Economics Bulletin

Volume 37, Issue 1

Sensitivity of directional technical inefficiency measures to the choice of the direction vector: a simulation study

Pedro Macedo
CIDMA - Center for Research & Development in Mathematics and Applications, University of Aveiro

Elvira Silva
Faculty of Economics & Center for Economics and Finance at University of Porto (CEF.UP)

Abstract

The purpose of this note is to provide new insights on the sensitivity of technical inefficiency scores, estimated using a directional distance function with small samples and the presence of outliers, to the choice of the direction vector. A simulation study with a geometric illustration is conducted considering several direction vectors. To the best of authors' knowledge, this is the first simulation work comparing 16 direction vectors, some of which are often employed in empirical studies. The four directional vectors that consistently provide the best results are identified and used in the empirical application discussed in this study.

We would like to express our gratitude to the Referee and to the Editor who offered valuable comments and suggestions for improvements. This work was supported in part by the Portuguese Foundation for Science and Technology (FCT - Fundação para a Ciência e a Tecnologia), through CIDMA - Center for Research and Development in Mathematics and Applications, within project UID/MAT/04106/2013.
Citation: Pedro Macedo and Elvira Silva, (2017) "Sensitivity of directional technical inefficiency measures to the choice of the direction vector: a simulation study", *Economics Bulletin*, Volume 37, Issue 1, pages 52-62
Contact: Pedro Macedo - pmacedo@ua.pt, Elvira Silva - esilva@fep.up.pt.
Submitted: July 05, 2016. **Published:** January 13, 2017.

1. Introduction

Selection of a direction vector when estimating a directional distance function, and the sensitivity of technical inefficiency scores to the choice of the direction vector have originated a few studies concerned explicitly with this issue (e.g., Zofio, Pastor & Aparicio, 2013; Färe, Grosskopf & Whittaker, 2013; Peyrache & Daraio, 2012). The flexibility gained with the introduction of the directional distance functions, as opposed to the radial (Shephard) distance functions, has the counterpart of the sensitivity of technical inefficiency scores to the choice of the direction vector along which inefficiency is evaluated.

The purpose of this note is to provide new insights on the sensitivity of technical inefficiency scores to the choice of the direction vector in models with small samples and the presence of outliers. A simulation study is conducted considering 16 direction vectors, including the one based on sample medians, which is not frequently used in the efficiency literature. Since it is widely-known in robust statistics that median is more robust than the mean, the idea is to explore it in the context of directional distance functions. The four directional vectors that consistently provide the best results are identified and used in the empirical application discussed in this study.

2. Which direction vectors have been used?

In the nineties, Chambers, Chung & Färe (1996, 1998) introduce the directional distance functions based on the benefit function and the shortage function developed by Luenberger (1992, 1995). These functions allow the construction of directional measures of technical and economic inefficiency, as opposed to the radial measures, based on Shephard's (or radial) distance functions. The directional distance functions are an important contribution to the measurement of productive efficiency and productivity. Based on the directional distance functions, Chambers (1998, 2002) introduce the Luenberger productivity indicators and Chung, Färe & Grosskopf (1997) propose the Malmquist-Luenberger productivity index.

A series of empirical studies appear in the literature using the directional distance functions with different research purposes. Some empirical studies attempt to assess technical and economic inefficiency (e.g., Koutsomanoli-Filippaki, Margaritis & Staikouras, 2012; Glass *et al.*, 2006; Färe, Grosskopf & Weber, 2004), productivity growth (e.g., Nakano & Magani, 2008; Guironnet & Peypoch, 2007; Chambers, Färe & Grosskopf, 1996), and investigate environmental issues (e.g., Beltrán-Esteve *et al.*, 2014; Rødseth, 2014; Picazo-

Tadeo, Beltrán-Esteve & Gómez-Limón, 2012; Domazlicky & Weber, 2003; Färe *et al.*, 2005; Färe, Grosskopf & Pasurka, 2001).

The selection of the direction vector in empirical studies can be classified in two general categories. The first one consists in a firm's specific direction vector, namely: (*i*) $(g_x, g_y) = (x, y)$; and (*ii*) $(g_x, g_y) = (\vec{0}, y)$ or $(g_x, g_y) = (x, \vec{0})$. The second category involves a direction vector common to all observations: (*iii*) $(g_x, g_y) = (\bar{x}, \bar{y})$, where $\bar{x}$ and $\bar{y}$ are, respectively, the sample means of *x* and *y;* (*iv*) $(g_x, g_y) = (\vec{1}, \vec{1})$; (*v*) $(g_x, g_y) = (\vec{1}, \vec{0})$, $(g_x, g_y) = (\vec{0}, \vec{1})$, or $(g_x, g_y) = (0, \sum_j y_j)$ where y_j is a vector of outputs and $\sum_j y_j$ is a vector of total outputs.[1]

In case (*i*), technical inefficiency of each observation is evaluated along its own input-output bundle (e.g., Chambers, Färe & Grosskopf, 1996; Färe, Grosskopf & Weber, 2004; Glass *et al.*, 2006). Case (*ii*) either measures output expansion in the firm's output direction (e.g., Chambers, Färe & Grosskopf, 1996; Nin *et al.*, 2003) or input contraction in the firm's input direction (e.g., Chambers, Färe & Grosskopf, 1996). Case (*ii*) leads to the directional output or input distance functions that are the analog of the Shephard output or input distance functions.

Cases (*iii*) (e.g., Chambers, Färe & Grosskopf, 1996; Guironnet and Peypoch, 2007), (*iv*) (e.g., Färe, Grosskopf & Pasurka, 2001; Domazlicky & Weber, 2003; Färe *et al.*, 2005), and (*v*) (e.g., Foltz *et al.*, 2012; Weber & Xia, 2012; Ferrier, Leleu & Valdmanis, 2009) imply that inefficiency of all observations are evaluated along the same direction vector. The direction vectors in (*iii*)-(*v*) facilitate aggregation of efficiency and productivity indicators across firms to form aggregate (e.g., industry) efficiency or productivity indicators (e.g., Briec, Dervaux & Leleu, 2003).

In the empirical studies using (*i*)-(*v*), the direction vector is pre-defined and its elements are treated as exogenous variables in the estimation of the directional distance functions. Recently, some studies (e.g., Zofio, Pastor & Aparicio, 2013; Färe, Grosskopf & Whittaker, 2013) propose the endogenization of the direction vector, avoiding in this way *ad hoc* choices of the researcher. Recently, a data-driven approach is proposed by Daraio & Simar (2016).

[1] In the study by Ferrier, Leleu and Valdmanis (2009), $\sum_j y_j$ is the vector of total outputs produced within each standard metropolitan statistical area.

3. Simulation studies

It is widely known that nonparametric, deterministic frontier estimators (e.g., Data Envelopment Analysis (DEA) estimators) are, by construction, highly sensitive to outliers (e.g., Charnes et al., 1992; Wilson, 1995; Zhu, 1996). Additionally, these estimators suffer from the curse of dimensionality (e.g., Simar and Wilson, 2008). In this section, several simulation models are performed, although only five of them are reported here, to evaluate the sensitivity of technical inefficiency scores to the choice of the direction vector in models with small samples and the presence of outliers.

Different production technologies with one output and one or two inputs, exhibiting constant (CRS) and variable returns to scale (VRS), are considered. To illustrate geometrically each of the simulation models, some of the ten Decision Making Units (DMU's) define the production frontier and the remaining ones are randomly generated in 1000 trials inside a specific triangle or polyhedron (see Figure 1 to Figure 5).[2] Table I presents the mean value of the distance to the production frontier for each model with ten DMU's.

Model 1 represents a VRS production technology considering one output and one input with "no outliers".[3] Considering the directional technology distance function, it is interesting to note that $(g_x, g_y) = (median\, x,\ median\, y)$ provides the shortest distance to the production frontier when compared with the other directional vectors defined in the input-output space.

Models 2 and 3 correspond to a VRS production technology considering one output and one input with, respectively, one outlier and two outliers. Focusing on the simultaneous expansion of the output and contraction of the input, the average shortest distance to the frontier, in both models, is generated with the direction vector $(g_x, g_y) = (x, y)$ followed by $(g_x, g_y) = (median\, x,\ median\, y)$.

CRS production technologies considering one output and two inputs with one outlier and two outliers are represented, respectively, by model 4 and model 5. Considering simultaneously the expansion of the output and the contraction of inputs, the shortest distance to the frontier is achieved, in both models, by the unit direction vector followed by

---

[2] All the computations are accomplished in MATLAB software.

[3] Depending on the random values generated, some observations may be considered as outliers.

$(g_x, g_y) = (median\, x,\ median\, y)$. Focusing in an input-oriented directional distance function, the direction vectors (1,0) and $(median\, x,\ 0)$ provide a similar average distance to the frontier.

Table I: Mean values of the distance to the production frontier.

Direction vector	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**
(*x*,*y*)	5.6001	4.9129	4.1765	0.8851	0.8801
(*mean x, mean y*)	5.8012	5.0701	4.5618	0.8712	0.8666
(1,1)	5.7109	5.0064	4.2577	0.8519	0.8421
(*median x, median y*)	5.5990	4.9342	4.2277	0.8682	0.8544
(*mean x*, 1)	5.9449	5.2308	4.4586	0.9244	0.9257
(*median x*,1)	5.9447	5.2180	4.4433	0.9306	0.9224
(1, *mean y*)	11.2968	10.0354	8.7017	1.0907	1.0907
(1, *median y*)	11.0621	9.7822	8.3947	1.0992	1.0802
(0,*y*)	12.1147	10.6202	9.0320	1.4755	1.4586
(0,1)	12.1146	10.6202	9.0320	1.4755	1.4586
(0, *mean y*)	12.1147	10.6202	9.0320	1.4755	1.4586
(0, *median y*)	12.1147	10.6202	9.0320	1.4755	1.4586
(*x*,0)	6.0573	5.3101	4.5160	1.1699	1.6404
(1,0)	6.0573	5.3101	4.5160	1.0433	1.0314
(*mean x*, 0)	6.0573	5.3101	4.5160	1.0436	1.0420
(*median x*, 0)	6.0573	5.3101	4.5160	1.0437	1.0323

It is important to note that the presence of outliers, its number and its severity may have a huge influence in the results, namely in models with small samples. For example, in model 5, if the two outliers are considered as severe outliers (farther away from the center of mass of the remaining observations in the polyhedron), the direction vector based on the firm's input-output bundle (or the firm's input vector), which is often used in empirical studies, may not be an adequate choice to evaluate firm's inefficiency. Also, the direction vector defined by the sample means of *x* and *y* perform poor. In contrast, the direction vector based on the sample medians performs better in the presence of severe outliers. As expected, the same occurs in other simulation models, whose results are not reported in this note.

The directional vector $(g_x, g_y) = (median\, x,\ median\, y)$ provides the best or the second best result in the simulation models discussed here. In the other simulation models, whose results are not reported in this note, the directional vector based on the sample medians is always on the first three top best results (out of 16 directional vectors analyzed) corresponding to the lowest technical inefficiency scores. To a certain extent, we can assert that this direction vector is robust to different production specifications, the presence of outliers and its severity, in models with small samples.

Ranking the inefficiency results, generated by each of the four directional vectors that consistently provide the best results, based on the shortest distance (1) to the highest distance (4) to the frontier in all simulation models, including the ones not reported here, the average ranking of the directional vector based on the sample medians is 1.9. The direction vector $(g_x, g_y) = (x, y)$ and the unit direction vector have an equal average ranking of 2.2 and the directional vector based on the sample means has, on average, a ranking equal to 3.8.

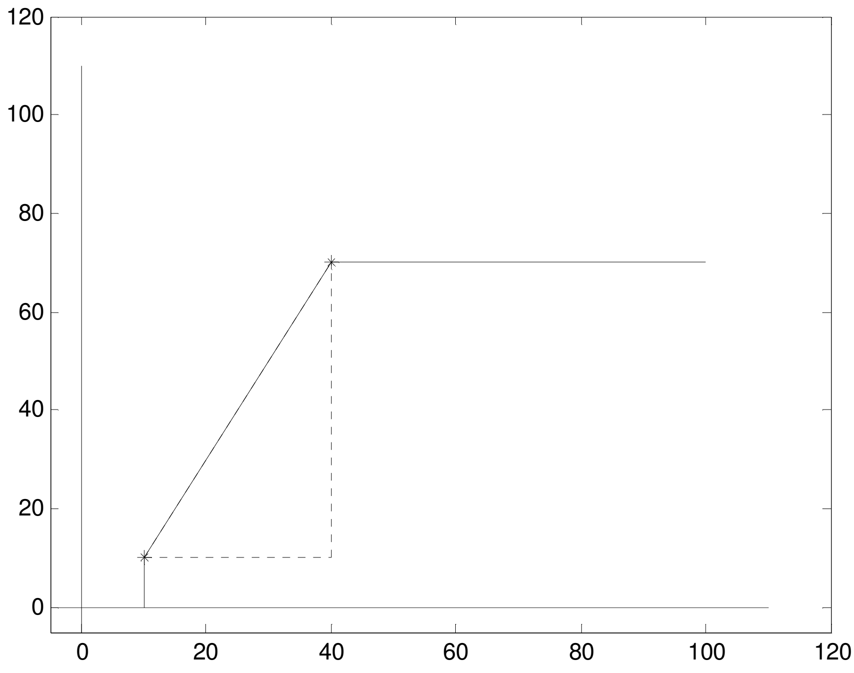

Figure 1. Model 1: VRS technology, one output and one input, with “no outliers”.

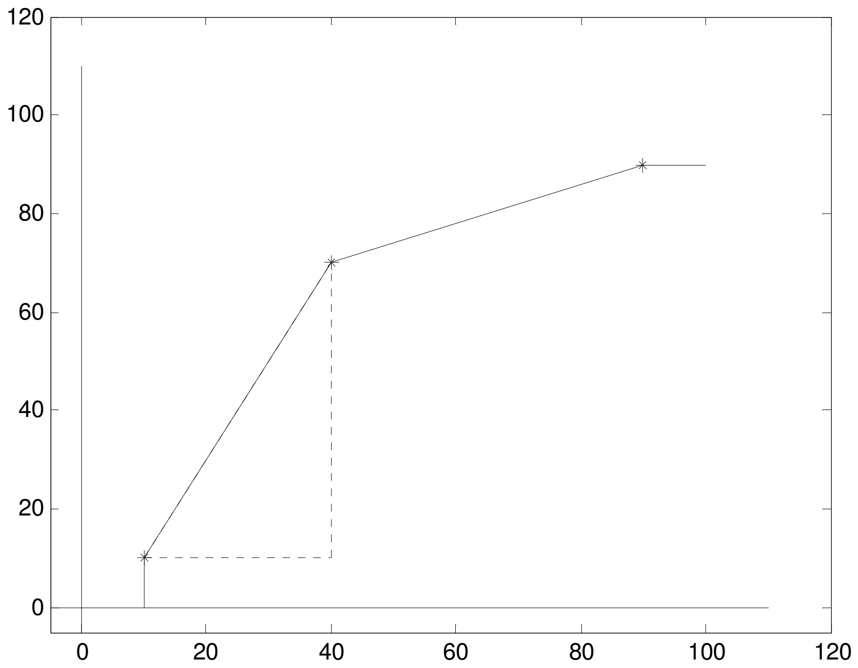

Figure 2. Model 2: VRS technology, one output and one input, with one outlier.

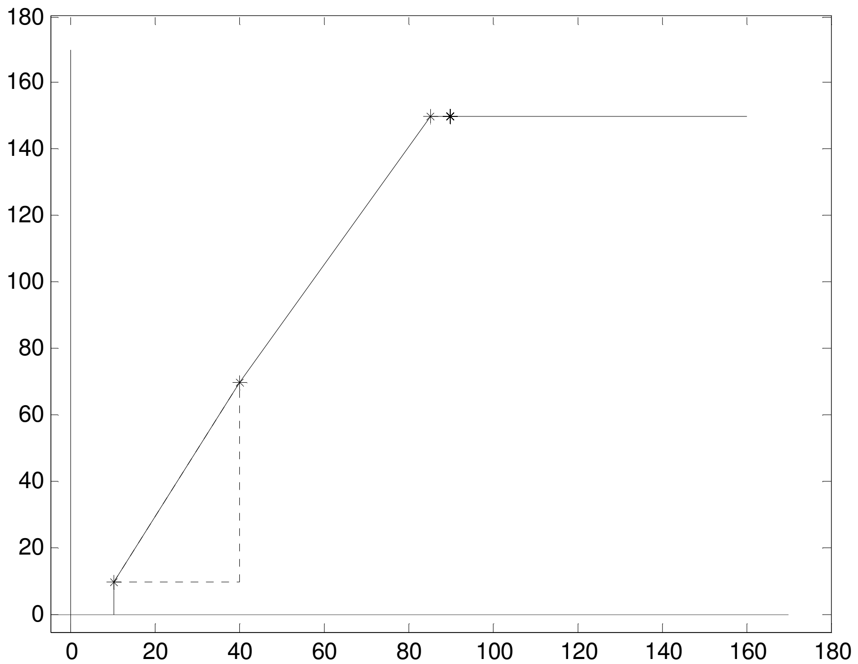

Figure 3. Model 3: VRS technology, one output and one input, with two outliers.

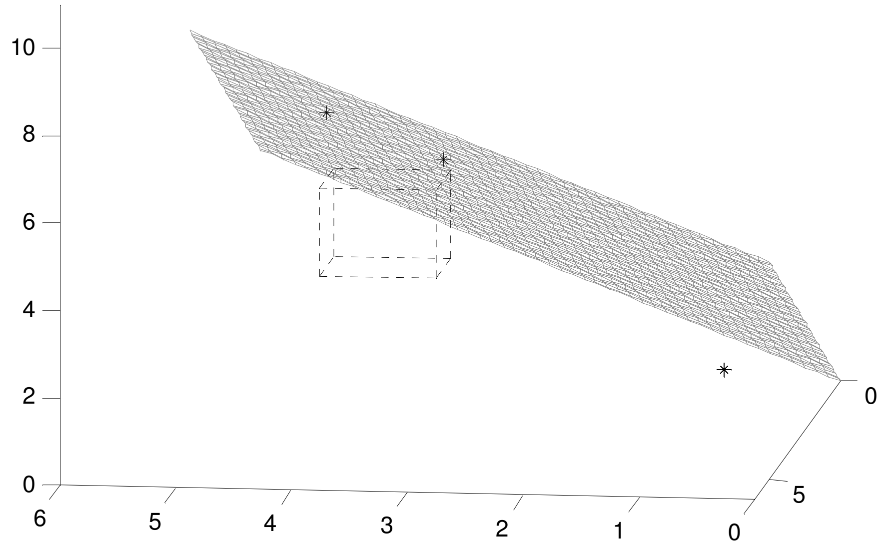

Figure 4. Model 4: CRS technology, one output and two inputs, with one outlier.

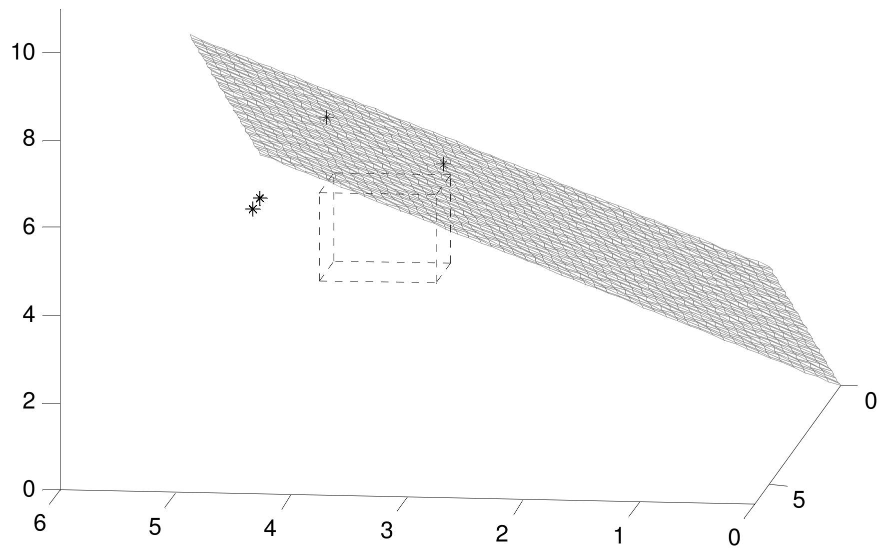

Figure 5. Model 5: CRS technology, one output and two inputs, with two outliers.

4. An empirical application

The four directional vectors that provide consistently the best results in the simulation study are used in an empirical application with the database "Philippines Rice Data" from Coelli et al. (2005), available on Package "frontier" (Coelli & Henningsen, 2013). For each rice producer, there is information on the output (tonnes of rice), area planted (hectares), labor used (man-days of family and hired labor), fertilizer (kg of active ingredients) and other inputs used. Considering 10 rice producers randomly selected in the year 1997, the information on 3 producers is modified accordingly to illustrate a possible outlier contamination in the sample. Figure 6 presents boxplots for the technology directional distance function, estimated using DEA under CRS and VRS.

As expected, the DEA inefficiency scores under VRS are smaller than or equal to the ones generated under CRS. The unit direction vector presents the worst results in this example. The direction vector $(g_x, g_y) = (x, y)$ and the directional vectors based on the sample medians and the sample means have a similar performance.

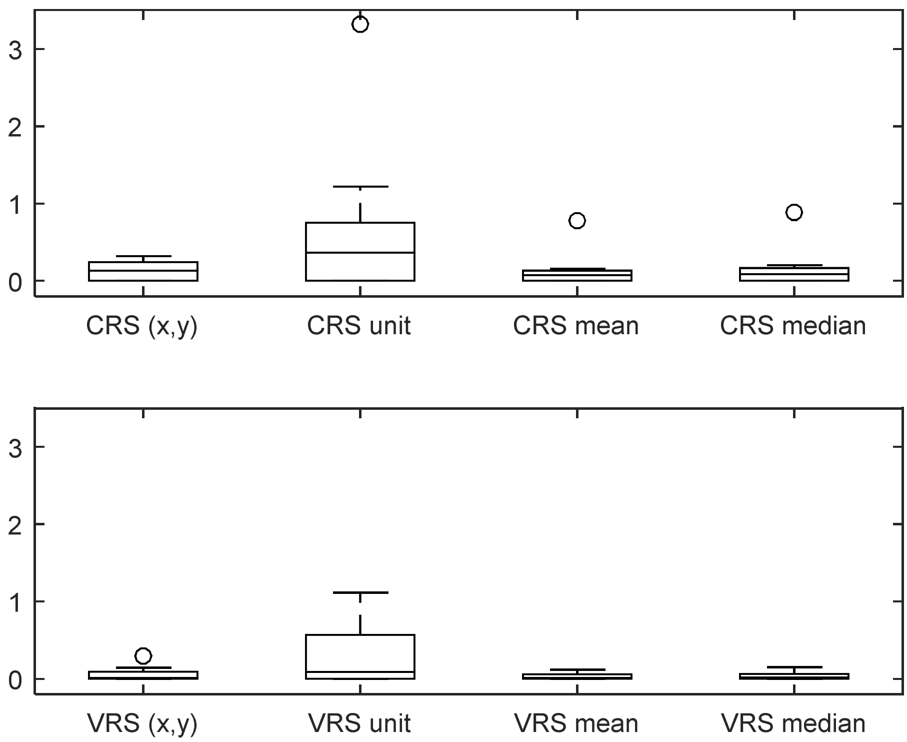

Figure 6. Value of the technology directional distance function under CRS and VRS.

5. Conclusions

This note evaluates the sensitivity of the technical inefficiency scores to the choice of 16 directional vectors, in models with small samples and the presence of outliers. To the best of authors' knowledge, this is the first simulation study with a comparison among several directional vectors, including the one based on sample medians, which is not frequently used in the efficiency literature. Considering all the simulation models, including the ones not reported here, the directional vector based on the sample medians obtains the best average position in the rankings of the four directional vectors that provide consistently the best results. To a certain extent, we can assert that this direction vector is robust to different production specifications, the presence of outliers and its severity, in models with small samples. However, these conclusions should be tempered with caution, since more complete simulation studies are needed in future research.

Acknowledgements

This work was supported in part by the Portuguese Foundation for Science and Technology (FCT - Fundação para a Ciência e a Tecnologia), through CIDMA - Center for Research and Development in Mathematics and Applications, within project UID/MAT/04106/2013.

References

Beltrán-Esteve, M., Gómez-Limón, J. A., Picazo-Tadeo, A. J., & Reig-Martínez, E. (2014). A metafrontier directional distance function approach to assessing eco-efficiency. *Journal of Productivity Analysis*, 41, 69-83.

Briec, W., Dervaux, B., & Leleu, H. (2003). Aggregation of directional distance functions and industry efficiency. *Journal of Economics*, 79, 237-261.

Chambers, R. G. (2002). Exact nonradial input, output and productivity measurement. *Economic Theory*, 20, 751-765.

Chambers, R. G. (1998). Input and output indicators. In R. Färe, S. Grosskopf, & R. R. Russell (Eds.), *Index Numbers: Essays in Honour of Sten Malmquist* (pp. 241-271). Boston: Kluwer Academic Publishers.

Chambers, R. G., Chung, Y., & Färe, R. (1996). Benefit and distance functions. *Journal of Economic Theory,* 70, 407-419.

Chambers, R. G., Chung, Y., & Färe, R. (1998). Profit, directional distance functions, and Nerlovian efficiency. *Journal of Optimization Theory*, 98, 351-364.

Chambers, R. G., Färe, R., & Grosskopf, S. (1996). Productivity growth in APEC countries. *Pacific Economic Review*, 1, 181-190.

Charnes, A., Haag, S., Jaska, P., & Semple, J. (1992). Sensitivity of efficiency classifications in the additive model of Data Envelopment Analysis. *International Journal of Systems Science,* 23, 789-798.

Chung, Y. H., Färe, R., & Grosskopf, S. (1997). Productivity and undesirable outputs: A directional distance function approach. *Journal of Environmental Management*, 51, 229-240.

Coelli, T., & Henningsen, A. (2013). frontier: Stochastic frontier analysis. R package version 1.0. (http://CRAN.R-Project.org/package=frontier).

Coelli, T. J., Rao, D. S. P., O'Donnell, C. J., & Battese, G. E. (2005). *An Introduction to Efficiency and Productivity Analysis*. New York: Springer.

Daraio, C. & Simar, L. (2016). Efficiency and benchmarking with directional distances: a data driven approach. *Journal of the Operational Research Society,* to appear.

Domazlicky, B. R., & Weber, W. L. (2003). Does environmental protection lead to slower productivity growth in the chemical industry? *Environmental and Resource Economics,* 28, 301-324.

Färe, R., Grosskopf, S., Noh, D.- W., & Weber, W. L. (2005). Characteristics of a polluting technology: Theory and practice. *Journal of Econometrics*, 126, 469-492.

Färe, R., Grosskopf, S., & Pasurka Jr., C. A. (2001). Accounting for air pollution emissions in measures of state manufacturing productivity growth, *Journal of Regional Science*, 41, 381-409.

Färe, R., Grosskopf, S., & Weber, W. L. (2004). The effect of risk-based capital requirements on profit efficiency in banking. *Applied Economics*, 36, 1731-1743.

Färe, R., Grosskopf, S., & Whittaker, G. (2013). Directional output distance functions: endogenous directions based on exogenous normalization constraints. *Journal of Productivity Analysis*, 40, 267-269.

Ferrier, G. D., Leleu, H., & Valdmanis, V. G. (2009). Hospital capacity in large urban areas: Is there enough in times of need? *Journal of Productivity Analysis*, 32, 103-117.

Foltz, J. D., Barham, L. B., Chavas, J.- P., & Kim, K. (2012). Efficiency and technological change at US research universities. *Journal of Productivity Analysis*, 37, 171-186.

Glass, J. C., McCallion, G., McKillop, D. G., & Stringer, K. (2006). A 'technically level playing-field' profit efficiency analysis of enforced competition between publicly funded institutions. *European Economic Review*, 50, 1601-1626.

Guironnet, J. P., & Peypoch, N. (2007). Human capital allocation and overeducation: A measure of French productivity (1987, 1999). *Economic Modelling,* 24, 398-410.

Koutsomanoli-Filippaki, A., Margaritis, D., & Staikouras, C. (2012). Profit efficiency in the European Union banking industry: a directional technology distance function approach. *Journal of Productivity Analysis*, 37, 277-293.

Luenberger, D. G. (1995). *Microeconomic Theory*. (International ed.). New York: McGraw-Hill, (chapters 2 and 4).

Luenberger, D. G. (1992). Benefit functions and duality. *Journal of Mathematical Economics*, 21, 461-481.

Nakano, M., & Managi, S. (2008). Regulatory reforms and productivity: An empirical analysis of the Japanese electricity industry. *Energy Policy*, 36, 201-209.

Nin, A., Arndt, C., Hertel, T. W., & Preckel, P. V. (2003). Bridging the gap between partial and total factor productivity measures using directional distance functions. *American Journal of Agricultural Economics*, 85(4), 928-942.

Peyrache, A., & Daraio, C. (2012). Empirical tools to assess the sensitivity of directional distance functions to direction selection. *Applied Economics*, 44, 933-943.

Picazo-Tadeo, A. J., Beltrán-Esteve, M., & Gómez-Limón, J. A. (2012). Assessing eco-efficiency with directional distance functions. *European Journal of Operational Research*, 220, 798-809.

Rødseth, K. L. (2014). Efficiency measurement when producers control pollutants: a non-parametric approach. *Journal of Productivity Analysis*, 42(2), 211-223.

Simar, L., & Wilson, P. (2008) Statistical inference in nonparametric frontier models: recent developments and perspectives, in *The Measurement of Productive Efficiency and Productivity Growth* by Harold. O. Fried, C. A. Knox Lovell and Shelton S. Schmidt, Eds., Oxford University Press, 421-521.

Weber, W. L., & Xia, Y. (2011). The productivity of nanobiotechnology research and education in U.S. universities. *American Journal of Agricultural Economics*, 93(4), 1151-1167.

Wilson, P. (1995). Detecting influential observations in data envelopment analysis. *Journal of Productivity Analysis*, 6, 27-45.

Zhu, J. (1996). Robustness of the efficient DMUs in data envelopment analysis. *European Journal of Operations Research*, 90, 451-460.

Zofio, J. L., Pastor, J. T., & Aparicio, J. (2013). The directional profit efficiency measure: On why profit inefficiency is either technical or allocative. *Journal of Productivity Analysis*, 40, 257-266.