Economics Letters 156 (2017) 18–21

Contents lists available at ScienceDirect

Economics Letters

journal homepage: www.elsevier.com/locate/ecolet

New insights into the stochastic ray production frontier

Arne Henningsen [a,*], Matěj Bělín [b], Géraldine Henningsen [c]

[a] Department of Food and Resource Economics, University of Copenhagen, Rolighedsvej 25, 1958 Frederiksberg C, Denmark
[b] Center for Economic Research and Graduate Education, Economics Institute, Politických vězňů 7, 111 21 Prague, Czech Republic
[c] Department of Management Engineering, Technical University of Denmark, Produktionstorvet Building 426, 2800 Kgs. Lyngby, Denmark

HIGHLIGHTS

- Suggestions for users of the Stochastic Ray Production Frontier.
- Calculate angles of polar coordinates by a non-recursive approach.
- Avoid taking logarithms of angles of polar coordinates.
- Address non-invariance to units of measurement of outputs.
- Address non-invariance to ordering of outputs.

ARTICLE INFO

Article history:
Received 6 January 2017
Received in revised form 1 April 2017
Accepted 6 April 2017
Available online 8 April 2017

JEL classification:
C13
C51
D22

Keywords:
Stochastic ray production frontier
Distance function
Multiple outputs
Primal approach
Zero output quantities

ABSTRACT

The stochastic ray production frontier was developed as an alternative to the traditional output distance function to model production processes with multiple inputs and multiple outputs. Its main advantage over the traditional approach is that it can be used when some output quantities of some observations are zero. In this paper, we briefly discuss – and partly refute – a few existing criticisms of the stochastic ray production frontier. Furthermore, we discuss some shortcomings of the stochastic ray production frontier that have not yet been addressed in the literature and that we consider more important than the existing criticisms: taking logarithms of the polar coordinate angles, non-invariance to units of measurement, and ordering of the outputs. We also give some practical advice on how to address the newly raised issues.

1. Introduction

In empirical analyses of production technologies with multiple inputs and multiple outputs, the most frequently used approach is probably the dual approach using the cost function (Kumbhakar and Lovell, 2000). However, if information on costs and prices is unavailable or prices do not vary between observations (Quiggin and Bui-Lan, 1984), the only remaining option is a primal approach using a distance function. Most parametric analyses with input or output distance functions use specifications that use logarithms of input and output quantities or the ratios between these quantities as explanatory variables, e.g., the Translog functional form (see, e.g., Kumbhakar and Lovell, 2000, equation 5.3.9). These "traditional" specifications require all input and output quantities to be strictly positive for every observation in the data set. However, in empirical applications, it is frequently observed that some firms do not produce all considered outputs, i.e., for some observations some output quantities are zero so that the traditional specifications cannot be used or have to be adjusted by questionable *ad hoc* modifications.

Löthgren (1997) suggested another primal approach, the stochastic ray production frontier (SRPF), which can be seen as a specific non-standard representation of an output distance function (Henningsen et al., 2015). Contrary to the traditional specifications, the specification suggested by Löthgren can handle zero

* Corresponding author.
E-mail addresses: arne@ifro.ku.dk (A. Henningsen), matej.belin@cerge-ei.cz (M. Bělín), gehe@dtu.dk (G. Henningsen).

http://dx.doi.org/10.1016/j.econlet.2017.04.006

values in the output quantities, which makes it a popular alternative to the traditional specifications.[1]

Nevertheless, the SRPF is not without its shortcomings. Although the SRPF has been thoroughly investigated (e.g., Whiteman, 1999; Fousekis, 2002; Zhang and Garvey, 2008; Henningsen et al., 2015), we point out further shortcomings that – to the best of our knowledge – have not yet been addressed in the literature and that we consider more important than the existing criticisms. First, we suggest using a non-recursive approach to calculate the angles of the polar coordinates because it reduces rounding errors. Second, we argue that taking logarithms of the angles of the polar coordinates should be avoided because it means that econometric assumptions are more likely to be fulfilled, makes the SRPF less dependent on the ordering of the outputs, and makes it capable of coping with zero values in all output quantities. Third, we point out that the SRPF is not invariant to the units of measurement of the outputs and we give several suggestions as to how to overcome this drawback. Fourth, we point out that the SRPF may not be invariant to the ordering of the outputs and we give suggestions about how to address this weakness. As many of these issues may have severe impacts on the quality of the results, we hope that the suggestions that we make to overcome these problems will be helpful for users of the SRPF.

The article is structured as follows: section two briefly explains the SRPF specification; and section three thoroughly discusses previously raised and new issues regarding the SRPF and gives suggestions as to how to handle them.

2. Specification of the stochastic ray production function

The SRPF can be seen as a specific functional form of a standard Shephard output distance function (Shephard, 1970):

$$D^o(x, y) = \min\{\lambda > 0 \,|\, (y/\lambda, x) \in T\}, \tag{1}$$

where $x = (x_1, x_2, \ldots, x_N)^\top$ is a vector of N input quantities, $y = (y_1, y_2, \ldots, y_M)^\top$ is a vector of M output quantities, and T is the technology set of all feasible input–output combinations. The basic idea of the SRPF is to express the vector of output quantities y through its magnitude $\|y\|$ and its direction $p(\vartheta)$ so that $y = \|y\| \cdot p(\vartheta)$. While the magnitude is expressed as Euclidean distance $\|y\| = \sqrt{\sum_{m=1}^{M} y_m^2}$, the direction is expressed through a vector of directional measures $p(\vartheta) = y/\|y\|$, which depend on polar-coordinates $\vartheta = (\vartheta_1, \ldots, \vartheta_{M-1})$ that are recursively defined by:

$$\vartheta_m(y) = \arccos\left(y_m \Big/ \left[\|y\| \prod_{j=0}^{m-1} \sin\vartheta_j \right] \right) \quad \forall\; m = 1, \ldots, M, \tag{2}$$

with arccos(.) denoting inverse of the cosine function ("arccosine") and $\sin(\vartheta_0) = \cos(\vartheta_M) = 1$.

Replacing y by $\|y\| \cdot p(\vartheta)$ in the Shephard output distance function (1), using the linear homogeneity property of this function, and re-arranging, we get:

$$D^o(x, y) = D^o(x, \|y\| \cdot p(\vartheta)) = \|y\| \cdot D^o(x, p(\vartheta)) \tag{3}$$

$$\|y\| = D^o(x, y)/D^o(x, p(\vartheta)). \tag{4}$$

By taking logarithms of both sides of (4), defining the inefficiency term $u = -\ln D^o(x, y)$, defining the SRPF as $f(x, \vartheta(y)) = -\ln D^o(x, p(\vartheta))$, and adding a noise term v, we get:

$$\ln(\|y\|) = f(x, \vartheta(y)) - u + v. \tag{5}$$

Löthgren (1997) suggests the following Translog functional form:

$$\begin{aligned}\ln(\|y\|) = {} & \alpha_0 + \sum_{m=1}^{M-1} \alpha_m \ln(\vartheta_m) \\ & + \frac{1}{2} \sum_{m=1}^{M-1} \sum_{j=1}^{M-1} \alpha_{mj} \ln(\vartheta_m) \ln(\vartheta_j) + \sum_{n=1}^{N} \beta_n \ln(x_n) \\ & + \frac{1}{2} \sum_{n=1}^{N} \sum_{l=1}^{N} \beta_{nl} \ln(x_n) \ln(x_l) \\ & + \sum_{m=1}^{M-1} \sum_{n=1}^{N} \gamma_{mn} \ln(\vartheta_m) \ln(x_n) - u + v,\end{aligned} \tag{6}$$

where α_0; $\alpha_m, m = 1, \ldots, M-1$; $\beta_n, n = 1, \ldots, N$; $\alpha_{mj}, m, j = 1, \ldots, M-1$ with $\alpha_{mj} = \alpha_{jm}\ \forall\ m, j$; $\beta_{nl}, n, l = 1, \ldots, N$ with $\beta_{nl} = \beta_{ln}\ \forall\ n, l$; and $\gamma_{mn}, m = 1, \ldots, M-1, n = 1, \ldots, N$ are parameters to be estimated.

While this specification was used in several empirical applications (e.g., Löthgren, 2000; Niquidet and Nelson, 2010; Bhattacharyya and Pal, 2013), others (e.g., Managi et al., 2006; Henningsen et al., 2015) did not take the logarithm of the polar coordinates and used the following Translog specification:

$$\begin{aligned}\ln(\|y\|) = {} & \alpha_0 + \sum_{m=1}^{M-1} \alpha_m \vartheta_m + \frac{1}{2} \sum_{m=1}^{M-1} \sum_{j=1}^{M-1} \alpha_{mj} \vartheta_m \vartheta_j + \sum_{n=1}^{N} \beta_n \ln(x_n) \\ & + \frac{1}{2} \sum_{n=1}^{N} \sum_{l=1}^{N} \beta_{nl} \ln(x_n) \ln(x_l) + \sum_{m=1}^{M-1} \sum_{n=1}^{N} \gamma_{mn} \vartheta_m \ln(x_n) \\ & - u + v.\end{aligned} \tag{7}$$

3. Issues

The SRPF specification has been criticized as follows: Murillo-Zamorano (2004) points out the need to specify a particular functional form to estimate the SRPF compared to non- or semi-parametric approaches. However, the problems that follow from functional misspecification are not exclusive to the SRPF, but arise in any other situation where parametric methods are used, including the parametric estimations of (other) output distance functions. We, therefore, suggest thoroughly checking the suitability of the chosen functional form, e.g., by using the Regression Equation Specification Error Test (RESET) suggested by Ramsey (1969).[2]

Fernández et al. (2000) and Ferreira and Steel (2007) criticize the SRPF asserting that it is a univariate single-equation approach that only gives a single inefficiency measure for each observation, while they suggest using a multivariate multi-equation model that gives a product-specific efficiency measure for each output. However, according to microeconomic theory, *all* output distance functions only give a single overall efficiency measure and anyway

[1] Another parametric approach that can be used when some of the output quantities of some observations are zero is a quadratic directional distance function. However, this approach is not often used in empirical applications. One major drawback of the quadratic directional distance function is that its noise and inefficiency terms represent *absolute* deviations from the frontier so that the these terms are – in our experience – strongly heteroskedastic in many empirical applications. In this case, the estimation of efficiency scores would require to explicitly model the heteroscedasticity of the noise and inefficiency terms, which complicates the estimation and bears the risk of misspecifying the parametrizations of the variances of these terms.

[2] Alternatively, one can use nonparametric estimation techniques that do not require the specification of a functional form (see, e.g., Kumbhakar et al., 2007; Sun, 2015).