RCE_k^t > 0 \right] \geq 6$) that are not persistently efficient; (iii) mostly non-improving firms that are not persistently inefficient and (iv) persistently inefficient firms (i.e., $\forall t : RCE_k^t \geq 10^{-2}$). Table 4.4 and

[20]The value of the test statistic is 0.1351, which corresponds to a p-value of 5.24×10^{-4}.

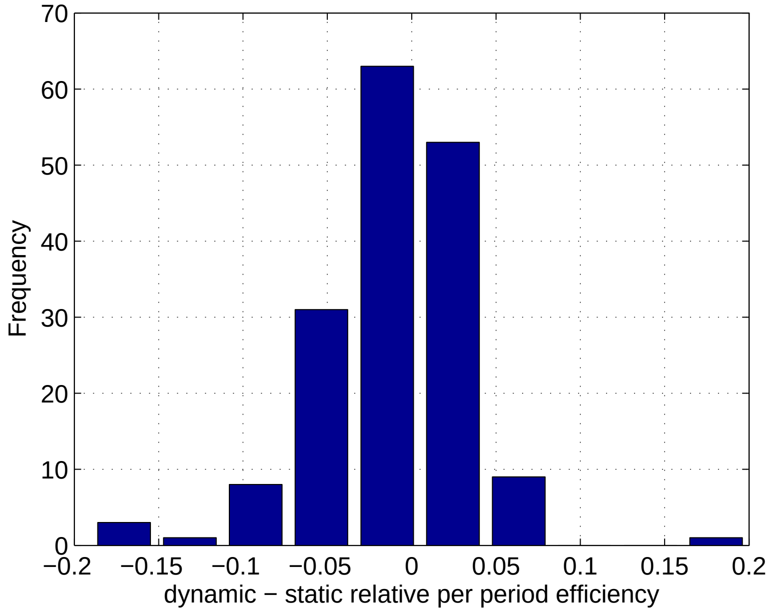


Figure 4.2: Histogram of dynamic - static RCE_k^t (> 0).

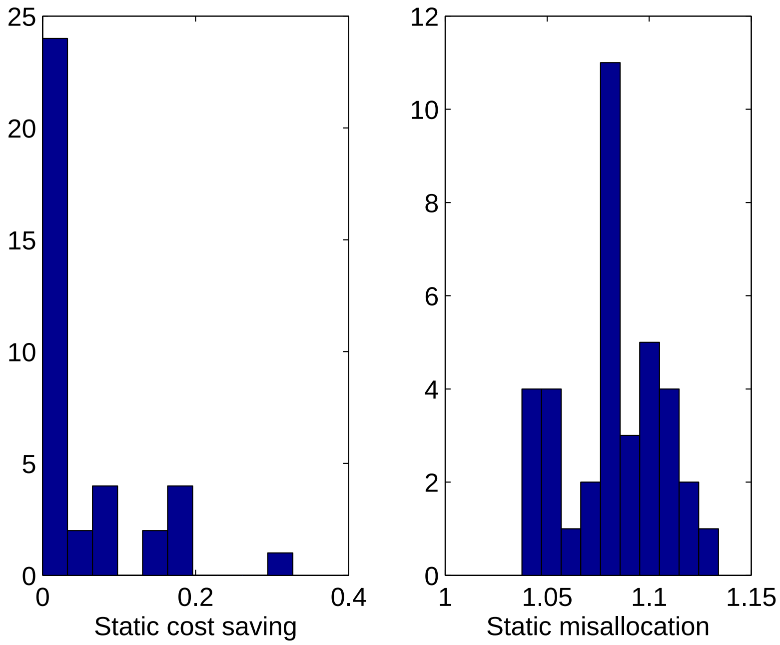


Figure 4.3: Decomposition of RCE_k^t in static cost saving and static misallocation components.

Figure 4.4 show this classification of every firm. This analysis reveals that a majority of firms ($22/37 \approx 59\%$) is persistently efficient or mostly improving over time.

	firm id
Persistently efficient firms	4, 5, 10, 14, 18, 23, 26, 31, 34, 37, 39, 41, 42, 45, 49
Mostly improving firms (i.e., at least for 6/12 periods)	2, 3, 9, 12, 24, 27, 46
Mostly non-improving firms (i.e., at least for 6/12 periods)	6, 7, 8, 21, 36, 43, 47, 48
Persistently inefficient firms	13, 15, 16, 17, 20, 22, 30

Table 4.4: Firm classification ($J = 2$).

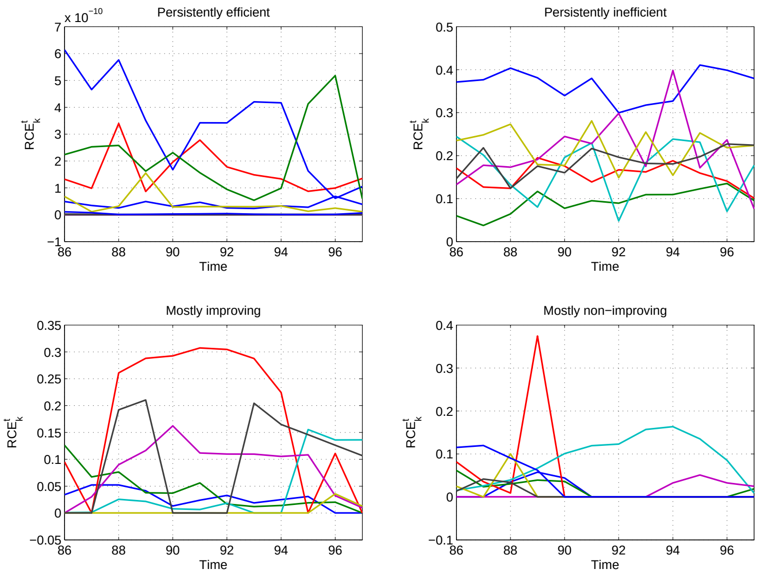


Figure 4.4: Classification of firms based on RCE_k^t over time: (i) persistently efficient; (ii) persistently inefficient; (iii) mostly improving and (iv) mostly non-improving.

As a following exercise, we redo our analysis by using shadow prices. As explained above, this effectively computes the technical inefficiency measure TE_k, which we can further use to calculate the aggregate allocative inefficiency measure AE_k (in (4.20)) as well as per-period allocative inefficiencies AE_k^t (in (4.21)). Table 4.5 shows the TE_k and AE_k results for all firms. We find that technical inefficiency is rather negligible for the firms under study: the maximal TE_k-value amounts to no more than 4.899×10^{-9}. In contrast, the AE_k-values are quite high for a number of firms: the worst performing firm has an allocative inefficiency of as much as 0.3636. Figure 4.5 also shows that there is substantial variation in the AE_k^t-values over time. One possible explanation for the large discrepancy between TE_k and RCE_k is that the assumption of perfect price foresight is too strong.

firm id	TE_k	rank	AE_k	rank
2	0,0000	26	0,0276	23
3	0,0000	24	0,0399	25
4	0,0000	16	0,0000	11
5	0,0000	8	0,0000	5
6	0,0000	17	0,0132	19
7	0,0000	23	0,0170	21
8	0,0000	28	0,0267	22
9	0,0000	29	0,1993	34
10	0,0000	5	0,0000	13
12	0,0000	25	0,0419	26
13	0,0000	36	0,3636	37
14	0,0000	33	0,0000	2
15	0,0000	27	0,0938	30
16	0,0000	37	0,1528	31
17	0,0000	10	0,1605	32
18	0,0000	31	0,0000	3
20	0,0000	3	0,2009	35
21	0,0000	30	0,0883	28
22	0,0000	6	0,2082	36
23	0,0000	2	0,0000	8
24	0,0000	20	0,0846	27
26	0,0000	22	0,0000	9
27	0,0000	9	0,0041	16
30	0,0000	32	0,1862	33
31	0,0000	1	0,0000	15
34	0,0000	35	0,0000	14
36	0,0000	11	0,0127	18
37	0,0000	7	0,0000	6
39	0,0000	12	0,0000	7
41	0,0000	34	0,0000	1
42	0,0000	21	0,0000	12
43	0,0000	15	0,0139	20
45	0,0000	13	0,0000	4
46	0,0000	4	0,0937	29
47	0,0000	18	0,0082	17
48	0,0000	14	0,0307	24
49	0,0000	19	0,0000	10

Table 4.5: TE_k and AE_k.

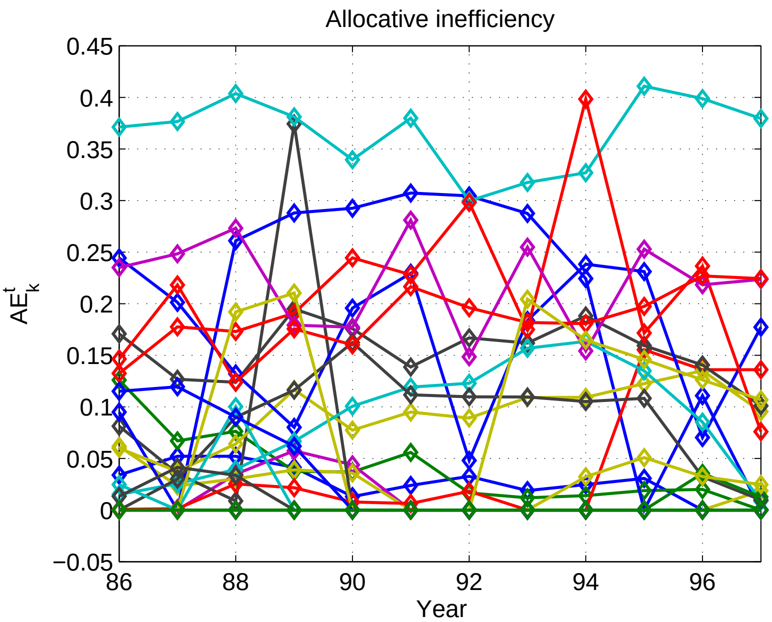


Figure 4.5: Allocative inefficiencies AE_k^t.

4.7.3 Robustness checks

We conclude our empirical analysis by conducting a number of robustness checks. These additional exercises will further illustrate the versatility of our general framework, in terms of relaxing or imposing particular assumptions in the intertemporal efficiency assessment. First, we consider the effect of specifying J on our results. Second, we investigate the effects of different environmental variables on the efficiency scores. Finally, we compute efficiency results when imposing particular (degressive and linear) structure on the write-off schemes used by the evaluated firms.

We begin by evaluating the sensitivity of the overall dynamic inefficiencies of our 37 firms to alternative specifications of J (> 1). Table 4.6 shows the scores and the relative rankings of all firms for varying choices of J. An interesting observation is that, although we observe some changes for different J-values, the firm rankings remain largely unchanged in general. In a similar vein, for most firms the inefficiency scores do not change much with J. This is confirmed by 6 Kolmogorov-Smirnov tests that verify equality of distribution of the inefficiency scores: all p-values ranged between $0.4787 - 0.9995$, so that we cannot reject the null hypothesis that the inefficiencies come from the same underlying distribution.

Next, we examine whether differences in firms' (observable) environments impact the efficiency results. In this respect, our dataset contains information on the length of the railway network (NETWORK) and the number of stations in the network (STOPS) (see Table 4.2). A priori, one may expect both variables to have a negative effect on efficiency, as larger networks can give rise to higher costs due to maintenance and, similarly, be-

firm id	J=2	rank	J=3	rank	J=4	rank	J=5	rank
2	0,0276	23	0,0215	23	0,0107	22	0,0016	22
3	0,0399	25	0,0323	25	0,0301	25	0,0236	25
4	0,0000	11	0,0000	12	0,0000	12	0,0000	7
5	0,0000	1	0,0000	16	0,0000	13	0,0000	1
6	0,0132	19	0,0129	22	0,0112	23	0,0064	23
7	0,0170	21	0,0091	21	0,0015	21	0,0004	21
8	0,0267	22	0,0233	24	0,0204	24	0,0220	24
9	0,1993	34	0,1899	35	0,1856	35	0,1568	34
10	0,0000	13	0,0000	18	0,0000	20	0,0000	19
12	0,0419	26	0,0407	26	0,0396	26	0,0406	26
13	0,3636	37	0,3677	37	0,3670	37	0,3616	37
14	0,0000	4	0,0000	9	0,0000	5	0,0000	16
15	0,0938	30	0,0813	28	0,0658	27	0,0572	27
16	0,1528	31	0,1428	32	0,1426	32	0,1379	32
17	0,1605	32	0,1421	31	0,1254	31	0,1148	31
18	0,0000	5	0,0000	7	0,0000	2	0,0000	20
20	0,2009	35	0,1847	33	0,1681	33	0,1553	33
21	0,0883	28	0,0855	29	0,0809	29	0,0672	28
22	0,2082	36	0,2092	36	0,2107	36	0,2081	36
23	0,0000	8	0,0000	10	0,0000	7	0,0000	6
24	0,0846	27	0,0799	27	0,0780	28	0,0682	29
26	0,0000	9	0,0000	6	0,0000	10	0,0000	8
27	0,0041	16	0,0000	8	0,0000	15	0,0000	5
30	0,1862	33	0,1854	34	0,1772	34	0,1788	35
31	0,0000	15	0,0000	3	0,0000	1	0,0000	10
34	0,0000	14	0,0000	15	0,0000	16	0,0000	9
36	0,0127	18	0,0000	2	0,0000	9	0,0000	13
37	0,0000	6	0,0000	5	0,0000	11	0,0000	4
39	0,0000	7	0,0000	11	0,0000	3	0,0000	14
41	0,0000	3	0,0000	13	0,0000	18	0,0000	3
42	0,0000	12	0,0000	19	0,0000	8	0,0000	17
43	0,0139	20	0,0000	17	0,0000	14	0,0000	2
45	0,0000	2	0,0000	14	0,0000	19	0,0000	11
46	0,0937	29	0,0916	30	0,0961	30	0,0895	30
47	0,0082	17	0,0000	1	0,0000	6	0,0000	12
48	0,0307	24	0,0032	20	0,0000	17	0,0000	18
49	0,0000	10	0,0000	4	0,0000	4	0,0000	15

Table 4.6: RCE_k for different J.

cause more stops imply additional (e.g., time-related) expenditures, all else equal. Next, the dataset contains a dummy variable indicating whether the network has a rack rail ("cremaillere"; represented by the binary variable RACK). Rack rails are special rails that are used to aid climbing of trains on steep terrain. Therefore, one may argue that the presence of rack rails effectively signals a less favorable operational environment.

We investigate how these variables affect our results by conducting, for each variable separately, an extra efficiency analysis in which we add the environmental variable $z_{s,t} \in \mathbb{R}_+$ as an additional output. Basically, this procedure implies that the dominating set (4.4) is modified to (only) include those peers that (1) produce at least the same output and (2) operate under the same, or worse, environmental conditions than the firm under examination. Following Ruggiero (1996), the modified dominating set is:

$$D_k^t = \left\{ s | \mathbf{y}_{s,t} \geq \mathbf{y}_{k,t} \right\} \cap \begin{cases} s | z_{s,t} = z_{k,t} \} & z_{s,t} \in \{0, 1\} \\ s | z_{s,t} \geq z_{k,t} \} & \text{Otherwise} \end{cases},$$

where $z_{s,t} \geq z_{k,t}$ implies s operates under worse conditions than k. By comparing these new efficiency results with the original ones for $J = 2$ (see Table 4.6), we can investigate the efficiency effect of the three contextual variables under study.

The results of these three exercises are summarized in Figures 4.6-4.7-4.8. Specifically, each of these figures sets out the firm ranks for the new exercises to the original firm ranks. Firms situated below the 45 degree line have a higher ranking (i.e., lower rank number) when the contextual variable (respectively, NETWORK, STOPS and RACK) is taken into account while, obviously, the opposite holds for firms above the 45 degree line. For each of our three environmental variables, we find that the firm ranks are fairly mildly affected, with the exception of a few firms. This is confirmed by Wilcoxon signed-rank tests, which check the statistical significance of the difference between the new and original rankings: it turns out that there is no significant difference for any of the three variables under evaluation.[21] We may thus conclude that none of our three contextual variables has a substantial effect on the efficiency patterns that we presented above.

Finally, at the end of Section 4.5 we indicated that an interesting feature of our methodology is that it allows for imposing specific hypotheses regarding the allocation of the costs of durables to individual time periods (i.e., putting structure on the write-off schemes). As a last robustness check, we compute efficiency results for the degressive scheme in (4.12) and the linear scheme in (4.13).

Our results are given in Table 4.7. We observe that, for a number of firms, the inefficiency values for the degressive write-off scheme are somewhat above the ones that we obtained in our original analysis (see the "Unconstrained" column), and the values for the linear scheme are always above those for the degressive scheme. Actually, this

[21] More specifically, the Wilcoxon signed-rank test compares the ranking of individual firms by first taking the difference in ranking. Observations with zero difference are dropped. These differences are then ranked and these ranks are summed. If there is no difference in ranking then the test statistic is zero. For NETWORKS the statistic equals 269 and the associated p-value is 0.3136, for STOPS the statistic is 258 and the p-value 0.1576, and for RACK the statistic amounts to 236 and the p-value 0.1246.

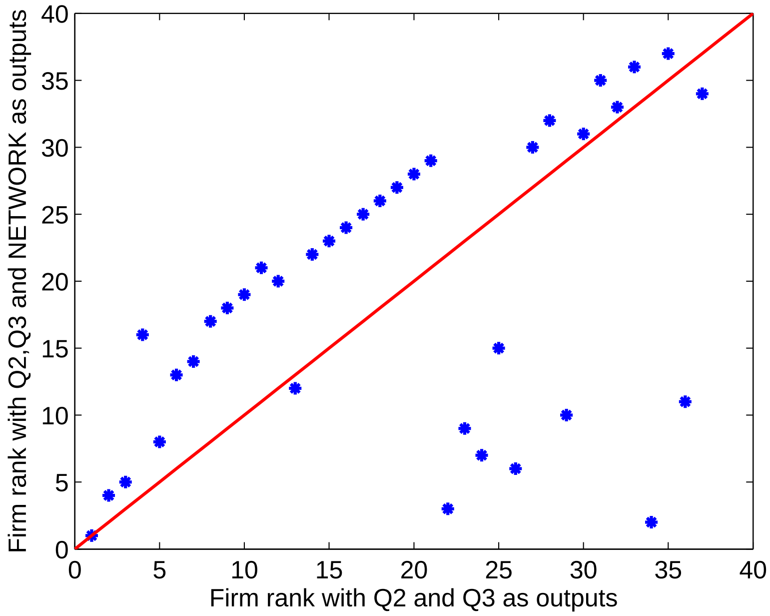


Figure 4.6: Firm rank comparison: with and without NETWORK as output.

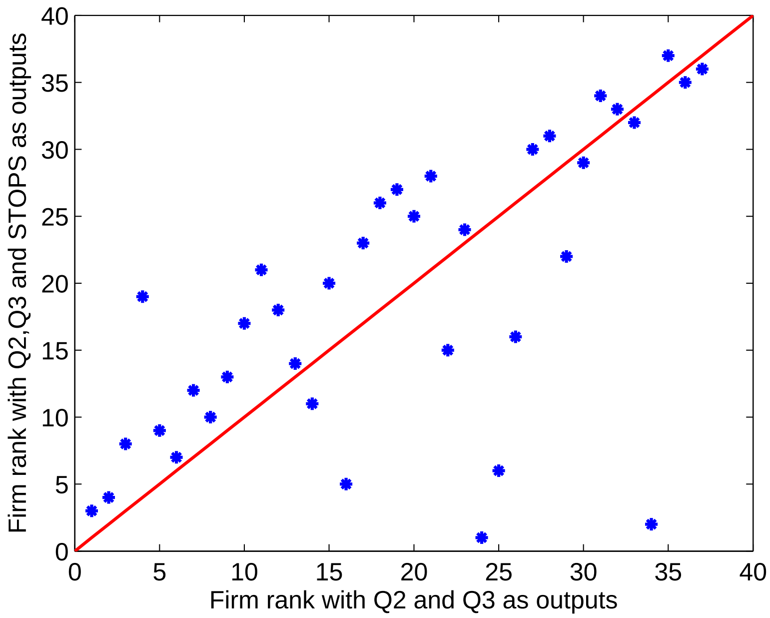


Figure 4.7: Firm rank comparison: with and without STOPS as output.

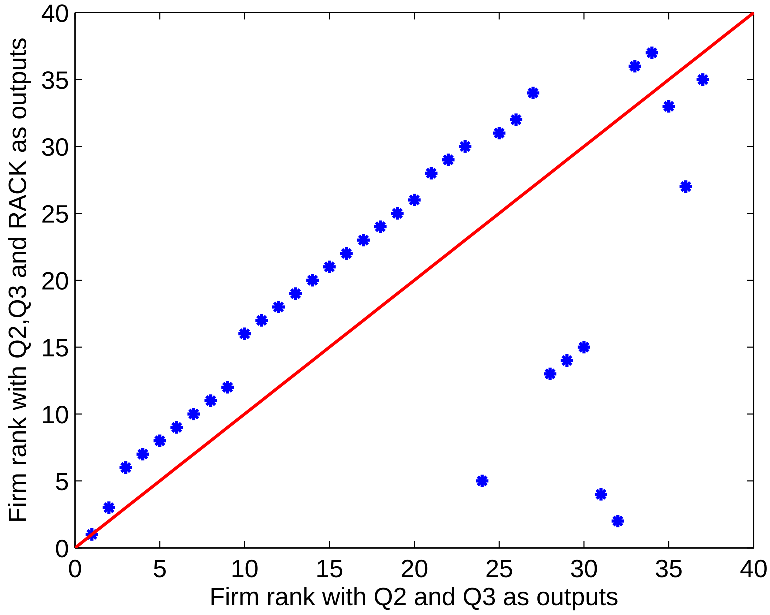


Figure 4.8: Firm rank comparison: with and without RACK as output.

could be expected a priori, as the degressive and linear models put increasingly stringent structure on possible allocations of capital costs to successive time periods. However, the firms' inefficiency differences are very small in general. In fact, the efficiency rankings hardly change. Thus, we may safely conclude that the efficiency results of our main analysis presented above are quite robust with respect to using degressivity or linearity for the write-off schemes of the durable capital input.

4.8 Conclusion

We have presented a methodology for intertemporal analysis of economically (cost) efficient production behavior that can account for intertemporal considerations related to the use of storable and durable inputs. The methodology is intrinsically nonparametric, which means that it does not require imposing (nonverifiable) functional structure on the production technology. The methodology is versatile in that it can account for production delays of durable inputs. In addition, it allows for defining a cost inefficiency measure that can be decomposed in period-specific inefficiencies. These cost inefficiencies can be computed through simple linear programming.

Our application to Swiss railway companies has shown the empirical usefulness of our methodology. Most notably, it showed that explicitly accounting for the dynamic nature of (in our case capital) inputs can significantly impact the efficiency results. For a considerable number of firms, we found that per-period inefficiencies for our model with capital investments as durable inputs differed substantially from the ones for the

firm id	Unconstrained	rank	Degressive	rank	Linear	rank
2	0,0276	23	0,0299	22	0,0319	19
3	0,0399	25	0,0397	25	0,0407	23
4	0,0000	11	0,0000	12	0,0000	1
5	0,0000	1	0,0000	13	0,0011	14
6	0,0132	19	0,0139	17	0,0152	17
7	0,0170	21	0,0176	19	0,0210	18
8	0,0267	22	0,0314	23	0,0438	24
9	0,1993	34	0,2129	36	0,2335	36
10	0,0000	13	0,0000	2	0,0000	4
12	0,0419	26	0,0434	26	0,0468	25
13	0,3636	37	0,3621	37	0,3595	37
14	0,0000	4	0,0014	15	0,0042	15
15	0,0938	30	0,1026	30	0,1119	30
16	0,1528	31	0,1578	31	0,1682	31
17	0,1605	32	0,1669	32	0,1803	32
18	0,0000	5	0,0000	14	0,0000	10
20	0,2009	35	0,2110	35	0,2225	35
21	0,0883	28	0,0912	28	0,0964	28
22	0,2082	36	0,2086	34	0,2117	34
23	0,0000	8	0,0000	11	0,0000	13
24	0,0846	27	0,0880	27	0,0908	27
26	0,0000	9	0,0000	3	0,0000	6
27	0,0041	16	0,0162	18	0,0338	20
30	0,1862	33	0,1891	33	0,1908	33
31	0,0000	15	0,0000	10	0,0000	12
34	0,0000	14	0,0000	1	0,0000	3
36	0,0127	18	0,0240	21	0,0527	26
37	0,0000	6	0,0000	9	0,0000	2
39	0,0000	7	0,0000	4	0,0000	9
41	0,0000	3	0,0000	8	0,0000	11
42	0,0000	12	0,0000	6	0,0000	5
43	0,0139	20	0,0226	20	0,0343	21
45	0,0000	2	0,0000	7	0,0000	7
46	0,0937	29	0,0960	29	0,0997	29
47	0,0082	17	0,0117	16	0,0123	16
48	0,0307	24	0,0349	24	0,0389	22
49	0,0000	10	0,0000	5	0,0000	8

Table 4.7: Dynamic efficiency scores RCE_k ($J = 2$) under degressive and linear write-off.

(static) model that ignores such intertemporal durability. At a more general level, these empirical findings demonstrate the practical relevance of our methodology for regulators: erroneously disregarding intertemporal aspects of firms' production decisions may substantially distort the efficiency assessment and, therefore, also the policy conclusions that are drawn from it.

We see multiple possible extensions. First, we have been considering a multi-output setting in the current paper but ignored any output interdependencies, mainly to simplify our exposition. In practice, however, interdependencies among outputs often exist in the form of joint inputs. From this perspective, it seems particularly interesting to combine our methodology for dynamic production analysis with the (nonparametric) methodology for multi-output production analysis that was recently developed by Cherchye et al. (2013, 2014). This multi-output framework accounts for interdependencies between different output production processes through jointly used inputs, which are formally similar in nature to the durable inputs on which we focus in the current paper (i.e., they capture inter-period interdependencies between production decisions). Combining the two methodologies will further enhance the realistic modeling of production interdependencies (across outputs as well as time periods).

Next, our cost and technical inefficiency measures can be used to measure productivity by combining it with various productivity measures such as the cost Malmquist index of Maniadakis and Thanassoulis (2004) or the Malmquist index of Caves et al. (1982), Bjurek (1996)'s Hicks-Moorsteen index or the Luenberger indicator of Chambers et al. (1996b), among others. These productivity measures have been proposed in the context of nonparametric (DEA) analysis of productive efficiency and, therefore, are easily combined with our novel methodology. This combination will lead to richer productivity analyses because it explicitly accounts for intertemporal production interdependencies through storable and durable inputs.

Third, the focus of this paper was on efficiency under the assumption of perfect price foresight: i.e., our efficiency measures are based on solutions of LPs that assume the evaluated firm correctly predicts the prevailing prices. In practice, this assumption rarely holds so that inefficiency can, at least partially, be explained by failure of this assumption. From a regulator perspective, it is therefore useful to consider efficiency measures allowing for price uncertainty where one considers efficiency under all possible (realistic) price situations. This would ensure that new regulation schemes by the regulator are not too harsh or too loose for the individual firms. A good starting point towards integration in our framework is Kuosmanen and Post (2002).

Finally, Varian (1982) has developed a nonparametric approach to consumer demand analysis that is formally analogous to the nonparametric approach to production analysis to which we adhere here. Following this analogy, we may translate the insights developed in the previous sections towards a consumption setting to obtain a more realistic modeling of intertemporal aspects of consumer behavior.[22] Specifically, our concept of storable inputs corresponds to the notion of infrequent purchases in a consumption con-

[22]See, for example Crawford (2010) and Crawford and Polisson (2014) for recent contributions to the nonparametric analysis of intertemporal consumer behavior.

text, and durable inputs are similar in spirit to durable consumption goods (for example, cars, houses, etc.) in a demand setting.

4.A The economic meaning of write-off schemes

In this short section we clarify the economic intuition behind the write-off schemes for the durable inputs. For a moment, let us rewrite (4.1) by replacing input requirement sets with production functions:

$$\min_{\substack{(\mathfrak{q}_t^1,\ldots,\mathfrak{q}_t^J)_{t=J}^T \\ (\mathfrak{P}_t^1,\ldots,\mathfrak{P}_t^J)_{t=J}^T}} \sum_{t=J}^{T} \sum_{j=t-J+1}^{t} \mathbf{p}_{k,j}\mathfrak{q}_j^{t-j+1} + \mathfrak{P}_j^{t-j+1}\mathbf{Q}_{k,j} \tag{4.22a}$$

$$\text{s.t. } \mathbf{F}_t\left(\sum_{j=1}^{J}\mathfrak{q}_{t-j+1}^{j}, \sum_{j=1}^{J}\mathbf{Q}_{k,t-j+1}\right) \geq \mathbf{y}_{k,t} \qquad \forall t = J,\ldots,T \tag{4.22b}$$

The first order conditions with respect to $\mathbf{Q}_{k,t}$, for all $t = J,\ldots,T$, are

$$\sum_{j=1}^{J}\mathfrak{P}_{k,t}^{j} - \sum_{j=1}^{J}\lambda_{t+j-1}\frac{\partial \mathbf{F}_{t+j-1}}{\partial \mathbf{Q}_{k,t}} \geq 0 \Leftrightarrow \mathbf{P}_{k,t} - \sum_{j=1}^{J}\lambda_{t+j-1}\frac{\partial \mathbf{F}_{t+j-1}}{\partial \mathbf{Q}_{k,t}} \geq 0,$$

which holds with equality if $\mathbf{Q}_{k,t} > 0$. Rearranging shows that, when a durable input $\mathbf{Q}_{k,t}$ is purchased at time t, the discounted market prices reflect the expected marginal benefits to production of the durable inputs over their entire lifetime, i.e.,

$$\mathbf{P}_{k,t} = \sum_{j=1}^{J}\lambda_{t+j-1}\frac{\partial \mathbf{F}_{t+j-1}}{\partial \mathbf{Q}_{k,t}}, \tag{4.23}$$

which reveals a write-off scheme that defines the valuation of the firm for durable inputs in terms of their marginal effects on productivity in periods t to $t+J-1$. It is as if the firm invests in this input and writes off this investment for J periods. We capture this interpretation by (implicit) period-specific prices

$$\mathfrak{P}_{k,t}^{j} = \lambda_j\frac{\partial \mathbf{F}_j}{\partial \mathbf{Q}_{k,t}}. \tag{4.24}$$

Intuitively, these period-specific prices attribute part of the cost of the durable inputs to different periods t in accordance to the inputs' marginal productivities.

Chapter 5

Peer screening with DEA: evaluating customer segments of a telecom operator

"Presenting DEA graphically, due to its multiple variable nature, has proven difficult and some have argued that this has left decision-makers and managers at a loss in interpreting the results."

— Nicole Adler and Adi Raveh[1]

5.1 Introduction

Management decisions are made keeping certain objectives in mind. These objectives are multi-dimensional in nature and realizing them all at once can be a complicated matter. Furthermore, not all objectives have equal priority. For instance, a manager may know that he could increase a certain output but the effort and resources spend on doing so can quite possibly outweigh the benefits. Moreover, realizing multiple objectives at once can require more complicated strategies. Technically, the direction of projection onto the efficient frontier determines the objectives to focus on. This begs the question how to choose this direction. This issue has long been neglected in the literature. Often, a direction vector is chosen radial to the evaluated observation or set to the unit vector without much justification apart from interpretational convenience. Furthermore, sensitivity of the results to direction vector choice is hardly checked. Peyrache and Daraio (2012) present empirical tools to assess sensitivity of efficiency measures to the choice of direction vectors in non-convex technologies.

Recently, some of the literature focused on determining the direction vectors in some "optimal" way. The interpretation of "optimality" broadly depends on whether price information is available. A first stream focuses on the case without prices. Some examples are: Färe et al. (2013) choose the direction vectors that maximize inefficiency conditional on a direction vector normalization; Hampf and Krüger (2015) extend their approach to a dynamic setting to calculate the Malmquist-Luenberger productivity index; Daraio

[1] Adler and Raveh (2008, p.716)

and Simar (2016) endogenously determine the direction vectors by regressing contextual factors on the angles of the polar coordinates of the inputs and outputs. Although much can be said in favor of these approaches, an important issue from a management perspective is with regard to the intuitive interpretation of these optimal direction vectors. Lack of intuitive interpretation could preclude widespread usage. A second stream uses available prices to determine optimal direction vectors. Examples here include: Zofío et al. (2013) determine direction vectors in the direction of profit-maximizing benchmarks; Atkinson and Tsionas (2016) estimate firm specific optimal directions consistent with first-order conditions of cost minimization and profit maximization in a parametric setting. Unfortunately, as in our empirical application, price data is often absent which render these methods useless. We refer to Wang et al. (2017) for an extensive literature review.

We take a different approach here. Liu et al. (2009) and Liu and Lu (2010) proposed a method to further differentiate among dominating peers using eigenvector centrality. Their approach consists of computing efficiency scores for all possible combinations of input and output dimensions, aggregating the corresponding intensity variables and computing the eigenvalue decomposition. This identifies key decision making units (DMUs) in the DMU network using partial technologies. These partial technologies might not all be technically sensible. We start from the observation that the choice of direction vector determines the part of the technology frontier to which inefficient DMUs are projected. Thus, it also determines the dominating peers against which an inefficient DMU is benchmarked. Therefore, we modify their method by computing efficiency scores for different possible direction vectors and use the key DMUs to identify the most interesting objectives to focus on in subsequent analysis. These objectives are identified through a comparative analysis of the key DMUs' characteristics. The rationale being that key DMUs are the successful DMUs where the others can learn from.

Typically an inefficient DMU has more than one dominating peer. How can we compare an inefficient DMU against its dominating peers? The second part of this paper proposes a visualization tool to compare different DMUs on their input-output mix and scale. The input and output levels of dominating peers are typically visualized on a radar plot to compare against the input and output levels of the assessed DMU (e.g., Daraio and Simar (2016)). Although radar plots can be illuminating, they have two drawbacks: (i) they quickly become overcrowded when there are many input (output) dimensions and/or when there are many observations; (ii) they do not easily allow for quickly comparing the input-output mix and scale of various DMUs. We are not aware of any visualization tool in the literature that allows to quickly compare DMUs in terms of their input-output mix similarity and scale. Therefore, we propose an easy to calculate input-output mix similarity and scale visualization tool which solves both drawbacks at the cost of the ability to compare individual input and output dimensions. The tool is therefore intended to be used in conjunction with radar plots and enriches the toolbox of efficiency analysis further to go beyond the conventional analysis.

The empirical application benchmarks customer segments of a large European telecom firm using Activity-Based-Costing (ABC) data. This dataset was previously used in Cherchye et al. (2017a) to analyze customer segments from a cost perspective using a multi-output framework. However, they did not analyze dominating peers in detail. We identify the most interesting output objectives using the modified Liu et al. (2009) and Liu and Lu (2010) method before analyzing the individual results in more detail. The proposed visualization tool is used to compare customer segments and their dominating peers. Our analysis highlights how results change under different direction vectors.

This paper is structured as follows. We start with introducing necessary notation and the directional distance function. We next describe the data used in the empirical application. The paper then essentially consists of two distinct parts with a common empirical application. The first part focuses on our first methodological contribution of selecting the direction vector using characteristics of key DMUs and applies it to the data. The second part focuses on the second methodological contribution by introducing our input-output mix similarity and scale visualization tool and applies it to the previously obtained results to compare evaluated DMUs and its dominating peers in more detail. The final section concludes.

5.2 Model

We have controllable inputs $\mathbf{X}^{C_x}$ and fixed inputs $\mathbf{X}^{F_x}$ such that $\mathbf{X} \equiv (\mathbf{X}^{C_x}, \mathbf{X}^{F_x}) \in \mathbb{R}^n_+$. Similarly, there are controllable outputs $\mathbf{Y}^{C_y}$ and fixed outputs $\mathbf{Y}^{F_y}$ such that $\mathbf{Y} \equiv (\mathbf{Y}^{C_y}, \mathbf{Y}^{F_y}) \in \mathbb{R}^m_+$. The distinction between fixed and controllable factors here reflects that fixed factors cannot be easily adjusted in the short term by the firm. Below, these fixed factors affect peer selection but do not determine the efficiency score itself. The dataset is $S = \{(\mathbf{X}_k^{C_x}, \mathbf{X}_k^{F_x}, \mathbf{Y}_k^{C_y}, \mathbf{Y}_k^{F_y})\}_{k=1,\dots,K}$. The production possibilities set is then defined:

$$\boldsymbol{\mathcal{Y}} = \left\{(\mathbf{X}^{C_x}, \mathbf{X}^{F_x}, \mathbf{Y}^{C_y}, \mathbf{Y}^{F_y}) \in \mathbb{R}^{n+m}_+ | (\mathbf{X}^{C_x}, \mathbf{X}^{F_x}) \text{ can produce } (\mathbf{Y}^{C_y}, \mathbf{Y}^{F_y})\right\}.$$

We assume that this technology is a closed, convex, free disposal of inputs and outputs production possibilities set. To maintain the dual profit interpretation (cfr. Section 1.4 in Chapter 1), we assume convexity of the technology. We use the directional distance function (Chambers et al., 1996b) to compute inefficiency for DMU 0:

$$D(\mathbf{X}_0, \mathbf{Y}_0; \mathbf{g}) = \sup\left\{\beta \in \mathbb{R} : (\mathbf{X}_0^{C_x} - \beta \mathbf{g}_x^{C_x}, \mathbf{X}_0^{F_x}, \mathbf{Y}_0^{C_y} + \beta \mathbf{g}_y^{C_y}, \mathbf{Y}_0^{F_y}) \in \boldsymbol{\mathcal{Y}}\right\}, \qquad (5.1)$$

if $(\mathbf{X}_0^{C_x} - \beta \mathbf{g}_x^{C_x}, \mathbf{X}_0^{F_x}, \mathbf{Y}_0^{C_y} + \beta \mathbf{g}_y^{C_y}, \mathbf{Y}_0^{F_y}) \in \boldsymbol{\mathcal{Y}}$ for some β and $D(\mathbf{X}_0, \mathbf{Y}_0; \mathbf{g}) = -\infty$ otherwise. Intuitively, the observation $(\mathbf{X}_0, \mathbf{Y}_0)$ is projected onto the efficient frontier in the direction $\mathbf{g} = (\mathbf{g}_x, \mathbf{g}_y) = (\mathbf{g}_x^{C_x}, \mathbf{0}^{|F_x|}, \mathbf{g}_y^{C_y}, \mathbf{0}^{|F_y|})$.

Although the dataset in our empirical application is well-suited, we do not work with output-specific or subprocess technologies such as in Chapter 2 and Cherchye et al. (2017a) because this would complicate matters further. In practice, we compute the linear program:

$$\max_{\beta_0,\boldsymbol{\lambda}} \beta_0 \tag{5.2a}$$
$$\sum_{k=1}^{K} \lambda_{0k}\mathbf{X}_k^{C_x} \leq \mathbf{X}_0^{C_x} - \beta_0\mathbf{g}_x^{C_x}, \tag{5.2b}$$
$$\sum_{k=1}^{K} \lambda_{0k}\mathbf{X}_k^{F_x} \leq \mathbf{X}_0^{F_x}, \tag{5.2c}$$
$$\sum_{k=1}^{K} \lambda_{0k}\mathbf{Y}_k^{C_y} \geq \mathbf{Y}_0^{C_y} + \beta_0\mathbf{g}_y^{C_y}, \tag{5.2d}$$
$$\sum_{k=1}^{K} \lambda_{0k}\mathbf{Y}_k^{F_y} \geq \mathbf{Y}_0^{F_y}, \tag{5.2e}$$
$$\sum_{k=1}^{K} \lambda_{0k} = 1, \tag{5.2f}$$

with direction vectors $\mathbf{g}_x^{C_x} \in \mathbb{R}_+^{|C_x|}, \mathbf{g}_y^{C_y} \in \mathbb{R}_+^{|C_y|}$ and intensity variables $\boldsymbol{\lambda} \in \mathbb{R}_+^{K\times K}$. This is basically the same linear program as (1.16) with $\Gamma = VRS$ and the direction vectors of the fixed inputs/outputs set to zero. We assume variable returns-to-scale as this provides the tightest inner bound approximation of the data and because we are ignorant on possible prevailing returns-to-scale. The direction vectors determine the directions in which to seek reduction (expansion) of inputs (outputs) and are specified by the empirical analyst. The efficient projection is then $(\mathbf{X}_0^{C_x} - \beta\mathbf{g}_x^{C_x}, \mathbf{X}_0^{F_x}, \mathbf{Y}_0^{C_y} + \beta_0\mathbf{g}_y^{C_y}, \mathbf{Y}_0^{F_y})$. The dominating peers of DMU 0 are characterized by:

$$D_0 = \{j|\lambda_{0j} > 0\}.$$

The set D_0 contains all comparison partners used in a convex combination to form a (hypothetical) benchmark DMU $\left(\sum_{k=1}^{K} \lambda_{0k}\mathbf{X}_k, \sum_{k=1}^{K} \lambda_{0k}\mathbf{Y}_k\right)$. It is possible that D_0 is not uniquely defined. This is the case if some of the dominating peers are only weakly efficient (i.e., there are non-zero slacks). For simplicity, we assume that D_0 is uniquely determined. We refer to Thrall (1996) for procedures to classify the various types of DMUs. We are interested in analyzing these individual dominating peers' characteristics which together compose the hypothetical benchmark DMU.

Quite often not much discussion is devoted to the exact choice of direction vector and one often opts for $(\mathbf{g}_x^{C_x}, \mathbf{g}_y^{C_y}) = (\mathbf{X}_0^{C_x}, \mathbf{Y}_0^{C_y})$ or $(\mathbf{g}_x^{C_x}, \mathbf{g}_y^{C_y}) = (\mathbf{1}_{|C_x|}, \mathbf{1}_{|C_y|})$ without much ado. The results of this benchmarking analysis serves to guide DMUs to become more profitable and therefore has severe managerial implications. These managerial implications can range from individual cost cutting on a DMU level to strategic reorientation. From a managerial perspective it is then curious that the choice of direction often receives little attention in empirical applications. A first objective of this paper is then to choose the direction vectors in such a way that they have an intuitive interpretation from a managerial perspective.

5.3 Evaluating customer segments of a telecom operator

The empirical application focuses on a benchmark analysis of customer segments for a large European telecom operator. This operator recently moved from a product centric to a customer centric strategy where customers rather than products are viewed as the most important asset. It is in this setting that we analyze the different customer segments using a panel data over 12 months.

5.3.1 Customer segments

The telecom operator segments its customers based on the product combination the customer has, the region in which the customer lives, and the socio-demographic category to which the customer belongs. The telecom operator offers fixed telephone, mobile telephone, digital television and internet. Customers can choose any possible combination of products. The main distinction for the product combination is the number of products, leading to product combinations, which we will label 'play packs', with 4, 3, 2, or 1 product respectively. Other subscriptions that do not fit these play packs are labeled as 0-play. These are legacy subscriptions which are no longer offered to new customers, but which are maintained for existing customers. Furthermore, the telecom operator distinguishes 11 regions and 6 socio-demographic groups. It is important to only compare customer segments that operate in a similar environment so that the benchmarking results are realistic. For that reason, we only compare customer segments within a particular play pack. Specifically, we only compare customer segments that have the same number of products (i.e., 1, 2, 3, or 4) in their play pack. Obviously, as the telecom operator offers 4 products, there are 4 different combinations for the 3-play pack, 6 different combinations for the 2-play pack and 4 different combinations for the 1-play pack. Combining the number of combinations within each play pack with the 11 regions and 6 socio-demographic groups leads to 66 customer segments for the 4-play pack, 264 customer segments for the 3-play pack, 396 customer segments for the 2-play pack, 264 customer segments for the 1-play pack and 66 customer segments for 0-play.

5.3.2 Data

The telecom operator provides us with data for the year 2014. For each month, we have detailed data on all costs and all revenues associated with every customer segment. We also have data on the total number of customers in each segment as well as on the migration of customers from one customer segment to another customer segment. The number of customers varied between $2,646,952$ and $2,719,896$ over the year.

We subdivide the inputs and outputs as follows:

- $C_x = \{$ repair costs, rental costs devices, call center costs, own shops costs, web costs, commissions, SAC costs, SDC costs $\}$
- $F_x = \{$ interconnection, roaming, content, billing, bad debt, IT $\}$

- $C_y = \{$ mobile revenues, fixed access revenues, fixed internet revenues, fixed digital TV revenues, other revenues$\}$
- $F_y = \{$ churn rate, upsells, number of customers $\}$

All these different outputs in C_y and F_y together reflect the current contribution of a customer segment and the future potential of a customer segment. The current contribution of a customer segment is reflected by five revenue streams in C_y: fixed access, mobile, fixed digital television, fixed internet and other revenues. The future potential of a customer segment is reflected by the churn rate and the number of upsells for each customer segment (i.e., F_y). The churn rate of a play pack represents the percentage of customers that cancel their subscription entirely. The number of upsells for every customer segment is constructed from the monthly migration data and is defined as the number of existing customers of the telecom operator that change their subscription to that particular customer segment. The number of customers is in F_y to ensure that the benchmarked customer segment is only benchmarked against customer segments with equal or more customers. The fixed outputs F_y are factors not under direct control of the firm. It can indirectly influence these by, for example, starting a marketing campaign or doing (temporary) price reductions. Since these would require more extensive customer behavior modeling we simply assume they are fixed in the short-run. Furthermore, this partitioning has the effect that all controllable outputs C_y are expressed in monetary terms.

The inputs in our model are the costs that the telecom operator makes to realize the outputs. These costs typically consists of controllable (C_x) and fixed (F_x) costs. The fixed costs F_x are composed of (i) long term factors which are not immediately adjustable (e.g., content, billing, bad debt and IT) and (ii) truly uncontrollable costs which are inherently linked to the usage pattern of the customer (e.g., interconnection and roaming costs). After consulting with the management team of the telecom operator, we decided to ignore the fixed costs for our analysis as these costs can never be used to realize profit improvements in the short run. The controllable costs include various operating expenditures, acquisition costs (SAC), and development costs (SDC). The acquisition (development) cost for every customer segment is constructed from the monthly migration data by multiplying the acquisition (development) cost for a particular customer segment with the number of customers that migrate to that particular customer segment. In total, we have 14 cost categories that serve as an input in our model.

Table 5.1 presents an overview of the descriptive statistics of the different inputs and outputs for each customer segment that we include in our model. The descriptive statistics learn that there is a lot of variation in both the inputs and the outputs across the different customer segments. On the cost side, repair costs and call center costs are generally the largest costs for the different play packs. The majority of both costs are probably labor costs which can be significant. This would explain why both costs account for a large share of total costs. Furthermore, 0-play has no acquisition or development costs which demonstrate that the subscription is only maintained for existing customers. On the revenue side, mobile revenues are on average largest for 4-play, 1-play and 0-play while fixed internet is on average largest for 3-play and 2-play. Although not shown here to conserve space, some revenues are negative for some observations. We follow Kerstens and Van de Woestyne (2011) to ensure (5.2) also finds profit improvements for these observations.

	4-play		3-play		2-play		1-play		0-play	
Input	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std
Repair costs	454499.29	506769.11	162391.00	307091.52	67038.17	121062.10	78552.40	240117.62	1292.63	1286.11
Rental costs devices	143.31	435.27	68.27	281.72	44.28	185.36	30.44	111.65	11.82	119.46
Call center costs	341731.08	392709.48	99664.88	180986.28	32838.94	56822.25	74216.30	145802.74	4032.03	3430.03
Shops costs	90802.42	104468.34	26351.87	48045.70	7950.84	14812.51	16251.70	34125.03	1089.01	930.04
Web costs	29349.02	33748.44	8493.28	15425.38	2534.66	4817.63	5244.59	11334.93	365.49	314.28
Commissions	32187.31	37880.92	6423.97	9096.37	1811.28	3000.56	9934.00	25645.21	844.17	719.37
SAC CHANNEL	84181.39	132197.61	53732.74	215607.91	15206.82	61934.33	83371.60	268187.65	0.00	0.00
SAC INSTALL	48508.31	76176.96	44440.04	181988.87	18849.24	77574.83	20579.95	91401.23	0.00	0.00
SAC TERMINAL	43939.58	69002.28	39712.93	164955.68	15039.30	69769.69	4148.08	24514.12	0.00	0.00
SAC CCA BACK OFFICE	9518.34	14947.51	8175.63	34831.72	2501.36	10691.17	2211.08	7725.90	0.00	0.00
SDC CHANNEL	197355.11	183100.06	32953.60	40251.84	7628.21	11691.40	1496.62	3466.20	0.00	0.00
SDC INSTALL	105024.60	88256.50	28910.82	36839.06	10537.02	18080.53	2113.03	8069.98	0.00	0.00
SDC TERMINAL	90139.03	75916.81	22959.44	32025.80	6450.71	15284.73	520.11	2204.44	0.00	0.00
SDC CCA BACK OFFICE	18854.18	15833.80	4141.30	6092.69	898.36	1493.36	165.50	665.20	0.00	0.00
Output	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std
Mobile revenues	1609028.16	1932590.72	211823.12	336505.18	49896.88	128735.41	915997.39	2417011.43	35109.44	27414.26
Fixed access revenues	1294707.19	1497576.23	463405.67	1019595.33	205638.11	470249.40	482764.07	1794354.21	3676.24	3428.55
Fixed internet revenues	1485713.15	1703256.65	592675.31	1147821.29	225368.95	491208.91	94561.38	300294.55	4188.88	4301.56
Fixed TV revenues	1119748.99	1282514.40	394958.02	809351.93	111568.24	309895.42	3400.10	5195.58	10101.49	12169.81
Other revenues	21045.20	25447.77	3693.96	6200.80	685.34	1358.57	1133.67	2469.05	412.51	570.50
Churn rate	-0.03	0.00	-0.06	0.04	-0.09	0.09	-0.22	0.10	-0.49	0.00
Upsells	59741.65	68229.48	24126.11	45126.88	12371.78	21999.89	60341.74	120496.68	6058.38	6008.06
Number of customers	60993.89	68906.95	24991.31	46034.43	12821.23	22390.94	61963.66	122203.20	6647.06	6132.40

Table 5.1: Summary statistics of controllable costs and revenues per play pack.

The efficiency scores are computed on a monthly basis by comparing each customer segment with all similar customer segments in all periods (i.e., we assume no change in technology over time). While this is a strong assumption, the advantage of this approach is that we have much more observations to compare with. Specifically, we have 792 peers for the 4-play packs, 3166 peers for the 3-play packs, 4489 peers for the 2-play packs, 3052 peers for the 1-play pack and 792 peers for the 0-play pack. We focus here on the efficiency results for the month December.

We define the set of dominating peers for customer segment 0 by:

$$D_0 = \{j|\lambda_{0j} > 0.001\} .$$

This leaves out the dominating peers that have only a marginal contribution in forming the hypothetical benchmark used in (5.2). All controllable inputs and outputs are expressed in monetary terms, such that output improvements and input reductions have the interpretation of a change in profit:

$$\Delta\text{Profit} \equiv \beta_0 \left(\mathbf{g}_y^{C_y}\mathbf{1}'_{|C_y|} + \mathbf{g}_x^{C_x}\mathbf{1}'_{|C_x|}\right) . \tag{5.3}$$

The common factor β_0 in (5.2) is the greatest common factor with which we can reduce controllable inputs and expand controllable outputs. Hence, it is limited by the factor with the smallest room for improvement and the choice of direction can make a substantial difference on the results. The first task is then to fix these direction vectors. To keep the exposition simple, we focus on identifying important output objectives.

5.4 Identifying important objectives from key DMUs

The underlying rationale of our proposal is that one can learn a lot about important objectives from efficient DMUs. Through a comparative analysis of these efficient DMUs' characteristics one can distill important information. For example: if all output objectives are monetary then one can compute the output shares of every controllable output (i.e., $Y^i / \sum_{\forall i \in C_y} Y^i$) and determine which output most frequently has the largest output share. One could then infer that the most successful DMUs focus more on this particular output and that inefficient DMUs should also focus on improving this output. The same reasoning applies for controllable inputs.

Whether a DMU is efficient requires, of course, a choice of direction vector and this is exactly what we want to determine in the first place. One approach is to solve (5.2) for many different choices of direction vector and keep track of the efficient DMUs through $\boldsymbol{\lambda}$. However, not all efficient DMUs are equally interesting: some are highly specialized while others are all-round efficient. Thus, we need a way of identifying one from the other. This is where the proposal of Liu et al. (2009) and Liu and Lu (2010) to identify key DMUs comes in.

5.4.1 Liu et al.(2009)'s method

We can interpret $\boldsymbol{\lambda}$ as describing a weighted graph: the rows/columns (DMUs) represent different vertices while every entry $\lambda_{ij} > 0$ represents a weighted edge from DMU i to DMU j. Counting the number of incoming vertices reveals the importance of a dominating DMU in the network. However, this only reveals the "all-round" dominating DMUs that perform well on all inputs and outputs, but does not reveal the "highly specialized" dominating DMUs which are only referenced by a small number of other DMUs. To further discriminate among these different types of dominating DMUs, Liu et al. (2009) proposed to calculate Shephard distance functions under a VRS technology for all $(2^n - 1) \times (2^m - 1)$ possible input-output specifications and then summing the intensity variables:

$$\boldsymbol{\Omega} = \sum_{t=1}^{(2^n-1)\times(2^m-1)} \lambda_{ij}^t,$$

where $\boldsymbol{\lambda}^t$ represents the intensity variables for specification t. In this way the merits of every DMU under various situations are considered: the "highly specialized" dominating DMUs are referenced more and with more weight for highly specialized specifications, while "all-round" dominating DMUs are more referenced in the all inputs – all outputs specification. They then use the eigenvector centrality concept of social network analysis in DEA. This allows for ranking of the DMUs in terms of importance in the network. Assume that the vector of rank scores $\mathbf{I} \in \mathbb{R}^K$ measures the importance of every DMU in the network. The basic idea of eigenvector centrality is to determine this I_j for DMU j such that the score is a weighted sum of the referencing DMUs' rank scores and weighted by the link weights Ω_{ij}:

$$cI_j = \sum_{i=1}^{K} \Omega_{ij} I_i,$$

where c is some constant. In matrix notation this gives:

$$c\mathbf{I} = \boldsymbol{\Omega}'\mathbf{I}.$$

It turns out that c is an eigenvalue and $\mathbf{I}$ the corresponding eigenvector. Although there are K solutions to this system of equations, the largest eigenvalue and corresponding eigenvector summarize most of the variation and are therefore retained. Key DMUs are then those with an eigenvector centrality score larger than zero. All-round DMUs k have the largest eigenvector centrality score cI_k while highly specialized DMUs have the lowest eigenvector centrality scores (larger than zero).

5.4.2 Our proposal

An important point to note is that Liu et al. (2009) effectively use all possible "partial technologies" by generating all possible input-output specifications (i.e., every input and output is left out in more than one specification t). One could doubt whether these partial technologies are sensible, because these partial technologies might describe physically impossible true technologies.[2] Instead of modifying the technology in the different specifications and in light of our earlier discussion regarding identification of important objectives from efficient DMUs, we propose to vary the direction vectors over the different specifications t. The choice of direction vector obviously selects the part of the technology frontier to which inefficient DMUs are projected and thus influences which dominating DMUs appear as peers in $\boldsymbol{\lambda}^t$. Varying the direction vector can then also identify all-round efficient DMUs and highly specialized efficient DMUs.

In a follow-up paper Liu and Lu (2010) propose a different aggregation which removes the bias due to scale differences among DMUs. We integrate their modification in Algorithm 1 which further only differs from theirs in that the specifications are generated by varying the direction vectors.

Once the key DMUs are identified one can do a comparative analysis of the most important key DMUs (i.e., those k with largest cI_k score). As mentioned earlier, in the empirical application we propose to compare the output shares of these key DMUs and select the output whose output share is most frequently largest. This simple comparative analysis selects one output objective to be radially expanded. This keeps the intuition simple and yields a clear message from a management perspective. Of course, nothing prevents us from selecting multiple outputs instead of only one. Naturally, the same strategy can be used on the cost side.

[2] To be fair, Liu et al. (2009) do not advocate one particular way of generating the specifications t and simply note that this is one possibility.

Algorithm 1 Identification of key DMUs.

1. Solve (5.2) with $(\mathbf{g}_{t,x}^{C_x}, \mathbf{g}_{t,y}^{C_y}) = \mathbf{g}_t$ for all specifications t and collect $\boldsymbol{\lambda}^t$.

2. Normalize the intensity variables for every evaluated DMU 0 as follows. First compute input weights

$$IW_j^i = \frac{\lambda_{0j}^t X_j^i}{\sum_{j\in D_0} \lambda_{0j}^t X_j^i} \qquad \forall i = 1, \ldots, n,$$

and output weights

$$OW_j^r = \frac{\lambda_{0j}^t Y_j^r}{\sum_{j\in D_0} \lambda_{0j}^t Y_j^r} \qquad \forall r = 1, \ldots, m.$$

We have $0 \leq IW_j^i, OW_j^r \leq 1$. Next, combine these input and output weights $\forall j \in D_0$:

$$IOW_{0j}^t = \frac{1}{n+m}\left(\sum_{i=1}^{n} IW_j^i + \sum_{r=1}^{m} OW_j^r\right).$$

3. Construct the adjacency matrix $\boldsymbol{\Omega} \in \mathbb{R}_+^{K\times K}$ of the network with elements:

$$\Omega_{0j} = \sum_t IOW_{0j}^t.$$

4. Solve $c\mathbf{I} = \boldsymbol{\Omega}'\mathbf{I}$ using eigenvalue decomposition. The largest eigenvalue and corresponding eigenvector correspond to c and $\mathbf{I}$ respectively.

5.4.3 Empirical application

What can the key customer segments in the network reveal about key strategic objectives?

We apply Algorithm 1 to compute eigenvector centrality $c\mathbf{I}$ for the direction vector choices:[3]

- $\left(\mathbf{g}_x^{C_x}, \mathbf{g}_y^{C_y}\right) = \left(|\mathbf{X}_0^{C_x}|, |\mathbf{Y}_0^{C_y}|\right)$;

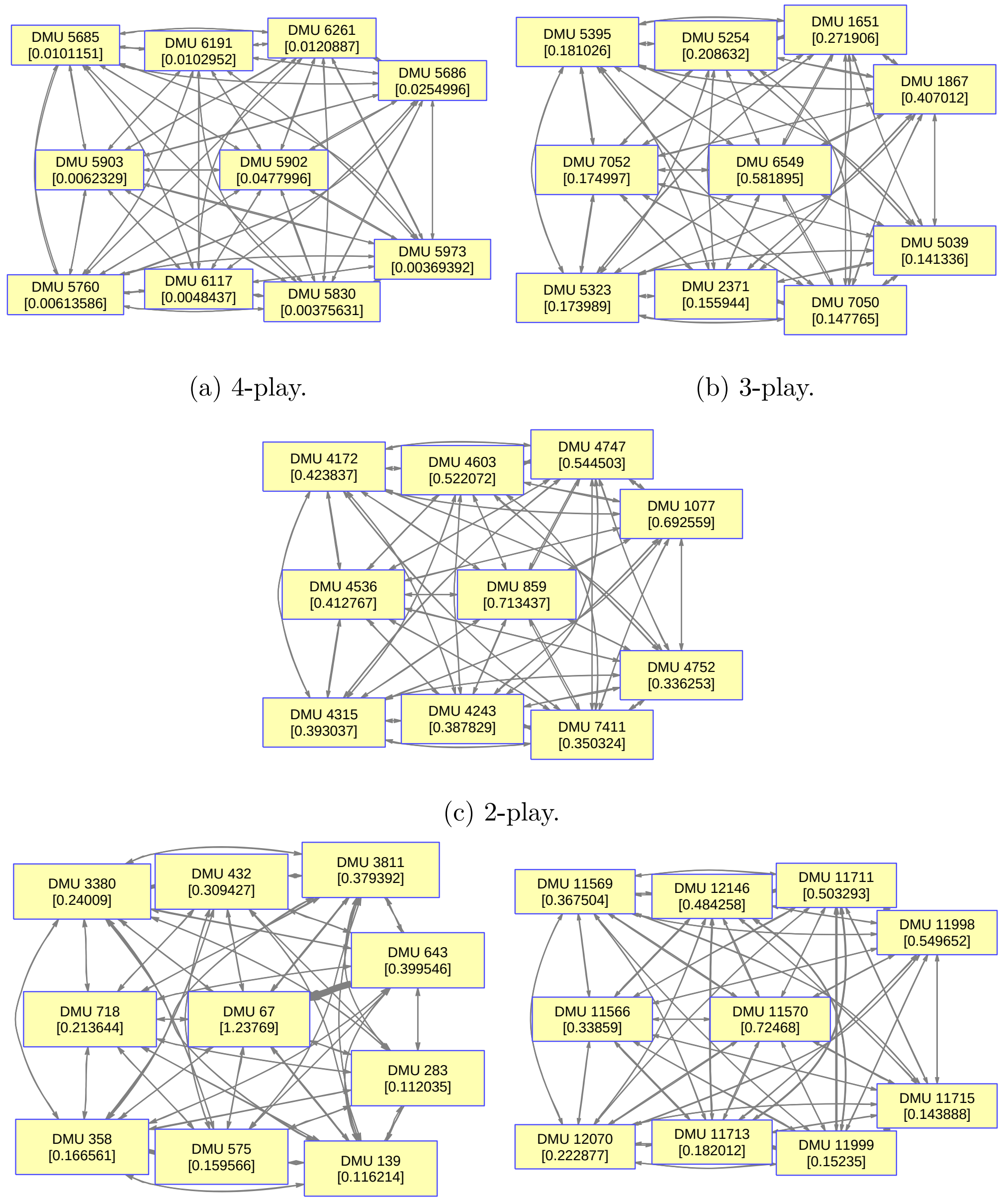
[3]Naturally, this specification of direction vectors is not exhaustive. There is the open question whether there exists a set of direction vectors such that adding any other direction vector to this set does not add any new key DMU. A positive answer to this question would ensure that we do not miss any key DMUs when identifying them through Algorithm 1. This is an interesting avenue for future research.

- $\forall p \in C_y : \left(\mathbf{g}_x^{C_x}, \mathbf{g}_y^{C_y}\right) = \left(|\mathbf{X}_0^{C_x}|, \mathbf{e}_p|\mathbf{Y}_0^{C_y}|\right)$ with $\mathbf{e}_p = (\underbrace{0,\ldots,0}_{p-1}, 1, \underbrace{0,\ldots,0}_{m-p}) \in \mathbb{R}_+^m$.

The first choice is a common choice in the literature which seeks radial improvements in both controllable outputs and controllable inputs. The other choices represent quite extreme managerial views in terms of important output objectives by subsequently giving each controllable output objective complete priority. From a management perspective it gives an idea of how focusing on every output separately affects profitability ceteris paribus. We use the absolute values of the observations in the direction vectors to ensure that (5.2) always finds profit improvements even when an observation contains negative values (Kerstens and Van de Woestyne, 2011).

We favor these direction vector choices, because they have simple managerial interpretations. Expanding only one output objective at a time yields the most straightforward managerial implications. These become harder to interpret when one focuses on more than one output objective at a time. However, a possible disadvantage is that none of the above direction vector choices result in a translation invariant directional distance function. The reason is that these direction vectors depend on the data sample itself: a translation of the data sample causes a translation of the direction vector and in turn affects the computed β_0 in (5.2) (Aparicio et al., 2016). An alternative way of choosing among direction vector candidates is through the properties of the directional distance function such as the translation invariance property. For example: a translation invariant alternative would be to work with variations of $\left(\mathbf{g}_x^{C_x}, \mathbf{g}_y^{C_y}\right) = \left(\mathbf{1}_{|C_x|}, \mathbf{e}_p\mathbf{1}_{|C_y|}\right)$ and divide all inputs and outputs by their respective sample mean.

The computed eigenvector centrality then reveals which customer segments are the key customer segments in the network and how they are linked. We can then draw the network of key customer segments as suggested by Liu et al. (2009). Figure 5.1 shows the top 10 key dominating DMUs in the network in terms of their eigenvector centrality $c\mathbf{I}$ and their links to the other key DMUs in this top 10. We limit this network to 10 as displaying more customer segments would quickly clutter the figures. Recall that all-round weakly efficient DMUs have the largest eigenvector centrality score as they more frequently act as dominating peers compared to highly specialized weakly efficient DMUs. We observe that the key DMUs of 4-play have much smaller eigenvector centrality scores than the key DMUs of the other play packs. However, we cannot compare these scores over different play packs as these scores are derived from the largest eigenvalue of different adjacency matrices $\mathbf{\Omega}$. Moreover, $\sum_{k=1}^{K} cI_k$ differs quite substantially over the play packs: from 0.16 for 4-play to 18.74 for 2-play. Next, these graphs have a directed edge from a key DMU i to j if $\Omega_{ij} > 0$. At first glance, one immediately notices that many of these edges are bidirectional. This might seem odd at first, but is explained by the fact that every edge Ω_{ij} is the aggregation result of 6 individual λ_{ij} for 6 different direction vector choices. Five of these direction vector choices favor extremely specialized customer segments as they only seek output improvements along one dimension at a time. Therefore, all-round weakly efficient DMUs will be inefficient for at least one of these direction vector choices. This explains why there are many bidirectional edges in these graphs.



(a) 4-play. (b) 3-play.

(c) 2-play.

(d) 1-play. (e) 0-play.

Figure 5.1: Network of top 10 DMUs according to their eigenvector centrality (in brackets). Higher is better.

Figure 5.2 shows histograms of the eigenvector centrality scores for every play pack. Only a limited minority of customer segments has an eigenvector centrality larger than zero. Hence, we can limit the analysis to only a number of key DMUs to determine important output objectives. We use these identified key customer segments to determine the important strategic output objectives. The key customer segments are the "successful" ones from which one can learn to improve the other customer segments. For example: determining which monetary outputs have the largest share in their total revenues can be an indication to focus on improving these particular outputs across the board. More concretely, we do:

1. Sort cI_k $\forall k = 1, \ldots, K$ in descending order. Choose $\alpha \in [0, 1]$ and solve
$$d = \arg \min_{d \in [1,K]} d \text{ s.t. } \frac{\sum_{k=1}^{d} cI_k}{\sum_{k=1}^{K} cI_k} \geq \alpha.$$
We use $\alpha = 0.75$.
2. Sort the output shares $\frac{Y_k^i}{\sum_{\forall i \in C_y} Y_k^i}$ $\forall i \in C_y$ in descending order for all $k = 1, \ldots, d$.
3. Count the number of times every output $i \in C_y$ has the largest output share over all $k = 1, \ldots, d$.
4. Set $\mathbf{g}_y^{C_y} = (0, \ldots, Y_0^i, \ldots, 0)$ for output $i \in C_y$ which most frequently has the largest output share.

Thus, we rank the 5 monetary outputs in terms of output share for the largest customer segments whose cumulative share of eigenvector centrality represents 75% of the total eigenvector centrality. Figure 5.3a shows histograms per play pack. These are constructed by computing, for every observation, the share of every output in total outputs and sorting these shares in descending order. We then count how many times every output appears in the ranking. This gives an idea which output objectives are most important for key dominating customer segments. Hence, they can give an indication of the most important output objectives to focus on. In addition, Figure 5.3b shows radar plots of the controllable output shares for these key DMUs. This gives an indication of how the output shares are distributed among the different outputs. It thus complements Figure 5.3a by telling us something about the relative size of the different outputs. Furthermore, Figure 5.3b shows that the key DMUs for all play packs (except 4-play) are (highly) specialized customer segments mostly focusing on a few outputs. The minimal output shares range from 0 to 1 over these play packs. 4-play is the exception here: every output but "other revenues" represents a significant share in the total output. The output shares vary between 0.14 to 0.37 for 4-play.

Overall, we find that mobile revenues and fixed internet revenues most frequently represent the largest share of total output. Following these conclusions, we next focus the analysis on two choices of direction vectors: $\left(\mathbf{g}_x^{C_x}, \mathbf{g}_y^{C_y}\right) = \left(|\mathbf{X}_0^{C_x}|, |\mathbf{Y}_0^{C_y}|\right)$ and $\left(\mathbf{g}_x^{C_x}, \mathbf{g}_y^{C_y}\right) = \left(|\mathbf{X}_0^{C_x}|, \mathbf{e}_1|\mathbf{Y}_0^{C_y}|\right)$ where $\mathbf{e}_1$ selects mobile revenues only. This final choice is very much in line with the developments in the industry which experiences year-on-year increases in mobile data usage and services. A summary of results for other choices of direction vector is in Appendix 5.A.

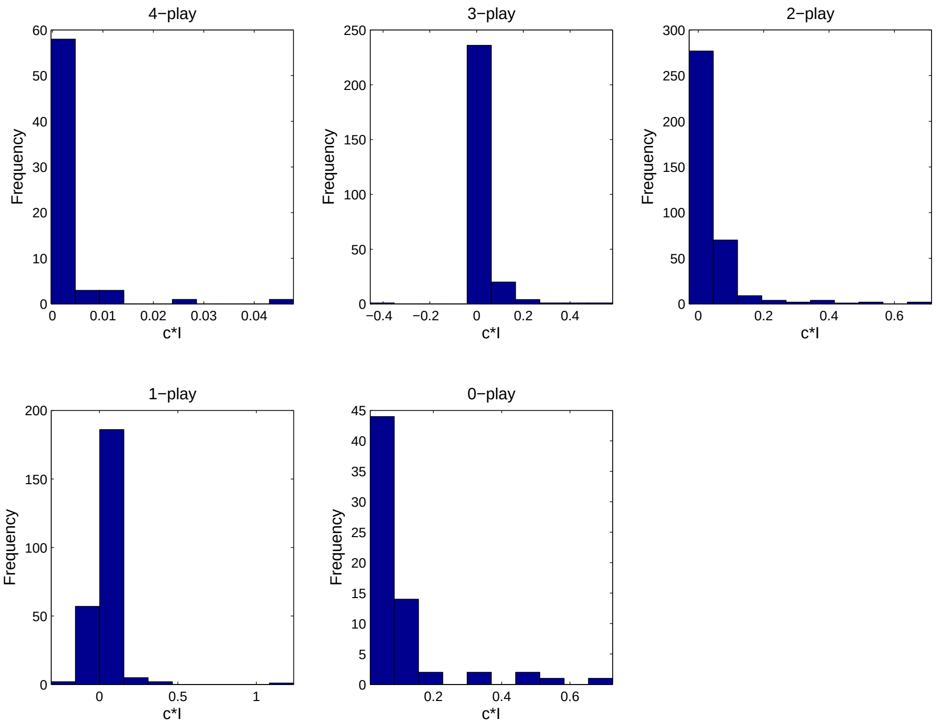
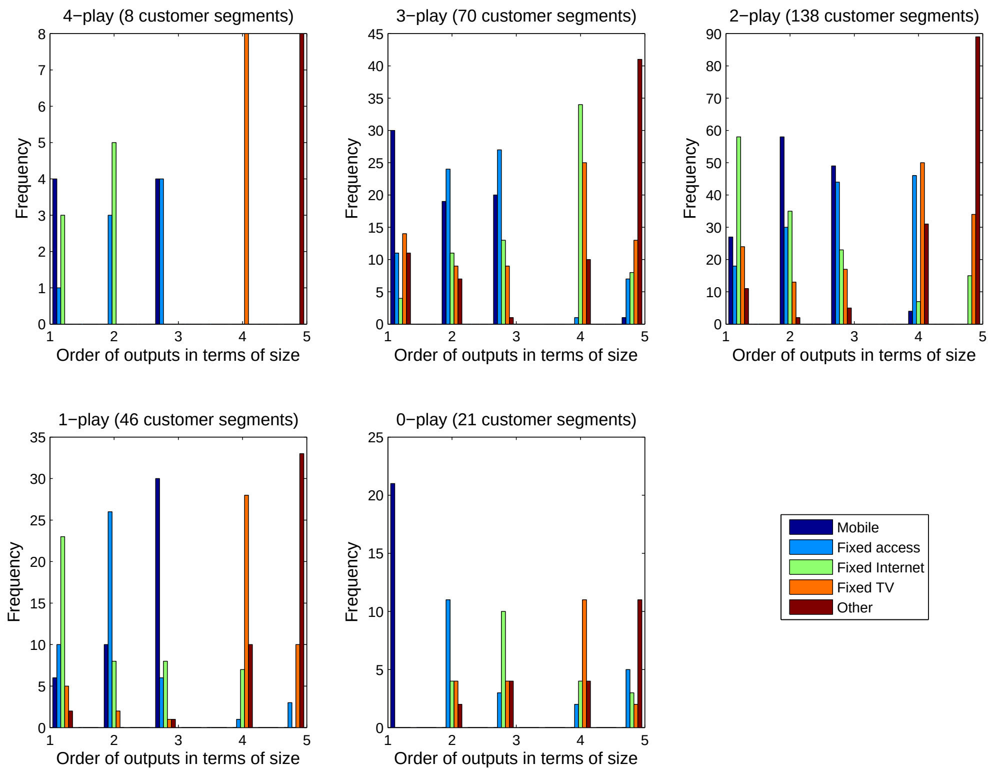


Figure 5.2: Histogram of eigenvector centrality scores $c\mathbf{I}$ for every play pack.

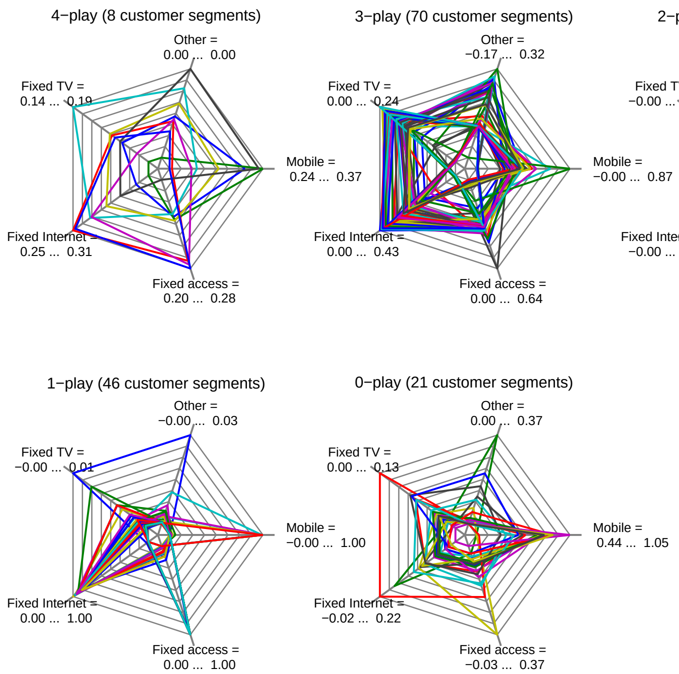
Main results

We compute (5.2) with $\left(\mathbf{g}_x^{C_x}, \mathbf{g}_y^{C_y}\right) = \left(|\mathbf{X}_0^{C_x}|, |\mathbf{Y}_0^{C_y}|\right)$. This common choice seeks radial improvements in both controllable outputs and controllable inputs. We also redo the analysis with $\left(\mathbf{g}_x^{C_x}, \mathbf{g}_y^{C_y}\right) = \left(|\mathbf{X}_0^{C_x}|, \mathbf{e}_1|\mathbf{Y}_0^{C_y}|\right)$ which seeks radial improvements in both mobile revenues and controllable inputs. Thus, it leaves the other outputs unchanged. We discuss the main results for both these direction vector choices together, because the pattern is the same.

Figures 5.4a-5.4b show the frequency of the inefficiency scores per play pack. For every play pack there are many efficient DMUs. The 4-play pack has the lowest inefficiency scores while 0-play has the highest. The proportion of inefficient customer segments in 0-play is largest. Overall a large number of DMUs are efficient. Figures 5.5a-5.5b show histograms of the potential change in profit per play pack. Contrary to what one might expect from looking at β, these figures show that quite some money is left on the table for individual play packs: the potential change in profit can go up to approximately 58000 EUR and 35000 EUR in 4-play depending on the direction vectors choice. Potential change in profit is far lower in 0-play than in the other play packs. However, this is not surprising since 0-play customer segments yield far lower revenues than the other play packs (cfr. Table 5.1).

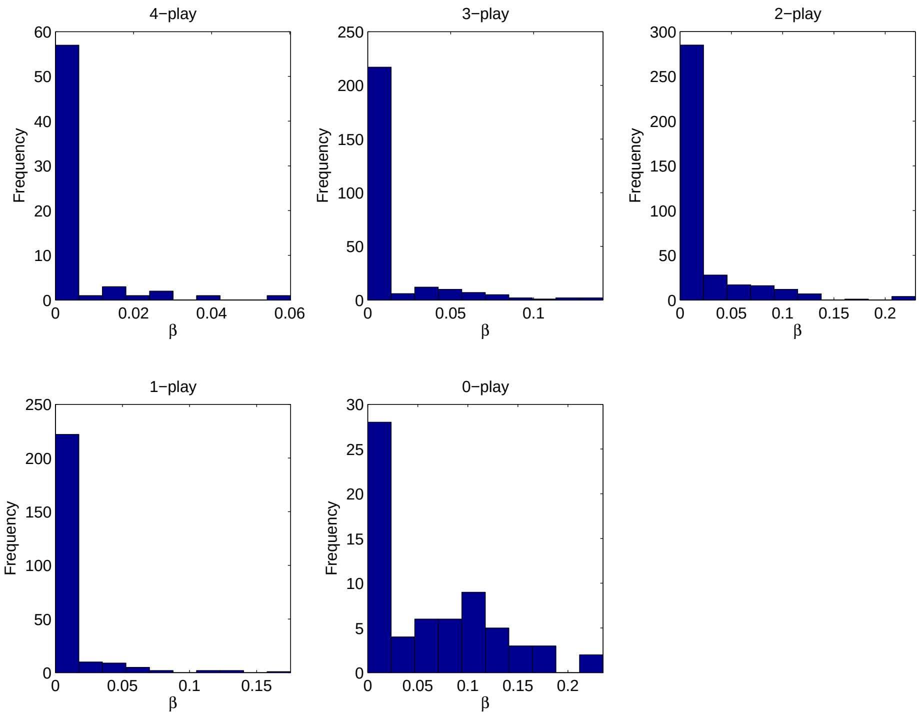


(a) Histogram of order of outputs measured in terms of output share.

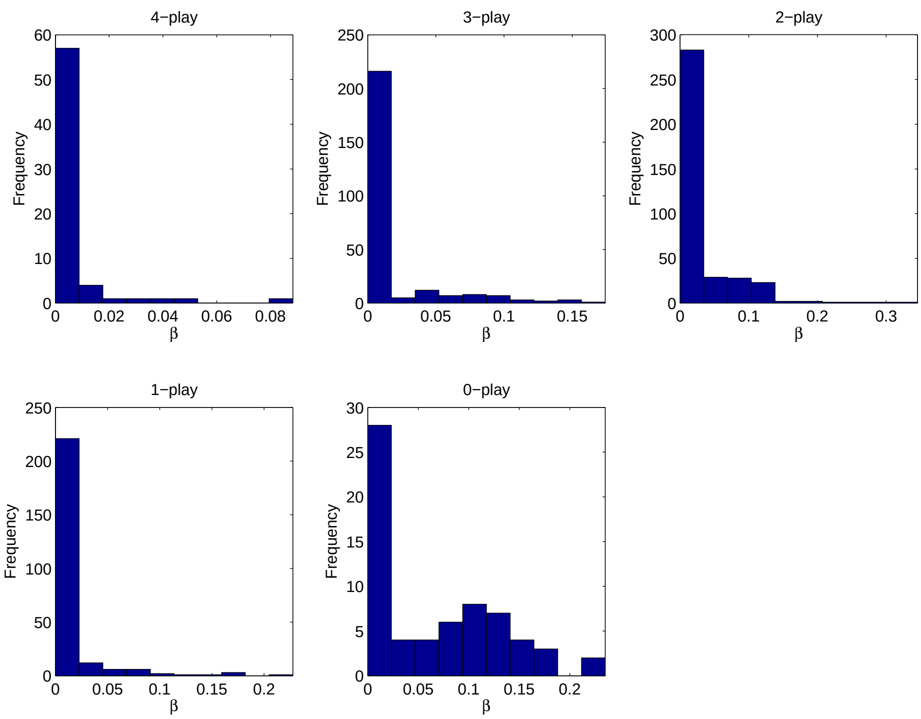


(b) Radar plots of controllable outputs' shares.

Figure 5.3: Characteristics of controllable outputs' shares for key DMUs representing 75% of total eigenvector centrality.

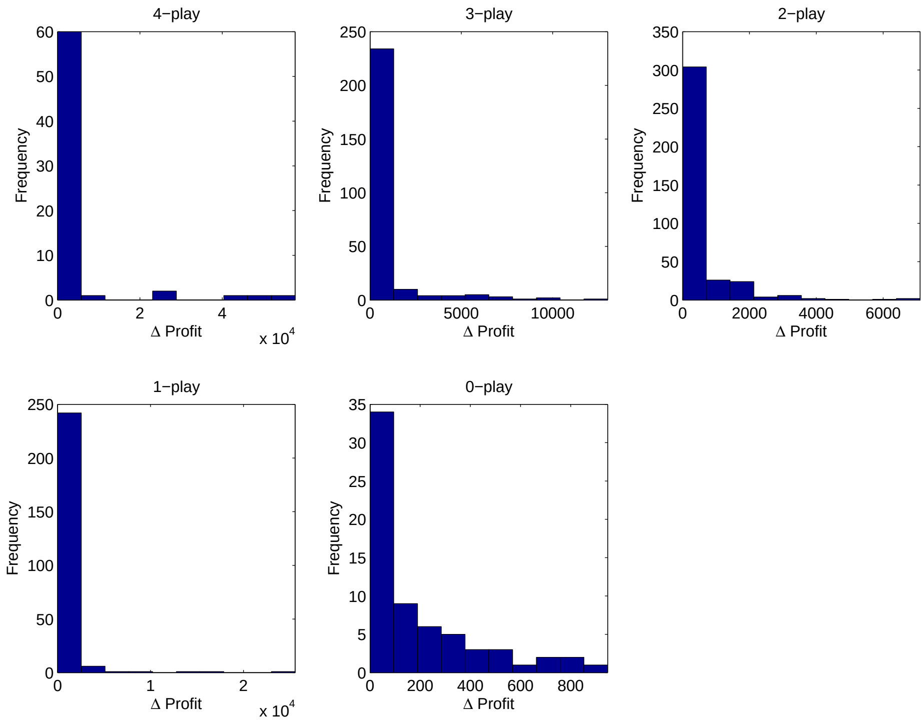


(a) $\left(\mathbf{g}_x^{C_x}, \mathbf{g}_y^{C_y}\right) = \left(|\mathbf{X}_0^{C_x}|, |\mathbf{Y}_0^{C_y}|\right)$.

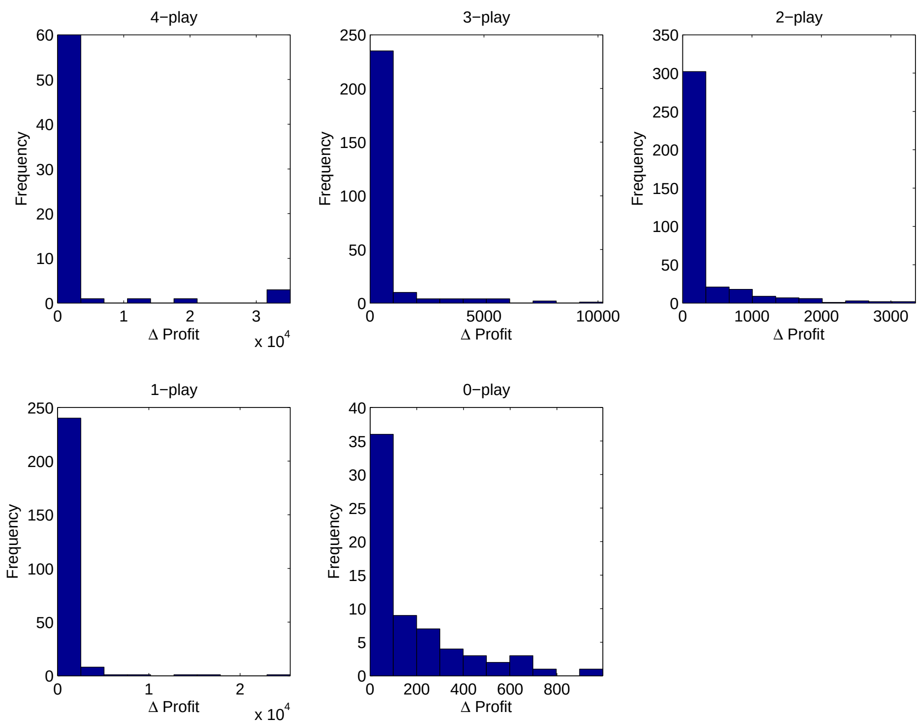


(b) $\left(\mathbf{g}_x^{C_x}, \mathbf{g}_y^{C_y}\right) = \left(|\mathbf{X}_0^{C_x}|, \mathbf{e}_1|\mathbf{Y}_0^{C_y}|\right)$.

Figure 5.4: Histograms of β_0 over all customer segments.



(a) $\left(\mathbf{g}_x^{C_x}, \mathbf{g}_y^{C_y}\right) = \left(|\mathbf{X}_0^{C_x}|, |\mathbf{Y}_0^{C_y}|\right)$.



(b) $\left(\mathbf{g}_x^{C_x}, \mathbf{g}_y^{C_y}\right) = \left(|\mathbf{X}_0^{C_x}|, \mathbf{e}_1|\mathbf{Y}_0^{C_y}|\right)$.

Figure 5.5: Histograms of ΔProfit over all customer segments.

Table 5.2 shows summary statistics on potential profit improvements per play pack category. It also shows the average potential profit improvement per customer in the last column. These average potential profit improvements per customer can be significant: from as little as 0.41 EUR (0.26 EUR) in the 4-play up to 2.58 EUR (2.63 EUR) in the 2-play category for $\left(\mathbf{g}_x^{C_x}, \mathbf{g}_y^{C_y}\right) = \left(|\mathbf{X}_0^{C_x}|, |\mathbf{Y}_0^{C_y}|\right)$ ($\left(\mathbf{g}_x^{C_x}, \mathbf{g}_y^{C_y}\right) = \left(|\mathbf{X}_0^{C_x}|, \mathbf{e}_1|\mathbf{Y}_0^{C_y}|\right)$ respectively). Furthermore, within different play packs there is a lot of variation in ΔProfit. Not surprisingly, the largest profit improvements are found in 4-play and the lowest in 0-play. Curiously, 1-play has larger potential profit improvements than 2-play and even larger than 3-play in some cases. We find the same pattern for $\left(\mathbf{g}_x^{C_x}, \mathbf{g}_y^{C_y}\right) = \left(|\mathbf{X}_0^{C_x}|, \mathbf{e}_1|\mathbf{Y}_0^{C_y}|\right)$. A first takeaway is then that there is quite a lot of room for improvement across all play packs.

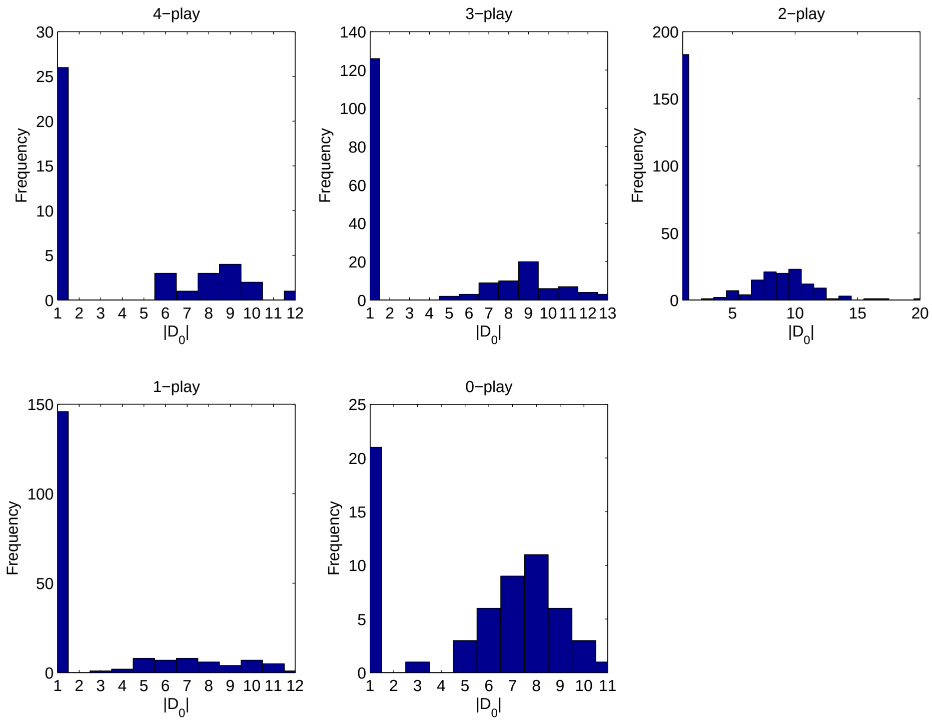
$\left(\mathbf{g}_x^{C_x}, \mathbf{g}_y^{C_y}\right) = \left(\lvert\mathbf{X}_0^{C_x}\rvert, \lvert\mathbf{Y}_0^{C_y}\rvert\right)$					
Play pack	Mean	Std.	Min	Max	Mean EUR/customer
4-play	3502.77	11528.26	-0.01	57692.16	0.41
3-play	600.04	1789.14	-0.01	13010.77	1.89
2-play	386.58	941.88	-0.00	7104.92	2.58
1-play	417.54	2264.94	-0.01	25547.35	0.77
0-play	184.35	237.03	-0.00	947.40	0.47
$\left(\mathbf{g}_x^{C_x}, \mathbf{g}_y^{C_y}\right) = \left(\lvert\mathbf{X}_0^{C_x}\rvert, \mathbf{e}_1\lvert\mathbf{Y}_0^{C_y}\rvert\right)$					
Play pack	Mean	Std.	Min	Max	Mean EUR/customer
4-play	2303.87	7704.09	-0.00	35120.06	0.26
3-play	440.96	1338.57	-0.00	10210.32	1.93
2-play	231.66	548.24	-0.00	3352.36	2.63
1-play	419.34	2279.72	-0.01	25496.78	0.79
0-play	172.07	226.08	-0.00	996.44	0.42

Table 5.2: Summary statistics of potential profit improvements per play pack.

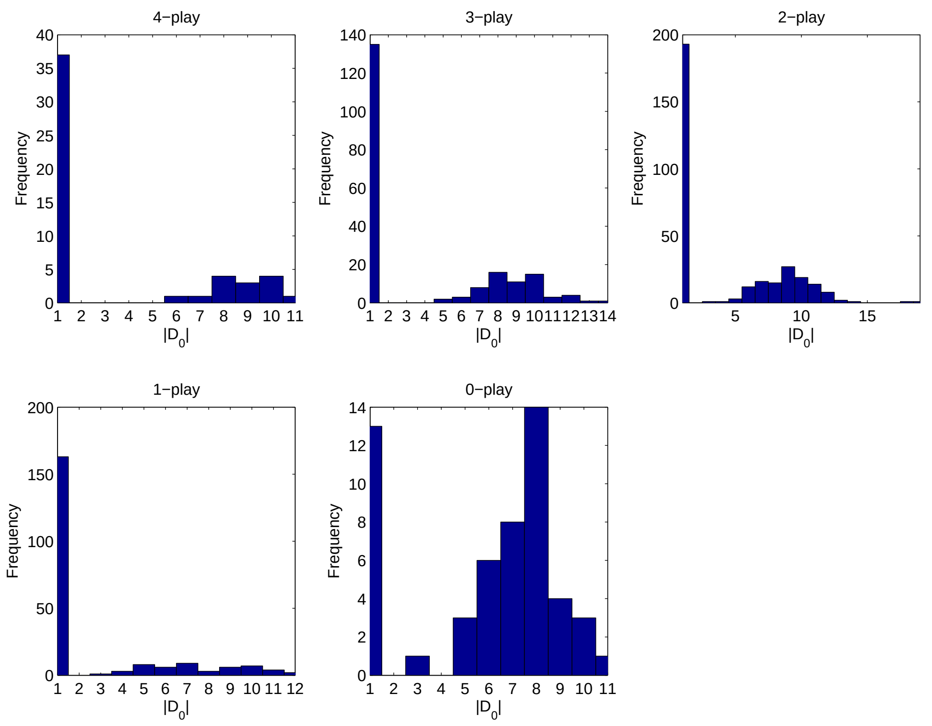
We next look at the number of dominating peers (i.e., $|D_0|$) of every inefficient customer segment. Figures 5.6a-5.6b show the frequency of the number of peers for every customer segment. For all plays (except 0-play) a majority of customer segments has 1 dominating peer while for the minority that does have more than one dominating peer, the number can amount to 20 dominating peers. For 0-play a majority of customer segments has more than one dominating peer. Fortunately, the number of customer segments with many dominating peers is relatively small such that it should still be doable to look into each of these separately. However, this requires much more knowledge of the actual firm operations so that it can only be done by the firm itself.

Individual results

The next step is to delve a bit deeper into these aggregate results and examine more closely the individual observations for every play pack. Improving every single customer



(a) $\left(\mathbf{g}_x^{C_x}, \mathbf{g}_y^{C_y}\right) = \left(|\mathbf{X}_0^{C_x}|, |\mathbf{Y}_0^{C_y}|\right)$.



(b) $\left(\mathbf{g}_x^{C_x}, \mathbf{g}_y^{C_y}\right) = \left(|\mathbf{X}_0^{C_x}|, \mathbf{e}_1|\mathbf{Y}_0^{C_y}|\right)$.

Figure 5.6: Histogram of $|D_0|$ over all inefficient customer segments 0.

segment might not be worthwhile from a cost-benefit perspective and can be time consuming. Therefore, we look into the top 5 of largest potential profit improvements per play pack in Table 5.3 for $\left(\mathbf{g}_x^{C_x}, \mathbf{g}_y^{C_y}\right) = \left(|\mathbf{X}_0^{C_x}|, |\mathbf{Y}_0^{C_y}|\right)$. This table also lists the percentage of ΔProfit the top 5 represents for every play pack. A large proportion of ΔProfit is concentrated in the top 5 for both 4-play (89.81%) and 1-play (66.9%) which should simplify the subsequent analysis to implement these improvements. For the other play packs the gains are much more diversified over a larger number of customer segments. Strikingly, socio segments E and B appear very often in the different top 5.

Next, we look at the top 5 of largest potential profit improvements per play pack in Table 5.4 for $\left(\mathbf{g}_x^{C_x}, \mathbf{g}_y^{C_y}\right) = \left(|\mathbf{X}_0^{C_x}|, \mathbf{e}_1|\mathbf{Y}_0^{C_y}|\right)$. This top 5 once again represents the majority of profit improvements for 4-play (90.26%) and 1-play (66.68%). The potential gains are distributed over much more customer segments for the other play packs. The entire top 5 of 4-play is from socio segment E while the entire top 5 of 0-play is from socio segment B. For the other play packs there is less of a pattern. We further notice that for 4-play, 1-play and 0-play the entire top 5 contains the same customer segments than in Table 5.3. Moreover, for 1-play and 0-play the order is even the same.

The frequency of socio segment E and B in these top 5 would suggest that inefficiency is much more systematic for certain socio demographic groups than for others. Only the firm itself can look into this pattern for explanations. One possibility is that this inefficiency is intentional by the firm as a way, for example, to bind certain demographic groups now to its services with the intention to monetize on their customer loyalty later on.

Now that we know the performance of the different customer segments and the potential profit improvements, the next step in the analysis is to learn how these customer segments can improve their performance. This can be done for every inefficient customer segment by looking at the characteristics of its dominating peers. This also allows management to determine a course of action and differentiate between feasible dominating peers and infeasible dominating peers. Management can then try to bring the inefficient customer segment more in line with the feasible dominating peers by taking appropriate action. The second goal of this paper is to propose a visualization tool that allows to compare input-output mix and scale of each DMU with its dominating peers (or other inefficient DMUs).

5.5 Visualization of input-output mix similarity and scale of DMUs

When the direction vectors are determined we can move on to the actual analysis of individual DMUs. For an individual DMU there are typically many different dominating peers that one needs to analyze by comparing their input and output mix with the benchmarked DMU. We next present a visualization tool that allows to compare input-output mix similarity and scale of DMUs.

4-play		
Total profit improvement = 231182.82 EUR, top 5 = 89.81%		
ΔProfit	Socio segment	Province
57692.16	E	9
50189.33	E	6
44235.23	E	1
28700.25	E	10
26800.67	E	8
3-play		
Total profit improvement = 158411.22 EUR, top 5 = 30.14%		
ΔProfit	Socio segment	Province
13010.77	E	6
9706.64	E	8
9587.96	E	6
7999.61	E	9
7442.99	B	4
2-play		
Total profit improvement = 143033.71 EUR, top 5 = 20.52%		
ΔProfit	Socio segment	Province
7104.92	E	1
7018.49	B	5
6368.99	C	5
4808.93	C	2
4043.11	B	10
1-play		
Total profit improvement = 105638.07 EUR, top 5 = 66.90%		
ΔProfit	Socio segment	Province
25547.35	E	6
16530.73	E	7
14384.66	E	8
8992.04	G	11
5217.68	G	1
0-play		
Total profit improvement = 12167.00 EUR, top 5 = 32.30%		
ΔProfit	Socio segment	Province
947.40	B	2
827.99	B	1
760.57	B	4
704.49	B	5
689.12	B	9

Table 5.3: Top 5 of largest potential profit improvements per play pack for $\left(\mathbf{g}_x^{C_x}, \mathbf{g}_y^{C_y}\right) = \left(|\mathbf{X}_0^{C_x}|, |\mathbf{Y}_0^{C_y}|\right)$.

4-play		
Total profit improvement = 152055.62 EUR, top 5 = 90.26%		
ΔProfit	Socio segment	Province
35120.06	E	6
34506.51	E	1
34140.39	E	9
19985.83	E	8
13485.56	E	10
3-play		
Total profit improvement = 116412.36 EUR, top 5 = 31.57%		
ΔProfit	Socio segment	Province
10210.32	E	6
7385.23	B	4
7384.35	E	8
5968.32	G	11
5803.91	H	1
2-play		
Total profit improvement = 85947.09 EUR, top 5 = 17.37%		
ΔProfit	Socio segment	Province
3352.36	E	1
3023.85	B	5
2996.31	A	8
2948.56	C	2
2608.02	A	3
1-play		
Total profit improvement = 106093.33 EUR, top 5 = 66.68%		
ΔProfit	Socio segment	Province
25496.78	E	6
16481.92	E	7
14616.64	E	8
8949.41	G	11
5193.40	G	1
0-play		
Total profit improvement = 11356.32 EUR, top 5 = 32.79%		
ΔProfit	Socio segment	Province
996.44	B	2
768.32	B	1
679.66	B	4
657.66	B	5
622.22	B	9

Table 5.4: Top 5 of largest potential profit improvements per play pack for $\left(\mathbf{g}_x^{C_x}, \mathbf{g}_y^{C_y}\right) = \left(|\mathbf{X}_0^{C_x}|, \mathbf{e}_1|\mathbf{Y}_0^{C_y}|\right)$.

5.5.1 Input-output mix similarity and scale plot

Consider the numerical example in Table 5.5 with 2 inputs – 2 outputs of 5 hypothetical DMUs. DMU A is a generalist one while B and C are highly specialized DMUs. DMU D is the least performing DMU and is more specialized with a focus on producing Y_1 using input X_1. DMU E is a generalist who divides his attention equally over both outputs and producing efficiently. In this stylized example it is easy to compare the 5 DMUs in terms of their efficiency and input-output mix. The input-output mix here refers to the proportions in which inputs and outputs are used.

DMU	X_1	X_2	Y_1	Y_2
A	2	2	1	1
B	2	0	1	0
C	0	1	0	2
D	0.75	0.25	0.75	0.25
E	0.75	0.75	0.75	0.75

Table 5.5: Numerical example illustrating radar plot and input-output mix similarity.

When analyzing the results of the efficiency computation we often are interested in comparing the input-output mix of various DMUs. In particular, we might be interested to compare the input-output mix of the dominating peers for a benchmarked DMU. The “radar plot” (or “spider plot”) is a visualization tool which can be used for this purpose (Daraio and Simar, 2016). In brief, one assigns different angles $\forall i = 1, \ldots, v$:

$$\rho_i = \frac{(i-1) \cdot 2\pi}{v-1}$$

to the v input (output) dimensions on a circle and then one plots the (normalized) input (output) vector $\mathbf{A} = (A_1, \ldots, A_v)$ with coordinates

$$(A_i \cos \rho_i, A_i \sin \rho_i),$$

$\forall i = 1, \ldots, v$ on the circle. This allows to quickly compare different DMUs along their input (output) dimensions.

However, there are two drawbacks to this visualization. First, if there are many input (output) dimensions and/or DMUs to compare then the plot quickly becomes overcrowded. Second, it is difficult to compare the exact input (output) mix of the various DMUs. The reason is simply that the angles ρ_i are always uniformly distributed over the entire circle and do not depend on $\mathbf{A}$ itself.

Figure 5.7 shows a radar plot for the 5 DMUs of our numerical example. This radar plot allows to compare the different input and output dimensions separately over all DMUs. While the number of dimensions is limited for this numerical example, it should be clear to the reader that adding a few more dimensions would quickly clutter the graph. Furthermore, it is not straightforward to compare the DMUs in terms of their input-output mix. From Table 5.5 one can see that the input-output mix of DMU E is the same as that of DMU A: both use inputs in the same proportion to produce outputs in the same proportion. In contrast, the radar plot does not convey this information.

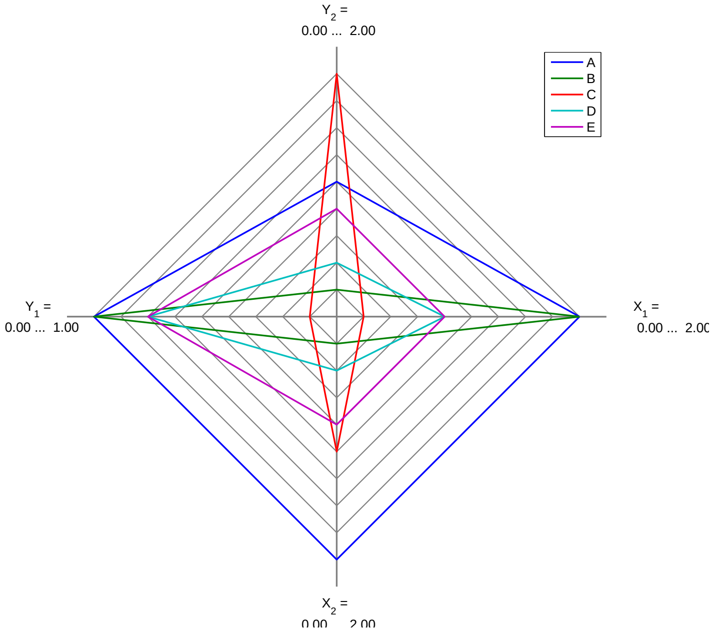


Figure 5.7: Radar plot for 2 inputs - 2 outputs numerical example.

We propose an alternative visualization tool which solves both problems. The drawback is that one looses the ability to compare the individual input (output) dimensions across DMUs. Therefore, this tool can be used in conjunction with the radar plot. In particular, assume that we compare two input (output) vectors **A** and **B** where **A** is the base (e.g., the input (output) vector of the benchmarked DMU) and **B** is some other input (output) vector (e.g., of a dominating peer). We compare (i) their input (output) mix using a similarity criterion and (ii) we also compare their similarity in scale. Both similarity criteria parametrize a 2-dimensional point in polar coordinates. For (i) we use the cosine similarity criteria which only compares the "orientation" of both vectors while ignoring their size. To measure (ii) we simply use the ratio of the 2-norm of both vectors. Summarizing, we define a two-dimensional "similarity" point (r, θ) in polar coordinates:

$$\theta = \arccos(\text{cosine similarity}) = \arccos \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\|_2 \|\mathbf{B}\|_2}, \tag{5.4a}$$

$$r = \frac{\|\mathbf{B}\|_2}{\|\mathbf{A}\|_2}, \tag{5.4b}$$

where $\cdot$ is the dot product, $\theta \in [0, \pi]$ is the angle and $r \in \mathbb{R}_+$ is the radius. The arccos function is used to retrieve the corresponding angle. A larger θ indicates that **A** and **B** are very dissimilar in terms of their input (output) mix.[4] The radius measures similarity

[4]An alternative is to follow Blancard et al. (2016) and use the Hamming distance to measure diffe-

in scale: $r > (<)1$ indicates that $\mathbf{B}$ is much larger (smaller) than $\mathbf{A}$. Naturally, $r = 1$ indicates equal scale.

An intuitive property of this parametrization is that a rescaled input (output) only differs in scale from the original input (output). This is shown as follows: assume $\mathbf{A} = \mathbf{X}_0$ and $\mathbf{B} = (1-\beta_0)\mathbf{X}_0$. Obviously, their mix is equal but their scale is different. The polar coordinates computed using (5.4) reflect this:

$$\begin{aligned}
\theta &= \arccos \frac{\mathbf{X}_0 \cdot (1-\beta_0)\mathbf{X_0}}{\|\mathbf{X}_0\|_2 \|(1-\beta_0)\mathbf{X}_0\|_2} = \arccos \frac{(1-\beta_0)\|\mathbf{X}_0\|_2^2}{(1-\beta_0)\|\mathbf{X}_0\|_2^2} = 0, \\
r &= \frac{(1-\beta_0)\|\mathbf{X}_0\|_2^2}{\|\mathbf{X}_0\|_2^2} = 1-\beta_0.
\end{aligned}$$

This is what one would expect. In contrast, this property breaks down when comparing an efficient projection $\mathbf{B} = [(1-\beta_0)\mathbf{X}_0, (1+\beta_0)\mathbf{Y}_0]$ of an observation with the original observation $\mathbf{A} = [\mathbf{X}_0, \mathbf{Y}_0]$ itself. Here, $[\cdot]$ denotes vector concatenation and using (5.4) yields

$$\begin{aligned}
\theta &= \arccos \frac{[\mathbf{X}_0, \mathbf{Y}_0] \cdot [(1-\beta_0)\mathbf{X}_0, (1+\beta_0)\mathbf{Y}_0]}{\|[\mathbf{X}_0, \mathbf{Y}_0]\|_2 \|[(1-\beta_0)\mathbf{X}_0, (1+\beta_0)\mathbf{Y}_0]\|_2} \\
&= \arccos \frac{[(1-\beta_0)\|\mathbf{X}_0\|_2^2, (1+\beta_0)\|\mathbf{Y}_0\|_2^2]}{\|[\mathbf{X}_0, \mathbf{Y}_0]\|_2 \|[(1-\beta_0)\mathbf{X}_0, (1+\beta_0)\mathbf{Y}_0]\|_2} \neq 0, \\
r &= \frac{\|[(1-\beta_0)\mathbf{X}_0, (1+\beta_0)\mathbf{Y}_0]\|_2}{\|(\mathbf{X}_0, \mathbf{Y}_0)\|_2}.
\end{aligned}$$

The efficient radial projection $\mathbf{B}$ by definition retains the same input-output mix as its original observation $\mathbf{A}$, but this is not reflected in the computed θ. In order to retain this property we propose to compute (5.4) separately for the inputs and the outputs before combining them as follows:

$$\begin{aligned}
\theta &= \frac{1}{2}\left\{\theta_X + \theta_Y\right\} && (5.5a) \\
&= \frac{1}{2}\left\{\arccos\left(\frac{\mathbf{X}_0 \cdot \mathbf{X}_j}{\|\mathbf{X}_0\|_2 \|\mathbf{X}_j\|_2}\right) + \arccos\left(\frac{\mathbf{Y}_0 \cdot \mathbf{Y}_j}{\|\mathbf{Y}_0\|_2 \|\mathbf{Y}_j\|_2}\right)\right\}, \\
r &= \frac{r_Y}{r_X} && (5.5b) \\
&= \frac{\|\mathbf{Y}_j\|_2}{\|\mathbf{Y}_0\|_2} \cdot \frac{\|\mathbf{X}_0\|_2}{\|\mathbf{X}_j\|_2}.
\end{aligned}$$

rences in input and output mix. The range of the Hamming distance is $[0, 1]$ where 0 means complete input (output) mix similarity. The arcsin function then retrieves the corresponding angle.

Now, comparing $\mathbf{B} = [(1-\beta_0)\mathbf{X}_0, (1+\beta_0)\mathbf{Y}_0]$ with $\mathbf{A} = [\mathbf{X}_0, \mathbf{Y}_0]$ yields the similarity point $(r, \theta) = \left(\frac{1+\beta}{1-\beta}, 0\right)$ as expected. Note that we always find $r \geq 1$ when comparing dominating peers with the benchmarked DMU.

We compare the input-output mix similarity of DMU B, C, D and E with respect to A in Table 5.6 and Figure 5.8. Looking at the data we make the following observations: the inputs and outputs of E are used in the same proportions as A, but at lower levels. Thus, E dominates A in the inputs ($r_X = \sqrt{2 \cdot 0.75^2}/2\sqrt{2} \approx 0.38 < 1$), but is being dominated by A in outputs ($r_Y = \sqrt{2 \cdot 0.75^2}/\sqrt{2} = 0.75 < 1$). However, it dominates A more in inputs than A dominates E in outputs. This is reflected in angle 0 and scale 2. Next, B and C are similar in that they are highly specialized DMUs. The difference here is that C clearly dominates A in both inputs and outputs while B does not dominate A. The 45° angle for both reflects their difference in mix compared to A. The radius of 1.0 shows that B operates at equal scale with A while the radius of 4.0 of C reflects that it dominates in both inputs and outputs. Finally, D has a more similar input-output mix to A than B and C, but operates at larger scale ($r = 2$). The larger scale is because D dominates A more in inputs ($r_X = \sqrt{0.75^2 + 0.25^2}/2\sqrt{2} \approx 0.28 < 1$) than it is being dominated by A in outputs ($r_Y = \sqrt{0.75^2 + 0.25^2}/\sqrt{2} \approx 0.56 < 1$). Both characteristics are reflected in its radius and angle.

DMU	r	θ (in rad)
A	1.00	0.00
B	1.00	0.79 ($\approx 45.00°$)
C	4.00	0.79 ($\approx 45.00°$)
D	2.00	0.4636 ($\approx 26.57°$)
E	2.00	0.00

Table 5.6: Input-output mix similarity of DMU B,C,D and E with respect to A of the numerical example.

In the empirical application we set $\mathbf{A} = (\mathbf{X}_0, \mathbf{Y}_0)$ and $\mathbf{B} = (\mathbf{X}_j, \mathbf{Y}_j)$ for all $j \in D_0$. The resulting points (r_j, θ_j) for all $j \in D_0$ are then visualized on a half circle. Together with the aforementioned radar plots this allows to compare the different dominating peers.

5.5.2 Empirical application

We can then compare the inputs and outputs for each of these customer segments with their individual benchmarks (i.e., $\forall j \in D_0$). Figure 5.9a uses radar plots to compare the different outputs of the most inefficient customer segment (4-play, socio segment E, province 9) with its target outputs ($(1+\beta_0)\mathbf{Y}_0^{C_y}$) and its 6 benchmark customer segments for $\left(\mathbf{g}_x^{C_x}, \mathbf{g}_y^{C_y}\right) = \left(|\mathbf{X}_0^{C_x}|, |\mathbf{Y}_0^{C_y}|\right)$. The output levels are relative to the output levels of the benchmarked customer segment in order to give them a percentage interpretation. Figure 5.9b shows similar radar plot for all inputs relative to the input levels of the benchmarked customer segment. Figure 5.10a shows the input-output mix similarity and

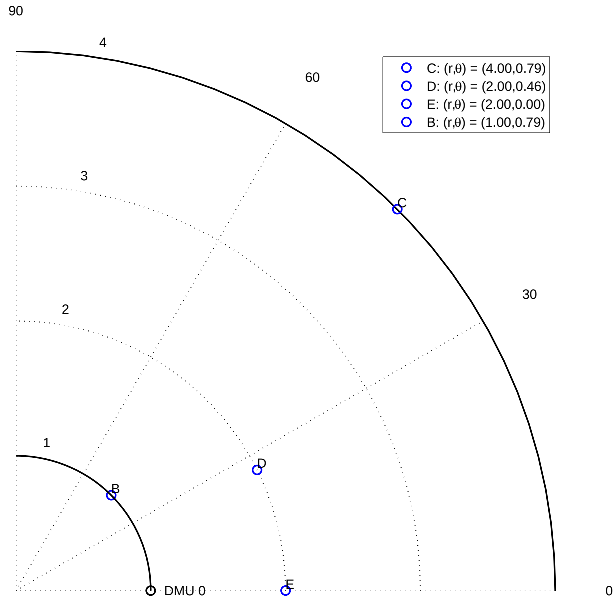
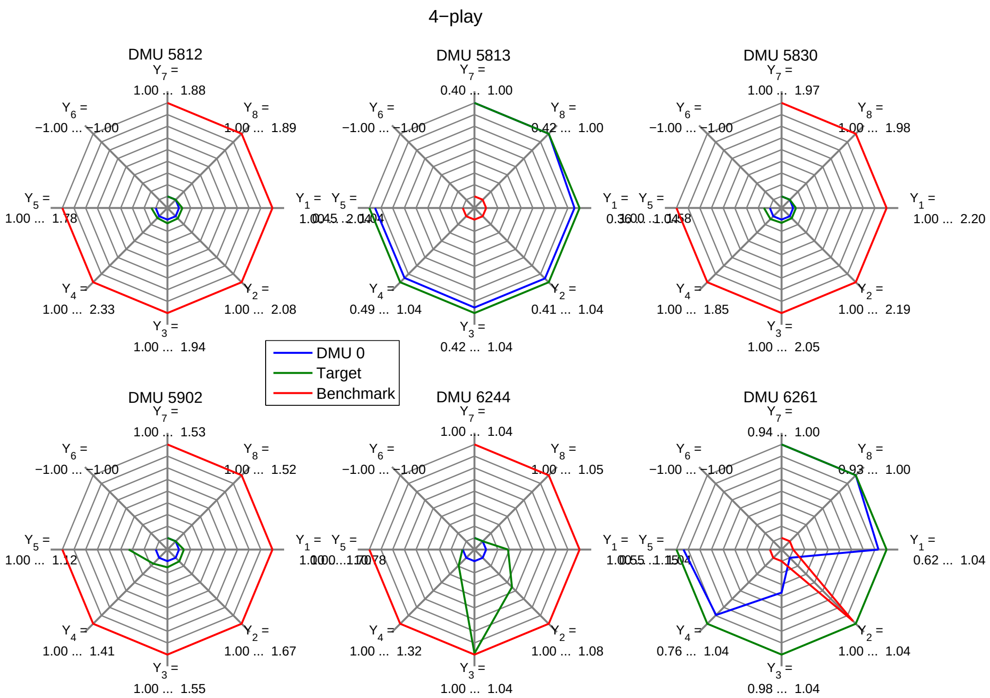


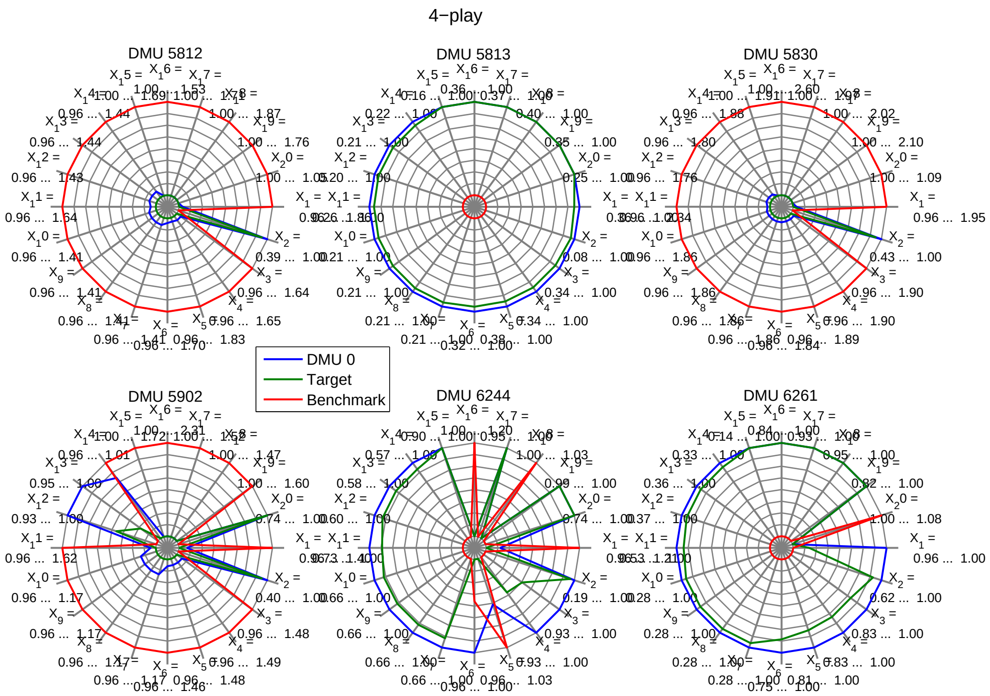
Figure 5.8: Input-output mix similarity plot for 2 inputs - 2 outputs numerical example.

scale plot (computed as in (5.5)) of this customer segment, its efficient projection and all of its dominating peers. This figure reveals some interesting information concerning the dominating peers. The dominating peers can be subdivided into three clusters according to their input-output mix similarity: (i) $5830, 5812, 5902$ and 6244 ($r \in [1.00, 1.21]$, $\theta \in [0.05, 0.11]$), (ii) 5813 ($r = 1.53, \theta = 0.24$) and (iii) 6261 ($r = 1.39, \theta = 0.45$). The customer segments in the first cluster are all rather similar to DMU 0. In contrast, DMU 5813 and 6261 are much more dissimilar than DMU 0. Looking back at Figures 5.9a-5.9b this is clearly visible: 6261 has a very different input-output mix from DMU 0 while 5813 has an identical output mix and a slightly different input mix.

As a following example we consider customer segment 1-play, socio segment E, province 6. This is the most inefficient customer segment within 1-play for $\left(\mathbf{g}_x^{C_x}, \mathbf{g}_y^{C_y}\right) = \left(|\mathbf{X}_0^{C_x}|, |\mathbf{Y}_0^{C_y}|\right)$. Figures 5.11a-5.11b show the radar plots of all outputs and inputs while Figure 5.10b shows the input-output mix similarity and scale plot. A first glance at Figure 5.10b immediately reveals that there are now three different clusters of peers according to their mix similarity: (i) $11497, 10777, 11425, 11349, 10835$ ($r \in [0.93, 2.89], \theta \in [0.01, 0.06]$) (ii) 10776 ($r = 4.06, \theta = 0.34$) and (iii) 10990 ($r = 0.21, \theta = 0.65$). Customer segment 10990 has a completely different input mix (Figure 5.11b) than DMU 0: 10990 uses large quantities of X_6 and X_7 whereas DMU 0 does not use these inputs. Customer segment 10990 also uses far less of the other inputs while generating far lower revenues than DMU 0. Output 5 (Y_5) is the exception: customer segment 10990 produces 7.08 times more of this output than DMU 0 (Figure 5.11a). Customer segment 10990 is then a highly specialized customer segment. This is also indicated by its radius of 0.21 and its

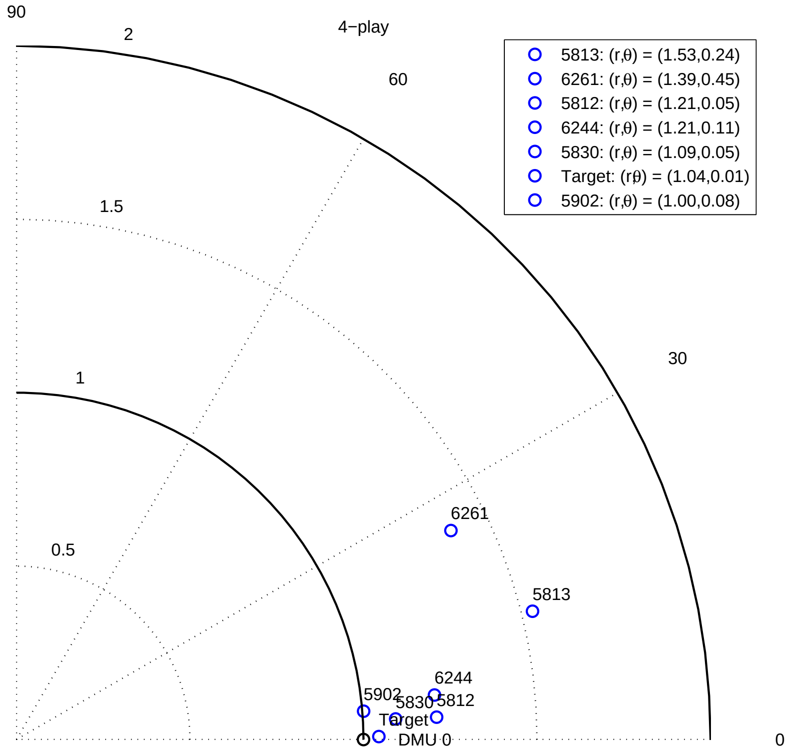


(a) Radar plot of all outputs. Output levels are relative to current output levels.

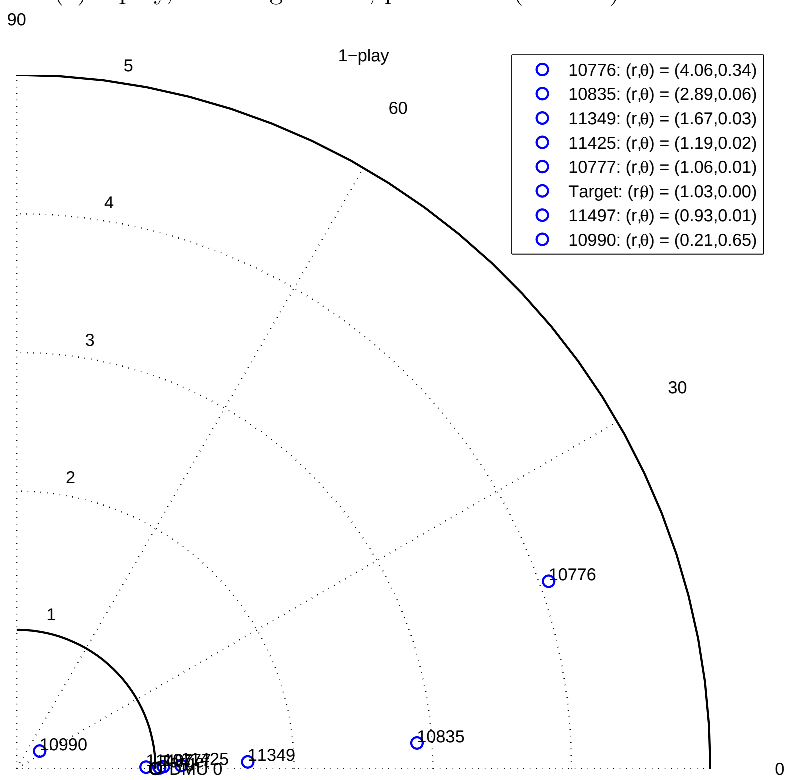


(b) Radar plot of all inputs. Input levels are relative to current input levels.

Figure 5.9: Radar plot of all inputs and outputs for customer segment 4-Play, socio segment E, province 9. From left to right and top to bottom we have benchmark DMU 5812, 5813, 5830, 5902, 6244 and 6261. The radar plot shows the current, target and benchmark input and output levels.



(a) 4-play, socio segment E, province 9 (DMU 0).



(b) 1-play, socio segment E, province 6 (DMU 0).

Figure 5.10: Input-output mix similarity and scale plot for selected customer segment with $\left(\mathbf{g}_x^{C_x}, \mathbf{g}_y^{C_y}\right) = \left(|\mathbf{X}_0^{C_x}|, |\mathbf{Y}_0^{C_y}|\right)$. The input-output mix similarity and scale of the dominating peers is computed relative to DMU 0.

angle 0.65 on Figure 5.10b. We see a similar pattern for customer segment 10776: both the input and output mix in Figure 5.11a-5.11b are very different compared to DMU 0. In contrast to 10990 it does operate at a much larger scale than DMU 0. Within the first cluster of similar customer segments DMU 11349 and 10835 operate at the larger scale. The input-output mix of the other DMUs 11425 and 10777 is only slightly different than DMU 11349 while their scale is much smaller. Finally, the input-output mix similarity of DMU 11497 equals that of DMU 10777 but its scale is smaller and lower than that of DMU 0. Both radar plots confirm this.

Therefore, it seems there are 3 options for DMU 0: (i) drastically change it input-output mix in line with 10990 at a possible cost of a reduction in scale; (ii) change its input-output mix and increase its scale in line with 10776; or (iii) increase its scale in line with 10835 or 11349 and only minor modifications to the input-output mix. The firm should look into these clusters to determine the more feasible input-output mix.

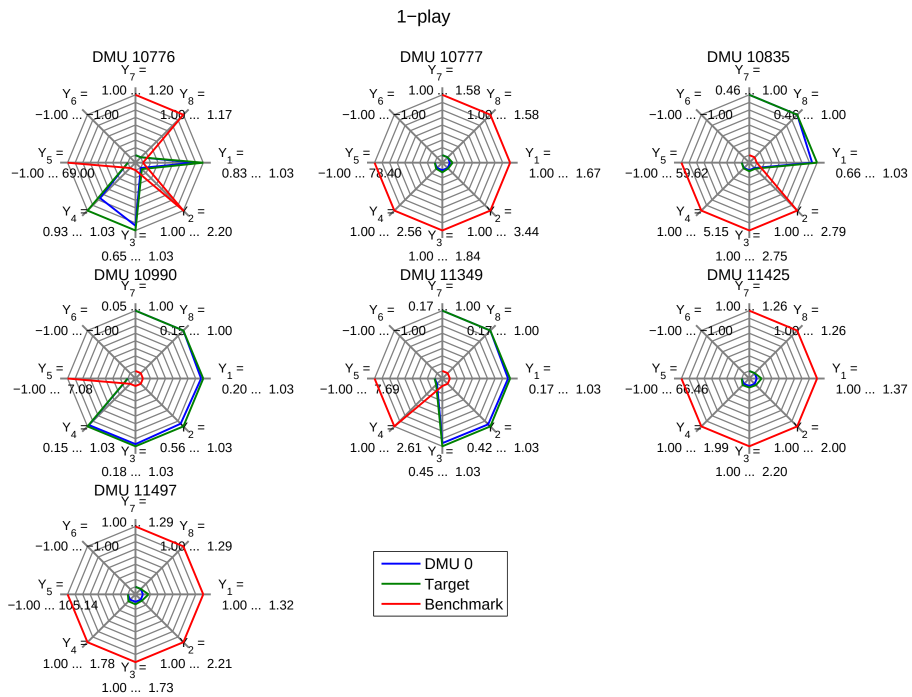
We again look at the most inefficient customer segment (4-play, socio segment E, province 6) for $\left(\mathbf{g}_x^{C_x}, \mathbf{g}_y^{C_y}\right) = \left(|\mathbf{X}_0^{C_x}|, \mathbf{e}_1|\mathbf{Y}_0^{C_y}|\right)$. To conserve space, we only consider the input-output mix similarity and scale plot in Figure 5.12a. This customer segment has 10 dominating peers. Here we see a lot of variation in both input-output mix similarity and scale such that we cannot distinguish different clusters as before. The most similar customer segment is 5830 ($r = 1.12, \theta = 0.05$) operating at a larger scale while the most dissimilar customer segment is 5763 ($r = 1.55, \theta = 0.44$). Between these extremes we find the other dominating peers. It seems that the more DMU 0 changes its input-output mix the more it could increase its scale. In the end it is up to the firm to look into these different dominating peers to decide which of them is the most feasible peer to learn from. Finally, we consider the input-output mix similarity and scale plot of customer segment 1-play, socio segment E, province 6 in Figure 5.12b. This plot is completely the same as Figure 5.10b with $\left(\mathbf{g}_x^{C_x}, \mathbf{g}_y^{C_y}\right) = \left(|\mathbf{X}_0^{C_x}|, |\mathbf{Y}_0^{C_y}|\right)$ so we will not repeat the discussion here.

5.5.3 Alternative definition of dominating peers

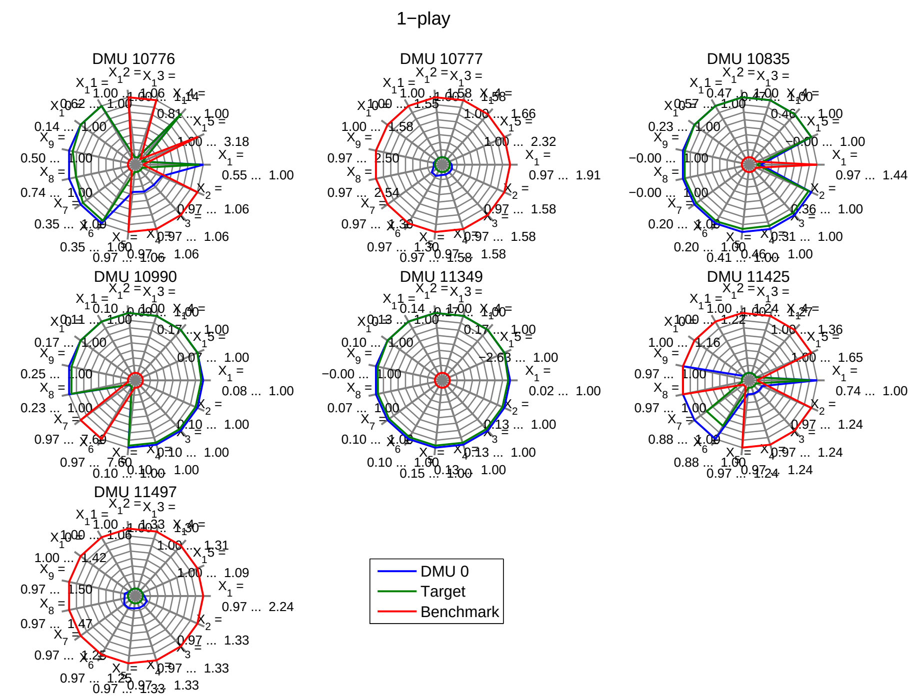
Recall the definition of D_0 as the set of peers used to form a convex combination of the hypothetical benchmark DMU. As already briefly mentioned, there is a genuine concern that not all these peers are truly dominating in a mathematical sense. Thus, one might question whether the previous analysis of comparing individual dominating peers makes sense. Moreover, one might expect that peers dominating in a mathematical sense are also more similar to DMU 0. We investigate this by modifying the definition of D_0. We modify this definition to

$$D_0^* = \{j|\mathbf{X}_j \leq \mathbf{X}_0, \mathbf{Y}_j \geq \mathbf{Y}_0\}.$$

It turns out that almost no customer segment dominates any of the evaluated customer segments (i.e., we almost always have $D_0 = \{0\}$). Therefore, we weaken the above dominance criteria by only requiring the peers to dominate on the controllable dimensions we seek to improve. First, let $\tilde{C}_x = \{j \in C_x|g_x^j > 0\}$ and $\tilde{C}_y = \left\{j \in C_y|g_y^j > 0\right\}$ be

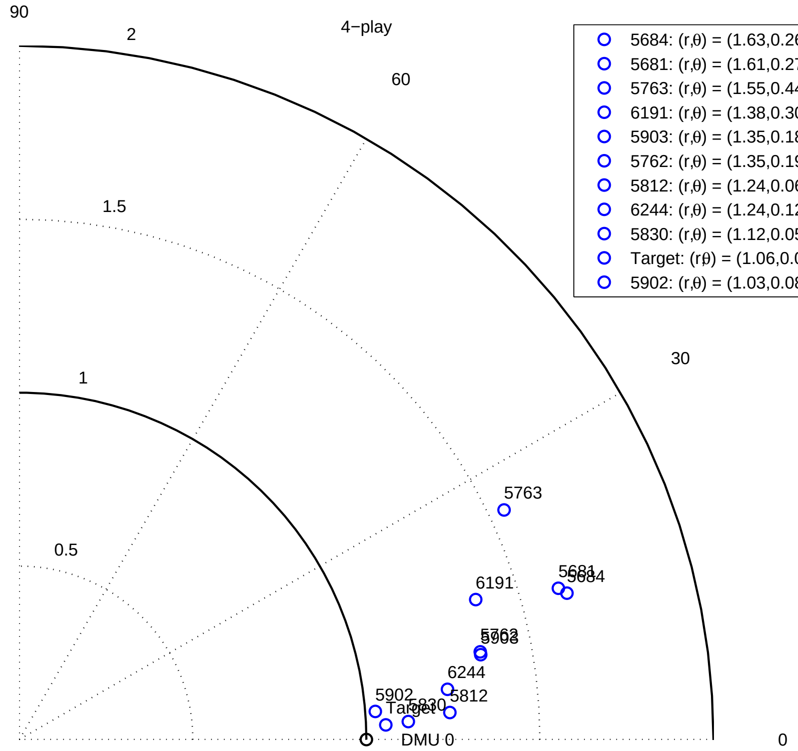


(a) Radar plot of all outputs. Output levels are relative to current output levels.

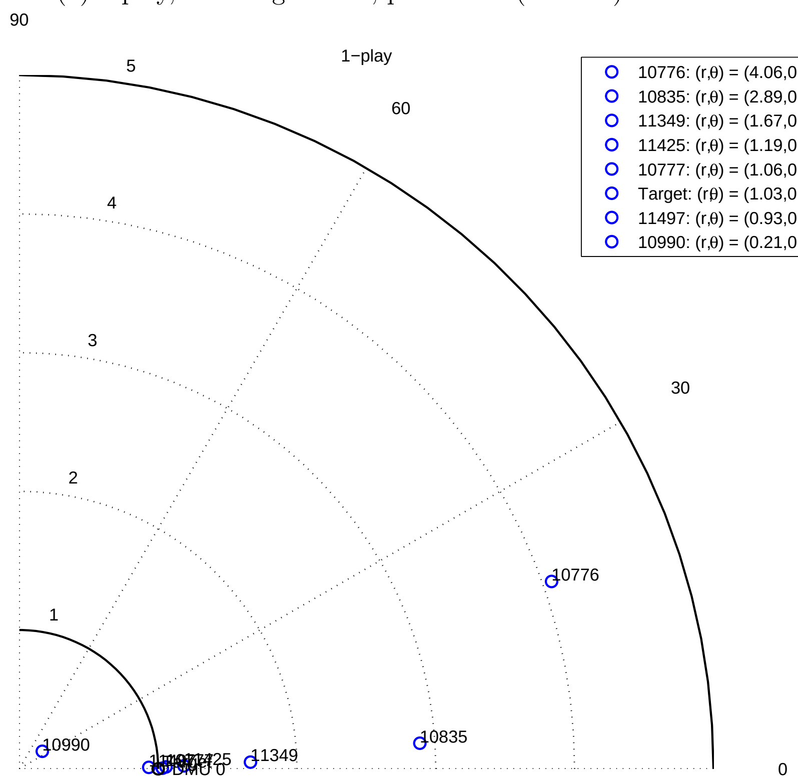


(b) Radar plot of all inputs. Input levels are relative to current input levels.

Figure 5.11: Radar plot of all inputs and outputs for customer segment 1-Play, socio segment E, province 6. From left to right and top to bottom we have benchmark DMU 10776, 10777, 10835, 10990, 11349, 11425 and 11497. The radar plot shows the current, target and benchmark input and output levels.



(a) 4-play, socio segment E, province 6 (DMU 0).



(b) 1-play, socio segment E, province 6 (DMU 0).

Figure 5.12: Input-output mix similarity and scale plot for selected customer segment with $\left(\mathbf{g}_x^{C_x}, \mathbf{g}_y^{C_y}\right) = \left(|\mathbf{X}_0^{C_x}|, \mathbf{e}_1|\mathbf{Y}_0^{C_y}|\right)$. The input-output mix similarity and scale of the dominating peers is computed relative to DMU 0.

the set of controllable inputs and outputs for which we seek improvements. Now we can define the dominating set:

$$\tilde{D}_0 = \left\{ j | \mathbf{X}_j^{\tilde{C}_x} \leq \mathbf{X}_0^{\tilde{C}_x}, \mathbf{Y}_j^{\tilde{C}_y} \geq \mathbf{Y}_0^{\tilde{C}_y} \right\}. \tag{5.6}$$

Note that (5.6) corresponds with D_0^* for $\left(\mathbf{g}_x^{C_x}, \mathbf{g}_y^{C_y}\right) = \left(|\mathbf{X}_0^{C_x}|, |\mathbf{Y}_0^{C_y}|\right)$. For the most inefficient customer segment per play pack in Table 5.4 we compare (r, θ) computed for all peers in D_0 and $\tilde{D}_0$. Summary statistics are reported in Table 5.7 for (r, θ). The number of peers in $\tilde{D}_0$ is smaller than the number of peers in D_0 for 3/5 cases and is generally small. The exception is $\tilde{D}_0$ for customer segment 2-play, socio segment E, province 1 which has 511 peers. The peers in $\tilde{D}_0$ are on average more similar and larger in scale than those in D_0 in 3/5 cases. The standard deviation column shows that there is also less variation in input-output mix similarity θ for peers in $\tilde{D}_0$ for all cases, but not in scale. Finally, the mean (r, θ) of D_0 and the mean (r, θ) of $\tilde{D}_0$ are always within two standard deviations of each other which seems to indicate that the difference is probably not statistically significant.[5]

		mean (r, θ)	std. (r, θ)
4-play, socio segment E,	$\|D_0\| = 10$	$(1.32, 0.18)$	$(0.21, 0.13)$
province 6	$\|\tilde{D}_0\| = 8$	$(1.63, 0.21)$	$(0.27, 0.09)$
3-play, socio segment E,	$\|D_0\| = 7$	$(1.21, 0.41)$	$(0.71, 0.40)$
province 6	$\|\tilde{D}_0\| = 1$	$(1.00, 0.00)$	$(0.00, 0.00)$
2-play, socio segment E,	$\|D_0\| = 12$	$(1.28, 0.22)$	$(0.48, 0.34)$
province 1	$\|\tilde{D}_0\| = 511$	$(1.97, 1.05)$	$(0.60, 0.18)$
1-play, socio segment E,	$\|D_0\| = 7$	$(1.63, 0.14)$	$(1.25, 0.23)$
province 6	$\|\tilde{D}_0\| = 2$	$(1.06, 0.01)$	$(0.08, 0.01)$
0-play, socio segment B,	$\|D_0\| = 7$	$(1.84, 0.24)$	$(0.84, 0.18)$
province 2	$\|\tilde{D}_0\| = 19$	$(1.87, 0.22)$	$(0.34, 0.06)$

Table 5.7: Comparison of input-output mix similarity and scale of D_0 and $\tilde{D}_0$ for $(\mathbf{g}_x^{C_x}, \mathbf{g}_y^{C_y}) = (|\mathbf{X}_0^{C_x}|, \mathbf{e}_1|\mathbf{Y}_0^{C_y}|)$.

5.6 Conclusions

Efficiency analysis can guide management in their (strategic) decisions to improve the firm's performance. Management decisions are made keeping certain multi-dimensional objectives in mind. Furthermore, not all objectives have equal priority and realizing

[5] We test whether the distributions of (r, θ) for D_0 and $\tilde{D}_0$ are equal (i.e., $H_0 : F_{D_0}(\cdot) = F_{\tilde{D}_0}(\cdot)$) using the nonparametric test of Li et al. (2009). This test uses kernel density estimates of the distributions to compute the test statistic. We reject the H_0 hypothesis in 4/5 cases. However, we question the outcomes of the test given that the power of the test is expected to be small in this case due to the small sample size in each of the 5 cases.

multiple objectives at once can require more complicated strategies. The direction of projection determines the objectives to focus on. Thus, the choice of these direction vectors is important. Surprisingly, empirical applications in efficiency analysis usually choose the direction vectors onto the efficient frontier without much ado. Recently, the literature has proposed some ways of determining these direction vectors in some optimal way. Although much can be said in favor of these approaches, an important issue from a management perspective is with regard to the intuitive interpretation of these optimal direction vectors. Lack of intuitive interpretation could preclude widespread usage. This paper takes a different approach. Liu et al. (2009) and Liu and Lu (2010) propose a method to further differentiate among dominating peers and identify key DMUs using eigenvector centrality. We start from the observation that the choice of direction vector determines the part of the technology frontier to which inefficient DMUs are projected. Thus, it also determines the dominating peers with which an inefficient DMU is benchmarked. Therefore, we modify their method by computing efficiency scores for different possible direction vectors and use the key DMUs to identify the most interesting objectives to focus on in subsequent analysis. These objectives are identified through a comparative analysis of the key DMUs' characteristics. The rationale being that key DMUs are the successful DMUs where the others can learn from.

After calculating efficiency scores, the next stage consists of looking into the dominating peers to determine the feasibility of the proposed benchmark. One way is to use radar plots of the inputs and outputs to compare the different dominating peers. Although radar plots can be illuminating, they have two drawbacks: (i) they quickly become overcrowded when there are many input (output) dimensions and/or when there are many observations; (ii) they do not easily allow for quickly comparing the input-output mix of various DMUs. We propose an input-output mix similarity and scale visualization tool which solves both drawbacks at the cost of the ability to compare individual input (output) dimensions. The tool is therefore intended to be used in conjunction with radar plots.

The empirical application benchmarked customer segments of a large European telecom firm using Activity-Based-Costing (ABC) data. Our preliminary analysis reveals that both mobile and fixed internet revenues are important output objectives. We then focus the main analysis on radial direction vectors with (i) all controllable inputs and outputs and (ii) all controllable inputs and mobile revenues only. This analysis learns that there is a lot of heterogeneity in inefficiency over the different customer segments. Although the level of inefficiency can be small, the associated potential profit improvements can be considerable. We next examined the individual results in more detail. We found that the number of dominating peers varied between $1 - 20$. Looking at the most inefficient customer segments of every play pack we found that socio segment E and B prevalently appear in these top 5. Only the firm can explain this pattern, but it could be intentional inefficiency by the firm as a way to tie certain demographic groups now to its services with the intention to monetize on their customer loyalty later on. Visualization of the input-output mix and scale of the dominating peers for selected customer segments using the newly proposed visualization tool revealed different courses of action for the firm to remove the present inefficiencies. We demonstrated how this new visualization tool in conjunction with radar plots enriches the toolbox of efficiency analysis to go a step beyond the conventional analysis.

We see a number of opportunities for future research. First of all, our proposed way of selecting objectives is limited to objectives with the same units of measurement. Future work could consider alternative ways of selecting these objectives when units of measurement are different. Second, it can be important to consider the direction vectors of earlier periods when choosing direction vectors. This would make the choice of direction vector an intertemporal problem where one accounts for the history of direction vector choices. This intertemporal dependence can be important as ignoring it could lead to conflicting management conclusions. Third, valuable information can be found in tracking the evolution of dominating peers over time. Visualization tools that present these temporal changes can aid decision makers. The proposed input-output mix similarity and scale visualization plot might be extended to show these temporal changes. Finally, extending the peer screening to models with output-specific technologies such as those of Cherchye et al. (2013, 2016) presents methodological challenges of its own. These arise from the fact that each output-specific technology has its own set of dominating peers. The main challenge is then to reconcile the individual results obtained from peer screening the dominating set of every output.

5.A Other choice of direction vectors

This section highlights how the efficiency results change with alternative choices for the direction vectors. To focus our discussion, we concentrate on 5 individual DMUs for every play pack. The first column of Table 5.8 lists the 5 individual DMUs of every play pack that leave the most money on the table in absolute terms. These are the same customer segments we discussed earlier in Table 5.3. The table also lists the potential change in profit of these DMUs relative to current profit as well as the individual β. The other columns show the same results for these DMUs but for other choices of direction vector. This allows to closely investigate the effect of a particular choice of direction vector on individual DMUs.

Focusing solely on mobile revenues, the company could realize similar, sometimes larger, profit improvements in many of these customer segments than when focusing on all output objectives simultaneously (i.e., column 2 versus column 1). This holds for all types of play packs. Thus, mobile represents an important source of future potential growth in profits. We next discuss some conclusions for each play pack separately.

For the 4-play customer segments similar potential profit improvements are realizable when focusing exclusively on the fixed access objective compared to focusing on the mobile objective. We note that none of these alternative direction vector choices yield similar potential profit improvements compared to focusing on all input and output objectives. There is not a clear single output objective which merits exclusive focus for the 3-play customer segments. Here it is advisable to focus on all output objectives simultaneously (column 1). In the 2-play customer segment focusing exclusively on profit improvements for fixed digital TV (column 5) generally leads to larger profit improvements than focusing exclusively on any of the other output objectives. To a lesser extent this is also true for the fixed internet objective (column 4). It is much more worthwhile to focus exclusively on mobile in the 1-play and 0-play customer segments as this delivers much larger gains in terms of profit. In terms of inefficiency scores, focusing exclusively on the other output objectives (column 3-6) leads to larger inefficiency scores than focusing exclusively on mobile output (column 2). This suggest that there is more room for improvement in those output objectives in relative terms but that mobile is a larger source of revenue. This is also clearly visible by the relative potential change in profit (denoted in brackets).

Table 5.9 lists D_0 for the same DMUs in the first column. One notices that every DMU has $6 - 12$ dominating DMUs with $\lambda_{0s} > 0.001$. The other columns show the dominating peers of these DMUs for the different choices of direction vector. This gives a better idea of the individual strengths of particular dominating peers and can provide important information for managers.

	$(\mathbf{g}_x^{C_x}, \mathbf{g}_y^{C_y}) = (\lvert\mathbf{X}_0^{C_x}\rvert, \lvert\mathbf{Y}_0^{C_y}\rvert)$		$(\mathbf{g}_x^{C_x}, \mathbf{g}_y^{C_y}) = (\lvert\mathbf{X}_0^{C_x}\rvert, \text{Mobile})$		$(\mathbf{g}_x^{C_x}, \mathbf{g}_y^{C_y}) = (\lvert\mathbf{X}_0^{C_x}\rvert, \text{Fixed access})$		$(\mathbf{g}_x^{C_x}, \mathbf{g}_y^{C_y}) = (\lvert\mathbf{X}_0^{C_x}\rvert, \text{Fixed internet})$		$(\mathbf{g}_x^{C_x}, \mathbf{g}_y^{C_y}) = (\lvert\mathbf{X}_0^{C_x}\rvert, \text{Fixed TV})$		$(\mathbf{g}_x^{C_x}, \mathbf{g}_y^{C_y}) = (\lvert\mathbf{X}_0^{C_x}\rvert, \text{Other})$	
4-play												
Rank	β	Δ Profit	β	Δ Profit	β	Δ Profit	β	Δ Profit	β	Δ Profit	β	Δ Profit
1	0.0362	57692.2	0.04513	34140.4	0.04983	31355.4	0.03708	25292.1	0.04983	28671.1	0.04912	17542.6
		(0.1211)		(0.07166)		(0.06581)		(0.05308)		(0.06018)		(0.03682)
2	0.0602	50189.3	0.08827	35120.1	0.07897	27017.7	0.06458	23328.2	0.09822	29688.7	0.09789	19233.9
		(0.2137)		(0.1495)		(0.115)		(0.09932)		(0.1264)		(0.08189)
3	0.0251	44235.2	0.04213	34506.5	0.05017	33964.6	0.02698	19678	0.04999	29704	0.05017	17956
		(0.07168)		(0.05591)		(0.05504)		(0.03189)		(0.04813)		(0.0291)
4	0.01651	28700.3	0.01657	13485.6	0.02066	13478.6	0.02066	14198.5	0.02066	11682.8	0.02056	6864.76
		(0.04354)		(0.02046)		(0.02045)		(0.02154)		(0.01772)		(0.01041)
5	0.02056	26800.7	0.03287	19985.8	0.0309	15575.2	0.02327	12274.4	0.03982	16809.1	0.03953	10215
		(0.05416)		(0.04039)		(0.03147)		(0.0248)		(0.03397)		(0.02064)
3-play												
Rank	β	Δ Profit	β	Δ Profit	β	Δ Profit	β	Δ Profit	β	Δ Profit	β	Δ Profit
1	0.07169	13010.8	0.1006	10210.3	0.2126	8403.59	0.12	10453.9	0.1676	11653.2	0.2126	8276.55
		(0.2484)		(0.1949)		(0.1604)		(0.1996)		(0.2224)		(0.158)
2	0.04355	9706.64	0.05922	7384.35	0.1101	5160.82	0.05423	5788.51	0.1099	9066.46	0.1101	5103.76
		(0.1356)		(0.1031)		(0.07209)		(0.08086)		(0.1266)		(0.07129)
3	0.02122	9587.96	0.02304	2071.53	0.02182	4767.31	0.02233	5160.45	0.02304	3950.45	0.02272	1997.2
		(0.04621)		(0.009985)		(0.02298)		(0.02487)		(0.01904)		(0.009626)
4	0.008944	7999.61	0.01062	2016.94	0.01066	4601.38	0.01066	4935.01	0.01061	3748.26	0.008944	1651.27
		(0.02044)		(0.005153)		(0.01176)		(0.01261)		(0.009577)		(0.004219)
5	0.03082	7442.99	0.05535	7385.23	0.08004	4041.66	0.03664	4309.97	0.07905	6856.29	0.08004	4023.56
		(0.09136)		(0.09065)		(0.04961)		(0.0529)		(0.08415)		(0.04939)
2-play												
Rank	β	Δ Profit	β	Δ Profit	β	Δ Profit	β	Δ Profit	β	Δ Profit	β	Δ Profit
1	0.02687	7104.92	0.05211	3352.36	0.05375	3482.68	0.02774	5171.93	0.04987	6683.1	0.05253	3277.05
		(0.07672)		(0.0362)		(0.03761)		(0.05585)		(0.07216)		(0.03538)
2	0.03013	7018.49	0.05237	3023.85	0.05454	3197.25	0.03284	5421.09	0.04906	5748.53	0.05446	3017.31
		(0.08675)		(0.03738)		(0.03952)		(0.06701)		(0.07105)		(0.03729)
3	0.02251	6368.99	0.02528	1996.01	0.02258	4736.69	0.02528	1983.25	0.02528	3803.05	0.02519	1971.26
		(0.1015)		(0.03182)		(0.07551)		(0.03162)		(0.06063)		(0.03142)
4	0.02135	4808.93	0.04379	2948.56	0.02152	3664.44	0.04442	2977.64	0.04442	5390.24	0.04442	2972.79
		(0.1146)		(0.07027)		(0.08733)		(0.07096)		(0.1285)		(0.07084)
5	0.02334	4043.11	0.03142	1275.3	0.03263	1318.73	0.02381	2929.73	0.03263	2800.51	0.03263	1272.07
		(0.06427)		(0.02027)		(0.02096)		(0.04657)		(0.04452)		(0.02022)
1-play												
Rank	β	Δ Profit	β	Δ Profit	β	Δ Profit	β	Δ Profit	β	Δ Profit	β	Δ Profit
1	0.02994	25547.3	0.02994	25496.8	0.05232	5781.87	0.05232	5782.86	0.05232	5773.43	0.05232	5751.05
		(0.08823)		(0.08806)		(0.01997)		(0.01997)		(0.01994)		(0.01986)
2	0.05823	16530.7	0.05823	16481.9	0.06681	1799.32	0.06681	1791.7	0.06681	1784.49	0.06681	1779.87
		(0.1091)		(0.1088)		(0.01188)		(0.01183)		(0.01178)		(0.01175)
3	0.02254	14384.7	0.02304	14616.6	0.03344	2240.32	0.03344	2232.14	0.03359	2219	0.03359	2209.48
		(0.04561)		(0.04635)		(0.007104)		(0.007078)		(0.007037)		(0.007006)
4	0.03555	8992.04	0.03555	8949.41	0.04308	1237.21	0.04308	1231.48	0.04308	1227.93	0.04308	1228.22
		(0.05624)		(0.05597)		(0.007738)		(0.007702)		(0.00768)		(0.007682)
5	0.01541	5217.68	0.01541	5193.4	0.01747	648.253	0.01747	650.845	0.01739	642.924	0.01747	640.104
		(0.02425)		(0.02414)		(0.003013)		(0.003025)		(0.002989)		(0.002976)
0-play												
Rank	β	Δ Profit	β	Δ Profit	β	Δ Profit	β	Δ Profit	β	Δ Profit	β	Δ Profit
1	0.1009	947.403	0.131	996.443	0.1176	299.674	0.1244	310.774	0.1419	353.145	0.1419	275.755
		(0.1281)		(0.1347)		(0.04051)		(0.04201)		(0.04774)		(0.03728)
2	0.1076	827.986	0.1161	768.319	0.1656	276.309	0.1719	284.697	0.1781	296.704	0.1781	232.307
		(0.1296)		(0.1202)		(0.04324)		(0.04455)		(0.04643)		(0.03635)
3	0.1278	760.566	0.1283	679.658	0.1798	207.654	0.1798	196.234	0.1798	219.894	0.1788	173.66
		(0.1486)		(0.1328)		(0.04058)		(0.03835)		(0.04297)		(0.03394)
4	0.1544	704.489	0.1596	657.659	0.1788	182.839	0.1788	176.696	0.1788	176.352	0.1755	170.333
		(0.1845)		(0.1722)		(0.04787)		(0.04626)		(0.04617)		(0.0446)
5	0.1741	689.124	0.1761	622.219	0.2317	209.588	0.2358	188.965	0.2358	210.283	0.234	177.217
		(0.2132)		(0.1925)		(0.06485)		(0.05847)		(0.06506)		(0.05483)

Table 5.8: Worst 5 DMUs per play pack in terms of ΔProfit for $\left(\mathbf{g}_x^{C_x}, \mathbf{g}_y^{C_y}\right) = \left(|\mathbf{X}_0^{C_x}|, |\mathbf{Y}_0^{C_y}|\right)$ (first column). For every choice of direction vectors the table shows β and ΔProfit for the same 5 DMUs as in the first column. Relative ΔProfit with respect to current profit is in brackets.

	$(\mathbf{g}_x^{C_x}, \mathbf{g}_y^{C_y}) = (\lvert\mathbf{X}_0^{C_x}\rvert, \lvert\mathbf{Y}_0^{C_y}\rvert)$	$(\mathbf{g}_x^{C_x}, \mathbf{g}_y^{C_y}) = (\lvert\mathbf{X}_0^{C_x}\rvert, \text{Mobile})$	$(\mathbf{g}_x^{C_x}, \mathbf{g}_y^{C_y}) = (\lvert\mathbf{X}_0^{C_x}\rvert, \text{Fixed access})$	$(\mathbf{g}_x^{C_x}, \mathbf{g}_y^{C_y}) = (\lvert\mathbf{X}_0^{C_x}\rvert, \text{Fixed internet})$	$(\mathbf{g}_x^{C_x}, \mathbf{g}_y^{C_y}) = (\lvert\mathbf{X}_0^{C_x}\rvert, \text{Fixed TV})$	$(\mathbf{g}_x^{C_x}, \mathbf{g}_y^{C_y}) = (\lvert\mathbf{X}_0^{C_x}\rvert, \text{Other})$
4-play						
Rank	D_0	D_0	D_0	D_0	D_0	D_0
1	5812, 5813, 5830, 5902, 6244, 6261,	5762, 5763, 5764, 5812, 5837, 5902, 6191, 6244,	5763, 5812, 5823, 5902, 6191, 6244,	5811, 5812, 5902, 6191, 6244, 6261,	5763, 5812, 5823, 5902, 6191, 6244,	5763, 5812, 5823, 5902, 6191, 6244,
2	5666, 5672, 5684, 5811, 5813, 5830, 5903, 6244, 6261,	5681, 5684, 5762, 5763, 5812, 5830, 5902, 5903, 6191, 6244,	5684, 5762, 5811, 5902, 5980, 6028, 6191, 6197, 6244,	5609, 5654, 5672, 5684, 5811, 5813, 5903, 6244,	5684, 5762, 5763, 5764, 5811, 5812, 5902, 6191, 6244,	5666, 5684, 5762, 5763, 5811, 5902, 5903, 6191, 6244,
3	5654, 5681, 5686, 5812, 5823, 5902, 6261,	5620, 5654, 5681, 5686, 5764, 5812, 5902, 6261,	5654, 5681, 5686, 5764, 5812, 5902, 6261,	5654, 5681, 5686, 5811, 5812, 5823, 5902, 6261,	5654, 5656, 5681, 5686, 5764, 5812, 5902, 6261,	5654, 5681, 5686, 5764, 5812, 5902, 6261,
4	5644, 5680, 5681, 5686, 5794, 5812, 5902, 6244, 6261,	5644, 5680, 5681, 5686, 5794, 5812, 5902, 6244, 6261,	5680, 5681, 5686, 5794, 5812, 5902, 6244, 6261,	5680, 5681, 5686, 5794, 5812, 5902, 6244, 6261,	5680, 5681, 5686, 5794, 5812, 5902, 6244, 6261,	5680, 5681, 5686, 5794, 5812, 5902, 6244, 6261,
5	5620, 5666, 5681, 5685, 5686, 5757, 5812, 5830, 6244,	5618, 5620, 5621, 5666, 5681, 5686, 5757, 5764, 5812, 6244,	5618, 5620, 5668, 5681, 5685, 5686, 5812, 6028, 6244,	5620, 5666, 5681, 5685, 5686, 5757, 5812, 5830, 6244,	5618, 5620, 5621, 5666, 5681, 5685, 5686, 5757, 5764, 5812, 6244,	5618, 5620, 5621, 5666, 5681, 5685, 5686, 5757, 5764, 5812, 6244,
3-play						
Rank	D_0	D_0	D_0	D_0	D_0	D_0
1	1669, 1706, 1710, 2000, 5421, 6446, 6464, 8459,	1642, 1669, 1706, 1715, 1854, 6464, 8459,	1634, 1706, 1853, 4756, 6412, 6446, 6467, 7942,	1649, 1706, 1854, 2000, 5481, 6446,	1634, 1706, 1853, 4756, 6412, 7954, 7973,	1634, 1706, 1853, 4756, 6412, 6446, 6467, 7942,
2	1706, 1852, 1854, 1862, 5482, 6607,	1706, 1708, 1718, 1792, 1852, 1854, 1924, 2097, 6446, 6556,	1706, 1718, 1852, 1854, 1862, 2097, 6465, 6556, 6970,	1706, 1718, 1852, 1854, 1862, 5194, 6980,	1706, 1718, 1850, 1852, 1854, 1862, 6465, 6556, 6970,	1706, 1718, 1852, 1854, 1862, 2097, 6465, 6556, 6970,
3	4830, 4964, 5039, 5109, 5188, 5253, 5254, 5476, 8129,	4828, 4830, 5039, 5109, 5111, 5181, 5188, 5254, 5405, 5476,	4828, 4830, 4964, 5039, 5109, 5111, 5188, 5254, 5476,	4830, 4893, 5039, 5109, 5254, 5476, 6914, 8129,	4828, 4830, 5039, 5109, 5111, 5181, 5188, 5254, 5405, 5476,	4828, 4830, 5039, 5109, 5111, 5181, 5188, 5254, 5405, 5476,
4	4882, 4894, 4978, 5038, 5109, 5111, 5254, 5476,	4882, 4894, 4978, 5038, 5109, 5111, 5254, 5476,	4882, 4894, 4978, 5038, 5109, 5111, 5254, 5476,	4882, 4894, 4978, 5038, 5109, 5111, 5254, 5476,	4882, 4894, 4978, 5038, 5109, 5111, 5254, 5476,	4882, 4894, 4978, 5038, 5109, 5111, 5254, 5476,
5	1706, 1718, 1854, 1862, 5194, 6411, 6446, 6464, 6980,	1706, 1718, 1852, 1854, 1862, 6464, 6465, 6556,	1718, 1854, 1862, 5194, 6411, 6464, 6465, 6556,	1718, 1854, 1862, 5194, 6411, 6446, 6464, 6980,	1706, 1718, 1854, 1862, 5194, 6411, 6446, 6465, 6980,	1718, 1854, 1862, 5194, 6411, 6464, 6465, 6556,
2-play						
Rank	D_0	D_0	D_0	D_0	D_0	D_0
1	914, 916, 929, 932, 934, 1491, 4473,	914, 916, 929, 932, 934, 1060, 1072, 1073, 1150, 1491, 4160, 4689,	914, 916, 929, 932, 934, 1060, 1077, 1491, 4160, 4473, 4689,	914, 916, 929, 932, 934, 1077, 1491, 4473,	914, 916, 929, 932, 934, 1060, 1072, 1073, 1491, 4160, 4689,	914, 916, 929, 932, 934, 1060, 1072, 1491, 4160, 4689, 4690,
2	914, 932, 1013, 1060, 1061, 1292, 1490, 1502, 4160, 4689, 4690,	914, 932, 1013, 1060, 1372, 1502, 1514, 1565, 4547, 4689,	914, 932, 1013, 1060, 1372, 1502, 1514, 1565, 4160, 4547, 4689,	914, 932, 1013, 1060, 1061, 1077, 1292, 1502, 4160, 4689, 4690,	914, 932, 1013, 1060, 1372, 1502, 1514, 4547, 4689,	914, 932, 1013, 1060, 1372, 1502, 1514, 1565, 4547, 4689,
3	1233, 3073, 3991, 4114, 7412, 7526, 7634, 7706, 8935, 10694,	1233, 3073, 3991, 4114, 7412, 7526, 7706, 7916, 8935, 10694,	1233, 3073, 3991, 4114, 7412, 7526, 7634, 7706, 8935, 10694,	1233, 3073, 3991, 4114, 7412, 7526, 7706, 7916, 8935, 10694,	1233, 3073, 3991, 4114, 7412, 7526, 7706, 7916, 8935, 10694,	1233, 3073, 3991, 4114, 7412, 7526, 7706, 7916, 8935, 10694,
4	1233, 4114, 7155, 7412, 7418, 7634, 7742, 7844, 9358,	1233, 4114, 7155, 7412, 7418, 7634, 7742, 7844, 9358,	1233, 4114, 7155, 7412, 7418, 7634, 7742, 7844, 9358,	1233, 4114, 7155, 7412, 7418, 7634, 7742, 7844, 8935, 9358,	1233, 4114, 7155, 7412, 7418, 7634, 7742, 7844, 8935, 9358,	1233, 4114, 7155, 7412, 7418, 7634, 7742, 7844, 8935, 9358,
5	932, 934, 1292, 1366, 1502, 4617, 4689,	932, 1077, 1490, 1502, 1565, 4547,	932, 1077, 1292, 1490, 1502, 4547,	932, 934, 1061, 1292, 1366, 1502, 4617, 4689,	932, 1077, 1292, 1490, 1502, 4547,	932, 1077, 1292, 1490, 1502, 4547,
1-play						
Rank	D_0	D_0	D_0	D_0	D_0	D_0
1	10776, 10777, 10835, 10990, 11349, 11425, 11497,	10776, 10777, 10835, 10990, 11349, 11425, 11497,	10777, 10847, 10849, 10990, 10992, 11349, 11425,	10777, 10847, 10849, 10990, 10992, 11349, 11425,	10777, 10847, 10849, 10990, 10992, 11349, 11425,	10777, 10847, 10849, 10990, 10992, 11349, 11425,
2	212, 10829, 10835, 10847, 11200, 11349, 11355, 11415, 11424,	212, 10829, 10835, 10847, 11200, 11349, 11355, 11415, 11424,	212, 10829, 10835, 10840, 10847, 10848, 11349, 11355,	212, 10829, 10835, 10840, 10847, 10848, 11349, 11355,	212, 10829, 10835, 10840, 10847, 10848, 11349, 11355,	212, 10829, 10835, 10840, 10847, 10848, 11349, 11355,
3	10775, 10776, 10829, 10835, 10840, 10847, 10849, 10981, 10984, 11424,	10775, 10776, 10829, 10835, 10840, 10847, 10849, 10981, 10984, 11424,	10829, 10835, 10840, 10843, 10847, 10849, 10981, 10992, 11424,	10829, 10835, 10840, 10843, 10847, 10849, 10981, 10992, 11424,	10829, 10835, 10840, 10843, 10847, 10849, 10981, 10992, 11424,	10829, 10835, 10840, 10843, 10847, 10849, 10981, 10992, 11424,
4	3198, 10775, 10776, 10779, 10829, 10835, 10847, 10848, 11200, 11202, 11355, 11424,	3198, 10775, 10776, 10779, 10829, 10835, 10847, 10848, 11200, 11202, 11355, 11424,	10757, 10775, 10779, 10840, 10847, 10848, 11343, 11349, 11355, 11424,	10757, 10775, 10779, 10840, 10847, 10848, 11343, 11349, 11355, 11424,	10757, 10775, 10779, 10840, 10847, 10848, 11343, 11349, 11355, 11424,	10757, 10775, 10779, 10840, 10847, 10848, 11343, 11349, 11355, 11424,
5	362, 10775, 10829, 10835, 10840, 10848, 10994, 11355, 11424,	362, 10775, 10829, 10835, 10840, 10848, 10994, 11355, 11424,	11, 10775, 10835, 10840, 10841, 10847, 10848, 11349, 11355, 11424,	11, 10775, 10835, 10840, 10841, 10847, 10848, 11349, 11355, 11424,	11, 10775, 10835, 10840, 10841, 10848, 11349, 11355, 11424,	11, 10775, 10835, 10840, 10841, 10847, 10848, 11349, 11355, 11424,
0-play						
Rank	D_0	D_0	D_0	D_0	D_0	D_0
1	11569, 11573, 11767, 11773, 11801, 11945, 12078, 12199,	11569, 11773, 11945, 12132, 12199, 12211, 12283,	11653, 11767, 11773, 11945, 12078, 12199,	11573, 11767, 11773, 11801, 11851, 11945, 12199,	11569, 11767, 11773, 11801, 11945, 12199, 12211,	11569, 11767, 11773, 11801, 11945, 12199, 12211,
2	11570, 11624, 11725, 12095, 12115, 12132, 12146, 12199, 12205,	11570, 11624, 11725, 12095, 12115, 12132, 12146, 12199, 12205,	11570, 11624, 11796, 12095, 12115, 12132, 12199, 12205, 12211,	11570, 11624, 11796, 12095, 12115, 12132, 12199, 12205, 12211,	11570, 11624, 11796, 12095, 12115, 12132, 12199, 12205, 12211,	11570, 11624, 11796, 12095, 12115, 12132, 12199, 12205, 12211,
3	11539, 11563, 11569, 11598, 11624, 12115, 12136, 12199,	11539, 11563, 11569, 11598, 11624, 12115, 12136, 12199,	11539, 11563, 11569, 11598, 11624, 12199, 12211, 12283,	11539, 11563, 11569, 11598, 11624, 12199, 12211, 12283,	11539, 11563, 11569, 11598, 11624, 12199, 12211, 12283,	11539, 11563, 11569, 11598, 11624, 12199, 12211, 12283,
4	11598, 11600, 11624, 11846, 11917, 12103, 12283,	11552, 11598, 11600, 11624, 11846, 11917, 12103, 12283,	11600, 11624, 11705, 11893, 11917, 12055, 12283,	11600, 11624, 11705, 11893, 11917, 12055, 12283,	11600, 11624, 11705, 11893, 11917, 12055, 12283,	11600, 11624, 11705, 11893, 11917, 12055, 12283,
5	11558, 11570, 11624, 11744, 11756, 11888, 11911, 11943, 12211,	11570, 11624, 11756, 11846, 11888, 11911, 11943, 12206, 12211,	11570, 11624, 11911, 11940, 11941, 11943, 12073, 12211,	11570, 11624, 11911, 11943, 12073, 12136, 12211,	11570, 11624, 11911, 11943, 12073, 12136, 12211,	11570, 11624, 11911, 11943, 12073, 12136, 12211,

Table 5.9: D_0 of the worst 5 DMUs per play pack in terms of potential change in profit for $\left(\mathbf{g}_x^{C_x}, \mathbf{g}_y^{C_y}\right) = \left(\lvert\mathbf{X}_0^{C_x}\rvert, \lvert\mathbf{Y}_0^{C_y}\rvert\right)$ (first column). For every choice of direction vectors the table shows dominating peers for the same 5 DMUs.

General conclusions

The aim of this thesis has been to go beyond black box modeling in production theory. This was done by enhancing the realism of the black box model itself and by providing some tools to analyze the driving factors behind efficiency and productivity results. Furthermore, we went beyond conventional efficiency analysis by analyzing what one can learn from dominating peers and offering a visualization tool in the process. Let us now reflect on our contributions and discuss some of the limitations before discussing some future avenues for research.

Contributions and limitations

Chapter 2 modeled mixed farms using network DEA and introduced a coordination productivity measure which measures potential gains in productivity due to reallocation of resources from one activity to the other. In addition, we decomposed this coordination productivity into a coordination technical change component and a coordination technical inefficiency change component which allows to assess the impacts of reallocation on the different sources of productivity. The empirical application focused on a large panel of English and Welsh farms over the period 2007–2013. The results showed that coordination inefficiency significantly increases with the proportion of resources allocated to livestock production in economic and statistical terms. Coordination inefficient farms should generally allocate more land to crop production. Depending on the region, the average coordination productivity growth ranges from -9.7% to 15.9% per year. It is driven by coordination technical change rather than coordination inefficiency change.

We see three main limitations to our work. First, we assumed perfect substitutability between crop land and livestock land. By combining current data with new data sources on land characteristics, future work could relax this assumption of perfect land substitutability and account for the fact that not all land is equally suited for both farming activities. Second, we did not really account for stochastic factors that impact agricultural production. Weather conditions can significantly impact farm production so that efficiency is biased upwards (downwards) as a result of good (bad) weather conditions. We partly accounted for this by running our nonparametric models per region. Third, we did not model intertemporal linkages between crop and livestock activities because of data limitations. One obvious intertemporal link is manure, a by-product of livestock, which is used as fertilizer in crop farming in a subsequent period. Recently, the Farm Business Survey started to include data on manure so that over time one could repeat the analysis with manure data.

Chapter 3 provided a decomposition of the Luenberger-Hicks-Moorsteen total factor productivity indicator into components of technical change, technical inefficiency change and scale inefficiency change. The decomposition requires only mild conditions on the underlying technology. This makes it applicable to a wide range of production technologies. The Luenberger-Hicks-Moorsteen total factor productivity indicator is additively complete: output contributions to TFP can be disentangled from input contributions to TFP. The empirical application focused on the agricultural sector at the state-level in the U.S. over the period $1960 - 2004$. In line with the literature we found that TFP increased substantially over the considered period and that technical change is the main driver of TFP, but that the importance of the technical inefficiency change and the scale inefficiency change components depends on the convexity assumption of the production technology. We also found that TFP growth is mostly due to output growth rather than input decline: the agricultural sector was able to produce more with the same amount of resources.

There are a number of limitations to our work. First, we did not account for possible intertemporal linkages in the modeling of technology in the empirical application. Capital and land are prime examples of "durable" inputs (see Chapter 4) which link production in subsequent periods. Accounting for this in modeling of the technology likely affects the TFP results and its decomposition. Second, the residual approach in defining the scale inefficiency change component differs from the conventional CRS-VRS approach. The accuracy of this residual approach depends on the "step-size", while the CRS-VRS approach uses a (hypothetical) CRS benchmark susceptible to a few (extreme) observations. The first question pertains to the definition of the scale inefficiency change component: can one use the CRS-VRS approach instead of the residual approach and, if so, what is the interpretation of the remaining component? Next, it is of practical interest to determine the conditions under which one should prefer one approach over the other via simulations and empirical applications.

Chapter 4 presented a nonparametric framework of dynamic cost minimization with durable and storable inputs. Both types of inputs link production in different time periods together. We explicitly modeled the possibility that different vintages of durables are used and we allow for production delays. We present a nonparametric test for consistency with dynamic cost minimization and quantify the degree of inefficiency. We further decompose this inefficiency measure into period-specific cost inefficiencies. The framework is illustrated with an application on Swiss railway companies.

Future work can expand on this work in multiple directions. First, one could further combine our dynamic cost minimization framework with the nonparametric methodology for multi-output production analysis of Cherchye et al. (2013, 2014). This would account for interdependencies between different output production processes through joint inputs while also enabling us to exclusively assign inputs to specific outputs. Second, our framework can be used in combination with various productivity measures (such as the LHM TFP indicator of Chapter 3). This combination will lead to richer productivity analyses because it explicitly accounts for intertemporal production interdependencies through storable and durable inputs. Third, the focus of this chapter was on efficiency under the assumption of perfect price foresight: i.e., our efficiency measures are based on solutions of LPs that assume the evaluated firm correctly predicts the prevailing prices. In practice, this assumption rarely holds so that inefficiency can, at least partially, be explained by failure of this assumption. From a regulator perspective, it is therefore useful to consider efficiency measures allowing for price uncertainty where one considers efficiency under all possible (realistic) price situations. This would ensure that new regulation schemes by the regulator are not too harsh or too loose for the individual firms. A good starting point towards integration in our framework is Kuosmanen and Post (2002).

Finally, Varian (1982) has developed a nonparametric approach to consumer demand analysis that is formally analogous to the nonparametric approach to production analysis to which we adhere here. Following this analogy, we may translate the insights developed in this chapter towards a consumption setting to obtain a more realistic modeling of intertemporal aspects of consumer behavior.[6] Specifically, our concept of storable inputs corresponds to the notion of infrequent purchases in a consumption context, and durable inputs are similar in spirit to durable consumption goods (for example, cars, houses, etc.) in a demand setting.

Chapter 5 identified direction vectors from key DMUs. These key DMUs are the important (dominating) peers among all DMUs in the sample. Thus, key DMUs are those DMUs which are frequently part of the virtual benchmark DMU constructed when benchmarking DMUs under a variety of different direction vectors. Through a comparative analysis of these key DMUs' characteristics we identified key objectives which are then used as the direction vectors. We further presented an easy to compute visualization tool to compare the input-output mix and scale of dominating peers to the benchmarked DMU. These tools were put into practice in the empirical application where we benchmarked all the customer segments of a telecom operator.

We see four immediate directions for future research. First of all, our proposed way of selecting objectives is limited to objectives with the same units of measurement. Future work could consider alternative ways of selecting these objectives when units of measurement are different. Second, it can be important to consider the direction vectors of earlier periods when choosing direction vectors. This would make the choice of direction vector an intertemporal problem where one accounts for the history of direction vector choices. This intertemporal dependence can be important as ignoring it could lead

[6]See, for example Crawford (2010) and Crawford and Polisson (2014) for recent contributions to the nonparametric analysis of intertemporal consumer behavior.

to conflicting management conclusions over time. Third, valuable information can be found in tracking the evolution of dominating peers over time. Visualization tools that present these temporal changes can aid decision makers. The proposed input-output mix similarity and scale visualization plot might be extended to show these temporal changes. Finally, extending the peer screening to models with output-specific technologies such as those of Cherchye et al. (2013, 2016) presents challenges of its own. These arise from the fact that each output-specific technology has its own set of dominating peers. The main challenge is then to reconcile the individual results obtained from peer screening the dominating set of every output.

Future research avenues

In addition to the already mentioned extensions, we see a number of interesting avenues for future research. Naturally, these potential future research avenues are but a small sample of interesting topics for future work. The selection below merely constitutes research topics which we find particularly interesting.

First of all, we could further increase the realism of our models by using output-specific technologies which have output-specific inputs and joint inputs as in Cherchye et al. (2013, 2016). These joint inputs link the different output-specific technologies. In particular, the integration of these output-specific technologies in the by-production approach of Murty et al. (2012) to model pollution-generating technologies comes to mind. One example which can serve as a starting point is Cherchye et al. (2015) who use the output-specific technologies of Cherchye et al. (2013) with bad outputs by applying a transformation to the bad outputs. Furthermore, where possible and appropriate we should account for the intertemporal dependencies that exist in production.

Second, we might need to concede that the inefficiency we measure can sometimes be intentional. Bogetoft and Hougaard (2003) argued that inefficiency can be an indirect, on-the-job compensation to agents in an organization to keep them motivated or to create loyal workers. Asmild et al. (2013), for example, find a systematic pattern in slacks between staff groups. In this respect slacks can be the result of a deliberate, rational choice. This provides an opportunity to combine revealed preferences methodology with nonparametric production methodology to separate “rational inefficiency” from “irrational inefficiency”. The revealed preferences methodology can be used to check whether the measured amounts of inefficiency are consistent with a utility maximizing individual or utility maximizing collective decision maker. Deviations from rational choices can then help in separating rational inefficiency from the irrational kind.

Third, competition analysis often relies on parametric assumptions, for example, to detect collusion. Recently, a number of papers such as Carvajal et al. (2013, 2014) started to explore nonparametric restrictions to test assumptions on market behavior. It is possible that economic efficiency measures could also be used in competition analysis. Profit efficiency analysis relies on the assumption of perfect competition while cost efficiency analysis relaxes this assumption for the output prices. The difference between profit efficiency and cost efficiency may convey information about the degree or type of competition. This research avenue may be of considerable interest to competition authorities and regulators.

Finally in a world of big data, economics faces – like nearly every other discipline – major opportunities and challenges. Ever more data becomes available in the future for researchers which is both a blessing and a curse. Massive amounts of data – hopefully – enable us to provide better answers to our research questions while posing challenges of its own. These problems include devising efficient algorithms that improve computational complexity, visualization of results, detecting patterns in results and translating this in clear cut messages for management. Tackling these issues will be critical in order to deal with the availability of ever more data in the future.

Visualization will become ever more important for data exploration and analysis of results. Due to the large amount of results and in order to summarize the most important managerial implications, it is worthwhile in future research to look into computational techniques to detect patterns. For example: Dai and Kuosmanen (2014) combine clustering methods with efficiency analysis to identify absolute and relative benchmarks within every cluster of DMUs.

As a consequence of duality the computational problems in the linear programs can differ: either the optimization variables or the number of constraints scale with the number of observations. Exponential amounts of data then result in exponential number of optimization variables or constraints. Both issues require different solutions. The problem is less of an issue for enumeration algorithms typically encountered with non-convex technologies. One possible solution therefore can be to use a non-convex technology as a preprocessing step to select a subsample of the data as comparison partners in the efficiency analysis. The subsample then only contains the observations marked as efficient under a non-convex technology which are potentially also efficient under a convex technology. This approach would reduce the number of constraints or optimization variables in the linear programs.

Bibliography

Adler, N. and A. Raveh (2008): "Presenting DEA graphically," *Omega*, 36, 715–729.

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

Andersen, J. L. and P. Bogetoft (2007): "Gains from quota trade: theoretical models and an application to the Danish fishery," *European Review of Agricultural Economics*, 34, 105–127.

Ang, F. and P. J. Kerstens (2016): "To Mix or Specialise? A Coordination Productivity Indicator for English and Welsh farms," *Journal of Agricultural Economics*, 67, 779–798.

——— (2017a): "Decomposing the Luenberger-Hicks-Moorsteen Total Factor Productivity Indicator: an application to U.S. agriculture," *European Journal of Operational Research*, 260, 359–375.

——— (2017b): "Exact and superlative measurement of the Luenberger-Hicks-Moorsteen productivity indicator," Tech. Rep. 17.02, KU Leuven, Department of Economics.

Aparicio, J., J. T. Pastor, and F. Vidal (2016): "The directional distance function and the translation invariance property," *Omega*, 58, 1–3.

Aragon, Y., A. Daouia, and C. Thomas-Agnan (2005): "Nonparametric Frontier Estimation: A Conditional Quantile-Based Approach," *Econometric Theory*, 21, 358–389.

Areal, F. J., K. Balcombe, and R. Tiffin (2012): "Integrating spatial dependence into Stochastic Frontier Analysis," *Australian Journal of Agricultural and Resource Economics*, 56, 521–541.

Asmild, M., P. Bogetoft, and J. L. Hougaard (2013): "Rationalising inefficiency: Staff utilisation in branches of a large Canadian bank," *Omega*, 41, 80–87.

Atkinson, S. E. and M. G. Tsionas (2016): "Directional distance functions: Optimal endogenous directions," *Journal of Econometrics*, 190, 301–314.

Balk, B., J. Blank, J. Bos, and P. Koot (2010): "Onderzoek naar de generieke productiviteitsontwikkeling in de reguleringsmethode van Tennet TSO BV: Rapport voor de Energiekamer van de NMa," Tech. rep., Universiteit Delft, Instituut voor Public Sector Efficiency.

Balk, B. M. (1998): *Industrial price, quantity, and productivity indices: the microeconomic theory and an application*, Springer Science & Business Media.

——— (2008): *Price and quantity index numbers: models for measuring aggregate change and difference*, Cambridge University Press.

Balk, B. M., R. Färe, and S. Grosskopf (2003): "The theory of economic price and quantity indicators," *Economic Theory*, 23, 149–164.

Ball, E., S. L. Wang, and R. Nehring (2016): "USDA ERS - Agricultural Productivity in the U.S." (Accessed on 2016-05-02).

Ball, V. E., R. Färe, S. Grosskopf, and D. Margaritis (2010): *The Economic Impact of Public Support to Agriculture*, New York, NY: Springer New York, chap. Productivity and Profitability of US Agriculture: Evidence from a Panel of States, 125–139.

Ball, V. E., C. Hallahan, and R. Nehring (2004): "Convergence of Productivity: An Analysis of the Catch-up Hypothesis within a Panel of States," *American Journal of Agricultural Economics*, 86, 1315–1321.

Banker, R. D. (1993): "Maximum likelihood, consistency and data envelopment analysis: a statistical foundation," *Management science*, 39, 1265–1273.

Banker, R. D. and H. Chang (2006): "The super-efficiency procedure for outlier identification, not for ranking efficient units," *European Journal of Operational Research*, 175, 1311–1320.

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

Barros, C. P., A. Ibiwoye, and S. Managi (2008): "Productivity change of Nigerian insurance companies: 1994–2005," *African Development Review*, 20, 505–528.

BEA (2016): "BEA," `http://http://www.bea.gov/iTable/index_regional.cfm`, (Accessed on 2016-12-01).

Bjurek, H. (1996): "The Malmquist total factor productivity index," *Scandinavian Journal of Economics*, 303–313.

Blackorby, C. and R. R. Russell (1999): "Aggregation of efficiency indices," *Journal of Productivity Analysis*, 12, 5–20.

Blancard, S., J.-P. Boussemart, J.-P. Chavas, and H. Leleu (2016): "Potential gains from specialization and diversification further to the reorganization of activities," *Omega*, 63, 60–68.

Bogetoft, P. (1996): "DEA on relaxed convexity assumptions," *Management Science*, 42, 457–465.

Bogetoft, P., K. Boye, H. Neergaard-Petersen, and K. Nielsen (2007): "Reallocating sugar beet contracts: can sugar production survive in Denmark?" *European Review of Agricultural Economics*, 34, 1–20.

Bogetoft, P. and J. L. Hougaard (2003): "Rational inefficiencies," *Journal of Productivity Analysis*, 20, 243–271.

Bogetoft, P. and L. Otto (2015): *Benchmarking with DEA and SFA*, r package version 0.26.

Briec, W. and K. Kerstens (2004): "A Luenberger-Hicks-Moorsteen productivity indicator: its relation to the Hicks-Moorsteen productivity index and the Luenberger productivity indicator," *Economic Theory*, 23, 925–939.

——— (2006): "Input, output and graph technical efficiency measures on non-convex FDH models with various scaling laws: An integrated approach based upon implicit enumeration algorithms," *Top*, 14, 135–166.

——— (2009): "Infeasibility and directional distance functions with application to the determinateness of the Luenberger productivity indicator," *Journal of Optimization Theory and Applications*, 141, 55–73.

——— (2011): "The Hicks–Moorsteen productivity index satisfies the determinateness axiom," *The Manchester School*, 79, 765–775.

Briec, W., K. Kerstens, and P. V. Eeckaut (2004): "Non-convex technologies and cost functions: definitions, duality and nonparametric tests of convexity," *Journal of Economics*, 81, 155–192.

Briec, W., K. Kerstens, and N. Peypoch (2012): "Exact relations between four definitions of productivity indices and indicators," *Bulletin of Economic Research*, 64, 265–274.

Briec, W. and Q. B. Liang (2011): "On some semilattice structures for production technologies," *European Journal of Operational Research*, 215, 740–749.

Briec, W. and N. Peypoch (2007): "Biased technical change and parallel neutrality," *Journal of Economics*, 92, 281–292.

Caballero, R. J. and M. L. Hammour (1998): "The macroeconomics of specificity," *Journal of Political Economy*, 106, 724–767.

Carvajal, A., R. Deb, J. Fenske, and J. K.-H. Quah (2013): "Revealed preference tests of the Cournot model," *Econometrica*, 81, 2351–2379.

——— (2014): "A nonparametric analysis of multi-product oligopolies," *Economic Theory*, 57, 253–277.

Caves, D. W., L. R. Christensen, and W. E. Diewert (1982): "The economic theory of index numbers and the measurement of input, output, and productivity," *Econometrica*, 1393–1414.

Cazals, C., J.-P. Florens, and L. Simar (2002): "Nonparametric frontier estimation: a robust approach," *Journal of Econometrics*, 106, 1 – 25.

Chambers, R. G. (2002): "Exact nonradial input, output, and productivity measurement," *Economic Theory*, 20, 751–765.

Chambers, R. G., Y. Chung, and R. Färe (1996a): "Benefit and distance functions," *Journal of Economic Theory*, 70, 407–419.

——— (1998): "Profit, directional distance functions, and Nerlovian efficiency," *Journal of Optimization Theory and Applications*, 98, 351–364.

Chambers, R. G., R. Färe, and S. Grosskopf (1996b): "Productivity growth in APEC countries," *Pacific Economic Review*, 1, 181–190.

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

Chavas, J.-P. (2008): "On the economics of agricultural production*," *Australian Journal of Agricultural and Resource Economics*, 52, 365–380.

Chavas, J.-P. and M. Aliber (1993): "An analysis of economic efficiency in agriculture: a nonparametric approach," *Journal of Agricultural and Resource Economics*, 1–16.

Chavas, J.-P. and K. Kim (2010): "Economies of diversification: A generalization and decomposition of economies of scope," *International Journal of Production Economics*, 126, 229–235.

Chen, C.-M. and J. van Dalen (2010): "Measuring dynamic efficiency: Theories and an integrated methodology," *European Journal of Operational Research*, 203, 749 – 760.

Chen, P.-C. (2012): “Measurement of technical efficiency in farrow-to-finish swine production using multi-activity network data envelopment analysis: evidence from Taiwan,” *Journal of Productivity Analysis*, 38, 319–331.

Cherchye, L., B. De Rock, B. Dierynck, P. J. Kerstens, and F. Roodhooft (2017a): “A benefit of the doubt approach to customer value with an application in telecom,” Working paper.

Cherchye, L., B. De Rock, B. Dierynck, F. Roodhooft, and J. Sabbe (2013): “Opening the “Black Box” of Efficiency Measurement: Input Allocation in Multioutput Settings,” *Operations research*, 61, 1148–1165.

Cherchye, L., B. De Rock, and V. Hennebel (2017b): “Coordination efficiency in multi-output settings: a DEA approach,” *Annals of Operations Research*, 250, 205–233.

Cherchye, L., B. De Rock, and B. Walheer (2015): “Multi-output efficiency with good and bad outputs,” *European Journal of Operational Research*, 240, 872–881.

——— (2016): “Multi-output profit efficiency and directional distance functions,” *Omega*, 61, 100–109.

Cherchye, L., T. Demuynck, B. De Rock, and K. De Witte (2014): “Non-parametric Analysis of Multi-output Production with Joint Inputs,” *The Economic Journal*, 124, 735–775.

Cherchye, L., T. Kuosmanen, and T. Post (2001): “FDH directional distance functions with an application to European commercial banks,” *Journal of Productivity Analysis*, 15, 201–215.

——— (2002): “Non-parametric production analysis in non-competitive environments,” *International Journal of Production Economics*, 80, 279–294.

Cherchye, L., W. Moesen, N. Rogge, and T. Van Puyenbroeck (2007): “An introduction to benefit of the doubt composite indicators,” *Social Indicators Research*, 82, 111–145.

Cook, W. D., M. Hababou, and H. J. Tuenter (2000): “Multicomponent efficiency measurement and shared inputs in data envelopment analysis: an application to sales and service performance in bank branches,” *Journal of Productivity Analysis*, 14, 209–224.

Cox, T. L. and M. K. Wohlgenant (1986): “Prices and quality effects in cross-sectional demand analysis,” *American Journal of Agricultural Economics*, 68, 908–919.

Crawford, I. (2010): “Habits Revealed,” *Review of Economic Studies*, 77, 1382–1402.

Crawford, I. and M. Polisson (2014): "Testing for intertemporal nonseparability," *Journal of Mathematical Economics*, 52, 46–49.

Dai, X. and T. Kuosmanen (2014): "Best-practice benchmarking using clustering methods: Application to energy regulation," *Omega*, 42, 179–188.

Dakpo, K. H., P. Jeanneaux, and L. Latruffe (2016): "Modelling pollution-generating technologies in performance benchmarking: Recent developments, limits and future prospects in the nonparametric framework," *European Journal of Operational Research*, 250, 347–359.

Daraio, C. and L. Simar (2005): "Introducing environmental variables in nonparametric frontier models: a probabilistic approach," *Journal of Productivity Analysis*, 24, 93–121.

——— (2007): *Advanced Robust and Nonparametric Methods in Efficiency Analysis: Methodology and Applications*, Springer.

——— (2014): "Directional distances and their robust versions: Computational and testing issues," *European Journal of Operational Research*, 237, 358–369.

——— (2016): "Efficiency and benchmarking with directional distances: a data-driven approach," *Journal of the Operational Research Society*, 67, 928–944.

de Witte, K. and M. Kortelainen (2013): "What explains the performance of students in a heterogeneous environment? Conditional efficiency estimation with continuous and discrete environmental variables," *Applied Economics*, 45, 2401–2412, cited By 12.

Diewert, W. E. (1976): "Exact and superlative index numbers," *Journal of Econometrics*, 4, 115–145.

——— (2005): "Index number theory using differences rather than ratios," *American Journal of Economics and Sociology*, 64, 311–360.

Diewert, W. E. and K. J. Fox (2010): "Malmquist and Törnqvist productivity indexes: returns to scale and technical progress with imperfect competition," *Journal of Economics*, 101, 73–95.

——— (2014): "Reference technology sets, Free Disposal Hulls and productivity decompositions," *Economics Letters*, 122, 238–242.

——— (2017): "Decomposing productivity indexes into explanatory factors," *European Journal of Operational Research*, 256, 275 – 291.

Diewert, W. E. and C. Parkan (1983): *Linear programming tests of regularity conditions for production functions*, Springer.

Epure, M. (2016): "Benchmarking for routines and organizational knowledge: a managerial accounting approach with performance feedback," *Journal of Productivity Analysis*, 46, 87–107.

Epure, M., K. Kerstens, and D. Prior (2011): "Bank productivity and performance groups: A decomposition approach based upon the Luenberger productivity indicator," *European Journal of Operational Research*, 211, 630 – 641.

Eurostat (2015): "Eurostat," `http://ec.europa.eu/eurostat`, (Accessed on 2015-11-30).

Fallah-Fini, S., K. Triantis, and A. L. Johnson (2013): "Reviewing the literature on non-parametric dynamic efficiency measurement: state-of-the-art," *Journal of Productivity Analysis*, 41, 51–67.

Färe, R. (1986): "A dynamic non-parametric measure of output efficiency," *Operations Research Letters*, 5, 83–85.

Färe, R., R. Grabowski, S. Grosskopf, and S. Kraft (1997): "Efficiency of a fixed but allocatable input: A non-parametric approach," *Economics Letters*, 56, 187–193.

Färe, R. and S. Grosskopf (1996): *Intertemporal production functions with dynamic DEA*, Kluwer Academic Publishers.

Färe, R. and S. Grosskopf (2000): "Network DEA," *Socio-Economic Planning Sciences*, 34, 35–49.

Färe, R. and S. Grosskopf (2012): *Cost and revenue constrained production*, Springer Science & Business Media.

Färe, R., S. Grosskopf, and C. K. Lovell (1985): *The Measurement of Efficiency of Production*, vol. 6, Springer Science & Business Media.

——— (1994a): *Production frontiers*, Cambridge University Press.

Färe, R., S. Grosskopf, and D. Margaritis (2010): "Time substitution with application to data envelopment analysis," *European Journal of Operational Research*, 206, 686–690.

Färe, R., S. Grosskopf, and D. Njinkeu (1988): "Note—On Piecewise Reference Technologies," *Management Science*, 34, 1507–1511.

Färe, R., S. Grosskopf, M. Norris, and Z. Zhang (1994b): "Productivity growth, technical progress, and efficiency change in industrialized countries," *American Economic Review*, 66–83.

Färe, R., S. Grosskopf, and P. Roos (1996): "On two definitions of productivity," *Economics Letters*, 53, 269–274.

——— (1998): "Malmquist productivity indexes: a survey of theory and practice," in *Index numbers: Essays in honour of Sten Malmquist*, Springer, 127–190.

Färe, R., S. Grosskopf, and G. Whittaker (2013): "Directional output distance functions: endogenous directions based on exogenous normalization constraints," *Journal of Productivity analysis*, 40, 267–269.

Färe, R. and D. Primont (1995): *Multi-output production and duality: theory and applications*, Springer Science & Business Media.

Färe, R. and G. Whittaker (1995): "An intermediate input model of dairy production using complex survey data," *Journal of Agricultural Economics*, 46, 201–213.

Färe, R. and V. Zelenyuk (2003): "On aggregate Farrell efficiencies," *European Journal of Operational Research*, 146, 615–620.

Farrell, M. J. (1957): "The measurement of productive efficiency," *Journal of the Royal Statistical Society. Series A (General)*, 253–290.

Farsi, M., M. Filippini, and W. Greene (2005): "Efficiency measurement in network industries: application to the Swiss railway companies," *Journal of Regulatory Economics*, 28, 69–90.

Fernandez-Cornejo, J., C. Gempesaw, J. Elterich, and S. Stefanou (1992): "Dynamic measures of scope and scale economies: an application to German agriculture," *American Journal of Agricultural Economics*, 74, 329–342.

Forsund, F. R. and L. Hjalmarsson (1974): "On the Measurement of Productive Efficiency," *The Scandinavian Journal of Economics*, 76, 141–154.

Färe, R., S. Grosskopf, and G. Whittaker (2007): "Network DEA," in *Modeling Data Irregularities and Structural Complexities in Data Envelopment Analysis*, ed. by J. Zhu and W. Cook, Springer US, 209–240.

Fried, H. O., S. S. Schmidt, and C. K. Lovell (1993): *The measurement of productive efficiency: techniques and applications*, Oxford university press.

Grifell-Tatjé, E. and C. K. Lovell (1995): "A note on the Malmquist productivity index," *Economics Letters*, 47, 169–175.

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

Guthrie, G. (2006): "Regulating infrastructure: The impact on risk and investment," *Journal of Economic Literature*, 44, 925–972.

Hackman, S. T. and R. C. Leachman (1989): "A general framework for modeling production," *Management Science*, 35, 478–495.

HADLEY, D. (2006): "Patterns in technical efficiency and technical change at the farm-level in England and Wales, 1982–2002," *Journal of Agricultural Economics*, 57, 81–100.

HADLEY, D., E. FLEMING, AND R. VILLANO (2013): "Is Input Mix Inefficiency Neglected in Agriculture? A Case Study of Pig-based Farming Systems in England and Wales," *Journal of Agricultural Economics*, 64, 505–515.

HAMPF, B. AND J. J. KRÜGER (2015): "Optimal directions for directional distance functions: an exploration of potential reductions of greenhouse gases," *American Journal of Agricultural Economics*, 97, 920–938.

HANEY, A. B. AND M. G. POLLITT (2013): "International benchmarking of electricity transmission by regulators: A contrast between theory and practice?" *Energy Policy*, 62, 267 – 281.

HANOCH, G. AND M. ROTHSCHILD (1972): "Testing the assumptions of production theory: a nonparametric approach," *Journal of Political Economy*, 256–275.

JAENICKE, E. C. (2000): "Testing for intermediate outputs in dynamic DEA models: Accounting for soil capital in rotational crop production and productivity measures," *Journal of Productivity Analysis*, 14, 247–266.

JEONG, S.-O., B. PARK, AND L. SIMAR (2010): "Nonparametric conditional efficiency measures: asymptotic properties," *Annals of Operations Research*, 173, 105–122.

JOHANSEN, L. (1959): "Substitution versus fixed production coefficients in the theory of economic growth: a synthesis," *Econometrica*, 27, 157–176.

JORGENSON, D. W. AND Z. GRILICHES (1967): "The explanation of productivity change," *The Review of Economic Studies*, 34, 249–283.

JUST, D. R. AND J. D. KROPP (2013): "Production Incentives from Static Decoupling: Land Use Exclusion Restrictions," *American Journal of Agricultural Economics*, 95, 1049–1067.

KAO, C. (2013): "Dynamic data envelopment analysis: A relational analysis," *European Journal of Operational Research*, 227, 325–330.

KAO, C. AND S.-N. HWANG (2008): "Efficiency decomposition in two-stage data envelopment analysis: An application to non-life insurance companies in Taiwan," *European Journal of Operational Research*, 185, 418–429.

KARAGIANNIS, G., P. MIDMORE, AND V. TZOUVELEKAS (2002): "Separating technical change from time-varying technical inefficiency in the absence of distributional assumptions," *Journal of Productivity Analysis*, 18, 23–38.

——— (2004): "Parametric decomposition of output growth using a stochastic input distance function," *American Journal of Agricultural Economics*, 86, 1044–1057.

Kasparis, I. and P. C. Phillips (2012): "Dynamic misspecification in nonparametric cointegrating regression," *Journal of Econometrics*, 168, 270–284.

Kerstens, K. and I. Van de Woestyne (2011): "Negative data in DEA: a simple proportional distance function approach," *Journal of the Operational Research Society*, 62, 1413–1419.

Koopmans, T. C. (1951): "Efficient allocation of resources," *Econometrica*, 455–465.

Kuosmanen, T. (2003): "Duality theory of non-convex technologies," *Journal of Productivity Analysis*, 20, 273–304.

Kuosmanen, T. and T. Post (2002): "Nonparametric efficiency analysis under price uncertainty: a first-order stochastic dominance approach," *Journal of Productivity Analysis*, 17, 183–200.

Lee, C.-Y. and A. L. Johnson (2015): "Measuring efficiency in imperfectly competitive markets: An example of rational inefficiency," *Journal of Optimization Theory and Applications*, 164, 702–722.

Leleu, H. (2006): "A linear programming framework for free disposal hull technologies and cost functions: Primal and dual models," *European Journal of Operational Research*, 168, 340–344.

Li, Q., E. Maasoumi, and J. S. Racine (2009): "A nonparametric test for equality of distributions with mixed categorical and continuous data," *Journal of Econometrics*, 148, 186–200.

Liu, J. S. and W.-M. Lu (2010): "DEA and ranking with the network-based approach: a case of R&D performance," *Omega*, 38, 453–464.

Liu, J. S., W.-M. Lu, C. Yang, and M. Chuang (2009): "A network-based approach for increasing discrimination in data envelopment analysis," *Journal of the Operational Research Society*, 60, 1502–1510.

Luenberger, D. G. (1992): "New optimality principles for economic efficiency and equilibrium," *Journal of Optimization Theory and Applications*, 75, 221–264.

Luh, Y.-H. and S. E. Stefanou (1991): "Productivity growth in US agriculture under dynamic adjustment," *American Journal of Agricultural Economics*, 73, 1116–1125.

Malmquist, S. (1953): "Index numbers and indifference surfaces," *Trabajos de estadística*, 4, 209–242.

Managi, S. (2010): "Productivity measures and effects from subsidies and trade: an empirical analysis for Japan's forestry," *Applied Economics*, 42, 3871–3883.

MANIADAKIS, N. AND E. THANASSOULIS (2004): "A cost Malmquist productivity index," *European Journal of Operational Research*, 154, 396–409.

MAS-COLELL, A., M. D. WHINSTON, J. R. GREEN, ET AL. (1995): *Microeconomic theory*, vol. 1, Oxford university press New York.

MCCLOSKEY, D. N. AND S. T. ZILIAK (1996): "The standard error of regressions," *Journal of Economic Literature*, 34, 97–114.

MIZOBUCHI, H. (2016): "A Superlative Index Number Formula for the Hicks-Moorsteen Productivity Index," Tech. rep., School of Economics, University of Queensland, Australia.

MOTAMED, M., L. MCPHAIL, AND R. WILLIAMS (2016): "Corn Area Response to Local Ethanol Markets in the United States: A Grid Cell Level Analysis," *American Journal of Agricultural Economics*, 98, 726–743.

MURTY, S., R. R. RUSSELL, AND S. B. LEVKOFF (2012): "On modeling pollution-generating technologies," *Journal of Environmental Economics and Management*, 64, 117–135.

NAKANO, M. AND S. MANAGI (2008): "Regulatory reforms and productivity: an empirical analysis of the Japanese electricity industry," *Energy Policy*, 36, 201–209.

NEMOTO, J. AND M. GOTO (1999): "Dynamic data envelopment analysis: modeling intertemporal behavior of a firm in the presence of productive inefficiencies," *Economics Letters*, 64, 51 – 56.

——— (2003): "Measurement of Dynamic Efficiency in Production: An Application of Data Envelopment Analysis to Japanese Electric Utilities," *Journal of Productivity Analysis*, 19, 191–210.

——— (2005): "Productivity, efficiency, scale economies and technical change: A new decomposition analysis of TFP applied to the Japanese prefectures," *Journal of the Japanese and International Economies*, 19, 617–634.

NICK, S. AND H. WETZEL (2015): "The hidden cost of investment: the impact of adjustment costs on firm performance measurement and regulation," *Journal of Regulatory Economics*, 1–23.

NIELSEN, R., J. L. ANDERSEN, AND P. BOGETOFT (2014): "Dynamic reallocation of marketable nitrogen emission permits in Danish freshwater aquaculture," *Marine Resource Economics*, 29, 219–239.

O'DONNELL, C. J. (2012a): "An aggregate quantity framework for measuring and decomposing productivity change," *Journal of Productivity Analysis*, 38, 255–272.

——— (2012b): "Nonparametric estimates of the components of productivity and profitability change in US agriculture," *American Journal of Agricultural Economics*, 873–890.

OECD (2009): "Measuring Capital - OECD Manual 2009," Tech. rep., Organisation for Economic Co-operation and Development.

Oude Lansink, A. and S. Stefanou (2001): "Dynamic Area Allocation and Economies of Scale and Scope," *Journal of Agricultural Economics*, 52, 38–52.

Oude Lansink, A., S. Stefanou, and T. Serra (2015): "Primal and dual dynamic Luenberger productivity indicators," *European Journal of Operational Research*, 241, 555–563.

Ouellette, P. and L. Yan (2008): "Investment and dynamic DEA," *Journal of Productivity Analysis*, 29, 235–247.

Panzar, J. C. and R. D. Willig (1981): "Economies of Scope," *American Economic Review*, 71, 268–272.

Pestieau, P. and H. Tulkens (1993): "Assessing and explaining the performance of public enterprises," *FinanzArchiv/Public Finance Analysis*, 50, 293–323.

Peters, W. (1985): "Can inefficient public production promote welfare?" *Journal of Economics*, 45, 395–407.

Petersen, N. C. (1990): "Data envelopment analysis on a relaxed set of assumptions," *Management Science*, 36, 305–314.

Peyrache, A. (2013): "Industry structural inefficiency and potential gains from mergers and break-ups: A comprehensive approach," *European Journal of Operational Research*, 230, 422–430.

——— (2014): "Hicks-Moorsteen versus Malmquist: a connection by means of a radial productivity index," *Journal of Productivity Analysis*, 41, 435–442.

Peyrache, A. and C. Daraio (2012): "Empirical tools to assess the sensitivity of directional distance functions to direction selection," *Applied Economics*, 44, 933–943.

Podinovski, V. (2005): "Selective convexity in DEA models," *European Journal of Operational Research*, 161, 552–563.

Pollitt, M. (2005): "The role of efficiency estimates in regulatory price reviews: Ofgem's approach to benchmarking electricity networks," *Utilities Policy*, 13, 279 – 288.

Portela, M. C. A. S. and E. Thanassoulis (2014): "Economic efficiency when prices are not fixed: disentangling quantity and price efficiency," *Omega*, 47, 36–44.

Ray, S. C. (2008): "The directional distance function and measurement of super-efficiency: an application to airlines data," *Journal of the Operational Research Society*, 59, 788–797.

Ruggiero, J. (1996): "On the measurement of technical efficiency in the public sector," *European Journal of Operational Research*, 90, 553–565.

Ryschawy, J., N. Choisis, J. P. Choisis, A. Joannon, and A. Gibon (2012): "Mixed crop-livestock systems: an economic and environmental-friendly way of farming?" *Animal*, 6, 1722–1730.

Shephard, R. (1970): *Theory of cost and production functions*, Princeton University Press.

Shephard, R. W. and R. Färe (1980): *Dynamic theory of production correspondences*, vol. 50, Verlag Anton Hain.

Shuttleworth, G. (2005): "Benchmarking of electricity networks: Practical problems with its use for regulation," *Utilities Policy*, 13, 310–317.

Silva, E. and S. E. Stefanou (2003): "Nonparametric dynamic production analysis and the theory of cost," *Journal of Productivity Analysis*, 19, 5–32.

——— (2007): "Dynamic efficiency measurement: theory and application," *American Journal of Agricultural Economics*, 89, 398–419.

Simar, L. and A. Vanhems (2012): "Probabilistic characterization of directional distances and their robust versions," *Journal of Econometrics*, 166, 342–354.

Skevas, T., A. Oude Lansink, and S. Stefanou (2012): "Measuring technical efficiency in the presence of pesticide spillovers and production uncertainty: The case of Dutch arable farms," *European Journal of Operational Research*, 223, 550 – 559.

Thrall, R. M. (1996): "Chapter 5 Duality, classification and slacks in DEA," *Annals of Operations Research*, 66, 109–138.

Tone, K. (2001): "A slacks-based measure of efficiency in data envelopment analysis," *European Journal of Operational Research*, 130, 498–509.

Törnqvist, L. and E. Törnqvist (1937): "Vilket är Förhällandet Mellan Finska Markens och Svenska Kronans Köpkraft?" *Ekonomiska Samfundets Tidskrift*, 39, 1–39.

Törnqvist, L. (1936): "The Bank of Finlands Consumption Price Index," *Bank of Finland Monthly Bulletin*, 10, 1–8.

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

USDA (2016): "USDA," `http://www.ers.usda.gov/data-products/agricultural-productivity-in-the-us.aspx`, (Accessed on 2016-11-24).

Varian, H. (1982): "The nonparametric approach to demand analysis," *Econometrica*, 50, 945–973.

——— (1984): "The nonparametric approach to production analysis," *Econometrica*, 52, 579–597.

Varian, H. R. (1990): "Goodness-of-fit in optimizing models," *Journal of Econometrics*, 46, 125–140.

Wang, K., Y. Xian, C.-Y. Lee, Y.-M. Wei, and Z. Huang (2017): "On selecting directions for directional distance functions in a non-parametric framework: a review," *Annals of Operations Research*, 1–34.

Wibe, S. (2008): "Efficiency: A dynamic approach," *International Journal of Production Economics*, 115, 86–91.

Wilson, P., D. Hadley, and C. Asby (2001): "The influence of management characteristics on the technical efficiency of wheat farmers in eastern England," *Agricultural Economics*, 24, 329 – 338.

Zofío, J. L. and C. A. K. Lovell (2001): "Graph efficiency and productivity measures: an application to US agriculture," *Applied Economics*, 33, 1433–1442.

Zofío, J. L., J. T. Pastor, and J. Aparicio (2013): "The directional profit efficiency measure: on why profit inefficiency is either technical or allocative," *Journal of Productivity Analysis*, 40, 257–266.

List of Figures

List of Tables

List of doctoral dissertations

A list of doctoral dissertations from the Faculty of Economics and Business can be found at the following website:
`http://www.kuleuven.ac.be/doctoraatsverdediging/archief.htm`.