



IFRO Working Paper

New Insights into the Stochastic Ray Production Frontier

Arne Henningsen
Matěj Bělín
Géraldine Henningsen

2017 / 01

IFRO Working Paper 2017 / 01

New insights into the stochastic ray production frontier

Authors: Arne Henningsen, Matěj Bělín, Géraldine Henningsen

JEL-classification: C13, C51, D22

Published: January 2017

See the full series IFRO Working Paper here:

www.ifro.ku.dk/english/publications/foi_series/working_papers/

Department of Food and Resource Economics (IFRO)

University of Copenhagen

Rolighedsvej 25

DK 1958 Frederiksberg DENMARK

www.ifro.ku.dk/english/

New Insights into the Stochastic Ray Production Frontier

Arne Henningsen^{a*}, Matěj Bělín^b, and 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

* Corresponding author, e-mail: arne@ifro.ku.dk

January 6, 2017

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.

Keywords: Stochastic Ray Production Frontier, Distance Function, Multiple Outputs, Primal Approach, Zero Output Quantities

JEL codes: C13, C51, D22

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

Nevertheless, the SRPF is not without its shortcomings. In this paper, we briefly discuss—and partly refute—a few existing criticisms of the SRPF. 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. 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; section three thoroughly discusses previously raised and new issues regarding the SRPF and gives suggestions as to how to handle them, and finally; section four concludes.

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 reason could be that the noise and inefficiency terms of this approach represent *absolute* deviations from the frontier so that in most data sets the noise and inefficiency terms of a quadratic directional distance function would be heteroskedastic.

2 Specification of the Stochastic Ray Production Function

The SRPF can be seen as a specific functional form of a standard Shepard output distance function:

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

where $x = (x_1, x_2, \dots, x_N)^\top$ is a vector of N input quantities, $y = (y_1, y_2, \dots, y_M)^\top$ 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 $\vartheta = (\vartheta_1, \dots, \vartheta_{M-1})$ that are recursively defined by:

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

with $\sin(\vartheta_0) = \cos(\vartheta_M) = 1$.

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

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

$$\|y\| = D^o(x, y) / D^o(x, p(\vartheta)) \quad (4)$$

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:

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

Löthgren (1997) suggests the following Translog functional form:

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

where α_0 ; $\alpha_m, m = 1, \dots, M-1$; $\beta_n, n = 1, \dots, N$; $\alpha_{mj}, m, j = 1, \dots, M-1$ with $\alpha_{mj} = \alpha_{jm} \forall m, j$; $\beta_{nl}, n, l = 1, \dots, N$ with $\beta_{nl} = \beta_{ln} \forall n, l$; and $\gamma_{mn}, m = 1, \dots, M-1, n = 1, \dots, N$ 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:

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

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).

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 it is questionable whether product-specific efficiency measures are meaningful if how much of each input quantity is used to produce each of the outputs is unknown.²

Fernández et al. (2000) claim that the SRPF assumes input-output separability, but this is incorrect. In the specification of the SRPF (equations (6) and (7)), the shape of the transformation curves and, thus, the feasible output combinations depend on the combinations of the input quantities unless the coefficients of all interaction terms between the polar coordinate angles and the input quantities ($\gamma_{mn}; m = 1, \dots, M-1, n = 1, \dots, n$) are zero.

We now present further issues that we consider far more important for practical applications of the SRPF, but that have—to the best of our knowledge—not yet been addressed in the literature.

² For a more thorough discussion of the many other aspects of efficiency measures in a multiple-output space see Fernández et al. (2000) or Kumbhakar and Lovell (2000).

3.1 Calculation of the Polar Coordinate Angles

The calculation of the angles of the polar coordinates in equation (2) is recursive, so that rounding errors, which are unavoidable on digital computers, accumulate and an increasing number of outputs makes these calculations increasingly imprecise. Furthermore, if one output j becomes very large compared to the other outputs, $\sin(\vartheta_j)$ and, thus, the denominator in equation (2) tends towards zero, which quickly increases rounding errors (Hermann, 2006). Hence, the more unequally the outputs are distributed, the larger the accumulated rounding errors.

We suggest replacing equation (2) by:

$$\vartheta_m(y) = \arccos \left(\frac{y_m}{\sqrt{\sum_{j=m}^M y_j^2}} \right) \quad (8)$$

(Petersen and Bernasconi, 1997), which avoids recursive calculations and, thereby, reduces the magnitude of rounding errors in ϑ .

3.2 Logarithms of Angles

The two models specified in (6) and (7) are simply two different (non-nested) functional forms, which only differ because the former uses the logarithmic, while the latter uses the non-logarithmic angles of the polar coordinates (ϑ). In many empirical applications with cross-sectional or panel data, input and output quantities have right-skewed distributions with many small values and few large values. Taking logarithms of these input and output quantities often makes their distributions less skewed and narrows their ranges so that the assumptions of standard (linear) regressions models (e.g., homoskedasticity) are more likely fulfilled and the estimation results are less sensitive to outliers and extreme values (Wooldridge, 2016, p. 172).

In contrast, the angles of polar coordinates are always in the range of $[0, \pi/2]$ and their logarithms usually have much wider ranges (up to $[-\infty, \ln(\pi/2)]$) and are often left-skewed with few large negative values. Hence, it is usually advantageous to take the logarithms of quantities, but not to take logarithms of the angles of the polar coordinates as proposed in the model specification in (7).

Further advantages of not taking logarithms of the angles of the polar coordinates are given in sections 3.4 and 3.5.

3.3 Units of Measurement

Unlike traditional input and output distance functions, the SRPF is not invariant to the units of measurement of the outputs because a change in the units of measurement of an

output implies non-linear changes in the Euclidean distance ($\|y\|$) and the angles of the polar coordinates (ϑ).³ This means that the estimated technical efficiencies and properties of the production technology depend on the (arbitrarily chosen) units of measurement of the outputs.

One solution could be to measure all output quantities in the same monetary unit because the SRPF is invariant to proportional changes in the units of measurement of *all* outputs (e.g., conversion from Euro to USD) due to its linear homogeneity in output quantities. Another solution could be to normalize all output quantities by their mean values because the normalized output quantities are invariant to the units of measurement of the original output quantities. One can also check the robustness of the results by trying out different combinations of units of measurement for the output quantities. If the results vary substantially across the chosen combinations of units of measurement, various indicators can be used to select the most suitable model (e.g., log-likelihood value, out-of-sample prediction performance, Ramsey's (1969) RESET test).

3.4 Ordering of the Outputs

Unlike traditional input and output distance functions, the SRPF is not invariant to the ordering of the outputs. This means that the SRPF has the undesirable feature that the estimated technical efficiencies and properties of the production technology depend on the (arbitrarily chosen) ordering of the outputs. However, if one does not take the logarithm of the polar coordinates (as in equation 7), the Translog SRPF becomes invariant to the ordering of the last two outputs, i.e., two analyses with $y = (y_1, y_2, \dots, y_{M-2}, y_{M-1}, y_M)^\top$ and $y^* = (y_1, y_2, \dots, y_{M-2}, y_M, y_{M-1})^\top$, respectively, give identical efficiency estimates and identical representations of the production frontier. This is because the polar coordinates that correspond to the output vector y^* are linearly related to the polar coordinates that correspond to the original output vector y : $\vartheta_1^* = \vartheta_1, \dots, \vartheta_{M-2}^* = \vartheta_{M-2}, \vartheta_{M-1}^* = \frac{\pi}{2} - \vartheta_{M-1}$. However, when one takes the logarithm of the polar coordinates (as in equation 6), there is no longer a linear relationship between the regressors $\ln \vartheta_{M-1}$ and $\ln \vartheta_{M-1}^* = \ln\left(\frac{\pi}{2} - \vartheta_{M-1}\right)$.

In the case of two outputs ($M = 2$), specification (7) is completely invariant to the ordering of the outputs. In the case of more than two outputs ($M > 2$), we suggest testing (all) the different orderings of the outputs and checking the robustness of the results regarding changes of the ordering of the outputs. As specification (7) is invariant to the ordering of the last two outputs, this specification has the advantage that there are only half as many different orderings of the outputs that result in different estimates

³This also applies to the directional distance function, unless the distance vector is specified in relative terms.

than specification (6). If the results considerably depend on the ordering of the outputs, various indicators could be used to select the most suitable model (see section 3.3).

3.5 Zero Output Quantities

Contrary to traditional input and output distance functions, the SRPF can handle observations where one or several output quantities are nil. However, when estimating the SRPF over M different outputs using specification (6), the last output quantity, y_M , needs to be strictly positive across all observations. If this is not the case and the last output quantity is zero ($y_M = 0$), ϑ_{M-1} becomes zero too (see equation 8), and hence, the logarithm of ϑ_{M-1} will be undefined.

In contrast, specification (7) only demands that at least one output quantity per observation—no matter which one—is strictly positive. Hence, if a data set contains zero values for at least one observation across all output quantities, specification (6) is not applicable and should be replaced by specification (7).⁴

4 Conclusion

The SRDF is a practical alternative to traditional specifications of output distance functions, particularly when some of the output quantities are nil for some observations. This article discusses several practical implications of the SRDF, which have not been examined before. 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.

References

Bhattacharyya, A., Pal, S., 2013. Financial reforms and technical efficiency in Indian commercial banking: A generalized stochastic frontier analysis. *Review of Financial Economics* 22 (3), 109–117.

⁴ We strongly advise against the malpractice of replacing true zeros in the output quantities by small values as this approach can lead to considerable distortions of the results (see Henningsen et al., 2015, for more details).

- Fernández, C., Koop, G., Steel, M., 2000. A Bayesian analysis of multiple-output production frontiers. *Journal of Econometrics* 98 (1), 47–79.
- Ferreira, J. T. A. S., Steel, M. F. J., 2007. Model comparison of coordinate-free multivariate skewed distributions with an application to stochastic frontiers. *Journal of Econometrics* 137 (2), 641–673.
- Fousekis, P., 2002. Distance vs. ray functions: An application to the inshore fishery of greece. *Marine Resource Economics* 17 (4), 251–267.
- Henningsen, G., Henningsen, A., Jensen, U., 2015. A Monte Carlo study on multiple output stochastic frontiers: A comparison of two approaches. *Journal of Productivity Analysis* 44 (3), 309–320.
- Hermann, M., 2006. *Numerische Mathematik*. Oldenbourg.
- Kumbhakar, S. C., Lovell, C. A. K., 2000. *Stochastic Frontier Analysis*. Cambridge University Press, Cambridge.
- Löthgren, M., 1997. Generalized stochastic frontier production models. *Economics Letters* 57, 255–259.
- Löthgren, M., 2000. Specification and estimation of stochastic multiple-output production and technical inefficiency. *Applied Economics* 32, 1533–1540.
- Managi, S., Opaluch, J. J., Jin, D., Grigalunas, T. A., 2006. Stochastic frontier analysis of total factor productivity in the offshore oil and gas industry. *Ecological Economics* 60 (1), 204–215.
- Murillo-Zamorano, L. R., 2004. Economic efficiency and frontier techniques. *Journal of Economic Surveys* 18 (1), 33–77.
- Niquidet, K., Nelson, H., 2010. Sawmill production in the interior of British Columbia: A stochastic ray frontier approach. *Journal of Forest Economics* 16 (4), 257–267.
- Petersen, W. P., Bernasconi, A., 1997. Uniform sampling from an n-sphere. TR 97-06, Swiss Center for Scientific Computing, ETH Zürich.
- Quiggin, J., Bui-Lan, A., 1984. The use of cross-sectional estimates of profit functions for tests of relative efficiency: A critical review. *Australian Journal of Agricultural Economics* 28 (1), 44–55.
- Ramsey, J. B., 1969. Tests for specification errors in classical linear least-squares regression analysis. *Journal of the Royal Statistical Society. Series B (Methodological)* 31 (2), 350–371.

- Whiteman, J. L., 1999. The measurement of efficiency where there are multiple outputs. General Paper G-134, Center of Policy Studies, Monash University.
- Wooldridge, J. M., 2016. Introductory Econometrics. A Modern Approach, 6th Edition. Cengage Learning.
- Zhang, T., Garvey, E., 2008. A comparative analysis of multi-output frontier models. Journal of Zhejiang University - Science A 9 (10), 1426–1436.