NOAA Technical Memorandum NMFS

JANUARY 2013

PROCEEDINGS OF THE NATIONAL MARINE FISHERIES SERVICE PRODUCTIVITY WORKSHOP (SANTA CRUZ, CA JUNE 11-12, 2012)

Edited By:
Aaron Mamula
John Walden

NOAA-TM-NMFS-SWFSC-503

U.S. DEPARTMENT OF COMMERCE
National Oceanic and Atmospheric Administration
National Marine Fisheries Service
Southwest Fisheries Science Center

NOAA Technical Memorandum NMFS

The National Oceanic and Atmospheric Administration (NOAA), organized in 1970, has evolved into an agency which establishes national policies and manages and conserves our oceanic, coastal, and atmospheric resources. An organizational element within NOAA, the Office of Fisheries is responsible for fisheries policy and the direction of the National Marine Fisheries Service (NMFS).

In addition to its formal publications, the NMFS uses the NOAA Technical Memorandum series to issue informal scientific and technical publications when complete formal review and editorial processing are not appropriate or feasible. Documents within this series, however, reflect sound professional work and may be referenced in the formal scientific and technical literature.

NOAA Technical Memorandum NMFS
This TM series is used for documentation and timely communication of preliminary results, interim reports, or special purpose information. The TMs have not received complete formal review, editorial control, or detailed editing.

JANUARY 2013

PROCEEDINGS OF THE NATIONAL MARINE FISHERIES SERVICE PRODUCTIVITY WORKSHOP (SANTA CRUZ, CA JUNE 11-12, 2012)

Edited by Aaron Mamula[1] and John Walden[2]

[1] NOAA National Marine Fisheries Service
SWFSC Fisheries Ecology Division
110 Shaffer Road
Santa Cruz, CA 95060

[2] NOAA National Marine Fisheries Service
Northeast Fisheries Science Center
Woods Hole, MA

NOAA-TM-NMFS-SWFSC-503

U.S. DEPARTMENT OF COMMERCE
Rebecca M. Blank, Acting Secretary

National Oceanic and Atmospheric Administration
Jane Lubchenco, Undersecretary for Oceans and Atmosphere

National Marine Fisheries Service
Samuel D. Rauch, Acting Assistant Administrator for Fisheries

Executive Summary

A workshop was held at the Southwest Fisheries Science Center in Santa Cruz, California, June 11-12, 2012, bringing together leading experts in the field of productivity, and National Marine Fisheries Service economists. The meeting was held to generate discussion about productivity concepts, and to begin thinking about how to measure productivity in our nation's fisheries. Productivity is a key driver of profitability, and has been identified as an important indicator for fishery performance reports.

Harvesting a regulated natural resource stock presents challenges for assessing productivity. Changing stock abundance, environmental variability fleet heterogeneity (such as what might arise when firms employ different production technologies), and regulatory change all affect productivity. In addition, fisheries economists are frequently tasked with measuring productivity with little data.

This document contains manuscripts submitted by the productivity experts and reflects the content of their presentations. It is hoped that this workshop proceedings document will serve as a springboard for further productivity research centered on our nation's fisheries. The variety of topics, and the depth in which they are presented, shows that we have merely scratched the surface in NOAA regarding productivity change. We still have many opportunities for further research in this exciting field.

Introduction.

A workshop was held at the Southwest Fisheries Science Center in Santa Cruz, California, June 11-12, 2012, which brought together leading experts in the field of productivity, and National Marine Fisheries Service Science Center economists. The purpose of the meeting was to generate discussion around the concepts of productivity measurement, and to begin thinking about how to best measure productivity in our nation's fisheries. Productivity is a key metric for understanding profitability change, and has been identified as a Tier II indicator for national reporting purposes.

Productivity measurement of fishing fleets has been conducted periodically throughout the years. One of the earliest works was by Comitini and Huang (1967) who used a Cobb-Douglas technology to characterize the production of 32 halibut fishing vessels in the North Pacific over a seven year period. Norton, Miller, and Kenney (1985) used aggregated data from vessels fishing in five U.S. fisheries to estimate an Economic Health Index, which contained a productivity component that could be examined separately. Squires (1987, 1992) published a study measuring productivity in the Pacific Coast Trawl Fishery using an index number approach. Weninger (2001) examined changes in productivity for surfclam vessels using a directional distance function model. Jin et al. (2002) measured total factor productivity in the New England Groundfish Fishery during the period 1964-2003. Felthoven and Paul (2004) reviewed past productivity studies, and suggested a methodology for productivity measurement to answer questions concerning economic performance. Fox et al. (2006) examined changes in capacity, quota trading and productivity after a license buyback in Australian fisheries. Hannesson (2007) used a growth accounting framework to measure productivity change in Norwegian fisheries. Squires, Reid, and Jeon (2008) examined productivity growth in the Korean tuna purse seine fishery operating in the Pacific Ocean. Felthoven, Paul, and Torres (2009) measured productivity in the Alaskan pollock fishery from 1994-2003 while incorporating environmental conditions, bycatch and stock effects. Eggert and Tveterås (2011) examined productivity change in Icelandic, Norwegian and Swedish fisheries between 1973 and 2003.

Productivity measurement in fisheries presents challenges that are different from traditional industries. Fundamentally, vessels are harvesting a natural resource stock where the Government sets the total harvest level allowed in any given time period. Whether the harvest is controlled directly through a total allowable catch, or through indirect measures such as limits on fishing time, total output is constrained. In many instances, regulations are intended to make vessels less productive. For example, more productive areas (biologically) may be closed, forcing vessels to fish less productive fishing grounds. This means vessels will need to use more inputs to catch the same amount of fish as they would in the more productive areas. Stock conditions and environmental factors can also influence productivity. External drivers, such as changing ocean temperatures, can impact overall stock conditions. Failing to recognize external conditions when setting catch limits may lead to over harvest in one year, and subsequent harvest reductions in the following years. Finally, different technologies (typically gear types) can be used to harvest the same resource, and this needs to be factored into productivity assessments.

The Santa Cruz workshop was designed to introduce NMFS economists to productivity concepts, and at the same time gather ideas from the productivity experts that were invited to attend the workshop. Some questions that were posed to the group included: Is there an alternative to the Malmquist index to measure productivity for fishing fleets? Do we measure productivity at an aggregate or individual vessel level? How do we account for stock conditions in our estimates? Do we consider undesirable outputs in our measures? Do binding catch limits matter, and if so how? How do we estimate productivity if cost data are lacking?

This document contains selected papers from participants in the workshop. There are a wide variety of topics covered in the manuscripts. By not focusing solely on fisheries, it is hoped that the ideas presented here will help inform and develop the next generation of productivity models in fisheries. A brief description of each paper follows. The presenter for each paper is in bold print.

E. Griffell-Tatjé and **Knox Lovell** presented a framework that links financial performance, price, productivity and capacity constraints using the DuPont triangle. Productivity is measured using both a price and technology based index. Although the DuPont triangle has not been applied in fisheries, the methods shown in their paper could be used in many commercial fisheries given availability of appropriate data. Thus, the paper presents a way forward in development of rigorous financial performance indicators.

Kristiaan Kerstens and Ignace Van de Woestyne explored the difference in Productivity measures using balanced and unbalanced panel data. This is important because in most productivity studies in fisheries, entry and exit of vessels in each year leads to unbalanced panel data. Determining how to handle unbalanced panel data is a challenging question for fishery researchers.

Eldon Ball and **Sun Ling Wang** presented the methods used by the U.S. Department of Agriculture Economic Research Service (ERS) to estimate productivity growth for the farm sector, based on the Törnqvist index. The ERS has been publishing estimates of total factor productivity since the 1960s. Their work can help inform the National Marine Fisheries Service on ways to carry out productivity estimates for our catch share fisheries.

Subhash Ray presented a method for decomposing the cost competiveness of a firm, relative to a rival, and showed how efficiency changes, relative price changes and technical change affect cost competitiveness. With the move to catch share fisheries management, where total output is fixed, the role that these factors play in overall profitability of vessels will become more important.

Trevor Collier, Aaron Mamula and **John Ruggiero** presented a model to incorporate exogenous factors into a DEA efficiency model, and then apply the model to vessels in the California groundfish trawl fishery. In a fisheries context, this is important because environmental conditions can change throughout, and between years. Their application examined crowding externalities, which can result from various regulations such as area closures and effort limitations.

Kristiaan Kerstens and **Niels Vestergaard** examined the impact of the convexity assumption on fishing capacity estimates. This is a topic which has not generated a lot of attention, but is very important in fisheries. Assuming convexity when it does not exist may lead to unrealistic capacity estimates. Decommissioning schemes based on unrealistic capacity estimates may result in retiring too many vessels, at a higher cost than is necessary.

Chris O'Donnell used Bayesian methods to compute and decompose a Färe-Primont index of total factor productivity change in the Australian Northern Prawn Fishery. A Bayesian approach offers some advantages for the study of fishing vessels, because inferences can be made about productivity with little data. This is often a problem facing researchers using data sets with aggregated data from fleets of vessels. A further advantage to this approach is that it solves the endogeneity problem in econometric estimation of multiple-input, multiple-output distance functions.

Ole Olesen, John Ruggiero, and Aaron Mamula estimated a nonparametric homothetic S-Shaped production relation for U.S. West Coast groundfish vessels. Their paper presents an approach which hasn't been attempted before in the fisheries literature, but which holds promise for future work. For example, scale characteristics of various fishing fleets can be estimated using this approach.

Carl Pasurka presented a paper on undesirable outputs in productivity estimates, and how the literature has evolved during the past 30 years. This is a topic which is very important in the context of fisheries, as discards from fishing vessels have been identified as a major contributor to fishing mortality worldwide. How to adjust productivity estimates to account for discards is a topic which needs further discussion.

We hope that the reader of this document will take full account of the depth and breadth of the manuscripts submitted by each author.

NMFS Productivity Workshop, June 11-12, 2012, Southwest Fisheries Science Center, Santa Cruz, CA.

Monday, June 11th.

8:30. Welcome and Introductions. Conference Logistics	Aaron Mamula, Southwest Fisheries Science Center, Santa Cruz.
8:45-9:00. NMFS National Performance Measure Initiative and Productivity	Eric Thunberg, NMFS Office of Science and Technology
9:00-9:20. Workshop Objectives. Productivity Issues in Fisheries	John Walden, Northeast Fisheries Science Center
9:20-10:00. Productivity, Capacity Constraints and their Financial Consequences.	C.A. Knox Lovell, U. of Queensland
10:00 – 10:30 Break	
10:30-11:10 Primal Productivity Indices: Exploring the Impact of Unbalanced Panel Data	Kris Kerstens, IESEG School of Management.
11:15-11:55 The U.S. Agricultural Productivity Slowdown: When and Why?	Sun Ling Wang, USDA Economic Research Service
12:00-1:00 Lunch	
1:00-1:15 Follow-up questions from Morning Session	
1:20-1:40 Overview of Alaska Region Fisheries	Ron Felthoven, Alaska Fisheries Science Center
1:45 – 2:25 Productivity Change over time and the Dynamics of Cost Competition	Subhash Ray, U. of Connecticut
2:30 -2:50 Hawaii Fisheries	Minling Pan, Pacific Islands Fisheries Science Center
3:00-3:30 Break	
3:30-4:10 Nonparametric Estimation of Efficiency in Commercial Fisheries in the Presence of Nondiscretionary Factors of Production	John Ruggiero, U. of Dayton
4:15-4:55 Primal and dual approaches to fishing capacity: The impact of the convexity assumption.	Niels Vestergaard, U. of Southern Denmark
5:00-5:15 Wrap up day One	

Tuesday June 12th.

9:00 – 9:15. Day two objectives. Outstanding questions from Day One	
9:15-10:00 Econometric Estimates of Productivity and Efficiency Change in the Australian Northern Prawn Fishery	Chris O'Donnell, U. of Queensland
10:00-10:30 Break	
10:30-10:50 Overview of Southeast Fisheries	Juan Agar, Southeast Fisheries Science Center
10:50-11:30 Estimating a nonparametric homothetic S-Shaped production relation for the U.S. West Coast Groundfish Production 2004-2007.	Ole Olesen, University of Southern Denmark
11:30-11:50 Overview of Northwest Fisheries	Erin Steiner, Northwest Fisheries Science Center
12:00-1:00 Lunch	
1:00-1:40 Adventures in Modeling the Joint Production of Good and Bad Outputs: A 30 Year retrospective	Carl Pasurka, Environmental Protection Agency
1:45 – 2:30 Technical Change and the Commons	Dale Squires, Southwest Fisheries Science Center, La Jolla
2:30-3:00 Measuring Productivity Change in the Northeast Multispecies Fishery using a Fisher Index.	John Walden, Northeast Fisheries Science Center
3:00 – 3:30 Break	
3:30 – 5:00 Wrap-up	

Recommendations

During the afternoon of June 12th, a discussion was held among NMFS economists and external participants about recommendations for constructing productivity metrics for national reporting purposes. The Malmquist index has initially been identified as the preferred metric for measuring productivity, and there needed to be further discussion about whether this was still appropriate given the workshop presentations.

Workshop participants seemed to agree that an aggregate metric, using a “Fisher form” index, with a fixed period base is the best way to proceed. This requires fixing the base to a mutually agreed upon time period. It was further recommended that a Lowe Index would be the best index to construct because it is a transitive index. The construction of the Lowe Index requires data on output quantities and prices, along with input quantities and prices. If price information is not available, then the panelists recommended that we pursue a Färe-Primont Index, which only requires output and input quantities.

The most difficult part of constructing the index will be getting input price information (and perhaps quantity for some inputs). If this situation arises, a mixed approach can be pursued where the Malmquist, or Färe-Primont Index is used to measure input growth for the denominator, and a Lowe index for output growth in numerator are combined. In order to aggregate inputs to the industry level in the denominator, individual scores can be aggregated using revenue shares.

Finally, a discussion needs to occur between regional economists about the data needed to construct these indices. That conversation should occur before the indices are actually constructed.

Acknowledgements

We would like to thank Rita Curtis, and the Office of Science and Technology for funding this workshop, along with the participants in the workshop who gave their time and expertise to help us learn more about productivity.

References

Comitini, Salvatore, and David S. Huang. 1967. "A Study of Production and Factor Shares in the Halibut Fishing Industry." *Journal of Political Economy* no. 75 (4):366-372.

Eggert, Håkan, and Ragnar Tveterås. 2011. "Productivity development in Icelandic, Norwegian and Swedish fisheries." *Applied Economics* no. 45 (6):709-720. doi: 10.1080/00036846.2011.610751.

Felthoven, Ronald G., and Catherine J. Morrison Paul. 2004. "Directions for productivity measurement in fisheries." *Marine Policy* no. 28 (2):161-169. doi: 10.1016/S0308-597X(03)00084-8.

Felthoven, Ronald G., Catherine J. Morrison Paul, and Marcelo Torres. 2009. "Measuring productivity and its components for fisheries: The case of the Alaskan pollock fishery, 1994–2003." *Natural Resource Modeling* no. 22 (1):105-136. doi: 10.1111/j.1939-7445.2008.00031.x.

Fox, Kevin J., R. Quentin Grafton, Tom Kompas, and Tuong Nhu Che. 2006. "Capacity reduction, quota trading and productivity: the case of a fishery*." *Australian Journal of Agricultural and Resource Economics* no. 50 (2):189-206. doi: 10.1111/j.1467-8489.2006.00331.x.

Hannesson, Rögnvaldur. 2007. "Growth accounting in a fishery." *Journal of Environmental Economics and Management* no. 53 (3):364-376. doi: 10.1016/j.jeem.2006.11.001.

Jin, Di, Eric Thunberg, Hauke Kite-Powell, and Kevin Blake. 2002. "Total Factor Productivity Change in the New England Groundfish Fishery: 1964–1993." *Journal of Environmental Economics and Management* no. 44 (3):540-556. doi: 10.1006/jeem.2001.1213.

Norton, Virgil J., Morton M. Miller, and Elizabeth Kenney. 1985. Indexing the Economic Health of the U.S. Fishing Industry's Harvesting Sector. edited by National Oceanic and Atmospheric Administration U.S. Department of Commerce. Woods Hole, Massachusetts: Northeast Fisheries Science Center.

Squires, Dale. 1987. Index Numbers and Productivity Measurement in Multispecies Fisheries: An application to the Pacific Coast Trawl Fleet. . edited by National Marine Fisheries Service. La Jolla, CA.

Squires, Dale. 1992. "Productivity Measurement in Common Property Resource Industries: An Application to the Pacific Coast Trawl Fishery." *The RAND Journal of Economics* no. 23 (2):221-236.

Squires, Dale, Christopher Reid, and Yongil Jeon. 2008. "Productivity growth in natural resource industries and the environment: an application to the Korean tuna purse-seine fleet in the Pacific Ocean." *International Economic Journal* no. 22 (1):81-93. doi: 10.1080/10168730801886978.

Weninger, Quinn. 2001. "An analysis of the efficient production frontier in the fishery: implications for enhanced fisheries management." *Applied Economics* no. 33 (1):71-79. doi: 10.1080/00036840122937.

Productivity, Price Recovery, Capacity Constraints and their Financial Consequences*

E. Grifell-Tatjé
Universitat Autònoma de Barcelona

C. A. K. Lovell
University of Queensland

Abstract

Mining and fishing are both extractive industries, although one resource is renewable and the other is not. Miners and fishers pursue financial objectives, although their objectives may differ. In both industries financial performance is influenced by productivity and price recovery. Finally, in both industries capacity constraints influence financial performance, perhaps but not necessarily through their impact on productivity, and both industries encounter external as well as internal capacity constraints.

The objective of this study is to develop an analytical framework that links all four phenomena. We use return on assets to measure financial performance, and the basic analytical framework is the duPont triangle. We measure productivity in two ways, with a theoretical technology-based index and with an empirical price-based index. We measure price change with an empirical quantity-based index. We measure internal capacity utilisation by relating a pair of output quantity vectors, actual output and full capacity output, and we develop physical and economic measures of internal capacity utilisation. External capacity constraints restrict the ability to reach full capacity output. The analytical framework has productivity change, price change and change in capacity utilisation influencing change in return on assets, the latter in two ways, directly and indirectly through its impact on productivity change.

JEL classification: D24

Keywords: duPont triangle, capacity utilization, productivity, price recovery

Productivity, Price Recovery, Capacity Constraints and their Financial Consequences

1. Introduction

Mining and fishing are both extractive industries, although one resource is renewable and the other is not. Miners and fishers pursue financial objectives, although their objectives may differ. In both industries productivity and prices influence financial performance. In both industries capacity constraints influence financial performance, perhaps but not necessarily through their impact on productivity, and both industries encounter external as well as internal capacity constraints.

We offer two relevant illustrations. First, global mining giant Rio Tinto has generated impressive, and volatile, financial results throughout the recent mining boom. Figure 1 shows five-year moving averages of return on assets and its two components, profit margin and asset turnover, from 2007 through 2011.[1] One would like to learn something about the sources of the observed volatility in return on assets that digs deeper than just variation in the profit margin and asset turnover. Variation in productivity, prices and capacity constraints are likely sources.

Second, ABARES (2012) publishes a fisheries surveys report. The report provides detailed boat-level financial information, averaged over boats, within each of two fisheries, and similarly detailed economic information for the fisheries themselves. The boat-level financial information includes alternative measures of profit and return on assets. The fishery economic information includes profit and net economic returns, which adjusts profit in several ways, including the incorporation of the costs of managing the fishery. One would like to know something about the sources of variation in profit and return on assets across boats within a fishery, and the sources of variation in net economic returns through time and across fisheries. Again, variation in productivity, prices and capacity constraints are likely sources.

The objective of this study is to develop an analytical framework that links all four phenomena, financial performance, productivity, prices and capacity constraints. We use return on assets ROA to measure financial performance, and the basic analytical framework is the duPont triangle depicted in Figure 1. We measure productivity change Y/X in two ways, with a theoretical technology-based index and with an empirical price-based index. We measure price recovery change P/W with an empirical quantity-based index. We measure capacity utilization CU by relating a pair of output quantity vectors, actual output and full capacity output, and we develop physical and economic measures of CU. The analytical framework has Y/X, P/W and CU influencing ROA, the latter in two ways, directly and indirectly through its impact on productivity.

The study unfolds as follows. In Section 2 we introduce the duPont triangle as a framework for financial performance evaluation. In Section 3 we attempt to incorporate Y/X and CU into the duPont triangle. We succeed with CU and fail with Y/X because, while CU is an absolute indicator, Y/X is a relative indicator that compares one situation with another. Accordingly, in Section 4 we compare ROA in two time periods by converting the analytical framework to an inter-temporal one, and we seek to exploit the duPont triangle format to attribute ROA change from one period to the next to CU change and productivity change. Once again we succeed with CU change and fail with productivity change. In Sections 2-4 we ignore price change as an ROA driver; we would have failed, for the same reason we fail with productivity change. In Section 5 we develop a pair of analytical frameworks within which CU change, productivity change and price recovery change drive ROA change. In Sections 2-5 CU is an *internal* measure associated with short run fixity of some inputs used by the firm. In Section 6 we introduce *external* capacity constraints resulting from regulation and other sources outside the firm, and we show how these external capacity constraints can render some or all internal capacity constraints redundant. Section 7 concludes.

2. The duPont Triangle

ROA is a widely used measure of financial performance. Bliss (1923), in discussing ROA, claims that "[f]rom the operating point of view as distinguished from the stockholders' point of view, the real measure of the financial return earned by a business is the percentage of operating profits earned on the total capital used in the conduct of such operations…regardless from what sources such capital may have been secured." Two duPont executives, Kline & Hessler (1952), concur, writing that "It is our considered opinion, which has been critically re-examined many times over three decades, that a manufacturing enterprise with large capital committed to the manufacture and sale of goods can best measure and judge the effectiveness of effort in terms of 'return on investment'." Amey (1969) calls ROA "the key index of business 'success'," even though he acknowledges that maximizing ROA and maximizing profit in absolute terms do not generally coincide. Amey continues, "…maximization of profits in absolute terms will be taken as the firm's objective; this can then be *expressed* as a rate of return." (italics in the original) Thus ROA is an observable consequence of the pursuit of a different (indeed, almost any) objective.

ROA sits atop the duPont triangle, a management accounting system developed at duPont and General Motors (GM) early in the 20th century. Even then both duPont and GM were diversified corporations, producing a variety of products in several locations, and management had to decide how to allocate capital investment, as well as other resources and managerial compensation, across product lines and among plants. The allocation criterion duPont and GM used was the return on those investments, ROA. The developers also devised a product pricing formula designed to set product prices that would yield a desired ROA when production was at standard volume, defined at GM to be two shifts per day.

To assist in the resource allocation and product pricing strategies, ROA = π/A was decomposed into a pair of financial ratios that drive π/A. This in turn enabled management to develop strategies intended to enhance either ratio, and hence π/A. The decomposition states that π/A is the product of the profit margin π/R, and asset turnover R/A. π/R indicates how much of sales revenue a firm retains as profit rather than absorbs as expense. An increase in π/R is consistent with an improvement in cost efficiency, the adoption of cost-saving technology, a reduction in input prices or an increase in output prices. R/A indicates the revenue productivity of a firm's assets. An increase in R/A suggests that capital is being allocated to higher-valued uses, or output prices are increasing.[2]

For our purposes it is important to note that the duPont triangle does not contain measures of CU, Y/X or P/W, any one of which is a potential driver of π/R and/or R/A. We incorporate CU in Section 3, we incorporate CU and Y/X in Section 4, and we incorporate CU, Y/X and P/W in Section 5.

3. Capacity Utilization

Incorporating CU into a duPont triangle requires a definition of capacity, and there are several to choose from. A generic approach is to write the triangle as

$$\pi/A = \pi/R \times R/A$$

$$= \pi/R \times (p^T y/p^T y^c) \times (p^T y^c/A), \qquad (1)$$

with output price vector $p \in R^M_{++}$, output quantity vector $y \in R^M_+$ and capacity output quantity vector $y^c \in R^M_+$. Weighting y and y^c by p maintains the financial structure of the triangle and, more significantly for our purposes, allows M>1. Expression (1) decomposes ROA into the product of three drivers: the profit margin, the rate of capacity utilization, and potential asset turnover, the turnover that would occur at full capacity output. We now consider how to define y^c.

Figure 2 supports three definitions of capacity and its rate of utilization. We observe output vector y and input vector x, with $y \in P(x)$ and feasible set P(x) bounded above by its frontier $P^F(x)$. All $y \in P^F(x)$ are maximum output vectors that can be produced with x and given technology. The technically efficient output vector associated with y is $y^a = y/D_o(x,y)$, with $D_o(x,y)$ an output distance function defined as $D_o(x,y) = \min\{\lambda: y/\lambda \in P(x)$, and the technical efficiency of y is $y/y^a = D_o(x,y) \leq 1$. We next partition x into fixed and variable sub-vectors, so that $x = (x_f,x_v)$, and by fixity of x_f we mean $x_f \leq \bar{x}_f$. Following Gold (1955) and Johansen (1968), we define $P(\bar{x}_f)$ as the set of feasible output vectors obtainable from $x_f \leq \bar{x}_f$ when no constraint is imposed on the availability and use of x_v. $P(\bar{x}_f)$ is bounded above by its frontier $P^F(\bar{x}_f)$, and all $y \in P^F(\bar{x}_f)$ are full capacity output vectors, given $x_f \leq \bar{x}_f$ and technology.[3]

Our first definition of capacity and its rate of utilization follows Gold and Johansen, and solves an output maximization problem. It is independent of prices, and defines capacity output as the largest feasible radial expansion of y. In Figure 2 $y^{GJ} = y/D_o(\bar{x}_f,y) \in P^F(\bar{x}_f)$ is the full capacity output vector associated with actual output vector y, with $D_o(\bar{x}_f,y) = \min\{\lambda: y/\lambda \in P(\bar{x}_f)\}$, and so the rate of capacity utilization is $CU^{GJ} = D_o(\bar{x}_f,y) \leq 1$. The superscript "^{GJ}" honors the two pioneers, Gold and Johansen. CU^{GJ} is measured holding the output mix constant, and so is useful without output price information even when M > 1. CU^{GJ} is a gross measure that can be decomposed into the product of an output-oriented technical efficiency term $[D_o(x,y) \leq 1]$ and a net capacity utilization term $[D_o(\bar{x}_f,y)/D_o(x,y) \leq 1]$. We refer to the two components of CU^{GJ} as *wasted capacity* and *excess capacity*, respectively.[4]

Our second definition follows Segerson & Squires (1995) and Lindebo *et al.* (2007), and solves a revenue maximization problem.[5] It is dependent on the output price vector p, and defines capacity output as the vector $y^r \in P^F(\bar{x}_f)$ that solves the revenue maximization problem $\max_y\{p^Ty: x_f \leq \bar{x}_f\}$, and so the rate of capacity utilization is $CU^r = p^Ty/p^Ty^r \leq 1$. In Figure 2 the vectors $y^a = y/D_o(x,y) \in P^F(x)$ and $y^{GJ} = y/D_o(\bar{x}_f,y) \in P^F(\bar{x}_f)$ divide revenue-based capacity utilization into three components, an output-oriented technical efficiency term $p^Ty/p^Ty^a = D_o(x,y) \leq 1$ and a pair of capacity utilization components, a radial capacity utilization term $p^Ty^a/p^Ty^{GJ} = D_o(\bar{x}_f,y)/D_o(x,y) \leq 1$ and an output mix term p^Ty^{GJ}/p^Ty^r. We refer to the three components as wasted capacity, excess capacity, and *misallocated capacity*, respectively. Wasted capacity and excess capacity have the same interpretations and magnitudes as in the output maximization problem, and misallocated capacity is new. It captures the economic value of an optimizing movement along $P^F(\bar{x}_f)$ to adapt the output mix to prevailing output prices.

Our third definition follows Coelli *et al.* (2002), and solves a variable profit maximization problem, with variable profit $\pi_v = p^Ty - w_v^Tx_v$, w_v being the variable input price vector and $w_v^Tx_v$ being variable cost. This definition is dependent on two price vectors, p and w_v. It defines capacity output as the output vector $y^{v\pi} \in P^F(\bar{x}_f,x_v^{v\pi})$ that, together with $x_v^{v\pi}$, solves the variable profit maximization problem $\max_{y,x_v}\{p^Ty - w_v^Tx_v: x_f \leq \bar{x}_f\}$, so that maximum $\pi_v^{v\pi} = p^Ty^{v\pi} - w_v^Tx_v^{v\pi}$. The rate of capacity utilization is $CU^{v\pi} = p^Ty/p^Ty^{v\pi}$. The vectors $y^a = y/D_o(x,y) \in P^F(x)$ and $y^b = y/D_o(\bar{x}_f, x_v^{v\pi},y) \in P^F(\bar{x}_f,x_v^{v\pi})$ divide $CU^{v\pi}$ into an output-oriented technical efficiency term $p^Ty/p^Ty^a = D_o(x,y) \leq 1$ and a pair of capacity utilization components, a radial capacity utilization term $p^Ty^a/p^Ty^b = D_o(\bar{x}_f,x_v^{v\pi},y)/D_o(x,y) \leq 1$, and an output mix term $p^Ty^b/p^Ty^{v\pi}$. As in the revenue maximization problem we refer to the three components as wasted capacity, excess capacity, and misallocated capacity, although excess capacity and misallocated capacity have different magnitudes in the two problems.[6]

We are now prepared to introduce capacity utilization into a duPont triangle. For the output maximization problem we have

$$\pi/A = \pi/R \times R/A$$

$$= \pi/R \times p^T y/[p^T y/D_o(\bar{x}_f,y)] \times [p^T y/D_o(\bar{x}_f,y)]/A, \qquad (2)$$

in which CU^{GJ} is $p^T y/[p^T y/D_o(\bar{x}_f,y)] = p^T y/p^T y^{GJ} = R/R^{GJ} = D_o(\bar{x}_f,y)$, and asset turnover is converted to potential asset turnover, defined as $[p^T y/D_o(\bar{x}_f,y)]/A = p^T y^{GJ}/A = R^{GJ}/A$. Although CU^{GJ} appears to be price-dependent, prices appear in CU^{GJ} to implement the division operator, and to maintain a revenue-based numerator in the potential asset turnover term. As above, CU^{GJ} decomposes into wasted capacity and excess capacity components, and so expression (2) contains four drivers of ROA.

For the revenue maximization problem we have

$$\pi/A = \pi/R \times R/A$$

$$= \pi/R \times p^T y/p^T y^r \times p^T y^r/A, \qquad (3)$$

in which CU^r is $p^T y/p^T y^r = R/R^r$ and potential asset turnover is $p^T y^r/A = R^r/A$. In this case CU^r is price-dependent, and decomposes into wasted capacity, excess capacity and misallocated capacity. Consequently expression (3) contains five drivers of ROA.

For the variable profit maximization problem we have

$$\pi_v/A = \pi_v/R \times R/A$$

$$= \pi_v^{v\pi}/p^T y^{v\pi} \times \pi_v/\pi_v^{v\pi} \times p^T y^{v\pi}/A, \qquad (4)$$

in which the profit margin is converted to a potential profit margin $\pi_v^{v\pi}/p^T y^{v\pi} = \pi_v^{v\pi}/R^{v\pi}$, $CU^{v\pi}$ is $\pi_v/\pi_v^{v\pi}$, and potential asset turnover is $p^T y^{v\pi}/A = R^{v\pi}/A$. $CU^{v\pi}$ remains price-dependent, and decomposes into wasted capacity, excess capacity and misallocated capacity. Expression (4) also contains five drivers of ROA.

The three CU measures are derived from an analytical framework in which $x_f \le \bar{x}_f$, and therefore $C_f = w_f^T x_f \le \bar{C}_f = w_f^T \bar{x}_f$. However it is possible to impose $C_f \le \bar{C}_f$ without imposing $x_f \le \bar{x}_f$, thereby allowing substitution among fixed inputs along a fixed input budget constraint $C_f \le \bar{C}_f$. This formulation is particularly appropriate if information on w_f is unavailable. However if this information is available, then the constraints $x_f \le \bar{x}_f$ collapse to a single constraint $w_f^T x_f \le \bar{C}_f \Leftrightarrow (w_f/\bar{C}_f)^T x_f \le 1$. This strategy allows the construction of three "fixed cost indirect" CU measures corresponding to the three direct measures in expressions (2) – (4). In this case $P(\bar{x}_f)$ is replaced by $P(w_f/\bar{C}_f) \supseteq P(\bar{x}_f)$, and so each indirect CU measure is smaller than its corresponding direct CU measure. Referring to Figure 8.2, $P^F(x)$ remains unchanged, $P^F(\bar{x}_f, x_v^{v\pi})$ expands to $P^F(w_f/\bar{C}_f, x_v^{v\pi})$, and $P^F(\bar{x}_f)$ expands to $P^F(w_f/\bar{C}_f)$. The full capacity output quantity vectors increase accordingly.[7]

The output maximization problem becomes $\max_y\{y: (w_f/\bar{C}_f)^T x_f \leq 1\}$, and the associated duPont triangle is

$$
\begin{aligned}
\pi/A &= \pi/R \times R/A \\
&= \pi/R \times p^T y/[p^T y/D_o(w_f/\bar{C}_f, y)] \times [p^T y/D_o(w_f/\bar{C}_f, y)]/A, \qquad (5)
\end{aligned}
$$

in which the fixed cost indirect CU^{GJ} simplifies to $D_o(w_f/\bar{C}_f, y)$.

The revenue maximization problem becomes $\max_y\{p^T y: (w_f/\bar{C}_f) \leq 1\}$, and the associated duPont triangle is unchanged from that in expression (3), with the proviso that $y^r \in P^F(w_f/\bar{C}_f)$. The variable profit maximization problem becomes $\max_{y,x_v}\{p^T y - w_v{}^T x_v: (w_f/\bar{C}_f) \leq 1\}$, and the associated duPont triangle is unchanged from that in expression (4), with the proviso that $y^{v\pi} \in P^F(w_f/\bar{C}_f, x_v^{v\pi})$.

The direct and fixed cost indirect analyses are structurally similar; the only difference is the expansion of the direct output sets $P^F(x_f, x_v{}^{v\pi})$ and $P^F(x_f)$ to the fixed cost indirect output sets $P^F(w_f/\bar{C}_f, x_v^{v\pi})$ and $P^F(w_f/\bar{C}_f)$, and the corresponding reductions in capacity utilization. The virtues of the fixed cost indirect approach are (i) at the producer level it offers flexibility in the allocation of fixed cost budgets, (ii) at the industry level it offers managers and/or regulators an alternative way of restricting capacity, by assigning quotas to a single variable $C_f \leq \bar{C}_f$ rather than several $x_f \leq \bar{x}_f$, and (iii) at the analyst level it shrinks the number of direct constraints in an optimization problem.

We have introduced direct and fixed cost indirect measures of capacity utilization into a duPont triangle. We now attempt to introduce productivity into a duPont triangle by extending expression (1) to

$$
\begin{aligned}
\pi/A &= \pi/R \times p^T y/p^T y^c \times p^T y^c/A \\
&= [1 - (C/R)] \times p^T y/p^T y^c \times p^T y^c/A, \qquad (6)
\end{aligned}
$$

in which C/R is the ratio of cost to revenue, also known as the operating ratio or the expense ratio. Gold argued, convincingly, that productivity was negatively related to C and positively related to R, both of which are positively related to π/R.

Gold's argument is persuasive, but analytically deficient. Y/X does not appear explicitly in expression (6) as a driver of π/R. Its role is played out behind the scenes. There is a reason for its absence. The components of the duPont triangle are absolute variables describing levels. But Y/X is a relative variable describing change from one situation to another. Any attempt to incorporate a relative variable into a relationship among absolute variables is destined to fail. Incorporating productivity into a duPont triangle requires construction of a pair of triangles, so that change from one to another may be driven in part by productivity change. We undertake this exercise in Section 4.

4. Drivers of ROA Change

In this Section we convert an atemporal duPont triangle to an intertemporal duPont triangle change. We then show how change in the rate of capacity utilization and productivity change affect ROA change.

The ratio of comparison period to base period duPont triangles is

$$(\pi/A)^1/(\pi/A)^o = (\pi/R)^1/(\pi/R)^o \times (R/A)^1/(R/A)^o. \qquad (7)$$

We consider three different strategies for incorporating change in the rate of capacity utilization and productivity change into expression (7). All three strategies are based on Gold's expression

$$Y/X = Y^c/X \times Y/Y^c, \qquad (8)$$

in which Y and X are output and input quantity indexes and Y^c is a full capacity output quantity index derived from any one of the six direct and indirect capacity output vectors defined in Section 3. The three quantity indexes can be either theoretical technology-based indexes or empirical price-based indexes. Expression (8) states that actual productivity change Y/X is the product of potential productivity change Y^c/X and change in capacity utilization Y/Y^c. Gold provides a detailed discussion of the relationship, and of the relative merits of Y/X and the less volatile Y^c/X as productivity indexes.

One strategy is to introduce $Y/X = Y^c/X \times Y/Y^c$ directly into the profit margin change leg of expression (7), generating a model in which CU change influences Y/X, which influences π/R change, which drives ROA change. In this strategy the analysis proceeds in two steps. In the first step we use any of the six optimization problems to create a full capacity output vector y^c. In the second step we use y^c to derive $CU = Y/Y^c$ and to derive (and perhaps decompose) the Y^c/X component of Y/X. This generates the expression

$$(\pi/A)^1/(\pi/A)^o = (\pi/R)^1/(\pi/R)^o \times (R/A)^1/(R/A)^o$$
$$\uparrow$$
$$Y/X = Y^c/X \times Y/Y^c. \qquad (9)$$

A second strategy starts with a duPont triangle that incorporates capacity utilization. It converts the triangle to a triangle change and continues by introducing Y/X into the profit margin change leg of the triangle. This strategy generates a model in which Y/X influences ROA change through the π/R change leg, and CU change influences ROA change independently, but CU change does not influence Y/X. Using generic expression (1) and writing $R^c = p^T y^c$ generates the expression

$$(\pi/A)^1/(\pi/A)^o = (\pi/R)^1/(\pi/R)^o \times CU^1/CU^o \times (R^c/A)^1/(R^c/A)^o$$
$$\uparrow \text{ Y/X.} \qquad (10)$$

A third strategy starts with a duPont triangle that incorporates capacity utilization, converts it to a triangle change, and continues by introducing $Y/X = Y^c/X \times Y/Y^c$ into the profit margin change leg of the triangle change. This strategy generates a model in which Y/X influences ROA change through the π/R change leg, and CU change influences ROA change twice, once through its impact on productivity change in the profit margin change leg, and again independently. Again using expression (1) to illustrate, we have[8]

$$(\pi/A)^1/(\pi/A)^o = (\pi/R)^1/(\pi/R)^o \times CU^1/CU^o \times (R^c/A)^1/(R^c/A)^o$$
$$\uparrow \; Y/X = Y^c/X \times Y/Y^c. \qquad (11)$$

These three strategies raise an issue. What is the most likely relationship linking CU change, productivity change and ROA change? The first strategy is preferred if CU change influences productivity change, which influences ROA change, but CU change has no independent influence on ROA change. The second strategy is preferred if CU change influences ROA change independently and has no influence on productivity change. The third strategy encompasses the first two, and is preferred if CU change influences ROA change independently, and again through its impact on productivity change, which influences ROA change. All three choices face the challenge, originally encountered by Gold, of showing analytically how Y/X influences the profit margin change leg of the duPont triangle. The driving relationships in expressions (9) – (11) are hypotheses rather than analytical demonstrations. We meet this challenge in Section 5.

5. Incorporating Productivity Change into a duPont Triangle Change

In this Section we introduce price change, and we show how productivity change and price change drive margin change, and thus ROA change. We have already shown that it is straightforward to incorporate change in the rate of capacity utilization into a duPont triangle change expression, and we write, using y^c as the solution vector to any of the six direct and indirect optimization problems in Section 3,

$$\frac{(\boldsymbol{\pi}/\mathbf{A})^{\mathbf{1}}}{(\boldsymbol{\pi}/\mathbf{A})^{\mathbf{o}}} = \frac{(\boldsymbol{\pi}/\mathbf{R})^{\mathbf{1}}}{(\boldsymbol{\pi}/\mathbf{R})^{\mathbf{o}}} \times \frac{(\mathbf{p}^{\mathbf{1T}}\mathbf{y}^{\mathbf{1}})/(\mathbf{p}^{\mathbf{1T}}\mathbf{y}^{\mathbf{1c}})}{(\mathbf{p}^{\mathbf{oT}}\mathbf{y}^{\mathbf{o}})/(\mathbf{p}^{\mathbf{oT}}\mathbf{y}^{\mathbf{oc}})} \times \frac{(\mathbf{p}^{\mathbf{1T}}\mathbf{y}^{\mathbf{1c}})/\mathbf{A}^{\mathbf{1}}}{(\mathbf{p}^{\mathbf{oT}}\mathbf{y}^{\mathbf{oc}})/\mathbf{A}^{\mathbf{o}}}, \qquad (12)$$

which attributes ROA change to profit margin change, change in the rate of capacity utilization, and change in potential asset turnover. Change in the rate of capacity utilization exerts an independent influence on ROA change, but neither productivity change nor price change appears in expression (12).

We now consider how price change and productivity change influence ROA change. The key is to acknowledge that change in the profit margin derives from price change and quantity change, and we write

$$\begin{aligned}\frac{\pi^1/R^1}{\pi^o/R^o} &= \left[\frac{\pi^1/R^1}{\pi_o^1/R_o^1}\right] \times \left[\frac{\pi_o^1/R_o^1}{\pi^o/R^o}\right] \\ &= \left[\frac{\pi_1^o/R_1^o}{\pi^o/R^o}\right] \times \left[\frac{\pi^1/R^1}{\pi_1^o/R_1^o}\right], \qquad (13)\end{aligned}$$

where $\pi_o^1 = p^{oT}y^1 - w^{oT}x^1$ and $R_o^1 = p^{oT}y^1$ in the first equality are comparison period profit and revenue evaluated at base period prices, and π_1^o and R_1^o in the second equality are base period profit and revenue evaluated at comparison period prices. We focus on the first equality, in which the first term on the right side is that part of the margin change that can be attributed solely to price change, since it compares nominal and real comparison period margins. The second term on the right side is that part of the margin change attributable solely to quantity change, since it compares the real comparison period margin with the nominal base period margin. We return to the second equality in Section 5.2.[9]

We develop two strategies for decomposing the margin change component of ROA change. In the first we express the quantity effect in terms of the theoretical productivity index proposed by Caves *et al*. (1982). In the second we express the quantity effect in terms of empirical Laspeyres, Paasche and Fisher quantity indexes. Both strategies decompose the quantity effect, but in different ways. Problems with the first strategy include (i) the CCD productivity index is not in Y/X form; (ii) decomposing the quantity effect in terms of a CCD productivity index requires cost allocation, so that $w^Tx = c^Ty$, with $c \in R_{++}^M$ a vector of unit costs of producing each output; (iii) it is not possible to express the price effect in terms of a CCD price recovery index that has a meaningful economic interpretation; and (iv) the CCD productivity index must be estimated, which requires degrees of freedom. The second strategy requires information on output and input prices. Of course drawbacks of one strategy are strengths of the other.

5.1 The Theoretical CCD Productivity Index Strategy

We focus on the quantity effect $\left[\frac{\pi_o^1/R_o^1}{\pi^o/R^o}\right]$ in the first equality in expression (13). Assuming that cost allocation is feasible, we can write[10]

$$
\begin{aligned}
\pi^o &= p^{oT}y^o - w^{oT}x^o \\
&= p^{oT}y^o - c^{oT}y^o \\
&= (p^o - c^o)^T y^o, \qquad (14)
\end{aligned}
$$

where $w^{oT}x^o = c^{oT}y^o$, c^o being a vector of base period unit costs of producing each output. Writing base period profit in this way enables us to rewrite the base period profit margin as

$$
\begin{aligned}
\frac{\pi^o}{R^o} &= [(p^o - c^o)^T y^o]/R^o \\
&= [(p^o - c^o)/R^o]^T y^o] \\
&= \rho^{oT}y^o, \qquad (15)
\end{aligned}
$$

where $\rho^o = (p^o - c^o)/R^o$. Similarly, we can rewrite the real comparison period profit margin as

$$
\begin{aligned}
\frac{\pi_o^1}{R_o^1} &= [(p^o - c_o^1)^T y^1]/R_o^1 \\
&= [(p^o - c_o^1)/R_o^1]^T y^1 \\
&= \rho_o^{1T} y^1, \qquad (16)
\end{aligned}
$$

where $c_o^{1T}y^1 = w^{oT}x^1$ and $\rho_o^1 = (p^o - c_o^1)/R_o^1$. Consequently the quantity effect can be rewritten as

$$\left[\frac{\pi_o^1/R_o^1}{\pi^o/R^o}\right] = \frac{\rho_o^{1T}y^1}{\rho^{oT}y^o}. \qquad (17)$$

The next step is to interpret expression (17), which we do with the assistance of Figure 3, in which T^o and T^1 are base period and comparison period production frontiers analogous to $P^{Fo}(x^o)$ and $P^{F1}(x^1)$. We have

$$\left[\frac{\pi_o^1/R_o^1}{\pi^o/R^o}\right] = \left[\frac{\rho_o^{1T}y^1/\rho_o^{1T}y^C}{\rho^{oT}y^o/\rho_o^{1T}y^A}\right] \times \left[\frac{\rho^{oT}y^B}{\rho^{oT}y^A}\right] \times \left[\frac{\rho_o^{1T}y^C}{\rho^{oT}y^B}\right], \qquad (18)$$

where $y^A = y^o/D_o^o(x^o,y^o)$, $y^B = y^o/D_o^1(x^o,y^o)$ and $y^C = y^1/D_o^1(x^1,y^1)$. We can rewrite expression (18) as

$$\left[\frac{\pi_o^1/R_o^1}{\pi^o/R^o}\right] = \left[\frac{D_o^1(x^1,y^1)}{D_o^1(x^o,y^o)}\right] \times \left[\frac{\rho_o^{1T}y^C}{\rho^{oT}y^B}\right]$$

$$= \left[\frac{D_o^1(x^1,y^1)}{D_o^o(x^o,y^o)}\right] \times \left[\frac{D_o^o(x^o,y^o)}{D_o^1(x^o,y^o)}\right] \times \left[\frac{\rho_o^{1T}y^C}{\rho^{oT}y^B}\right], \tag{19}$$

where $\left[\frac{D_o^1(x^1,y^1)}{D_o^1(x^o,y^o)}\right] = \left[\frac{D_o^1(x^1,y^1)}{D_o^o(x^o,y^o)}\right] \times \left[\frac{D_o^o(x^o,y^o)}{D_o^1(x^o,y^o)}\right]$ is an output-oriented comparison period CCD productivity index. We know from Caves *et al.* (1982) that the two components $D_o^1(x^1,y^1)/D_o^o(x^o,y^o)$ and $D_o^o(x^o,y^o)/D_o^1(x^o,y^o)$ measure technical efficiency change and technical change respectively, as is apparent from Figure 3. Consequently[11]

$$\left[\frac{\pi_o^1/R_o^1}{\pi^o/R^o}\right] = M_{CCD}^1(y^o, y^1, x^o, x^1) \times \left[\frac{\rho_o^{1T}y^C}{\rho^{oT}y^B}\right]. \tag{20}$$

The term $[\rho_o^{1T}y^C/\rho^{oT}y^B]$ measures the productivity impact of size change that is absent from $M_{CCD}^1(y^o,y^1,x^o,x^1)$, and corresponds to the movement along T^1 from (x^o,y^B) to (x^1,y^C) in Figure 3. Thus the quantity effect $\left[\frac{\pi_o^1/R_o^1}{\pi^o/R^o}\right]$ is a measure of productivity change, because it includes the impact of size change along with the impacts of technical efficiency change and technical change.[12]

Substituting expression (20) into expression (12) yields a decomposition of ROA change incorporating (and decomposing and augmenting) a theoretical CCD productivity index

$$\frac{\pi^1/A^1}{\pi^o/A^o} = \left[\frac{\pi^1/R^1}{\pi_o^1/R_o^1}\right] \times \left[\frac{D_o^1(x^1,y^1)}{D_o^o(x^o,y^o)}\right] \times \left[\frac{D_o^o(x^o,y^o)}{D_o^1(x^o,y^o)}\right] \times \left[\frac{\rho_o^{1T}y^C}{\rho^{oT}y^B}\right]$$

$$\times \left[\frac{(p^{1T}y^1)/(p^{1T}y^{1c})}{(p^{oT}y^o)/(p^{oT}y^{oc})}\right] \times \left[\frac{R^{1c}/A^1}{R^{oc}/A^o}\right], \tag{21}$$

where $R^{tc} = p^{tT}y^{tc}$, t=o,1. Expression (21) attributes ROA change to price recovery change, three components of productivity change, change in capacity utilization and change in potential asset turnover. Although capacity utilization change influences ROA change, it does so without influencing productivity change.

Starting with the first equality in expression (13) leads to a decomposition of ROA change in expression (21) built on a comparison period CCD productivity index and a size change term measured along comparison period technology. Starting with the second equality in expression (13) and following the same procedures generates a decomposition of ROA change built on a base period CCD productivity index and a size change term measured along base period technology. Omitting intermediate steps, this decomposition is

$$\frac{\pi^1/A^1}{\pi^o/A^o} = \left[\frac{\pi_1^o/R_1^o}{\pi^o/R^o}\right] \times \left[\frac{D_o^1(x^1,y^1)}{D_o^o(x^o,y^o)}\right] \times \left[\frac{D_o^o(x^1,y^1)}{D_o^1(x^1,y^1)}\right] \times \left[\frac{\rho^{1T}y^D}{\rho_1^{oT}y^A}\right]$$

$$\times \left[\frac{(p^{1T}y^1)/(p^{1T}y^{1c})}{(p^{oT}y^o)/(p^{oT}y^{oc})}\right] \times \left[\frac{R^{1c}/A^1}{R^{oc}/A^o}\right], \qquad (22)$$

in which $\rho_1^o = [(p^1 - c_1^o)/R_1^o]^T y^o$, $R_1^o = p^1 y^o$ and y^D is located on T^o in Figure 3. Expression (22) decomposes ROA change into price recovery change, a base period CCD productivity index, a size change term measured along base period technology, change in capacity utilization and change in potential asset turnover. The capacity and turnover terms are the same as in expression (21). It is straightforward to calculate the geometric mean of expressions (21) and (22) to create an ROA change decomposition based on a geometric mean price recovery effect, a geometric mean CCD productivity index, and a geometric mean size change effect.

Kendrick & Grossman (1980) have argued, and demonstrated empirically, that productivity change is *positively* related to change in the rate of capacity utilization at the aggregate level. Many subsequent writers concur. Our objective is to introduce capacity utilization change as a driver of productivity change in expression (21).

The key to relating the two is contained in Gold's expression $Y/X = Y^c/X \times Y/Y^c$, which states that Y/X depends on change in CU, which is nice because a lot of empirical evidence supports the linkage, and the sign of the impact of CU change on Y/X is indeterminate, which is also nice because it makes pro-cyclicality a testable hypothesis. Suppose, as seems reasonable but not certain, that $Y^c/X \gtrless 1 \gtrless Y/Y^c$, so that potential productivity and capacity utilization move in opposite directions. Then productivity is pro-cyclical if $(Y/Y^c) \gtrless 1 \Rightarrow [(Y^c/X) \times (Y/Y^c)] \gtrless 1$ and counter-cyclical if $(Y/Y^c) \gtrless 1 \Rightarrow [(Y^c/X) \times (Y/Y^c)] \lesseqgtr 1$. Alternatively, if causation moves in the opposite direction, productivity is pro-cyclical if $(Y^c/X) \gtrless 1 \Rightarrow [(Y^c/X) \times (Y/Y^c)] \lesseqgtr 1$ and counter-cyclical if $(Y^c/X) \gtrless 1 \Rightarrow [(Y^c/X) \times (Y/Y^c)] \gtrless 1$. In words, productivity change is pro-cyclical if CU adjusts more than proportionately to change in potential productivity.[13]

Referring to Figure 2, in each period $p^{tT}y^t/p^{tT}y^{at} = p^{tT}y^t/p^{tT}y^{GJt} \div p^{tT}y^{at}/p^{tT}y^{GJt}$, t=o,1, which states that wasted capacity (technical inefficiency) can be expressed as the ratio of gross excess capacity to net excess capacity. Change in wasted capacity coincides with the technical efficiency change component of the CCD productivity index. Following De Borger & Kerstens (2000), we replace the technical efficiency change component of the CCD productivity index with the ratio of gross excess capacity to net excess capacity to obtain

$$\left[\frac{D_o^1(x^1,y^1)}{D_o^1(x^o,y^o)}\right] = \left[\frac{D_o^1(x_f^1,y^1)}{D_o^o(x_f^o,y^o)}\right] \times \left[\frac{D_o^1(x^1,y^1)/D_o^1(x_f^1,y^1)}{D_o^o(x^o,y^o)/D_o^o(x_f^o,y^o)}\right] \times \left[\frac{D_o^o(x^o,y^o)}{D_o^1(x^o,y^o)}\right], \tag{23}$$

which states that a CCD productivity index can be expressed as the product of technical efficiency change relative to $P^{Fo}(\bar{x}_f^o)$ and $P^{F1}(\bar{x}_f^1)$, change in the net rate of capacity utilization, and technical change. Substituting expression (23) into expression (19) yields

$$\left[\frac{\pi_o^1/R_o^1}{\pi^o/R^o}\right] = \left[\frac{D_o^1(x_f^1,y^1)}{D_o^o(x_f^o,y^o)}\right] \times \left[\frac{D_o^1(x^1,y^1)/D_o^1(x_f^1,y^1)}{D_o^o(x^o,y^o)/D_o^o(x_f^o,y^o)}\right] \times \left[\frac{D_o^o(x^o,y^o)}{D_o^1(x^o,y^o)}\right] \times \left[\frac{\rho_o^{1T}y^C}{\rho_o^{oT}y^B}\right], \tag{24}$$

which decomposes actual productivity change into change in net capacity utilization and potential productivity change (the CCD productivity index analogue to Y^c/X). Gold's expression (8) is embedded in expressions (23) and (24), in theoretical index number form. Finally, substituting expression (24) into expression (21) yields the complete CCD decomposition of ROA change

$$\frac{\pi^1/A^1}{\pi^o/A^o} = \left[\frac{\pi^1/R^1}{\pi_o^1/R_o^1}\right] \times \left[\frac{D_o^1\left(x_f^1,y^1\right)}{D_o^o\left(x_f^o,y^o\right)}\right] \times \left[\frac{D_o^1(x^1,y^1)/D_o^1\left(x_f^1,y^1\right)}{D_o^o(x^o,y^o)/D_o^o\left(x_f^o,y^o\right)}\right] \times \left[\frac{D_o^o(x^o,y^o)}{D_o^1(x^o,y^o)}\right]$$

$$\times \left[\frac{\rho_o^{1T}y^C}{\rho_o^{oT}y^B}\right] \times \left[\frac{(p^{1T}y^1)/(p^{1T}y^{1c})}{(p^{oT}y^o)/(p^{oT}y^{oc})}\right] \times \left[\frac{R^{1c}/A^1}{R^{oc}/A^o}\right], \tag{25}$$

which attributes ROA change to price change, potential productivity change, change in the rate of capacity utilization, and change in potential asset turnover. Change in capacity utilization plays a dual role, as an independent driver of ROA change, and as a driver of potential productivity change, which in turn drives ROA change. A similar decomposition can be derived from expression (22), and the geometric mean of expression (25) and the decomposition based on expression (22) can be calculated.[14]

5.2 The Empirical Index Number Strategy

Expressions (23) – (25) use a pair of augmented CCD productivity indexes to interpret the quantity effect as a productivity effect, on the assumption that cost allocation is feasible. Although these expressions do provide an augmented CCD productivity index interpretation of the quantity effect, they do not provide an analogous interpretation of the price recovery effect. This requires empirical quantity-based and price-based indexes.

A few mathematical manipulations enable us to write the price recovery effect in the first equality of expression (13) as

$$\left[\frac{\pi^1/R^1}{\pi_o^1/R_o^1}\right] = \frac{\pi^1}{R^1 - (\frac{P_P}{W_P})w^{1T}x^1}\ , \tag{26}$$

in which P_P/W_P is a quantity-based Paasche price recovery index, with $\left[\frac{\pi^1/R^1}{\pi_o^1/R_o^1}\right] \gtreqless 1$ $\Leftrightarrow$ $P_P/W_P \gtreqless 1$. Expression (26) contains comparison period and base period prices, but only comparison period quantities, and shows the contribution of P_P/W_P to profit margin change.

We follow the same strategy to write the quantity effect in the first equality of expression (13) as

$$\left[\frac{\pi_o^1/R_o^1}{\pi^o/R^o}\right] = \frac{\pi_o^1}{R_o^1 - (\frac{Y_L}{X_L})w^{oT}x^1}\ , \tag{27}$$

in which Y_L/X_L is a price-based Laspeyres productivity index, with $\left[\frac{\pi_o^1/R_o^1}{\pi^o/R^o}\right] \gtreqless 1 \Leftrightarrow$ $Y_L/X_L \gtreqless 1$. Expression (27) contains comparison period and base period quantities, but only base period prices, and shows the contribution of Y_L/X_L to profit margin change.

Substituting expressions (26) and (27) into expression (12) yields a decomposition of ROA change based on empirical price and quantity indexes

$$\frac{\pi^1/A^1}{\pi^o/A^o} = \frac{\pi^1}{R^1 - (\frac{P_P}{W_P})w^{1T}x^1} \times \frac{\pi_o^1}{R_o^1 - (\frac{Y_L}{X_L})w^{oT}x^1)}$$

$$\times \left[\frac{(p^{1T}y^1)/(p^{1T}y^{1c})}{(p^{oT}y^o)/(p^{oT}y^{oc})}\right] \times \left[\frac{R^{1c}/A^1}{R^{oc}/A^o}\right], \tag{28}$$

which attributes ROA change to price change, productivity change, change in capacity utilization and change in potential asset turnover. The difference between expressions (25) and (28) is that the augmented CCD productivity index decomposes by economic driver, while the Paasche price recovery index and the

Laspeyres productivity index decompose by variable. In both expressions price recovery change, productivity change and change in the rate of capacity utilization exert independent influences on ROA change. Change in the rate of capacity utilization is a driver of productivity change in expression (25), but not in expression (28). We explore this relationship next.

The key to relating productivity change to capacity utilization change is, as in Section 5.1, Gold's expression $Y/X = Y^c/X \times Y/Y^c$. If the quantity indexes Y, Y^c and X are empirical indexes we can write

$$\frac{p^Ty^1/p^Ty^o}{w^Tx^1/w^Tx^o} = \frac{p^Ty^{c1}/p^Ty^{co}}{w^Tx^1/w^Tx^o} \times \frac{p^Ty^1/p^Ty^o}{p^Ty^{c1}/p^Ty^{co}}$$

$$= \frac{p^Ty^{c1}/p^Ty^{co}}{w^Tx^1/w^Tx^o} \times \frac{p^Ty^1/p^Ty^{c1}}{p^Ty^o/p^Ty^{co}}\,, \tag{29}$$

where p and w can be base period or comparison period price vectors. The first term on the right side of expression (29) is Y^c/X and the second is Y/Y^c. The second equality rewrites and clarifies the capacity utilization change term. Expression (29) is interpreted exactly as Gold's expression (8), in empirical index number form. Substituting a Laspeyres version of expression (29) into expression (27) yields

$$\left[\frac{\pi_0^1/R_0^1}{\pi^o/R^o}\right] = \frac{\pi_0^1}{R_0^1 - \left[\left(\frac{Y_L}{Y_L^c}\right)\left(\frac{Y_L^c}{X_L}\right)\right]w^{oT}x^1}, \tag{30}$$

which expresses actual productivity change in terms of change in capacity utilization and potential productivity change. Substituting expression (30) into expression (28) yields

$$\frac{\pi^1/A^1}{\pi^o/A^o} = \frac{\pi^1}{R^1 - (\frac{P_P}{W_P})w^{1T}x^1} \times \frac{\pi_0^1}{R_0^1 - \left[\left(\frac{Y_L}{Y_L^c}\right)\left(\frac{Y_L^c}{X_L}\right)\right]w^{oT}x^1}$$

$$\times \left[\frac{(p^{1T}y^1)/(p^{1T}y^{1c})}{(p^{oT}y^o)/(p^{oT}y^{oc})}\right] \times \left[\frac{R^{1c}/A^1}{R^{oc}/A^o}\right], \tag{31}$$

which attributes ROA change to price recovery change, productivity change, change in the rate of capacity utilization and change in potential asset turnover. Change in the rate of capacity utilization appears twice, as an independent driver of ROA change, and as a driver of productivity change.

The first equality in expression (13) generates a Paasche price recovery effect and a Laspeyres quantity index that eventually make their way into the ROA change decomposition in expression (31). We now return to the second line, in which the first term is a price recovery effect and the second term is a quantity effect. It is easy to manipulate the two effects to generate

$$\left[\frac{\pi_1^o/R_1^o}{\pi^o/R^o}\right] = \frac{\pi_1^o}{R_1^o - (\frac{P_L}{W_L})w^{1T}x^o}\ , \tag{32}$$

which is a Laspeyres price recovery effect in which P_L/W_L is a Laspeyres price recovery index, with $\left[\frac{\pi_1^o/R_1^o}{\pi^o/R^o}\right] \gtreqless 1 \Leftrightarrow P_L/W_L \gtreqless 1$.[15] Similarly,

$$\left[\frac{\Pi^1/R^1}{\pi_1^o/R_1^o}\right] = \frac{\pi^1}{R^1 - (\frac{Y_P}{X_P})w^{1T}x^1}\ , \tag{33}$$

which is a Paasche productivity effect in which Y_P/X_P is a Paasche productivity index, with $\left[\frac{\Pi^1/R^1}{\pi_1^o/R_1^o}\right] \gtreqless 1 \Leftrightarrow Y_P/X_P \gtreqless 1$. Noting that $\frac{Y_P}{X_P} = \left(\frac{Y_P}{Y_P^C}\right)\left(\frac{Y_P^C}{X_P}\right)$, substituting this expression into expression (33), and replacing the price and quantity effects in expression (31) with those in expressions (32) and (33) generates

$$\frac{\pi^1/A^1}{\pi^o/A^o} = \frac{\pi_1^o}{R_1^o - (\frac{P_L}{W_L})w^{1T}x^o} \times \frac{\pi^1}{R^1 - \left(\frac{Y_P}{Y_P^C}\right)\left(\frac{Y_P^C}{X_P}\right)w^{1T}x^1}$$

$$\times \left[\frac{(p^{1T}y^1)/(p^{1T}y^{1c})}{(p^{oT}y^o)/(p^{oT}y^{oc})}\right] \times \left[\frac{R^{1c}/A^1}{R^{oc}/A^o}\right], \tag{34}$$

which is an alternative decomposition of ROA change based on a Laspeyres price recovery recovery index and a Paasche productivity index with capacity utilization appearing twice.

Taking the geometric mean of expressions (26) and (32) generates a Fisher price recovery effect, and taking the geometric mean of expressions (27) and (33) generates a Fisher productivity effect. It does not, however appear possible to express the Fisher price recovery effect in terms of P_F/W_F or the Fisher productivity effect in terms of Y_F/X_F.

The quantity vectors needed to implement the ROA change decompositions in expressions (31) and (34) (and also in expression (25) in Section 5.1) are either observed (y^1,y^o,x^1,x^o) or solutions to optimization problems specified above (y^{c1},y^{co}). All that is required is to specify a functional form for the index numbers in expressions (31) and (34) (and specify base period or comparison period technology and conditioning variables for the distance functions in expression (25) in Section 5.1). The two decompositions are interpreted in exactly the same way; the only difference is that one uses distance functions and the other uses prices to decompose productivity change and to measure change in capacity utilization.

6. External Capacity Constraints

Thus far we have treated capacity utilization as a short run phenomenon created by a fixed input constraint $x_f \leq \bar{x}_f$ or by a weaker fixed input expenditure constraint $C_f \leq \bar{C}_f$. These capacity constraints are internal to the firm. However firms also face external capacity constraints that have financial consequences. Mining firms are constrained by health, safety and environmental regulations, by weather conditions, by a lack of social infrastructure (e.g., housing and schools), and also by inadequate transport infrastructure that inhibits their ability to move minerals to ports to satisfy demand in a timely fashion.[16] Fishers are constrained by a variety of fishery management policies intended to limit catch in a fishery in pursuit of maximum economic yield. Input-oriented policies constrain fisher fixed input use, or “effort,” and output-oriented policies impose total allowable catch (TAC) limits on the fishery, often combined with individual transferrable quota (ITQ) allocation among fishers.[17] In both industries external capacity constraints may make at least some internal capacity constraints redundant for at least some firms at least some of the time.[18]

Figure 4, a simultaneously simplified and augmented version of Figure 2, illustrates the potential impact of external capacity constraints. Two internal frontiers, $P^F(x)$ and $P^F(\bar{x}_f)$, remain, and the third internal frontier, $P^F(\bar{x}_f, x_v^{v\pi})$ remains as well, but for expositional simplicity is replaced by a new external frontier $P^F(Z)$. The three internal frontiers are interpreted as before. The external frontier $P^F(Z)$ represents the collective impacts of industry management practices and regulations, supply chain bottlenecks and other production-limiting capacity constraints unrelated to $\bar{x}_f$ or $\bar{C}_f$.

Using the output maximization framework of Gold and Johansen, output vector y has wasted capacity p^Ty/p^Ty^a and excess capacity p^Ty^a/p^Ty^{GJ}. It also has over-capacity p^Ty^{GJ}/p^Ty^E. In mining overcapacity may be due to the transport infrastructure constraint, and in fishing it may be due to the imposition of TAC and

ITQ. The interpretation is similar in the revenue maximization and variable profit maximization frameworks, although y^E would not be a revenue maximizing or profit maximizing output mix given output price vector p. Since $P(Z) \subset P(\bar{x}_f)$, the internal capacity constraints associated with the output maximization and revenue maximization frameworks are rendered redundant by Z. The external capacity constraints have eliminated overcapacity by reducing capacity, thereby increasing capacity utilization from p^Ty/p^Ty^{GJ} to p^Ty/p^Ty^E. $P^F(Z)$ is not a neutral contraction of $P^F(\bar{x}_f)$, and may constrain some outputs proportionally more than others. $P^F(Z)$ may also constrain some firms more than others, inducing exit by relatively weak forms that creates a more efficient industry structure.

7. Summary and Conclusions

Change in the financial health of a business depends on trends in its price recovery, its productivity, its rate of capacity utilization, and in the external capacity constraints it faces. We have developed a pair of analytical frameworks with which to examine the relationship between change in financial health and its four drivers. We measure financial health with return on assets, and both analytical frameworks begin with the duPont triangle. The first framework exploits a theoretical productivity index, and the second is based on empirical price and quantity index numbers. Both frameworks provide valuable information to management concerning the likely sources of changes in its financial performance. The two frameworks have offsetting strengths. The first does not require price information, and decomposes the productivity effect into three economic drivers of productivity change, technical efficiency change, technical change, and size change. The second framework decomposes both the productivity effect and the price recovery effect into the contributions of individual quantity changes and price changes. The second does not require cost allocation, and it is calculated rather than estimated, so it does not face a degrees of freedom constraint. Both frameworks include change in capacity utilization twice, once as an independent driver of ROA change and again as a driver of productivity change. We also show how external capacity constraints influence capacity output.

Figures

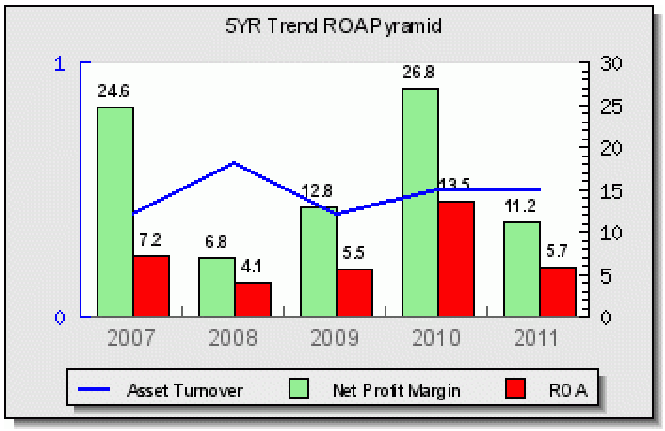
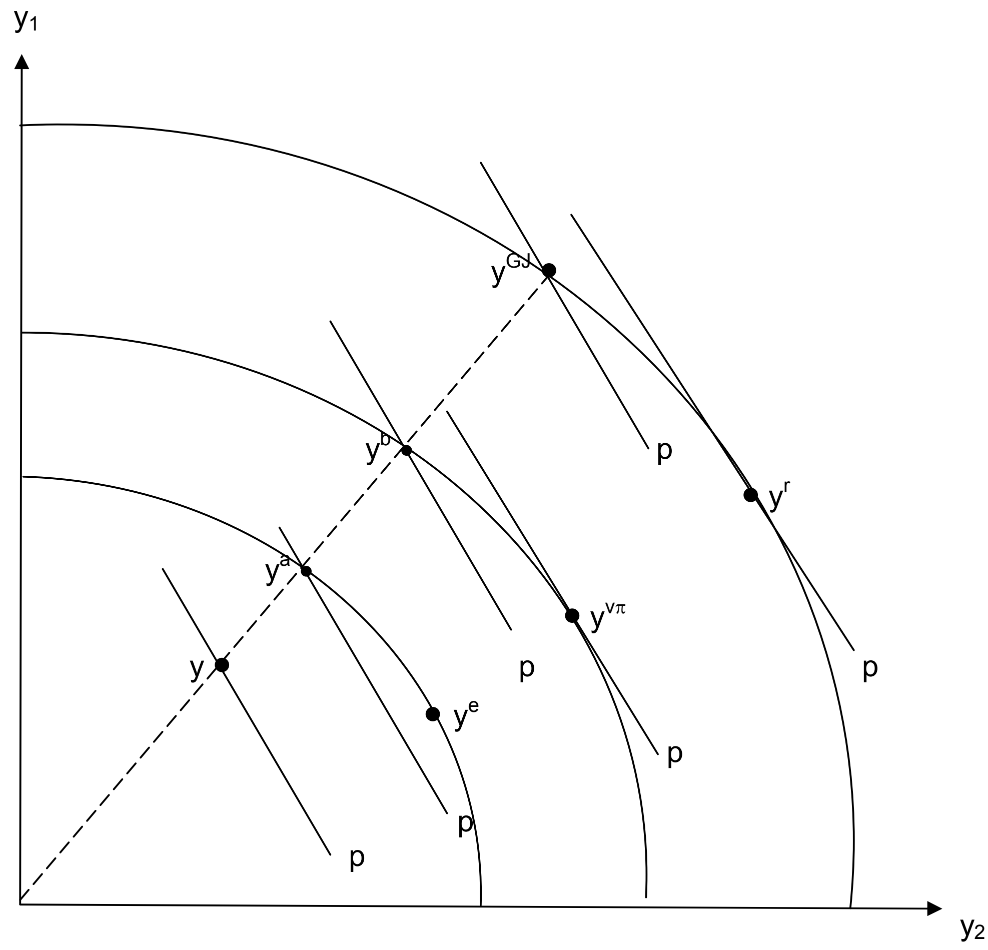


Figure 1 The duPont Triangle at Rio Tinto



$P^F(x)$ $P^F(\bar{x}_f, x_v^{v\pi})$ $P^F(\bar{x}_f)$

Figure 2 Capacity and its Rate of Utilization

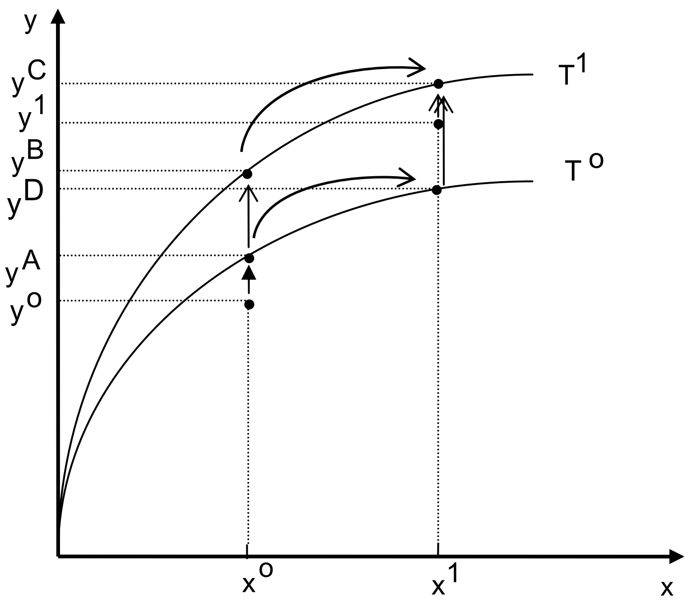


Figure 3 Output-Oriented Productivity Effect Decomposition

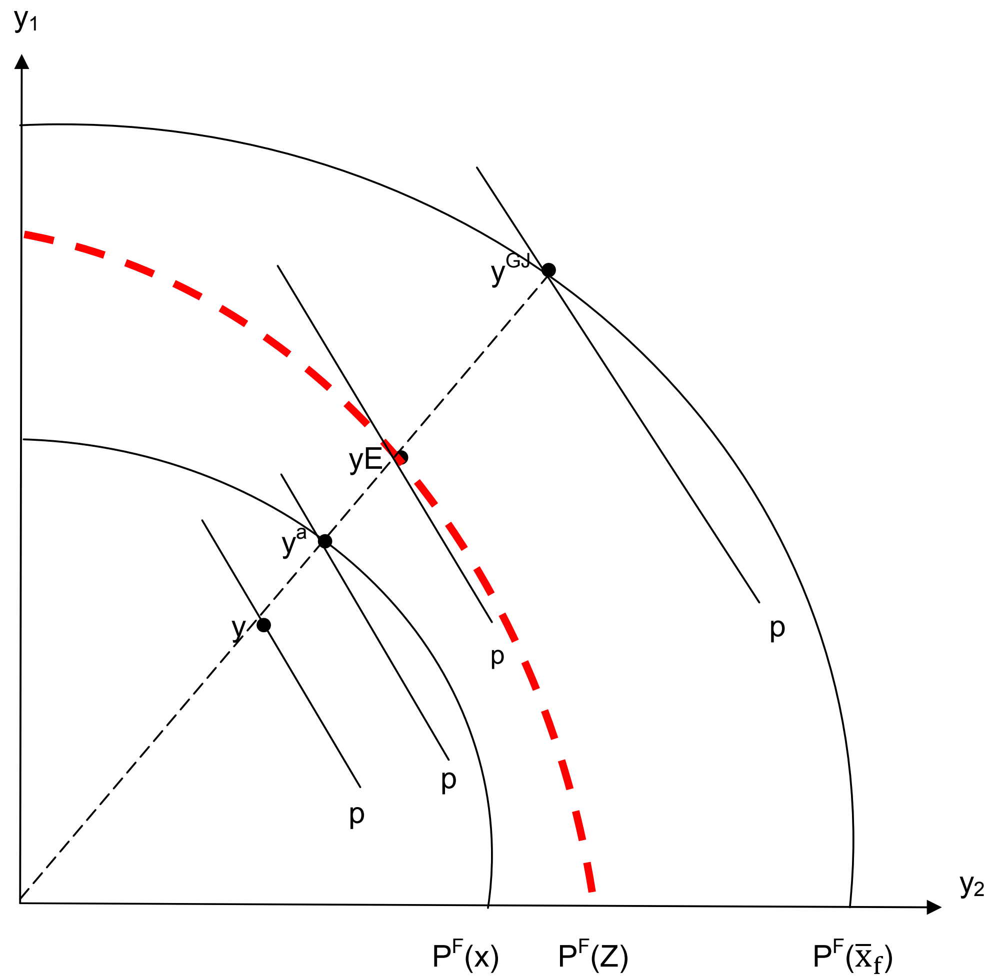


Figure 4 Internal and External Capacity Constraints

References

ABARES (Australian Bureau of Agricultural and Resource Economics and Sciences) (2012), *Australian Fisheries Surveys Report 2011: Results for Selected Fisheries 2008-09 to 2010-11.* www.daff.gov.au/abares/publications_remote_content/publication_topics/fisheries_and_aquaculture

Amey, L. R. (1969), *The Efficiency of Business Enterprises*. London: George Allen and Unwin Ltd.

Berndt, E. R., and C. J. Morrison (1981), "Capacity Utilization Measures: Underlying Economic Theory and an Alternative Approach," *American Economic Review* 71:2 (May), 48-52.

Bjurek, H. (1996), "The Malmquist Total Factor Productivity Index," *Scandinavian Journal of Economics* 98:2, 303-13.

Bliss, J. H. (1923), *Financial and Operating Ratios in Management*. New York: The Ronald Press Co.

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* 50:6 (November), 1393-1414.

Chandler, A. D. (1962), *Strategy and Structure: Chapters on the History of the Industrial Enterprise.* Cambridge, MA: The MIT Press.

Coelli, T., E. Grifell-Tatjé and S. Perelman (2002), "Capacity Utilisation and Profitability: A Decomposition of Short-Run Profitability," *International Journal of Production Economics* 79:3 (October), 261-78.

DeBorger, B., and K. Kerstens (2000), "The Malmquist Productivity Index and Plant Capacity Utilization," *Scandinavian Journal of Economics* 102:2 (June), 303-10.

Estache, A., and E. Grifell-Tatjé (2011), "Assessing the Impact of Mali's Water Privatization Across Stakeholders," ECARE Working Paper 2011-031, Université Libre de Bruxelles. www.ecares.org

FAO (2000), "Report of the Technical Consultation on the Measurement of Fishing Capacity," FAO Fisheries Report No. 615, Rome.

Färe, R., S. Grosskopf and J. Kirkley (2000), "Multi-Output Capacity Measures and their Relevance for Productivity," *Bulletin of Economic Research* 52:2 (April), 101-12.

Färe, R., S. Grosskopf, M. Norris and Z. Zhang (1994), "Productivity Growth, Technical Progress and Efficiency Change in Industrialized Countries," *American Economic Review* 84:1 (March), 66-83.

Frankel, M. (1963), "Review of Kendrick & Creamer (1961)," *Journal of the American Statistical Association* 58:301 (March), 258-59.

Gold, B. (1955), *Foundations of Productivity Analysis*. Pittsburgh, PA: University of Pittsburgh Press.

Grifell-Tatjé, E., and C. A. K. Lovell (1999a), "A Generalized Malmquist Productivity Index," *Top* 7:1, 81-101.

Johansen, L. (1968), "Production Functions and the Concept of Capacity," in *Recherches Récent sur la Function de Production. Collection Économie Mathématique et Économétrie 2*, Centre d'Études et de la Recherche, Université de Namur, Belgium. [Reprinted in F. R. Førsund, ed., *Collected Works of Leif Johansen, Vol 1.* Amsterdam: North-Holland (1987)].

Johnson, H. T. (1975), "Management Accounting in an Early Integrated Industrial: E. I. duPont de Nemours Powder Company, 1903-1912," *Business History Review* 49:2 (Summer), 184-204.

Johnson, H. T. (1978), "Management Accounting in an Early Multidivisional Organization: General Motors in the 1920s," *Business History Review* 52:4 (Winter), 490-517.

Kendrick, J. W., and E. S. Grossman (1980), *Productivity in the United States: Trends and Cycles*. Baltimore: Johns Hopkins University Press.

Klein, L. R. (1960), "Some Theoretical Issues in the Measurement of Capacity," *Econometrica* 28:2 (April), 272-86.

Kline, C. A. Jr., and H. L. Hessler (1952), "The du Pont Chart System for Appraising Operating Performance," *N. A. C. A. Bulletin* (August, Section 3), 1595-1619.

de Leeuw, F. (1962), "The Concept of Capacity," *Journal of the American Statistical Association* 57:300 (December), 826-40.

Lindebo, E., A. Hoff and N. Vestergaard (2007), "Revenue-based Capacity Utilisation Measures and Decomposition: The Case of Danish North Sea Trawlers," *European Journal of Operational Research* 180:1 (July), 215-27.

Pincus, J., and H. Ergas (2008), "Improving the Efficiency of Australian Coal Chains," Report to the Queensland Competition Authority. www.qca.org.au

Ray, S. C., and E. Desli (1997), "Productivity Growth, Technical Progress, and Efficiency Change in Industrialized Countries: Comment," *American Economic Review* 87:5 (December), 1033-39.

Schultze, C. L. (1963), "Uses of Capacity Measures for Short-Run Economic Analysis," *American Economic Review* 53:2 (May), 293-308.

Segerson, K., and D. Squires (1995), "Measurement of Capacity Utilization for Revenue-Maximizing Firms," *Bulletin of Economic Research* 47:1 (January), 77-84.

Shephard, R. W. (1974), *Indirect Production Functions.* Mathematical Systems in Economics, No. 10. Meisenheim Am Glan, Germany: Verlag Anton Hain.

Shubik, M. (2011), "A Note on Accounting and Economic Theory: Past, Present and Future," *Accounting, Economics, and Law* 1:1 www.bepress.com/ael/vol1iss1/1

Squires, D., R. Yongil Jeon, Q. Grafton and J. Kirkley (2010), "Controlling Excess Capacity in Common-Pool Resource Industries: The Transition from Input to Output Controls," *Australian Journal of Agricultural and Resource Economics* 54:3 (July), 361-77.

*This paper is an outgrowth of a presentation by the second author at the National Marine Fisheries Service Productivity Workshop, Southwest Fisheries Science Center, Santa Cruz, CA, June 11-12, 2012. We are grateful to John Walden for the invitation to attend the Workshop, and to Aaron Mamula and John Walden for organizing the Workshop.

[1] Source: http://au.advfn.com.

[2] Chandler (1962) and Johnson (1975, 1978) detail the development and use of the ROA triangle at duPont and GM.

[3] Gold and Johansen proposed virtually identical physical definitions of y^c, and their definition of y^c was given a managerial slant akin to the use of standard volume at GM. Gold emphasized "practically sustainable capacity," determined by "the customary number of shifts and the normally acceptable length of work day and work week," and with allowance made for breakdowns, repairs and maintenance. Johansen conditioned his definition on the assumption that the firm is "operating under normal conditions with respect to number of shifts, hours of work etc."

[4] The United Nations Food and Agriculture Organization (FAO) (2000) has endorsed the physical measure of capacity utilization proposed by Gold and Johansen, in part due to the shortage of reliable information on output and variable input prices.

[5] Segerson & Squires justify a revenue maximization objective on the grounds that in the short run *all* inputs are quasi-fixed, so that $x = x_f$. Their CU analysis is based on a dual shadow price approach.

[6] If M=1 the solution to the variable profit maximization problem is very similar to the solution to the short run average cost minimization problem proposed by Klein (1960) and Berndt & Morrison (1981) and widely used in the fisheries literature. Sources of the difference are (i) price ≠ minimum short run average cost and (ii) minimum short run average cost ≠ minimum short run average variable cost. An overlooked definition of full capacity output was proposed by de Leeuw (1962), who defined capacity output as that level of output at which short run marginal cost exceeds minimum short run average cost by some percent, the logic being that at that output level marginal cost is well above minimum average cost, signalling upward pressure on output price.

[7] The theory of cost indirect and return indirect production was developed by Shephard (1974). Empirical applications are regrettably rare. A fixed cost indirect capacity utilization measure was proposed by Färe *et al.* (2000).

[8] Schultze (1963) summarizes the theory behind and evidence for the argument that changes in capacity utilization influence productivity change and profit margin change.

[9] It does not appear possible to implement decomposition (13) into pure price and quantity effects using Edgeworth-Marshall arithmetic mean price and quantity vectors $(\bar{p}, \bar{w})$ and $(\bar{y}, \bar{x})$ because this introduces three pairs of price vectors (p^o,w^o), (p^1,w^1) and $(\bar{p},\bar{w})$, and three pairs of quantity vectors (y^o,x^o), (y^1,x^1) and $(\bar{y}, \bar{x})$.

[10] Cost allocation is a contentious issue. Allocating operating cost is feasible, although the allocation may not be optimal, but allocating overhead cost is difficult; Shubik (2011) calls it an open problem in economic theory and accounting. Estache & Grifell-Tatjé (2011) compromise by ignoring overhead cost, or general expenses, and allocating operating cost to three activities in a sample of Mali water utilities.

[11] Although the CCD productivity index is not in Y/X form, we can calculate $M_{CCD}(y,x)$ and $M^c_{CCD}(y^c,x)$ and define change in capacity utilization residually as $M_{CCD}(y,x)/M^c_{CCD}(y^c,x)$.

[12] Expression (20) augments the CCD productivity index with what we call a size change term, in an effort to introduce a size-related driver of productivity change that might capture

the joint impacts of economies of scale and diversification. Our effort has several antecedents; Färe *et al.* (1994), Ray & Desli (1997) and Grifell-Tatjé & Lovell (1999) all augment the CCD productivity index, which ignores the potential impact of size change on productivity change, with a size change term, although these terms differ.

[13] The indexes Y, Y^c and X, and therefore Y/X, Y^c/X and Y/Y^c, must equal unity in the base period. Thus, for example, CU grows or shrinks from an initial value of unity. However we observe or solve for the underlying output quantity vectors. This allows us to calculate $CU_m^t = y_m^t/y_m^{ct}$, m=1,…,M, t=o,1, for each output individually, or we can calculate an aggregate price-dependent measure $CU^t = R^t/R^{ct} = p^{tT}y^t/p^{tT}y^{ct}$.

[14] We base our decompositions on a CCD productivity index. We prefer to decompose the Malmquist productivity index proposed by Bjurek (1996), in part because it is in Y/X form. This index decomposes as

$$\frac{D_O(x,y^1)/D_O(x,y^o)}{D_I(y,x^1)/D_I(y,x^o)} = \frac{D_O(x,y^{c1})/D_O(x,y^{co})}{D_I(y,x^1)/D_I(y,x^o)} \times \frac{D_O(x,y^1)/D_O(x,y^{c1})}{D_O(x,y^o)/D_O(x,y^{co})},$$

where $D_I(y,x)$ is an input distance function. The first term on the right side is Y^c/X and the second is Y/Y^c. Unfortunately it does not appear possible to link this productivity index with the quantity effect $\left[\frac{\pi_O^1/R_O^1}{\pi^o/R^o}\right]$.

[15] Frankel (1963) recommends use of Paasche quantity indexes (and, to satisfy the product test, Laspeyres price indexes) because, being based on comparison period weights, they are better suited to a company's current needs than are the more popular Laspeyres quantity indexes.

[16] Mining Australia reports that floods in 2011 reduced Queensland's coal exports by 20%. http://www.miningaustralia.com.au/news/qld-flood-damage-confirmed. Pincus & Ergas (2008) analyze Australian mining supply infrastructure bottlenecks, due in part to diffuse and uncoordinated ownership of port terminals, tracks and rolling stock. They cite a study commissioned by the Queensland government that estimated that revenues in excess of a billion AUD per year were being sacrificed to inefficiencies in a single coal supply chain.

[17] Squires *et al.* (2010) provide evidence on the capacity-reducing and distributional impacts of TAC and ITQ in the British Columbia halibut fishery.

[18] Overcapacity in a fishery results from lack of ownership, which creates a tragedy of the commons; external capacity constraints such as TAC and ITQ are intended to create property rights and alter fisher incentives. Overcapacity in mining results from the opposite problem, diffuse and uncoordinated ownership of links in the transport infrastructure.

Primal Productivity Indices: Unbalanced vs. Balanced Panel Data

Kristiaan Kersten*   Ignace Van de Woestyne†

June 2012

Abstract

We explore the effect of balancing unbalanced panel data when estimating primal productivity indices using non-parametric frontier estimators. First, we explore a series of pseudo-solutions aimed at making an unbalanced panel balanced. Then, we discuss some intermediate solutions (e.g., balancing 2-years by 2 years). We empirically illustrate these issues comparing both Malmquist and Hicks-Moorsteen productivity indices. Furthermore, we link this problem with a variety of literatures on infeasibilities, statistical inference of non-parametric frontier estimators, and the index theory literature focusing on the dynamics of entry and exit in industries. Finally, we draw up a list of remaining issues that could benefit from further exploration.

Keywords: Malmquist productivity index, Hicks-Moorsteen productivity index, Balanced panel, Unbalanced panel.

1 Introduction

Traditionally, Total Factor Productivity (TFP) growth is estimated by the traditional Solow residual and yields an index number representing technology shifts from output growth that remains unexplained by input growth (see Hulten (2001) or Van Beveren (2010)). In the last decades, economists have become conscious that ignoring inefficiency may well bias TFP measures. Nishimizu and Page (1982) is probably the seminal article

*CNRS-LEM (UMR 8179), IESEG School of Management, 3 rue de la Digue, F-59000 Lille, France. k.kerstens@ieseg.fr, Corresponding author.

†Hogeschool-Universiteit Brussel, Warmoesberg 26, B-1000 Brussel, Belgium.

suggesting to decompose TFP into a technical change component as well as a technical efficiency change component. Caves, Christensen, and Diewert (1982) have analysed discrete time Malmquist input, output and productivity indices using distance functions as general technology representations. Since these Malmquist indices require a precise knowledge on the technology, these authors relate Malmquist and Törnqvist productivity indices, the latter depending on both price and quantity information (without need of exact knowledge on the technology).

Integrating the two-part Nishimizu and Page (1982) decomposition of TFP, Färe, Grosskopf, Norris, and Zhang (1994) propose to estimate the output distance functions in the Malmquist output productivity index by exploiting their inverse relation with the radial output efficiency measures evaluated relative to multiple input and output non-parametric technologies. Meanwhile, parametric estimates of the underlying distance functions of this Malmquist productivity index approach have also been reported in the literature (see, e.g., Atkinson, Cornwell, and Honerkamp (2003) or Tsekouras, Pantzios, and Karagiannis (2004)). Bjurek (1996) offers an alternative Hicks-Moorsteen TFP index, defined as a ratio of a Malmquist output over a Malmquist input index (see also O'Donnell (2010) and O'Donnell (2012a)). Finally, it is good to indicate that these primal productivity indices have become relatively popular in empirical work in comparison with more traditional TFP measures.

This paper concentrates on a seemingly rather widespread misconception that these primal productivity indices require balanced panel data and cannot cope with unbalancedness. Just to cite one example, Hollingsworth and Wildman (2003) state that "DEA based Malmquist techniques are unable to cope with unbalanced panel estimation procedures" (page 497). One reason for such beliefs could be that some of the popular software options around to compute these productivity indices cannot handle unbalanced panels. For instance, the still popular DEAP software of Coelli (1996) explicitly requires a balanced panel (see p. 31 of the manual). Another example of the same explicit requirement is the R-package "Nonparaeff" (version 0.5-3: page 14). Such software restrictions may induce people to believe balanced panels are a prerequisite for this Malmquist productivity index approach. This is to some extent surprising given that some of the seminal articles in the literature on the Malmquist productivity index have clearly pointed out that the use of an unbalanced panel is possible, "although the index will be undefined for missing observations" (see fn 14 on page 73 of Färe, Grosskopf, Norris, and Zhang (1994)).

While the notion of a potential unbalancedness bias due to unplanned missing data is quite standard in the statistical literature (see, e.g., Baltagi and Song (2006) or Frees (2004)), to the best of our knowledge nobody has so far analysed the extent of the differences between computing primal productivity indices using balanced and unbalanced panel data. In this contribution, we intend to systematically start exploring the consequences of computing these primal productivity indices using a balanced panel when initially an unbalanced panel data set is available. In particular, this paper is structured as follows. Section 2 provides some basic definitions of the technology, and of the Malmquist productivity index as well as the Hicks-Moorsteen TFP index. Section 3 offers a structured overview of different “solutions” advanced in the literature to cope with unbalanced panel data when computing these primal productivity indices. We argue against most of these pseudo-solutions. In Section 4, the effect of the balancedness or unbalancedness of the sample is illustrated using an existing data set. The final Section 5 concludes and outlines future research issues.

2 Definitions of Technology and Primal Productivity Indices

We first introduce the assumptions on technology and the definitions of the required distance functions. The latter provide the components for computing the primal productivity indices.

2.1 Technology and Distance Functions

A production technology describes how a vector of inputs $x = (x_1, \ldots, x_n) \in \mathbb{R}^n_+$ is transformed into a vector of outputs $y = (y_1, \ldots, y_p) \in \mathbb{R}^p_+$. For each time period t, the production possibility set (or technology for short) T^t summarises the set of all feasible vectors of input and output. It is defined as follows:

$$T^t = \left\{(x^t, y^t) \in \mathbb{R}^{n+p}_+ : x^t \text{ can produce } y^t\right\}. \quad (1)$$

Throughout this contribution, technology is assumed to satisfy the following conventional assumptions:

$(T.1)$ $(0,0) \in T^t,\ (0, y^t) \in T^t \Rightarrow y^t = 0$.
$(T.2)$ The set $A(x^t) = \{(u^t, y^t) \in T^t : u^t \leq x^t\}$ of dominating observations is bounded $\forall x^t \in \mathbb{R}^n_+$.
$(T.3)$ T^t is closed.
$(T.4)$ $\forall (x^t, y^t) \in T^t : (x^t, -y^t) \leq (u^t, -v^t)$ and $(u^t, v^t) \geq 0$ implies that $(u^t, v^t) \in T^t$.
The first axiom creates the possibility of inaction and also states that there is no free lunch. The second axiom of boundedness (i.e., infinite outputs can not be obtained

from a finite input vector) is just a mathematical regularity condition, as is closedness of technology assumed in the third axiom. The fourth axiom of strong disposal of inputs and outputs implies that fewer outputs can always be produced with more inputs, and inversely.

Sometimes, the following two additional axioms are assumed in various combinations with the preceding ones as well:

$(T.5)$ T^t is a convex set.
$(T.6)$ $\delta T^t \subseteq T^t, \forall \delta > 0$.
Convexity of technology in the fifth axiom allows for linear combinations of activities to remain feasible. The sixth axiom imposes constant returns to scale rather than a more flexible variable returns to scale hypothesis that is traditionally maintained.

Efficiency is estimated relative to technologies using distance or gauge functions. Distance functions are related to the efficiency measures defined by Farrell (1957). In general, the Farrell efficiency measure $E_t(x^t, y^t)$ is defined as the inverse of the Shephardian distance function. In the input-orientation, this Farrell efficiency measure $E_t^i(x^t, y^t)$ indicates the minimum contraction of an input vector by a scalar λ while still remaining on the boundary of the technology:

$$E_t^i(x^t, y^t) = \inf_{\lambda} \left\{ \lambda : (\lambda x^t, y^t) \in T^t, \lambda \geq 0 \right\}. \tag{2}$$

In the output-orientation, the Farrell efficiency measure $E_t^o(x^t, y^t)$ searches for the maximum expansion of the output vector by a scalar θ to the boundary of the technology:

$$E_t^o(x^t, y^t) = \sup_{\theta} \left\{ \theta : (x^t, \theta y^t) \in T^t, \theta \geq 1 \right\}. \tag{3}$$

For all $(a, b) \in \{t, t+1\}^2$, the time-related versions of the Farrell input efficiency measure are given by

$$E_a^i(x^b, y^b) = \inf_{\lambda} \left\{ \lambda : (\lambda x^b, y^b) \in T^a \right\} \tag{4}$$

if there is some λ such that $(\lambda x^b, y^b) \in T^a$ and $E_a^i(x^b, y^b) = +\infty$ otherwise. Similarly, in the output case, $E_a^o(x^b, y^b) = \sup_{\theta} \left\{ \theta : (x^b, \theta y^b) \in T^a \right\}$ if there is some θ such that $(x^b, \theta y^b) \in T^a$ and $E_a^o(x^b, y^b) = -\infty$ otherwise.

2.2 Malmquist Productivity Index

Following Caves, Christensen, and Diewert (1982), using the input Farrell measures one can define the input-oriented Malmquist productivity index in base period t as follows:

$$M_t^i(x^t, y^t, x^{t+1}, y^{t+1}) = \frac{E_t^i(x^t, y^t)}{E_t^i(x^{t+1}, y^{t+1})}. \tag{5}$$

Values of this base period t input-oriented Malmquist productivity index below (above) unity reveal productivity growth (decline).

Similarly, a base period $t+1$ input-oriented Malmquist productivity index is defined as follows:

$$M_{t+1}^i(x^t, y^t, x^{t+1}, y^{t+1}) = \frac{E_{t+1}^i(x^t, y^t)}{E_{t+1}^i(x^{t+1}, y^{t+1})}. \tag{6}$$

Again, values of this base period $t+1$ input-oriented Malmquist productivity index below (above) unity reveal productivity growth (decline).

To avoid an arbitrary selection among base years, the input-oriented Malmquist productivity index is defined as a geometric mean of a period t and a period $t+1$ index:

$$M_{t,t+1}^i = \sqrt{M_t^i \cdot M_{t+1}^i}, \tag{7}$$

whereby the arguments of the functions are suppressed to save space. Note again that when the geometric mean input-oriented Malmquist productivity index is smaller (larger) than unity, it points to a productivity growth (decline).

2.3 Hicks-Moorsteen Productivity Index

Following the seminal article by Bjurek (1996), a Hicks-Moorsteen productivity (or Malmquist TFP) index with a base period t is defined as the ratio of a Malmquist output quantity index in base period t over a Malmquist input quantity index in the same base period t:

$$HM_t(x^t, y^t, x^{t+1}, y^{t+1}) = \frac{MO_t(x^t, y^t, y^{t+1})}{MI_t(x^t, x^{t+1}, y^t)} \tag{8}$$

whereby the output quantity index is defined as $MO_t(x^t, y^t, y^{t+1}) = \frac{E_t^o(x^t, y^t)}{E_t^o(x^t, y^{t+1})}$ and the input quantity index is defined as $MI_t(x^t, x^{t+1}, y^t) = \frac{E_t^i(x^t, y^t)}{E_t^i(x^{t+1}, y^t)}$. If the Hicks-Moorsteen productivity index is larger (smaller) than unity, then it indicates a gain (loss) in productivity.

Similarly, a base period $t+1$ Hicks-Moorsteen productivity index is defined as follows:

$$HM_{t+1}(x^t, y^t, x^{t+1}, y^{t+1}) = \frac{MO_{t+1}(x^{t+1}, y^{t+1}, y^t)}{MI_{t+1}(x^t, x^{t+1}, y^{t+1})} \tag{9}$$

where we now have for the output quantity index $MO_{t+1}(x^{t+1}, y^{t+1}, y^t) = \frac{E^o_{t+1}(x^{t+1}, y^t)}{E^o_{t+1}(x^{t+1}, y^{t+1})}$ and for the input quantity index $MI_{t+1}(x^t, x^{t+1}, y^{t+1}) = \frac{E^i_{t+1}(x^t, y^{t+1})}{E^i_{t+1}(x^{t+1}, y^{t+1})}$. Again, when the Hicks-Moorsteen productivity index is larger (smaller) than unity, it points to a productivity gain (loss).

To avoid a choice of base year, it is customary to take a geometric mean of these two Hicks-Moorsteen productivity indices (see Bjurek (1996)):

$$HM_{t,t+1} = \sqrt{HM_t \cdot HM_{t+1}}, \tag{10}$$

where arguments of the functions are suppressed for reasons of space. Note once more that when the geometric mean Hicks-Moorsteen productivity index is larger (smaller) than unity, it points to a productivity gain (loss).

A final observation can be made. The denominator of both the Malmquist output and input quantity indices in the base period t Hicks-Moorsteen productivity index compares a "hypothetical" or pseudo-observation consisting of inputs and outputs observed from different periods to a technology in period t. The same remark applies to the numerator for the corresponding Malmquist output and input quantity indices in base period $t+1$. Such "hypothetical" observations do not appear in the Malmquist productivity index, which makes for a somewhat easier interpretation.

2.4 Primal Productivity Indices: A Comparison

We end with some remarks regarding the relative popularity as well as the properties of both these primal productivity indices (see also O'Donnell (2012a) for more details).

First, the Malmquist productivity index has recently become very popular. By contrast, the Hicks-Moorsteen productivity index has so far found rather limited use in applied research (e.g., Arora and Arora (2012), Nemoto and Goto (2005), O'Donnell (2012b), O'Donnell (2012a) or Zaim (2004)).

Second, both ratio-based productivity indices can be related to one another under rather stringent conditions. Indeed, an analytical relation is established in Färe, Grosskopf, and Roos (1996): both indices coincide under constant returns to scale and inverse homotheticity. Empirical studies comparing both indices are extremely rare: for example, Bjurek, Førsund, and Hjalmarsson (1998) report minor differences between both indices. This limited empirical evidence could be taken as an indication that the conditions on technology under which both indices coincide do not seem to hold exactly.

Third, one well-known pitfall of the Malmquist productivity index is that it is not always a TFP index. For instance, while its TFP properties are maintained under constant returns to scale, as illustrated by Grifell-Tatjé and Lovell (1995), these are not preserved in the presence of variable returns to scale (i.e., a more general technology). By contrast, already Bjurek (1996) states that the Hicks-Moorsteen productivity index has a TFP interpretation. More recently, O'Donnell (2010) shows that profitability change can be decomposed into the product of a total factor productivity (TFP) index and an index measuring relative price changes. Many TFP indices can be decomposed into measures of technical change and technical efficiency change (following Nishimizu and Page (1982)), but furthermore into scale efficiency change and mix efficiency change components. Indices that can be decomposed in this way include the Fisher, Törnqvist and Hicks-Moorsteen TFP indices, but not the Malmquist productivity index. In fact, Grosskopf (2003) suggests to call the Malmquist productivity index a technology index. In other words, it just measures local technical change, not TFP change.

Fourth, another problem known since the beginning of this literature is that some of the distance functions constituting the Malmquist productivity index may well be undefined when estimated using general technologies (see Färe, Grosskopf, Norris, and Zhang (1994), footnote 15). However, empirical studies often ignore reporting on this infeasibility problem. Briec and Kerstens (2009) prove that infeasibilities can occur for an even more general productivity indicator based upon more general distance functions. Thus, even this more general indicator does not satisfy the determinateness property in index theory. By contrast, the Hicks-Moorsteen index satisfies the determinateness axiom, as conjectured by Bjurek (1996) and proven in Briec and Kerstens (2011) under mild conditions (i.e., mainly strong disposability of inputs and outputs).[1]

Fifth, as mentioned in the introduction, both these primal productivity indices can be computed on balanced and unbalanced panel data alike. However, in view of the preceding remark it is critical to distinguish between an infeasibility due to unavailable data (e.g., related to the unbalanced nature of the panel) and a computational infeasibility. The former case could probably better be called a logical impossibility because one simply cannot measure the adjacent period efficiency measures.

Overall, the TFP nature of the Hicks-Moorsteen index and the fact that it can easily be

[1]Zaim (2004) employs a Hicks-Moorsteen index to measure environmental performance imposing weak disposal in the bad outputs that are jointly produced with the good outputs. Not entirely surprisingly, he reports some infeasibilities for this Hicks-Moorsteen environmental performance index.

made transitive (underscored by O'Donnell (2012b)) make it undoubtedly deserve greater attention. Transitivity allows for meaningful multi-lateral and multi-temporal instead of only binary comparisons. The reader is referred to O'Donnell (2012b) for more details on the economically-relevant axioms a suitable choice of base for the Hicks-Moorsteen index can yield. This assessment echoes the conclusion earlier made by Lovell (2003) in the same context.

3 Treatments for Unbalanced Panel Data and Critiques

One basic strategy found in the empirical literature employing these primal productivity indices consists in making the unbalanced panel somehow balanced. In fact, a variety of strategies can be discerned in the literature.

First, a straightforward strategy consists in simply dropping the observations that are not balanced. One example -already cited- is the article by Hollingsworth and Wildman (2003). Other examples of studies seemingly applying this strategy include Matthews and Zhang (2010) or Sturm and Williams (2004), among others.

Second, sometimes a kind of natural remedy is employed to make the unbalanced panel a balanced one. One example is the backward merger of units: units that merge at some point in time are also treated as merged for the years in the sample preceding the year of the merger. An example of a study adopting this remedy is Tortosa-Ausina, Grifell-Tatjé, Armero, and Conesa (2008).

Third, alternatively some authors resort to a more artificial remedy to make the initially unbalanced panel balanced. One example is the creation of artificial units in an effort to make the panel balanced. One study implementing such approach is Hongliang and Pollitt (2009).

Other strategies are more elaborate and involve some kind of partial balancing of the data set. For instance, one kind of intermediate solution found in the literature is to balance on a 2-years by 2-years basis. In such a setting, all firms present in each of the adjacent two-year comparison periods (the adjacent-year sample) are maintained (see, e.g., Cummins and Rubio-Misas (2006) for an empirical paper). More in general, one can note that some proposals to average these productivity indices over a variety of base periods are at least partially motivated by the desire to accommodate the case of unbalancedness in panel data (for instance, Asmild and Tam (2007)).

It is well-known that unbalancedness can occur due to delayed entry, early exit, or intermittent nonresponse. Another important distinction is that the lack of balance can be either planned (designed) as, for instance, in the case of rotating panels, or unplanned. In the latter unplanned case, non-responses are called missing data and these represent a potential source of bias. This is in particular the case in situations in which the mechanisms for missingness are related to the phenomenon being modelled (i.e., attrition bias). See Baltagi and Song (2006) or Frees (2004) for more details.

In the context of productivity measurement, attrition bias is a known issue (Van Beveren (2010) offers a survey of estimation issues) and it has regularly been reported in some parts of the literature (see, e.g., Foster, Haltiwanger, and Syverson (2008) for a recent example).[2] However, unbalancedness is in practice an unknown mix of unplanned and planned elements. Furthermore, the exact reason for the missing data (i.e., delayed entry, early exit, or intermittent nonresponse) is rarely known to the empirical researcher. If the exact reason for the missing data is known to the analyst, then it seems obvious that one should exploit this knowledge to measure the contribution of entering and exiting firms to productivity growth (see Griliches and Regev (1995) or more recently Diewert and Fox (2010)).

In general, it would seem useful to at least document the eventual impact of unbalancedness versus balancedness in productivity measurement. Only when the impact would be negligible, one could envision ignoring the issue. In the next section, we turn to this empirical exercise.

4 Data, Methodology, and Empirical Illustration

In this section, we first present the sample used for the empirical illustration. Then, we present the various technologies employed to compute the efficiency measures underlying both the Malmquist and Hicks-Moorsteen productivity indices. Thereafter, we provide the empirical results.

4.1 Data Description

The data base for this empirical part is a rather short unbalanced panel of three years (1984-1986) of French fruit producers based on annual accounting data collected in a survey (see Ivaldi, Ladoux, Ossard, and Simioni (1996) for details). Farms are selected on mainly two criteria: (i) the production of apples must be positive, and (ii) the acreage

[2] However, we are unaware of any article reporting attrition bias while employing the primal productivity indices analysed in this study. Byrnes (1991) is the only study we are aware of explicitly analysing selectivity bias in an efficiency context.

of the orchard must be at least five acres. As a description of technology, three aggregate inputs produce two aggregate outputs. The three inputs are: (i) capital (including land), (ii) labour, and (iii) materials. The two aggregate outputs are (i) the production of apples, and (ii) an aggregate of alternative products. For all three inputs, also input prices are available.

In total, 184 farms are available in the data base of which 130, 135 and 140 have records in 1984, 1985 and 1986, respectively. Thus, the unbalanced panel contains 405 observations in total. The balanced panel, containing only those farms for which records are available for all years, consists of only 92 farms. This yields an overall total of 276 observations. Thus, imposing balancedness amounts to eliminating about 32% of the information in the sample. Further summary statistics for all observations and details on the definitions of all variables are available in Appendix 2 in Ivaldi, Ladoux, Ossard, and Simioni (1996).

4.2 Specifications of Technologies for the Efficiency Computations

For the empirical application, we employ a variety of non-parametric technologies. In particular, we use both convex and non-convex technologies and both constant and variable returns to scale assumptions. Let K be the number of units. A unified algebraic presentation for a technology satisfying some combination of the above axioms is:

$$T^{\Lambda,\Gamma} = \Bigg\{ (x,y) \in \mathbb{R}_+^{n+p} : y_i \leq \sum_{k=1}^{K} \delta z_k y_{ki},\ (i = 1,\dots,p), \\ \sum_{k=1}^{K} \delta z_k x_{kj} \leq x_n,\ (j = 1,\dots,n), z \in \Lambda, \delta \in \Gamma \Bigg\},$$

where $\Lambda \in \{C, NC\}$, with $C = \{z \in \mathbb{R}_+^K : \sum_{k=1}^K z_k = 1\}$ and $NC = \{z \in \mathbb{R}_+^K : \sum_{k=1}^K z_k = 1 \text{ and } \forall k = 1,\dots,K : z_k \in \{0,1\}\}$, and where $\Gamma \in \{CRS, VRS\}$, with $CRS = \mathbb{R}_+$ and $VRS = \{1\}$.

From activity analysis, z is the vector of activity variables that indicates the intensity at which a particular activity is employed in constructing the reference technology by forming convex or non-convex combinations of observations constituting the best practice frontier (see Briec, Kerstens, and Vanden Eeckaut (2004)).

Axioms (T.1)-(T.4) are maintained in the non-convex case, while the convex case also imposes (T.5). In addition, both these technologies can impose constant returns to scale (T.6) rather than flexible returns to scale. This unified specification is non-linear, but it can be straightforwardly linearised in the convex case. For the non-convex case, it basically involves solving either some non-linear mixed integer programs, or some scaled vector dominance algorithms.

4.3 Empirical Results for the Primal Productivity Indices

Table 1 contains basic descriptive statistics for both the Malmquist and Hicks-Moorsteen productivity indices with the balanced and unbalanced panel data and using several technologies. This table is structured as follows: (i) The first four columns list the Malmquist, the last four columns report the Hicks-Moorsteen results. (ii) Within the latter distinction, the first two columns always contain the results for the unbalanced panel, while columns three and four each time display the balanced panel results. (iii) Horizontally, we first distinguish between convex and non-convex technologies. (iv) Then, we separately report both constant and variable returns to scale assumptions imposed on a given technology.

Computation of these descriptive statistics is performed over the productivity indices available. To give an example, the Malmquist index for the unbalanced panel results in valid results for 110 farms for the period 1984-85. Consequently, all corresponding descriptive statistics are computed for these 110 valid results. Obviously, due to a-priori removal of data, the number of valid results is 92 for the balanced panel, unless computational infeasibilities occur. E.g., in case of the specification $T^{C,VRS}$ only 89 valid computations are recorded for the periods 1984-85 and 1985-86 because of 3 such computational infeasibilities in each of these periods.

Several conclusions jump out. First, Malmquist and Hicks-Moorsteen productivity indices often seem to disagree on the nature of productivity change: while the Malmquist index points to productivity decline (except under the specification $T^{C,VRS}$), the Hicks-Moorsteen measures always productivity growth. Second, the descriptive statistics for both indices are different when comparing the balanced and the unbalanced cases (see infra). Third, these descriptive statistics seem rather robust across the several specifications of technology, again with the exception of the specification $T^{C,VRS}$.

Table 2 reports on the relative presence of infeasibilities due to unavailable data (denoted "na") and the computational infeasibilities (denoted "Inf"). Three conclusions emerge from studying this table. First, infeasibilities due to unavailable data amount to 50% in the balanced case, while these vary around 40% depending on the exact year in the unbalanced case. This amounts to a gain of about 10% in the amount of information included in the estimates. Second, despite this gain in the amount of information, the percentage of computational infeasibilities seems rather stable when comparing the balanced and the unbalanced cases. For the Malmquist index, the computational infeasibilities vary between 0.00% to 2.72% in both the unbalanced and the balanced cases depending on the technology specification. In the $T^{C,VRS}$ specification, the amount of computational infeasibilities remains stable at 1.63% for both periods and in both the balanced and unbalanced cases. No computational infeasibilities occur for the Malmquist index with the $T^{C,CRS}$ and $T^{NC,CRS}$ specifications. Third, the Hicks-Moorsteen index does not have a single computational infeasibility for all the technology and panel specifications over all periods. This is why it is not reported in Table 2.

To appreciate the observed differences in more detail we also plot kernel densities for a selection of productivity indices for a variety of frontier specifications. Figure 1 plots the densities for the balanced and unbalanced Malmquist index for the pair of years 1984-85 under $T^{C,CRS}$ and $T^{NC,CRS}$. For the same pair of years and technology specifications, Figure 2 plots the densities for the balanced and unbalanced Hicks-Moorsteen index. Note that to facilitate comparison, the densities with the balanced and unbalanced index are estimated with a common bandwidth for each technology specification. We ignore the plots of densities for the VRS assumption for reasons of space. In general, for a given index the densities for the balanced and unbalanced cases seem to resemble one another rather closely.

Table 5 formally tests for the differences between the densities of these productivity indices on both balanced and unbalanced samples with a test statistic proposed by Li (1996) (see also Fan and Ullah (1999) for a refinement) that is valid for both dependent and independent variables. Note that dependency is a characteristic for these frontier estimators (e.g., efficiency levels depend, among others, on sample size). The null hypothesis states the equality of both balanced and unbalanced distributions for a given productivity index and underlying specification of technology. The differences in densities between both balanced and unbalanced data sets turn out to be non-significant for this sample.

5 Conclusions

Using data on French fruit producers, this contribution is -to the best of our knowledge- the first to empirically illustrate the differences in between using either unbalanced or balanced panel data when computing frontier estimates for the primal Malmquist and Hicks-Moorsteen productivity indices. In particular, the main empirical results regarding the effect of balancing an unbalanced panel data is that in the balanced case one can loose substantial amounts of information (around 10% in our sample). Having documented the non-negligible impact of balancedness on productivity measurement, it is no longer an option to ignore this issue.

Obviously, there remain open challenges for future research. While inferential issues have been extensively studied when using parametric technology specifications estimated using unbalanced panel data, it remains somehow an open issue in the case of non-parametric specifications as employed in this study. When using unbalanced data, a key benefit is a larger sample. However, the technology per year depends on varying numbers of observations such that the precision of the estimates varies over the years. When balanced data is used, the drawback is a smaller sample, but at least the precision of the estimates does not vary over the years.[3]

References

Arora, H., and P. Arora (2012): "Bank Productivity Measurement using Hicks-Moorsteen Indices: Evidence from Indian Public Sector Banks," *International Journal of Business Performance Management*, 13(3), 386–407.

Asmild, M., and F. Tam (2007): "Estimating Global Frontier Shifts and Global Malmquist Indices," *Journal of Productivity Analysis*, 27(2), 137–148.

Atkinson, S., C. Cornwell, and O. Honerkamp (2003): "Measuring and Decomposing Productivity Change: Stochastic Distance Function Estimation versus Data Envelopment Analysis," *Journal of Business & Economic Statistics*, 21(2), 284–294.

Baltagi, B., and S. Song (2006): "Unbalanced Panel Data: A Survey," *Statistical Papers*, 47(4), 493–523.

Bjurek, H. (1996): "The Malmquist Total Factor Productivity Index," *Scandinavian Journal of Economics*, 98(2), 303–313.

Bjurek, H., F. Førsund, and L. Hjalmarsson (1998): "Malmquist Productivity Indices: An Empirical Investigation," in *Index Numbers: Essays in Honour of Sten Malmquist*, ed. by R. Färe, S. Grosskopf, and R. Russell, pp. 217–239. Kluwer, Boston.

Briec, W., and K. Kerstens (2009): "Infeasibilities and Directional Distance Functions: With Application to the Determinateness of the Luenberger Productivity Indicator," *Journal of Optimization Theory and Applications*, 141(1), 55–73.

——— (2011): "The Hicks-Moorsteen Productivity Index Satisfies the Determinateness Axiom," *Manchester School*, 79(4), 765–775.

[3]In fact, this problem has explicitly motivated some authors (e.g., Odeck (2008)) to estimate these primal productivity indices using balanced data even when an unbalanced panel were available.

Briec, W., K. Kerstens, and P. Vanden Eeckaut (2004): "Non-convex Technologies and Cost Functions: Definitions, Duality and Nonparametric Tests of Convexity," *Journal of Economics*, 81(2), 155–192.

Byrnes, P. (1991): "Estimation of Cost Frontiers in the Presence of Selectivity Bias: Ownership and Efficiency of Water Utilities," in *Advances in Econometrics (Volume 9)*, ed. by G. Rhodes, pp. 121–137. JAI, Greenwich.

Caves, D., L. Christensen, and W. Diewert (1982): "The Economic Theory of Index Numbers and the Measurement of Inputs, Outputs and Productivity," *Econometrica*, 50(6), 1393–1414.

Coelli, T. (1996): "A Guide to DEAP Version 2.1: A Data Envelopment Analysis (Computer) Program," Discussion Paper CEPA WP 96/08, Centre for Efficiency and Productivity Analysis (CEPA), University of Queensland, School of Economics.

Cummins, J., and M. Rubio-Misas (2006): "Deregulation, Consolidation, and Efficiency: Evidence from the Spanish Insurance Industry," *Journal of Money Credit and Banking*, 38(2), 323–356.

Diewert, W., and K. Fox (2010): "On Measuring the Contribution of Entering and Exiting Firms to Aggregate Productivity Growth," in *Price and Productivity Measurement: Volume 6 – Index Number Theory*, ed. by W. Diewert, B. Balk, D. Fixler, K. Fox, and A. Nakamura, pp. 41–66. Trafford Press, Bloomington.

Fan, Y., and A. Ullah (1999): "On Goodness-of-fit Tests for Weakly Dependent Processes Using Kernel Method," *Journal of Nonparametric Statistics*, 11(1), 337–360.

Färe, R., S. Grosskopf, M. Norris, and Z. Zhang (1994): "Productivity Growth, Technical Progress, and Efficiency Change in Industrialized Countries," *American Economic Review*, 84(1), 66–83.

Färe, R., S. Grosskopf, and P. Roos (1996): "On Two Definitions of Productivity," *Economics Letters*, 53(3), 269–274.

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

Foster, L., J. Haltiwanger, and C. Syverson (2008): "Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability?," *American Economic Review*, 98(1), 394–425.

Frees, E. (2004): *Longitudinal and Panel Data: Analysis and Applications in the Social Sciences*. Cambridge University Press, Cambridge.

Grifell-Tatjé, E., and C. Lovell (1995): "A Note on the Malmquist Productivity Index," *Economics Letters*, 47(2), 169–175.

Griliches, Z., and H. Regev (1995): "Firm Productivity in Israeli Industry 1979–1988," *Journal of Econometrics*, 65(1), 175–203.

Grosskopf, S. (2003): "Some Remarks on Productivity and its Decompositions," *Journal of Productivity Analysis*, 20(3), 459–474.

Hollingsworth, B., and J. Wildman (2003): "The Efficiency of Health Production: Re-Estimating the WHO Panel Data Using Parametric and Non-Parametric Approaches to Provide Additional Information," *Health Economics*, 12(6), 493–504.

Hongliang, Y., and M. Pollitt (2009): "Incorporating both Undesirable Outputs and Uncontrollable Variables into DEA: The Performance of Chinese Coal-Fired Power Plants," *European Journal of Operational Research*, 197(3), 1095–1105.

Hulten, C. (2001): "Total Factor Productivity: A Short Biography," in *New Developments in Productivity Analysis*, ed. by C. Hulten, E. Dean, and M. Harper, pp. 1–47. University of Chicago Press, Chicago.

Ivaldi, M., N. Ladoux, H. Ossard, and M. Simioni (1996): "Comparing Fourier and Translog Specifications of Multiproduct Technology: Evidence from an Incomplete Panel of French Farmers," *Journal of Applied Econometrics*, 11(6), 649–668.

Li, Q. (1996): "Nonparametric Testing of Closeness between Two Unknown Distribution Functions," *Econometric Reviews*, 15(1), 261–274.

Lovell, C. (2003): "The Decomposition of Malmquist Productivity Indexes," *Journal of Productivity Analysis*, 20(3), 437–458.

Matthews, K., and N. X. Zhang (2010): "Bank Productivity in China 1997–2007: Measurement and Convergence," *China Economic Review*, 21(4), 617–628.

Nemoto, J., and M. Goto (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(4), 617–634.

Nishimizu, M., and J. Page (1982): "Total Factor Productivity Growth, Technological Progress and Technical Efficiency Change: Dimensions of Productivity Change in Yugoslavia, 1965-78," *Economic Journal*, 92(368), 920–936.

Odeck, J. (2008): "How Efficient and Productive Are Road Toll Companies? Evidence from Norway," *Transport Policy*, 15(4), 232–241.

O'Donnell, C. (2010): "Measuring and Decomposing Agricultural Productivity and Profitability Change," *Australian Journal of Agricultural and Resource Economics*, 54(4), 527–560.

——— (2012a): "An Aggregate Quantity-Price Framework for Measuring and Decomposing Productivity and Profitability Change," *Journal of Productivity Analyis*, p. forthcoming.

——— (2012b): "Nonparametric Estimates of the Components of Productivity and Profitability Change in U.S. Agriculture," *American Journal of Agricultural Economics*, p. forthcoming.

Sturm, J.-E., and B. Williams (2004): "Foreign Bank Entry, Deregulation and Bank Efficiency: Lessons from the Australian Experience," *Journal of Banking & Finance*, 28(7), 1775–1799.

Tortosa-Ausina, E., E. Grifell-Tatjé, C. Armero, and D. Conesa (2008): "Sensitivity Analysis of Efficiency and Malmquist Productivity Indices: An Application to Spanish Savings Banks," *European Journal of Operational Research*, 184(3), 1062–1084.

Tsekouras, K., C. Pantzios, and G. Karagiannis (2004): "Malmquist Productivity Index Estimation with Zero-Value Variables: The Case of Greek Prefectural Training Councils," *International Journal of Production Economics*, 89(1), 95–106.

Van Beveren, I. (2010): "Total Factor Productivity Estimation: A Practical Review," *Journal of Economic Surveys*, 26(1), 98–128.

Zaim, O. (2004): "Measuring Environmental Performance of State Manufacturing through Changes in Pollution Intensities: A DEA Framework," *Ecological Economics*, 48(1), 37–47.

		Malmquist				Hicks-Moorsteen			
		Unbalanced		Balanced		Unbalanced		Balanced	
		1984-85	1985-86	1984-85	1985-86	1984-85	1985-86	1984-85	1985-86
$T^{C,CRS}$	n	110	111	92	92	110	111	92	92
	Average	1.1368	1.2213	1.1181	1.2297	1.1793	1.0965	1.1934	1.1070
	Stand. Dev.	0.5439	0.7576	0.5222	0.7978	1.1556	0.7540	1.2234	0.7938
	Min	0.0855	0.1435	0.0854	0.1435	0.3418	0.1913	0.3264	0.1919
	Max	2.9785	5.2625	3.0830	5.2365	11.6472	6.8672	11.6385	6.8672
$T^{C,VRS}$	n	107	108	89	89	110	111	92	92
	Average	1.1369	0.9536	1.1210	0.9432	1.1500	1.1675	1.1611	1.1812
	Stand. Dev.	0.2683	0.3004	0.2811	0.2965	1.1368	0.7237	1.1999	0.7772
	Min	0.5282	0.5583	0.5280	0.5518	0.3318	0.1893	0.3472	0.1925
	Max	1.9068	2.6097	1.9512	2.5140	11.5799	6.4165	11.5746	6.4861
$T^{NC,CRS}$	n	110	111	92	92	110	111	92	92
	Average	1.1429	1.1605	1.1289	1.1707	1.1300	1.1003	1.1359	1.0955
	Stand. Dev.	0.5883	0.5870	0.5711	0.6014	0.8003	0.7421	0.8191	0.7836
	Min	0.1471	0.1252	0.1428	0.1305	0.2507	0.2625	0.2507	0.2829
	Max	4.3777	3.5131	4.3777	3.5803	6.7987	7.0934	7.0040	7.0399
$T^{NC,VRS}$	n	105	107	87	87	110	111	92	92
	Average	1.1116	1.0025	1.1101	1.0161	1.0992	1.1402	1.0995	1.1215
	Stand. Dev.	0.3326	0.3015	0.3579	0.3544	0.6649	0.6579	0.6649	0.6714
	Min	0.4652	0.5062	0.4210	0.4974	0.4015	0.1668	0.4022	0.1668
	Max	2.8566	2.0174	2.8566	2.3650	5.3857	5.7232	5.2018	5.7377

Table 1: Descriptive Statistics for Malmquist and Hicks-Moorsteen Productivity Indices under Various Specifications

		Unbalanced			Balanced		
		1984-85	1985-86	Overall	1984-85	1985-86	Overall
	% na	40.22	39.67	39.95	50.00	50.00	50.00
Malmquist							
$T^{C,CRS}$	% Inf	0.00	0.00	0.00	0.00	0.00	0.00
$T^{C,VRS}$	% Inf	1.63	1.63	1.63	1.63	1.63	1.63
$T^{NC,CRS}$	% Inf	0.00	0.00	0.00	0.00	0.00	0.00
$T^{NC,VRS}$	% Inf	2.72	2.17	2.45	2.72	2.72	2.72

Table 2: Malmquist and Hicks-Moorsteen Productivity Indices under Various Specifications: Non-Availabilities ("na") and Infeasibilities ("Inf")

		Malmquist		Hicks-Moorsteen	
		1984-85	1985-86	1984-85	1985-86
$T^{C,CRS}$	z-value	−1.1264	−0.9661	−1.0471	−0.9330
	p-value	0.1300	0.1670	0.1475	0.1754
$T^{C,VRS}$	z-value	−1.0497	−0.8951	−1.0001	−0.9975
	p-value	0.1469	0.1854	0.1586	0.1593
$T^{NC,CRS}$	z-value	−1.0439	−1.0173	−0.9673	−0.8615
	p-value	0.1483	0.1545	0.1667	0.1945
$T^{NC,VRS}$	z-value	−0.8854	−0.7358	−0.9764	−0.8456
	p-value	0.1880	0.2309	0.1644	0.1989

Table 3: Li-test Results of Density Comparison between Balanced and Unbalanced Malmquist and Hicks-Moorsteen Productivity Indices under Various Specifications

Figure 1: Kernel Density of Balanced and Unbalanced Malmquist Index (1984-85) under $T^{C,CRS}$ and $T^{NC,CRS}$

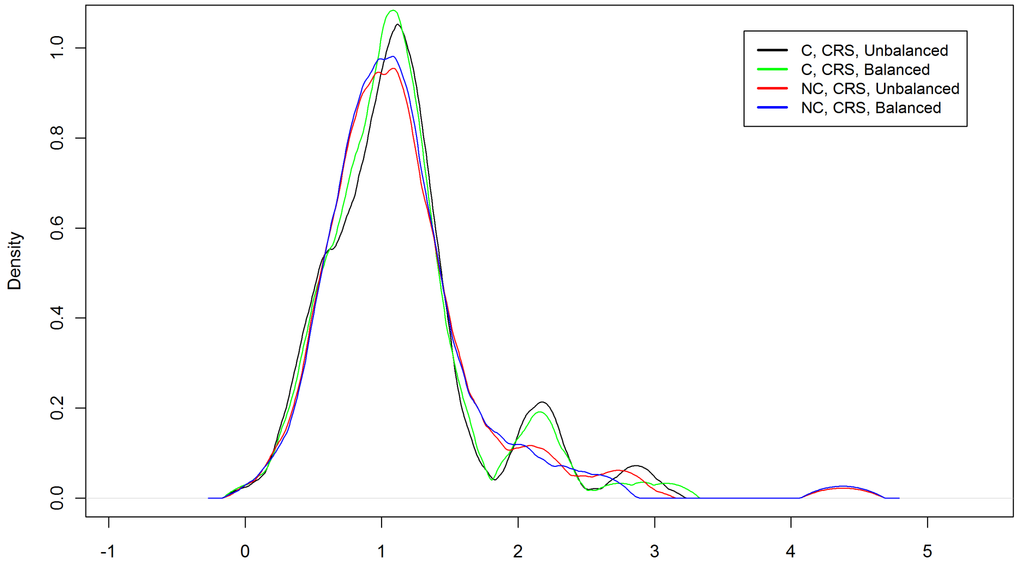
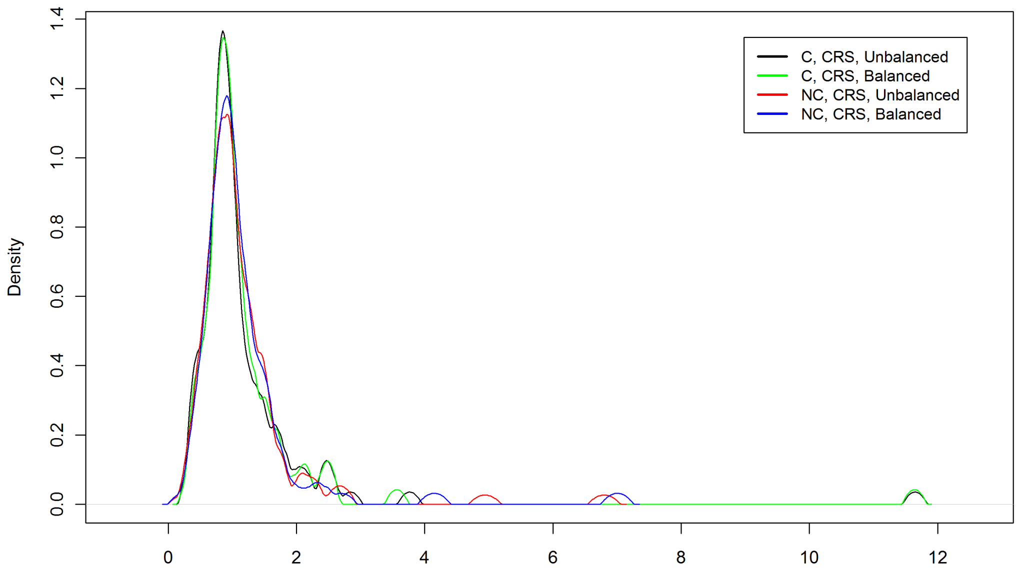


Figure 2: Kernel Density of Balanced and Unbalanced Hicks-Moorsteen Index (1984-85) under $T^{C,CRS}$ and $T^{NC,CRS}$



Total Factor Productivity Growth in the United States Farm Sector: 1948-2009[1]

Eldon Ball and Sun Ling Wang

Economic Research Service

US Department of Agriculture

I. Introduction

The rise in agricultural productivity has long been chronicled as the single most important source of economic growth in the U.S. farm sector. Though their methods differ in important ways, the major sectoral productivity studies by Kendrick and Grossman (1980), Jorgenson, Gollop, and Fraumeni (1987), and Jorgenson and Gollop (1992) share this common conclusion. In a more recent study, Jorgenson, Ho, and Stiroh (2005) find that productivity growth in agriculture averaged 1.9 percent per year over the 1977-2000 period. Output grew at a 3.4 percent average annual rate over this period. Thus productivity growth accounted for almost 80 percent of the growth in output in the farm sector. Only three of the forty-four sectors covered by the Jorgenson, Ho, and Stiroh (2005) study achieved higher rates of productivity growth than did agriculture.

The U.S. Department of Agriculture (USDA) has been monitoring the sector's productivity performance for decades. In fact, the USDA in 1960 was the first agency to introduce a multifactor productivity measure into the Federal statistical program. Today, the Department's Economic Research Service (ERS) routinely publishes total factor productivity (TFP) measures based on a sophisticated system of production accounts. The official TFP statistics are based on

[1] The views expressed herein are those of the authors, and not necessarily those of the U.S. Department of Agriculture.

the translog transformation frontier. The translog model relates the growth rates of multiple outputs to the cost-share weighted growth rates of labor, capital, and intermediate goods. [2]

The applied USDA model is quite detailed. The changing demographic character of the agricultural labor force is used to construct a quality-adjusted index of labor input.[3] Similarly, much asset specific detail underlies the measure of capital input. For example, aggregation across different parcels of land is at the county level. The contributions of feed and seed, energy, and agricultural chemicals to output growth are captured in the index of intermediate inputs. An important innovation is the use of hedonic price indexes in constructing measures of fertilizers and pesticides consumption. Also included in intermediate input are a number of purchased services. The result is a time series of productivity indexes now spanning the years 1948 to 2009

II. Total Factor Productivity Indexes

The measured rates of productivity growth reported by the USDA are formed from Törnqvist indexes of outputs and inputs. A sector's total factor productivity (TFP) growth over some period is defined as:

(1) $$\ln\left[\frac{TFP_t}{TFP_{t-1}}\right] = \sum\left[\frac{R_{it} + R_{i,t-1}}{2}\right]\ln\left[\frac{Y_{it}}{Y_{it-1}}\right] - \sum\left[\frac{W_{jt} + W_{j,t-1}}{2}\right]\ln\left[\frac{X_{jt}}{X_{jt-1}}\right]$$

where the Y_i are output indexes, the X_j are input indexes, the R_i are output revenue shares, and the W_j are input cost shares.

[2] A complete description of the USDA model can be found in Ball et al. (1999).

[3] See Jorgenson and Griliches (1967) for a discussion of input quality.

The above expression can be derived from a homogenous translog transformation function. The translog function itself can provide a second-order approximation to an arbitrary linearly homogenous function.

III. The Production Accounts

The USDA's Economic Research Service has constructed production accounts consistent with a gross output model of production. Output is defined as gross production leaving the farm as opposed to real value added. Inputs are not limited to labor and capital but include intermediate inputs as well. The text in this section provides an overview of the sources and methods used to build the annual production accounts for the 1948-2009 period.

Output. The development of a measure of output begins with disaggregated data for physical quantities and market prices of crops and livestock. The output quantity for each crop and livestock category includes the quantities of commodities sold off the farm, additions to inventory, and quantities consumed in farm households as part of final demand during the calendar year. The price corresponding to each disaggregated output reflects the value of that output to the producer; that is, subsidies are added and indirect taxes are subtracted from market values.

One unconventional aspect of our measure of total output is the inclusion of goods and services from certain 'non-agricultural' or secondary activities. These activities are defined as activities closely linked to agricultural production for which information on output and input use cannot be separately observed. Two types of secondary activities are distinguished. The first represents a continuation of the agricultural activity, such as the processing and packaging of agricultural products on the farm, while services relating to agricultural production, such as machine services for hire, are typical of the second. The index of total output reported in table 1 is formed by aggregating over agricultural goods and the output of goods and services from secondary activities.

Intermediate Input. Intermediate input consists of goods used in production during the calendar year whether withdrawn from beginning inventories or purchased from outside the farm sector. Open-market purchases of feed, seed, and livestock inputs enter the intermediate goods accounts. Withdrawals from producers' inventories are also measured in output, intermediate input, and capital input. Beginning inventories of crops and livestock represent capital inputs and are treated in the discussion of capital below. Additions to these inventories represent deliveries to final demand and are treated as part of output. Goods withdrawn from inventory are symmetrically defined as intermediate goods and recorded in the farm input accounts.

Data on current dollar consumption of petroleum fuels, natural gas, and electricity are compiled by the Economic Research Service. Prices of individual fuels are taken from the Energy Information Administration's Monthly Energy Review. The index of energy consumption is formed implicitly as the ratio of total expenditures (less State and Federal excise tax refunds) to the corresponding price index.

Pesticides and fertilizers are important intermediate inputs but their data require adjustment since these inputs have undergone significant changes in input quality over the 1948-2009 period. Since input price and quantity series used in a study of productivity must be denominated in constant-efficiency units, we construct price indexes for fertilizers and pesticides using hedonic methods. Under this approach, a good or service is viewed as a bundle of characteristics which contribute to the productivity derived from its use. Its price represents the valuation of the characteristics "that are bundled in it", and each characteristic is valued by its "implicit" price (Rosen, 1974). However, these prices are not observed directly and must be estimated from the hedonic price function.

A hedonic price function expresses the price of a good or service as a function of the quantities of the characteristics it embodies. Thus, the hedonic price function for, say pesticides, may be expressed as $w = W(X, D)$, where w represents the price of a particular pesticide or fertilizer, X is a vector of characteristics or quality variables and D is a vector of other variables. Kellogg et al. (2002) have compiled data on characteristics that capture differences in pesticide quality. These characteristics include toxicity, persistence in the environment, and leaching potential, among others.

Other variables (denoted by D) are also included in the hedonic equation, and their selection depends not only on the underlying theory but also on the objectives of the study. If the main objective of the study is to obtain price indexes adjusted for quality, as in our case, the only variables that should be included in D are time dummy variables, which will capture all price effects other than quality. After allowing for differences in the levels of the characteristics, the part of the price difference not accounted for by the included characteristics will be reflected in the time dummy coefficients.

Economic theory places few if any restrictions on the functional form of the hedonic price function. We adopt a generalized linear form, where the dependent variable and each of the continuous independent variables is represented by the Box-Cox transformation. This is a mathematical expression that assumes a different functional form depending on the transformation parameter, and which can assume both linear and logarithmic forms, as well as intermediate non-linear functional forms.

Thus the general functional form of our model is given by:

$$w(\lambda_0) = \sum_{n=1}^{N} \alpha_n X_n(\lambda_n) + \sum_{m=1}^{M} \gamma_m D_m + \varepsilon, \qquad (2)$$

where $w(\lambda_0)$ is the Box-Cox transformation of the dependent price variable

$$w(\lambda_0) = \begin{cases} \dfrac{w^{\lambda_0} - 1}{\lambda_0}, \lambda_0 \neq 0, \\ \ln w, \lambda_0 = 0. \end{cases} \qquad (3)$$

Similarly, $X_n(\lambda_n)$ is the Box-Cox transformation of the continuous quality variable X_n where $X_n(\lambda_n) = (X_n^{\lambda_n} - 1)/\lambda_n$ if $\lambda_n \neq 0$ and $X_n(\lambda_n) = \ln X_n$ if $\lambda_n = 0$. Variables represented by *D* are time dummy variables, not subject to transformation, α and γ are unknown parameter vectors, and ε is a stochastic disturbance.

Finally, price and implicit quantity indexes are calculated for a range of purchased services, such as contract labor services, custom machine services, and machine and building repairs and maintenance. Contract labor services are becoming increasingly important in agricultural production. Since farmers contract with labor brokers to assemble crews, there is little data on hours worked. Only data on nominal expenditures for contract labor are collected. In order to account for the contribution of contract labor services to agricultural production we must construct an appropriate deflator. Given that the compensation of contract workers will likely vary with differences in demographic characteristics such as age, experience, gender, and education, we construct a deflator for contract labor using hedonic methods based on data from the National Agricultural Workers Survey (NAWS). The general form for the hedonic model is given in equation (2) above. In the case of contract labor, however, the dependent variable is the wage rate per hour for the contract labor service. The characteristic variables X_n include farm work experience, education, gender, age, among others.[4]

A Törnqvist index of total intermediate input is formed by weighting the growth rates of each category of intermediate input described above by their value shares in the overall value of intermediate inputs.

Labor Input. The USDA labor accounts incorporate the demographic cross-classification of the agricultural labor force developed by Jorgenson, Gollop, and Fraumeni (1987). Matrices of hours worked and compensation per hour have been developed for laborers cross-classified by sex, age, education, and employment class—employee versus self-employed and unpaid family workers.

Labor compensation data for self-employed and unpaid family workers are not observed. As a result, self-employed and unpaid family workers are imputed the mean wage earned by hired workers with the same demographic characteristics. Indexes of labor input are constructed using the demographically cross-classified hours and compensation data. Under the Törnqvist approach, labor hours having higher marginal productivity (wages) are given higher weights in forming the index of labor input than are hours having lower marginal productivities. Doing so explicitly adjusts indexes of labor input for quality change in labor hours.

---

[4] See Wang et al. (2011) for a detailed description of the characteristics included in the hedonic model.

Capital Input. The measurement of productivity growth requires time series measures of capital input and service prices. Construction of these series begins with estimating the capital stock and the rental price for each asset type. For depreciable assets, the perpetual inventory method is used to develop capital stocks from data on investment. The stocks of land and inventories are measured as implicit quantities derived from balance sheet data. Implicit rental prices for each asset are based on the correspondence between the purchase price of the asset and the discounted value of future service flows derived from that asset.

***Depreciable assets*.** The perpetual inventory method cumulates investment data measured in constant prices into a measure of capital stock. Current dollar investment data for each depreciable asset are obtained from the USDA's Agriculture Resource Management survey. The Bureau of Labor Statistics producer price indexes for passenger cars, motor trucks, wheel-type farm tractors, and agricultural machinery excluding tractors are employed as investment deflators. For non-residential structures, the implicit price deflator is taken from the U.S. national income and product accounts.

Under the perpetual inventory method, capital stock at the end of each period, say K_t , is measured as the sum of all past investments, each weighted by its relative efficiency d_τ:

(4)

where   is approximated by hyperbolic efficiency function,

(5)
$$d_\tau = (L - \tau)/(L - \beta\tau), 0 \leq \tau \leq L$$
$$d_\tau = 0, \tau \geq L,$$

L is the service life of the asset, τ represents the asset's age, and ß is a curvature or decay parameter.[5]

Little empirical evidence is available to suggest a precise value for ß. However, two studies (Penson, Hughes and Nelson, 1977; Romain, Penson and Lambert, 1987) provide evidence that efficiency decay occurs more rapidly in the later years of service, corresponding to a value of ß in the 0 to 1 interval. For purposes of this study, it is assumed that the efficiency of a structure declines slowly over most of its service life until a point is reached where the cost of repairs exceeds the increased service flows derived from the repairs, at which point the structure is allowed to depreciate rapidly (ß=0.75). The decay parameter for durable equipment (ß=0.5) assumes that the decline in efficiency is more uniformly distributed over the asset's service life.

The other critical variable in the efficiency function (5) is the asset lifetime L. For each asset type, there exits some mean service life L around which there exists some distribution of actual service lives. It is assumed that the underlying distribution is the normal distribution truncated at points two standard deviations above and below the mean service life.

Rental rates. Firms will add to the capital stock as long as the present value of the net revenue generated by an addition unit of capital exceeds the purchase price of the asset. This can be stated algebraically as:

(6) $$\sum_{t=1}^{\infty}\left(p\frac{\partial y}{\partial K} - w_K\frac{\partial R_t}{\partial K}\right)(1+r)^{-t} > w_K$$

[5] The value of ß is restricted only to values less than or equal to one. For values of ß greater than zero, the efficiency of the asset approaches zero at an increasing rate. For values less than zero, efficiency approaches zero at a decreasing rate.

where p is the price of output, w_K is the price of investment goods, and r is the real discount rate. To maximize net present value, firms will continue to add to capital stock until this equation holds as an equality:

(7) $$p\frac{\partial y}{\partial K} = rw_K + r\sum_{t=1}^{\infty} w_K \frac{\partial R_t}{\partial K}(1+r)^{-t} = c$$

where c is the implicit rental price of capital.

The rental price consists of two components. The first term, rw_K, represents the opportunity cost associated with the initial investment. The second term $r\sum_{t=1}^{\infty} w_K \frac{\partial R_t}{\partial K}(1+r)^{-t}$ is the present value of the cost of all future replacements required to maintain the productive capacity of the capital stock. Let F denote the present value of the stream of capacity depreciation on one unit of capital according to the mortality distribution m:

(8) $$F = \sum_{t=1}^{\infty} m_r(1+r)^{-r}$$

where $m_r = -(d_r - d_{r-1}), \tau = 1, \dots, t.$

Since replacement at time t is equal to capacity depreciation at time t:

(9) $$\sum_{t=1}^{\infty} \frac{\partial R_t}{\partial K}(1+r)^{-t} = \sum_{t=1}^{\infty} F^t = \frac{F}{(1-F)}$$

and

(10) $$c = \frac{rw_K}{1-F}$$

The real rate of return r in the above expression is calculated as the nominal yield on investment grade corporate bonds less the rate of inflation as measured by the implicit deflator for gross domestic product. An *ex ante* rate is then obtained by expressing observed real rates as an ARIMA process. We then calculate *F* holding the required real rate of return constant for that vintage of capital goods. In this way, implicit rental prices c are calculated for each asset type.

Land input. To obtain a constant-quality stock of land stock, we first construct intertemporal price indexes of land in farms. The stock of land is then constructed implicitly as the ratio of the value of land in farms to the intertemporal price index. We assume that land in each county is homogeneous, hence aggregation is at the county level.

Inventories. Beginning inventories of crops and livestock are treated as capital inputs. The number of animals on farms is available from annual surveys, as are the stocks of grains and oilseeds. December average prices are used to value commodities held in inventory.

Indexes of capital input are formed by aggregating over the various capital assets using as weights the cost shares based on asset-specific rental prices. Service prices for capital input are formed implicitly as the ratio of the total current dollar value of capital service flows to the quantity index. As is the case for labor input, the resulting measure of capital input is adjusted for changes in asset quality.

IV. Productivity Growth

Input growth typically has been the dominant source of economic growth for the aggregate economy and for each of its producing sectors. Jorgenson, Gollop, and Fraumeni (1987) find this to be the case for the aggregate economy in every subperiod over the period 1948-79. Denison (1979) draws a similar conclusion over the longer interval 1926-76. In their sectoral analyses, Jorgenson, Gollop, and Fraumeni (1987) find that output growth relies most heavily on input growth in forty-two of forty-seven private business sectors in the 1948-79 period, and in a more aggregated study (Jorgenson and Gollop, 1992) through 1985 in eight of nine sectors.

Agriculture turns out to be one of the few exceptions. Productivity growth dominates input growth. This is confirmed in table 2 which reports the source decomposition of output growth in the farm sector for the full 1948-2009 period and twelve subperiods. Applying equation (1), output growth equals the sum of contributions of labor, capital, land, and materials inputs and TFP growth (The contribution of each input equals the product of the inputs growth rate and its respective share in total cost.).

The singularly important role of productivity growth in agriculture is made all the more remarkable by the dramatic contraction in labor input in the sector, a pattern that persists through every subperiod. Over the full 1948-2009 period, labor input declined at an annual rate of 2.51%, a rate unmatched by any of the 50 nonfarm sectors evaluated by Jorgenson, Gollop, and Fraumeni (1987). When weighted by its 0.20 share in total cost, the contraction in labor input contributes an average -0.52 percentage points to output growth.

Capital input (excluding land) exhibits a different pattern. Its contribution to output growth increased markedly through the early 1980s, but declined following the spike in real rates. Producers increasingly substituted purchased machine services for own capital. On average, however, capital input expanded over the full period. Its positive growth contributes an annual 0.02 percentage points to output growth.

Land input declined at a -0.55% average annual rate; its contribution to output growth averaged -0.08 percentage points per year.

Material input's contribution, as reported in table 2, oscillated between positive and negative values over the sample period but averaged a substantial positive rate of 0.69% per year. This positive contribution was sufficient to outweigh the negative contributions through labor and land. Still, the net contribution of all three inputs was only 0.11 percentage points per year, leaving responsibility for the 1.63% average annual rate growth in farm sector output almost entirely to productivity growth.

V. Concluding Remarks

Productivity growth is the single most important source of economic growth in the U.S. farm sector. The major sectoral productivity studies share this common conclusion (Kendrick and Grossman, 1980; Ball, 1985; Jorgenson, Gollop, and Fraumeni, 1987; Jorgenson and Gollop, 1992; Ball et al., 1997, 1999; Jorgenson, Ho, and Stiroh, 2005). In this paper, we provide further evidence in support of what appears a consensus point of view. More specifically, we provide an overview of the methods and data used by the USDA's Economic Research Service to measure the growth in farm sector output and the contributions of input growth and growth in total factor productivity. The official USDA statistics show that output grew at an average annual rate of 1.63% over the period 1948 to 2009. The net contribution of capital, labor, and materials inputs was a modest 0.11 percentage points per year, while growth in total factor productivity added an annual 1.52 percentage points to output growth. Thus productivity growth accounted for 93% of the growth in farm sector output.

References

Ball, V. E., F. Gollop, A. Kelly-Hawke, and G. Swinand. "Patterns of Productivity Growth in the U.S. Farm Sector: Linking State and Aggregate Models," *American Journal of Agricultural Economics* 81 (February 1999): 164-179.

Ball, V. E., J-C Bureau, R. Nehring, and A. Somwaru. "Agricultural Productivity Revisited," *American Journal of Agricultural Economics* 79 November 1997): 1045-63.

Ball, V. E. "Output, Input, and Productivity Measurement in U.S. Agriculture, 1948-79," American Journal of Agricultural Economics 6 (August 1985): 475-86.

Jorgenson, D., M. Ho, and K. Stiroh. *Productivity: Information Technology and the American Growth Resurgence*. Cambridge: The MIT Press, 2005.

Jorgenson, D. and F. Gollop. "Productivity Growth in U.S. Agriculture: A Postwar Perspective," *American Journal of Agricultural Economics* 74 (August 1992): 745-50.

Jorgenson, D., F. Gollop, and B. Fraumeni. *Productivity and U.S. Economic Growth*. Cambridge MA: Harvard University Press, 1987.

Kellogg, R., R. Nehring, A. Grube, D. Goss, and S. Plotkin. "Environmental Indicators of Pesticide Leaching and Runoff from Farm Fields," in V.E. Ball and G. Norton, eds. *Agricultural Productivity: Measurement and Sources of Growth*, Boston: Kluwer Academic Publishers, 2002.

Kendrick, J.W., and E.S. Grossman. *Productivity in the United States, Trends and Cycles*. Baltimore MD: The Johns Hopkins University Press, 1980.

Penson, J., D. Hughes, and G. Nelson. “Measurement of Capacity Depreciation Based on Engineering Data,” *American Journal of Agricultural Economics* 59 (May 1977): 321-29.

Romain, R., J. Penson, and R. Lambert. “Prices and Investment for Farm Tractors in Canada,” *Canadian Journal of Agricultural Economics* 35 (July 1987): 373-78.

Rosen, S. “Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition,” Journal of Political Economy 82(1974): 34-55.

Table 1. Average annual rates of growth (percent), 1948-2009 (2005=1)

Period	Total output	Total farm input	Capital	Labor	Intermediate inputs					TFP
					All	Farm origin	Energy	Agricultural chemicals	Purchased services	
1948-2009	1.63	0.11	-0.21	-2.51	1.43	1.15	0.85	2.54	1.15	1.52
1948-1953	1.18	1.34	1.75	-3.34	3.72	2.23	4.61	2.87	2.40	-0.16
1953-1957	0.96	0.28	-0.10	-4.58	2.86	3.73	0.15	1.35	1.88	0.68
1957-1960	4.03	0.50	-0.43	-3.74	2.95	2.58	0.06	5.95	5.42	3.53
1960-1966	1.21	0.05	0.05	-3.75	1.72	1.73	1.65	5.54	-0.67	1.16
1966-1969	2.24	-0.08	0.34	-2.78	0.88	2.13	0.43	-2.99	-0.75	2.32
1969-1973	2.65	0.46	-0.44	-1.84	2.02	1.70	-0.42	8.22	0.40	2.19
1973-1979	2.26	1.64	1.04	-1.06	2.76	1.83	4.09	3.29	4.84	0.62
1979-1981	1.53	-1.85	0.39	-1.39	-3.05	-2.68	-3.35	2.54	-7.50	3.39
1981-1990	0.96	-1.22	-2.16	-2.79	-0.13	-0.19	-1.59	-0.75	0.39	2.19
1990-2000	1.84	0.31	-0.75	-1.63	1.64	1.20	0.96	2.85	2.38	1.53
2000-2007	0.77	0.14	-0.16	-1.56	0.88	0.48	-0.74	1.55	1.18	0.63
2007-2009	1.88	-1.80	0.88	-3.69	-2.41	-3.27	5.24	1.32	-5.19	3.68

Note: The subperiods are measured from cyclical peak to peak in aggregate economic activity as defined by the National Bureau of Economic Research (see http://www.nber.org/cycles.html).

Table 2. Sources of Growth in the U.S. Farm Sector, 1948-2009 (average annual growth rates in percent)

	1948-2009	1948-1953	1953-1957	1957-1960	1960-1966	1966-1969	1969-1973	1973-1979	1979-1981	1981-1990	1990-2000	2000-2007	2007-2009
Output growth	1.63	1.18	0.96	4.03	1.21	2.24	2.65	2.26	1.54	0.96	1.84	0.77	1.88
Sources of growth													
Input growth	0.11	1.34	0.28	0.50	0.05	-0.08	0.46	1.64	-1.85	-1.22	0.31	0.14	-1.80
Labor	-0.52	-0.81	-1.08	-0.83	-0.81	-0.61	-0.38	-0.19	-0.22	-0.43	-0.34	-0.35	-0.64
Capital	0.02	0.54	0.15	0.03	0.08	0.32	0.14	0.32	0.23	-0.61	-0.21	0.05	0.35
Land	-0.08	0.02	-0.17	-0.16	-0.07	-0.22	-0.29	0.00	-0.12	-0.09	0.00	-0.08	-0.12
Materials	0.69	1.58	1.38	1.45	0.85	0.43	0.99	1.50	-1.74	-0.09	0.87	0.52	-1.39
Total factor productivity	1.52	-0.16	0.68	3.53	1.16	2.32	2.19	0.62	3.39	2.19	1.53	0.63	3.68

Note: The subperiods are measured from cyclical peak to peak in aggregate economic activity.

Productivity Change over Time and the Dynamics of Cost Competitiveness

Subhash C Ray

University of Connecticut

Storrs CT 06269-1063

subhash.ray@uconn.edu

Abstract: Firms within a state, states within a country, and countries across the world are continuously striving to enhance their competitiveness in the present age of globalization. This paper defines competitiveness of a production unit as the relative cost of production per unit of output. Basic concepts from neoclassical production economics are used to provide a detailed decomposition of cost competitiveness of a firm relative to a rival. It is also shown how changes in efficiency and relative input prices along with technical change affect how cost competitiveness of a firm evolves over time. State level data from the U.S. Census of Manufacturers from the years 1992, 1997, 2002, and 2007 are used in a empirical application of the proposed methodology using Data Envelopment Analysis.

This paper has benefited from comments from the participants at the National Marine Fisheries Service Productivity Workshop held in Santa Cruz, CA on June 11-12, 2012. Responsbility for any errors rests on the author.

July 12, 2012

Productivity Change over Time and the Dynamics of Cost Competitiveness

Subhash C Ray
University of Connecticut
Storrs CT 06269-1063
subhash.ray@uconn.edu

In the present age of globalization, and especially after the Great Recession, countries all over the world are engaged in a 'race to the bottom' endeavoring to outperform others in terms of lowering their production costs in an effort to capture a bigger share of the global market. While enhanced *competitiveness* has become the holy grail of public policy and features frequently in political discourse, calibration of competitiveness for comparison across regions or over time has remained rather vague. In fact, more often than not, it is not clear enough what the different regions are assumed to be competing for.

The World Economic Forum, in its Global Competitiveness Report (2011), defines competitiveness as

> "*the set of institutions, policies, and factors that determine the level of productivity of a country*" (WEF 2011, p 4).

The Report goes on to emphasize

> "*The productivity level also determines the rate of return obtained by investments in an economy, which in turn are the fundamental drivers of its growth rate*". (WEF 2011, p 4).

In a sub-national context, a wider price-cost margin can be seen to be a competitive advantage of one region relative to another one. In particular, if the output price is uniform across states in a nationally integrated product market, lower average cost makes a state more attractive to entrepreneurs thereby making it more competitive. Even at the sub-national level, states and provinces within the country try to remain cost-competitive. In U.S. manufacturing, for example, there has been considerable migration of industries from the high cost Rustbelt states to the lower cost Sunbelt states. As unit costs change over time, firms (or states) move up or down the list in the relative competitiveness chart.

Differences in average cost between two firms at a given point in time (or for the same firm at two different points in time) arise primarily out of two sources: (i) differences in input prices, and (ii) productivity differences. In an inter temporal context, the productivity change component can be further decomposed into several factors showing (a) differences in cost efficiency, (b) differences in scale efficiency, (c) autonomous shift in the cost function due to technical change, and (d) scale effects of technical change. This paper builds on a comparable decomposition of the Fisher productivity Index by Ray and Mukherjee (1996) on one hand, an earlier static decomposition of cost competitiveness due to Ray and Mukherjee (2000), and a differential measure of the contribution of productivity in cost change due to Grifell-Tatje and Lovell (2000). State level data on output quantities and input prices and quantities constructed from the U.S. Census of Manufacturers for different years are used in an empirical application of the proposed DEA model.

The rest of the paper unfolds as follows. Section 2 presents the conceptual background of cost competitiveness based on standard economic theory. Section 3 briefly describes the relevant nonparametric DEA methodology. Section 4 uses an empirical application using U.S. manufacturing Data. Section 5 is the conclusion.

2. The Conceptual Background

A Static Measure of Competitiveness

Consider two firms from two regions, *A and B*, producing a scalar output *y* from a vector of *n* inputs $x \in R_n^+$. It is assumed that both firms have access to the same production technology and are observed at the same point in time. The input price vectors in the two regions are w^A and w^B, respectively. The input-output bundles of the two firms are (x^A, y_A) and (x^B, y_B). The actual costs of the two firms are

$C_A = w^{A\prime} x^A$ and $C_B = w^{B\prime} x^B$. The corresponding average costs are $AC_A = \frac{C_A}{y_A}$ and $AC_B = \frac{C_B}{y_B}$. As argued above, firm *A* has a competitive advantage over firm *B* if AC_A is lower than AC_B. A measure of the cost competitiveness index of *A* over *B* is

$$CCI_{A|B} = \frac{AC_B}{AC_A}. \qquad (1)$$

When this index exceeds unity *A* is more competitive than *B*. The greater the value of this index, the greater is its competitiveness. On the other hand, if the index is less than 1, *A* stands behind *B* in respect of competitiveness.

<u>Productivity and Competitiveness</u>

For simplicity we first consider the 1-input 1-output case. *A* uses x_A of the scalar input to produce y_A units of the output. Similarly, *B* uses x_B to produce y_B. In this simple case, the average productivities of the two firms are $AP_A = \frac{y_A}{x_A}$ and $AP_B = \frac{y_B}{x_B}$. A measure of the relative productivity of *A* when compared to *B* is

$$\pi_{A|B} = \frac{AP_A}{AP_B} = \frac{\frac{y_A}{x_A}}{\frac{y_B}{x_B}} = \frac{\frac{x_B}{y_B}}{\frac{x_A}{y_A}}. \qquad (2)$$

Now suppose both firms paid the same price $\bar{w}$ for the input. In that case, $C_A = \bar{w}x_A$ and $C_B = \bar{w}x_B$. In this case, it is obvious, that

$$CCI_{A|B} = \frac{AC_B}{AC_A} = \frac{\frac{\bar{w}x_B}{y_B}}{\frac{\bar{w}x_A}{y_A}} = \pi_{A|B}. \qquad (3)$$

In other words, cost competitiveness reflects relative productivity. When the firms pay two different prices (w_A and w_B),

$$CCI_{A|B} = \frac{AC_B}{AC_A} = \frac{\frac{w_B x_B}{y_B}}{\frac{w_A x_A}{y_A}} = \frac{w_B}{w_A}\pi_{A|B}. \qquad (4)$$

In this case, cost competitiveness is determined by both productivity and input price differences. Even when *A* has a lower productivity than *B*, a sufficiently lower input price may make it more competitive.

Of course, in the more realistic case of multiple inputs, even with a scalar output, the simple average productivities and the input price ratio defined above are meaningless. We now show how the relation between competitiveness index, productivities, and input prices can be retained by using a *total factor productivity index* and *input price index*. Note that the cost competitiveness index can be expressed alternatively as either

$$\frac{AC_B}{AC_A} = \frac{\frac{w^{B\prime}x^B}{y_B}}{\frac{w^{A\prime}x^A}{y_A}} = \frac{\frac{y_A}{y_B}}{\frac{w^{A\prime}x^A}{w^{B\prime}x^B}}$$

$$= \frac{\frac{y_A}{y_B}}{\frac{w^{A\prime}x^A}{w^{A\prime}x^B}} \cdot \frac{w^{B\prime}x^B}{w^{A\prime}x^B} \tag{5}$$

or

$$\frac{AC_B}{AC_A} = \frac{\frac{w^{B\prime}x^B}{y_B}}{\frac{w^{A\prime}x^A}{y_A}} = \frac{\frac{y_A}{y_B}}{\frac{w^{A\prime}x^A}{w^{B\prime}x^B}}$$

$$= \frac{\frac{y_A}{y_B}}{\frac{w^{B\prime}x^A}{w^{B\prime}x^B}} \cdot \frac{w^{B\prime}x^A}{w^{A\prime}x^A} \tag{6}$$

Taking the geometric mean of the two, we get

$$\frac{AC_B}{AC_A} = \frac{\frac{y_A}{y_B}}{\left[\frac{w^{A}x^{A}}{w^{A\prime}x^B} . \frac{w^{B\prime}x^A}{w^{B\prime}x^B}\right]^{\frac{1}{2}}} \cdot \left[\frac{w^{B\prime}x^A}{w^{A\prime}x^A} \cdot \frac{w^{B\prime}x^B}{w^{A\prime}x^B}\right]^{\frac{1}{2}} \tag{7}$$

Define the Fisher input quantity index

$$Q^x_{A|B} = \left[\frac{w^{A\prime}x^A}{w^{A\prime}x^B} . \frac{w^{B\prime}x^A}{w^{B\prime}x^B}\right]^{\frac{1}{2}} \tag{8}$$

and the Fisher input price index

$$W_{B|A} = \left[\frac{w^{B\prime}x^A}{w^{A\prime}x^A} \cdot \frac{w^{B\prime}x^B}{w^{A\prime}x^B}\right]^{\frac{1}{2}} \tag{9}$$

Then the cost competitiveness index becomes

$$CCI_{A\}B} = \frac{AC_B}{AC_A} = \frac{\frac{y_B}{y^A}}{Q_x} . W_{B|A}. \tag{10}$$

Finally define the output quantity index

$$Q^{y}_{A|B} = \frac{y_A}{y_B}. \qquad (11)$$

Then

$$CCI_{A\}B} = \frac{AC_B}{AC_A} = \frac{Q^{y}_{A|B}}{Q^{x}_{A|B}}.W_{B|A}. \qquad (12)$$

The first factor on the right hand side of (12) is the Fisher Total Factor Productivity index while the second is the input price index (of *B* relative to *A*). Again, as in the single input case, *A* may be found to be cost competitive relative to *B* even when it has a lower total factor productivity if the input price index (of *B* relative to *A*) is sufficiently greater than unity.

<u>Decomposition of the Cost Competitiveness Index:</u>

We now identify the various factors that contribute to competitiveness through a multiplicative decomposition of a number of normative components of the index along the lines of Ray and Mukherjee (2000). For this we need to define a benchmark technology in terms of the production possibility set:

$$T = \{(x, y): y \text{ can be produced from } x\}. \qquad (13)$$

Any $(x , y) \in T$ is a feasible input output bundle. Obviously, both (x^A, y_A) and (x^B, y_B) are elements of the set *T*. The minimum cost of producing a specific output level y_0 at input prices w^0 is

$$C(w^0, y_0) = \min w^{0\prime} x : (x, y_0) \in T.$$

Clearly, the actual cost of firm *A* cannot be any lower than the minimum cost of producing y_A at input prices w^A. That is

$$C(w^A, y_A) \le C_A = w^{A\prime} x^A.$$

The cost efficiency of firm *A* can be measured as

$$CE_A = \frac{C(w^A, y_A)}{C_A} \leq 1. \qquad (14)$$

Similarly, the cost efficiency of firm *B* is

$$CE_B = \frac{C(w^B, y_B)}{C_B} \leq 1. \qquad (15)$$

Whenever cost efficiency is less than unity, the observed output of a firm can be produced at a lower cost. This would automatically lower the average cost and enhance its cost competitiveness *ceteris paribus.*

It is possible to decompose the cost competitiveness index as

$$\begin{aligned} CCI_{A|B} &= \frac{AC_B}{AC_A} = \frac{\frac{C_B}{y_B}}{\frac{C_A}{y_A}} \\ &= \frac{\left(\frac{C_B}{C(w^B,y_B)}\right)\cdot\left(\frac{C(w^B,y_B)}{y_B}\right)}{\left(\frac{C_A}{C(w^A,y_A)}\right)\cdot\left(\frac{C(w^A,y_A)}{y_A}\right)} \\ &= \frac{\left(\frac{C(w^A,y_A)}{C_A}\right)\cdot\left(\frac{C^B(w^B,y_B)}{y_1}\right)}{\left(\frac{C(w^B,y_B)}{C_B}\right)\cdot\left(\frac{C(w^A,y_A)}{y_A}\right)} \qquad (16) \\ &= \left[\frac{CE_A}{CE_B}\right]\cdot\left[\frac{\left(\frac{C(w^B,y_B)}{y_B}\right)}{\left(\frac{C(w^A,y_A)}{y_A}\right)}\right]. \end{aligned}$$

The first factor on the right hand side is the ratio of the cost efficiencies of the two firms. A higher cost efficiency of firm *A* adds to its cost competitiveness.

We focus now on the second factor which would measure the cost competitiveness of *A* if the two firms had the same level of cost efficiency. The two terms in the numerator and the denominator measure the average cost of producing the output levels y_B and y_A , respectively, at the applicable input prices . Now, unless the technology exhibits constant returns to scale globally, the average cost varies with the level of output. The output level where the average cost for a given vector of input prices reaches a minimum is known as the efficient scale of production. Let y_A^* be the efficient scale for prices w^A.

Then $\frac{C(w^A, y_A^*)}{y_A^*} \le \frac{C(w^A, y_A)}{y_A}$ and

$$SE_A = \frac{\frac{C(w^A, y_A^*)}{y_A^*}}{\frac{C(w^A, y_A)}{y_A}} \text{ is the scale efficiency of firm } A.$$

Similarly, y_B^* is the efficient scale for prices w^B and

$$SE_B = \frac{\frac{C(w^B, y_B^*)}{y_B^*}}{\frac{C(w^B, y_B)}{y_B}} \text{ is the scale efficiency of firm } B.$$

Hence, this second factor in (16) can be expressed as

$$\left[\frac{\left(\frac{C(w^B, y_B)}{y_B}\right)}{\left(\frac{C(w^A, y_A)}{y_A}\right)}\right] = \left[\frac{\frac{\frac{C(w^B, y_B)}{y_B}}{\frac{C(w^B, y_B^*)}{y_B^*}}}{\frac{\frac{C(w^A, y_A)}{y_A}}{\frac{C(w^A, y_A^*)}{y_A^*}}}\right]\left[\frac{\frac{C^{1}w^B, y_B^*)}{y_B^*}}{\frac{C(w^A, y_A^*)}{y_A^*}}\right] \qquad (17)$$

$$= \left[\frac{SE_A}{SE_B}\right] \cdot \left[\frac{\frac{C(w^B, y_B^*)}{y_B^*}}{\frac{C(w^A, y_A^*)}{y_A^*}}\right].$$

Next

$$\left[\frac{\frac{C(w^B, y_B^*)}{y_B^*}}{\frac{C(w^A, y_A^*)}{y_A^*}}\right] = \left[\frac{C(w^B, y_B^*)}{C(w^A, y_B^*)}\right] \cdot \left[\frac{\frac{C(w^A, y_B^*)}{y_B^*}}{\frac{C(w^A, y_A^*)}{y_A^*}}\right]. \qquad (18)$$

Alternatively,

$$\left[\frac{\frac{C(w^B, y_B^*)}{y_B^*}}{\frac{C(w^A, y_A^*)}{y_A^*}}\right] = \left[\frac{C(w^B, y_A^*)}{C(w^A, y_A^*)}\right] \cdot \left[\frac{\frac{C(w^B, y_B^*)}{y_B^*}}{\frac{C(w^B, y_A^*)}{y_A^*}}\right]. \qquad (19)$$

Taking the geometric mean of the two

$$\left[\frac{\frac{C(w^B,y_B^*)}{y_1^*}}{\frac{C(w^A,y_A^*)}{y_A^*}}\right]=\left[\frac{C(w^B,y_A^*)}{C(w^A,y_A^*)}\frac{C(w^B,y_B^*)}{C(w^A,y_B^*)}\right]^{\frac{1}{2}}.\left[\frac{\frac{C(w^B,y_B^*)}{y_B^*}}{\frac{C(w^B,y_A^*)}{y_A^*}}.\frac{\frac{C(w^A,y_B^*)}{y_B^*}}{\frac{C(w^A,y_A^*)}{y_A^*}}\right]^{\frac{1}{2}}. \qquad (20)$$

The first factor of the right hand side of (20) is a ‘cost of production index’ or a *true* input price index. A value of this term greater than unity implies that it costs more to produce the same output quantity (y_A^* or y_B^*) at input prices w^B than at prices w^A. Naturally, it shows an input price advantage for *A*.

Finally, the last factor can also be expressed as

$$\left[\frac{\frac{C(w^B,y_B^*)}{y_B^*}}{\frac{C(w^B,y_A^*)}{y_A^*}}.\frac{\frac{C(w^A,y_B^*)}{y_B^*}}{\frac{C(w^A,y_A^*)}{y_A^*}}\right]^{\frac{1}{2}}=\left[\frac{\frac{\frac{C(w^B,y_B^*)}{y_B^*}}{\frac{C(w^B,y_A^*)}{y_A^*}}}{\frac{\frac{C(w^A,y_A^*)}{y_A^*}}{\frac{C(w^A,y_B^*)}{y_B^*}}}\right]^{\frac{1}{2}}. \qquad (21)$$

This can be regarded as a scale bias of input price change. If this ratio exceeds unity, y_A^* has a higher scale efficiency at input prices w^B than the scale efficiency of y_B^* at prices w^A. A value of this factor different from unity reflects non-homotheticity of the technology. In the case of homotheticity, the cost function is multiplicatively separable in input prices and the output quantity. As a result, when input prices change, the total (and average) cost curves experience a neutral shift. In particular, the average costs at prices w^A and w^B reach their respective minima at the same level output. In that case, y_A^* and y_B^* are identical. Hence, this last factor equals unity.

Thus, a complete decomposition of the cost competitiveness index is

$$CCI=\frac{AC_B}{AC_A}=\frac{\frac{C_B}{y_B}}{\frac{C_A}{y_A}}$$

$$=\frac{\frac{C(w^A,y_A)}{C_A}}{\frac{C(w^B,y_B)}{C_B}}.\left[\frac{C(w^B,y_A^*)}{C(w^A,y_A^*)}.\frac{C(w^B,y_B^*)}{C(w^A,y_B^*)}\right]^{\frac{1}{2}}.\frac{\frac{AC(w^A,y_A^*)}{AC(w^A,y_A)}}{\frac{AC(w^B,y_B^*)}{AC(w^B,y_B)}}.\left[\frac{AC(w^A,y_B^*)}{AC(w^A,y_A^*)}.\frac{AC(w^B,y_B^*)}{AC(w^B,y_A^*)}\right]^{\frac{1}{2}} \qquad (22)$$

$$=(CEI).(IPI).(SEI).(SBIPC)$$

Here

$$CEI = \left[\frac{CE_A}{CE_B}\right]$$ is the cost efficiency index,

$$IPI = \left[\frac{C(w^B, y_A^*)}{C(w^A, y_A^*)} \cdot \frac{C(w^B, y_B^*)}{C(w^A, y_B^*)}\right]^{\frac{1}{2}}$$ is the input price index,

$$SEI = \left[\frac{SE_A}{SE_B}\right]$$ is the scale efficiency index, and

$$SBIPC = \left[\frac{AC(w^A, y_B^*)}{AC(w^A, y_A^*)} \cdot \frac{AC(w^B, y_B^*)}{AC(w^B, y_A^*)}\right]^{\frac{1}{2}}$$ is the scale bias of input price change.

Of course, following the usual Farrell decomposition, one can further separate the technical and allocative efficiency components of overall cost efficiency. The input oriented technical efficiency of *A* is

$$TE_A = \min \theta : (\theta x^A, y_A) \in T. \qquad (23)$$

The corresponding allocative efficiency is

$$AE_A = \frac{CE_A}{TE_A}. \qquad (24)$$

Similarly,

$$TE_B = \min \theta : (\theta x^B, y_B) \in T \qquad (25)$$

and

$$AE_B = \frac{CE_B}{TE_b}. \qquad (26)$$

Cost competitiveness index and its various components are explained diagrammatically in Figures 1-4. In Figure 1 the points A_0 and B_0 show the actual output levels and the costs of the two firms *A* and *B*. As is apparent from the slopes of the two lines OB_0 and OA_0 , *A* has a lower average cost and, hence, is cost competitive relative to *B*. The cost competitiveness index of *A* can be measure by the ratio $\frac{B_0 y_B}{D y_B}$. However, the minimum cost of producing y_A is $A_1 y_A$. Thus, cost efficiency of *A* is $\frac{A_1 y_A}{A_0 y_A}$. Similarly, the cost efficiency of B is $\frac{B_1 y_B}{B_0 y_B}$. For the cost function $C(w^A, y)$ average cost reaches a minimum at the output level y_A^*. Scale efficiency of A is $SE_A = \frac{\frac{A_2 y_A^*}{y_A^*}}{\frac{A_1 y_A}{y_A}}$. Similarly, $SE_B = \frac{\frac{B_2 y_B^*}{y_B^*}}{\frac{B_1 y_B}{y_B}}$.

Figure 2 shows the Farrell decomposition of cost efficiency into technical and allocative efficiency. The points x^A, x_A^T, and x_A^* are, respectively, the actual, technically efficient, and cost efficient input bundles of the firm A. Accordingly, $TE_A = \frac{Ox_A^T}{Ox_A}$, $AE_A = \frac{Ox_A^T}{OC_A^*}$, and $CE_A = \frac{OC_A^*}{Ox_A}$. Similarly, $TE_B = \frac{Ox_B^T}{Ox_B}$, $AE_B = \frac{Ox_b^T}{OC_B^*}$, and $CE_B = \frac{OC_B^*}{Ox_B}$. In Figure 3, the curves $C(w^A, y)$ and $C(w^B, y)$ are the total cost curves for input prices w^A and w^B. The ratio $\frac{Cy_A^*}{Ay_A^*}$ shows the relative cost of producing output y_A^* at input prices w^B and w^A and is a 'true' cost of production index. Similarly, $\frac{By_B^*}{Dy_B^*}$ is the 'true' cost of production index for output y_B^*. The geometric mean of the two is a Fisher-type cost of production index and reflects the contribution of input price differences towards the cost competitiveness of A. The scale bias of input price change is shown in Figure 4. Along the average cost curve $AC(w^A, y)$ y_A^* is the efficient production scale. Similarly, y_B^* is the efficient scale for $AC(w^B, y)$. The ratio $\frac{Fy_A^*}{Cy_A^*}$ is the cross price scale efficiency of y_A^*. Similarly, $\frac{Ey_B^*}{Dy_B^*}$ is the cross price scale efficiency of y_B^*. The scale bias of input price change is the square root of the ratio of these two cross price scale efficiencies. If the ratio is greater than unity, y_A^* can be broadly interpreted as relatively more scale efficient than y_B^*. Of course, when the technology is homothetic, both average cost curves attain their respective minimum points at the same level of output. In that case, there is no scale bias of input price difference.

Dynamics of Cost Competitiveness over Time:

Although firms continuously strive to become more competitive by lowering their average costs, how the cost competitiveness index of a firm changes over time depends upon the rates of change in the average costs of both itself and its rival. The dynamics of cost competitiveness of A relative to B can be captured by the ratio

$$\frac{CCI_{A|B}^{t+1}}{CCI_{A|B}^{t}} = \frac{\frac{AC_B^{t+1}}{AC_A^{t+1}}}{\frac{AC_B^{t}}{AC_A^{t}}} = \frac{\frac{AC_B^{t+1}}{AC_B^{t}}}{\frac{AC_A^{t+1}}{AC_A^{t}}}. \qquad (27)$$

Define the average cost index of the firm *j* (*j* = *A*, *B*)

$$ACI_j = \frac{AC_j^{t+1}}{AC_j^{t}} (j = A, B) \qquad (28)$$

In order to track the change in the cost competitiveness of *A* relative to *B*, we need to focus on the average cost index of the two individual firms. In what follows, we focus on the change in the average cost of an individual firm between two periods: 0 and 1. Hence, a subscript identifying the firm is no longer used except where necessary.

Let x^t and w^t be the input quantity and price vectors for the firm in period *t* and y_t be its output quantity. Then its cost competitiveness of the firm (relative to itself) over time can be measured as

$$CCI_{1|0} = \frac{\frac{w^{0\prime}x^0}{y_0}}{\frac{w^{1\prime}x^1}{y^1}}. \qquad (29)$$

In an inter temporal context one must accommodate the possibility of technological change over time and define the period-specific production possibility set

$$T^t = \{(x, y): y \text{ can be produced from } x \text{ in period } t\}. \qquad (30)$$

Correspondingly,

$$C^t(w, y) = \min w'x : (x, y) \in T^t. \qquad (31)$$

In a two period comparison,

$$
\begin{aligned}
CCI_{1|0} &== \frac{\frac{C_0}{y_0}}{\frac{C_1}{y_1}} \\
&= \frac{\left(\frac{C_0}{C^0(w^0,y_0)}\right)\cdot\left(\frac{C^0(w^0,y_0)}{y_0}\right)}{\left(\frac{C_1}{C^1(w^1,y_1)}\right)\cdot\left(\frac{C^1(w^1,y_1)}{y_1}\right)} \\
&= \frac{\left(\frac{C^1(w^1,y_1)}{C_1}\right)\cdot\left(\frac{C^0(w^0,y_0)}{y_0}\right)}{\left(\frac{C^0(w^0,y_0)}{C_0}\right)\cdot\left(\frac{C^1(w^1,y_1)}{y_1}\right)} \qquad (32) \\
&= \left[\frac{CE_1}{CE_0}\right]\cdot\left[\frac{\left(\frac{C^0(w^0,y_0)}{y_0}\right)}{\left(\frac{C^1(w^1,y_1)}{y_1}\right)}\right].
\end{aligned}
$$

Now,

$$
\begin{aligned}
\left[\frac{\left(\frac{C^0(w^0,y_0)}{y_0}\right)}{\left(\frac{C^1(w^1,y_1)}{y_1}\right)}\right] &= \left[\frac{\frac{\frac{C^1(w^1,y_1^*)}{y_1^*}}{\frac{C^1(w^1,y^1)}{y_0^*}}}{\frac{\frac{C^0(w^0,y_0^*)}{y^1}}{\frac{C^0(w^0,y_0)}{y_0}}}\right]\left[\frac{\frac{C^0(w^0,y_0^*)}{y_0^*}}{\frac{C^1(w^1,y_1^*)}{y_1^*}}\right] \qquad (33) \\
&= \left[\frac{SE_1}{SE_0}\right]\cdot\left[\frac{\frac{C^0(w^0,y_0^*)}{y_0^*}}{\frac{C^1(w^1,y_1^*)}{y_1^*}}\right].
\end{aligned}
$$

Next

$$
\left[\frac{\frac{C^0(w^0,y_0^*)}{y_0^*}}{\frac{C^1(w^1,y_1^*)}{y_1^*}}\right] = \left[\frac{C^0(w^0,y_0^*)}{C^0(w^1,y_0^*)}\right]\cdot\left[\frac{\frac{C^0(w^1,y_0^*)}{y_0^*}}{\frac{C^1(w^1,y_0^*)}{y_1^*}}\right]. \qquad (34)
$$

Alternatively,

$$
\left[\frac{\frac{C^0(w^0,y_0^*)}{y_0^*}}{\frac{C^1(w^1,y_1^*)}{y_1^*}}\right] = \left[\frac{C^1(w^0,y_1^*)}{C^1(w^1,y_1^*)}\right]\cdot\left[\frac{\frac{C^0(w^0,y_0^*)}{y_0^*}}{\frac{C^1(w^0,y_1^*)}{y_1^*}}\right]. \qquad (35)
$$

Taking the geometric mean of the two

$$\left[\frac{\frac{C^0(w^0,y_0^*)}{y_0^*}}{\frac{C^1(w^1,y_1^*)}{y_1^*}}\right]=\left[\frac{C^0(w^0,y_0^*)}{C^0(w^1,y_0^*)}\cdot\frac{C^1(w^0,y_1^*)}{C^1(w^1,y_1^*)}\right]^{\frac{1}{2}}\cdot\left[\frac{\frac{C^0(w^1,y_0^*)}{y_0^*}}{\frac{C^1(w^1,y_0^*)}{y_1^*}}\cdot\frac{\frac{C^0(w^0,y_0^*)}{y_0^*}}{\frac{C^1(w^0,y_1^*)}{y_1^*}}\right]^{\frac{1}{2}}. \qquad (36)$$

The first factor on the right hand side of (36) is a 'cost of production index' and measures the effect of input price change. A value greater than unity implies that prices are, in general, lower in period 1 than in period 0 improving the competitiveness of the firm over time.

The other factor can be broken up as

$$\begin{aligned}&\left[\frac{\frac{C^0(w^1,y_0^*)}{y_0^*}}{\frac{C^1(w^1,y_0^*)}{y_1^*}}\cdot\frac{\frac{C^0(w^0,y_0^*)}{y_0^*}}{\frac{C^1(w^0,y_1^*)}{y_1^*}}\right]^{\frac{1}{2}} \\ &=\left[\frac{C^0(w^1,y_0^*)}{C^1(w^1,y_0^*)}\frac{C^0(w^0,y_1^*)}{C^1(w^0,y_1^*)}\right]^{\frac{1}{2}}\cdot\left[\frac{\frac{C^1(w^1,y_1^*)}{y_1^*}}{\frac{C^1(w^1,y_0^*)}{y_0^*}}\cdot\frac{\frac{C^0(w^0,y_1^*)}{y_1^*}}{\frac{C^0(w^0,y_0^*)}{y_0^*}}\right]^{\frac{1}{2}}.\end{aligned} \qquad (37)$$

The first factor is a geometric mean of autonomous shift in the cost function (measured at two output levels, y_0^* and y_1^*) due to technical change. A value of this ratio greater than unity implies that the cost function has shifted downwards between period 0 and period 1 due to technical progress. It is a cost-indirect measure of technical change.

Finally, the other factor on the right hand side of (37) can be expressed differently as

$$\left[\frac{\frac{C^1(w^1,y_1^*)}{y_1^*}}{\frac{C^1(w^1,y_0^*)}{y_0^*}}\cdot\frac{\frac{C^0(w^0,y_1^*)}{y_1^*}}{\frac{C^0(w^0,y_0^*)}{y_0^*}}\right]^{\frac{1}{2}}=\frac{\left[\frac{C^0(w^0,y_1^*)}{C^0(w^0,y_0^*)}\cdot\frac{C^1(w^1,y_1^*)}{C^1(w^1,y_0^*)}\right]^{\frac{1}{2}}}{\frac{y_1^*}{y_0^*}}. \qquad (38)$$

The denominator is the ratio of the efficient output scale in the two periods. When $y_1^* > y_0^*$, it implies that the region of scale economies along the cost curve $C^1(w^1, y)$ is bigger than the corresponding region of scale economies along the cost curve $C^0(w^0, y)$. The reverse is true if the output ratio in the denominator is less than unity. It is important to note however, that because the cost function is (at least weakly) monotonic in the output, if the denominator is greater (less) than unity, so must be each ratio in the numerator of (38). Consider, for example, the ratio $\frac{C^0(w^0,y_1^*)}{C^0(w^0,y_0^*)}$. When $y_1^* > y_0^*$, $C^0(w^0, y_1^*) > C^0(w^0, y_0^*)$. Thus, $\frac{C^0(w^0,y_1^*)}{C^0(w^0,y_0^*)}$ is a gross measure of the change in the total cost as the output increases. A similar interpretation applies also to the other ratio $\frac{C^1(w^1,y_1^*)}{C^1(w^1,y_0^*)}$. If the whole expression is less than unity, we can see that although the efficient scale of output is bigger in period 1, the total cost does not increase proportionately. This factor may be interpreted as the *scale bias of technical change.* Note that unlike the other factor that measures the shift in the cost function, this one relates more to the curvature of the cost function.

Putting all the pieces together, a full blown decomposition of the cost-competitiveness index is:

$$CCI_{1|0} = \frac{AC_0}{AC_1} = [CEC].[SEC].[IPC].[TC].[SBTC] \qquad (39)$$

In this decomposition

$$CEC = \frac{CE_1}{CE_0} = \frac{\left(\frac{C^1(w^1,y_1)}{C_1}\right)}{\left(\frac{C^0(w^0,y_0)}{C_0}\right)}$$ is Cost Efficiency Change.

Cost Efficiency Change can be further broken up as

$$CEC = \frac{CE_1}{CE_0} = \left[\frac{TE_1}{TE_0}\right].\left[\frac{AE_1}{AE_0}\right]$$

showing changes in technical and allocative efficiencies.

$$SEC = \left[\frac{SE_1}{SE_0}\right] = \frac{\dfrac{\frac{C^1(w^1,y_1^*)}{y_1^*}}{\frac{C^1(w^1,y_1)}{y_1}}}{\dfrac{\frac{C^0(w^0,y_0^*)}{y_0^*}}{\frac{C^0(w^0,y^0)}{y_0}}}$$ is Scale Efficiency Change.

$$IPC = \left[\frac{C^0(w^1,y_0^*)}{C^0(w^0,y_0^*)}\frac{C^0(w^1,y_0^*)}{C^0(w^0,y_0^*)}\right]^{\frac{1}{2}}$$ is a measure of input price change.

$$TC = \left[\frac{C^1(w^1, y_0^*)}{C^o(w^1, y_0^*)} \frac{C^1(w^0, y_1^*)}{C^0(w^0, y_1^*)} \right]^{\frac{1}{2}}$$ is a measure of Technical Change.

Finally,

$$SBTC = \frac{\left[\frac{C^0(w^0, y_1^*)}{C^0(w^0, y_0^*)} \cdot \frac{C^1(w^1, y_1^*)}{C^1(w^1, y_0^*)} \right]^{\frac{1}{2}}}{\frac{y_1^*}{y_0^*}}$$ is the Scale Bias of Technical Change.

The two new concepts that become relevant in the context of changes in competitiveness over time are technical change and scale bias of technical change. Figure 5 shows the total cost curves for two different time periods, 0 and 1, and for two different input price vectors, w^0 and w^1. The output level, y_0^* and y_1^* are the efficient production scales (where respective average cost curves reach their minima) for $C^0(w^0, y)$ and $C^1(w^1, y)$. Note that y_0^* need not be the efficient scale for $C^0(w^1, y)$. Similarly, y_1^* need not be the efficient scale for $C^1(w^0, y)$. The ratio $\frac{D_1 y_0^*}{D_0 y_0^*}$ measures the shift in the cost curve at constant input prices w^1 between the periods 0 and 1 at the output level y_0^*. Similarly, $\frac{E_1 y_1^*}{E_0 y_1^*}$ measures the shift in the cost curve at constant input prices w^0 between the periods 0 and 1 at the output level y_1^*. Technical change is measured by the geometric mean of the two ratios.

Scale bias of technical change has two features. The first relates to a change in the range of output over which economies of scale are present. As shown in Figure 5, in period 0 and at input prices w^0 economies of scale prevail up to the output level y_0^*. By contrast, in period 1 scale economies (at input prices w^1) are not exhausted until the higher output level y_1^* is reached. The other aspect of the scale bias relates to the change in the total cost along the cost curves relevant for the two periods as the efficient scales of production change. As can be seen from Figure 5, The proportionate increase in the cost along the total cost curve $C^0(w^0, y)$ as output increases from y_0^* to y_1^* is $\frac{B_0 y_1^*}{D_1 y_0^*}$ where as along the total cost curve $C^1(w^1, y)$ it is $\frac{B_0 y_1^*}{E_0 y_0^*}$. A geometric mean of the two results in a much smaller value that what we get from the period-0 cost curve. The overall ratio represented by the scale bias of technical change is less than unity. In this example, technical change favors period 1 over period 0 because the region of scale economies is extended and also the proportionate increase in total cost (between the smaller and the large efficient scale) is less than the in increase in output scales. One needs to remember here that the technical change and the scale bias both relate to the entire technology and are not specific to the actual input outputs of the firm in the two periods.

3. The Nonparametric Methodology

In order to carry out the decomposition of the dynamics of cost competitiveness described above one needs to estimate the various cost functions at different input prices and for different technologies from actual output quantity and input price and quantity data. The two commonly employed empirical methods are the parametric econometric approach of stochastic frontier analysis (SFA) and the nonparametric linear programming method of Data Envelopment Analysis (DEA). In SFA one specifies an explicit form of the cost function and uses the maximum likelihood procedure to estimate the parameters of the specified function. In DEA, by contrast, one makes a number of fairly weak assumptions about the technology but leaves the functional form of the fitted cost frontier unspecified.

In DEA one starts from a sample data set of observed input-output bundles (x^j, y_j) $\left(j = 1, 2, ..., N \right)$ from a number of firms from some industry and makes the following assumptions about the technology:

- All observed input-output bundles are feasible.
- The production possibility set is convex.
- Inputs are freely disposable.
- Outputs are freely disposable.

Based on these assumptions, an empirical estimate of the production possibility set (T) would be

$$S = \left\{ (x, y) : x \geq \sum_{1}^{N} \lambda_j x^j ; y \leq \sum_{1}^{N} \lambda_j y_j ; \sum_{1}^{N} \lambda_j = 1; \lambda_j \geq 0; (j = 1, 2, ..., N) \right\}. \quad (40)$$

This set is often described as the free disposal convex hull of the observed input-output vectors. In the inter temporal context, one constructs two different sets S^0 and S^1 using only the data points from that time period.

The cost function $C^t(w, y)$ evaluated at some specific output level and at given input prices *w* is obtained by solving the linear programming problem:

$$\begin{aligned} & \min \; w'x \\ & s.t. \sum_1^N \lambda_j x_t^j \le x; \\ & \quad \sum_1^N \lambda_j y_{tj} \ge y; \qquad (41) \\ & \quad \sum_1^N \lambda_j = 1; \\ & \quad \lambda_j \ge 0; \; (j = 1, 2, \ldots, N). \end{aligned}$$

In the problem above, x_t^j is the input bundle and y_{tj} is the output of firm *j* in period *t*.

Nonparametric Measure of the Efficient Production Scale

The question of a minimum point of the average cost curve arises only when the technology exhibits variable returns to scale. However, two important points need to be noted. First, locally Constant Returns to Scale holds at the input-output bundle (x^*, y^*) where the average cost reaches a minimum. Second, when the technology exhibits Constant Returns to Scale globally, average cost is a constant at all output level. Taking advantage of these two properties of the cos function, Ray (2011) proposed the following approach to measuring the efficient output scale.

Solve the following CRS cost minimization problem for the unit output level:

$$\begin{aligned} & \min \; w'x \\ & s.t. \sum_1^N \lambda_j x^j \le x; \\ & \quad \sum_1^N \lambda_j y_j \ge 1; \qquad (42) \\ & \quad \lambda_j \ge 0; \; (j = 1, 2, \ldots, N) \end{aligned}$$

Suppose that the optimal solution of (42) is $\left\{x^*; \lambda_j^*(j = 1, 2, \ldots, N)\right\}$. Define $k^* = \sum_1^N \lambda_j^*$.

As shown in Ray (2011), the efficient scale of output where the average cost reaches a minimum is

$$y^* = \frac{1}{k^*}. \qquad (43)$$

4. An Empirical Application with U.S. Manufacturing Data

In this example we use data constructed from the U.S. Census of Manufacturers for the years 1992, 1997, 2002, and 2007 to examine the dynamics of cost competitiveness for a number of states from the Continental U.S. In particular, we focus on the sates of: California (CA), Indiana (IN), Massachusetts (MA), Michigan (MI), New Jersey (NJ), New York (NY), North Carolina (NC), Texas (TX), and Virginia (VA). While not an exhaustive list of leading manufacturing states in the country, this group includes states like MI which has long been an important manufacturing state in the Rustbelt as well as emerging industrial stets like NC and VA from the Sunbelt.

We conceptualize a 1-output, 5-input production technology and use data constructed from the various Census years[1]. It is assumed that in any given year there is no difference in the technology across the states within the continental US. Given the fact that the market for manufactured goods is nationally integrated, we assume that the output price does not vary across states so that the value of output is a reasonable measure of the quantity produced. Input prices, however, do vary across states.

We treat the average (i.e. per firm) input-output bundle *as a data point* from each state and the production possibility set is constructed as the free disposal convex hull of these points. Output is measured by the gross value of production. The gross output value is adjusted by the producer's price index and output quantities from different years are measured in constant 1992 dollars. The inputs included are (a) production labor (L_1), (b) non-production labor (L_2), (c) capital (K), (d) energy (E), and (e) materials (M). Production labor is measured by the number of hours worked. The corresponding input price is wage paid per hour to production workers (w_1). The other labor input is the number of non-production employees. The corresponding wage rate (w_2) is total

[1] A similarly constructed (cross section) data set for the year 2002 was used in Ray, Chen, and Mukherjee (2008) and was included in the data appendix to the paper by the authors.

emolument per employee. The capital input is the average of beginning-of-the year and end-of the year (nominal) values of gross fixed assets. The value of fixed assets was inflation-adjusted using the producer's price index of machinery and transport equipment. The capital input price (i.e. user cost), p_K, is measured by the sum of depreciation, rent, and (imputed) interest expenses per dollar of gross value of capital. The quantity of the energy input (*E*) is constructed by deflating the expenditure on purchased fuels and electricity by state specific energy price (p_E)[2]. Total expenditure on materials, parts, and containers is used as a measure of the materials input quantity (*M*). By implication, materials price (p_M) was set equal to unity for every state. However, for inter temporal compatibility the material input quantity was adjusted by the overall producer's price index for primary articles.

Table 1 shows the production cost per dollar of output in the selected states over the different Census years. In 1992 VA had the lowest average cost of 64.5 cents while NY came second with 65.1 cents. At the other end, MI (78.7 cents) and TX (76.5 cents) led IN and CA (also with average cost exceeding 70 cents). By the year 2007, NC had clearing emerged as the most competitive of all of the states in this group with a significant decline in average cost over the years. This is in sharp contrast with the other states most of which experienced increase in average cost. By 2007, average cost in TX exceeded that in NC by a wide margin of nearly 20 cents per dollar of output.
Table 2 tracks the change in average cost in the different states between successive Census years. Between 1992 and 1997 all states in the group considered except NY and VA experienced a decline in average cost. The biggest decline was in NJ. Between 1997 and 2002 average cost increased in 6 out of 9 states while it declined in NY, NC, and VA. Finally, between 2002 and 2007, again average cost increased in all states except NC and TX. Overall, NC is the only state that saw a decline in average cost over the entire period. Table 3 provides a snapshot picture of cost competitiveness of NC relative to the other states. Note that the CCI reported in the last column of this table is the ratio of the average cost in the other state to the average cost in NC and can be calculated from the

---

[2] We use the industrial sector total energy price for the relevant years (measured in nominal dollars per million Btu). Source: US Energy Information Administration.

top row of Table 1. NC was cost competitive relative to all other states except NY and VA. The other columns of this table show the factors that contribute to cost competitiveness of NC. Overall cost efficiency was higher in NC than in the rival states except for NY, VA and MA (although slightly). Interestingly, lower cost efficiency relative to NY and VA was entirely due to worse allocative efficiency because technical efficiency in NC was equal or better than the rival state in all cases. Scale efficiency in NC was worse than in all states except IN. The column headed IPC shows that the cost of production (for the same output level) would be higher in all of the other states than in NC. Thus, NC enjoyed a large input price advantage. Finally, there was no scale bias of input price differences. Although average costs would be different for the different input prices in the different states, in all cases the minimum point of the average cost was attained at the same output level ($611.4 million) in 1992.

Table 4 shows the decomposition of cost competitiveness index of the individual states in 2007 relative to 1997. The index is measured for each state as

$$CCI_{07|97} = \frac{AC_{1997}}{AC_{2007}}.$$

As was seen in Table 1, average cost in 2007 was higher in all of the states except NC where it fell from 64.5 cents to 60.3 cents. Among the other states NJ saw the worst decline in competitiveness (by 11.2%). All states saw an improvement in cost efficiency. Improvement by more than 20% was found in NC and TX. Increase in cost efficiency in virtually all states was driven mainly by increase in allocative efficiency. Technical efficiency increased in some states and declined in others but the change in either direction was much smaller. Scale efficiency also improved for all states. In several cases (IN, NY, NC, and VA) the improvement exceeded 25%. However, input price change, technical change, and scale bias of technical change adversely affected cost competitiveness in all states. As shown by the input price change (IPC) column, compared to 2007 input prices in 1997 were lower by nearly 10%. Also, the technical change (TC) column shows that even if input prices remained unchanged, the same output level would cost at least 20% less in 1997 than in 2007. This implies a significant upward shift in the cost function due to technical regress. Finally, the scale bias of technical change also hindered competitiveness in all states. In fact, the efficient production scale of an average firm declined from $11.4 million in 1997 to $9.5 million in 2007. However, while the output declined by 17.6%, the cost declined by only 14.7%. The overall conclusion is that while improvement in cost efficiency (due mainly to higher allocative efficiency) and in scale efficiency contributed positively to competitiveness, increase in input prices, technological regress, and scale bias of technical change had a negative impact. The end result was that while NC alone significantly improved cost competitiveness all other states fell behind.

The analytical format proposed in this paper may become quite useful for policy. As is apparent, there are several components of cost competitiveness that are within the control of the firm while others are not. For example, improving cost efficiency through elimination of technical and allocative inefficiency is clearly a task for the management. Increasing scale efficiency would require altering the level of output. Whether that would be profitable would depend on market demand factors. At the state or regional level, however, one can think about meeting the given market demand by altering the number of firms through consolidation of multiple small firms or breaking up large ones. Of the remaining components of competitiveness, the input price factor is what business owners complain most loudly about. Virtually in all states, Chambers of Commerce in different communities lament that high wage and energy costs that they claim to be undermining their competitiveness. The model proposed here can allow one to measure the extent a certain percentage reduction in one input price (like a 10% wage cut) would lower the true cost of production index of the firms in the state. As for the technical change and the accompanying scale bias, there is little that any individual firm or state can do. It needs to be clarified that the popular perception of a switch from labor intensive to capital intensive processes *is not technical change*. It is merely a change in input proportions that occurs when a firm moves from one point to another on the same isoquant. Moreover, technical change, as already mentioned above, is a property of the entire production function. Because no firm has exclusive access to the technology, it basically is a public good. To what extent, a firm can benefit from technical change when it occurs, is a matter of efficiency.

5. Conclusion

Cost competitiveness of a nation or a region within a nation depends as much on its own productivity as that of its rivals. Improvement in efficiency in resource utilization and also full exploitation of scale economies improves cost and scale efficiency making the region more cost competitive. To the extent that higher input prices are due to market imperfections, lowering the cost of production would help. At the sub-national level a state can gain limited competitive advantage from a lower wage rate. In the international context, where technological differences may exist across countries, fostering technological progress should be an important part of long term economic policy.

Table 1 Production cost per $ of Gross Output (1992constant $)

Year	CA	IN	MA	MI	NJ	NY	NC	TX	VA
1992	0.701	0.737	0.684	0.787	0.697	0.651	0.669	0.765	0.645
1997	0.668	0.721	0.650	0.750	0.645	0.669	0.645	0.719	0.648
2002	0.697	0.726	0.670	0.786	0.675	0.631	0.630	0.789	0.627
2007	0.703	0.735	0.706	0.801	0.726	0.671	0.603	0.779	0.674

Table 2 Change from the previous Census Year

			Average Cost Index						
Year	CA	IN	MA	MI	NJ	NY	NC	TX	VA
92-97	0.954	0.979	0.951	0.953	0.925	1.027	0.963	0.94	1.004
97-02	1.043	1.007	1.031	1.049	1.047	0.943	0.976	1.097	0.968
02-07	1.009	1.013	1.053	1.018	1.075	1.064	0.958	0.988	1.075

Table 3 Decomposition of Cost Competitiveness of North Carolina

year	State	CE	TE	AE	SE	IPC	N_HOMO	CCI
1992	CA	1.044	1.015	1.028	0.922	1.087	1	1.047
1992	IN	1.056	1.06	0.997	1.002	1.04	1	1.1
1992	MA	0.996	1	0.996	0.932	1.1	1	1.021
1992	MI	1.096	1.04	1.054	0.968	1.108	1	1.175
1992	NJ	1.027	1	1.027	0.932	1.087	1	1.041
1992	NY	0.976	1	0.976	0.918	1.086	1	0.973
1992	NC	1	1	1	1	1	1	1
1992	TX	1.148	1.019	1.126	0.969	1.027	1	1.143
1992	VA	0.976	1	0.976	0.976	1.011	1	0.964

Table 4 Decomposition of Cost Competitiveness Index (2007 over 1997)

State	CEC	TEC	AEC	SEC	PC	TC	SBTC	$CCI_{07\|97}$
CA	1.148	0.996	1.153	1.162	0.931	0.794	0.965	0.951
IN	1.101	1.026	1.073	1.267	0.941	0.774	0.965	0.981
MA	1.176	1.000	1.176	1.129	0.905	0.795	0.965	0.921
MI	1.133	0.943	1.201	1.192	0.926	0.778	0.963	0.937
NJ	1.136	0.981	1.158	1.150	0.902	0.784	0.963	0.888
NY	1.126	1.021	1.102	1.262	0.918	0.790	0.967	0.996
NC	1.215	1.000	1.215	1.265	0.908	0.791	0.969	1.070
TX	1.222	1.046	1.169	1.117	0.892	0.785	0.965	0.923
VA	1.114	0.958	1.163	1.255	0.908	0.784	0.966	0.961

References:

Grifell-Tatje, E. and C.A.K. Lovell (2000) “Cost and Productivity”; *Managerial Decision Economics* Vol. 21, No 1 (Jan-Feb), pp 19-30.

Ray, S.C. (2011) “Scale Economies and Capacity Utilization in U.S. Manufacturing: A Nonparametric Analysis”; Paper presented at the 9th International DEA Conference held in Thessaloniki, Greece.

Ray, S.C., L. Chen, and K. Mukherjee (2008) “Input price variation across locations and a generalized measure of cost efficiency”; *International Journal of Production Economics*; 116, 208-218.

Ray, S.C. and K. Mukherjee (2000) “ Decomposition of Cost Competitiveness in U.S. Manufacturing: Some State-by-State Comparisons”; *Indian Economic Review*; Vol. XXXV; No 2, pp 133-153.

US Census Bureau. Economic Census, Manufacturing, Subject Series, General Summary. (various years)

World Economic Forum (2012) *The Global Competitiveness Report* 2011-2012.

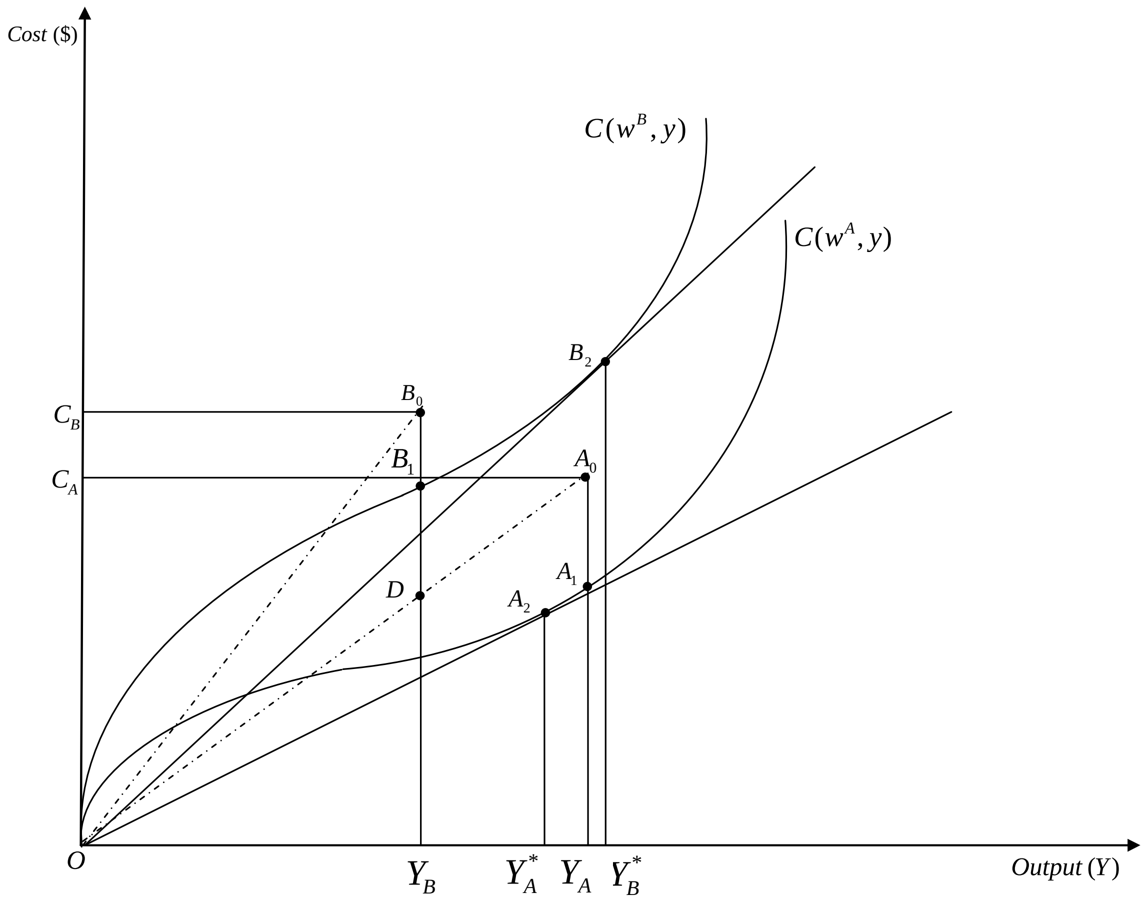


Figure 1: Cost Competitiveness Index

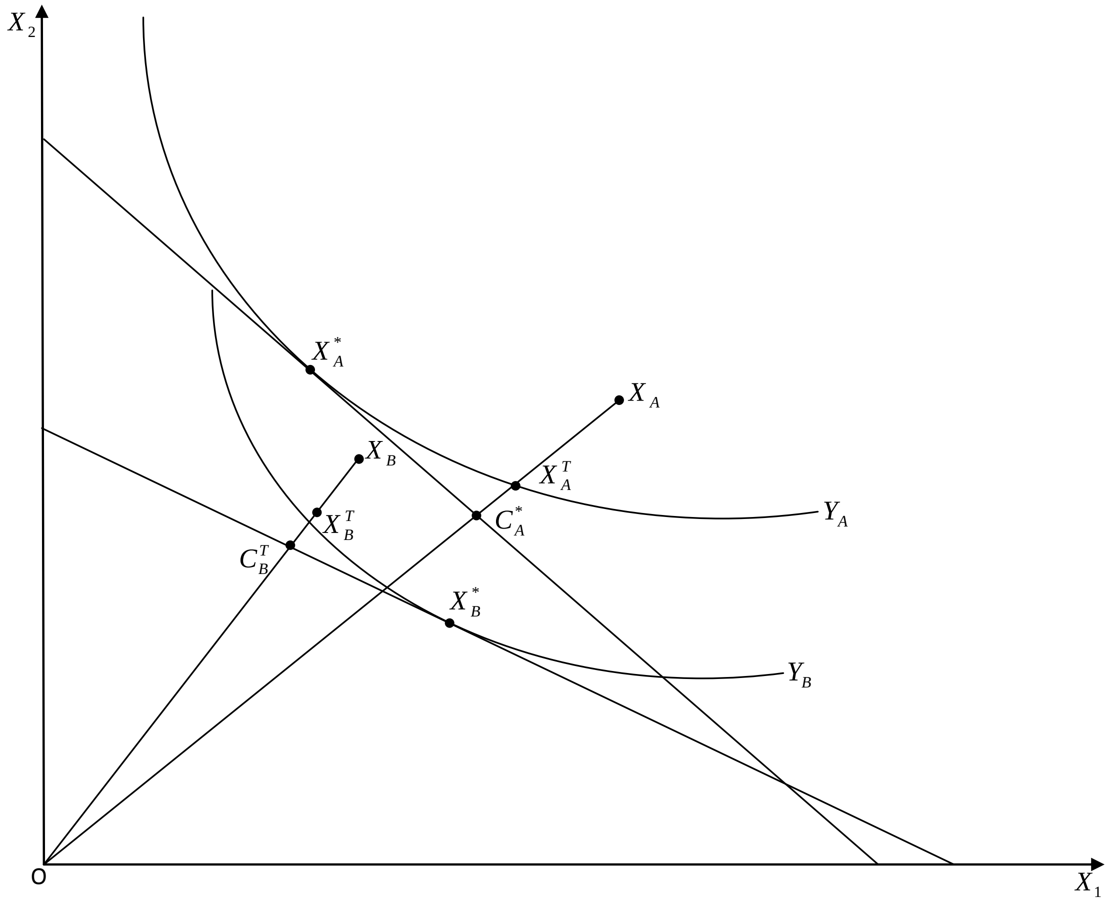


Figure 2: Technical and Allocative Efficiency

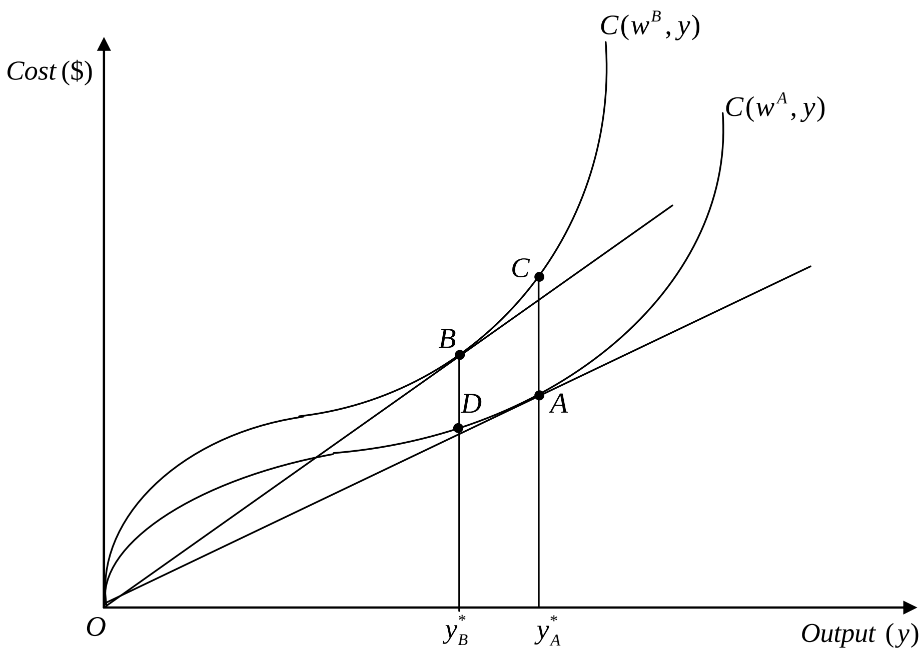


Figure 3: Input Price Change

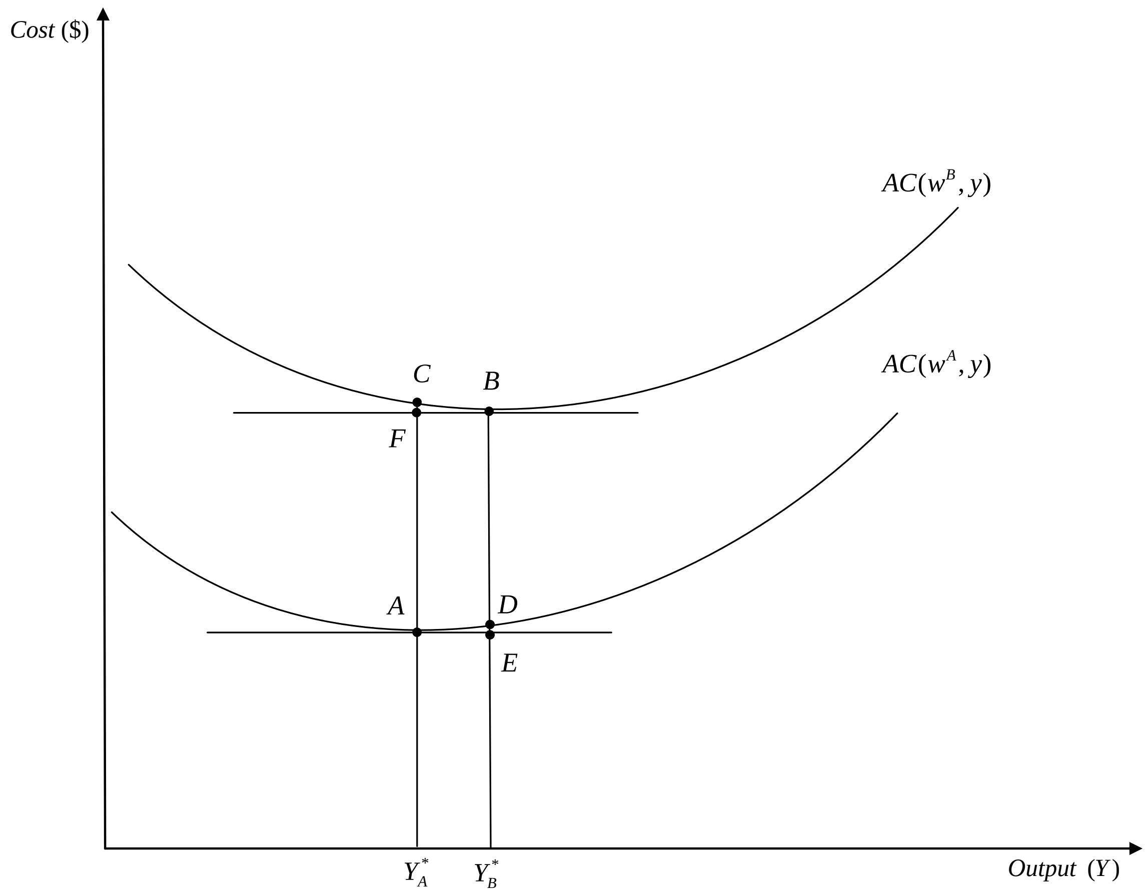


Figure 4: Scale Bias of Input Price Change

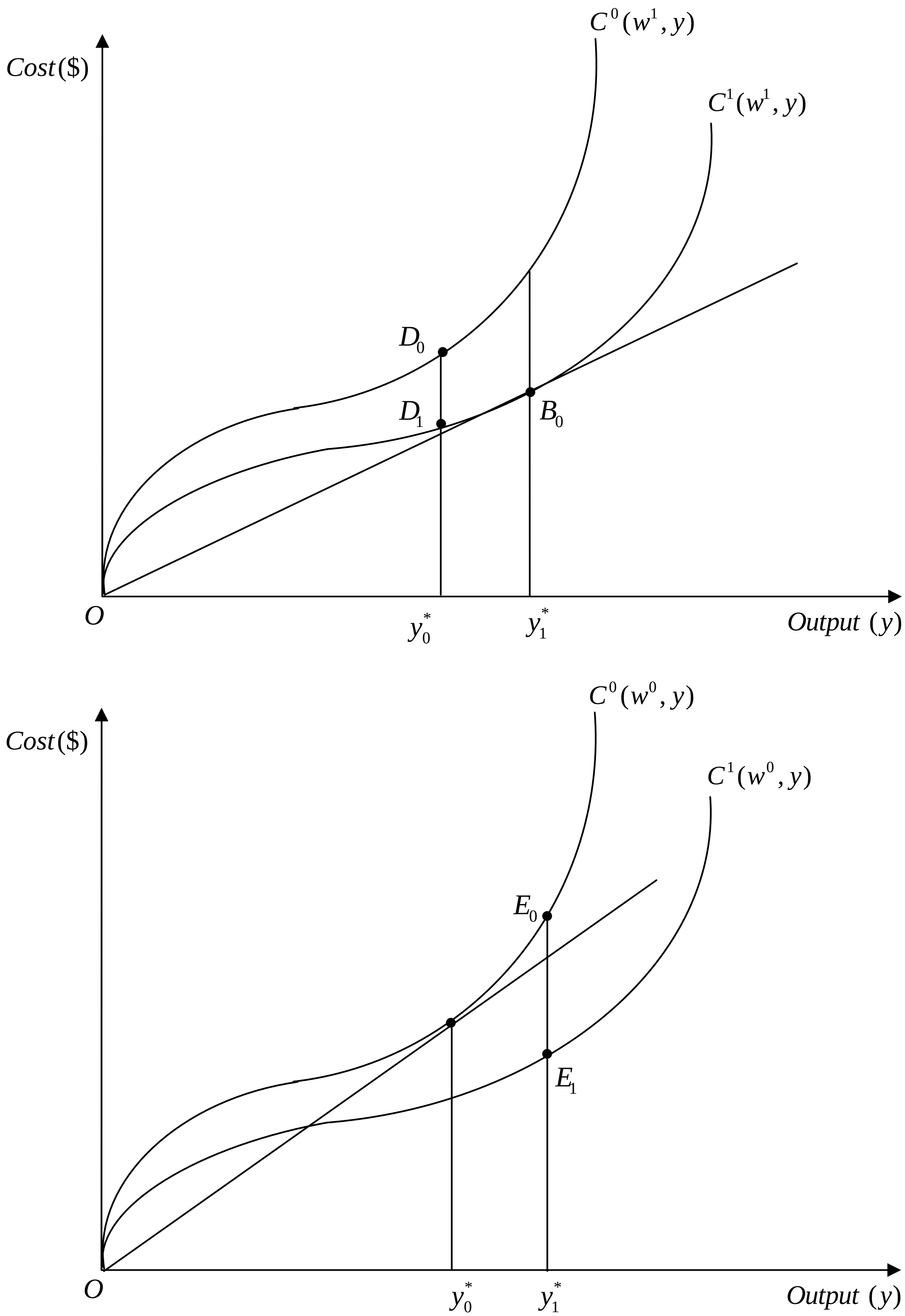


Figure 5 Technical Change and Scale Bias

Measuring Efficiency in Fisheries
in the Presence of Nondiscretionary Inputs

Trevor C. Collier
University of Dayton

Aaron T. Mamula
NOAA Fisheries

John G. Ruggiero
University of Dayton

May 15, 2012

Abstract

Most empirical Data Envelopment Analysis (DEA) models analyzing efficiency and productivity of vessels in multi-species fisheries typically assume that all inputs are discretionary. However, some factors influencing output are exogenous and beyond the control of the vessels. In this paper, we present models useful for controlling nondiscretionary inputs and employ a conditional estimator of technical efficiency. For illustrative purposes, we apply the model to analyze vessel efficiency while controlling for vessel size and a measure of depletion due to competition.

Keywords: Efficiency, nondiscretionary inputs, data envelopment analysis,

1. Introduction

The principal challenge in productivity estimation of fishing vessels is the limited data that exist on vessel operations. In particular the lack of input price data for most fisheries constrains the models available to the analyst. Regression-based approaches to the problem of estimating productivity, technical efficiency and/or harvesting capacity in the absence of input prices have been employed by Segerson and Squires (1990), Felthoven and Morrison-Paul (2004) and Felthoven, Horrace and Schnier (2009) among others.

A popular alternative to the regression based approaches is Data Envelopment Analysis (DEA), a linear programming model that evaluates each production possibility relative to a piecewise linear frontier. DEA, coined by Charnes, Cooper and Rhodes (1984), built on the pioneering work of Farrell (1957) by allowing multiple inputs and outputs assuming constant returns to scale. Banker, Charnes and Cooper (1984) extended DEA to the variable returns to scale technology of Afriat (1972). In addition to allowing multiple outputs, the approach is axiomatic and hence does not require *a priori* specification of the production function. A sample of applications of DEA to the problem of estimating fish production functions include Kirkley et al. (2003), Walden et al. (2003) and Fare et al. (2006).

A second challenge that arises in the analysis of productivity and efficiency among commercial fishermen is how to incorporate non-discretionary inputs to the production function. The potential importance of non-discretionary inputs in fish production was recognized by Kirkley, Squires and Strand (1995). Since then, particular attention has been paid to the impact of one particular non-discretionary input: the resource stock (Kirkley et al. (1998), Pascoe et al.

(2001) and Andersen (2005)). This paper develops a general DEA model which incorporates non-discretionary inputs into productivity analysis and provides a fisheries application.

The standard DEA models allow discretionary inputs and outputs. Banker and Morey (1986) extended the model to allow nondiscretionary variables. However, their extension assumed convexity with respect to the nondiscretionary inputs. Ray (1991) provided a two-stage model; DEA was applied to the discretionary variables and a second-stage regression controlled for the nondiscretionary inputs. Ruggiero (1996) extended the Banker and Morey model with a conditional estimator that did not assume convexity. A limitation of this model was the curse of dimensionality – as the number of nondiscretionary inputs increased, the model tended to find efficiency by default. Ruggiero (1998) developed a three stage model that used Ray's model to construct an environmental harshness index which was incorporated into a third stage using the Ruggiero (1996) model.

In this paper, we explore the measurement of efficiency in the California Limited Entry Groundfish Trawl (groundfish trawl) fishery using data envelopment analysis. The groundfish trawl has historically been one of California's most valuable fisheries[1] and has been analyzed previously in the literature (see *e.g.* Squires and Kirkley (1999), Mason *et al*. (2011) and Collier *et al.* (2012).) In this paper, we focus on the measurement of efficiency and returns to scale in the presence of nondiscretionary factors of production.

Our paper contributes to the applied production economics literature by controlling for nondiscretionary inputs in fishery applications. We present the conditional estimator to control for the fixed factors. We apply the method to estimate efficiency and the environmental

[1] Using data on West Coast commercial fishing revenues from the PacFIN database we summed total revenue for each fishery, for all landings in California from the period 1981 – 2007. The groundfish trawl fishery ranked second among all commercial fisheries in this database.

harshness of vessels operating in multi-species fisheries for illustrative purposes. The rest of the paper is organized as follows. In the next section, we introduce the technology and discuss estimation of efficiency in the presence of nondiscretionary factors. In section 3, we apply the model to the fishery data. The last section concludes.

2. Technology with Nondiscretionary Inputs

We assume that each of n vessels uses a vector $X = (x_1, ..., x_m)$ of m discretionary inputs to produce a vector $Y = (y_1, ..., y_s)$ of s outputs while facing an environment represented by an exogenous (nondiscretionary) variable z. Individual vessel production data for vessel j $(j = 1, ..., n)$ are given by $X_j \equiv (x_{1j}, ..., x_{mj})$, $Y_j = (y_{1j}, ..., y_{sj})$ and z_j. The empirical production possibility set assuming variable returns to scale is given as:

$$
\begin{aligned}
T_V(z) = \{(Y, X, z) : & \sum_{j=1}^{n} \lambda_j y_{kj} \geq y_k, \; k = 1, ..., s; \\
& \sum_{j=1}^{n} \lambda_j x_{lj} \leq x_l, \; l = 1, ..., m; \\
& \sum_{j=1}^{n} \lambda_j = 1; \\
& \lambda_j = 0 \text{ if } z_j > z, \; j = 1, ..., n; \\
& \lambda_j \geq 0, \; j = 1, ..., n\}.
\end{aligned} \tag{1}
$$

The technology in (1) allows variable returns to scale for any given level of the nondiscretionary variable in the standard sense of changing the scale of operation with respect to the discretionary inputs. Also, we assume that output is monotonic with respect to the nondiscretionary input; larger values of z imply a favorable operating environment where the school should produce at least as much output for any given mix of discretionary inputs.

Based on (1), technical efficiency of vessel $i\ (i=1,\dots,n)$ is estimated as the solution to the following linear program:

$$
\begin{aligned}
& TE_i = Min\ \theta \\
& s.t. \\
& \qquad \sum_{j=1}^{n} \lambda_j y_{kj} \ge y_{ki},\ k=1,\dots,s; \\
& \qquad \sum_{j=1}^{n} \lambda_j x_{lj} \le \theta x_{li},\ l=1,\dots,m; \\
& \qquad \sum_{j=1}^{n} \lambda_j = 1; \\
& \qquad \lambda_j = 0 \text{ if } \mathrm{z}_j > z_i,\ j=1,\dots,n; \\
& \qquad \lambda_j \ge 0,\ j=1,\dots,n.
\end{aligned}
\tag{2}
$$

Here, the frontier is defined for each level of the non-discretionary input assuming variable returns to scale with respect to the discretionary inputs.[2]

The measurement of efficiency relative to the technology characterized by exogenous factors of production is illustrated in Figure 1. Here we assume that one discretionary input (x_1) is used to produce one output (y_1) with one nondiscretionary factor of production (z_1). Vessels *A* - D are observed with the most favorable environment with the highest level of the exogenous factor (z_A) while vessels *E* - *H* face a harsher envirnonment $(z_E < z_A)$. As shown, the vessels with the better environment are able to produce a given level of output with lower levels of the discretionary input. Alternatively, for a given level of the discretionary input, vessels with a better environment are able to produce more output. It is assumed that vessels *A* – *D* and F – H are technically efficient, each producing observed output with the least amount of discretionary inputs possible given its environment.

[2]Essentially, this model, due to Ruggiero (1996), assumes selective convexity for a given level of nondiscretionary input. See Podinovski (2005) for further discussion.

The only vessel that is technically inefficiency in Figure 1 is *E*. Given nondiscretionary factor z_E, *E* should be able to produce its observed output y_{1E} using only $TE_E x_{1E}$ of the discretionary input. As shown, this is a convex combination of two vessels (*F* and *G*) that also have the same environment z_E. If we employ the standard DEA model with only the discretionary variables, vessel *E* would be projected to an infeasible point defined by a convex combination of *A* and *B*, both of whom have the more favorable environment.

Following Ruggiero (2000), we can obtain information on the effect of the exogenous variable by also solving the DEA model for each vessel $i \ (i = 1, \ldots, n)$ using only the discretionary variables:

$$\begin{aligned}
&\theta_i = Min\ \theta \\
&s.t. \\
&\qquad \sum_{j=1}^{n} \lambda_j y_{kj} \geq y_{ki},\ k = 1, \ldots, s; \\
&\qquad \sum_{j=1}^{n} \lambda_j x_{lj} \leq \theta x_{li},\ l = 1, \ldots, m; \\
&\qquad \sum_{j=1}^{n} \lambda_j = 1; \\
&\qquad \lambda_j \geq 0,\ j = 1, \ldots, n.
\end{aligned} \tag{3}$$

Model (3) differs from (2) with the exclusion of the constraint on the nondiscretionary variable. Returning to Figure 1, we see the resulting projection to the convex combination of *A* and *B* for inefficient vessel *E* discussed above. The solution reveals the minimum level of discretionary input necessary to produce output y_{1E} *if vessel E had the most favorable environment.* For technically efficient vessels *F* – *H* we find that $\theta_i < 1$ meaning that these vessels have to use more of the discretionary input to compensate for the harsher environment.

Combining (2) and (3), we obtain information regarding the effect the exogenous variable has on the production environment; for vessel j $(TE_j - \theta_j)x_{lj}$ measures the extra amount of input l that is needed to compensate the vessel for its environment. If $TE_j = \theta_j$ then no extra inputs are necessary. Likewise, the ratio $(\theta_j / TE_j) \leq 1$, which we call the environmental index provides a measure of environmental harshness and indicates the reduction in discretionary inputs that could have been possible if the vessel faced the best environment, after eliminating inefficiency.

3. Illustrative Example

The groundfish trawl fishery is one of California's largest commercial fisheries, both in terms of pounds of fish landed and revenue generated by the fleet. The fishery is regulated by species-specific catch limits, gear restrictions, and area closures. Recently the fishery transitioned into quota management, although our analysis utilizes data from before this management shift[3].

The data for this analysis consist of trawl logbook[4] reports collected by California Department of Fish and Game for boats with limited entry groundfish permits in 2007. We utilize only trips from the most recent year (2007) in our data set and only trips fishing exclusively in California waters. Logbooks contain trawl locations defined by start and end points of each tow, hours towed, the weight and market category of fish landed from the tow, and landing port. Although our data provide an extremely fine breakdown of species harvested, we use landings and reported output prices to construct a single output: groundfish revenue. In

[3] Under the catch limits management regime each vessel was allowed to land only a limited weight of key targeted species. These limits varied by two-month period and were occasionally adjusted during the year to ensure that total catch of any one species did not exceed the target.

[4] Logbooks are an important data source for economists studying fisheries issues. For more detailed information on how these data are collected and organized see Sampson and Crone (1997, Chapter 4) or the Pacific States Marine Fisheries Commission website: http://pacfin.psmfc.org/pacfin_pub/trawllog.php.

analysis of multi-species fisheries it is common to aggregate individual outputs to a sensible dimensionality (Terry et al., 2008; Fare et al., 2006; Dupont et al., 2002). In this application we choose to simplify the problem to one involving only a single output in order to retain focus on the key modeling issue of incorporating non-discretionary inputs to the fish production function[5].

Using the logbook data, we created a new variable to account for environmental factors outside the control of the vessel operator. Using longitude and latitude data from the log books, we divide the study area into square blocks one sixteenth of a degree by one sixteenth of a degree. We then calculated the number of tow hours each vessel spent in each grid on each trip. Then we were able to calculate the total effort (measured in tow-hours) in each grid at a given time for the fishery as a whole. Finally, we created a measure of crowding for each vessel based on the weighted average of effort in each grid that a vessel fished in a given trip, where the weights are the portion of the total trip spent in each grid. In empirical illustration we utilize this measure of crowding over a 14-day period.

Much of the previous work on incorporating non-discretionary inputs into the fish production function has focused on ways to condition output estimation on the state of the resource stock. A complication for regression based models in this endeavor is finding fishery independent estimates of stock abundance in order to avoid potential endogeneity problems associated with using contemporaneous catch rates to proxy for stock size. Kirkely et al. (1995, 2002 and 2004) provide three possibilities for using observed catch and effort data to measure fish stocks in an SPF framework.

---

[5] The use of total fishery revenue as a multi-species aggregation method has been employed by Sharma and Lueng (1999) and Pascoe and Coglan (2002).

Our analysis takes a slightly more general view of non-discretionary inputs into fish production. We note that our variable indicating the level of crowding faced by a particular vessel on a trip is an attempt to control for spatial variations in resource availability and is consistent with the economic literature on crowding externalities in commercial fisheries, which began with Scott (1957) and Smith (1969). Crowding externalities, also sometimes referred to as interference competition, have also been addressed in the marine ecology literature by Hilborn (1985), Gillis et al. (1993) and Rijinsdorp and Poos (2007). Of particular note, with respect to our application, is the work of Dalton and Ralston (2004). The authors found crowding externalities associated with spatial closures to have significant impacts on the cost structure of groundfish harvesters on California's Central Coast.

We ran model (2) on monthly sub-samples of the data described above for each month in 2007 at the trip level. Using monthly sub-sample allows us to avoid problems with seasonality and differing catch restrictions. Results are displayed in Table 2. Average vessel efficiency over the twelve separate monthly estimates was 0.844, with a high of 0.898 in November and a low of 0.717 in September. Average environmental harshness was 0.84 over the entire twelve months, with a high of 0.937 in December and a low of 0.773 in May. This measure tells us the reduction in discretionary inputs that could have been possible if the vessel faced the best environment, after eliminating inefficiency. Thus, vessels operating in May could have reduced their input usage by 77-percent if they were able to operate in the best environment.

4. Conclusions

DEA has emerged as a popular methodology for estimating efficiency and productivity in fisheries applications because it provides a straightforward framework in which to model multi-output technologies. Past work on incorporating non-discretionary inputs into fish production have tended to rely on regression-based approaches because of the ease with which exogenous variables may be included in these models. In this paper, we presented a nonparametric model of efficiency estimation where the technology set is conditioned on nondiscretionary inputs. Failure to properly control for these exogenous factors leads to technical efficiency estimates that are biased downward. The model was applied to analyze vessel efficiency using 2007 data. We argue that vessel size and crowding (recent competition in the general area) could be beyond the control of the vessels. The results indicate that production is influenced by these factors. Our model allows us to further identify an index of environmental harshness that measures the extra discretionary inputs necessary to compensate for adverse conditions.

Although the practical implications of our empirical illustration are limited by a number of factors[6], we believe the methodology presented here has some important and valuable extensions. First, Johnson and Ruggiero (2011) show how the conditional estimator presented in (2) here can be used to construct a corrected Malmquist productivity index. The Malmquist productivity index has been identified as potentially important tool for tracking productivity change in commercial fisheries over time (Walden et al. 2012). Johnson and Ruggiero show how a DEA model incorporating non-discretionary inputs is used to construct an Environmental Malmquist Index which can be decomposed into changes in efficiency, changes in the operating environment, changes in technology and environmental technical change.

In addition, Ouellette and Vierstraete (2004) utilize the Ruggiero's (1996) conditional estimator, presented in (2) here, to incorporate quasi-fixed inputs as non-discretionary inputs.

[6] In order to retain focus on the incorporation of non-discretionary inputs we employ a rather simple method of multi-species aggregation. Extensions of this work will benefit from a more rigorous definition of the output set. Additionally, we did not investigate all possible sources of vessel heterogeneity in order to define the reference set of vessels. In particular, many groundfish trawlers are also active participants in the West Coast crab pot fishery. We recognize that, in DEA applications such as this, it may be important to distinguish between vessels which are primarily groundfish vessels and vessels which are primarily dependent on the crab fishery, participating in the groundfish fishery infrequently.

The implications of this procedure to fisheries applications are potentially useful. In DEA models of commercial fishing it is common for the analyst to select a reference set of relatively homogenous vessels for each vessel in the sample. This process usually involves selection of peer vessels based on similar vessel length, weight or engine horsepower. The use of the conditional estimator to incorporate inputs fixed in the short-run may help reduce the amount of data preprocessing required in DEA models of commercial fishing.

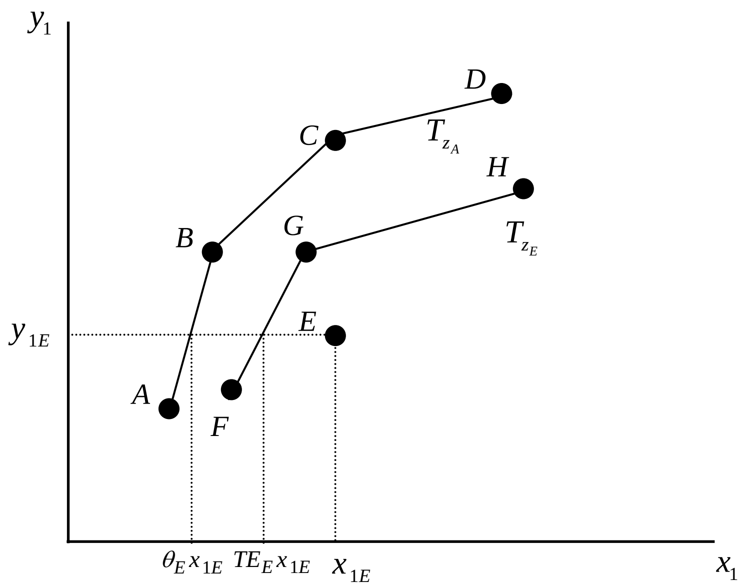


Figure 1: Production with Nondiscretionary Inputs

Table 1: Descriptive Statistics (N = 989)

Variable	Average	Std. Dev.
Output	6,267.11	4,842.21
Tow Hours	16.15	12.07
Days at Sea	2.13	1.05
Vessel Length	57.87	10.69
Crowding Measure	23.44	21.05

Calculations by authors.

Table 2: Average Results, Year 2007

		Efficiency		Environmental Index	
Month	N	Average	Std. Dev.	Average	Std. Dev.
January	70	0.813	0.202	0.769	0.216
February	41	0.801	0.259	0.840	0.201
March	69	0.896	0.183	0.883	0.202
April	79	0.865	0.193	0.768	0.219
May	144	0.865	0.205	0.773	0.257
June	92	0.876	0.215	0.808	0.221
July	98	0.810	0.272	0.894	0.186
August	110	0.869	0.202	0.847	0.191
September	96	0.717	0.297	0.905	0.184
October	97	0.840	0.206	0.827	0.193
November	64	0.898	0.170	0.851	0.173
December	29	0.877	0.174	0.937	0.119

Calculations by authors.

References

Afriat, S. (1972), "Efficiency estimation of production functions," *International Economic Review* 13, 568-598.

Andersen, J.P. (2005), "The inclusion of stocks in multi-species fisheries: The case of the Danish Seiners," *Marine Resource Economics*, 20, 163 – 184.

Banker, R., Charnes, A. and W.W. Cooper (1984), "Some models for estimating technical and scale inefficiencies in data envelopment analysis," *Management Science* 30, 1078- 1092.

Banker, R. and R. Morey (1986), "Efficiency Analysis for Exogenously Fixed Inputs and Outputs," *Operations Research* 34, 513-521.

Charnes, A., Cooper, W.W. and E. Rhodes (1978), "Measuring the inefficiency of decision making units," *European Journal of Operational Research* 2, 429-444.

Collier, T., Mamula, A. and J. Ruggiero (2012), "Estimation of Multi-Output Production Functions in Commercial Fisheries: A New Empirical Approach," under second review at Omega.

Dupont, D., Grafton, R.Q., Kirkley, J. and D. Squires (2002), "Capacity utilization measures and excess capacity in multi-product privatized fisheries," *Resource and Energy Economics* 24, 193-210.

Färe, R., Kirkley, J. and J. Walden (2006), "Adjusting technical efficiency to reflect discarding: The case of the U.S. Georges Bank Multi-species Otter Trawl Fishery," *Fisheries Research* 78, 257-265.

Farrell, M.J. (1957), "The measurement of productive efficiency," *Journal of the Royal Statistical Society Series A* 120, 253-281.

Felthoven, R.G., Horrace, W. and K.E. Schnier (2009), "Estimating heterogeneous capacity and capacity utilization measures in a multi-species fishery," *Journal of Productivity Analysis* 32, 173-189.

Felthoven, R.G., and C.J. Morrison Paul (2004), "Multi-output, nonfrontier primal measures of capacity and capacity utilization," *American Journal of Agricultural Economics* 86, 619-633.

Gillis, D.M., R.M. Peterman and A.V. Taylor (1993), "Movement dynamics in a fishery: Application of the Ideal Free Distribution to spatial allocation of effort," *Canadian Journal of Fisheries and Aquatic Sciences*, 50, 323-333.

Hilborn, R. (1985), “Fleet dynamics and individual variation: why some people catch more fish than others,” *Canadian Journal of Fisheries and Aquatic Sciences*, 42(1), 2-13.

Johnson, A.L. and J. Ruggiero (2011), “Nonparametric measurement of productivity and efficiency in education,” *Annals of Operations Research*, online, DOI: 10.1007/s10479-011-0880-9.

Kirkley, J.E., C.J. Morrison-Paul and D. Squires (2002), “Capacity and capacity utilization in common-pool resource industries,” *Environmental and Resource Economics*, v.22, 71-97.

Kirkley, J.E., C.J. Morrison-Paul, S. Cunningham and J. Catanzano (2004), “Embodied and disembodied technical change in fisheries: An analysis of the Sete trawl fishery, 1985-1999”, *Environmental and Resource Economics*, 29, 191-217.

Kirkley, J.E., I. Strand and D. Squires (1995), “Assessing technical efficiency in commercial fisheries: the Mid-Atlantic sea scallop fishery,” *American Journal of Agricultural Economics*, 77, 686-697.

Kirkley, J.E., I. Strand and D. Squires (1998), “Characterizing managerial skill and technical efficiency in a fishery,” *Journal of Productivity Analysis*, 9, 145 – 160.

Kirkley, J.E., D. Squires, M.F. Alam and H.O. Ishak (2003), “Excess capacity and asymmetric information in developing country fisheries: The Malaysian purse seine fishery,” *American Journal of Agricultural Economics*, 85, 647-662.

Mason, J., Kosaka, R., Mamula, A. and C. Speir (2011), “Effort changes around a marine reserve: The case of the California Rockfish Conservation Area,” *Marine Policy*, v.36(5): 1054-1063.

Ouellette, P. and V. Vierstraete (2004), “Technological change and efficiency in the presence of quasi-fixed inputs: A DEA application to the hospital sector,” *European Journal of Opertations Research*, 154, 755-763.

Pascoe, S., J.L. Andersen and J.W. de Wilde (2001), “The impact of management regulation on the technical efficiency of vessels in the Dutch Beam Trawl Fishery, *European Review of Agricultural Economics*, 28(2): 187 – 206.

Podinovski, V. (2005), “Selective convexity in DEA models,” *European Journal of Operational Research, 161* (552-563).

Poos, J.J. and A.D. Rijnsdorp (2007), “An “experiment” on effort allocation of fishing vessels: the role of interference competition and area specialization,” *Canadian Journal of Fisheries and Aquatic Sciences*, 64(2), 304-313.

Ray, S. C. (1991), “Resource-use efficiency in public schools: A study of Connecticut data,”*Management Science, 37*(12), 1620-1628.

Ruggiero, J. (1996), “On the measurement of technical efficiency in the public sector,”. *European Journal of Operational Research, 90*, 553-565.

Ruggiero, J. (1998), “Nondiscretionary inputs in Data Envelopment Analysis,” *European Journal of Operational Research* 111: 461-469.

Ruggiero, J. (2000), “Nonparametric Estimation of Returns to Scale in the Public Sector with na Application to the Provision of Edcuational Services,” *Journal of the Operational Research Society* 51, 906-912.

Scott, A. (1957), “Optimal utilization and control of a fishery,” In, *The Economics of Fisheries*, eds. Ralph Turvey and Jack Wiseman, Rome: Food and Agriculture Organization of the United Nations.

Segerson, K. and D. Squires (1990), “On the measurement of economic capacity utilization for multiproduct industries,” *Journal of Econometrics* 44, 81-95.

Smith, V.L. (1969), “On models of commercial fishing,” *Journal of Political Economy*, 77, 181-198.

Squires, D. and J. Kirkley (1999), “Skipper skill and panel data in fishing industries,” *Canadian Journal of Fisheries and Aquatic Sciences* 56 (11) 2001-2018.

Terry, J., Walden, J. and J. Kirkley (2008), “National assessment of excess harvesting capacity in federally managed commercial fisheries,” U.S. Department of Commerce, NOAA Tech. Memo, NMFS-F/SPO-93.

Walden, J., Kirkley, J.E. and A. Kitts (2003), “A limited economic assessment of the Northeast Groundfish Fishery Buyout Program,” *Land Economics*, 79, 426-439.

Walden, J.B., J.E. Kirkley, R. Fare and P. Logan (2012), “Productivity change under an individual transferable quota management system,” *American Journal of Agricultural Economics*, online: *doi: 10.1093/ajae/aas025.*

Primal and Dual Approaches to Fishing Capacity: The Impact of the Convexity Assumption

Kristiaan Kerstens*

Niels Vestergaard**

July 2012

Abstract:

Application of primal non-parametric approaches to estimation of fishing capacity provides useful disaggregated information about fishing firm's capacities utilizations. A potentially serious issue is that the estimated capacity utilization rates can be relatively low. This may call for reductions in the fishing fleet that are political impossible to defend. In this paper two modifications of the traditional approach are explored. First, non-convex technologies are introduced and it is shown how the primal non-parametric approach leads to different capacity utilization rates. Then capacity utilization measures using cost functions are specified for both convex and non-convex technologies. It is illustrated how the convexity assumption impacts capacity utilization rates and how this dual approach differs from the primal approach. Second, the effect of utilizing these different convex versus non-convex capacity utilization rates in the short-run Johansen industry production model is explored in terms of the resulting policy conclusions. This model has been used to formulate realistic plans to implement fishery policies (decommissioning schemes, zone and time restrictions, etc). Empirically, we illustrate these capacity measures on the Pacific albacore tuna fishery to document the impact of convexity.

Keywords: Capacity measures; Convexity.

JEL classification: C61, D24.

* CNRS-LEM (UMR 8179), IESEG School of Management, 3 rue de la Digue, F-59000 Lille, France. Corresponding author. k.kerstens@ieseg.fr

** Department of Environmental and Business Economics, Centre of Fisheries and Aquaculture Management and Economics (FAME), University of Southern Denmark, Esbjerg, Denmark.

1. INTRODUCTION

Excess capacity of fishing fleets is one of the most pressing problems facing the world's fisheries and the sustainable harvesting of resource stocks. Since 1990, both world marine fish catches and the world-wide number of vessels have leveled off, albeit that vessel productivity has kept increasing. This has resulted in a situation where many species are fully or over-exploited and with a general excess number of vessels. Adoption of the Precautionary Principle (= calling for resource stocks higher than those for maximal sustainable yield and correspondingly lower sustainable catch levels) by FAO exacerbates the excess capacity problem.

The current situation generates pressure to harvest past the point of sustainability to keep the fishing fleet economically viable. With many vessels operating under little or no profits, reductions in fleet size become politically and socially harder to implement. Excess capacity encourages inefficient allocation and constitutes a major waste of economic resources. Overinvestment occurs: i.e., excessive amounts of variable inputs are used. Excess capacity also complicates the fishery management process, particularly in regulated open access fisheries, by frequently leading to micro-regulation. Excess capacity substantially reinforces the tendency for management decisions to become primarily (re)allocation decisions.

The empirical analyses of technologies and related value functions (e.g., cost functions) have become standard methods of the applied economist. The traditional parametric, semi-parametric and non-parametric specifications of technologies and value functions almost all maintain the axiom of convexity. However, it is well-known that a variety of reasons may generate non-convexities in technology (see Farrell (1959) for an early overview). One example is indivisibilities: the fact that inputs and outputs in production are not perfectly divisible and thus cannot be adjusted in a continuous way. Furthermore, indivisibilities limit the up- and especially the downscaling of production processes. In addition, economies of scale and economies of specialization as well as externalities are all well-known features violating the convexity of technology. Non-convexities create issues about the role of prices in defining equilibria.

Furthermore, the impact of convexity is not limited to estimates based on the technology. Indeed, while it is widely ignored, already Jacobsen (1970) and Shephard (1970) indicated that the cost function is non-decreasing and convex (non-convex) in outputs when technology is convex (non-convex). Briec et al. (2004) proof a similar point: the cost function estimated on a convex technology is always smaller or equal to a cost function computed relative to on non-convex technology. Both cost functions are only identical under a single output and constant returns to scale. Though this potential impact of convexity is probably widely underestimated, it seems clear that convexity can then only be maintained if there is well-established empirical evidence that its impact on most or some specific applications is negligible. This evidence is largely lacking, simply because few studies have explicitly tested for the impact of convexity.

Non-convexities have been documented for some particular industries. In electricity generation, non-convexities exist due to minimum up and down time constraints, multi-fuel effects, etc. This naturally leads to non-convex and non-differentiable variable cost function (e.g., Park et al. (2010)). In car manufacturing cost are non-convex due to changes in the number of shifts and the shutting down or not of plants for some time: Copeland and Hall (2011) find that a non-convex model fits their data best. Non-convexities have been applied in the non-parametric productivity and efficiency literature on some occasions. Cummins and Zi (1998) as well as Grifell-Tatjé and Kerstens (2008) offer cost frontier estimates and cost efficiency ratios for USA life insurance and Spanish electricity distribution respectively that are different from convex results. For oil field petroleum data, Kerstens and Managi (2012) report substantial differences in a Luenberger productivity indicator between convex and non-convex technologies and only find convergence in productivity levels and variations for the latter technology.

However, in the non-parametric literature on capacity the only paper on measuring fishing capacity under non-convex technologies is Kerstens, Squires and Vestergaard (2005). Thus, there is certainly a need for more studies in this direction. In fisheries, there have been parametric applications using both primal and dual approaches (see Morrison (1985) and Segerson and Squires (1990)). Also the non-parametric approach has been applied by numerous authors to fisheries all over the world (see e.g. Pascoe and Greboval (2003), Kirkley, Morrison-Paul and Squires (2002) and Vestergaard (2005)). In several instances the calculated capacity utilization rates are relatively low. One could argue that this is expected under inefficient fishery management. But, when these low capacity utilization rates are plugged into planning models with the aim of restructuring the fleet this inevitably leads to high decommissioning rates. The latter results may not be easy to "sell" to policy makers. For researchers, it is important that the results are robust with respect to changes in the basic assumptions and to be sure that the conclusions obtained reflect the fishery technology and behavior of fishermen. Some considerations in that respect could be:

- What is the relevant production period? (single trip, or longer?)
- What are meaningful choices of inputs and outputs? (aggregation, bycatch, etc.).

- Is there any difference between the primal and dual approach?
- Is the assumption of convexity appropriate?

In this paper we primarily investigate the assumption of convexity in relation to existing notions of capacity utilisation. A multitude of other issues related to implementing capacity notions in fisheries have been discussed among others by Kirkley and Squires (1999). In Section two we develop some more detailed criticisms of convexity. Section three defines the technology and the cost function and explores the often neglected impact of convexity on the latter. An overview of economic and technical capacity utilisation concepts is discussed in Section four. The next section discusses the short-run Johansen industry model that can be used for planning purposes. Section six offers some preliminary empirical evidence of the impact of convexity on some capacity notions as well as on the short-run Johansen industry model. Throughout this analysis, we focus on non-parametric approaches.

It should be stressed that as the first paper of its kind, we sketch an overview of the potential impact of convexity on capacity estimates in fisheries as well as on the plans resulting from the short-run Johansen industry model. Clearly, more detailed work is needed to explore the full set of implications.

2. CRITICISMS OF THE CONVEXITY ASSUMPTION

While the convexity assumption is often maintained in economics, one can find criticisms of the convexity assumption both in consumption and in production theory. Ignoring the issue of consumption, a critique of the convexity assumption in production theory can consider a variety of arguments.

First, Hackman (2008, p. 39) interprets convexity of technology solely in terms of time divisibility of technologies and sees no other justification for its use (this is in line with Shephard (1970)).[1] This time divisibility argument ignores lead times and the associated setup costs that make switching between the underlying activities costly. Thus, time divisibility is a questionable assumption that ideally needs empirical validation. Furthermore, even if time would be perfectly divisible, this does not imply that activities themselves are perfectly divisible. In other words, the question about time divisibility and divisibility of production factors are in principle independent of one another.

[1] It can also be questioned to which extent the time divisibility assumption makes time enter into what essentially is a static production theory. In this interpretation, arguments related to time can only enter into a dynamic theory of production (see also Hackman (2008) for the latter).

Second, convexity is sometimes not considered as a primitive axiom, but it is implied by divisibility and additivity. Of course, then the plausibility of divisibility and additivity separately are at the heart of the debate. First, perfect divisibility of inputs and/or outputs is probably the most debatable assumption. Many if not most operations management problems in industry and distribution involve some forms of indivisibilities and input fixities. This typically results in complex integer and possibly non-linear optimization problems. More in general, all production processes seem to have some lower limit below which a process cannot possibly be scaled down realistically. Consequently, perfect divisibility is highly problematic (see Scarf (1994) or Winter (2008) for more detailed criticisms). Second, while additivity (defined as the possibility of summing two or more input-output bundles) is crucial to define free entry and is considered a plausible axiom in many textbooks, it is not beyond any criticism. For instance, Winter (2008) argues that it presupposes spatial separation and non-interaction, both of which are highly implausible. Since additivity relates to the aggregation of results of activities occurring in geographically distinct places, transportation costs must be small to be safely ignored under spatial separation. But, when activities are close to one another for transportation costs to be negligible (i.e., spatial separation is low), then the risk of production externalities looms when activities get "too close" to create interactions. Furthermore, location matters for quite some outputs (e.g., Italian and Californian lemons are considered different). Third, additivity and divisibility taken together do not only imply convexity, but constant returns to scale as well. The latter returns to scale assumption is in conflict with the existence of indivisibilities and the lower bounds on the scaling of almost all production processes (see supra and Scarf (1994) for a sharp critique).

Summing up, convexity is widely known to be maintained for analytical convenience solely (e.g., Hackman (2008: p. 2)). But, ideally it may require testing in a production context, especially if key results would happen to depend on its validity. Observe that when convexity of technology is questionable, then also the particular assumption of convexity of either the input sets or the output sets is doubtful.

3. TECHNOLOGY AND COST FUNCTION

3.1. Technology: Definitions

We start out by introducing some basic notation and by defining technology and the cost function. A production technology is characterised by the production possibility set: $T = \{(x,y) \mid x$ can produce $y\}$. The input set associated with this technology T denotes all input vectors $x \in \Upsilon^{n}_{+}$

that are capable of generating a given output vector $y \in \Upsilon^m_+$: $L(y) = \{x \mid (x,y) \in T\}$. Often, it is useful to partition the input vector into a fixed and variable part ($x = (x^v, x^f)$) and to make the same distinction regarding the input price vector ($w = (w^v, w^f)$).

The input set $L(y)$ associated with T satisfies some combination of the following standard assumptions (e.g., Hackman (2008)):

L1: $\forall\, y \geq 0$ with $y \neq 0$, $0 \notin L(y)$ and $L(0) = \Upsilon^m_+$.

L2: Let $\{y_n\}_{n\in\infty}$ be a sequence such that $\lim_{n\to\infty} \Pi y_n \Pi = \infty$, then $\cap_{n\in\infty} L(y_n) = 0$.

L3: $L(y)$ is closed $\forall\, y \in \Upsilon^m_+$.

L4: $\forall\, x \in L(y)$, $u \geq x \Rightarrow u \in L(y)$.

L5: $L(y)$ is a convex set $\forall\, y \in \Upsilon^m_+$.

L6: $L(\lambda\, y) = \lambda\, L(y)$, $\forall\, \lambda \geq 0$.

Apart from the traditional regularity conditions (i.e., no free lunch and the possibility of inaction (*L1*), the boundedness (*L2*) and closedness (*L3*) of the input set, and strong (or free) disposal of inputs (*L4*)), there are two other assumptions that are sometimes invoked. Axiom (*L5*) imposes the traditional assumption convexity of the input set. Finally, axiom (*L6*) presents the special case of a homogenous or constant returns to scale (*CRS*) input correspondence contrasting with a more flexible variable returns to scale (*VRS*).

We first define the input distance function that offers a complete characterization of technology. Indeed the input distance function characterizes the input set $L(y)$ as follows:

$$D_i(x,y) \quad = \quad \max\{\lambda : \lambda \geq 0,\, x/\lambda \in L(y)\}. \quad (1)$$

Next, we define the radial input efficiency measure as:

$$DF_i(x,y) = \min\{\lambda \mid \lambda \geq 0, (\lambda\, x) \in L(y)\}. \qquad (2)$$

This measure is simply the inverse of the input distance function ($DF_i(x,y) = [D_i(x,y)]^{-1}$). Its most important properties are: (i) $0 < DF_i(x,y) \leq 1$, with efficient production on the boundary (isoquant) of $L(y)$ represented by unity; (ii) it has a cost interpretation (see Hackman (2008) for details).

Alternatively, technology can also be represented by the output distance function defined over the output set. This formulation is particular useful in the primal approach to capacity. Let the output set associated with technology T denote all output vectors $y \in \Upsilon^m_+$ that can be obtained from a given input vector $x \in \Upsilon^n_+$: $P(x) = \{y \mid (x,y) \in T\}$. Similar to (*L1*)-(*L6*), equivalent assumptions on the output set are available. The output distance function is defined as follows:

$$D_o(x,y) \quad = \quad \min\{\theta : \theta \geq 0,\, y/\theta \in P(x)\}. \qquad (3)$$

We next define the radial output efficiency measure as its inverse:

$$DF_o(x,y) = \max\left\{\theta \;\middle|\; \theta \geq 0, (\theta\, y) \in P(x)\right\}. \qquad (4)$$

This radial output efficiency measure is – just like the input distance function - the inverse of the output distance function ($DF_o(x,y) = \left[D_o(x,y)\right]^{-1}$). Key properties are: (i) 1 < $DF_o(x,y)$, with efficient production on the boundary (isoquant) of $P(x)$ represented by unity; (ii) it has a revenue interpretation (see Hackman (2008)).

3.2. Cost Function: Definition and Impact Convexity

The cost function is a dual representation of technology linked to the input distance function. The cost function defines the minimum expenditures to produce a given output vector y for a given vector of semi-positive input prices $w \in \Upsilon^{n}_{+}$:

$$C(y,w) = \min\ \{wx \mid x \in L(y)\}. \qquad (5)$$

Briec et al (2004) prove an important property with regard to the impact of convexity on the cost function. Costs evaluated on non-convex technologies ($C^{NC}(y,w)$) are higher or equal to costs evaluated on convex technologies ($C^{C}(y,w)$):

$$C^{NC}(y,w) \quad \geq \quad C^{C}(y,w). \qquad (6)$$

Only in case of CRS and a single output (see Briec et al. (2004): Proposition 4), both these cost functions are identical:

$$C^{C}\left(y,w \middle| C\right) = C^{NC}\left(y,w \middle| C\right). \qquad (7)$$

This relation (6) simply reflects the property of the cost function stating that costs are non-decreasing and convex (non-convex) in the outputs if and only if technology is convex (non-convex) (see Jacobsen (1970): Proposition 5.2, or Shephard (1970: p. 227)).

4. ECONOMIC AND TECHNICAL CAPACITY UTILISATION CONCEPTS[2]

A variety of capacity notions exist in the literature. Specifically, it is customary to distinguish between technical (engineering) and economic (mainly cost) capacity concepts (see, e.g., Johansen (1968), Nelson (1989)). Limiting ourselves to capacity notions that can be

[2] This section draws heavily on De Borger et al. (2012).

estimated using non-parametric specifications of technology or value functions, we first treat the economic concepts using a cost frontier approach, and then the technical or engineering notion.

Note that following Squires (1987), Briec et al. (2010) show that it is possible to develop dual capacity measures for the profit maximisation case using non-parametric technologies. In the case of revenue maximisation several proposals are around: e.g., Segerson and Squires (1995), Färe, Grosskopf and Kirkley (2000), and Lindebo, Hoff and Vestergaard (2007), among others. However, it should be stressed that the eventual relations between this large variety of capacity concepts remain to be developed.

It is tradition to distinguish between three basic ways of defining a cost-based notion of capacity (see Nelson (1989)). Each of these notions has the purpose to isolate the effect of excessive or inadequate utilisation of existing fixed inputs (e.g., capital stock) in the short-run.

Advocated by Hickman (1964), among others, the first notion of potential or capacity output is defined in terms of the output produced at short-run minimum average total cost given existing plant and factor prices. It stresses the need to exploit the short-run technology and the shape of the average total cost function is determined by the law of diminishing returns.

Following, e.g., Segerson and Squires (1990), the second definition corresponds to the output at which short and long run average total costs curves are tangent to one another. This corresponds to the intersection point of short and long run expansion paths. This gives this notion a particular theoretical appeal. Both notions presented so far coincide under *CRS*, since minimum of short and long run average total costs is tangent to one another. In fact, there are two variations of this tangency point notion depending on which variables one assumes to be decision variables. One notion assumes that outputs are constant and determines optimal variable and fixed inputs. Another notion assumes that fixed inputs cannot adjust, but outputs, output prices and fixed input prices do adjust.

A third definition of economic capacity, advocated by Cassels (1937) and Klein (1960), among others, focuses on the output at the minimum of the long run average total costs. This notion has been little used, however, probably because it creates confusion with the notion of scale economies.

We first characterise the above three economic capacity notions, one of which has two variants, in a multiple output context in the following definition.

Definition 1: *Reference points of economic capacity notions in the multiple output case are defined as the quantities and prices corresponding to:*

1) *Minimum of short-run total cost function* $C(y,w^v,x^f \mid V)$*:* $C(y,w^v,x^f \mid C)$.

2) *Tangency cost with modified fixed inputs* $C^{tang1}(y,w,x^{f*} \mid V)$*:* $C^{tang1}(y,w,x^{f*} \mid V) = C(y,w \mid V) = C(y,w^v,x^{f*} \mid V)$.

3) *Tangency cost with modified outputs* $C^{tang2}(y(p,w^f,x^f),w,x^f \mid V)$*:* $C^{tang2}(y(p,w^f,x^f),w,x^f \mid V) = C(y(p,w^f,x^f),w \mid V) = C(y(p,w^f,x^f),w^v,x^f \mid V)$.

4) *Minimum of long run total cost function* $C(y,w \mid V)$*:* $C(y,w \mid C)$,

where x^{f*} represents optimal fixed inputs; p is a vector of output prices ($p \in \Upsilon^m_+$); $y(p,w^f,x^f)$ represents a vector of outputs that is adjusted for given output prices, fixed input prices, and the given fixed inputs; and C (V) is a shorthand for *CRS* (*VRS*).

Obviously, this definition does not exhaust all proposals for economic capacity notions available in the literature. For instance, Coelli, Grifell-Tatjé and Perelman (2002) define an alternative ray economic capacity measure based on a short-run profit maximization model whereby the output mix is held constant. This notion is estimated using non-parametric frontiers. While this notion is not without appeal, it has so far been rarely applied. We also remark that a transposition in a cost function context is still missing and its position relative to the above more traditional capacity notions remains to be explored.

Without going into technicalities, we briefly indicate how the four capacity notions of Definition 1 can be estimated using non-parametric frontier specifications. More details are available in the Appendix to De Borger et al. (2012).

First, the minimum of the single output short-run average total cost function can be estimated in the multiple output case by solving for a variable cost function relative to a *CRS* technology ($VC(y,w^v,x^f \mid C)$). Thereafter, one simply adds observed fixed costs ($FC = w^f x^f$) to obtain a short-run total cost function $C(y,w^v,x^f \mid C)$ ($= VC(y,w^v,x^f \mid C) + FC$) as reference point for this first capacity notion.

Second, a tangency notion of capacity aiming at finding a common point between short and long run costs can also be estimated using non-parametric cost frontiers. It is possible to distinguish two types of tangency notions depending on what one considers to be the decision variables. A first tangency cost notion keeps outputs constant and determines the corresponding optimal variable and fixed inputs ($C^{tang1}(y,w,x^{f*} \mid V)$). This notion can be indirectly solved by minimising a long run total cost function $C(y,w \mid V)$ yielding optimal fixed inputs (x^{f*}). The short-run total cost function with fixed inputs equal to these optimal fixed inputs yields by definition exactly the same solution in terms of optimal costs and optimal variable inputs

($C(y,w^v,x^{f*} \mid V) = VC(y,w^v,x^{f*} \mid V) + FC(y,w^v,x^{f*} \mid V)$). Thus, the optimal solution for $C(y,w \mid V)$ generates this first tangency notion.

A second tangency cost notion, favoured by Nelson (1989: 277) and analysed in detail in Briec et al. (2010), keeps fixed inputs constant and adjusts outputs, output prices ($p \in \Upsilon^m_+$) and fixed input prices such that the installed capacity is ex post utilised at a tangency cost level ($C^{tang2}(y(p,w^f,x^f),w,x^f \mid V)$). While outputs are normally assumed to be exogenous in a competitive cost minimisation model, this tangency notion aims to indicate the output quantities and prices as well as the fixed input prices at which existing fixed inputs are optimally utilised. This tangency cost level may imply an output level ($y(p,w^f,x^f)$) below or above the current outputs for a given observation. These tangency costs require for each observation the solution to a non-linear system of inequalities (see Briec et al. (2010)).

Third, the minimum of the run average total costs can be easily indirectly determined by solving for a long run total cost function relative to a *CRS* technology ($C(y,w \mid C)$).

Turning to a technical (engineering) capacity concept, Johansen (1968) pursued a technical approach focusing on a plant capacity notion.[3] Plant capacity is defined as the maximal amount that can be produced per unit of time with existing plant and equipment without restrictions on the availability of variable inputs. Färe, Grosskopf and Kokkelenberg (1989) include this notion into a frontier framework using output efficiency measures (see also Färe, Grosskopf and Lovell (1994: § 10.3)). An output-oriented measure of plant capacity utilisation requires solving an output efficiency measure relative to both a standard technology and the same technology without restrictions on the availability of variable inputs. Plant capacity utilisation in the outputs ($PCU_o(x,x^f,y)$) is defined as:

$$PCU_o(x,x^f,y) = \frac{DF_o(x,y)}{DF_o(x^f,y)}, \qquad (10)$$

where $DF_o(x,y)$ and $DF_o(x^f,y)$ are output efficiency measures relative to technologies including respectively excluding the variable inputs. Defining the technology excluding variable inputs, let the output set associated with technology S denote all output vectors $y \in \Upsilon^m_+$ that can be obtained from a given fixed input vector $x^f \in \Upsilon^n_+$: $P(x^f) = \{y \mid (x^f,y) \in S\}$. Now we can define $DF_o(x^f,y) = \max\{\theta \mid \theta \geq 1, (\theta\, y) \in P(x^f)\}$. Notice that $PCU_o(x,x^f,y) \leq 1$, since $1 \leq DF_o(x,y) \leq DF_o(x^f,y)$.

[3] Johansen (1968) also defines a synthetic capacity concept as the maximal output with existing plant and equipment as well as the currently available variable inputs. This amounts to a basic notion of technical efficiency.

It is important to make a remark on the impact of convexity on the plant capacity utilisation measure. First, note that the component efficiency measures can be signed as follows: $1 \leq DF_o^{NC}(x,y) \leq DF_o^{C}(x,y)$ and $1 \leq DF_o^{NC}(x^f,y) \leq DF_o^{C}(x^f,y)$. In other words, output inefficiency is always lower or equal under non-convexity compared to convexity. Second, the plant capacity utilisation measure in the outputs ($PCU_o(x,x^f,y)$) under convexity versus non-convexity cannot be signed, since it is a ratio of two efficiency measures.

Two more remarks on this issue. First, while the effect of convexity on the cost capacity notions in Definition 1 are unambiguous, the effect on the cost based capacity utilisation notions remains to be explored. Second, do note that the second relation above (i.e., $DF_o^{NC}(x^f,y) \leq DF_o^{C}(x^f,y)$) clearly reveals the impact of convexity when delivering capacity estimates that feed into the industry level models (see infra).

Once a capacity utilisation notion has been selected, one faces the challenge to formulate a capacity utilisation measure. For single output technologies, a straightforward primal capacity utilisation measure is the ratio between actual output and the optimal output corresponding to the capacity notion. Alternatively, one can define dual capacity utilisation measure in terms of the costs due to the input fixity. For multi-output technologies, primal capacity utilisation measures are not straightforward (but Segerson and Squires (1990) have formulated some proposals), and dual measures are readily available.[1] There is little agreement on how to define capacity utilisation measures: some define it as a ratio of observed to "optimal" costs, while others define it the reverse way (see, e.g., Segerson and Squires (1990)).

Notice that for the purpose of the short-run Johansen industry model, it is not only necessary to compute the cost level corresponding to some capacity notion, but one also needs the corresponding optimal variable inputs, fixed or optimal fixed inputs as well as the optimal outputs.

5. INDUSTRY-LEVEL CAPACITY AND PLANNING MODELS

Summing firm-level capacity outputs offers an estimate of the aggregate industry capacity output, hence a measure of overcapacity at the industry level. But, just summing firm-level capacity levels precludes insight into the optimal restructuring and configuration of the industry. For example, the plant capacity measure implicitly assumes that production of capacity output is feasible and that the necessary variable inputs are available. However, the availability of variable inputs may be limited at the industry level.

In fisheries, overall production is limited by the productivity of the fish stock and among the relevant questions at the industry level are:

- What is the optimal firm-structure given current aggregate outputs?
- How can reallocation of inputs and outputs be organised between firms?
- How does this reallocation changes if certain policy concerns are included?
- What are the costs of pursuing these policy issues in terms of allocating more inputs than necessary?

The focus of the short-run Johansen (1972) industry model is on reallocation of resources between production units in an industry by explicitly allowing improvements in technical efficiency and capacity utilization rates. Using a primal approach, the industry model is developed in two steps as follows:

Step 1: From the firm level model $DF_o(x^f,y)$ an optimal activity vector z^{*k} is provided for firm k and hence capacity output and its optimal use of fixed and variable inputs can be computed:

$$y_{km}^* = \sum_{j=1}^{J} z_j^{*k} y_{jm}; \quad x_{kf}^* = \sum_{j=1}^{J} z_j^{*k} x_{jf}; \quad x_{kv}^* = \sum_{j=1}^{J} z_j^{*k} x_{jv}$$

Step 2: These "optimal" frontier estimates at the firm level are used as parameters in the industry model. In particular, the industry model can minimize the industry use of fixed inputs such that total production remains at the current total level (or at a desired target level) by reallocation of resources between firms. Define Y_m as the industry output level of output m and X_f (X_v) as the aggregate fixed (variable) inputs available to the sector of factor f (v):

$$Y_m = \sum_{j=1}^{J} y_{jm}, \quad X_f = \sum_{j=1}^{J} x_{fj} \text{ and } X_v = \sum_{j=1}^{J} x_{vj}$$

The formulation of the second step short-run Johansen (1972) industry model is:

$$\min_{\theta, w, X_v} \theta$$

$$\begin{aligned}
\text{s.t.} \quad & \sum_{j=1}^{J} y_{jm}^* w_j \geq Y_m, && m = 1,..,M, \\
& \sum_{j=1}^{J} x_{fj}^* w_j \leq \theta\, X_f, && f = 1,...,F, \\
& \sum_{j=1}^{J} x_{vj}^* w_j \leq X_v, && v = 1,...,V, \\
& 0 \leq w_j \leq 1, \quad \theta \geq 0, && j = 1,...,J.
\end{aligned}$$

The weight variables w now indicate which firms' capacity is utilized and by how much. The components of the vector w are bounded above at unity, such that current capacities can never be exceeded.[4] The first constraint prevents total production by a combination of firm capacities from falling below the current output levels. The second constraint means that the total use of fixed inputs on the right-hand side cannot be less than the use by a combination of firms. The third constraint calculates the resulting total use of variable inputs. Note that the total amount of variable inputs is a decision variable. The objective function is a radial input efficiency measure focusing on the aggregate use of fixed inputs solely. This input efficiency measure (θ) has a fixed-cost interpretation at the industry level. The activity vector w indicates which portions of the line segments representing the firm capacities are effectively used to produce outputs from given inputs.

To sum up, the optimal solution to this short-run Johansen (1972) industry model gives the combination of firms or branches that can produce the same or more outputs with less or the same use of fixed inputs in aggregate. It measures the combined impact of the removal of any inefficiency, the exploitation of existing plant capacities, and the reallocation of inputs and outputs. The model specifies the optimal production plan of each active firm in the industry with the optimal use of variable and fixed inputs.[5] Notice that an alternative formulation could be to have an output efficiency measure focusing on the expansion of industry outputs that has a revenue interpretation. But, this formulation is not that relevant for a resource constrained industry.

This industry capacity model approach has been applied to several important fisheries, mainly in Europe.[6] One application is reported by Kerstens, Vestergaard and Squires (2006) on the Danish fishing fleet. Different policy options were adopted (e.g., protecting certain vessels groups) and all results show there is substantial overcapacity. Lindebo (2005) offers the first transnational application of this model to the flatfish fishery in the North Sea. The analysis estimates that the same catch could be harvested with a fleet at 77% of its current size, and suggests an optimal reallocation of fixed inputs for each national fleet. Simulations of the impact of possible quota reductions and restrictions of equal capacity reduction across nations were also

[4] In fact, this short-run industry model is geometrically speaking a set consisting of a finite sum of line segments known as a zonotope.
[5] The optimal use of fixed inputs is – in case of no slacks – the same as the current use by definition. The issue of slack is a major issue in DEA and will not be investigated further here.
[6] Apart from fishing industry, other application areas of this revised frontier-based short-run Johansen (1972) industry model known to us are hospitals (see Dervaux, Kerstens and Leleu (2000)) and a bank branch network (see Kerstens et al. (2010)).

considered. Tingley and Pascoe (2005) apply the model to the Scottish fleet and noted the economic benefits to the industry and employment impacts of achieving this capacity reduction and restructuring. Kerstens, Squires and Vestergaard (2005) develop the model using non-convex technologies and observe that, compared with convex technologies, more units remain active in the optimal industry solution. Yagi and Managi (2011) apply the Kerstens, Vestergaard and Squires (2006) framework of policy measures to Japan.

So far, the industry model has not been developed based on economic capacity notions. A straight forward application would be to generate in the first step the optimal frontier data based on the capacity cost approach and then in the second step minimizes the aggregate industry cost subject to the current production. As before the solution will specify a production plan that produces the same or more aggregate outputs at the same or lower overall industry cost. The resulting production plan will tell how outputs and inputs can be optimally reconfigured between the active firms. Another application could be to set up the model where overall costs are minimised in a reallocation problem, see Andersen and Bogetoft (2007) for a similar approach. Our a priori assessment is that the solution could mimic an Individual Transferable Quota system and the shadow values of the output constraints could provide information about the equilibrium quota prices.

6. EMPIRICAL ILLUSTRATIONS

In this section we illustrate the implication of the assumptions convexity and non-convexity on some test data sets using the primal approach as well as the cost-based approach. First, we explore the eventual differences between primal and cost-based approaches to capacity measurement using a sample of French fruit producers. This sample is mainly selected because it has input prices. Therefore, it allows illustrating the effect of convexity on some cost-based capacity notions as well as on the traditional plant capacity notion. Then, we apply the primal approach to annual survey cost data from the US Albacore industry.

6.1. Some Capacity Estimates and Convexity: Preliminary Evidence

As a preliminary empirical analysis, we apply three capacity notions to a small panel of three years of French fruit producers. The sample is based on annual accounting data collected in a survey (Ivaldi et al. (1996)). These fruit producers cultivate two outputs: (i) production of apples, and (ii) aggregate of alternative products. There is also price and quantity information for three inputs: (i) capital (including land), (ii) labour, and (iii) materials. While in total 184 farms are available, just 130, 135 and 140 farms have records in 1984, 1985 and 1986 respectively. This yields a total of 405 observations in this unbalanced panel.

Table 1 reports for three notions of capacity utilisation some basic descriptive statistics of the estimates resulting from a convex versus a non-convex estimation. In particular, we estimate the tangency cost with modified fixed inputs (i.e., $C^{tang1}(y,w,x^{f*} \mid V)$), the minimum of long run total cost function (i.e., $C(y,w \mid V)$), and the output efficiency measure (i.e., $DF_o(x^f,y)$). In particular, this table reports a 10% trimmed mean, as well as the 10^{th}, 25^{th}, 50^{th}, 75^{th}, and 90^{th} percentiles.

The descriptive statistics in Table 1 lead to the following conclusions. First, the distribution of some capacity estimates are quite dispersed, but more importantly the differences between the traditional convex and non-convex estimates are markedly. This is visually confirmed by Figures 1 to 3 displaying the kernel density estimate of the convex and the non-convex distribution for these three capacity notions. Note that each figure uses a bandwidth common to the convex and non-convex distributions to facilitate comparison. Second, the differences between the densities of these convex and non-convex estimates can be tested with a test statistic proposed by Li (1996). This Li test statistic is also valid for dependent variables, whereby dependency is distinctive for frontier estimators. As the bottom line of Table 1 indicates, the null hypothesis of the equality of both convex and non-convex estimates can be rejected for all three capacity notions.

6.2. Plant Capacity Estimates and Short-Run Industry Model: Impact Convexity

In the annual survey data from the US Albacore industry there are 122 observations. The survey data is very detailed with numerous outputs and inputs. We have aggregated the data into two outputs, two variable inputs and two fixed inputs. The purpose is just to illustrate the potential use of the primal plant capacity approach. The outputs and inputs are: (i) catches of albacore and (ii) an aggregate output of other species; two variable inputs (labor and oil); and two fixed inputs (capital and other fixed cost).[7] We compute both the plant capacity measures and the short-run industry model under convex and non-convex technologies. The technology for computing plant capacity assumes *VRS*.

In Table 2 the plant output efficiency measure is reported for different percentiles. Based on this example the estimates of plant capacity utilization rates are lower under convex than

[7] For a detailed description of the dataset: see Squires and Vestergaard (2009).

under non-convexity technologies. At the median, the utilization rate is 1/1.17 under non-convex versus 1/1.95 under convex technologies.

The number of vessels operating at full capacity (i.e., being fully efficient for $DF_o(x^f,y)$) is reported in the last row of Table 2. There are 31 efficient observations under convex technologies, while under non-convex technologies the number is 54. This implies an increase by nearly 75%. Since the fishery is fully developed with a mature fleet (Squires and Vestergaard 2009) one would expect a relative high number of vessels operating at full plant capacity. This probably makes the non-convex results more credible in terms of their realism.

The optimal industry configuration under convex vs. non-convex capacity estimates are shown in Table 3. While the number of active vessels under convex technologies is 74 the number is increased to 80 vessels under non-convex technologies. This amounts to an increase of around 8%.

These illustrative results show the impact of the assumption of convexity. The results also indicate that the assumption is critical and therefore needs to be justified. One could argue that the assumption of non-convexity is more appropriate because the approximation of the frontier is closer to the data and hence therefore does only use the available information provided by the data.

7. CONCLUSIONS

We have raised questions concerning the assumption of convexity in the estimation of primal and dual capacity utilisation notions. Methodologically, it is possible to apply the primal approach under both convexity and non-convexity at the firm as well as at the industry level. This approach can obviously handle both the single output and the multiple output cases. Dual cost function approaches exist under both convexity and non-convexity (the latter have hitherto been totally ignored), and from an economic point of view these non-convex approaches have some advantages. Note that alternative dual capacity utilisation measures (revenue- or profit-function-based) exist in the literature that have been largely ignored in this contribution.

In the current paper, we have tried to provide some preliminary tests on the impact of convexity from an empirical point of view. In general, the assumption of non-convexity seems to give more “conservative” results implying less overcapacity at both firm and industry level.

We note the following limitations of this preliminary study. First, not all capacity notions have been empirically assessed in terms of the impact of convexity. Some alternative cost-based capacity utilisation measures remain to be tested for the effect of convexity.

Furthermore, the repercussions of convexity for most of these capacity measures except the plant capacity measures have not yet been developed at the industry level. Therefore, the assumption of non-convexity needs definitely to be further explored at the industry level.

REFERENCES

Andersen, J.L., P. Bogetoft (2007) Gains from Quota Trade: Theoretical Models and an Application to the Danish Fishery. *European Review of Agricultural Economics*, 34(1), 105-127.

Briec, W., K. Kerstens, P. Vanden Eeckaut (2004) Non-convex Technologies and Cost Functions: Definitions, Duality and Nonparametric Tests of Convexity, *Journal of Economics*, 81(2), 155-192.

Briec, W., K. Kerstens, D. Prior, I. Van de Woestyne (2010) Tangency Capacity Notions Based upon the Profit and Cost Functions: A Non-Parametric Approach and a Comparison, *Economic Modelling*, 27(5), 1156-1166.

De Borger, B., K. Kerstens, D. Prior, I. Van de Woestyne (2012) Static Efficiency Decompositions and Capacity Utilisation: Integrating Technical and Economic Capacity Notions, *Applied Economics*, 44(31), 4125-4141.

Cassels, J.M. (1937) Excess Capacity and Monopolistic Competition, *Quarterly Journal of Economics*, 51(3), 426-443.

Coelli, T., E. Grifell-Tatjé, S. Perelman (2002) Capacity Utilisation and Profitability: A Decomposition of Short-Run Profit Efficiency, *International Journal of Production Economics*, 79(3), 261-278.

Copeland, A., G. Hall (2011) The Response of Prices, Sales, and Output to Temporary Changes in Demand, *Journal of Applied Econometrics*, 26(2), 232–269.

Cummins, D., H. Zi (1998) Comparison of Frontier Efficiency Methods: An Application to the U.S. Life Insurance Industry, *Journal of Productivity Analysis*, 10(2), 131-152.

Dervaux, B., K. Kerstens, H. Leleu (2000) Remedying Excess Capacities in French Surgery Units by Industry Reallocations: The Scope for Short and Long Term Improvements in Plant Capacity Utilization, in: J.L.T. Blank (ed) *Public Provision and Performance: Contributions from Efficiency and Productivity Measurement*, Amsterdam, Elsevier, 121-146.

Färe, R., S. Grosskopf, E. Kokkelenberg (1989) Measuring Plant Capacity, Utilization and Technical Change: A Nonparametric Approach, *International Economic Review*, 30(3), 655-666.

Färe, R., S. Grosskopf, J. Kirkley (2000) Multi-Output Capacity Measures and Their Relevance for Productivity, *Bulletin of Economic Research*, 52(2), 101-112.

Färe, R., S. Grosskopf, C.A.K. Lovell (1994) *Production Frontiers*, Cambridge, Cambridge University Press.

Grifell-Tatjé, E., K. Kerstens (2008) Incentive Regulation and the Role of Convexity in Benchmarking Electricity Distribution: Economists versus Engineers, *Annals of Public and Cooperative Economics*, 79(2), 227-248.

Hackman, S.T. (2008) *Production Economics: Integrating the Microeconomic and Engineering Perspectives*, Berlin, Springer.

Hickman, B.G. (1964) On a New Method for Capacity Estimation, *Journal of the American Statistical Association*, 59(306), 529-549.

Ivaldi, M., N. Ladoux, H. Ossard, M. Simioni (1996) Comparing Fourier and Translog Specifications of Multiproduct Technology: Evidence from an Incomplete Panel of French Farmers, *Journal of Applied Econometrics*, 11(6), 649-668.

Jacobsen, S.E. (1970) Production Correspondences, *Econometrica*, 38(5), 754–771.

Johansen, L. (1968) Production Functions and the Concept of Capacity, Namur, Recherches Récentes sur la Fonction de Production (Collection "Economie Mathématique et Econometrie", n° 2) [reprinted in F.R. Førsund (ed.) (1987) Collected Works of Leif Johansen, Volume 1, Amsterdam, North Holland, 359-382].

Johansen, L. (1972). *Production Functions: An Integration of Micro and Macro, Short Run and Long Run Aspects.* Amsterdam: North Holland Publishing Company.

Kerstens, K., S. Managi (2012) Total Factor Productivity Growth and Convergence in the Petroleum Industry: Empirical Analysis Testing for Convexity, *International Journal of Production Economics*, 139(1), 196-206.

Kerstens, K., B.A. Moulaye Hachem, I. Van De Woestyne, N. Vestergaard (2010) Optimal Capacity Utilization and Reallocation in a German Bank Branch Network: Exploring Some Strategic Scenarios, in: A. Tavidze (ed) *Progress in Economics Research (Volume 16)*, New York, Nova Science, 35-61.

Kerstens, K., D. Squires, N. Vestergaard (2005) Methodological Reflections on the Short-Run Johansen Industry Model in Relation to Capacity Management, *Marine Resource Economics*, 20(4), 425-443.

Kerstens, K., N. Vestergaard, D. Squires (2006) A Short-Run Johansen Industry Model for Common-Pool Resources: Planning a Fishery's Industrial Capacity to Curb Overfishing, *European Review of Agricultural Economics*, 33(3), 361-389.

Klein, L.R. (1960) Some Theoretical Issues in the Measurement of Capacity, *Econometrica*, 28(2), 272-286.

Kirkley, J., D. Squires (1999) *Measuring Capacity and Capacity Utilization in Fisheries,* FAO Fisheries Technical Paper, no. 386.

Kirkley, J., C. Morrison-Paul, D. Squires (2002) Capacity and Capacity Utilization in Common-pool Resource Industries: Definition, Measurement, and a Comparison of Approaches. *Environmental and Resource Economics*, 22(1), 71–97.

Li, Q. (1996) Nonparametric Testing of Closeness between Two Unknown Distribution Functions, *Econometric Reviews*, 15(1), 261–274.

Lindebo, E. (2005) Multi-national Industry Capacity in the North Sea Flatfish Fishery, *Marine Resource Economics*, 20(4), 385-406.

Lindebo, E., A. Hoff, N. Vestergaard (2007) Revenue-based Capacity Utilisation Measures and Decomposition: The Case of Danish North Sea Trawlers, *European Journal of Operational Research*, 180(1), 215–227.

Morrison, C.J. (1985) Primal and Dual Capacity Utilization: An Application to Productivity Measurement in the U.S. Automobile Industry. *Journal of Business & Economic Statistics,* 3(4), 312-324.

Nelson, R. (1989) On the Measurement of Capacity Utilization, *Journal of Industrial Economics*, 37(3), 273-286.

Park, J.-B., Y.-W. Jeong, J.-R. Shin, K.Y. Lee (2010) An Improved Particle Swarm Optimization for Nonconvex Economic Dispatch Problems, *IEEE Transactions on Power Systems*, 25(1), 156-166.

Pascoe, S., D. Gréboval (eds.) (2003) *Measuring Capacity in Fisheries*. FAO Fisheries Technical Paper, No. 445. Rome, FAO. 2003. 314p.

Scarf, H.E. (1994) The Allocation of Resources in the Presence of Indivisibilities, *Journal of Economic Perspectives*, 8(4), 111-128.

Segerson, K., D. Squires (1990) On the Measurement of Economic Capacity Utilization for Multi-Product Industries, *Journal of Econometrics*, 44(3), 347-361.

Segerson, K., D. Squires (1995) Measurement of Capacity Utilization for Revenue-Maximizing Firms, *Bulletin of Economic Research*, 47(1), 77-85.

Squires, D. (1987) Long-Run Profit Functions for Multiproduct Firms, *American Journal of Agricultural Economics*, 69(3), 558-569.

Squires, D., N. Vestergaard (2009) Technical Change and the Commons, Center for Environmental Economics (Working Paper 09-01), University of California San Diego.

Shephard, R.W. (1970) *Theory of Cost and Production Functions*, Princeton, Princeton University Press.

Tingley, D., S. Pascoe (2005) Eliminating Excess Capacity: Implications for the Scottish Fishing Industry, *Marine Resource Economics*, 20(4), 407-424.

Vestergaard, N. (2005) Fishing Capacity in Europe: Special Issue Introduction, *Marine Resource Economics*, 20(4), 323-326.

Winter, S. (2008) Scaling Heuristics Shape Technology! Should Economic Theory Take Notice?, *Industrial and Corporate Change*, 17(3), 513-531.

Yagi, M., S.Managi (2011) Catch Limits, Capacity Utilization and Cost Reduction in Japanese Fishery Management, *Agricultural Economics*, 42(5), 577-592.

Table 1: Convex vs. Non-convex Estimates for Three Capacity Notions

Descriptive statistics	Opt. Cost CRS		Tangency cost with modified fixed inputs		PCU	
	Convex	Non-Convex	Convex	Non-Convex	Convex	Non-Convex
Trimmed mean†	302046.0	430273.8	406443.6	646937.8	22.27	16.41
10th Percentile	59639.3	90648.5	171701.0	211163.7	2.48	1.97
25th Percentile	104092.9	159629.8	225669.2	310735.6	4.47	3.39
50th Percentile	201810.8	311058.8	302346.4	494061.9	9.94	6.98
75th Percentile	428748.3	655480.3	510322.6	845695.3	23.49	17.88
90th Percentile	834635.9	1127155.8	886693.0	1681899.8	95.77	75.02
Li – test	Ho rejected		Ho rejected		Ho rejected	

† 10% trimming level.

Table 2: Plant Output Efficiency Estimates $DF_o(x^f,y)$

	Output efficiency estimate $DF_o(x^f,y)$	
	Convexity	Non-Convexity
10th Percentile	1.00	1.00
25th Percentile	1.12	1.00
50th Percentile	1.95	1.17
75th Percentile	3.49	1.96
90th Percentile	5.58	4.10
# Efficient observations	31	54

Table 3: Optimal Industry Configuration using the Short Run Johansen Industry Model

Status of Vessel	Number of Vessels	
	Convexity	Non-Convexity
Active	74	80
Partly active	3	3
Not active	45	39

Figure 1: Density of Convex vs. Non-convex Cost under CRS

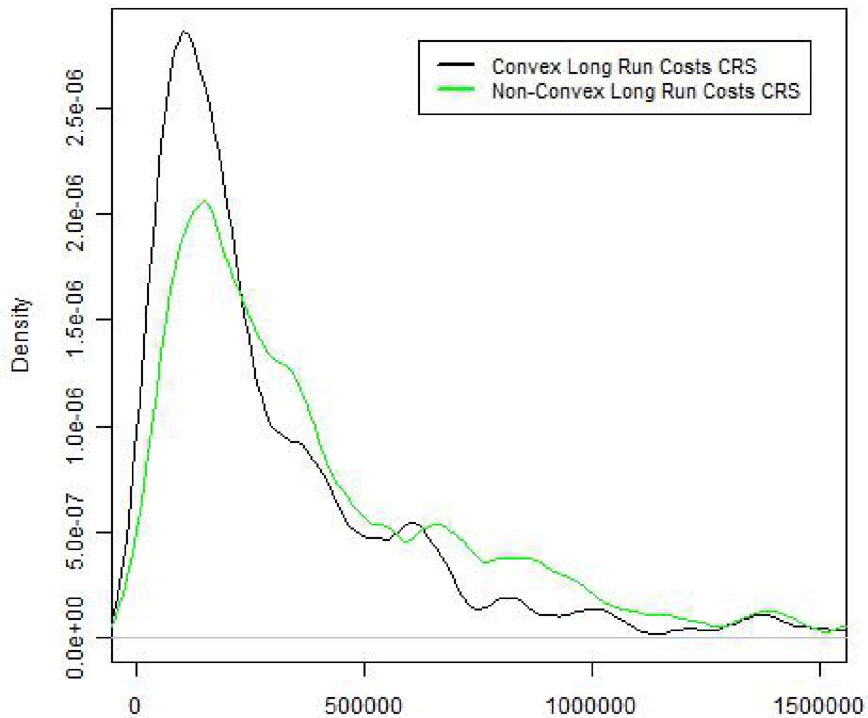


Figure 2: Density of Convex vs. Non-convex Tangency Cost with Modified Fixed Inputs

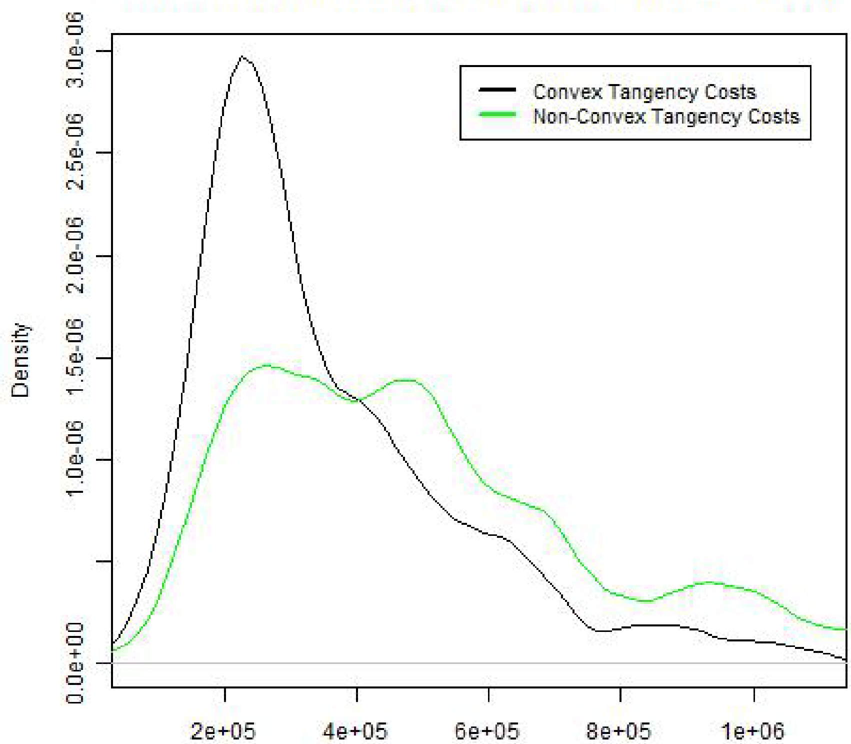
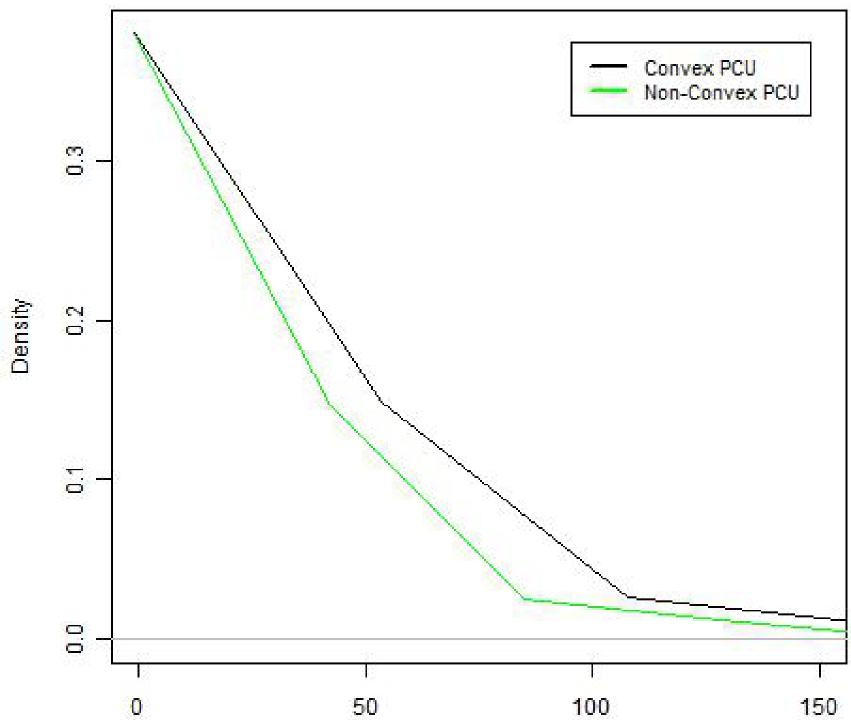


Figure 3: Density of Convex vs. Non-convex Output Efficiency Measure without Limits on Variable Inputs



Econometric Estimates of Productivity and Efficiency Change in the Australian Northern Prawn Fishery[1]

C. J. O'Donnell

Abstract

Bayesian methods are used to compute and decompose Färe-Primont indexes of total factor productivity (TFP) change in the Australian Northern Prawn Fishery (NPF). Färe-Primont indexes are used because i) they satisfy all basic axioms from index number theory, ii) they can be exhaustively decomposed into measures of environmental change and efficiency change (i.e., there are no residual 'effects'), and iii) they can be computed using only quantity data (i.e., no prices are needed). Bayesian estimation methodology is used because i) it can solve an endogeneity problem that arises in the econometric estimation of multiple-input multiple-output distance functions, ii) it can be used to draw valid finite-sample inferences concerning nonlinear functions of the model parameters (e.g., measures of technical efficiency), and iii) non-sample information (e.g., information provided by economic theory) can be easily incorporated into the estimation process. Results for the NPF for the period 1974–2010 are summarised in terms of characteristics of estimated posterior pdfs for measures of TFP change, environmental change, technical efficiency change, and scale efficiency change.

Key words: Total Factor Productivity; Färe-Primont Index; Technical Efficiency; Mix Efficiency; Scale Efficiency; Output Distance Function; Markov Chain Monte Carlo.

1. Introduction

The Australian Northern Prawn Fishery (NPF) is a multi-species fishery covering an area of 771,000 square kilometres off Australia's northern coast. Banana prawns and tiger prawns account for approximately 80% of the landed catch. The total catch of all prawn species peaked at more than 13,800 tonnes in 1974. Since that time, input controls (e.g., spatial closures, gear controls, restrictions on entry and vessel replacement) and changes in environmental conditions have seen the total catch fall by more than 45%. The main aim of this paper is to estimate associated changes in total factor productivity (TFP). A second aim is to decompose these changes into measures of environmental (or "technical") change, technical efficiency change, and scale efficiency change. Early empirical studies of the NPF were only concerned with estimating technical efficiency change [e.g., Kompas, Che, and Grafton (2004)]. Only relatively recently have NPF researchers and policy-makers recognised the importance of also measuring TFP change and other types of efficiency change [e.g., Pascoe et al. (2012)].

Several index formulas are available for measuring TFP change. Irrespective of the index formula chosen, decomposing TFP change into measures of environmental change and efficiency change involves estimating a functional representation of the production technology. Alternative functional representations of multiple-input multiple-output technologies include cost, revenue, profit and distance functions. If no price data are available (as in this paper) then the least restrictive representation is a distance function. This paper represents the NPF production technology using the output and input distance functions of Shephard (1970).

Distance functions are typically estimated using one of two estimation methodologies: stochastic frontier analysis (SFA) or data envelopment analysis (DEA). Environmental variables can be incorporated into SFA analyses by simply treating them like any other exogenous variables. Incorporating environmental variables into DEA analyses is slightly more complicated. To estimate distance functions (equivalently, measures of technical efficiency) using DEA and at the same time allow for changes in the production environment it is necessary to use non-overlapping subsets of observations (i.e., observations peculiar to different production environments) to estimate separate frontiers. If the production environment changes over time and only time-series data are available (as in this paper) then this approach is infeasible.[2] Thus, this paper estimates the production technology using SFA methodology.

In O'Donnell (2012) I identify a class of TFP indexes that can be exhaustively decomposed into various measures of environmental change and efficiency change. Indexes that can be decomposed in this way include Laspeyres, Paasche, Fisher, Törnqvist, Lowe, Hicks-Moorsteen and Färe-Primont TFP indexes. If no price data are available (as in this paper) then the price-based Laspeyres, Paasche, Fisher, Törnqvist and Lowe indexes are unavailable. Of the two remaining indexes, the Hicks-Moorsteen index is unsuitable because it is intransitive. Transitivity means that a direct comparison of two observations will yield the same measure of TFP change as an indirect comparison through a third observation. For example, if TFP increases by 10% between 2000 and 2001, and by another 10% between 2001 and 2002, then the index that directly compares TFP in 2000 and 2002 should indicate that TFP has increased by 21%. The Hicks-Moorsteen TFP index does not generally satisfy this common sense property. This paper uses the Färe-Primont index to measure TFP change because, unlike the Hicks-Moorsteen index, it is transitive.

Within an SFA framework, different econometric estimators are available for estimating (and decomposing) Färe-Primont measures of TFP change. Possible sampling theory estimators include the maximum likelihood (ML) and generalised method of moments (GMM) estimators. These estimators are popular because they have good asymptotic properties (e.g., consistency, asymptotic normality). However, their finite sample properties are generally unknown. Thus, if sample sizes are small (as in this paper), an alternative estimation methodology is generally required. This paper uses a Bayesian estimation approach ostensibly because valid inferences can still be made when sample sizes are small. The Bayesian approach can also be used to overcome an endogeneity problem that arises in the econometric estimation of multiple-input multiple-output distance functions [O'Donnell (2011)].

The structure of the paper is as follows. Section 2 uses Shephard (1970) output and input distance functions to represent a Hicks neutral and homothetic production technology. Section 3 describes a class of output and input quantity indexes that satisfy a number of common sense properties, including transitivity. Ratios of these so-called "proper" indexes can be used to measure TFP change. Section 4 describes one such TFP index—the Färe-Primont index. Section 5 explains how the Färe-Primont TFP index can be exhaustively decomposed into a measure of environmental change and output-oriented measures of technical and scale efficiency change. The decomposition methodology doesn't require any restrictive assumptions concerning the production technology or market structure—this makes it suitable for use in regulated industries (e.g., the NPF). Section 6 develops the econometric model. Section 7 explains how a Bayesian methodology developed by Fernandez, Koop, and Steel (2000) can be used to estimate the unknown model parameters and associated inefficiency effects. Section 8 describes the NPF data used in the empirical work. Section 9 discusses point and interval estimates of the parameters of the distance function and different components of TFP change. The paper is concluded in Section 10.

2. The Production Technology

A common and very general representation of a production technology is the production possibilities set:

(1) $T(z) = \{(x,q) : x \text{ can produce } q \text{ in environment } z\}$

where $x = (x_1, \ldots, x_M)' \in \mathbb{R}^M_+$ is a nonnegative vector of inputs, $q = (q_1, \ldots, q_N)' \in \mathbb{R}^N_+$ is a nonnegative vector of outputs, and $z = (z_1, \ldots, z_J)' \in \mathbb{R}^J_{++}$ is a positive vector of exogenous variables measuring characteristics of the production environment. In this paper I maintain the basic regularity properties of Färe and Primont (1995, pp., 26, 27):

T1: $(x,0) \in T(z)$ for all $x \in \mathbb{R}^M_+$ (inactivity);

T2: $P(x,z) = \{q : (x,q) \in T(z)\}$ is bounded for all $x \in \mathbb{R}^M_+$ (boundedness);

T3: $q \geq 0 \Rightarrow (0,q) \notin T(z)$ (weak essentiality);

T4: $(x,q) \in T(z)$ and $0 \leq \lambda \leq 1 \Rightarrow (x, \lambda q) \in T(z)$ (weak disposability of outputs);

T5: $(x,q) \in T(z)$ and $\lambda \geq 1 \Rightarrow (\lambda x, q) \in T(z)$ (weak disposability of inputs); and

T6: $P(x,z) = \{q : (x,q) \in T(z)\}$ is closed for all $x \in \mathbb{R}^M_+$ and $L(q,z) = \{x : (x,q) \in T(z)\}$ is closed for all $q \in \mathbb{R}^N_+$ (output and input closedness).

Assumption T1 (inactivity) says it is possible to do nothing. T2 (boundedness) says there are limits to what can be produced using a finite amount of inputs. T3 (weak essentiality) says positive outputs cannot be produced without a positive amount of at least one input. T4 (weak disposability of outputs) says that if an input vector can produce a particular output vector then it can also be used to produce a scalar contraction of that output vector (i.e., fewer outputs in the same mix). T5 (weak disposability of inputs) says that if an output vector can be produced using a particular input vector then it can also be produced using a scalar magnification of that input vector (i.e., more inputs in the same mix). Finally, T6 (output and input closedness) is a mathematical property that guarantees the existence of the input distance function of Shephard (1970, pp. 206): $D_I(x,q,z) = \sup\{\rho > 0 : (x/\rho, q) \in T(z)\}$. This function gives the maximum factor by which a firm can radially contract its input vector and still produce the same output vector. If it exists then it is nonnegative (NN) and linearly homogeneous (HD1) in inputs. Assumptions T2 and T6 together guarantee the existence of the Shephard (1970, pp. 207) output distance function: $D_O(x,q,z) = \inf\{\delta > 0 : (x, q/\delta) \in T(z)\}$. The output distance function gives the reciprocal of the largest factor by which a firm can radially expand its output vector while holding its input vector fixed. If it exists then it is NN and HD1 in outputs.

In this paper, a technology is said to be *regular* if T1–T6 hold. If a technology is regular then the output and input distance functions exist. However, if they are to be used to construct meaningful TFP indexes then the following stronger versions of T4 and T5 are required:

T4s: $(x,q) \in T(z)$ and $0 \leq q^1 \leq q \Rightarrow (x,q^1) \in T(z)$ (strong disposability of outputs); and

T5s: $(x,q) \in T(z)$ and $x^1 \geq x \Rightarrow (x^1,q) \in T(z)$ (strong disposability of inputs).

Assumption T4s (strong disposability of outputs) says that it is possible to use the same inputs to produce fewer outputs. If T4s holds then the output distance function is nondecreasing (ND) in outputs. T5s (strong disposability of inputs) says it is possible to produce the same outputs using more inputs. If T5s holds then the input distance function is ND in inputs. Strong disposability implies weak disposability (i.e., T4s⇒T4 and T5s⇒T5).

Assumptions T1–T6, T4s and T5s are enough to construct a meaningful TFP index and decompose it into a measure of environmental change and various measures of efficiency change (details are provided in Section 4). However, for a particularly simple decomposition, in this paper I also assume

HD*r*: $(x,q) \in T(z) \Leftrightarrow (\lambda x, \lambda^r q) \in T(z)$ for all $\lambda > 0$ (homogeneity of degree *r*),

EOH: $D_O(x,q,z) = g(\mu_x) D_O(\mu_x,q,z)/g(x)$ (extended output homotheticity),

EIH: $D_I(x,q,z) = h(\mu_q) D_I(x,\mu_q,z)/h(q)$ (extended input homotheticity),

EHON: $D_O(x,q,z) = b(\mu_z) D_O(x,q,\mu_z)/b(z)$ (extended Hicks output neutrality) and

EHIN: $D_I(x,q,z) = a(z) D_I(x,q,\mu_z)/a(\mu_z)$ (extended Hicks input neutrality)

where $h(.)$, $g(.)$, $b(.)$ and $a(.)$ are scalar-valued functions with properties that are consistent with the properties of the respective distance functions [e.g., T4s means $h(.)$ must be nondecreasing]. The HDr property says that a one percent increase in inputs provides for an r percent increase in outputs. The technology is said to exhibit decreasing returns to scale (DRS), constant returns to scale (CRS) or increasing returns to scale (IRS) as r is less than, equal to, or greater than one. The homotheticity and Hicks neutrality properties have important implications for marginal rates of technical transformation (MRTTs) and marginal rates of technical substitution (MRTSs): if a regular technology is EOH then MRTSs are independent of output quantities and environmental variables; if the technology is EIH then MRTTs are independent of input quantities and environmental variables; and if the technology is either EHON or EHIN then both MRTSs and MRTTs are independent of environmental variables. In this paper, a regular technology is said to be extended homothetic (EH) if and only if it is both EOH and EIH. It is also said to be extended Hicks neutral (EHN) if and only if it is both EHON and EHIN. In the single output case, a regular production technology is EHN if and only if the production function can be written $F(x,z) = b(z)\bar{F}(x)$ where $\bar{F}(x) = 1/D_O(x, b(\mu_z), \mu_z)$ (Appendix, Propositions D1 and D2). This corresponds to the definition of EHN in Blackorby, Lovell, and Thursby (1976, p. 848, Lemma 3).

If a regular technology is EH, EHN and HDr (as assumed in this paper) then and only then the output and input distance functions take the form (Appendix, Propositions D8 and D9):

(2) $D_O(x,q,z) \propto h(q)^r / \left[b(z)g(x)\right]$

and

(3) $D_I(x,q,z) \propto \left[b(z)g(x)\right]^{1/r} / h(q)$

where $h(.)$ is NN, ND and homogeneous of degree $1/r$ and $g(.)$ is NN, ND and HDr.

3. Proper Output and Input Quantity Indexes

It is convenient at this point to introduce a time subscript t into the notation so that, for example, $x_t = (x_{1t}, \ldots, x_{Mt})'$ and $q_t = (q_{1t}, \ldots, q_{Nt})'$ represent the input and output quantity vectors in period t. An index that compares q_t with q_s using the latter as the reference (or base) vector is any variable of the form [O'Donnell (2012)]

$$(4) \qquad QI_{st} \equiv \frac{Q(q_t)}{Q(q_s)}$$

where $Q(.)$ is an NN, ND and HD1 scalar aggregator function. If $Q(.)$ is differentiable then $Q(q_t) = \sum_n a_{nt} q_{nt}$ where $a_{nt} \equiv \partial Q(q_t)/\partial q_{nt} \geq 0$. It follows that the index (4) can be written

$$(5) \qquad QI_{st} = \frac{\sum_n a_{nt} q_{nt}}{\sum_n a_{ns} q_{ns}} \equiv QI(q_s, q_t, a_s, a_t)$$

where $a_t = (a_{1t}, \ldots, a_{Nt})' \geqq 0$ can be interpreted as a vector of weights measuring the relative importance of different outputs to the decision maker. Moreover, it is easily shown that:[3]

Q1: $q_r \geq q_t \Rightarrow QI_{sr} \geq Q_{st}$ and $q_r \geq q_s \Rightarrow QI_{rt} \leq Q_{st}$ (weak monotonicity);

Q2: $QI(q_s, \lambda q_t, a_s, a_t) = \lambda QI(q_s, q_t, a_s, a_t)$ for $\lambda > 0$ (linear homogeneity);

Q3: $QI(q_t, q_t, a_t, a_t) = Q_{tt} = 1$ (identity);

Q4: $QI(\lambda q_s, \lambda q_t, a_s, a_t) = QI(q_s, q_t, a_s, a_t)$ for $\lambda > 0$ (homogeneity of degree 0);

Q5: $QI(\Lambda q_s, \Lambda q_t, \Lambda^{-1} a_s, \Lambda^{-1} a_t) = QI(q_s, q_t, a_s, a_t)$ where Λ is a diagonal matrix with diagonal elements strictly greater than zero (commensurability);

Q6: $QI(q_s, \lambda q_s, a_s, a_s) = \lambda$ for $\lambda > 0$ (proportionality); and

Q7: $QI_{st} = QI_{sr} QI_{rt}$ (transitivity).

Property Q1 (weak monotonicity) says that the index will not decrease with i) an increase in any element of the comparison vector and/or ii) a decrease in any element of the base vector. Q2 (linear homogeneity) says that a proportionate increase in the comparison vector will translate into the same proportionate increase in the index. Q3 (identity) says that if the comparison and base vectors are identical then the index takes the value one. Q4 (homogeneity of degree zero) says that if the comparison and base vectors are multiplied by the same constant then the index doesn't change. Q5 (commensurability) says that the index is robust to changes in units of measurement (i.e., the index doesn't change if, for example, a variable is measured in tonnes instead of kilograms). Q6 (proportionality) says that if all variables change λ-fold then the index that measures this change is equal to λ. Q7 (transitivity) says that the index that directly compares two observations is equal to the index obtained when the comparison is made indirectly via a third observation. In this paper, an output quantity index is said to be *proper* if and only if it satisfies Q1–Q7. If there are only two observations in the dataset then the transitivity property Q7 is redundant. Thus, an output quantity index for binary comparisons (i.e., a comparison involving only two observations) is proper if and only if it satisfies Q1–Q6.

Input quantity indexes have a similar structure. Specifically, an index that compares x_t with x_s using the latter as the base is any variable of the form [O'Donnell (2012)]

$$(6) \qquad XI_{st} \equiv \frac{X(x_t)}{X(x_s)}$$

where $X(.)$ is an NN, ND and HD1 scalar aggregator function. If $X(.)$ is differentiable then

$$(7) \qquad XI_{st} = \frac{\sum_m b_{mt} x_{mt}}{\sum_m b_{ms} x_{ms}} \equiv XI(x_s, x_t, b_s, b_t)$$

where $b_{mt} \equiv \partial X(x_t)/\partial x_{nt} \geq 0$ and $b_t = (b_{1t}, \ldots, b_{Nt})' \geqq 0$. This index has properties that are analogous to those of the output index:

X1: $x_r \geq x_t \Rightarrow XI_{sr} \geq X_{st}$ and $x_r \geq x_s \Rightarrow XI_{rt} \leq X_{st}$ (weak monotonicity);

X2: $XI(x_s, \lambda x_t, b_s, b_t) = \lambda XI(x_s, x_t, b_s, b_t)$ for $\lambda > 0$ (linear homogeneity);

X3: $XI(x_t, x_t, b_t, b_t) = X_{tt} = 1$ (identity);

X4: $XI(\lambda x_s, \lambda x_t, b_s, b_t) = XI(x_s, x_t, b_s, b_t)$ for $\lambda > 0$ (homogeneity of degree 0);

X5: $XI(\Lambda x_s, \Lambda x_t, \Lambda^{-1} b_s, \Lambda^{-1} b_t) = XI(x_s, x_t, b_s, b_t)$ where Λ is a diagonal matrix with diagonal elements strictly greater than zero (commensurability);

X6: $XI(x_s, \lambda x_s, b_s, b_s) = \lambda$ for $\lambda > 0$ (proportionality); and

X7: $XI_{st} = XI_{sr} XI_{rt}$ (transitivity).

Again, an input quantity index for multiple comparisons is said to be proper if and only if it satisfies X1–X7. An input quantity index for binary comparisons is proper if and only if it satisfies X1–X6.

4. TFP Indexes

In O'Donnell (2012) I define the TFP of the firm to be $TFP_t = Q_t/X_t$ where $Q_t \equiv Q(q_t)$ and $X_t \equiv X(x_t)$. It follows that the index that compares TFP in period t with TFP in period s is

$$(8) \qquad TFPI_{st} = \frac{TFP_t}{TFP_s} = \frac{Q_t/X_t}{Q_s/X_s} = \frac{Q_t/Q_s}{X_t/X_s} = \frac{QI_{st}}{XI_{st}}$$

where QI_{st} and XI_{st} are the proper output and input quantity indexes defined in Section 3. In this paper, a TFP index is said to be proper if and only if it can be written as the ratio of a proper output index to a proper input index.

Any NN, ND and HD1 function can be used as an aggregator function for purposes of constructing proper output, input and TFP indexes. If T1–T6, T4s and T5s hold then the output (input) distance function is NN, ND and HD1 in outputs (inputs). Thus, suitable output and input aggregator functions are $Q(q_t) \propto D_O(\mu_x, q_t, \mu_z)$ and $X(x_t) \propto D_I(x_t, \mu_q, \mu_z)$ where $\mu_x \in \mathbb{R}^M_+$, $\mu_q \in \mathbb{R}^N_+$ and $\mu_z \in \mathbb{R}^J_{++}$ are arbitrary. The associated output, input and TFP indexes are

$$(9) \qquad QI_{st} = \frac{D_O(\mu_x, q_t, \mu_z)}{D_O(\mu_x, q_s, \mu_z)},$$

$$(10) \quad XI_{st} = \frac{D_I(x_t, \mu_q, \mu_z)}{D_O(x_s, \mu_q, \mu_z)}$$

and

$$(11) \quad TFPI_{st} = \frac{D_O(\mu_x, q_t, \mu_z)}{D_O(\mu_x, q_s, \mu_z)} \frac{D_I(x_s, \mu_q, \mu_z)}{D_O(x_t, \mu_q, \mu_z)}.$$

In this paper I refer to the TFP index (11) as a Färe-Primont index because the component indexes (9) and (10) can be traced back at least as far as Färe and Primont (1995, pp. 36, 38). In practice, evaluating Färe-Primont indexes usually involves choosing vectors μ_x, μ_q and μ_z that are relevant to all the observations that are being compared (e.g., if comparisons are being made between all the observations in a dataset then any measure of central tendency would be suitable). However, different assumptions concerning the production technology can sometimes obviate the need to identify vectors μ_x, μ_q and μ_z that are "relevant". For example, if the technology is EH, EHN and HD*r* then the output and input distance functions are given by (2) and (3) and the TFP index (11) takes the form

$$(12) \quad TFPI_{st} = \frac{h(q_t)^r}{h(q_s)^r} \frac{g(x_s)^{1/r}}{g(x_t)^{1/r}}.$$

Observe that this index does not depend on μ_x, μ_q or μ_z.

5. The Components of TFP Change

In O'Donnell (2010, 2012) I explain how TFP indexes that take the form (8) can be decomposed into a measure of technical (or environmental) change and various measures of efficiency change. A distinctive feature of the methodology is the way in which measures of efficiency are expressed in terms of aggregate quantities. For example, for a firm operating in a production environment characterised by z_t, measures of technical, mix and scale efficiency include

$$(13) \quad OTE_t = \frac{Q_t}{\bar{Q}_t} = D_O(x_t, q_t, z_t)$$

$$(14) \quad OME_t = \frac{\bar{Q}_t}{\hat{Q}_t}$$

(15) $OSE_t = \frac{\bar{Q}_t/X_t}{\tilde{Q}_t/\tilde{X}_t}$ and

(16) $OSME_t = \frac{\bar{Q}_t/X_t}{TFP_t^*}$

where $\bar{Q}_t \equiv Q_t D_O(x_t, q_t, z_t)^{-1}$ is the maximum aggregate output possible when using x_t to produce a scalar multiple of q_t, $\hat{Q}_t$ is the maximum aggregate output possible when using x_t to produce *any* output vector, $\tilde{Q}_t$ and $\tilde{X}_t$ are the aggregate output and input obtained when TFP is maximized subject to the constraint that the output and input vectors are scalar multiples of q_t and x_t respectively, and TFP_t^* is the maximum TFP possible (in an environment characterised by z_t). Output-oriented technical efficiency (OTE) is a measure of the increase in TFP that can be obtained when holding the input vector and the output mix fixed. Output-oriented mix efficiency (OME) is a measure of the increase in TFP that can be obtained through economies of scope. Output-oriented scale efficiency (OSE) is a measure of the increase in TFP possible through economies of scale. Finally, output-oriented scale-mix efficiency (OSME) is a combined measure of the increase in TFP that can be achieved through economies of both scale and scope. More details concerning these and related measures of efficiency can be found in O'Donnell (2010, 2012).

Equations (13)–(16) can be used to decompose the TFP index (8) into several economically-meaningful components. For example, using (13) and (16):

(17) $TFPI_{st} = \left(\frac{TFP_t^*}{TFP_s^*}\right)\left(\frac{OTE_t}{OTE_s}\right)\left(\frac{OSME_t}{OSME_s}\right).$

The term TFP_t^*/TFP_s^* compares the maximum TFP possible in an environment characterised by z_t with the maximum TFP possible in an environment characterised by z_s. This is a natural measure of environmental change. The remaining components in (17) are output-oriented measures of technical efficiency change and scale-mix efficiency change. Thus, equation (17) reveals that TFP change is driven by three intrinsically different components: an environmental change component that measures movements in the production frontier; a technical efficiency change component that measures movements towards or away from the frontier; and a scale-mix efficiency change component that measures movements around the frontier to capture economies of scale and scope.

Different aggregator functions and different assumptions concerning the production technology typically give rise to different measures of (the components of) TFP change. For example, if the aggregator functions are $Q(q_t) \propto D_O(\mu_x, q_t, \mu_z)$ and $X(x_t) \propto D_I(x_t, \mu_q, \mu_z)$ and if the technology is EH, EHN and HDr then the TFP index is given by (12). In this case the decomposition (17) takes the specific form

$$(18) \quad TFPI_{st} = \left(\frac{b(z_t)}{b(z_s)}\right) \left(\frac{h(q_t)^r}{b(z_t)g(x_t)} \frac{b(z_s)g(x_s)}{h(q_s)^r}\right) \left(\frac{g(x_s)}{g(x_t)}\right)^{(1-r)/r}.$$

The first component is still a measure of environmental change, the second component is still a measure of output-oriented technical efficiency change, but the last component is in fact a measure of pure scale efficiency change. There is no mix efficiency change component in this decomposition because the output and input aggregator functions are proportional to the output and input distance functions (O'Donnell 2012, Section 3.7). If the technology exhibits CRS then $r = 1$ and the last component vanishes.

6. The Econometric Model

To derive an econometric model it is convenient to first define

$$(19) \quad v_{dt} \equiv r \ln h(q_t) - \ln b(z_t) - \ln g(x_t) - \ln D_O(x_t, q_t, z_t)$$

$$(20) \quad v_{xt} \equiv \ln g(x_t) - \sum_{m=1}^{M} \beta_m \ln x_{mt}$$

$$(21) \quad v_{zt} \equiv \ln b(z_t) - \sum_{j=1}^{J} \gamma_j \ln z_{jt} - \gamma_0 \qquad \text{and}$$

$$(22) \quad v_{qt} \equiv \ln\left(\sum_{n=1}^{N} \alpha_n q_{nt}\right) - r \ln h(q_t)$$

where $\alpha_n \geq 0$, $\sum_n \alpha_n = 1$, $\beta_m \geq 0$ and $\sum_m \beta_m = r$. The variable v_{dt} can be viewed as a specification error that vanishes if the technology is EH, EHN and HDr. Similarly, the variables v_{xt}, v_{zt} and v_{qt} can be viewed as errors that vanish if the associated functions $g(.)$, $b(.)$ and $h(.)$ are in fact Cobb-Douglas and linear. If the error terms vanish then the inequality constraints $\alpha_n \geq 0$ and $\beta_m \geq 0$ are needed to ensure that the technology satisfies T1–T6. Other inequality constraints may be appropriate in some empirical contexts. For example, all other things being equal, if the j-th environmental variable has a nonnegative effect on output then $\gamma_j \geq 0$.

Equations (19)–(22) are definitions, not assumptions. Thus, without making any assumptions concerning firm optimising behaviour or the functional form of the output distance function, it is possible to write

$$(23) \quad \ln Q_t = \gamma_0 + \sum_{j=1}^{J} \gamma_j \ln z_{jt} + \sum_{m=1}^{M} \beta_m \ln x_{mt} + v_t - u_t$$

where $Q_t \equiv \sum_n \alpha_n q_{nt}$ can be viewed as an aggregate output, $v_t \equiv v_{dt} + v_{zt} + v_{xt} + v_{qt}$ represents approximation errors but may also subsume other sources of statistical noise (e.g., measurement errors), and $u_t \equiv -\ln D_O(x_t, q_t, z_t) \geq 0$ is an output-oriented technical inefficiency effect. Equation (23) is in the form of the stochastic frontier model of Aigner, Lovell, and Schmidt (1977). The full set of T observations in the dataset can be written

$$(24) \quad y = X\beta + v - u$$

where $y = (y_1, \ldots, y_T)'$ and $y_t \equiv \ln Q_t$. The remaining definitions are obvious, although it is worth noting that X is $T \times (J + M + 1)$.

To estimate the unknown parameters is it necessary to make some assumptions concerning the error terms. In this paper I follow Aigner, Lovell, and Schmidt (1977, pp. 24, 29) and assume the idiosyncratic errors (v_t) and the inefficiency errors (u_t) are independent and identically distributed (iid) normal and exponential random variables respectively. Under these assumptions, ML estimation would be straightforward if it were not for the fact that if $N \geq 2$ the dependent variable is unobserved. In this paper I solve the problem using the Bayesian methodology of Fernandez, Koop, and Steel (2000). This involves sampling from the joint posterior probability density function (pdf) of the unknown parameters and unobserved inefficiency effects. The next section describes the conditional likelihood function, prior pdf, and conditional posterior pdfs needed to implement the sampling algorithm.

7. Bayesian Estimation

If the idiosyncratic errors are iid normally distributed with mean zero and precision h then the conditional joint density for the unobserved dependent variable vector is[4]

(25) $p(y|\beta,u,h) = f_N(y|X\beta - u, h^{-1}I_T)$

where I_T denotes an identity matrix of order T. Unfortunately, this T-variate density is not enough to define a sampling density for the $N \times T$ matrix of observed outputs $Q = (q_1,\ldots,q_T)$. To overcome this problem I follow Fernandez, Koop, and Steel (2000) and introduce $N-1$ new random variables into the model to generate stochastics in another $N-1$ dimensions. Specifically, I introduce the unobserved shadow revenue shares

(26) $r^*_{nt} = \dfrac{\partial \ln D_O(x_t,q_t,z_t)}{\partial \ln q_{nt}} = \dfrac{\alpha_n q_{nt}}{\sum_{k=1}^N \alpha_k q_{kt}}$

for $n = 1,\ldots,N$. These unobserved shares sum to one and so it is natural to assume $r^*_t = (r^*_{1t},\ldots,r^*_{Nt})'$ is iid with a Dirichlet pdf:[5] $p(r^*_t|s) = f_D(r^*_t|s)$ where $s = (s_1,\ldots,s_N)' \in \Re^N_+$. Given $\alpha = (\alpha_1,\ldots,\alpha_N)'$ there is a one-to-one mapping between the observed output vector $q_t \in \Re^N_+$ and the unobserved vector $(y_t, r^*_{2t},\ldots,r^*_{Nt})'$. Thus, the conditional likelihood function for the matrix of observed outputs $Q = (q_1,\ldots,q_T)$ is (Fernandez, Koop, and Steel 2000, p. 55, eq. 2.7):

(27) $p(Q|\alpha,\beta,h,s,u) = f_N(y|X\beta - u, h^{-1}I_T)\prod_{t=1}^T f_D(r^*_t|s)\prod_{t=1}^T |J_t|$

where $|J_t| = \prod_n (r^*_{nt}/q_{nt})$ is the absolute value of the Jacobian of the transformation from $(y_t, r^*_{2t},\ldots,r^*_{Nt})'$ to q_t. I also follow Fernandez, Koop, and Steel (2000) and specify a prior pdf of the form $p(\alpha,\beta,h,s,u) = p(\alpha)p(\beta)p(h)p(s)p(u)$ where each of the component priors is proper:[6]

(28) $p(\alpha) = f_D(\alpha|\iota_N)$

(29) $p(\beta) = f_N(\beta|0_K, k_2 I_K)I(\beta \in R)$

(30) $p(h) = f_G(h|1, k_1)$

(31) $p(s) = \prod_{n=1}^{N} f_G(s_n|1, k_3)$

(32) $p(u|\lambda) = \prod_{t=1}^{T} f_G(u_t|1, \lambda)$ and

(33) $p(\lambda) = f_G(\lambda|1, -\ln(\tau))$

where $I(.)$ is an indicator function that takes the value one if the argument is true and zero otherwise, 0_M is a $M \times 1$ zero vector, ι_N is an $N \times 1$ unit vector, R is the region of the parameters space where constraints of the type discussed at the beginning of Section 5 are satisfied, and $K = J + M + 1$. For the empirical work in this paper I set $k_1 = 10^{-4}$, $k_2 = 10^4$ and $k_3 = 10^{-2}$ to ensure the priors for β, h and s are relatively noninformative. The pdf (33) is centred on $-\ln(\tau)$ where $\tau = 0.9$ (a prior estimate of the average level of efficiency).

The prior pdf combines with the conditional likelihood function (27) to yield a joint posterior pdf for the unknown parameters and the unobserved inefficiency effects. Characteristics of marginal posterior pdfs (e.g., means and variances) are obtained by integrating this joint posterior. Unfortunately, analytical integration is impossible. In this paper, integration is conducted using a Markov Chain Monte Carlo (MCMC) sampling algorithm—the Gibbs sampler. The Gibbs sampler involves partitioning the vector of unknown parameters and inefficiency effects into blocks, then simulating from the conditional posterior pdf for each block. For details on the Gibbs sampler see Casella and George (1992) and Koop (2003).

In this paper, the conditional posterior pdfs needed to make the Gibbs sampler operational are:

$$(34)\quad p(h|\alpha,\beta,s,\lambda,u,Q) \propto f_G(h|1+0.5T, k_1+0.5e'e)$$

$$(35)\quad p(\lambda|\alpha,\beta,h,s,u,Q) \propto f_G(\lambda|T+1, u'\iota_T - \ln(\tau))$$

$$(36)\quad p(\beta|\alpha,h,s,\lambda,u,Q) \propto f_N(\beta|hVX'(y+u), V)I(\beta \in R)$$

$$(37)\quad p(u|\alpha,\beta,h,s,\lambda,Q) \propto f_N(u|X\beta - y - h^{-1}\lambda^{-1}\iota_T, h^{-1}I_T)I(u \geq 0_T)$$

$$(38)\quad p(s_n|\alpha,\beta,h,s_{-n},\lambda,u,Q)$$
$$\propto \Gamma\left(\sum_{k=1}^{N} s_k\right)^T \Gamma(s_n)^{-T} \exp\left(-s_n\left[k_1 - \sum_{t=1}^{T}\ln r^*_{nt}\right]\right)I(s_n \geq 0)$$

and

$$(39)\quad p(\alpha|\beta,h,s,\lambda,u,Q)$$
$$\propto \prod_{n=1}^{N} \alpha_n^{s_n\theta T} \prod_{t=1}^{T}\left(\sum_{n=1}^{N}\alpha_n q_{nt}\right)^{-\sum_{n=1}^{N} s_n} \exp(-0.5h^{-1}e'e)I(\alpha \in R)$$

where $e \equiv y - X\beta + u$; $V \equiv (hX'X + k_2^{-1}I_K)^{-1}$; and s_{-n} is the vector comprising all the elements of s except s_n. Simulating from (34) and (35) is straightforward using built-in functions in common computer packages. Simulating from the remaining pdfs is slightly more complicated. In this paper, a Metropolis-Hastings (MH) algorithm was used to simulate from (36)–(38). For details on the MH algorithm see Chib and Greenberg (1996) and Koop (2003). When simulating from (38) it was necessary to constrain $\sum_k s_k < 150$ to avoid overflow errors in the evaluation of $\Gamma(\sum_k s_k)$. Finally, simulating from (39) involved drawing $N-1$ elements of α, computing the N-th element from the adding up constraint $\sum_n \alpha_n = 1$, then rejecting the entire vector if any elements fell outside the unit interval.

8. Data

The dataset comprises $T = 37$ annual observations on $N = 4$ outputs, $M = 3$ inputs and $J = 4$ environmental variables. The sample period extends from 1974 to 2010. Brief variable descriptions and summary statistics are provided in Table 1. Data on outputs, inputs, vessel length and engine size were provided by the Australian Fisheries Management Authority (AFMA). Most of these data were originally sourced from daily logbooks and seasonal vessel returns. Data on the Southern Oscillation Index (SOI) were obtained from the Bureau of Meteorology.

Observe from Table 1 that the prawn harvest has been disaggregated by major species type (banana, tiger, endeavour, king). This is partly because the NPF fleet tends to target different species at different times of the year (i.e., production is nonjoint). For example, the fleet usually targets banana prawns from the start of April until the start of June, and tiger prawns from the start of August until the end of November. Another reason for disaggregating total prawn output by species is that different trawling methods are used for different species types. For example, banana prawns are found in dense concentrations in water that is approximately ten metres deep. They emerge from the mud on the sea bed early in the morning and create 'mud boils' about the size of a tennis court. Light aircraft are used to spot these mud boils and pass the GPS co-ordinates to vessels in the fleet. Vessels then 'sound' the location and identify a mark to 'shoot'. In contrast, tiger prawns are nocturnal, so vessels trawl for tiger prawns at night. Turtle excluding devices (TEDs) and by-catch reduction devices (BRDs) are used on each net to minimise the impact of tiger prawn fishing on non-target species (mainly turtles, dolphins and sharks). Observe from Table 1 that trawl effort has been disaggregated into banana fishery effort and tiger fishery effort. This allows for the fact that harvesting tiger prawns involves significantly more trawl effort than target fishing of banana prawns.

Other variables used in the analysis include the number of vessels, average vessel length and average engine size. Whether these variables should be treated as inputs or environmental variables is debatable. In this paper, the number of vessels is treated as an input. Vessel length and engine size are treated as characteristics of the environment (skippers working on small vessels with low engine power are viewed as operating in a different production environment to skippers working on large vessels with more engine power). All three variables are expected to have a positive effect on output and so all three associated coefficients are constrained to be nonnegative. Thus, irrespective of whether they are treated as inputs or environmental variables, all vessel-related variables enter the estimating equation (23) in logarithmic form with nonnegative coefficients.

Two other environmental variables are included in the model: the SOI is included because it is highly correlated with rainfall and, in turn, seasonal variations in rainfall are believed to be strongly linked to seasonal variations in prawn stocks [Staples and Vance (1986)]; and a time trend is included to account for any omitted variables that vary systematically over time.

Finally, data on several important variables were unavailable (e.g., communications equipment, headrope length, number of aircraft). Anecdotal evidence suggests that many of these omitted variables are highly correlated with included inputs (e.g., the number of aircraft is correlated with banana effort days). Thus, the omission of these variables is not expected to significantly bias the results.

9. Results

The MCMC sampling methods described in Section 6 were used to obtain 1.1 million draws on the unknown parameters and inefficiency effects. The first 100,000 draws were discarded as a "burn-in" and 10,000 of the remaining draws were retained (using a 1-in-100 systematic sampling scheme) for purposes of inference. Representative chains are plotted in Figure 1 and are clearly stationary. The estimates reported in this section are the means of (functions of) these sample observations.

Table 2 reports estimated characteristics of the marginal posterior pdfs of key parameters. The columns labelled "2.5%" and "97.5%" are estimated lower and upper bounds of 95% highest posterior density (HPD) intervals.[7] The prior pdf incorporates the inequality constraints discussed in Sections 5 and 7: $\alpha_n \geq 0$ for $n = 1, \ldots, 4$; $\beta_m \geq 0$ for $m = 1, 2, 3$; and $\gamma_j \geq 0$ for $j = 3, 4$. Thus, estimates of these parameters are guaranteed to be "correctly" signed. The relative magnitudes of $\hat{\alpha}_1, \ldots, \hat{\alpha}_4$ indicate that it is difficult for technically-efficient vessels to substitute banana prawns for other prawn species.[8] This reflects the fact that banana prawns are generally harvested at a different time of the year and using different trawling techniques to other species. The estimated coefficient of the time trend ($\hat{\gamma}_1 = -0.019$) indicates that the production environment has been deteriorating at an average rate of 1.9% p.a. due to factors that could not be included in the analysis. This is consistent with stock assessments that show significant depletion of prawn stocks prior to the introduction of input controls in the second half of the sample period. Tighter controls on omitted inputs (e.g., a 25% reduction in total allowable headrope length in 2002) may themselves help to explain a deterioration in the "production environment". The estimated elasticity of scale ($\hat{\eta} = 1.651$) indicates that, in any given production environment, fishers experience significant increasing returns to scale. Management controls designed to ensure long-term sustainability of the fishery currently limit the ability of fishers to operate on a scale that would maximise these returns (in any given year).

Table 3 reports estimates of TFP change (ΔTFP), environmental change (ΔENV), output-oriented technical efficiency change (ΔOTE) and output-oriented scale efficiency change (ΔOSE) over the sample period. These are the environmental and efficiency change components in (18). These estimates are also plotted in Figure 2. The interpretation of these results is straightforward. For example, the estimates reported in the last row of Table 3 reveal that TFP in 2010 was nearly three times higher than it had been in 1974 due to the combined effects of a $(4.21-1)=3.21=321\%$ improvement in environmental conditions, a 21% increase in technical efficiency, and a 44% fall in scale efficiency (i.e., $\Delta\text{TFP} = \Delta\text{ENV} \times \Delta\text{OTE} \times \Delta\text{OSE} = 4.21 \times 1.21 \times 0.56 = 2.86$). Further insights into these changes can be obtained from Figures 3 and 4.

Figure 3 shows how variations in the production environment (ΔENV) can be attributed to changes in omitted variables that vary systematically over time (Δtime), changes in the Southern Oscillation Index (ΔSOI), and variations in average vessel length and engine size (Δlength and Δengine). The numbers behind this figure reveal that environmental conditions in 2010 were 4.2 times better than they had been in 1974 due mainly to a 177% increase in average vessel length and a 197% increase in average engine size (i.e., $\Delta\text{ENV} = \Delta\text{time} \times \Delta\text{SOI} \times \Delta\text{length} \times \Delta\text{engine} = 0.51 \times 1.00 \times 2.77 \times 2.97 = 4.21$). Steady increases in average vessel length and engine size over the sample period are a rational response to the introduction of input controls in an industry that everywhere operates in a region of increasing returns to scale ($\hat{\eta} = 1.651$).

Figure 4 illustrates the relationship between scale efficiency and input use. The estimated technology everywhere exhibits increasing returns to scale, so scale efficiency is an increasing function of aggregate input use. It is evident from Figure 4 that aggregate input use (and scale efficiency) in the NPF has fallen steadily since the introduction of input controls in the late 1970s. These input controls have included vessel replacement controls (since 1980), vessel buy-back schemes (since the mid-1980s), bans on daylight trawling (since 1987), and time and area closures (since 2002).

One of the advantages of the Bayesian approach is that it is possible to draw valid finite-sample inferences concerning nonlinear functions of the unknown parameters (e.g., the environmental change and efficiency change components of TFP change). To illustrate, Figure 5 presents point and interval estimates of output-oriented technical efficiency levels over the sample period (again, the labels "2.5%" and "97.5%" refer to the estimated lower and upper bounds of 95% HPD intervals). Observe that the HPD intervals are unusually wide. This indicates that the the dataset (comprising only $T = 37$ time-series observations) conveys an unusually small amount of information about the value of the output distance function (a complicated nonlinear function of $M + N + J + 1 = 12$ parameters). Another noticeable feature of Figure 5 is that the technical efficiency estimates are generally much lower than ML estimates reported in earlier studies [the arithmetic average of the estimates reported in Figure 5 is less than 0.5; Pascoe et al. (2012) estimate the level of OTE to be 0.8 in 2007; Kompas, Che, and Grafton (2004) report an average OTE score of 0.725 for the period 1990–1996]. This is partly due to the fact that most of the estimated marginal posterior pdfs are skewed to the left and so the means (i.e., the estimates depicted in Figure 5) are less than the modes (i.e., the values that maximise the likelihood). To give a better sense of this effect, the bottom-right-hand panel in Figure 6 presents the estimated posterior pdf for the level of technical efficiency in 1995. The mean of this estimated posterior pdf is 0.59 but the mode is closer to 0.8.

Finally, Figure 6 also presents estimated posterior pdfs for the elasticity of scale (η) and measures of productivity change (ΔTFP) and scale efficiency change (ΔOSE) over the sample period. Again, the relatively large variances of these estimated pdfs reflect the relatively small amount of information contained in the data. The estimated variances would have been larger if binding non-sample information (i.e., the inequality constraints on the parameters) had not been included in the estimation process.

10. Conclusion

The NPF is one of Australia's most valuable trawl fisheries. The Australian Fisheries Management Authority (AFMA) manages the fishery with the aim of maximising sustainable economic returns. Important input controls include area closures (e.g., in 2002–2004 the season was shortened to only 134 days) and gear restrictions (e.g., in 2005 the allowable headrope length was reduced by 25%). Input controls and vessel buyback schemes have together seen fleet numbers fall from more than 250 in the early 1980s to 52 in 2010. The main aim of this paper has been to identify associated changes in TFP.

Several methods can be used to measure TFP change. The method used in this paper was more or less dictated by the type of data that were available. The dataset consisted of a mere $T = 37$ time-series observations on $N = 4$ output quantities, $M = 3$ input quantities, and $J = 4$ characteristics of the production environment. Lack of price data ruled out the use of price-based TFP indexes (e.g., Fisher, Törnqvist, Lowe), lack of cross-section data ruled out the use of nonparametric estimation methods (e.g., DEA, FDH), and the small number of observations ruled out the use of econometric estimators that only have asymptotic (i.e., large sample) justification (e.g., maximum likelihood). In this paper, the most sensible (and possibly the most difficult) way forward involved using Bayesian econometric methods to estimate a Färe-Primont TFP index. Unlike the sampling theory approach to inference, the Bayesian approach can be used to make valid finite-sample inferences concerning nonlinear functions of the model parameters (e.g., measures of TFP change). Unlike the well-known Fisher, Törnqvist and Hicks-Moorsteen TFP indexes, the Färe-Primont index is proper in the sense that the component output and input quantity indexes satisfy all economically-relevant axioms from index number theory.

The main results were summarised in terms of characteristics (e.g., means, standard deviations) of estimated posterior pdfs for measures of TFP change, environmental change, technical efficiency change, and scale efficiency change. Between 1974 and 2010, productivity in the fishery is estimated to have increased by 186% due to the combined effects of a 321% improvement in the production environment, a 21% increase in output-oriented technical efficiency, and a 44% fall in output-oriented scale efficiency (i.e., $\Delta\text{TFP} = \Delta\text{ENV} \times \Delta\text{OTE} \times \Delta\text{OSE} = 4.21 \times 1.21 \times 0.56 = 2.86$). Improvements in the “production environment” were associated with increases in average vessel length and engine size (skippers working on large vessels with significant engine power are today working in a better operating environment than skippers who worked on smaller vessels in 1974). The reduction in scale efficiency was attributed to downsizing in an industry that experiences increasing returns to scale (the estimated elasticity of scale was 1.651). Policy-makers will be interested (and possibly pleased) to know that variations in the components of TFP change were strongly linked to changes in fishery management.

Appendix

Proposition D1: If the output distance function exists and the technology is EHON ($\Leftarrow$ EHN) then the production function can be written $F(x,z) = b(z)/D_O(x,b(\mu_z),\mu_z)$.

Proof: If the output distance function exists then $F(x,z) = \{q : D_O(x,q,z) = 1\} = \{q : q = 1/D_O(x,1,z)\} \Rightarrow F(x,z) = 1/D_O(x,1,z)$. Then EHON $\Rightarrow D_O(x,q,z) = b(\mu_z)D_O(x,q,\mu_z)/b(z) \Rightarrow D_O(x,1,z) = b(\mu_z)D_O(x,1,\mu_z)/b(z) = D_O(x,b(\mu_z),\mu_z)/b(z) \Rightarrow 1/D_O(x,1,z) = b(z)/D_O(x,b(\mu_z),\mu_z) \Rightarrow F(x,z) = b(z)/D_O(x,b(\mu_z),\mu_z)$ $\square$

Proposition D2: If the production function can be written $F(x,z) = b(z)/D_O(x,b(\mu_z),\mu_z)$ then the technology is EHN.

Proof: $F(x,z) = \{q : D_O(x,q,z) = D_I(x,q,z) = 1\} \Rightarrow F(x,z) = 1/D_O(x,1,z)$. If $F(x,z) = b(z)/D_O(x,b(\mu_z),\mu_z)$ then $1/D_O(x,1,z) = b(z)/D_O(x,b(\mu_z),\mu_z) \Rightarrow D_O(x,1,z) = D_O(x,b(\mu_z),\mu_z)/b(z) = b(\mu_z)D_O(x,1,\mu_z)/b(z) \Rightarrow D_O(x,q,z) = b(\mu_z)D_O(x,q,\mu_z)/b(z)$ (i.e., EHON). Furthermore, $D_O(x,q,z) = D_I(x,q,z) = 1 \Rightarrow D_I(x,q,z) = b(\mu_z)D_O(x,q,\mu_z)/b(z) \Rightarrow D_I(x,q,\mu_z) = b(\mu_z)D_O(x,q,\mu_z)/b(\mu_z) \Rightarrow D_I(x,q,\mu_z) = D_O(x,q,\mu_z) \Rightarrow D_I(x,q,z) = b(\mu_z)D_I(x,q,\mu_z)/b(z) = a(z)D_I(x,q,\mu_z)/a(\mu_z)$ where $a(z) \equiv 1/b(z)$ (i.e., EHIN) $\square$

Proposition D3: The following statements are equivalent:

A: $(x,q) \in T(z) \Leftrightarrow (\lambda x, \lambda^r q) \in T(z)$ for all $\lambda > 0$

B: $P(\lambda x,z) = \lambda^r P(x,z)$ for all $\lambda > 0$

C: $L(\lambda q,z) = \lambda^{1/r} L(q,z)$ for all $\lambda > 0$

Proof:

(A $\Rightarrow$ B): Let $\delta = \lambda^r q$. Then A $\Rightarrow P(x,z) = \{q : (x,q) \in T(z)\} = \{q : (\lambda x, \lambda^r q) \in T(z)\} = \{\lambda^{-r}\lambda^r q : (\lambda x, \lambda^r q) \in T(z)\} = \{\lambda^{-r}\delta : (\lambda x, \delta) \in T(z)\} = \lambda^{-r}\{\delta : (\lambda x, \delta) \in T(z)\} = \lambda^{-r}P(\lambda x, z) \Rightarrow P(\lambda x, z) = \lambda^r P(x,z) \quad \square$

(B $\Rightarrow$ A): It is always the case that $(x,q) \in T(z)$ if and only if $q \in P(x,z)$. Thus, $(x,q) \in T(z) \Leftrightarrow q \in P(x,z) \Leftrightarrow \lambda^r q \in \lambda^r P(x,z) \Leftrightarrow \lambda^r q \in P(\lambda x, z)$ (using B) $\Leftrightarrow (\lambda x, \lambda^r q) \in T(z) \quad \square$

(A $\Rightarrow$ C): Let $\delta = \lambda^{1/r}x$. Then A $\Rightarrow L(q,z) = \{x : (x,q) \in T(z)\} = \{x : (\lambda^{1/r}x, \lambda q) \in T(z)\} = \{\lambda^{-1/r}\delta : (\delta, \lambda q) \in T(z)\} = \lambda^{-1/r}\{\delta : (\delta, \lambda q) \in T(z)\} = \lambda^{-1/r}L(\lambda q, z) \Rightarrow L(\lambda q, z) = \lambda^{1/r}L(q,z) \quad \square$

(C $\Rightarrow$ A): It is always the case that $(x,q) \in T(z)$ if and only if $x \in L(q,z)$. Thus, $(x,q) \in T(z) \Leftrightarrow x \in L(q,z) \Leftrightarrow \lambda^{1/r}x \in \lambda^{1/r}L(q,z) \Leftrightarrow \lambda^{1/r}x \in L(\lambda q, z)$ (using C) $\Leftrightarrow (\lambda^{1/r}x, \lambda q) \in T(z) \quad \square$

Proposition D4: If the output distance function exists and outputs are weakly disposable (T4) then the following statements are equivalent:

B: $P(\lambda x, z) = \lambda^r P(x,z)$ for all $\lambda > 0$

D: $D_O(\lambda x, q, z) = \lambda^{-r} D_O(x,q,z)$ for all $\lambda > 0$

Proof: It is always the case that $(x,q) \in T(z)$ if and only if $q \in P(x,z)$. T4 is necessary and sufficient for the validity of the following statement: $q \in P(x,z)$ if and only if $D_O(x,q,z) \leq 1$ [Färe and Primont (1995, pp. 15, 16, 22)]. A corollary is that if T4 holds then and only then $P(x,z) = \{q : D_O(x,q,z) \leq 1\}$.

(B $\Rightarrow$ D): Let $\kappa = \delta\lambda^r$. Then B $\Rightarrow D_O(\lambda x, q, z) = \inf\{\delta > 0 : q/\delta \in P(\lambda x, z)\} = \inf\{\delta > 0 : q/\delta \in \lambda^r P(x,z)\} = \inf\{\delta > 0 : q/(\delta\lambda^r) \in P(x,z)\} = \inf\{\lambda^{-r}\kappa > 0 : q/\kappa \in P(x,z)\} = \lambda^{-r}\inf\{\kappa > 0 : q/\kappa \in P(x,z)\} = \lambda^{-r}D_O(x,q,z) \quad \square$

(D $\Rightarrow$ B): Let $\kappa = \lambda^{-r}q$. Then D $\Rightarrow P(\lambda x,z) = \{q : D_O(\lambda x,q,z) \leq 1\} = \{q : \lambda^{-r}D_O(x,q,z) \leq 1\} = \{q : D_O(x,\lambda^{-r}q,z) \leq 1\} = \{\lambda^r\kappa : D_O(x,\kappa,z) \leq 1\} = \lambda^r\{\kappa : D_O(x,\kappa,z) \leq 1\} = \lambda^r P(x,z)$ □

Proposition D5: If the input distance function exists and inputs are weakly disposable (T5) then the following statements are equivalent:

C: $L(\lambda q,z) = \lambda^{1/r}L(q,z)$ for all $\lambda > 0$

E: $D_I(x,\lambda q,z) = \lambda^{-1/r}D_I(x,q,z)$ for all $\lambda > 0$

Proof: It is always the case that $(x,q) \in T(z)$ if and only if $x \in L(q,z)$. T5 is necessary and sufficient for the validity of the following statement: $x \in L(q,z)$ if and only if $D_I(x,q,z) \geq 1$ [Färe and Primont (1995, p.22)]. A corollary is that if T5 holds then and only then $L(q,z) = \{x : D_I(x,q,z) \geq 1\}$.

(C $\Rightarrow$ E): Let $\kappa = \rho\lambda^{1/r}$. Then C $\Rightarrow D_I(x,\lambda q,z) = \sup\{\rho > 0 : x/\rho \in L(\lambda q,z)\} = \sup\{\rho > 0 : x/\rho \in \lambda^{1/r}L(q,z)\} = \sup\{\rho > 0 : x/(\rho\lambda^{1/r}) \in L(q,z)\} = \sup\{\lambda^{-1/r}\kappa > 0 : x/\kappa \in L(q,z)\} = \lambda^{-1/r}\sup\{\kappa > 0 : x/\kappa \in L(q,z)\} = \lambda^{-1/r}D_I(x,q,z)$ □

(E $\Rightarrow$ C): Let $\kappa = \lambda^{-1/r}x$. Then E $\Rightarrow L(\lambda q,z) = \{x : D_I(x,\lambda q,z) \geq 1\} = \{x : \lambda^{-1/r}D_I(x,q,z) \geq 1\} = \{x : D_I(\lambda^{-1/r}x,q,z) \geq 1\} = \{\lambda^{1/r}\kappa : D_I(\kappa,q,z) \geq 1\} = \lambda^{1/r}\{\kappa : D_I(\kappa,q,z) \geq 1\} = \lambda^{1/r}L(q,z)$ □

Proposition D6: If the output and input distance functions exist, outputs are weakly disposable (T4) and the technology is HDr then $D_O(x,q,z) = D_I(x,q,z)^{-r}$.

Proof: T4 is necessary and sufficient for the validity of the following statement: $q \in P(x,z)$ if and only if $D_O(x,q,z) \leq 1$ [Färe and Primont (1995, pp. 15, 16, 22)]. It is always the case that $(x,q) \in T(z)$ if and only if $q \in P(x,z)$. Thus, T4 $\Rightarrow D_I(x,q,z) = \sup\{\rho > 0 : (x/\rho,q) \in T(z)\} = \sup\{\rho > 0 : q \in P(x/\rho,z)\} = \sup\{\rho > 0 : D_O(x/\rho,q,z) \leq 1\}$. T4 and HD$r$ $\Rightarrow D_O(\lambda x,q,z) = \lambda^{-r}D_O(x,q,z)$

for all $\lambda > 0$ (Propositions D3 and D4). Thus, T4 and HD$r \Rightarrow D_I(x,q,z) = \sup\{\rho > 0 : D_O(x/\rho,q,z) \leq 1\} = \sup\{\rho > 0 : \rho^r D_O(x,q,z) \leq 1\} = \sup\{\rho > 0 : \rho \leq D_O(x,q,z)^{-1/r}\} = D_O(x,q,z)^{-1/r}$. Equivalently, $D_O(x,q,z) = D_I(x,q,z)^{-r}$ □

Proposition D7: If the output and input distance functions exist, inputs are weakly disposable (T5) and the technology is HDr then $D_O(x,q,z) = D_I(x,q,z)^{-r}$.

Proof: T5 is necessary and sufficient for the validity of the following statement: $x \in L(q,z)$ if and only if $D_I(x,q,z) \geq 1$ [Färe and Primont (1995, p.22)]. It is always the case that $(x,q) \in T(z)$ if and only if $x \in L(q,z)$. Thus, T5 $\Rightarrow D_O(x,q,z) = \inf\{\delta > 0 : (x,q/\delta) \in T(z)\} = \inf\{\delta > 0 : x \in L(q/\delta,z)\} = \inf\{\delta > 0 : D_I(x,q/\delta,z) \geq 1\}$. T5 and HD$r \Rightarrow D_I(x,\lambda q,z) = \lambda^{-1/r} D_I(x,q,z)$ for all $\lambda > 0$ (Propositions D3 and D5). Thus, T5 and HD$r \Rightarrow D_O(x,q,z) = \inf\{\delta > 0 : D_I(x,q/\delta,z) \geq 1\} = \inf\{\delta > 0 : \delta^{1/r} D_I(x,q,z) \geq 1\} = \inf\{\delta > 0 : \delta \geq D_I(x,q,z)^{-r}\} = D_I(x,q,z)^{-r}$ □

Proposition D8: If a regular technology is EH, EHN and HDr then $D_O(x,q,z) \propto h(q)^r / \big(b(z)g(x)\big)$ and $D_I(x,q,z) \propto \big(b(z)g(x)\big)^{1/r}/h(q)$ where $h(.)$ is NN, ND and homogeneous of degree $1/r$ and $g(.)$ is NN, ND and HDr.

Proof: EOH $\Rightarrow D_O(x,q,\mu_z) = g(\mu_x) D_O(\mu_x,q,\mu_z)/g(x)$ (A). EHON $\Rightarrow D_O(x,q,z) = b(\mu_z) D_O(x,q,\mu_z)/b(z) = b(\mu_z) g(\mu_x) D_O(\mu_x,q,\mu_z)/[b(z)g(x)]$ (using A) $\Rightarrow D_O(x,\mu_q,\mu_z) = g(\mu_x) D_O(\mu_x,\mu_q,\mu_z)/g(x) \Rightarrow D_O(x,\mu_q,\mu_z)^{-1/r} = g(\mu_x)^{-1/r} D_O(\mu_x,\mu_q,\mu_z)^{-1/r} g(x)^{1/r}$ (B). HD$r \Rightarrow D_I(x,q,z) = D_O(x,q,z)^{-1/r}$ (from D6 or D7) $\Rightarrow D_I(x,\mu_q,\mu_z) = D_O(x,\mu_q,\mu_z)^{-1/r}$ (D). Thus, EOH, EHON and HD$r \Rightarrow D_I(x,\mu_q,\mu_z) = g(\mu_x)^{-1/r} D_O(\mu_x,\mu_q,\mu_z)^{-1/r} g(x)^{1/r}$ (using B and D) (E). EIH $\Rightarrow D_I(x,q,\mu_z) = h(\mu_q) D_I(x,\mu_q,\mu_z)/h(q)$ (F). EHIN $\Rightarrow D_I(x,q,z) = a(z) D_I(x,q,\mu_z)/a(\mu_z)$ (G). Thus, EH, EHN and HD$r \Rightarrow D_I(x,q,z) = a(z)h(\mu_q) D_I(x,\mu_q,\mu_z)/[h(q)a(\mu_z)]$ (using F and G) $\Rightarrow D_I(x,q,z) = a(z)h(\mu_q) g(\mu_x)^{-1/r} D_O(\mu_x,\mu_q,\mu_z)^{-1/r} g(x)^{1/r}/[h(q)a(\mu_z)]$ (using E). The left-hand side is independent of (μ_x,μ_q,μ_z) so the right-hand side must also also be independent of $(\mu_x, \mu_q, \mu_z) \Rightarrow D_I(x,q,z) \propto a(z)g(x)^{1/r}/h(q) = \left[b(z)g(x)\right]^{1/r}/h(q)$ where $b(z) = a(z)^r \Rightarrow D_O(x,q,z) \propto h(q)^r/\left[b(z)g(x)\right]$ (from D6 or D7). If the technology is regular (i.e., T1-T6 hold) then the output (input) distance function is NN, ND and HD1 in outputs (inputs) $\Rightarrow h(.)$ is NN, ND and homogeneous of degree $1/r$ and $g(.)$ is NN, ND and HDr $\square$

Proposition D9: If the distance function representations of a regular technology are $D_O(x,q,z) \propto h(q)^r/\big(b(z)g(x)\big)$ and $D_I(x,q,z) \propto \big(b(z)g(x)\big)^{1/r}/h(q)$ where $h(.)$ is NN, ND and homogeneous of degree $1/r$ and $g(.)$ is NN, ND and HDr then the technology is EH, EHN and HDr.

Proof: $D_O(x,q,z) = \kappa h(q)^r/\big(b(z)g(x)\big)$ and $D_I(x,q,z) = \tau\big(b(z)g(x)\big)^{1/r}/h(q)$ where κ and τ are factors of proportionality.

(HDr): $D_O(x,q,z) = \kappa h(q)^r/\big(b(z)g(x)\big) \Rightarrow D_O(\lambda x,q,z) = \kappa h(q)^r/\big(b(z)g(\lambda x)\big) = \lambda^{-r}\kappa h(q)^r/\big(b(z)g(x)\big) = \lambda^{-r}D_O(x,q,z) \Rightarrow$ (from D4 and $g(.)$ being HDr).

(EOH): $D_O(x,q,z) = \kappa h(q)^r/\big(b(z)g(x)\big) \Rightarrow D_O(\mu_x,q,z) = \kappa h(q)^r/\big(b(z)g(\mu_x)\big) \Rightarrow g(\mu_x)D_O(\mu_x,q,z)/g(x) = \kappa h(q)^r/\big(b(z)g(x)\big) = D_O(\mu_x,q,z)$.

(EIH): $D_I(x,q,z) = \tau\big(b(z)g(x)\big)^{1/r}/h(q) \Rightarrow D_I(x,\mu_q,z) = \tau\big(b(z)g(x)\big)^{1/r}/h(\mu_q) \Rightarrow h(\mu_q)D_I(x,\mu_q,z)/h(q) = \tau\big(b(z)g(x)\big)^{1/r}/h(q) = D_I(x,q,z)$.

(EHON): $D_O(x,q,z) = \kappa h(q)^r/\big(b(z)g(x)\big) \Rightarrow D_O(x,q,\mu_z) = \kappa h(q)^r/\big(b(\mu_z)g(x)\big) \Rightarrow b(\mu_z)D_O(x,q,\mu_z)/b(z) = \kappa h(q)^r/\big(b(z)g(x)\big) = D_O(\mu_x,q,z)$.

(EHIN): $D_I(x,q,z) = \tau\big(b(z)g(x)\big)^{1/r}/h(q) \Rightarrow D_I(x,q,\mu_z) = \tau\big(b(\mu_z)g(x)\big)^{1/r}/h(q) \Rightarrow b(z)^{1/r}D_I(x,\mu_q,z)/b(\mu_z)^{1/r} = \tau\big(b(z)g(x)\big)^{1/r}/h(q) = D_I(x,q,z)$. Equivalently, $a(z)D_I(x,\mu_q,z)/a(\mu_z) = D_I(x,q,z)$ where $a(z) = b(z)^{1/r}$. $\square$

Notes

[1]This paper was prepared for presentation in the Workshop on Productivity Measurement in Regulated Industries, Santa Cruz, CA, 11–12 June 2012. Valuable comments were provided by John Walden, Aaron Mamula, Kris Kerstens and several other participants. I gratefully acknowledge the work of NPF fishers, the CSIRO and AFMA in reporting, collecting and managing the data.

[2]In this case, only one observation will be available to estimate the frontier in each time period, and all observations will be found to lie on the frontier. Thus, the environmental change and technical change components of TFP change cannot be identified.

[3]ND $\Rightarrow$ if $q_r \geq q_t$ then $Q(q_r) \geq Q(q_t)$. Then NN $\Rightarrow Q(q_r)/Q(q_s) \geq Q(q_t)/Q(q_s) \Leftrightarrow QI_{sr} \geq QI_{st}$ (i.e., Q1). Proofs of Q2, Q4 and Q6 are straightforward using HD1 $\Rightarrow Q(\lambda q_t) = \lambda Q(q_t)$. Proofs of Q3 and Q7 are trivial. Finally, $QI(\Lambda q_s, \Lambda q_t, \Lambda^{-1} a_s, \Lambda^{-1} a_t) = (a_t' \Lambda^{-1} \Lambda q_t)/(a_s' \Lambda^{-1} \Lambda q_s) = (a_t' q_t)/(a_s' q_s) = QI(q_s, q_t, a_s, a_t)$ (i.e., Q5).

[4]If $E(v_t) = \mu_v \neq 0$ then (23) can be reparameterised in terms of $v_t^* = v_t - \mu_v$ and $\gamma_0^* = \gamma_0 + \mu_v$ (i.e., the nonzero mean can be subsumed into the intercept term). The notation $f_N(a|b, C)$ is used for a normal pdf with mean vector b and covariance matrix C.

[5]The notation $f_D(a|b)$ is the notation for a Dirichlet distribution used by Poirier (1995, p.132). If $a = (a_1, \ldots, a_N)'$ and $b = (b_1, \ldots, b_N)'$ then $E(a_n) = b_n/b_0$ and $Var(a_n) = b_n(b_0 - b_n)/(b_0^3 + b_0^2)$ where $b_0 = \sum_n b_n$.

[6]The notation $f_G(a|b, c)$ is used for a gamma pdf with mean vector b/c and variance b/c^2. If $b = 1$ then $f_G(a|b, c)$ is an exponential pdf.

[7]An HPD interval is the Bayesian analogue of a confidence interval: a $100(1-\alpha)\%$ HDR is the interval of shortest length that contains $100(1-\alpha)\%$ of the area under the pdf.

[8]The (k,n)-th marginal rate of technical transformation measures the rate at which output n can be substituted for output k in a technically efficient production process, holding inputs, environmental variables and all other outputs fixed. If the error terms in (19)–(22) vanish then $MRTT_{kn} = \alpha_n/\alpha_k$.

References

Aigner, D., C. Lovell, and P. Schmidt. 1977. "Formulation and Estimation of Stochastic Frontier Production Function Models." *Journal of Econometrics* 6:21–37.

Blackorby, C., C.A.K. Lovell, and M. Thursby. 1976. "Extended Hicks Neutral Technical Change." *The Economic Journal* 86:845–852.

Casella, G., and E. George. 1992. "Explaining the Gibbs Sampler." *The American Statistician* 46:167–73.

Chib, S., and E. Greenberg. 1996. "Markov Chain Monte Carlo Simulation Methods in Econometrics." *Econometric Theory* 12:409–431.

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

Fernandez, C., G. Koop, and M. Steel. 2000. "A Bayesian analysis of multiple-output production frontiers." *Journal of Econometrics* 98:47–79.

Kompas, T., T. Che, and R. Grafton. 2004. "Technical efficiency effects of input controls: evidence from Australia's banana prawn fishery." *Applied Economics* 36:1631–1641.

Koop, G. 2003. *Bayesian Econometrics*. Chichester: John Wiley and Sons.

O'Donnell, C. 2012. "An aggregate quantity framework for measuring and decomposing productivity change." *Journal of Productivity Analysis* (doi:10.1007/s11123-012-0275-1).

—. 2011. "Econometric Estimation of Distance Functions and Associated Measures of Productivity and Efficiency Change." Centre for Efficiency and Productivity Analysis Working Papers No. WP01/2011, University of Queensland.

—. 2010. "Measuring and Decomposing Agricultural Productivity and Profitability Change." *The Australian Journal of Agricultural and Resource Economics* 54:527–560.

Pascoe, S., L. Coglan, A. Punt, and C. Dichmont. 2012. "Impacts of Vessel Capacity Reduction Programmes on Efficiency in Fisheries: the Case of Australia's Multispecies Northern Prawn Fishery." *Journal of Agricultural Economics* doi: 10.1111/j.1477-9552.2011.00333.x.

Poirier, D. 1995. *Intermediate Statistics and Econometrics: A Comparative Approach*. London: MIT Press.

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

Staples, D., and D. Vance. 1986. "Emigration of juvenile banana prawns, Penaeus merguiensis, from a mangrove estuary and a recruitment to offshore areas in the wet-dry tropics of the Gulf of Carpentaria, Australia." *Marine Ecology Progress Series* 27:239–252.

Figures

Figure 1. Selected MCMC Chains

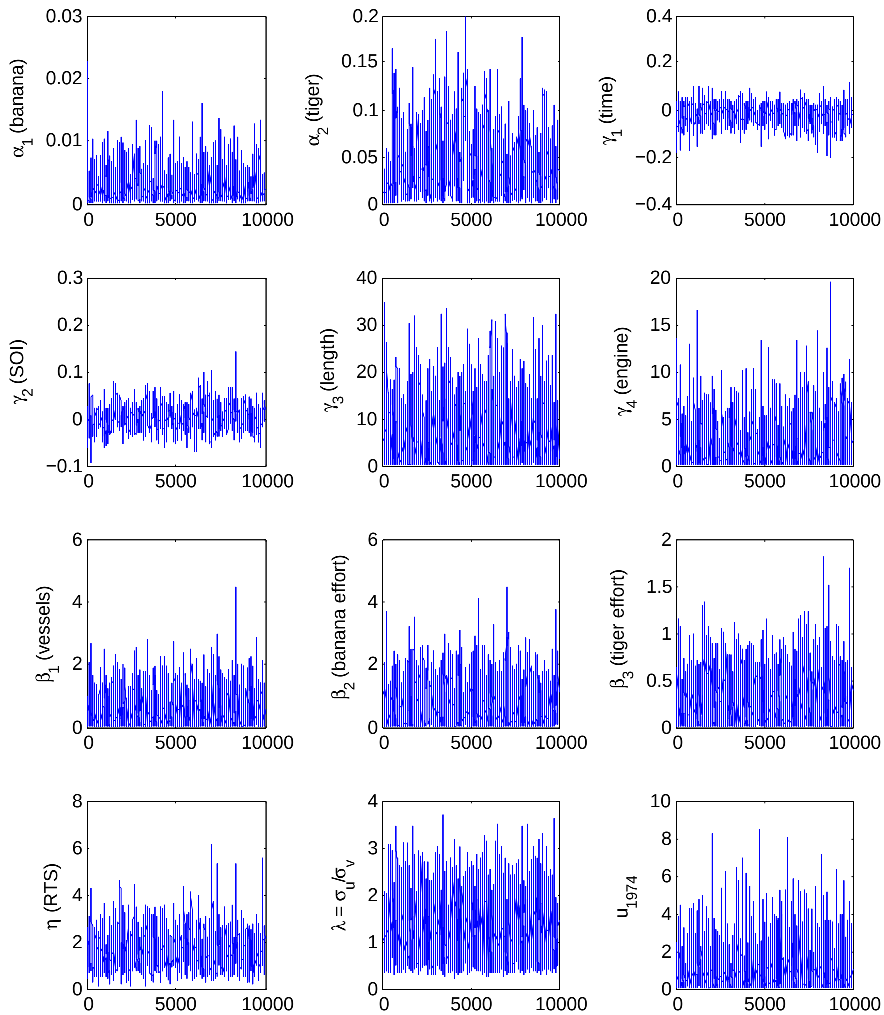


Figure 2. The Components of TFP Change

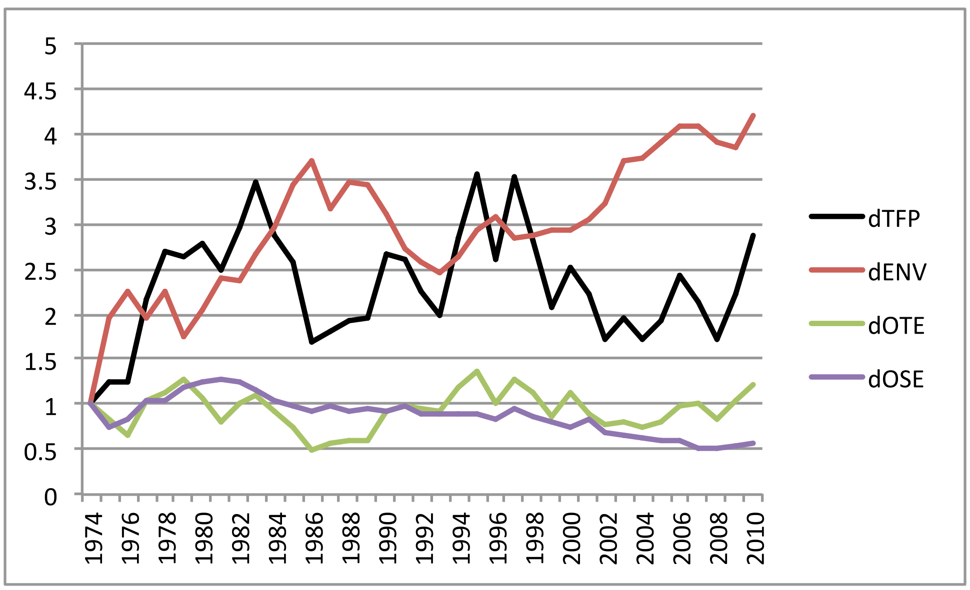


Figure 3. The Components of Environmental Change

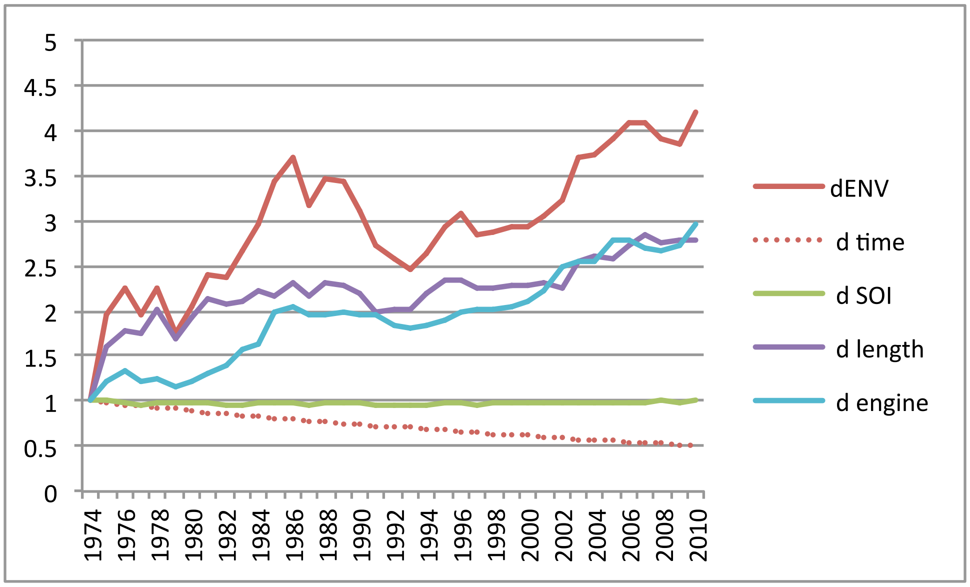


Figure 4. Changes in Output-Oriented Scale Efficiency

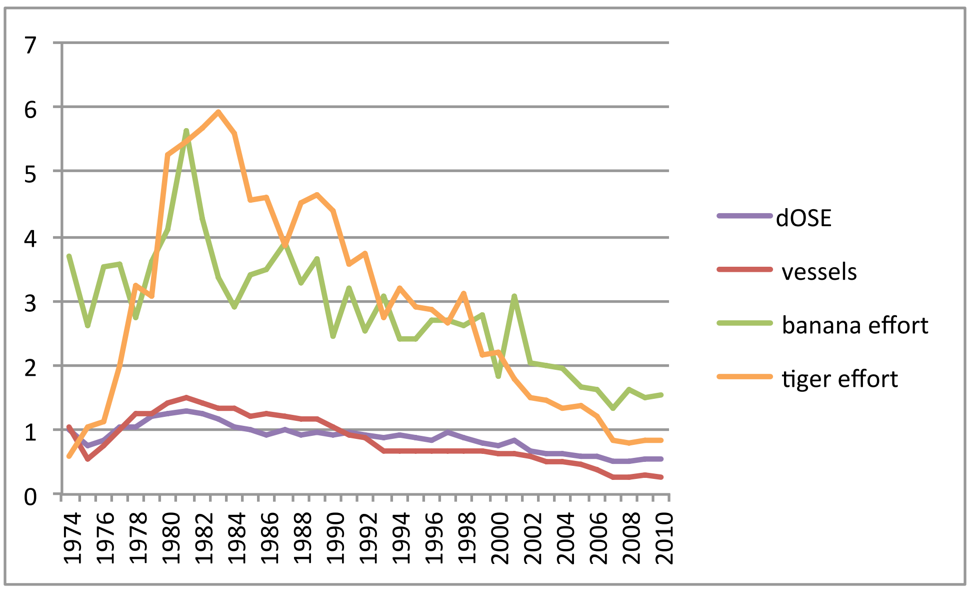


Figure 5. Levels of Output-Oriented Technical Efficiency

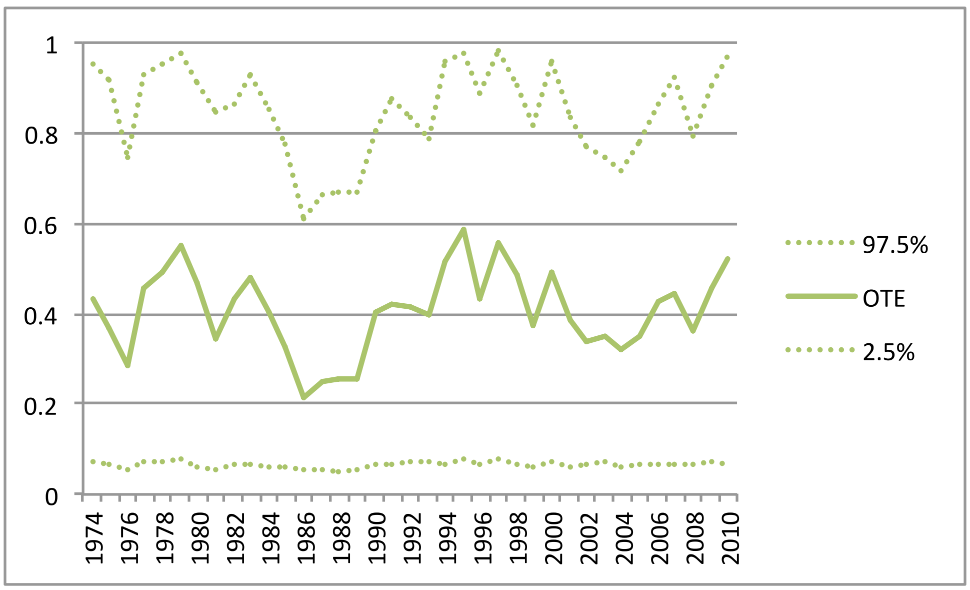
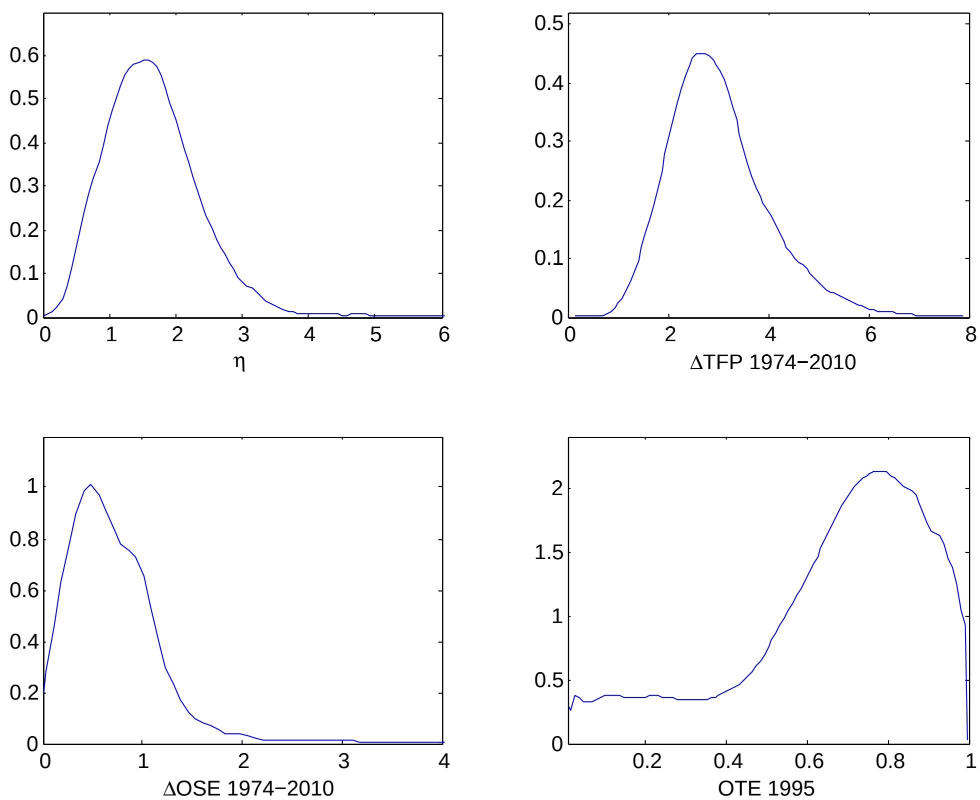


Figure 6. Estimated Posterior Pdfs



Tables

Table 1. Descriptive Statistics

	Mean	St. Dev.	Min.	Max.
q_1 = banana prawns (tonnes)	4290.51	1960.58	2157	12711
q_2 = tiger prawns (tonnes)	2847.84	1332.76	666	5751
q_3 = endeavour prawns (tonnes)	944.22	540.49	196	2124
q_4 = king prawns (tonnes)	52.14	47.42	3	207
x_1 = number of vessels	161.03	70.71	51	286
x_2 = banana fishery effort (days)	5777.24	1898.81	2696	11524
x_3 = tiger fishery effort (days)	16763.27	9502.22	3439	34551
$\ln z_1$ = time	23	11	5	41
$\ln z_2$ = SOI	-1.24	7.23	-13.1	13.6
z_3 = average vessel length (metres)	21.59	0.60	19.22	22.47
z_4 = average engine power (kw)	325.56	40.31	245.22	394.94

Table 2. Parameter Estimates

	Mean	St.Dev.	2.5%	97.5%
$\hat{\alpha}_1$ (banana)	0.002	0.002	0.000	0.006
$\hat{\alpha}_2$ (tiger)	0.042	0.032	0.004	0.124
$\hat{\alpha}_3$ (endeavour)	0.120	0.095	0.009	0.352
$\hat{\alpha}_4$ (king)	0.836	0.090	0.625	0.965
$\hat{\gamma}_0$	-39.566	20.457	-86.976	-9.089
$\hat{\gamma}_1$ (time)	-0.019	0.035	-0.099	0.040
$\hat{\gamma}_2$ (SOI)	0.002	0.022	-0.039	0.050
$\hat{\gamma}_3$ (length)	6.715	5.230	0.239	19.265
$\hat{\gamma}_4$ (engine)	2.281	2.152	0.069	8.030
$\hat{\beta}_1$ (vessels)	0.571	0.469	0.018	1.714
$\hat{\beta}_2$ (banana effort)	0.789	0.594	0.032	2.158
$\hat{\beta}_3$ (tiger effort)	0.291	0.221	0.011	0.818
$\hat{s}_1$	2.096	0.869	0.912	4.286
$\hat{s}_2$	24.794	20.148	2.318	79.795
$\hat{s}_3$	23.129	20.233	1.516	77.063
$\hat{s}_4$	8.022	5.653	0.828	22.743
$\hat{\eta}$ (RTS)	1.651	0.704	0.533	3.197
$\hat{\lambda}$	1.275	0.491	0.447	2.372
$\hat{h}$ (precision)	899.620	2640.400	0.232	6850.300

Table 3. The Components of TFP Change

Year	ΔTFP	ΔENV	ΔOTE	ΔOSE
1974	1	1	1	1
1975	1.240	1.954	0.846	0.750
1976	1.261	2.255	0.659	0.848
1977	2.162	1.956	1.055	1.047
1978	2.691	2.262	1.135	1.049
1979	2.639	1.738	1.270	1.196
1980	2.796	2.045	1.084	1.261
1981	2.477	2.402	0.799	1.291
1982	2.955	2.370	0.998	1.250
1983	3.450	2.679	1.104	1.167
1984	2.872	2.950	0.929	1.048
1985	2.581	3.435	0.755	0.995
1986	1.700	3.706	0.493	0.930
1987	1.819	3.180	0.574	0.997
1988	1.914	3.479	0.593	0.928
1989	1.965	3.439	0.597	0.956
1990	2.671	3.110	0.926	0.927
1991	2.605	2.726	0.979	0.976
1992	2.245	2.580	0.966	0.901
1993	2.000	2.463	0.916	0.887
1994	2.852	2.641	1.198	0.901
1995	3.565	2.940	1.356	0.894
1996	2.594	3.071	1.000	0.844
1997	3.519	2.858	1.292	0.953
1998	2.799	2.874	1.119	0.871
1999	2.065	2.937	0.861	0.816
2000	2.525	2.943	1.143	0.751
2001	2.232	3.056	0.890	0.820
2002	1.714	3.218	0.781	0.682
2003	1.952	3.701	0.815	0.647
2004	1.718	3.718	0.735	0.629
2005	1.916	3.908	0.815	0.602
2006	2.437	4.087	0.987	0.604
2007	2.141	4.083	1.022	0.513
2008	1.710	3.913	0.842	0.519
2009	2.231	3.860	1.050	0.550
2010	2.861	4.205	1.212	0.561

Estimating a nonparametric homothetic S-shaped production relation for the US West Coast groundfish production 2004-2007

O. B. Olesen(1), J. Ruggiero(2) and A. T. Mamula(3),

(1) Department of Business and Economics

The University of Southern Denmark, Odense, Denmark, ole@sam.sdu.dk

(2) School of Business Administration,

University of Dayton, Dayton, Ohio 45469-2251, USA, jruggiero1@udayton.edu

(3) National Oceanic and Atmospheric Administration,

NMFS/SWFSC/FED, Santa Cruz, CA. aaron.mamula@noaa.gov

October 4, 2012

Abstract

The use of the convex hull estimation in data envelopment analysis (DEA) models is often characterized as a variable returns to scale model. However it is well known that standard microeconomic production theory posits a nonconvex S-shaped production frontier, i.e. a production technologies that obey the Regular Ultra Passum Law with a monotone decreasing scale elasticity along any expansion path. Recently for a homothetic production relation a nonparametric estimation approach that allows for a nonconvex S-shaped scaling law has been proposed. In this paper we use this approach to analyze and estimate scale characteristics of the US West Coast groundfish production for the period 2004-2007.

1 Introduction

The non-parametric DEA approach involves the construction of a piecewise linear envelopment of observed data. The characteristics of a DEA model are derived from a number of maintained hypotheses imposed as part of the model. An estimator often used in DEA is the convex hull estimator also denoted the BCC-estimator (Banker, Charnes and Cooper 1984), which is a variable returns to scale estimator, and where the estimated production possibility set is polyhedral. The convex hull estimator maintains convexity of the production possibility set. Convexity is a convenient assumption in the sense that easy estimation procedures are available based on linear programming. But it is important to stress that this assumption is not in general supported by microeconomic theory. The assumption of convexity of input sets and output sets is supported in microeconomics but the scaling law can in general have many different shapes. In a recent paper (Olesen and Ruggiero 2012) suggest an approach that allows for an estimator which reflects possible segments with increasing returns to scale. This suggested approach is in contrast to the convex hull estimator which rely on an assumption of non-increasing marginal products, thereby violating standard microeconomic theory. The estimation approach suggested in (Olesen and Ruggiero 2012) is designed to uncover non-convexities that reflects the Regular Ultra Passum (RUP) law, ((Frisch 1965) Chapter 8) stating that for the case of one output along any expansion path in factor space, optimal scale size is unique (or possibly connected intervals of sizes) and that the scale elasticity is monotonic decreasing:

Definition 1 *The RUP law. Let a single output y be produced from a vector of m inputs x according to a production function $F(x,y) = 0$. This production function obeys the RUP law if $\frac{\partial \varepsilon(x,y)}{\partial x_i} < 0, i = 1, \cdots, m$ where the function $\varepsilon(x,y)$ is the scale elasticity, and for some point (x_1, y_1) we have $\varepsilon(x_1, y_1) > 1$, and for some point (x_2, y_2), where $x_2 > x_1, y_2 > y_1$, we have $\varepsilon(x_2, y_2) < 1$.*

A non-convex S-shaped technology is characterized as follows: along any expansion path an expanding DMU with low activity will have a high scale elasticity greater than one. As the unit expands its activity the scale elasticity will decrease and will approach optimal scale size with an elasticity equal to one. Further expansion will imply decreasing returns with a scale elasticity less than one and approaching zero.

Maintaining the RUP law requires that the scale elasticity is monotonically decreasing for increasing production. (Førsund and Hjalmarsson 2004a) have demonstrated how well established core concepts from neoclassical theory such as scale elasticity can be fruitfully translated and applied within the non parametric DEA approach. However, (Førsund and Hjalmarsson 2004b) argue that while the theoretical concepts as such carry over to the piecewise linear frontier, the RUP-law simply cannot be obeyed, not even with data generated in a process consistent with the law. This is a simple consequence of marginal productivity being constant while average productivity is decreasing when passing along a decreasing returns to scale facet[1]. Notice however, the piecewise linear frontier is an outer approximation to a true smooth frontier. Hence, only asymptotically will we see a smooth estimator of the frontier and only if the estimated piecewise linear frontier is S-shaped will we see a smooth estimator that obeys the RUP law. In other words, the violation discussed in (Førsund and Hjalmarsson 2004b) disappears asymptotically, see (Olesen and Petersen 2011).

One particular problem with the BCC model is the use of the convex hull estimator in the envelopment of data points. Using supporting hyperplanes for envelopment can overestimate inefficiency for points that should be projected to the local non-convex segments of the true frontier characterized by increasing returns to scale. Further, existing measures of scale efficiency are biased due to the improper projection. Maintaining the RUP-law adds structure to the estimation process that will allow us to recover the scale elasticity and the inefficiency of such DMUs.

The main contribution of (Olesen and Ruggiero 2012) is the development of an approach that is capable of measuring scale elasticities and inefficiencies for production possibilities in a non-convex homothetic and S-shaped technology. The proposed model assumes one output and multiple inputs. The assumption of homotheticity allows for an aggregation procedure of multiple inputs into an aggregated input index.

A homothetic production function was introduced in ((Shephard 1953),page 30). The notion was generalized in (Shephard 1970) to the multi-output case. A homothetic production function is a monotonic transformation of a linear homogenous production function. Assuming an input homothetic production structure allows us to estimate the isoquants because homotheticity implies that the shape of the isoquants are identical. This allows us to maintain convexity in input (and output space) and to allow non-convexities in input-output space. The proposed approach relies on the order-m estimation procedures, (Cazals, Florens and Simar 2002) with local convexity as proposed and formalized in (Daraio and Simar 2005). The order-m models are derived in quite general terms in the sense that no formal structure beyond the minimal DEA assumptions is imposed. For the purpose of this paper we do have additional structure that should be imposed on the estimation procedure. In addition to the traditional maintained hypothesis of strong disposability, returns to scale, minimal extrapolation we focus on the situation where it is reasonable impose i) a maintained hypothesis of homotheticity and ii) a maintained hypothesis of a monotonic declining scale elasticity along any expansion path.

The illustrative application of the model presented below in Section 6 is based on a data set with catch records from fishery in the Pacific Ocean during the period 2004-2007. The output data is aggregated into four different outputs, primarily based on the condition of the fishery. One variable input "days at sea" and one quasi-fixed input "vessel length" are available. With multiple outputs we change focus to an output homothetic relationship. Hence we estimate the S-shaped production relation based on an assumption of output homotheticity which allows us to aggregate the outputs into an output index. In the application we either ignore the quasi-fixed input or incorporate indirectly the impact from this input on the estimated shape of the base isoquant.

We have chosen not to implement a full multiple input multiple output model for several reasons:

- It is not clear how the long run substitution between the variable and the quasi fixed inputs works. It is not a traditional microeconomic substitution with positive substitution elasticities in the full range of the quasi-fixed input.

- The approach suggested in (Olesen and Ruggiero 2012) assumes one output and multiple inputs with input homotheticity. It is straight forward to change focus to one input multiple outputs maintaining output homotheticity. However, maintaining input *and* output homotheticity with multiple inputs and outputs requires a more complicated estimation procedure of the isoquants because we simultaneously have to estimate both a base input and a base output isoquant.

In order to move to a one input aggregated output space, we need to estimate the base output isoquant. Assuming selective output convexity we use a simplified order-m estimation procedure (Cazals et al. 2002) where replications are avoided. The order-m estimation procedures include a conditional estimation model maintaining selective convexity of the output sets[2]. Under the assumption of output homotheticity, we can aggregate output allowing us to move to a two dimensional input aggregate-output space.

As mentioned above, the traditional DEA approach using a convex technology, as originally presented in the BCC model, fails since part of a technology satisfying the RUP law is non-convex. It is argued in (Olesen and Ruggiero 2012) that several non-convex models exists (see e.g. (Petersen 1990),(Bogetoft 1996)), but these models are not well-suited to estimate an S-shaped production structure. These general non-convex models tend to pick up many non-convexities that in fact may be a consequence of fitting data to closely. The approach suggested in (Olesen and Ruggiero 2012) and applied in this paper introduces structure into the estimation procedure by allowing *only* non-convexities that are reflected in an S-shaped production structure. More precisely, the following axioms are maintained on the production possibility set $T = \{(x, y) : x \text{ can produce } y\}$, with a scalar output $y \in \mathbb{R}_+$:

Axiom 1 *(Feasibility of observed data). For any* $j = 1, \dots, n, (X_j, Y_j) \in T$

Axiom 2 *(Free disposability).*$(X, Y) \in T$, $Y \geq \widehat{Y} \geq 0$ *and* $X \leq \widehat{X}$ *implies* $\left(\widehat{X}, \widehat{Y}\right) \in T$.

Axiom 3 *(Convexity of input set). Let* $\left(\widetilde{X}\right) \in L(Y)$ *and* $\left(\widehat{X}\right) \in L(Y)$. *Then* $\lambda\left(\widetilde{X}\right) + (1-\lambda)\left(\widehat{X}\right) \in L(Y)$, *for any* $\lambda \in [0, 1]$, *where* $L(Y) = \{X \in \mathbb{R}^s_+ : X \text{ can produce } Y\}$

Axiom 4 *(Input Homotheticity). The technology* T *is input homothetic*

In addition to these axioms it is assumed that the technology satisfies the RUP law in Definition 1.

Notice that convexity of the input set by Axiom 3 does not imply convexity of the production possibility set T. Conversely, maintaining convexity of T using e.g. the BCC-model implies convexity of the input set $L(Y)$.

The rest of the paper is organized as follows. In section 2-3 we define the production technology, from an input orientation using an input distance function and from an output orientation using an output distance function. The assumption of homotheticity is presented and the implication for input or output aggregation is discussed. Specifically, the assumption of output homotheticity allows us to generate any output isoquant from a base output isoquant and hence, derive a well-defined index of aggregate output. Section 4 is devoted to the estimation of the base isoquant using a conditional estimator. We also discuss criteria for selecting a well-estimated isoquant among all possible base isoquants to aggregate output. In section 5 we focus on how to estimate the piecewise linear S-shaped frontier. In section 6 we apply this approach to data from US West Coast groundfish production for the period 2004-2007. The focus is on estimating an S-shaped input consumption function based on an assumption of output homotheticity. The last section concludes with directions for future research.

2 Production Technology, multiple inputs and one output

Recently (Olesen and Ruggiero 2012) have suggested an approach based on an assumption of input homotheticity that allows for an S-shaped estimator which reflects possible segments with increasing returns to scale. Focus is on the classical production function with one output Y and multiple inputs $X = (x_1, \ldots, x_s)$. Even though we in the application of the model in this paper will focus on the case with one input multiple outputs we will briefly mention how the suggested method works for a production function. We represent the production technology with the input set $L(Y) = \{X \in \mathbb{R}^s_+ : X \text{ can produce } Y\}$ which has isoquant $IsoqL(Y) = \{X : X \in L(Y), \lambda X \notin L(Y), \lambda \in [0,1)\}$. Since we assume that only one output is produced, we can define a production function as $\phi(X) = \max\{Y : X \in L(Y)\}$. The input distance function is then defined as $D_I(Y,X) = \max\{\gamma : X/\gamma \in L(Y)\}$, which provides an alternative characterization of the technology since $D_I(Y,X) \geq 1 \Leftrightarrow X \in L(Y)$. Finally, the index of technical efficiency proposed by Debreu (1951) and Farrell (1957) is given as $F_I(Y,X) = \min\{\gamma : \gamma X \in L(Y)\}$, where $F_I(y,x) = D_I(y,x)^{-1}$. Additional structure is imposed by assuming that production is input homothetic.

Definition 2 *A production function $\phi(X)$ is input homothetic*

$$Y = \phi(X) = F(g(X))$$

where $F() : R_+ \to R_+$ is monotonic and $g(\lambda X) = \lambda g(X)$ i.e. $g()$ is positive homogeneous of degree one and continuously differentiable (see (Shephard 1970)). $g()$ is denoted the kernel function.

From the definition, we see that a homothetic production function can be represented as a production process whereby the input vector X can be aggregated into a one dimensional input index $g(X)$, i.e. output is determined from the level of aggregate input (see (Färe and Lovell 1988) for a more general result). From (Olesen and Ruggiero 2012) we have the following proposition.

Proposition 5 *Assume a homothetic technology with one output. The distance function evaluated at* $(1, X)$ *is equal to aggregate input defined from the core function in the homothetic production function multiplied by a constant, i.e.*

$$D_I(1, X) = k_I \times g(X), k_I \in \mathbb{R}_+$$

The dimensionality of DEA models can according to Proposition 1 be reduced under the assumption of homotheticity. Homotheticity allows us to span the production technology from $L(1)$ (see (Shephard 1970), page 34). Let $H(Y)$ be a scaling function then $L(Y) = H(Y)L(1)$. Any input sets can in this way be generated from a base input set by a scaling function $H(Y)$ depending only on the level of output and not the input mix. From this follows that $IsoqL(Y) = H(Y)IsoqL(1)$. More generally, we could choose any output level and its associated isoquant to serve as the base.

3 Production Technology, multiple output and one input

In Section 6 we apply the approach outlined above to estimate scale characteristics of the US West Coast groundfish production for the period 2004-2007. This application involves a production with multiple outputs and one variable input. For that purpose we need to change focus to an output homothetic relationship. Let us consider a production environment where a vector of p outputs $Y = (y_1, \ldots, y_p)$ is produced using one input X. We represent the production technology with the output set $P(X) = \{Y \in \mathbb{R}^p_+ : X \text{ can produce } Y\}$ which has isoquant

$$IsoqP\,(X) = \{Y : Y \in P(X), \lambda Y \notin P(X), \lambda \in (1, \infty)\}. \tag{1}$$

Since we assume that only one input is consumed, we can define an input consumption function as

$$\phi\,(Y) = \min\{X : Y \in P(X)\} \tag{2}$$

The output distance function is then defined as

$$D_O(Y, X) = \min\{\gamma : Y/\gamma \in P(X)\}, \tag{3}$$

which provides an alternative characterization of the technology since $D_O(Y, X) \leq 1 \Leftrightarrow Y \in P(X)$. Finally, the output oriented index of technical efficiency that serves as basis for DEA is given as

$$F_O(Y, X) = \max\{\gamma : \gamma Y \in P(X)\}, \tag{4}$$

where $F_O(y, x) = D_O(y, x)^{-1}$.

We now assume that the input consumption function is output homothetic.

Definition 3 *An input consumption function $\phi(Y)$ is output homothetic if*

$$X = \phi(Y) = F(g(Y))$$

where $F() : R_+ \rightarrow R_+$ is monotonic and $g(\lambda Y) = \lambda g(Y)$ i.e. $g()$ is positive homogeneous of degree one and continuously differentiable. $g()$ is denoted the kernel function.

From the definition, we see that an output homothetic input consumption function can be represented as a consumption process whereby the output vector Y can be aggregated into a one dimensional output index $g(Y)$, i.e. input is determined from the level of aggregate output. We can now state a similar proposition generated from the assumption of output homotheticity:

Proposition 6 *Assume an output homothetic technology with one input. The output distance function evaluated at $(Y, 1)$ is equal to aggregate output defined from the core function in the homothetic input consumption function multiplied by a constant, i.e.*

$$D_O(Y, 1) = k_O \times g(Y), k_O \in \mathbb{R}_+$$

Proof. Let $\phi(Y) = F(g(Y))$ with $F^{-1} = f$. We know that

$$\begin{aligned} P(x) &= \{Y : F(g(Y)) \leq x\} \\ &= \{Y : g(Y) \leq f(y)\} \end{aligned}$$

Furthermore,

$$\begin{aligned}
D_O(Y,1) &= \min\{\gamma : Y/\gamma \in P(1)\} \\
&= \min\{\gamma : Y/\gamma \in \{Y : g(Y) \leq f(1)\}\} \\
&= \min\{\gamma : g(Y/\gamma) \leq f(1)\} \\
&= \min\left\{\gamma : \gamma^{-1} g(Y) \leq f(1)\right\} \\
&= \min\{\gamma : g(Y) \leq \gamma f(1)\} \\
&= \{\gamma : g(Y) = \gamma f(1)\} \\
&= (f(1))^{-1} \times g(Y)
\end{aligned}$$

■

Proposition 2 establishes that the dimensionality of DEA models can be reduced under the assumption of output homotheticity. In addition, homotheticity allows us to span the production technology from $P(1)$ (see (Shephard 1970)).

4 Estimating the Base Output Isoquant

One useful method for estimating any isoquant is the order-m estimation procedure (Daraio and Simar 2005). The output distance function $D_O(y,x)$, defined in (3) is expressed relative to the output set $P(x)$ and the basic idea in the order-m procedure is to regard this output set $P(x)$ as the support of a conditional density function

$$P(x) = \left\{y : F_{Y|X}(y|x) > 0\right\}$$

The corresponding support for the joint input output density $H_{X,Y}(x,y)$ is the production possibility set T, i.e.

$$\begin{aligned}
T &= \{(x,y) : H_{X,Y}(x,y) > 0\}, \\
H_{X,Y}(x,y) &= \Pr(Y \geq y, X \leq x) = \Pr(Y \geq y | X \leq x) \Pr(X \leq x) \\
&= F_{Y|X}(y|x) S_X(x),
\end{aligned}$$

where

$$S_X(x) = \Pr(X \leq x).$$

For a fixed level of input x_o let $Y_1, \ldots, Y_m$ be m i.i.d. random output vectors generated from $F_{Y|X}(.|x_o)$, i.e. all output vectors $Y_i, i = 1, \ldots, m$ are random variables that can be produced with x_o with a strict positive probability. Assuming selective (local) convexity of the output sets, the random output set of order-m for units consuming x_o, $P_m^C(x_o)$ is defined as:

$$P_m^C(x_o) = Conv\left[\{y|y \leq Y_i, i = 1, \ldots, m\}\right] \tag{5}$$

The locally convex order-m output efficiency $\theta_m^{LC}(x, y)$ can be defined as (Daraio and Simar 2005):

$$\theta_m^{LC}(x, y) = E_{Y|X}\left[(\widetilde{\theta}_m^{LC}(x, y)\,|X \leq x\right]$$

where

$$\widetilde{\theta}_m^{LC}(x, y) = \sup\left\{\theta|\theta y \in P_m^C(x)\right\}$$

To obtain the estimator

$$\widehat{\theta}_m^{LC}(x, y) = \widehat{E}_{Y|X}\left[(\widetilde{\theta}_m^{LC}(x, y)\,|X \leq x\right]$$

based on a sample of n observations we plug in the empirical version of $F_{Y|X}(.|x_o)$ as

$$\widehat{F}_{Y|X,n}(y|x) = \frac{\sum_{i=1}^{n} 1(, Y_i \geq y, X_i \leq x)}{\sum_{i=1}^{n} 1(X_i \leq x)},$$

where 1() is the indicator function. $\widehat{\theta}_m^{LC}(x, y)$ can be approximated by a Monte-Carlo procedure: Sample m observations $Y_{1,b}, \ldots, Y_{m,b}$ conditional on input being less than $x_o = 1$ with replacement. For each of the n observations find the inverse output distance function value $\widetilde{\theta}_m^{LC,b}(1, Y_l)$ relative to an output set

$$Conv\left[\{y|y \leq Y_{i,b}, i = 1, \ldots, m\}\right].$$

Redo this estimation $b = 1, \ldots, B$ and take the average of the obtained scores as the estimator, i.e. $\widehat{\theta}_m^{LC}(1, Y_l) \approx B^{-1}\sum_b \widetilde{\theta}_m^{LC,b}(1, Y_l)$. From these scores we obtain an estimated input set $\widehat{P}_m^{LC}(1)$ as

$$\widehat{P}_m^{LC}(1) = \left(Conv\left[\widehat{\theta}_m^{LC}(1, Y_1) \times Y_1, \ldots, \widehat{\theta}_m^{LC}(1, Y_n) \times Y_n\right] - \mathbb{R}_+^p\right) \cap \mathbb{R}_+^p. \tag{6}$$

A simplification of the order-m estimator is the conditional estimator of the base isoquant, which avoids the replications by choosing $m = n$.

In the application presented in section 6, we use this conditional estimator instead of the order-m estimator. The base isoquant can be estimated using this conditional model solving the following linear programs

$$\widehat{\theta}_C^{LC}(X_{base}, Y_l) = \begin{cases} \max & \theta - \varepsilon(1, \ldots, 1)\, s & & \\ s.t. & \theta Y_l - \sum_{j=1}^{n} \lambda_j Y_j + s & = & 0 \\ & \sum_{j=1}^{n} \lambda_j & = & 1 \\ & \lambda_j = 0 \text{ if } X_j > X_{base} & \forall j & \\ & \lambda \in \mathbb{R}_+^n, s \in \mathbb{R}_+^p & & \end{cases} \tag{7}$$

$l = 1, ..., n$, where ε is a non-Archimedian, $X_{base} = 1$ in this section and where the estimator $\widehat{P}_C^{LC}(1)$ of the output set is derived as

$$\widehat{P}_C^{LC}(1) = \left(Conv \left[\widehat{\theta}_C^{LC}(1, Y_1) \times Y_1, \ldots, \widehat{\theta}_C^{LC}(1, Y_n) \times Y_n \right] - \mathbb{R}_+^p \right) \cap \mathbb{R}_+^p. \tag{8}$$

A related model appears in the efficiency literature to control for exogenous inputs (Ruggiero 1996), selective convexity (Podinovski 2005) and as the condition estimator (Daraio and Simar 2005). In this formulation, units that are not observed consuming at most the base amount (in this case, one) are not allowed in the solution space. Hence, we simply envelop all output vectors with observed input at most equal to one. Notably, we replace the standard assumption of convexity with selective output convexity of the output sets:

Axiom 7 *Selective output convexity: If* $(X', Y') \in T, (X'', Y'') \in T, X'' < X' \Rightarrow \lambda(X', Y') + (1 - \lambda)(X', Y'') \in T, \lambda \in [0, 1]$

Our primary reason for using the conditional model is *not* to estimate efficiencies but to exploit homotheticity to aggregate multiple outputs into a one-dimensional output index. Hence, we estimate each output isoquant using the conditional estimator and choose the "best" isoquant that has good coverage in the sense that i) we want as many observations playing an active role of spanning the frontier, ii) we want the cone spanned by these observation in output space to be as large as possible and iii) we want the observations to be spread out across the cone as uniformly as possible. After choosing the output isoquant that best meets these desirable criteria, we estimate the output distance of each observation to this isoquant as an index of aggregated output.

To ease the presentation of the proposed methodology, we chose the unit output isoquant as the base in our discussion above. We now provide guidelines for how to choose the input level with the most useful information. Using the conditional estimator relative to a given input level x we only include output vectors from observations with an input level at most equal to this x. We would like to have as many observations as possible available for spanning the isoquant, which tends to suggest a high input level. However, observations consuming a large amount of input may not provide any additional information. Two relevant criteria for selecting an isoquant could be i) many observed points on or just below the isoquant and ii) the points are spread out evenly along the full isoquants. If we knew the positions and the shape of the true isoquants we would look for the "best" output isoquants according to these criteria. Unfortunately, we do not know the locations and the shape of the true isoquants. Hence, we have to rely on an estimator, and in this case we will use the conditional estimator defined above. For each observed input level $X_j, j = 1, \ldots, n$, we use the conditional estimators $\widehat{\theta}_C^{LC}(X_j, Y_l), l = 1, \ldots, n$ which provides us with the estimators $\widehat{P}_C^{LC}(X_j), j = 1, \ldots, n$ of all n output sets corresponding to all n inputs levels. Finally we choose as base output isoquant the specific input level which performs reasonably well according to the following three criteria:

1. A distribution of the angle coordinates of the observed data points on the conditional piecewise linear estimator of the isoquant, which mimics the uniform distribution on the empirical support of the angle coordinates for the whole data set. As a measure of the amount of deviation of the empirical distribution from the uniform distribution we suggest the volume between the two distribution functions.

2. A large number of observed data points is located on the conditional piecewise linear estimator of the output isoquant.

3. As many as possible of all the data points are radially projected to the envelopment of the points on the frontier, i.e. are located in output space within the cone spanned by the points that spans the output isoquant.

5 An estimator of a piecewise linear S-shaped frontier and the inflection point.

Our estimate of aggregate output allows us to analyze an estimator of the S-shaped technology in the single (aggregate) output single input case. Let the true production possibility set (PPS) be denoted T^S, and assume that the boundary of T^S is S-shaped in the sense that we can divide the input axis into two parts $[0, x^*]$ and $[x^*, \infty)$ where the production function is convex (concave) on the first (second) interval. The marginal product is monotonically non-decreasing in $[0, x^*]$ and monotonically non-increasing in $[x^*, \infty)$. As discussed in the introduction we know that the convex hull estimator $\widehat{T}^{BCC}$ (Banker et al. 1984) of the PPS is too large below x^*, but we also know that for input and output above the inflection point x^* this estimator works well, because of the true concave shape of the production function. Hence, we design a procedure that remove or "dig out" the part of the estimator $\widehat{T}^{BCC}$, that violates the S-shape. We focus on a certain convex hull of observed data point that satisfies the following:

- only points below (and on) the inflection point are part of the convex hull, i.e. points that are supposed to reflect the convex IRS part of the technology
- no point is located above the frontier (or equivalently, no points are located in the interior of this hull)

Figure 1 illustrates this idea using 5 input output observations generated from an "S-shaped" data generating process (DGP). Observations, A, D and E are BCC-efficient and observation D is most productive scale size (mpss). In this small illustrative example we use the mpss as an estimator of the inflection point[3]. In other words we assume that the production function is convex up to data point D and concave to the right of this point. The basic idea behind the digging approach is to determine a subset of all FDH-efficient DMUs "below" mpss which determines a convex hull $\widehat{T}^{Dig}$, where none of these DMUs belongs to the interior of this hull. An estimator $\widehat{T}^{S}$of the PPS with an S-shape with an efficient boundary being piecewise linear is now available as $\widehat{T}^{BCC} \backslash \widehat{T}^{Dig} \equiv \widehat{T}^{S}$, i.e. the convex hull BCC estimator of the PPS minus the convex hull $\widehat{T}^{Dig}$. In Figure 1 the estimated $\widehat{T}^{BCC}$ is the convex hull of observations A, D, E set added to $\mathbb{R}_+ \times \mathbb{R}_-$ (strong input and output disposability). $\widehat{T}^{S}$ is estimated as $\widehat{T}^{BCC} \backslash \widehat{T}^{Dig}$, where $\widehat{T}^{Dig}$ is the convex hull of the observations A, B, C and D.

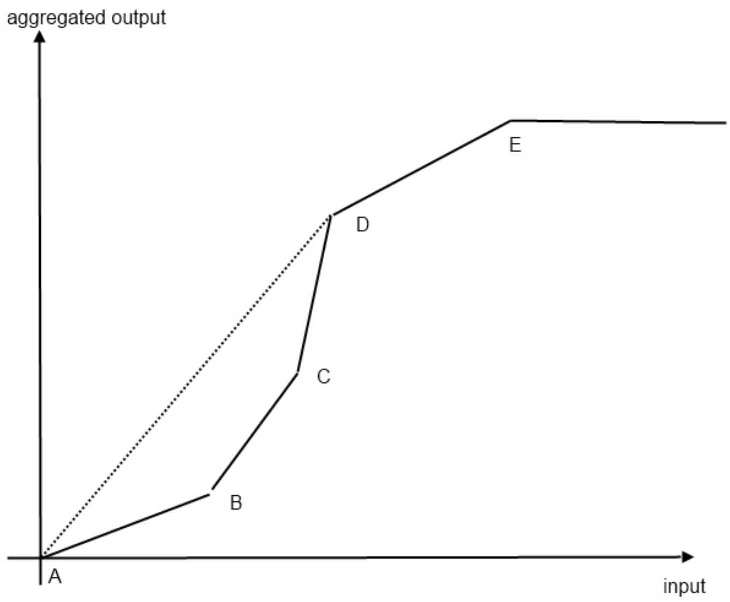


Figure 1. The S-shaped frontier.

Unfortunately, the determination of $\widehat{T}^{Dig}$ is not unique. Examples of non-uniqueness are provided in (Olesen and Ruggiero 2012), and a more general estimation procedure is suggested. Specifically, it is suggested to choose that specific solution to the piecewise linear S-shape that maximizes the number of FDH-efficient points located on the frontier.

6 Application to US West Coast Groundfish production.

We now turn to an application of the proposed estimation of a possibly S-shaped production relationship the West Coast Groundfish production 2004-2007. One main focus for our analysis is: Can we find indications of a purely increasing return to scale segment of the production of harvest of fish? Our estimation procedure relies on a maintained hypothesis of an output homothetic production. We avoid any specific function form of the production relation but using a non-parametric estimation procedure implies that we need many observations, especially is we want to control for many exogenous factors.

Performance measurement of West Coast groundfish production is complicated by the fact that it is a regulated industry. The landings limits vary in different regions and the fleet deployed includes both large and small vessels. This of course questions a comparison of the performance of vessels of different sizes of vessels of equal size harvesting in areas with different limits. One way to approach these problems is to subdivide the sample into smaller groups of more homogeneous sub-fleets with regard to the quasi-fixed input vessel-size and the West Coast areas targeted for harvest. However, it is well know that subdividing the sample worsen the impact of the curse of dimensionality making it difficult to get accurate estimation results. It is also well known that non-parametric estimation procedures as the ones used in this study is particularly sensitive to the curse of dimensionality. Without a priory imposed structure we have to extract structure from data which implies the need for large data sets.

In this study data allows us to specify a model with one variable input (days at sea), one quasi-fixed input (vessel size) and then at least four different outputs (types of catch). We will however in this paper focus on a simplified model where we partly ignore the quasi fixed input vessel size and we will use two aggregated outputs. Furthermore, we use only data from California vessels fishing south of Humboldt Bay, where the landings limits are the same for each vessel. A simple approach to control for the quasi-fixed input would be to expand model (7) with the following additional constants:

$$\lambda_j = 0 \text{ if } X_j^{\text{quasi-fixed}} > X_{base}^{\text{quasi-fixed}}, \forall j$$

In other words, for each estimated output isoquant at input level X_{base} discard the $j'th$ observations if

$X_j > X_{base}$ or $X_j^{\text{quasi-fixed}} > X_{base}^{\text{quasi-fixed}}$. Preliminary sensitivity results indicate that the isoquants with good coverage are unaffected by this additional control for the quasi-fixed input.

The application of the model is based on a data set with catch records from fishery in the Pacific Ocean during the period 2004-2007. The entire catch record for each trip from each vessels is used. Primarily, based on the condition of the fishery the catch is aggregated into the following 4 outputs,

i) the catch in lbs. of dover sole, sablefish and thoryhead rockfish (generally encountered at depths greater than 200 fathoms),

ii) the catch in lbs. of a nearshore mix species aggregate (mainly petrale sole, rex sole, sanddabs and other flatfish),

iii) the catch in lbs. of various rockfish species which are harvested mainly along the continental shelf at depths between 75 and 150 fathoms, and finally

iv) the catch in lbs. of California halibut which is harvested in shallow water less than 75 fathoms.

We have chosen to define a DMU as a vessel in a specific year. Hence, we sum the catch in these four categories and the variable input "days at sea" for fixed vessel for each of the four years. The total data set comprises 192 vessel-year combinations. We have removed five outlier observations which brings us to a final sample size of 187.

To facilitate a geometric presentation of the aggregation procedure we aggregate i) and iii) by simply adding the catch in lbs. in these two categories. The same procedure is used to aggregate ii) and iv).

Using the guidelines from Section 4 for how to choose the input level with the most useful information we will now use the conditional estimator relative to a given input level x and only include output vectors from observations with an input level below or equal to this x. We would like to have as many observations as possible available for spanning the isoquant. We look for a specific isoquant (a x level) which performs well on all the criteria outlined in Section 4.

We sort the data on the variable input "days at sea" and estimate the conditional model (7) for each of the 187 input levels; the solution space for each estimated isoquant is conditioned such that only DMUs with inputs less than or equal to the i'th DMUs input, $i = 1, \ldots, 187$ are included. We thus obtain 187 output oriented scores for each isoquant. If the output oriented score has additional slack in any of the output dimensions the score is assigned the value "missing". Based on the these results, we identify for each potential base isoquant only those points that span the conditional isoquants (i.e. only observations with output oriented score equal to one with no additional slack are included). For this application, we consider three criteria for choosing our base isoquant.

Firstly, we are looking for the particular isoquant with as many observation on the frontier as possible. Secondly, focussing on the points spanning the isoquant we search for sets of points with an empirical distribution of the angles as close as possible to a uniform distribution. Thirdly, we want the spanning points to span as large a cone as possible in output space compared to the cone spanned by the full sample of data points.

Figure 2 illustrates the first criterion. On the horizontal axis we have the 187 different estimators of the output isoquants. On the vertical axis we measure the number of data points on or close to the estimated frontier. The red curve includes the number of points with an output oriented score in the band [1,1.001]. Isoquant 60 seems to be performing well on this criterion. Figure 3 illustrates the second criterion. On the vertical axis we measure the deviation of the output mix from the uniform distribution. For a description of the exact measure of deviation from the uniform distribution we refer the reader to section 3 in (Olesen and Ruggiero 2012). Finally, Figure 4 illustrates the third criterion, where for each of the 187 estimated isoquants we have counted the number of sample points being projected to the envelopment of the points spanning the frontier, but without the presence of slacks. In other words we have counted the number of observation being in output space inside the cone spanned by the output vectors spanning each of the 187 estimators of the isoquant. Again we see that isoquant 60 is a promising candidate, but the isoquants 102-115 do seem promising too, especially if we expand the analysis by replacing the simple conditional estimator with the more complicated order m estimator that can take advantage of the fact, that we have additional observations present close to the envelopment.

Figure 2: The 1. criterion: The number of data points located on (blue) or on or close below (red) the estimated frontier.

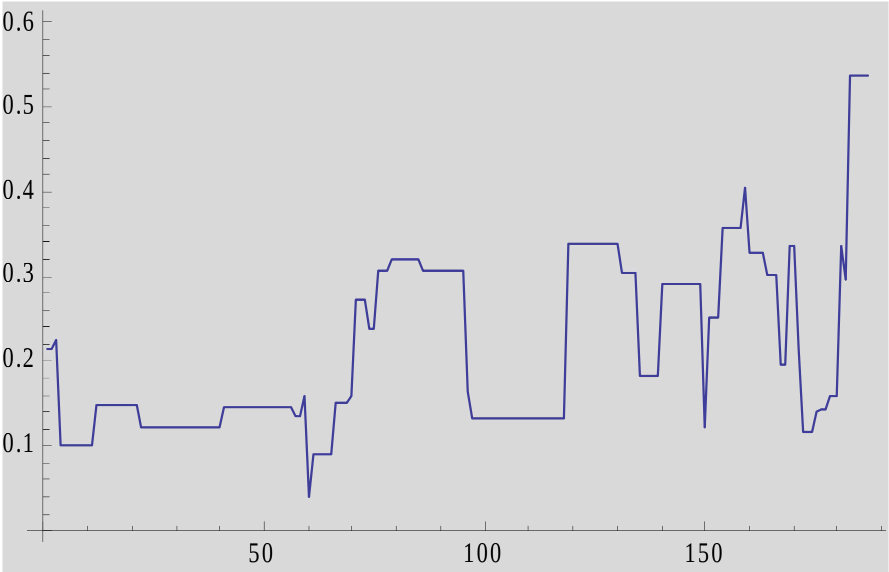


Figure 3: The 2. criterion. The deviation of output mix from the uniform distribution.

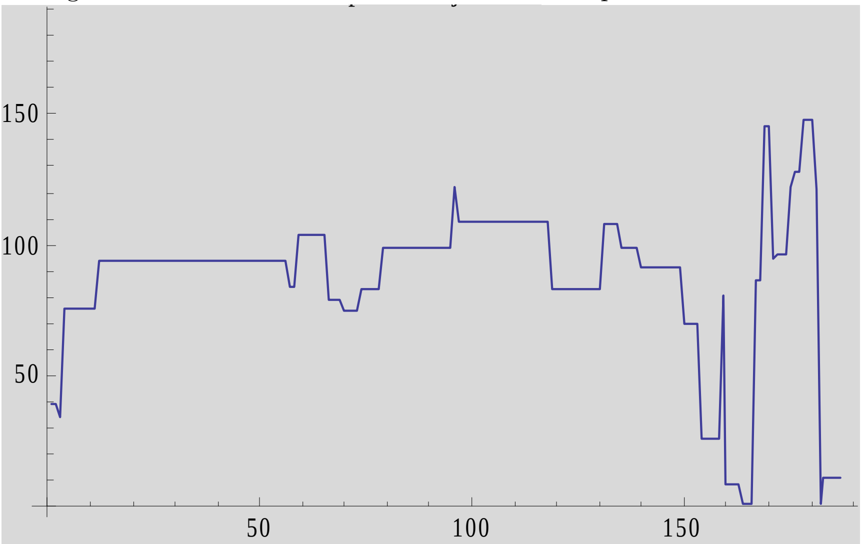


Figure 4. The 3. criterion: The number of sample points projected to the estimator of the isoquant (without slacks).

Taken together, we are looking for the isoquants spanned by many of points that uniformly span the isoquant. Several isoquants perform well on all criteria; for our analysis, we chose isoquant 60 as our base isoquant.

The inverse of the output oriented efficiency scores relative to isoquant 60 are used as indexes of aggregated output. Initiating the estimation of the inflection point and the S-shaped piecewise linear production function (one input, one aggregated output) we first remove the 83 observations with positive slacks present in the estimation of the aggregate output index and 88 observations that are FDH inefficient. That leaves us with a data set of 18 observations of which 8 (10) are above (on or below) most productive scale size (mpss). We estimate the BCC efficiency scores based on the sample of observations from sample point 10 to 18 since this concave part of the production function is unaffected of the exact choice of inflection point. Scores of one of course indicate observations that contribute to the spanning of this concave part of the frontier. We remove 5 of the 9 points below mpss because these points turn out to be in conflict with the requirement that the marginal product along each of the facets making up the piecewise linear estimator moving from origin to mpss must be monotonic increasing. Finally, we add the origin as a feasible production plan. This is illustrated in Figure 5d, where the origin and the 5 remaining points are numbered $1, \ldots, 6$.

A general approach that provides a piecewise linear estimator of the S-shaped technology using the inverted convex hull that maximizes the number of FDH efficient points on the S-shaped frontier, i.e that maximizes the number of "S-shaped efficient" points is presented in model (24) in (Olesen and Ruggiero 2012). An integral part of this procedure is an estimation of the inflection point; we seek as an inflection point on or below mpss a FDH efficient observation that allows for an estimated S-shaped frontier with a maximum number of FDH efficient points on the frontier. Testing a given point as a candidate for the inflection point involves several conditions: i) the marginal products along the facets from the origin to the inflection point must be increasing, ii) the marginal products must be non-increasing on facets above the inflection point and iii) all the points have to be on or below the frontier.

The general model (24) in (Olesen and Ruggiero 2012) is estimating a convex shape as a graph through a subset of the FDH efficient points starting at $(0, 0)$ and ending at the estimator of the inflection point. This convex shape is constructed as a graph in a network going through a subset of these points and leaving no points on the wrong side of the graph. However with the present data set the solution to the S shape is quite obvious which means that we do not need this complicated procedure. Looking at Figure 5d only the point labeled "5" is feasible as the inflection point.

In general, without an obvious candidate for the inflection point, we would have to get the solutions corresponding to the inflection point located at point "1", "2", "3","4" and "5". Hence, formally we would have to go through the following steps for $i \in \{1, \ldots, 5\}$:

1. The BCC model is solved including only points above point "$6 - i$". We count the number of points on the frontier and estimate the termination marginal product of the facet from point "$6 - i$" to point "$6 - i + 1$".

2. Focusing on "$6-i$" as the candidate for the inflection point we solve (24) in (Olesen and Ruggiero 2012) (including the sequence of cuts) to determine a sequence of binary variables indicating a path through a number of points below point "$6 - i$" starting at the origin and ending at point "$6 - i$" and with a

monotone non decreasing marginal product along the path and no points above the path. The optimal solution (assuming that one exists) will provide the count of points on this convex part of the frontier.

3. In addition we require that the marginal product on the facet from point "$6 - i - 1$" to point "$6 - i$" is greater than or equal to the termination marginal product estimated in step 1.

4. Finally, we add the counts of points on the frontier from step 1 and step 2.

The discussion above is illustrated in Figure 5a,b,c,d. Figure 5a illustrates the total sample of data points with non-missing output oriented efficiency scores relative to output isoquant 60. The final estimated S-shaped frontier solution is indicated in this figure as well as a piecewise linear non-convex shape. The estimation of this S-shaped frontier requires the following step: First and foremost we delete all FDH inefficient observations which leaves us with the data points in Figure 5b. Clearly, any FDH inefficient DMU can not contribute to the spanning of the S-shaped frontier. Second we determine the most productive scale size (mpss). The subset of FDH-efficient point on and above the mpss will be enveloped using the traditional convex hull estimator used in the BCC-DEA model. Thirdly, to estimate the part of the S-shape on and below mpss we focus on the subsample of FDH-efficient points shown in Figure 5c. We impose the requirement that the marginal product along the facets must be monotonically increasing up to mpss. Hence of the 10 points in Figure 5c we delete the 2th, the 6th,the 7th, the 8th and the 9th observations. These 5 observations can not be on the frontier if we insist that the marginal product along facets must be monotonically increasing. Finally, we are left with 5 observations and the origin as the set of points to be used in the procedure for estimating the best estimator of the inflection point. However, as noticed above in this specific case it is evident that observation 5 in Figure 5d is the only feasible candidate for the inflection point. If we try to find a solution insisting that e.g. point 4 is the inflection point then no feasible solution exist (if we use a facet spanned from point 4 to point 6 then point 5 is on the wrong side of the frontier). The final S-shape solution seems to indicate an increasing returns to scale segment from the origin up to approximately 20 days at sea.

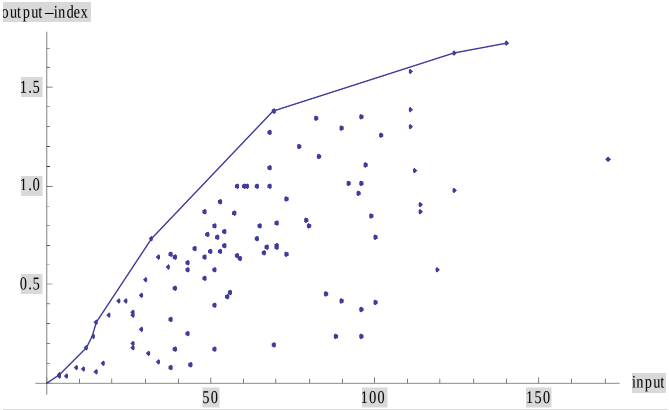


Figure 5a.

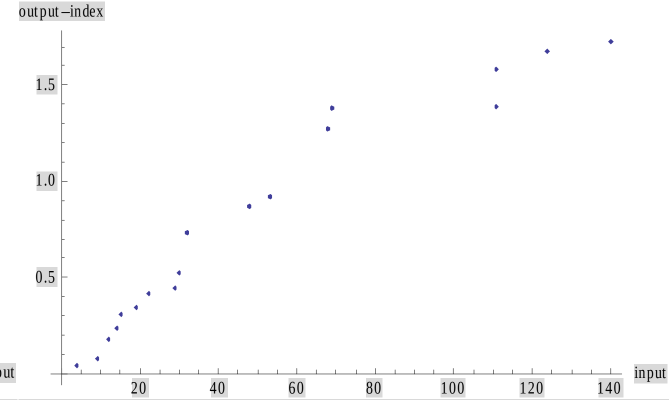


Figure 5b.

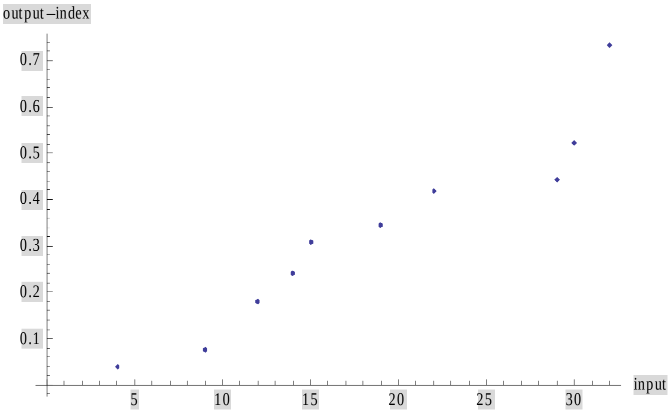


Figure 5c.

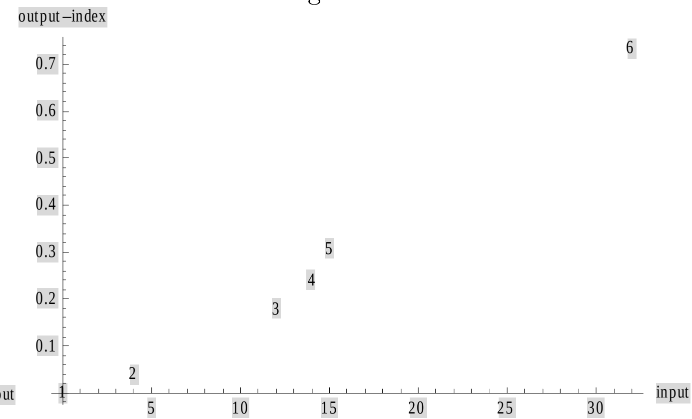


Figure 5d.

Figure 5. Graphical illustrations of the construction of the S-shaped input consumption function

7 Conclusion and further research

A maintained hypothesis of convexity in input-output space is often used in DEA estimations of efficiency scores. However, convexity is not consistent with standard microeconomic production theory that posits an S-shape for the production frontier. In a recent paper, (Olesen and Ruggiero 2012) propose an approach that allows for an estimation of efficiency from an S-shaped technology for the multiple inputs and one output case. The approach relies on an assumption of input homotheticity for the case of one output and multiple inputs. This assumption has allowed us to split the estimation procedure into two parts, i) an aggregation procedure based on the structure of input homotheticity, and ii) a joint estimation of the inflection point and a piecewise linear S-shaped structure for one aggregated input and one output.

The illustrative application in this paper is based on catch records with multiple output, one variable input and one quasi fixed input. We have chosen in this paper not to implement a full multiple input multiple output model. The approach suggested in (Olesen and Ruggiero 2012) assumes one output and multiple inputs with input homotheticity. It is straight forward to change focus to one input multiple outputs maintaining output homotheticity, which is the strategy chosen in this paper. Notice however, that maintaining input *and* output homotheticity with multiple inputs and outputs requires a more complicated estimation procedure of the isoquants because we simultaneously have to estimate both a base input and a base output isoquant.

The illustrative application presented in this paper is based on a data set with catch records from fishery in the pacific ocean of the US West Coast groundfish production for the period 2004-2007. We have aggregated outputs into two different outputs and used one variable input. Hence we estimate the S-shaped production relation based on an assumption of output homotheticity which allows us to aggregate the outputs into an output index.

Taking advantage of the reduced dimensionality (one aggregated output and one input) we have used the model proposed in (Olesen and Ruggiero 2012) to estimate a piecewise linear S-shaped frontier. In other words, we have assumed that the boundary of the true PPS is S-shaped in the sense that we can divide the input axis into two parts, where the frontier is convex (concave) on the first (second) part. Consequently, the convex hull estimator is too large and we have used a "digging approach" where we remove the part of the PPS that violates this S-shape. This digging approach is formulated as a joint estimation of the inflection point and the convex part of the frontier from the origin to the inflection point.

In Section 6 we have used this approach to analyze and estimate scale characteristics of the US West Coast groundfish production. One main focus for our analysis is: Can we find indications of a purely increasing return to scale segment of the production of harvest of fish? Our estimation procedure rely on a maintained hypothesis of an output homothetic production. The empirical analysis shows that only very modest increasing returns seems to exists. With "days at sea" in the interval [1,20] we do see increasing returns to scale. With "days at sea" above 30 days decreasing returns emerge, and becomes severe especially above 60 days.

As an estimation procedure for the base output isoquants we propose in this study, assuming selective output convexity, the use of a simplified order-m estimation procedure. An important extension of the analysis presented here would be to use a full order-m estimation procedure and allow more than two outputs.

The approach used in this paper has two apparent shortcomings. First and foremost we have assumed output homotheticity which may or may not be a reasonable assumption. An important extension of the analysis presented here is to allow at a less restrictive output structure. Secondly, only one variable input is used, although preliminary sensistivity results seems to indicate that controlling for the quasi-fixed input: vessel size will not alter the shape of the estimated base isoquant. An important theoretical extension would be to allow for a full multiple input and multiple output analysis based on an assumption of joint input and output homotheticity.

Endnotes

1. Observe that the law is satisfied for movements along increasing returns to scale facets, since average productivity in this case increases with marginal productivity unaffected.
2. See also (Ruggiero 1996) and (Podinovski 2005).
3. In (Olesen and Ruggiero 2012) we propose a more general estimator of the inflection point.

References

Banker, R. D., Charnes, A. and Cooper, W. W.: 1984, Some models for estimating technical and scale inefficiencies in Data Envelopment Analysis, *Management Science* **30**(9), 1078–1092.

Bogetoft, P.: 1996, DEA on relaxed convexity assumptions., *Management Science* **42**, 457–465.

Cazals, C., Florens, J. P. and Simar, L.: 2002, Nonparametric frontier estimation: A robust approach, *Journal of Econometrics* **106**, 1–25.

Daraio, C. and Simar, L.: 2005, Conditional nonparametric frontier models for convex and nonconvex technologies: A unified approach., *Journal of Productivity Analysis* **28**, 13–32.

Färe, R. and Lovell, C. A. K.: 1988, Aggregation and efficiency, *in Measurement in Economics ed. by W. Eichhorn. Physica-Verlag Heidelberg 1988* pp. 639–647.

Frisch, R.: 1965, *Theory of Production.*, D. Reidel Publ. Comp., Dodrecht-Holland.

Førsund, F. R. and Hjalmarsson, L.: 2004a, Are all scales optimal in DEA? theory and empirical evidence., *Journal of Productivity Analysis* **21**(1), 25–48.

Førsund, F. R. and Hjalmarsson, L.: 2004b, Calculating scale elasticity in DEA models., *Journal of Operational Research Society* **55**(10), 1023–1038.

Olesen, O. B. and Petersen, N. C.: 2011, Scale properties in data envelopment analysis, *Workingpaper, forthcomming in Omega* pp. 1–36.

Olesen, O. B. and Ruggiero, J.: 2012, Maintaining the regular ultra passum law in data envelopment analysis, *Submitted workingpaper: See publications at www.sdu.dk under Department of Business and Economics* pp. 1–25.

Petersen, N. C.: 1990, Data envelopment analysis on a relaxed set of assumptions, *Management Science* **36**(3), 305–314.

Podinovski, V. V.: 2005, Selective convexity in dea models, *European Journal of Operational Research* **161**, 552–563.

Ruggiero, J.: 1996, On the measurement of technical efficiency in the public sector., *European Journal of Operational Research* pp. 553–565.

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

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

Adventures in Modeling Good and Bad Outputs:
A 30 Year Retrospective

by

Carl A. Pasurka, Jr.*
(Corresponding author)

*U.S. Environmental Protection Agency (1809T)
Office of Policy, Economics and Innovation
1200 Pennsylvania Ave., N.W.
Washington, D.C. 20460

Phone: (202) 566-2275
FAX: (202) 566-2373

E-Mail: PASURKA.CARL@EPA.GOV

July 14, 2012

An earlier draft of this paper was presented at the National Marine Fisheries Productivity Workshop in Santa Cruz, CA (June 12, 2012). I wish to thank Rolf Färe for helpful comments on an earlier draft of this paper, and discussions that improved my understanding of several topics covered in this paper. All errors, opinions, and conclusions are those of the author and should not be attributed to the U.S. Environmental Protection Agency.

Abstract

The environmental movement of the late 1960s coincided with interest in modeling the consequences of environmental regulation on economic activity. In order to avoid problems associated with using survey data of the cost of inputs assigned to pollution abatement, modeling the joint production of good and bad output production has emerged as an alternative strategy for determining the cost and productivity consequences of pollution abatement. After a brief survey of early research on the joint production of good and bad outputs, the paper focuses on the evolution of data envelopment models based on the environmental technology proposed by Färe and Grosskopf (1983). After surveying important changes in the specification of the production technology and selected applications, the paper concludes by highlighting challenges confronting this strategy for assessing the effect of pollution abatement.

I. Introduction

The environmental movement of the late 1960s coincided with interest in modeling the consequences of environmental regulation on economic activity. The growing interest in implementing regulations to restrict bad outputs (i.e., emissions of pollutants) associated with producing and consuming good outputs (i.e., marketed desirable goods and services) led to concerns about the effect of these regulations on the production of the marketed good outputs. [1,2] This concern developed because for a fixed technology and input vector, reductions in emissions are achieved at the cost of reduced production of marketed goods and services. As a result, most early studies of the productivity, technical change, and technical efficiency consequences of pollution abatement focused on the good outputs of production. Over time, several pollution abatement strategies for reducing bad output production emerged.

The first abatement strategy available to a decision making unit (DMU) involves simply reducing good output production.[3] If a producer is restricted to a single process (e.g., a Leontief fixed coefficient technology), the only option for reducing bad output production is a proportional contraction of good and bad outputs along the process ray. Due to the lack of

[1] Throughout this study, "good" output refers to the desirable marketed good produced by a decision making unit (DMU), while "bad" output refers to the undesirable byproducts (i.e., emissions of a pollutant) of good output production.

[2] While Stone (1972, p. 412) referred to "Goods" and "Evils," "desirable" and "undesirable" outputs or "goods" and "bads" emerged as the preferred terminology for discussing marketed outputs and their undesirable by-products.

[3] Pollution abatement does not eliminate the undesirable byproduct of good output production. Instead, abatement activities transform the byproduct from - for example - one media (air) to another media (solid waste) where it constitutes a reduced threat to human health and the environment.

flexibility, this strategy yields the highest opportunity cost (i.e., foregone production of the good output) of pollution abatement.

The second pollution abatement strategy involves reassigning inputs from producing good outputs to pollution abatement. This strategy has two subcategories. First, end-of-pipe (EOP) abatement processes require a separate technology to be installed on a pipe or smokestack (e.g., a flue gas desulfurization system) prior to the release of the bad output. The second subcategory of reassigning inputs is a change-in process (CIP) strategy in which abatement activities are integrated into the production process.[4]

Because pollution abatement reallocates inputs from producing marketed goods and services to activities that produce outputs without prices (i.e., reduced levels of emissions), there is likely to be a decline in the production of marketed goods and services used to satisfy household consumption. While a short-run (i.e., fixed technology and fixed input vector) analysis finds that pollution abatement results in a decline in the material living standard of a society, pollution abatement positively affects its standard of living by reducing bad output production which improves its environmental quality.

The third pollution abatement strategy requires improving the quality of inputs. For example, if an electric power plant wants to reduce SO_2 emissions, it can switch from high-sulfur to low-sulfur coal (i.e., improving the fuel quality). However, low-sulfur coal has lower heat content (i.e., fewer BTUs per ton) than high-sulfur coal. Because the switch to low- sulfur coal effectively reduces fuel inputs (i.e., in BTUs), the opportunity cost of reducing SO_2 emissions is

[4] A major problem associated with the introduction of CIP activities is determining which inputs in an integrated production process are assigned to pollution abatement.

the reduced level of good output production (i.e., electricity) resulting from the decline in the heat content of the fuel.

While few dispute the benefits that environmental regulations provide society in the form of improved environmental quality and the resulting improvement in the quality of life, concerns remain about how pollution abatement affects the economic viability of firms, industries, and nations. There are two major concerns about costs of pollution abatement. First, there are concerns about whether the regulations are optimal in the sense that marginal benefits equal marginal costs. However, most empirical assessments of the welfare effects of environmental regulations rely not on incremental (i.e., marginal) costs and benefits, but on the total costs and benefits of a regulation. Hence, regulations pass the benefit-cost test when total benefits exceed total costs.

A second concern about the costs of environmental regulations is their impact on competitiveness (see Pasurka 2008 for a recent survey). The hypothesis is that imposing regulations results in higher production costs and declining the competitiveness of the DMUs that are subject to the regulations.

While substantial resources have been devoted to the development of techniques to assess the benefits of environmental regulations, fewer resources have been devoted to models that calculate the opportunity costs of regulations. While early efforts to assess the cost effects of environmental regulations often relied on engineering models, in later years many studies used data from surveys that collected information on the cost of inputs assigned to pollution abatement (see Pasurka 2008).

Data from surveys of pollution abatement costs constitute the basis of “assigned input” models. These models investigate the association between the cost of inputs assigned to pollution abatement and productivity measures focused solely on good output production. These models require information on the production of good outputs, the inputs assigned to the production of the good output, and inputs assigned to pollution abatement. However, they require no information on bad outputs. One version of the assigned input model is implemented in the following manner. Using only the inputs assigned to good output production, a production function is estimated. The parameters from this function are then used to calculate the good output production if inputs assigned to pollution abatement are made available for good output production. The difference in good output production represents the opportunity cost of pollution abatement. For example, Aiken, et al. (2009) implemented the assigned input model using a data envelopment analysis (DEA) framework by employing information on the capital stock assigned to pollution abatement in manufacturing industries in Germany, Japan, the Netherlands, and the USA to investigate the effect of pollution abatement on changes in good output productivity.

While surveys of pollution abatement costs remain a popular strategy for identifying abatement costs, concerns have been expressed about the quality of information generated by these surveys. Initially, EOP technologies were the preferred strategy for reducing bad outputs. Because EOP technologies are separate technologies, this simplifies the task of determining which input costs are assigned to pollution abatement. Over time, producers shifted from EOP strategies to CIP strategies. Because of the difficulty associated with determining the share of the cost of an integrated technology that represent pollution abatement costs, estimates of pollution abatement costs for CIP technologies are more problematic.

As a result of the increased use of CIP abatement technologies to reduce bad outputs, modeling a separate pollution abatement technology became an increasingly difficult task.

Because of the challenges associated with assigning inputs to either good output production or pollution abatement, there was an incentive to develop another framework for measuring the association between pollution abatement and productivity. Instead of assigning inputs to good output production and pollution abatement, the alternative method involves modeling the joint production of good and bad output production by regulated and unregulated technologies. We refer to these as "joint production" models. As is the case for the assigned input model, joint production models view the reduced good output production associated with abatement activities as the opportunity cost of pollution abatement. However, there are several advantages to estimating pollution abatement costs (PAC) by modeling the joint production of good and bad outputs. One advantage is that it does not require information on which inputs employed by a DMU are assigned to pollution abatement, nor is it necessary to know the quantity of reduced bad output (i.e., abated emissions). Instead, PAC is the value of the foregone good output production associated with reducing bad output production (i.e., the "output" of pollution abatement). Another advantage of modeling the joint production is it avoids the difficulties associated with survey efforts to estimate the cost of abating pollution when CIP abatement techniques are installed. A third advantage is synergies that exist when abating two or more pollutants are automatically captured by the joint output technology, while explicit pollution abatement functions require information about synergies existing among different pollution abatement processes.

While there was immediate interest among economists in employing models that use the cost of inputs assigned to pollution abatement, interest in modeling the joint production of good and bad outputs took longer to gain acceptance. Unlike Zhou, Ang, and Poh (2008a), who compiled the only existing survey of the characteristics of nonparametric DEA models with good and bad outputs, this paper attempts to describe the evolution of the art of modeling good and bad outputs. By focusing on differences in model specification, this paper bears a closer resemblance to the Tyteca (1996, 1997) surveys of strategies to measure the environmental performance of firms. However, this paper does not compare empirical results generated by different models (see Zhou, Ang, and Poh 2008b). The remainder of this survey is organized in the following manner. Section II reviews pre-DEA non-parametric models of good and bad output production. Section III reviews DEA non-parametric models of good and bad output production and selected applications beyond calculating technical efficiency, productivity change, and the opportunity costs of regulations. Section IV outlines extensions of the joint production model, while Section V reviews parametric joint production models. Section VI surveys cost function models of good and bad output production, and Section VII outlines challenges confronting efforts to model the joint production good and bad outputs.

II. Pre-DEA Non-parametric Models of Goods and Bads

Ayres and Kneese (1969) and Leontief (1970) represent the initial efforts to incorporate bad output production and pollution abatement into a general equilibrium framework.[5] Subsequent research primarily used an input-output framework in which each sector (i.e., DMU) had a single process at its disposal. As a result, the sole strategies available to reduce bad output production were (1) reducing good output production by the sector or (2) assigning inputs to a pollution abatement sector whose output was reduced bad output production. Using a joint production input-output model, with a separate pollution abatement (i.e., "anti-pollution") sector,

[5] Ethridge (1973) developed a theoretical model of a firm producing good and bad outputs.

Lowe (1979) represents an early attempt to calculate the shadow price of reducing bad output production within a joint production model.

An early attempt to model good and bad output production within an activity analysis framework is illustrated in Figure 1 (Kohn, 1975, p. 29-33). Unlike input-output models, which assumed a single process for each sector, Kohn's model accommodates multiple processes in which each process generates one good output (y) and one bad output (b) from a vector of inputs. In Figure 1, three processes (P_1, P_2, and P_3) are available to the DMU. Each process is represented by a ray from the origin and represents different combinations of good and bad outputs that can be produced by a given technology and input vector. A steeper ray (e.g., P_1) represents a process that assigns relatively more inputs to pollution abatement, while a relatively flat ray (e.g., P_2) represents a process with fewer inputs assigned to pollution abatement. Once the technology and input vector is known, it is possible to determine the combination of good and bad outputs associated with a process (i.e., points d, e, f), and construct linear combinations of the different processes.[6] For the case of one good and one bad output, the producer can use one or two processes to produce a variety of combinations of good and bad outputs. In Figure 1, process P_1 can be used to produce the bundle represented by point S and process P_2 can be used to produce bundle represented by point V. Combining these output bundles yields point Z on the production frontier line segment de. Implementing this procedure for the three processes yields the production frontier 0def0.

The activity analysis joint production framework would later be incorporated into computable general equilibrium (CGE) models. Willett (1985) and Shortle and Willet (1987)

[6] In Figures 1, 3-6, 8, and 10, lower case letters represent observations, while upper case letters are points generated by the models.

specified theoretical general equilibrium models using an activity analysis specification of the joint production technology to incorporate bad outputs into a CGE model. Komen and Peerlings (2001) also used this approach to specify input-output vectors representing three technologies (two active and one latent), which represented the availability of different mixes of good and bad outputs in the dairy farming sector.

Another variation of the joint production model used in CGE models assumes bad output production is a fixed proportion of good output production (see Smith and Espinosa, 1996, and Lee and Roland-Holst, 1997). Because only one process is available to each polluting industry, the only abatement strategy is a proportional reduction in good and bad output production.

III. DEA Non-parametric Modeling of the Joint Production of Goods and Bads

A. Environmental Technology

The early efforts to model the joint production of good and bad outputs failed to gain traction as a foundation for subsequent research. This changed when Färe, Grosskopf, and assorted co-authors proposed modeling the joint production of good and bad outputs within a data envelopment analysis (DEA) framework.[7] All joint production models start with several premises. From there, the framework can be modified to address a variety of research questions.

First, we specify the environmental technology. This technology incorporates weak disposability of outputs and null-jointness. The later concept tells us that producing good outputs requires producing bad outputs.

[7] An alternative approach that has been proposed involves modeling emissions as "bad" inputs in the production process. Koopmans (1951, p. 38, footnote 5) mentions the possibility of modeling undesirable byproducts as negative outputs (i.e., inputs).

Before proceeding, some notation must be introduced. Inputs are denoted by $x = (x_1, ..., x_N) \in \Re^N_+$ good outputs by $y = (y_1, ..., y_M) \in \Re^M_+$ and bad or undesirable outputs by $b= (b_1, ..., b_J) \in \Re^J_+$.

We apply output sets to model the environmental technology, i.e.,

$$P(x) = \{(y, b): x \text{ can produce } (y, b)\}, x \in \Re^N_+$$

For each input vector x, the output set P(x) is the combinations of good and bad outputs (y, b) that can be produced by that vector. In order to model the opportunity cost of reducing the bad outputs (i.e., the quantity of good output that must be foregone), we impose the assumption that good and bad outputs (y, b) are together weakly disposable.[8] This allows us to model the technology when bad outputs are regulated.

This environmental technology must satisfy the following standard axioms:

P.1. $\{0\} \in P(x)$ for all $x \in \Re^N_+$

P.2. P(x) is compact $x \in \Re^N_+$

P.3. $P(x) \subseteq P(x')$ if $x' \geq x$

These axioms tell us that inactivity is always possible (P.1.), that finite inputs can only produce finite outputs (P.2.), and that inputs are freely disposable (P.3.).

In order to specify the environmental technology, the technology must also meet two environmental axioms. First, the weak disposability of outputs:

P.4.W. $(y, b) \in P(x)$ and $0 \leq \theta \leq 1$ imply $(\theta y, \theta b) \in P(x)$

---

[8] Weak disposability is the proportional reduction in all good and bad outputs. Shephard (1970) introduced this concept.

Hence, if x can produce outputs (y, b), then it is feasible to proportionally reduce the outputs. This axiom can be contrasted with the usual strong disposability condition:

P.4.S. $(y,b) \in P(x)$ and $(y',b') \leq (y,b)$ imply $(y',b') \in P(x)$

This condition allows for non-proportional reduction in both good and bad outputs. In principle one can costlessly dispose of outputs. While this may make sense for the good output, it does not for the bads since we assume the existence of environmental regulations. If regulations do not exist, then bad outputs can be treated as being freely disposable. This represents the technology when the bad outputs are unregulated.

The second environmental axiom is null-jointness or the by-product axiom.

P.5. $(y,b) \in P(x)$ and $b=0$ imply $y=0$

Here the bad outputs, b, are by-products of the good outputs, y. This axiom tells us that if we produce good outputs then some bad outputs will also be produced. In summation, "there is no fire without smoke."

We assume, for simplicity, that good outputs are freely disposable, i.e.,[9]

P.6. $(y,b) \in P(x)$ and $y' \leq y$ imply $(y',b) \in P(x)$

In summation, the environmental technology assumes good outputs are freely disposable and good and bad outputs are jointly weakly disposable. We can illustrate the environmental technology using an output set P(x).

The environmental technology illustrated in Figure 2 meets the two environmental axioms. First, for any observed (y, b) in P(x) its proportional contraction (θy, θb) is also

[9] When specifying the environmental technology, Färe and Grosskopf (1983) and Färe, Grosskopf, and Pasurka (1986) assumed the good and bad outputs were jointly weakly disposable. Färe et al. (1989) introduced the environmental technology that assumed the good output was freely disposable.

feasible, i.e., it belongs to P(x). Second the only point in common between the good output (y-axis) and the output set P(x) is the origin 0, i.e., b is a by-product of y, or y is null-joint with b.

B. Technical Efficiency and Pollution Abatement Costs

The environmental technology can be specified as either a non-parametric or parametric model. The non-parametric DEA models employ a piece-wise linear specification of the joint production technology composed of observations reflecting the actual behavior of DMUs. This allows us to construct a piece-wise linear production frontier for a given technology and input vector. Using a distance function, Färe and Grosskopf (1983) and Färe, Grosskopf, and Pasurka (1986) were the first to formally model the technical efficiency of DMUs that jointly produce good and bad outputs. In these papers, good and bad outputs are treated symmetrically. In other words, a DMU is credited for simultaneously expanding production of both its good and bad outputs. In order to calculate the PAC, it is necessary to specify two technologies whose specification differs in their treatment of the disposability of the bad outputs.

The first technology imposes weak disposability on good and bad outputs. Under this assumption, the DMU may not freely dispose of its undesirable by-products (i.e., bad outputs). As a result, reducing its bad output production comes at the cost of reducing its good output production. Hence, the weak disposability technology can be viewed as the regulated technology. The second technology assumes bad outputs are freely disposable, i.e., a DMU is assumed to ignore the undesirable outputs that it produces. The free disposable technology can be viewed as the unregulated technology. [10] The difference in technical efficiency scores for the regulated and

[10] Because it is likely that DMUs are regulated prior to the year(s) included in the sample used to construct the production frontier, the unregulated frontier can be thought of as the least-regulated frontier.

unregulated technologies yields the foregone good output (i.e., the opportunity cost) associated with reducing bad output production.

These models evolved independently from the earlier activity analysis models, and closely resemble DEA models that impose weak disposability on inputs (i.e., allow for a backward bending isoquant). In fact, Färe and Grosskopf (1983) referred to the "congestion" of outputs when imposing weak disposability on the outputs.

The unregulated and regulated technologies can be specified as distance functions, which can be solved as LP problems. Because Färe, Grosskopf, and Pasurka (1983, 1986) assumed variable returns to scale (VRS), they needed to introduce an additional variable that is not present in equation (4). For DMU k′ in period t, the constant returns to scale weak disposability (i.e., regulated) technology is:[11]

$$\left(D_O^t(x^{t,k'}, y^{t,k'}, b^{t,k'})\right)^{-1} = \max \quad \beta \tag{4}$$
$$\begin{aligned} s.t. \quad & \sum_{k=1}^{K} z_k^t y_{km} = (\beta) y_{k'm}^t, && m = 1,\ldots,M \\ & \sum_{k=1}^{K} z_k^t b_{kj} = (\beta) b_{k'j}^t, && j = 1,\ldots,J \\ & \sum_{k=1}^{K} z_k^t x_{kn} \le x_{k'n}^t, && n = 1,\ldots,N \\ & z_k^t \ge 0, && k = 1,\ldots,K \end{aligned}$$

where β is the measure of technical efficiency. The LP problem calculates the maximal β by which production of all good and bad outputs can be expanded for a given technology and input vector. If outputs cannot be expanded (i.e., β=1), this indicates the observation is on the frontier (i.e., is technically efficient). If the outputs can be expanded (i.e., β>1), this indicates the observation is inside the frontier (i.e., is technically inefficient). If there is a single good output, the amount of forgone good output production due to technical inefficiency is $(\beta-1)\times y_{k'}$.

[11] While it is possible to specify an input-based measure of efficiency, output-based efficiency specifications are used throughout this paper.

The second technology assumes the good and bad outputs are freely disposable, i.e., a DMU is can costlessly dispose of its good and bad outputs. The free disposable technology or unregulated technology is modeled by converting the sign on the good output constraint and bad output constraint to "≥". For the free disposal technology, the LP program maximizes β*. If there is a single good output, the value of $(\beta^*-\beta)\times y_{k'}$ represents the total foregone good output production associated with pollution abatement by DMU k′.

The regulated technology specified in equation (4) yields the regulated (0fcd0) production frontier depicted in Figure 3. The unregulated technology, which is specified by modifying the constraint on the bad output, yields the unregulated (0AcdE0) frontier in Figure 3. The technical efficiency of observation (y, b) is calculated relative to a point on segment fc of the regulated frontier. The downward sloping line segment (cd) for both the unregulated and regulated frontiers is troublesome in that it allows a decision making unit to be on the frontier and increase good output production while simultaneously decreasing bad output production. The counter-intuitive shadow price for the bad output associated with the downward sloping portion of the frontier would remain a source of concern in the coming years.

Färe, Grosskopf, and Pasurka (1986, p. 183) viewed the opportunity cost of regulations as "indirect costs." This led to some confusion when interpreting their results (Färe, Grosskopf, and Pasurka, 1986, p. 184):

"Even though most of the current federal pollution control regulations had not become effective at that time, we found that lack of disposability 'cost' an average of roughly 16 million kilowatt-hours in lost potential output for each plant in our sample. This is in addition to any direct outlays on pollution control equipment, for example, suggesting that simple outlay measures of the costs of pollution control understate the social cost of improving environmental quality. "

This confusion would disappear in subsequent papers.

Unlike Pittman (1981, 1983), who assumed an asymmetric relationship between good and bad outputs, a drawback of the distance function model is its symmetric treatment of good and bad outputs. Färe et al. (1989) addressed this concern by specifying a hyperbolic function to model the joint production of good and bad output production. The advantage of the hyperbolic model is that it allows good and bad outputs to be treated asymmetrically. In other words, a DMU is credited for simultaneously expanding production good output and contracting of bad outputs. This allows the calculation of an adjusted measure of technical efficiency. Although the hyperbolic model calculates technical efficiency differently than the distance function, it calculates the opportunity cost of pollution abatement in the same manner as the distance function model proposed by Färe and Grosskopf (1983) and Färe, Grosskopf, and Pasurka (1986).

The hyperbolic function (see Färe et al. 1989) is calculated as a solution to nonlinear programming (NLP) problem.[12] For the regulated technology, we have for DMU k′ at t

$$
\begin{aligned}
H^{t}_{O}(x^{t,k'}, y^{t,k'}, b^{t,k'}) = \max \quad & \lambda && (5)\\
\text{s.t.} \quad & \sum_{k=1}^{K} z^{t}_{k} y_{km} \geq (\lambda) y^{t}_{k'm}, \qquad m = 1,\ldots,M \\
& \sum_{k=1}^{K} z^{t}_{k} b_{kj} = (1/\lambda) b^{t}_{k'j}, \qquad j = 1,\ldots,J \\
& \sum_{k=1}^{K} z^{t}_{k} x_{kn} \leq x^{t}_{k'n}, \qquad n = 1,\ldots,N \\
& z^{t}_{k} \geq 0, \qquad k = 1,\ldots,K
\end{aligned}
$$

(λ=1 indicates the observation is on the frontier, while λ>1 reveals the observation is inside the production frontier. As was the case for the distance function, we model the unregulated

---

[12] Boyd and McClelland (1999) show how the NLP problem can be converted into a LP problem, and argue that finding technical inefficiency when treating goods and bads asymmetrically is evidence of a potential “win-win” situation similar to the Porter hypothesis. Färe, Grosskopf, and Zaim (2002) show that a constant returns to scale hyperbolic function can be specified as a LP problem.

technology by converting the sign on the bad output constraint to "≥". For the unregulated technology, the LP program maximizes λ^*. Calculating the value of $(\lambda^* - \lambda)\times y_{k'}$ yields a measure of the total foregone good output production associated with pollution abatement by DMU k′.

The constraints on the good and bad outputs yield the regulated and unregulated production frontiers depicted in Figure 4. The change associated with the good output constraint for the regulated technology yields a new regulated frontier of 0fcdE0, while the unregulated frontier remains 0AcdE0. The technical efficiency of observation (y, b) is calculated relative to a point on segment fc of the regulated frontier. As with Figure 3, the downward sloping line segment (cd) exists in Figure 4.

Färe et al. (1989) proposed a second strategy for calculating the opportunity cost of reducing the bad output. While the regulated frontier is identical to that specified by equation (4), the constraint on the bad output in the unregulated frontier is eliminated. These constraints yield the regulated and unregulated production frontiers depicted in Figure 5. The change associated with the bad output constraint for the unregulated technology results in a new unregulated production frontier of 0AGE0, while the regulated frontier remains 0fcdE0. As with Figure 4, the downward sloping line segment (cd) remains for regulated frontier; however, the downward sloping portion of the frontier no longer exists for the unregulated technology.

While the hyperbolic function represents an important innovation, in the future it would be overshadowed by an alternative approach for modeling good and bad outputs asymmetrically – the directional distance function (see Luenberger 1995, and Chambers, Chung, and Färe 1996).

C. Adjusted Productivity Change with Good and Bad Outputs

After Färe et al. (1989) introduced the asymmetric treatment of good and bad outputs using data from a single year, the next step in the evolution of modeling good and bad output production was extending the model to allow for shifts in the production frontier. Chung, Färe, and Grosskopf (1997), which specified DEA directional distance functions to model the joint production of good and bad outputs, introduced a model that allowed for the possibility of shifts in the production frontier (i.e., technical change). The primary advantage of the directional distance function is that it possesses duality properties the hyperbolic function lacks. By treating good and bad outputs asymmetrically, these models can specify a Malmquist-Luenberger productivity index that credits a DMU for simultaneously expanding production of both its good output and contracting its production of bad outputs. They also demonstrated how the adjusted productivity change can be decomposed into technical change and changes in efficiency.

In order to calculate the adjusted productivity, four LP problems must be solved. Two LP problems are solved in which all of the observations are from the same periods. As an example we have for DMU k′ at t

$$
\begin{aligned}
&\vec{D}^{\,t}_{O}(x^{t,k'}, y^{t,k'}, b^{t,k'}; y^{t,k'}, -b^{t,k'}) = \max \quad \beta && (6)\\
&s.t. \quad \sum_{k=1}^{K} z_k^t y_{km} \geq (1+\beta) y_{k'm}^t, && m = 1,\ldots,M\\
&\sum_{k=1}^{K} z_k^t b_{kj} = (1-\beta) b_{k'j}^t, && j = 1,\ldots,J\\
&\sum_{k=1}^{K} z_k^t x_{kn} \leq x_{k'n}^t, && n = 1,\ldots,N\\
&z_k^t \geq 0, && k = 1,\ldots,K
\end{aligned}
$$

There are also two mixed period LP problems. These resemble (6), except that the time superscripts on the right-hand side of the constraints differ from the time superscripts on the left-hand side of the constraints. For example, in one of mixed-period LP problems, the output set (i.e., the production possibilities frontier) is determined by all of the observations from period t. However, the DMU under evaluation--the DMU denoted k′, on the right-hand-side of the constraints is from period t+1.

A computational problem that emerged with calculating productivity change and technical change was the imposition of weak disposability on the outputs resulted in the occurrence of infeasible LP problems for the mixed-period LP problems when employing contemporaneous frontiers. A visual representation of an infeasible LP problem can be seen in Figure 6 which represents the case where an LP problem is unable to find a solution when using the period t technology (0fcdE0) to evaluate an observation from period t+1 (point h).

The initial strategy employed to reduce the incidence of infeasible LP problems involved using multiple year "windows" of data as the reference technology (see Färe, Grosskopf, and Pasurka 2001). This strategy assumed the reference technology (i.e., the production frontier) consists of observations from a given year plus the previous two years. Using either a sequential technology or a minimum of two-year windows (i.e., the reference technology consists of observations from a given year plus the previous year) eliminates infeasible LP problems for the case consisting of the technology of period t+1 evaluating a DMU from period t. However, neither windows nor a sequential technology guarantee the elimination of infeasible LP problems for a mixed-period LP problem with the technology of period t being used to determine the efficiency of a DMU from period t+1.

In addition to the multiplicative environmental directional distance function specified in equation (6), there is also an additive environmental directional distance function (see Färe, Grosskopf, and Pasurka 2007a). This version of the directional distance function is specified as:

$$\vec{D}_o(x^{k'}, y^{k'}, b^{k'}; g_y, g_b) = \max \beta^{k'} \tag{7}$$

$$\begin{aligned} s.t. \quad & \sum_{k=1}^{K} z_k y_{km} \geq y_{k'm} + \beta^{k'} g_{y_m}, \qquad m = 1,\dots,M \\ & \sum_{k=1}^{K} z_k b_{kj} = b_{k'j} - \beta^{k'} g_{b_j}, \qquad j = 1,\dots,J \\ & \sum_{k=1}^{K} z_k x_{kn} \leq x_{k'n}, \qquad n = 1,\dots,N \\ & z_k \geq 0 \qquad k = 1,\dots,K \end{aligned}$$

When the good and bad outputs are treated asymmetrically, the direction vectors are assigned values of unity, i.e., $g_y = 1_M$, $g_b = 1_J$.

While selecting an additive or multiplicative directional distance function does not affect the shape of the frontier, the additive environmental directional distance function model has the following advantages: (1) its results are easy to aggregate and (2) it provides a clear connection to the traditional production function. However, if g_y and g_b are both non-zero then the results for the additive environmental directional distance function are affected by how the data are scaled, while results for the multiplicative models (see equation 6) are unaffected by data scaling. Therefore, caution must be exercised when interpreting the empirical results of the additive environmental directional distance function.

D. Traditional Productivity

Some of the earliest reservations expressed about environmental regulations concerned the adverse effect of pollution abatement on measures of productivity that credit a DMU solely for expanding good output production. Studies addressing this issue traditionally rely on assigned input models. In order to modify the joint production model to address this issue, Färe, Grosskopf, and Pasurka (2007a) specified the environmental production function, which they demonstrated is a special case of the environmental directional distance function. When employing an environmental production function, the DMU is credited solely for expanding good output production. This can be written as the following LP problem:

$$
\begin{aligned}
F(x^{k'}, b^{k'}) = \max \quad & \sum_{k=1}^{K} z_k y_k && && (8) \\
s.t. \quad & \sum_{k=1}^{K} z_k b_{kj} = b_{k'j}, && j = 1,\ldots,J \\
& \sum_{k=1}^{K} z_k x_{kn} \leq x_{k'n}, && n = 1,\ldots,N \\
& z_k^t \geq 0, && k = 1,\ldots,K
\end{aligned}
$$

Figure 7 shows the environmental production function. For a given technology, input vector (x^0), and level of the bad output (b^0), $F(x^0, b^0)$ is the maximum feasible good output production. This corresponds to point f on the production frontier.

Färe, Grosskopf, and Pasurka (2007b) used the environmental production function to model the joint production of good and bad outputs in which the DMU is credited solely for expanding good output production. The difference in good output production found by the regulated and unregulated production frontiers allows us to calculate the effect of pollution abatement on productivity growth. As was the case for Chung, Färe, and Grosskopf (1997) and Färe, Grosskopf, and Pasurka (2001), productivity change can be decomposed into technical change and changes in technical efficiency.

Färe, Grosskopf, and Pasurka (2007b) represent the next stage in the specification of the production technology. In this paper, infeasible LP problems are eliminated by following a two step procedure. First, they use 3-year windows of data as the reference technology.. Second, due to the inability to eliminate the occurrence of infeasible LP problems for the mixed-period LP problem with the technology of period t evaluating DMUs from period t+1, this mixed-period LP problem was not used to calculate productivity change and technical change. The drawback to this strategy is the arbitrary exclusion of period t as a reference technology.

E. Tradable Permits

In addition to using joint production models to calculate the opportunity costs and productivity consequences of pollution abatement, they can also be employed to analyze issues associated with tradable permits of bad outputs. For example, Brännlund et al. (1998) used the joint production model to forecast the potential increase in profits if a tradable permit system was implemented for the bad outputs of the Swedish pulp and paper industry. Recently, Färe, Grosskopf, and Pasurka (2012a) modified this framework to investigate unrealized gains from trade that exist in tradable permit system for SO_2 emissions of power plants in the United States for 1995 to 2005.

Färe, Grosskopf, and Pasurka (2012a) represent the most recent advance in the specification of the production technology. The LP problem specified the good and bad output constraints for the regulated technology as:

$$
\begin{aligned}
F(x^{k'}, b^{k'}) = \max \quad & \sum_{k=1}^{K} z_k y_k && && (9)\\
\text{s.t.} \quad & \sum_{k=1}^{K} z_k y_k \geq y_{k'}, && \\
& \sum_{k=1}^{K} z_k b_{kj} \leq b_{k'j}, && j = 1,\ldots,J \\
& \sum_{k=1}^{K} z_k x_{kn} \leq x_{k'n}, && n = 1,\ldots,N \\
& z_k^t \geq 0, && k = 1,\ldots,K
\end{aligned}
$$

The change in the bad output constraint yields a P(x) that is not bounded. These constraints yield the regulated production frontiers depicted in Figure 8. The change associated with the bad output constraint for the regulated technology results in a new regulated production frontier of

0fcGE0, while the unregulated frontier remains 0AGE0.[13] The technical efficiency of observation (y, b) is calculated relative to the segment fc of the regulated frontier.. With this most recent advance, the downward sloping line segment of the regulated technology is eliminated. Intuitively, the revised constraint on the bad output restricts the shadow price of the bad output for the regulated frontier to be non-positive. This improves the modeling good and bad outputs because the downward sloping portion of the frontier had been especially troublesome for interpreting the results of models that treated the good and bad output symmetrically or models that focused on expanding good outputs while holding bad outputs constant. However, the correct interpretation of the downward sloping portion of the frontier remains unresolved. The definitive explanation for the existence of a positive sloped frontier remains a mystery. For a perspective on interpreting the results generated by the downward sloping frontier, see Picazo-Tadeoa and Prior (2009).

F. Change in Pollution Abatement Cost

We mentioned earlier the concern about the association between pollution abatement costs and competitiveness. The traditional perspective emphasizes optimizing firms that confront trade-offs between producing marketed goods and environmental quality**.** While reducing emissions may improve the overall welfare of a nation, the traditional perspective argues that a unilateral increase in PAC results in some firms and industries becoming less competitive. For a polluting firm or industry that is producing on its production frontier, pollution abatement reduces both emissions and production of marketed goods as inputs are moved from production of marketed goods to pollution abatement. Hence, for a given

[13] Färe, Grosskopf, and Pasurka (2012a) do not specify an unregulated technology. Hence, the unregulated frontier in Figure 8 is taken from Figure 5.

technology, pollution abatement is associated with declining productivity and increasing opportunity costs of reducing bad outputs. If the model allows "induced innovation" from environmental regulations to occur, the change in the price (i.e., the opportunity cost) of bad outputs relative to the price of the good output spurs R&D expenditures, which develop products and processes that reduce bad outputs while maintaining or increasing good output production.

We have started to investigate the potential usefulness of joint production models to examine the role of technical change on the opportunity costs of pollution abatement. Pasurka (2001) and Färe, Grosskopf, and Pasurka (2012b) employed joint production models to address this issue. The Färe, Grosskopf, and Pasurka (2012b) framework extends Pasurka (2001) by specifying both regulated and unregulated frontiers in order to decompose changes in the opportunity cost of pollution abatement into (1) change in the scale of operation (i.e., changes in inputs), (2) technical change, and (3) changes in emission-intensity.

While modifying the specification of the regulated technology in equation 9 was not intended to eliminate infeasible mixed-period LP problems, Färe, Grosskopf, and Pasurka (2012b) found no infeasible mix-period LP problems. The decline in infeasible LP problems even occurs when employing contemporaneous frontiers for the mixed-period LP problem with the technology of period t evaluating DMUs from period t+1. If these preliminary results are found in subsequent studies, it may be possible to use both mixed-period LP problems without windows.

G. Changes in Bad Output Production

While pollution abatement reduces good output production, its primary objective is to reduce bad output production. As a result, there are continuing efforts to identify the relative importance of factors associated with changes in bad output production. One approach involves using econometric models to estimate an Environmental Kuznets Curve (EKC), which model the relationship between economic growth and bad output production. Taskin and Zaim (2000), Zaim and Taskin (2000), and Färe, Grosskopf, and Zaim (2005) introduced the use of joint production DEA models to investigate the existence of an EKC.

In addition to the EKC model, both input-output models and index number models are used to decompose changes in bad output production by identifying the factors that affect changes in bad output production. Li and Chan (1998), which is discussed in Grosskopf (2003), extended the standard DEA productivity decomposition framework by decomposing changes in good output production into changes in technical efficiency, technical change, and changes in inputs. Pasurka (2006) modified the Li and Chan framework to account for changes in bad output production associated with (1) changes in technical efficiency, (2) change in the scale of operation (i.e., changes in inputs), (3) technical change, and (4) changes in emission-intensity.

IV. Extensions of the DEA joint production model

A. Environmental Performance Indexes

An alternative measure of environmental performance, which requires only information on good and bad output production, is an Environmental Performance Index (EPI). The EPI calculates the inverse of the change in the emission-intensity ratio. Färe, Grosskopf and Hernando-Sanchez (2004), and Färe, Grosskopf, and Pasurka (2006) implemented this index for the case with a single bad output in which changes in the good-bad output ratio over time is proposed as a measure of environmental performance. In the most recent extension of the EPI, Färe, Grosskopf, and Pasurka (2010) specified an index that accommodates more than one bad output.

B. No data on PA inputs or bad outputs

In many instances both data on inputs assigned to pollution abatement and data on bad output production are unavailable. For these cases, it is necessary to know when a new technology was installed or when a regulation was implemented. If this information is available, it is possible to specify an alternative definition of the regulated and unregulated production technologies in order to calculate the effects of environmental regulations on technical efficiency in which only data on good output production are available. For example, Färe, Grosskopf, and Pasurka (1989) compared changes in technical efficiency before and after the introduction of precipitators for a sample of electric power plants in the United States.

C. Good and Bad Outputs in Network Models

Both assigned input and joint production models treat the transformation of inputs into good and bad outputs as a black box. A network technology of sub-technologies must be introduced (see Färe, Grosskopf, and Pasurka, 2012c) to allow an investigation of the consequences of ignoring the transformation process.[14] The network technology looks inside the black box, which consists of a set of sub-technologies or sub-processes. These sub-technologies are connected in a network that forms the joint production frontier. For example, the network technology for a coal-fired power plant consists of two sub-technologies. The first sub-technology produces electricity and bad outputs. The second sub-technology is an EOP abatement technology (e.g., a scrubber) in which the bad output produced by the first sub-technology is an input. The output of the second sub-technology is the transformation of some of bad output into a form that society views as less undesirable. After treatment, the remaining bad output is emitted into the environment.

[14] Hua and Bien (2008) present an alternative specification of a network model with good and bad outputs.

Processes inside the dashed box in Figure 9 are the subtechnologies inside the black box. Exogenous inputs (i.e., capital, labor, and fuel) - box A - are employed by the subtechnologies. The first subtechnology (box B) produces a good output (${}^{C}_{B}y + {}^{D}_{B}y$) and a bad output (${}^{C}_{B}b$).[15] The good output is consumed as an intermediate input (${}^{C}_{B}y$) by the EOP pollution abatement subtechnology (box C) or as a final output (${}^{D}_{B}y$). Hence, the difference between the gross production of the good output and the good output consumed by the pollution abatement subtechnology yields good output production - the net good output production, exclusive of plant use (box D).

Gross bad output production (${}^{C}_{B}b$) reflects the level of bad output production after CIP abatement activities or no treatment at all.[16] These bad outputs can be discharged from the plant or sent to the pollution abatement subtechnology for additional processing. For the pollution abatement subtechnology, exogenous inputs (${}^{C}_{A}x$), intermediate inputs (${}^{C}_{B}y$), and the gross production of the bad output (${}^{C}_{B}b$) are inputs whose output is net bad output production (${}^{E}_{C}b$), which is the bad output released by the DMU (box E).[17] While both the assigned input and joint production models view the reduced good output production associated with abatement activities as the opportunity cost of pollution abatement, there has been no attempt to establish the theoretical relationship between these models. The network model provides a theoretical framework that might permit a comparison of the assigned input and joint production models.

[15] When discussing the inputs and outputs in Figure 9, the subscript represents the source box, while the superscript represents the destination box.

[16] The U.S. Department of Commerce (1996, p. 72) found 20.4 percent of air pollution abatement capital expenditures by electric power plants in 1994 were production process enhancements.

[17] For plants that do not undertake abatement activities, gross SO_2 production equals net SO_2 production.

Recently, another variation of combining network models with bad outputs has been proposed. tenRaa (1995) extended the traditional input-output framework with no bad output production to calculate macroeconomic technical inefficiency. Böhm and Luptáčik (2006) and Luptáčik and Böhm (2010) extended tenRaa's framework to calculate technical inefficiency in the presence of a constraint on bad output production (i.e., emissions of air pollutants). In their model, inefficiency is determined by the extent to which it is possible to proportionally contact primary input (i.e., capital and labor) use while maintaining the original final demand vector or proportionally expanding the final demand vector with the original vector of primary inputs.

Pasurka (2012) extends the Prieto and Zofío (2007) model by introducing bad outputs into an input-output activity analysis model. Unlike the tenRaa which can be implemented with a single input-output table, the Prieto and Zofio model requires multiple input-output tables. Directional distance functions are then used to calculate adjusted measures of productivity change, technical change, and changes in technical efficiency by crediting an economy for simultaneously expanding good outputs and contracting bad outputs. By allowing primary factors of production (i.e., labor and capital) to be mobile among sectors, plus using the intermediate inputs from the input-output table, this approach provides a more general equilibrium framework than the standard joint production model.

V. Parametric joint production models and their application

A. Shadow Prices of Bad Outputs

While the non-parametric models specified in previous sections can calculate the total quantity of good output production foregone to reduce bad output production to a given level, the next stage in the evolution of the joint production model was calculating the shadow price (i.e., the marginal abatement cost) of reducing bad output production. Rather than calculating the distance between the unregulated and regulated frontiers, calculating the shadow price of bad outputs requires specifying a (regulated) technology that imposes weak disposability on the good and bad outputs and then determines the shadow price by calculating the slope of the production frontier. Färe et al. (1993) specified a translog distance function (a parametric model), while Ball et al. (1994) employed a piece-wise linear DEA model to calculate the shadow price of reducing production of a bad output. Both models treat good and bad outputs symmetrically. Because the piece-wise linear models yield production frontiers that are not smooth, difficulties were encountered when calculating the shadow prices of some observations. As a result, the parametric approach for calculating shadow prices emerged as the dominant framework. Färe et al. (1993) – and subsequent parametric models of goods and bad outputs – avoided the problem of a downward sloping production frontier, which yield counter-intuitive shadow prices of bad outputs, by imposing a constraint in the LP problem requiring the shadow prices of bad outputs to be non-positive.[18] Figure 10 provides a simple example of how the shadow price of a bad output is calculated. An inefficient DMU (point c) is projected to a point on the production frontier (point H). The shadow price of the bad output for the DMU is determined by the slope of the frontier at point H.

Färe, et al. (2005) extended the shadow price framework by specifying a quadratic directional distance function instead of a translog distance function. Unlike the translog distance function that requires good and bad outputs be treated symmetrically, the quadratic directional distance function also allows the good and bad outputs to be treated asymmetrically.[19] In

[18] See Hetemäki (1996, pp. 64-65) for an example of a translog distance function with no constraint imposed on the shadow price of the bad outputs.

[19] The translog function cannot be used when specifying a directional distance function.

addition, bootstrapping was employed to add a stochastic perspective to the analysis. Vardanyan and Noh (2006) investigated the effect of different directional vectors on shadow prices. Interest in the appropriate parametric functional form to employ when modeling good and bad outputs led Färe, Martins-Filho, and Vardanyan (2010) to investigate the relative strengths and weaknesses of the translog distance function and quadratic directional distance function. While their work led them to favor using a quadratic directional distance function, one shortcoming of the quadratic function is that its direction vector is additive. As a result, changes in the scaling of the data affect the parameters and results unless the direction vector is changed in a manner that offsets the change in the data.

B. Additional Applications of Parametric Joint Production Models

In addition to identifying the shadow prices of bad outputs, two applications of the parametric joint production models emerged. First, Aiken and Pasurka (2001) demonstrated an alternative strategy for calculating adjusted changes in productivity in which the value of good output production is adjusted by subtracting the "value" of bad output production. In their paper, the price of a bad output is determined by calculating its shadow price via a translog distance function (see Färe et al. 1993). With this approach, the value of good output production is reduced by the product of the shadow price of the bad output and the quantity of the bad output produced.

Second, the parametric models can be used to determine the extent of substitutability among outputs. Färe, et al. (2005) employed the quadratic directional distance function to calculate output elasticities of transformation in order to investigate the extent of substitutability between electricity (the good output) and SO_2 emissions (a bad output). Färe et al. (2012) extended Färe et al. (2005) to calculate output elasticities of transformation that determine the extent of substitutability or complementary among undesirable by-products. These elasticities have the potential to be useful when analyzing the ancillary benefits (or co-benefits) of a regulation. For example, if two bad outputs are complements then regulations intended to reduce production of one bad output will result in reduced production of its complementary bad output. If they are substitutes, then reducing production of one bad output will increase production of the other bad output. The output elasticities of transformation generated by the quadratic directional distance function are subject to the same caveats as its shadow prices that were discussed in the previous section.

VI. Cost Functions with Bad Outputs

An alternative to using hyperbolic functions, distance functions, or directional distance functions is employing cost functions to model the joint production of good and bad outputs. However, only a limited number of studies have employed cost functions when modeling good and bad outputs (see Tran and Smith 1983, McClelland and Horowitz 1999, Ball et al. 2005, Chapple, Paul and Harris 2005 and 2006, and Mosheim 2006).[20] Of these papers, only Ball, et al. (2005) explicitly employed the null-jointness and weak disposability assumptions imposed by papers employing hyperbolic functions, distance functions, and directional distance functions. Ball et al. (2005) specified a nonparametric cost function to calculate productivity change via a Malmquist productivity index that treated the good and bad outputs symmetrically. The remaining papers specified cost functions with good and bad outputs, but did not discuss imposing null-jointness and weak disposability.

[20] Gollop and Roberts (1983, 1985) used bad output production to construct measures of regulatory intensity when estimating cost functions for electric power plants. In addition, Kolstad and Turnovsky (1998) used the ash and sulfur content of fuels when estimating a cost functions for electric power plants.

VII. Challenges confronting efforts to model good and bad outputs

A. Bads – are they inputs or outputs?

While this survey has focused on the literature that treats bads as outputs, there is a view that sees bads as inputs. Cropper and Oates (1992) present a simplified theoretical model in which bads are introduced as inputs. Keilbach (1995) and Reinhard, Lovell, and Thijssen (1999) are two examples of empirical papers that modeled bads as inputs in a production function. One drawback to modeling bads as inputs when the bads are regulated is the assumption that inputs are freely disposable is violated.

B. Should we worry about a hypothetical unregulated universe?

The free disposability technology specified in joint production models is constructed from the observed behavior of DMUs. Does ignoring levels of labor, capital stock, and technical change that might have existed if regulations had not been introduced result in our opportunity cost calculations being downward biased?

C. Violation of null-jointness

While null-jointness is a reasonable assumption for modeling undesirable by-products generated by manufacturing and power plants, it may present issues with other types of economic activities. For example, the complete elimination of all undesirable by-products for a power plant may be unrealistic; however, there may instances when commercial fishing vessels have no bycatch (i.e., the undesirable by-product). Although the non-existence of any by-catch violates null-jointness, the LP problems will generate results.

D. Material Balances

The theoretical model developed by Ayres and Kneese (1969) had an empirical counterpart in a series of empirical studies on residuals generation and management from a

Resources for the Future (RFF) project. Bower (1975) provides a survey of these studies of residual management, followed by a Paul MacAvoy critique of the residuals management model. After the RFF project concluded - Russell and Vaughn (1974, 1976) appear to be its last published works – interest in material balances models waned.

If joint production models are viewed as reduced form models, then network models can be viewed as the true underlying technology. Recently, several researchers have expressed reservations about the joint production models due to their failure to account for material balances. Pethig (2006) investigated the material balances issues from a non-DEA perspective. Subsequently, Coelli, Lauwers, and van Huylenbroeck (2007), Ebert and Welsch (2007), Lauwers (2009), Førsund (2009), Van Meensel et al. (2010), Rødseth (2011), and Murty, Russell, and Levkoff (forthcoming) have undertaken theoretical and empirical investigations of the issue.[21]

Network models may address some of the concerns expressed by those championing the use of joint production models that incorporate materials balances into the specification of the model.

E. Data Availability

Perhaps the greatest challenge confronting efforts to model good and bad outputs can be summed up by the phrase -"It's the Data, Stupid." The lack of data has been - and remains - an ongoing problem confronting researchers. While substantial efforts have been made to collect data on the cost of inputs assigned to pollution abatement (see Pasurka 2008), less effort has been devoted to collecting data on the bad outputs that regulatory activity attempts to reduce. At

[21] Baumgärtner et al. (2001) discuss the relationship between joint production and thermodynamics.

the national and industry level, systems of environmental and resource accounts might prove to be a useful source of data in the coming years. The recently released World Input-Output Database (http://www.wiod.org/) represents an effort to link bad output data to input-output tables. In addition, the *Handbook of National Accounting: Integrated Environmental and Economic Accounting for Fisheries* (United Nations, 2004) represents a useful framework for organizing data related to fisheries. Unfortunately, the availability of plant-level data remains problematic.

F. Peer Acceptance

In addition to the increased use of DEA joint production models, joint production CGE models are also implicitly used to investigate the consequences of reducing greenhouse gas emissions. Despite growing interest in the topic, some members of the research community remain ambivalent. Recently, the editor of *Resource and Energy Economics* (REE), which published the second DEA article that modeled good and bad output production (see Färe, Grosskopf, and Pasurka, 1986), published the following note:

"We also want to emphasize here that REE does not normally publish the following type of papers, which are beyond the scope of REE and will be returned to authors without review:

.
.
.

- The development of purely statistical techniques or the application of (standard) statistical techniques without strong links to the theory. We consider techniques like VAR or DEA as tools to analyze well-defined economic research questions, rather than aims in themselves.

.
.
."

(http://www.journals.elsevier.com/resource-and-energy-economics/journal-news/note-from-the-editor-of-resource-and-energy-economics/):

The adventure continues.

References

Aiken, Deborah Vaughn and Carl Pasurka (2001), "Adjusting the Measurement of U.S. Manufacturing Productivity for the Generation of Air Pollutants," *Resource and Energy Economics*, 25, No. 4 (October), 329-351.

Aiken, Deborah Vaughn, Rolf Färe, Shawna Grosskopf, and Carl Pasurka (2009), "Pollution Abatement Activities and Traditional Productivity Growth in Germany, Japan, the Netherlands, and the United States," *Environmental and Resource Economics*, 44, No. 1 (September), 11-28.

Ayres, Robert and Allen Kneese (1969), "Production, Consumption, and Externalities," *American Economic Review*, 59, No. 3 (June), 282-297.

Ball, V. Eldon, C.A. Knox Lovell, Richard Nehring, and Agapi Somwaru (1994), "Incorporating Undesirable Outputs into Models of Production: An Application to U.S. Agriculture," *Cahiers d'Économie et Sociologie Rurales*, 31, 59-74.

Ball, Eldon, Rolf Färe, Shawna Grosskopf, Osman Zaim (2005), "Accounting for Externalities in the Measurement of Productivity Growth: the Malmquist Cost Productivity Measure," *Structural Change and Economic Dynamics*, 16, No. 3 (September), 374-394.

Baumgärtner, Stefan, Harald Dyckhoff, Malte Fabera, John Proops, and Johannes Schiller (2001), "The Concept of Joint Production and Ecological Economics" *Ecological Economics*, 36, No. 3 (March) 365-371.

Böhm, Bernhard and Mikulas Luptáčik (2006), "Measuring Eco-efficiency in a Leontief Input-Output Model," pp. 121-135 in: "*Multiple-Criteria Decision Making '05*", T. Trzaskalik (ed.), Publisher of the Karol Adamiecki University of Economics in Katowice, Katowice (2006).

Bower, Blair T. (1975), "Studies of Residuals Management in Industry," pp. 275-324, in: *Economic Analysis of Environmental Problems*, Edwin S. Mills (ed.), National Bureau of Economic Research, distributed by Columbia University Press. http://www.nber.org/books/mill75-1

Boyd Gale A. and McClelland John D. (1999), "The Impact of Environmental Constraints on Productivity Improvement in Integrated Paper Plants," *Journal of Environmental Economics and Management*," 38, No. 2 (September), 121-142.

Brännlund, Runar, Yangho Chung, Rolf Färe, and Shawna Grosskopf (1998), "Emissions Trading and Profitability: The Swedish Pulp and Paper Industry," *Environmental and Resource Economics*," 12, No. 3 (October), 345-356.

Chambers, Robert, Yangho Chung and Rolf. Färe (1996), "Profit, Directional Distance Functions, and Nerlovian Efficiency," *Journal of Optimization Theory and Applications*, 98, No. 2 (August), 351-364.

Chapple, Wendy, Catherine Morrison Paul, Richard Harris (2005), "Manufacturing and Corporate Environmental Responsibility: Cost Implications of Voluntary Waste Minimisation," *Structural Change and Economics Dynamics*, 16, No. 3 (September), 347-373.

Chapple, Wendy, Catherine Morrison Paul, Richard Harris (2006), "The Cost Implications of Waste Reduction: Factor Demand, Competitiveness and Policy Implications," *Journal of Productivity Analysis*, 26, No. 3 (December), 245-258.

Chung, Yangho, Rolf Färe, and Shawna Grosskopf (1997), "Productivity and Undesirable Outputs: A Directional Distance Function Approach," *Journal of Environmental Management*, 51, No. 3 (November), 229-240.

Coelli, Tim, Ludwig Lauwers, and Guido van Huylenbroeck (2007), "Environmental Efficiency Measurement and the Materials Balance Condition," *Journal of Productivity Analysis*, 28, No. 1-2 (October), 3-12.

Cropper, Maureen L. and Wallace E. Oates (1992), "Environmental Economics: A Survey," *Journal of Economic Literature*, 30, No. 2 (June), 675-740.

Ebert, U. and H. Welsch (2007), "Environmental Emissions and Production Economics: Implications of the Materials Balance," *American Journal of Agricultural Economics*, 89, No. 2 (May), 287-293.

Ethridge, Don (1973), "The Inclusion of Wastes in the Theory of the Firm," *Journal of Political Economy*, 81, No. 6 (Nov./Dec.), 1430-1441.

Färe, Rolf and Shawna Grosskopf (1983), "Measuring Output Efficiency," *European Journal of Operational Research*, 13, No. 2 (June), 173-179.

Färe, R., S. Grosskopf and F. Hernando-Sanchez (2004), "Environmental Performance: An Index Number Approach," *Resource and Energy Economics*, 26, No. 4 (December), 343-352.

Färe, R., Carlos Martins-Filho, and M. Vardanyan (2010), "On Functional Form Representation of Multi-Output Production Technologies," *Journal of Productivity Analysis,* 33, No. 2 (April), 81-96.

Färe, Rolf, Shawna Grosskopf, and Carl Pasurka (1986), "Effects on Relative Efficiency in Electric Power Generation Due to Environmental Controls," *Resources and Energy*, 8, No. 2 (June), 167-184.

Färe, Rolf, Shawna Grosskopf, and Carl Pasurka (1989), "The Effect of Environmental Regulations on the Efficiency of Electric Utilities: 1969 vs. 1975," *Applied Economics*, 21, No. 2 (February), 225-235.

Färe, Rolf, Shawna Grosskopf, and Carl Pasurka (2001), "Accounting for Air Pollution Emissions in Measures of State Manufacturing Productivity Growth," *Journal of Regional Science*, 41, No. 3 (August), 381-409.

Färe, Rolf, Shawna Grosskopf, and Carl Pasurka (2006), "Social Responsibility: U.S. Power Plants 1985-1998," *Journal of Productivity Analysis*, 26, No. 3 (December), 259-267.

Färe, Rolf, Shawna Grosskopf, and Carl Pasurka (2007a), "Environmental Production Functions and Environmental Directional Distance Functions," *Energy - The International Journal*, 32, No. 7 (July), 1055-1066.

Färe, Rolf, Shawna Grosskopf, and Carl Pasurka (2007b), "Pollution Abatement Activities and Traditional Productivity," *Ecological Economics*, 62, No. 3-4 (15 May), 673-682.

Färe, Rolf, Shawna Grosskopf, and Carl Pasurka (2010), "Toxic Releases: An Environmental Performance Index for Coal-fired Electric Power Plants," *Energy Economics*, 32, No. 1 (January), 158-165.

Färe, Rolf, Shawna Grosskopf, and Carl Pasurka (2012a), "Tradable Permits and Unrealized Gains from Trade," mimeo.

Färe, Rolf, Shawna Grosskopf, and Carl Pasurka (2012b), "Sources of Change in Pollution Abatement Costs: A Joint Production Perspective," mimeo.

Färe, Rolf, Shawna Grosskopf, and Carl Pasurka (2012c), "Joint Production of Good and Bad Outputs with a Network Application," *Encyclopedia of Energy, Natural Resource and Environmental Economics*, edited by Jason Shogren, Elsevier Academic Press.

Färe, Rolf and Shawna Grosskopf, and Osman Zaim (2002), "Hyperbolic Efficiency and Return to the Dollar," *European Journal of Operational Research*, 136, No. 3 (1 February) 671-679.

Färe, Rolf, Shawna Grosskopf, and Osman Zaim (2005), "An Environmental Kuznets Curve for the OECD Countries," in *New Directions: Efficiency and Productivity*, pp. 77-90, eds. Rolf Färe and Shawna Grosskopf, Kluwer Academic Publishers: Boston/London/Dordrecht.

Färe, Rolf, Shawna Grosskopf, C. A. Knox Lovell, and Carl Pasurka (1989), "Measuring Efficiency When Some Outputs are Undesirable: A Nonparametric Approach," *Review of Economics and Statistics*, 71, No.1 (February), 90-98.

Färe, Rolf, Shawna Grosskopf, C.A. Knox Lovell, and Suthathip Yaisawarng (1993), "Derivation of Shadow Prices for Undesirable Outputs: A Distance Function Approach," *Review of Economics and Statistics,* 75, No. 2 (May), 374-380.

Färe, R., S. Grosskopf, D. Noh and W. Weber (2005), "Characteristics of a Polluting Technology: Theory and Practice," *Journal of Econometrics*, 126, No. 2 (June), 469-492.

Färe, Rolf, Shawna Grosskopf, Carl Pasurka and William Weber (2012), "Substitutability Among Undesirable Outputs," *Applied Economics*, 44, No. 1, 39-47.

Førsund, Finn R. (2009), "Good Modelling of Bad Outputs: Pollution and Multiple-Output Production," *International Review of Environmental and Resource Economics*, 3, 1-38.

Gollop, Frank M. and Mark J. Roberts (1983), "Environmental Regulations and Productivity Growth: The Case of Fossil-fueled Electric Power Generation," *Journal of Political Economy*, 91, No. 4 (August), 654-674.

Gollop, Frank M. and Mark J. Roberts (1985), "Cost-Minimizing Regulation of Sulfur Emissions: Regional Gains in Electric Power," *Review of Economics and Statistics*, 67, No. 1, 81-90.

Grosskopf, Shawna (2003), "Some Remarks on Productivity and its Decompositions," *Journal of Productivity Analysis*, 20, 459-474.

Hetemäki, Lauri (1996), *Essays on the Impact of Pollution Control on a Firm: A Distance Function Approach*, Finnish Forest Research Institute, Research papers 609.

Hua, Zhongsheng and Yiwen Bien (2008), "Performance Measurement for Network DEA with Undesirable Factors," *International Journal of Management and Decision Making*, 9, No. 2, 141-152.

Keilback M. (1995), "Estimation of the Value of the Marginal Product of Emission in a Country Where Emissions Output is Regulated - an Empirical Study," *Environmental and Resource Economics*, 5, No. 3 (April), 305-319.

Kohn, Robert E. (1975), *Air Pollution Control: A Welfare Economic Interpretation*, Lexington Books, D.C. Heath and Company: Lexington, MA.

Kolstad, Charles D., and Michelle H. L. Turnovsky (1998), "Cost functions and nonlinear prices: estimating a technology with quality-differentiated inputs", *Review of Economics and Statistics,* 80, 444–53.

Komen, Marinus H.C. and Jack H.M. Peerlings (2001), "Endogenous Technological Switches in Dutch Dairy Farming under Environmental Restrictions," *European Review of Agricultural Economics*, 28, No. 2 (June), 117-142.

Koopmans, Tjalling J. (1951), "Analysis of Production as an Efficient Combination of Activities," in *Activity Analysis of Production and Allocation*, pp. 33-98, ed. Tjalling J.

Koopmans, Commission for Cowles Research in Economics, Monograph No. 13, John Wiley & Son, Inc.: New York.

Lauwers, Ludwig (2009), "Justifying the Incorporation of the Materials Balance Principle into Frontier-based Eco-efficiency Models," *Ecological Economics*, 68, No. 6 (15 April), 1605-1614.

Li, Sung Ko and Hing Lin Chan (1998), "Decomposing Output Growth in the Presence of Multiple Outputs," presented at the University of Georgia Productivity Workshop, mimeo.

Luenberger, David G. (1995), *Microeconomic Theory*, McGraw-Hill, Inc.: New York.

Lee, H. and D. Roland-Holst (1997), "The Environment and Welfare Implications of Trade and Tax Policy," *Journal of Development Economics*, 52, No. 1 (February), 65-82.

Leontief, Wassily (1970), "Environmental Repercussions and the Economic Structure: An Input-Output Approach," *Review of Economics and Statistics*, 52, No. 3 (August), 262-271.

Lowe, Peter D. (1979), "Pricing Problems in an Input-Output Approach to Environmental Protection," *Review of Economics and Statistics*, 61, No. 1 (February), 110-117.

Luptáčik, Mikuláš and Bernhard Böhm (2010), "Efficiency Analysis of a Multisectoral Economic System," *Central European Journal of Operations Research*, 18, No. 4 (December), 609-619.

McClelland, John D. and John K. Horowitz (1999), "The Costs of Water Pollution Regulation in the Pulp and Paper Industry," *Land Economics*, 75, No. 2 (May), 220-232.

Mosheim, Roberto (2006), "A Shadow Cost Function Model of the US Water Industry Incorporating Water Quality and Ownership Effects" in *Performance Measurement and Regulation of Network Utilities*, Chapter 9, pp. 243-262, edited by Tim Coelli and Denis Lawrence, Edward Elgar Publishing Inc.: Northhampton, MA.

Murty, Sushama, Robert R. Russell, and Steven B. Levkoff (forthcoming), "On Modeling Pollution-Generating Technologies," *Journal of Environmental Economics and Management*.

Pasurka, Carl (2001), "Technical Change and Measuring Pollution Abatement Costs: An Activity Analysis Framework," *Environmental and Resource Economics*, 18, No. 1 (January), 61-85.

Pasurka, Carl (2006), "Decomposing Coal-fired Electric Power Plant Emissions within a Joint Production Framework," *Energy Economics*, 28, No. 1 (January), 26-43.

Pasurka, Carl (2008), "Pollution Abatement and Competitiveness: A Review of Data and Analyses" *Review of Environmental Economics and Policy*, 2, No. 2 (Summer), 194-218.

[reprinted in *Economics of the Environment*, 6th edition (2012), Robert Stavins, ed., published by W. W. Norton & Company of New York]

Pasurka, Carl (2012), "Productivity Change in a Network Input-Output Model with Bad Outputs," presented at North American Productivity Workshop, Rice University, Houston, TX, June.

Pethig, Rutiger (2006), "Non-linear Production, Abatement, Pollution and Materials Balance Reconsidered," *Journal of Environmental Economics and Management*, 51, No. 2 (March), 185-204.

Picazo-Tadeoa, Andrés J. and Diego Prior (2009), "Environmental Externalities and Efficiency Measurement," *Journal of Environmental Management*, 90, No. 11 (August), 3332-3339.

Pittman, Russell W. (1981), "Issues in Pollution Control: Interplant Cost Differences and Economies of Scale", *Land Economics*, 57, No. 1 (February), 1-17.

Pittman, Russell W. (1983), "Multilateral Productivity Comparisons with Undesirable Outputs," *Economic Journal*, 93, No. 372 (December), 883-891.

Prieto, Angel M. and José L. Zofío (2007), "Network DEA Efficiency in Input–Output Models: With an Application to OECD countries," *European Journal of Operational Research*," 178, 292-304.

Reinhard, S. C.A.K. Lovell, and G. Thijssen (1999), "Econometric Estimation of Technical and Environmental Efficiency: An Application to Dutch Dairy Farms," American Journal of Agricultural Economics, 81, No. 1 (February), 44-60.

Rødseth, Kenneth L. (2011), *Treatment of Undesirable Outputs in Production Analysis: Desirable Modelling Strategies and Applications*, Unpublished Ph.D. dissertation, Department of Economics and Resource Management, Norwegian University of Life Sciences.

Russell, Clifford S. and William J. Vaughn (1974), "A Linear Programming Model of Residuals Management for Integrated Iron and Steel Production," *Journal of Environmental Economics and Management*, 1, No. 1 (May) 17-42.

Russell, Clifford S. and William J. Vaughn (1976), *Steel Production: Processes, Products, and Residuals*, published for Resources for the Future by The Johns Hopkins University Press: Baltimore.

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

Shortle, James and Keith Willett (1987), “A Computable General Equilibrium Model with Markets for Transferable Discharge Permits,” *Managerial and Decision Economics*, 8, No. 4 (December), 263-270.

Smith, V. Kerry and J. Andrès Espinosa (1996), “Environmental and Trade Policies: Some Methodological Lessons,” *Environment and Development Economics*, 1, No. 1 (February), 19-40.

Stone, Richard (1972), “The Evaluation of Pollution: Balancing Gains and Losses,” *Minerva*, 10, No. 3, 412-425.

Taskin, Fatma and Osman Zaim (2000) “Searching for a Kuznets Curve in Environmental Efficiency using Kernel Estimation,” *Economic Letters*, 68, No. 2 (August), 217–223.

ten Raa, Thijs (1995), *Linear Analysis of Competitive Economies*, LSE Handbooks in Economics, Prentice Hall-Harvester Wheatsheaf, Hemel Hempstead.

Tran, Ngoc-Bich and V. Kerry Smith (1983), “The Role of Air and Water Residuals for Steam Electric Power Generation,” *Journal of Environmental Economics and Management*, 10, No. 1 (March), 18-34.

Tyteca, Daniel (1996), “On the Measurement of the Environmental Performance of Firms – A Literature Review and a Productive Efficiency Perspective,” *Journal of Environmental Management*, 46, No. 3 (March), 281-308.

Tyteca, Daniel (1997), “Linear Programming Models for the Measurement of Environmental Performance of Firms - Concepts and Empirical Analysis,” *Journal of Productivity Analysis*, 8, No. 2 (May), 183–197.

United Nations (2004), *Handbook of National Accounting: Integrated Environmental and Economic Accounting for Fisheries*, Studies in Methods, Handbook of National Accounting, Series F, No.97 (ST/ESA/STAT/SER.F/97).
http://unstats.un.org/unsd/envaccounting/Fish_final_whitecover.pdf

U.S. Department of Commerce, Bureau of the Census (1996), *Pollution Abatement Costs and Expenditures: 1994*, Current Industrial Reports, MA200, Washington, D.C.: U.S. Government Printing Office. (http://www.census.gov/prod/2/manmin/ma200x94.pdf).

Van Meensel J., Ludwig Lauwers, G. Van Huylenbroeck, and S. Van Passel (2010), “Comparing Frontier Methods for Economic–environmental Trade-off Analysis,” *European Journal of Operational Research*, 207, No. 2, 1027-1040.

Vardanyan, Michael and Dong-Woon Noh (2006), “Approximating Pollution Abatement Costs via Alternative Specifications of a Multi-Output Production Technology: A Case of the US Electric Utility Industry,” *Journal of Environmental Management*, 80, No. 2 (July), 177-190.

Willett, Keith (1985), "Environmental Quality Standards: A General Equilibrium Analysis," *Managerial and Decision Economics*, 6, No. 1 (March), 41-49.

Zaim, Osman and Fatma Taskin (2000) "A Kuznets Curve in Environmental Efficiency: An Application on OECD countries," *Environmental and Resource Economics*, 17, No. 1 (September), 21–36.

Zhou, P., B.W. Ang, and K.L. Poh (2008a), "A Survey of Data Envelopment Analysis in Energy and Environmental Studies," *European Journal of Operational Research*, 189, 1 (16 August), 1-18.

Zhou, P., B.W. Ang, and K.L. Poh (2008b), "Measuring Environmental Performance under Different Environmental DEA Technologies," *Energy Economics*, 30, No.1 (January), 1-14.

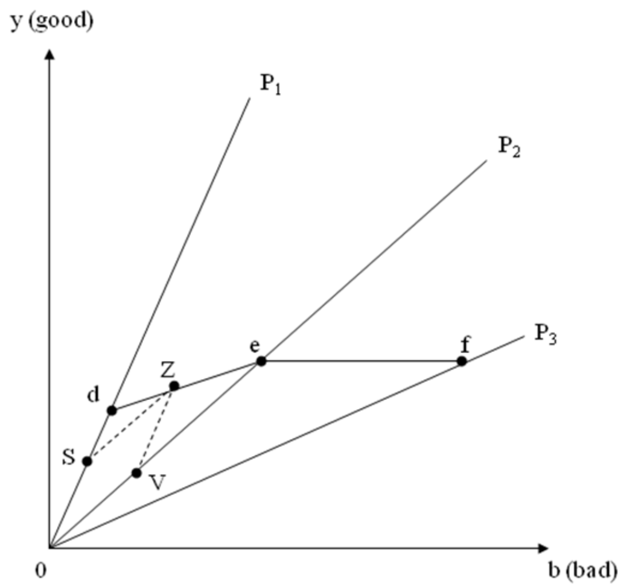


Figure 1. Kohn (1975)

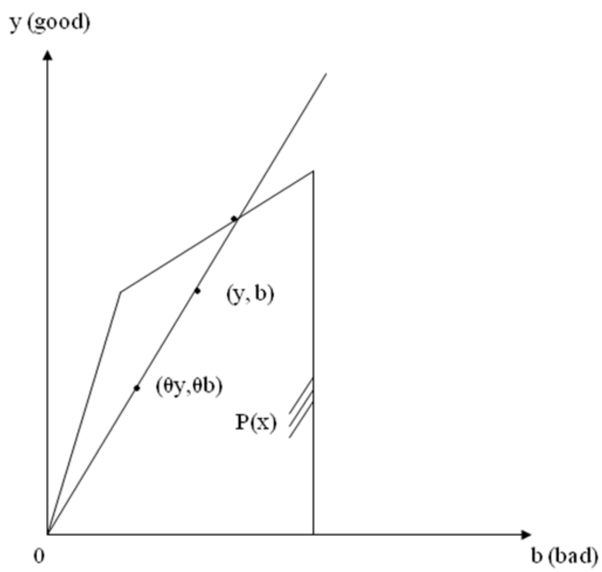


Figure 2. The Environmental Technology

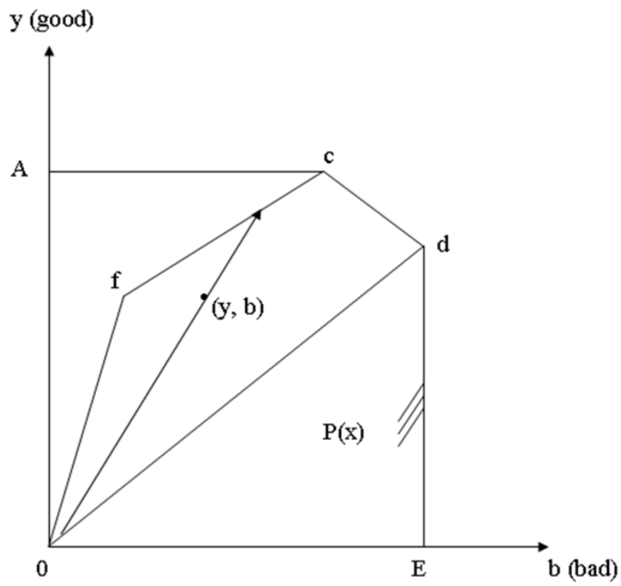


Figure 3. Färe, et al. (1986) frontier

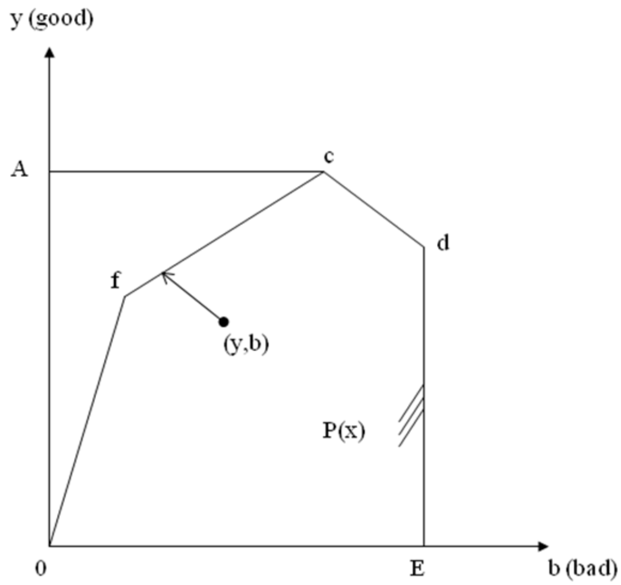


Figure 4. Färe, et al. (1989) frontier

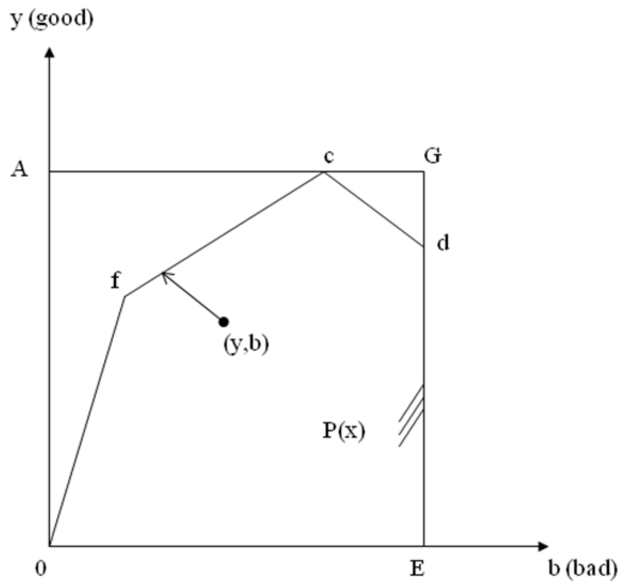


Figure 5. Färe, et al. (1989) frontier

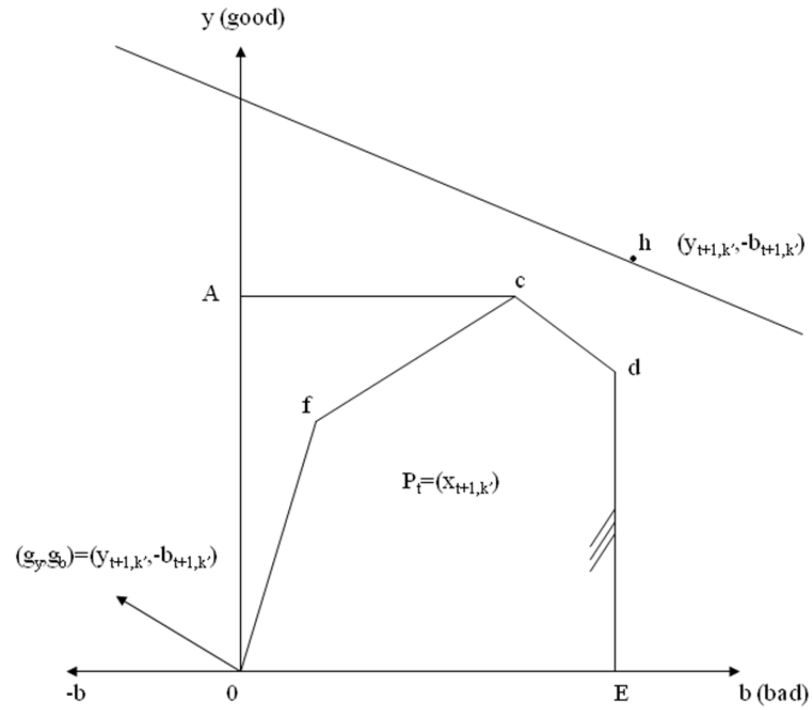


Figure 6. No Solution Example

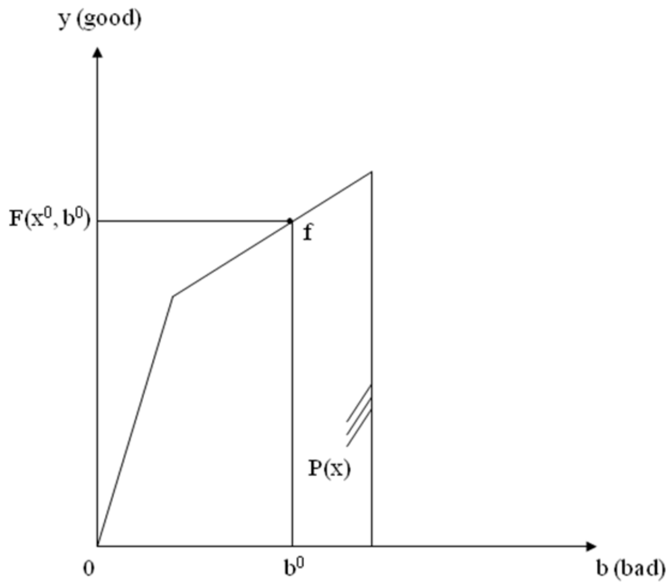


Figure 7. The Environmental Production Function

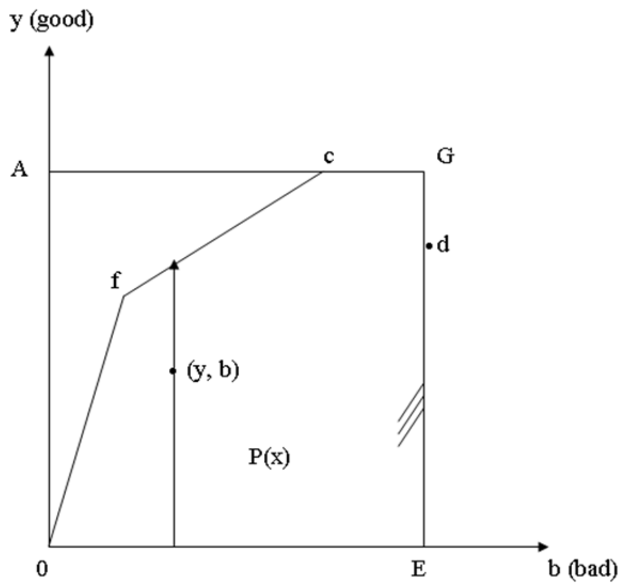


Figure 8. Färe, Grosskopf, Pasurka (2012a) frontier

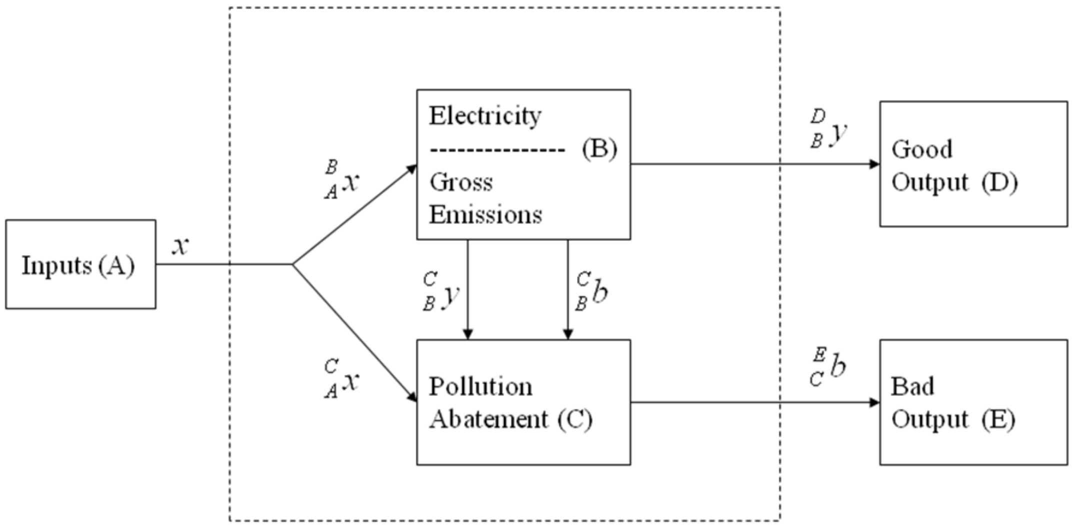


Figure 9. Network Technology of Subtechnologies

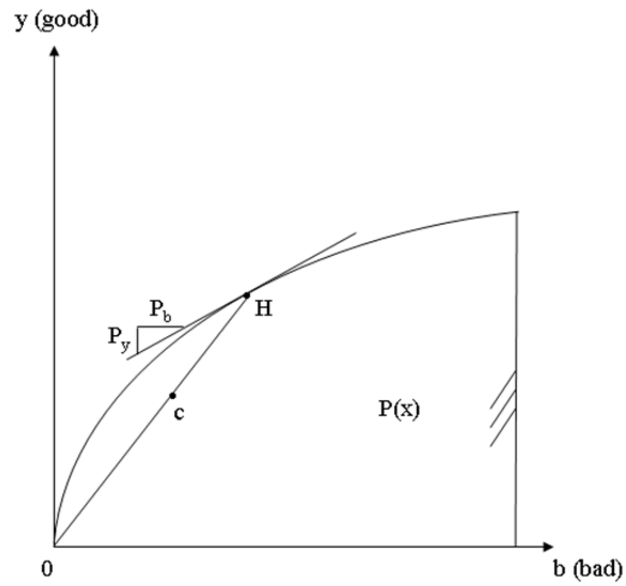


Figure 10. Parametric Joint Production Technology

SWFSC Technical Memorandums

SWFSC Technical Memorandums are available online at the SWFSC web site (http://swfsc.noaa.gov). Copies are also available from the National Technical Information Service, 5285 Port Royal Road, Springfield, VA 22161 (http://www.ntis.gov).