Productivity Measurement with Data Envelopment Analysis and Stochastic Frontier Analysis: A Review Article on *Measurement of Productivity and Efficiency: Theory and Practice*

Bert M. Balk[1]
Rotterdam School of Management, Erasmus University

The backside blurbs on this voluminous volume entitled *Measurement of Productivity and Efficiency: Theory and Practice* by Robin C. Sickles and Valentin Zelenyuk (600 pages) and published by Cambridge University Press in 2019 are intimidating to any potential reviewer: "the most comprehensive book on production theory and the measurement of productivity", "an amazingly comprehensive survey of the field", "a complete and thorough introduction", "a comprehensive survey and creative extension", "this monumental book on theory and practice". Maybe this is why it took this journal so long to find a suitable reviewer. And even the present one is not an expert on all the topics covered in this book, if this were possible at all. Thus, to start with, I sincerely apologize for any bias, mistake, or oversight.

Productivity and efficiency are distinct, but related concepts. Productivity has to do with the relation between the quantities of inputs and outputs of an economic agent, the change of this relation over time, or the differences between this relation for comparable economic agents. A typical productivity measure is output quantity divided by input quantity. This sounds simple, but is not. "The" productivity, of an enterprise, an industry, or an economy, does not exist. There is a myriad of questions to answer before computation can start. What is the output concept: revenue (gross output) or (real) value added? Do we take into account all the inputs, resulting in so-called total factor productivity (TFP), or only, say, labor inputs, resulting in labour productivity? Then, except in the case of a single input and a single output, quantities must be aggregated; hence, one needs weights, probably prices, but from where?

1 Bert M. Balk is Professor Emeritus Economic Measurement and Economic-Statistical Research at the Rotterdam School of Management of Erasmus University. Email: bert.m.balk@outlook.com.

There is the choice of the proper accounting period. How does one account for underutilization of inputs or overproduction of outputs, peak- and off-peak production, seasonality? *Etcetera*.

Efficiency has to do with the relation between quantities of inputs and/or outputs of an agent relative to some benchmark. The benchmark could be another agent or some mean of agents. For such a comparison to make sense the agents of course must be comparable. For example, the efficiency of an enterprise could improve by producing the same output quantity with less inputs, or by producing more outputs from the same input quantity. Improving efficiency means improving productivity- *ceteris paribus*, since productivity can also improve without efficiency improvement, for instance by raising the scale of the operations of the enterprise. Thus productivity has more components than efficiency alone.

In the usual, neo-classical theories firms, or more aggregate economic entities, are assumed to act efficiently. In essence this implies that the determining component of a firm's productivity turns out to be the technology, understood in the broadest sense of the word, by which the operations of this and similar firms are governed. Productivity change then becomes (almost) identical to technological change, and the interest becomes focused on the drivers behind. In the book under review the central role is given back to efficiency and the many ways of measuring this in a given data situation. In the actual world *in*efficiency is everywhere!

The stage of the play is occupied by a certain number of so-called decision-making units (DMUs), which can be plants, establishments, industries, groups of industries, or economies, with or without a time dimension. The central place is given to the concept of technology, defined as the (temporal) set of all the feasible combinations of input quantities (contained in an N-dimensional vector x) and output quantities (contained in an M-dimensional vector y). All the DMUs are supposed to have access to this technology, which thus serves as the benchmark relative to which efficiency is defined.

Chapters 1 and 2 discuss all the basic theoretical concepts, assumptions, and properties: technology set, output set, input set, distance functions, returns to scale, scale elasticity. Then cost function, revenue function, and profit function.

Chapter 3 turns to the precise definition of efficiency. Several alternatives have been developed in the course of time. They all boil down to measuring the distance between the actual operations of a DMU, summarized in its (x,y) combination, and the frontier of the current technology set, which determines the best input-output combinations. As only quantities are involved, the inverse of this distance is called technical efficiency. The farther away from the frontier, the less efficient the DMU is. If instead of the (primal) technology frontier one of the dual representations (such as cost, revenue, or profit function) is chosen there results economic efficiency, as prices are involved. The gap between economic efficiency and technical efficiency, measured in ratio or difference form depending on the particular functional form of the distance measures involved, is called allocative efficiency.

Chapter 4 formally defines productivity change. Also here is some choice. Productivity change can be measured by means of the Solow residual, or by means of a member of the Moorsteen-Bjurek, Caves-Christensen-Diewert, or Malmquist-Luenburger families of indices. Their decompositions are discussed, as well as their properties. Noteworthy is that all these indices are intransitive; that is, productivity change from period 0 to period 1, times productivity change from period 1 to period 2, is generally unequal to productivity change from period 0 to period 2. The underlying reason for this is not difficult to understand. Any satisfactory measure of productivity change requires input and/or output prices for the aggregation of input and/or output quantities. Prices adequate for comparing period 1 to period 0 are usually different from prices adequate for comparing period 2 to period 1, and both sets of prices are usually different from prices adequate for comparing period 2 to period 0. All this can of course be formalized in neat mathematical impossibility theorems.

Usually practitioners are not only interested in the performance of this or that individual agent, but also in weighted or unweighted means of groups of agents. The topic of aggregation thus demands attention. Chapter 5 treats aggregate efficiency, scale elasticity, and (dual) productivity index. The basic assumption here is that all the DMUs in each time period face the same prices.

Theory is beautiful, and one can spend a lifetime in devising subtle generalizations, but for actual application a bit more detail is necessary. In particular, one needs a specification of the technologies involved. Chapter 6 reviews a large number of alternatives: functional forms for the production function (in case of single output, M=1), the output distance function, the cost function, the revenue function, and the profit function. Special attention is paid to so-called flexible functional forms; that is, functions that can provide second-order approximations to unknown target functions at specific points.

Still, this leaves us with the problem that parameters of the particular functional form one has selected must be known. Fortunately, for some functional forms and under the assumption of optimal behavior of the agent – thus, under full economic efficiency – productivity indices reduce to statistical indices, computable from observable prices and quantities. The Fisher index is a great example. This then leads us in the area of index number theory, a topic discussed in Chapter 7.

An alternative to maintaining all those assumptions is to estimate the unknown technologies directly from sample data. There are several approaches possible, one of them being Data Envelopment Analysis (DEA). The basic idea of DEA is relatively simple to understand. Given a sample of data points (x,y), it is natural to assume that each data point is a member of the current technology, defined as the set of all the feasible input-output combinations. For any data point (x,y) it is next assumed that all the points (x',y') with x' larger than or equal to x and y' smaller than or equal to y also belong to the technology set. The union of all those south-eastern data sets then constitutes an inner approximation of the unknown technology set, and connecting the outermost data points provides an envelopment of the sample data.

Under the acronym DEA a myriad of more or less sophisticated models has been developed. All these models are implemented by means of linear programming techniques. The most important, imposing constant or variable returns to scale, are explained and discussed in Chapter 8.

As any envelopment model provides an estimator of the true, but unknown, technology, there are statistical issues to discuss. This is done at length in Chapters 9 and 10, first focusing on individual efficiency scores, and next on aggregates and distributions. There is also an introduction to two-stage DEA, where DEA-based efficiency scores of individual agents are regressed on potentially explanatory variables.

Chapters 11-14 cover stochastic frontier analysis (SFA). The basic, canonical model states that (single) output of a DMU is equal to the outcome of a production function (of input variables) minus a (positive) stochastic term representing technical inefficiency plus a stochastic term representing random disturbance. It is clear that there are some choices to make here before estimation can take place: specification of the production function, and specification of the distributions of inefficiency and disturbance. Since its inception, in the 1970s, this idea has generated a tremendous literature, hard to keep up with by a single person. Thus any survey is probably a bit biased by the individual research interests of its author.

After a detailed discussion of the basic model in Chapter 11, the developments in the field of SFA are surveyed in the next chapters. Chapter 12 covers endogenous growth models, time-varying inefficiency, panel models, the incorporation of environmental factors, and how to distinguish between wanted and unwanted outputs. Chapter 13 continues the increasing sophistication with a discussion of latent class models, and the incorporation of spatial effects. This chapter concludes that "it is not at all clear … which method is the best and thus which method(s) should be viewed as the gold standard for panel efficiency analyses" (page 448). Chapter 14 discusses the endogeneity problem (caused by correlation between explanatory variables and stochastic factors) in the estimation of production functions and SFA models.

Chapter 15 introduces dynamics in DEA-based efficiency measurement and in SFA-based panel models. Think of convergence over time of the productivity of economies; or distinguishing between short-run and long-run inefficiency.

Chapter 16 is the final theoretical chapter. A number of remaining topics are discussed: How to impose theoretically required regularity conditions on estimated production, cost, or other functions? Can semi- or nonparametric methods help us? How to obtain a consensus model out of a number of competitors? Model averaging sounds great, but what weights to use?

Chapter 17 starts with a brief (two pages only) section on data measurement issues and then turns to papers from the 2016 World KLEMS Conference, as published in the Fall 2017 issue of this journal, the *International Productivity Monitor*. All these papers were circling around productivity, the topic of the book. Thus, one would expect a certain degree of cross-referencing. However, the conclusion that "there is not one reference given in this collection of papers to any studies in the literature on measuring productivity and efficiency by other leading productivity scholars whose contributions have been discussed at length in this book." sounds like a cry from over an abyss, from one group of scholars to another. By and large this corresponds to my own impression that researchers brought up in the neo-classical tradition of thinking about productivity usually do not even read stuff from researchers who dare to dispute their cherished articles of faith. Neither do they engage into a public discussion.

This is not just a matter of using different techniques by protagonists of the two camps. It has more the traits of an ideological difference. The typical neo-classical economist starts with a bunch of assumptions, *e.g.* that agents (enterprises, industries, economies) display optimizing behavior of some kind (such as revenue maximizing, cost minimizing, or profit maximizing). Protagonists of the other camp avoid assumptions as much as possible, behavioral assumptions above all, and try to let the data speak. Kwak (2018:146) relates the majority view to academic upbringing:

> "Economics 101, ..., assumes that firms always rationally maximize their profits. ... But anyone who has ever worked at a large company – or read *Dilbert*, for that matter – knows that corporations are not ruthlessly efficient profit maximization machines, but collections of fallible and often self-interested human beings."

An interesting thought experiment would be to give protagonists of the two camps access to the same data set. It is highly likely that at the end of the day they will be looking at the same table containing, say, growth percentages derived from annual Fisher-type TFP index numbers. The neo-classical economist considers these as measures of technological change, and is surprised about the strikingly irregular features of the time-series. The other economist just takes these outcomes at face value, as measures of real profit or profitability change, and starts looking for explanatory components: technological change, efficiency change, scale change, *etcetera*

Back to the book. Chapter 17 closes with a list of publicly available data sets and a list of software. Such an undertaking is always a bit tricky, as the material is never exhaustive – for example, Statistics Netherlands is missing in the list of data sets, or can easily exhibit bias – for example, by taking on board only software developed by or related to the authors.

This concludes my review of the contents of this book. There is much to praise in this book. The explanations of models and estimation procedures are generally very clear. Each chapter contains a number of exercises, which makes the book very suitable as material for (advanced) academic or professional courses. But also researchers will want to have this book on their desk for reference purposes. It provides a synthesis of and access to a large literature.

A negative point is that, except for the two pages mentioned above, the book does not contain a serious discussion of measurement issues; that is, all the issues that must be solved before some formal model can be applied. All the models are based on input quantities x and gross output quantities y (if $M > 1$) or real value added y (if M=1), as if those data are simply observable. Except in the most simple cases, however, quantities are not observable due to the sheer number of commodities involved. At the lowest level of aggregation one thus has to work with sums of quantities and nominal values. At higher aggregation levels quantities, as required by theory, are usually substituted by real values, that is, nominal values deflated by some more or less suitable price index; which means that there is some normalization and functional choice involved. Finally, real value added is not observable at all, but a construct, obtained by deflating revenue, deflating intermediate input cost, and subtracting the outcomes.

Especially the second part of the book shows the progress made over the last twenty years since the appearance of the standard SFA book by Kumbhakar and Lovell (2000). The increased sophistication of the SFA models and the accompanying estimation techniques is remarkable.

For the first part, especially Chapters 1-7, I am not so certain that there has been much progress. A global comparison with Balk (1998) reveals that theoretical differences are almost negligible. Notable is, however, that a number of topics are missing in the Sickles-Zelenyuk book, the most important being the use of so-called indirect models, where input cost is related to a target revenue instead of output quantities, or where revenue is related to a given input budget rather than input quantities. A new topic is the aggregation of DMUs in Chapter 5. Unfortunately, the treatment is restricted to DMUs operating without interaction or reallocation of resources between them. Extensions of the theory, as referenced in the concluding remarks of the chapter, only go to 2014.

A feature of the whole book is that all the models follow the (x,y)-format; that is, on the assumption that x covers all the inputs and y all the outputs of a DMU, only the measurement of *total* factor productivity is treated. There is hardly any explicit attention for labour productivity, other partial productivity concepts, or the relations between them. Also the concept of Consistency-in-Aggregation, important when working with index numbers and actual data, is not touched.

The authors must be praised for providing throughout the book so many references to the literature. When it comes to historically interesting sources many articles are qualified as "seminal". Based on having four or more lines in the Author Index, the giants in the field of productivity and efficiency measurement appear to be Erwin Diewert, Rolf Färe, Alois Kneip, Peter Schmidt, Robin Sickles, Leopold Simar, Paul Wilson, and Valentin Zelenyuk. I wonder whether in circles of neo-classical economists sufficient attention – more than lip-service – has been paid to their work. It is thus appropriate that for all this work a monument has now been erected.

References

Balk, B. M (1998) *Industrial Price, Quantity, and Productivity Indices, The Micro-Economic Theory and An Application,* Kluwer Academic Publishers.

Kumbhakar, S. C. and C. A. K. Lovell. (2000) *Stochastic Frontier Analysis,* Cambrdge University Press.

Kwak, J. (2018) *Economism: Bad Economics and the Rise of Inequality,* Vintage Books.