European Journal of Operational Research 237 (2014) 358–369

Contents lists available at ScienceDirect

European Journal of Operational Research

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

Interfaces with Other Disciplines

Directional distances and their robust versions: Computational and testing issues

Cinzia Daraio [a], Léopold Simar [a,b,*]

[a] Department of Computer, Control and Management Engineering Antonio Ruberti (DIAG), University of Rome "La Sapienza", Rome, Italy
[b] Institute of Statistics, Biostatistics et Actuarial Sciences, Université Catholique de Louvain, Louvain-la-Neuve, Belgium

ARTICLE INFO

Article history:
Received 16 August 2013
Accepted 29 January 2014
Available online 6 February 2014

Keywords:
Directional distances
Free Disposal Hull (FDH)
Conditional efficiency measures
Nonparametric frontiers
Significance test
Bootstrap

ABSTRACT

Directional distance functions provide very flexible tools for investigating the performance of Decision Making Units (DMUs). Their flexibility relies on their ability to handle undesirable outputs and to account for non-discretionary inputs and/or outputs by fixing zero values in some elements of the directional vector. Simar and Vanhems (2012) and Simar, Vanhems, and Wilson (2012) indicate how the statistical properties of Farrell–Debreu type of radial efficiency measures can be transferred to directional distances. Moreover, robust versions of these distances are also available, for conditional and unconditional measures. Bădin, Daraio, and Simar (2012) have shown how conditional radial distances are useful to investigate the effect of environmental factors on the production process. In this paper we develop the operational aspects for computing conditional and unconditional directional distances and their robust versions, in particular when some of the elements of the directional vector are fixed at zero. After that, we show how the approach of Bădin et al. (2012) can be adapted in a directional distance framework, including bandwidth selection and two-stage regression of conditional efficiency scores. Finally, we suggest a procedure, based on bootstrap techniques, for testing the significance of environmental factors on directional efficiency scores. The procedure is illustrated through simulated and real data.

1. Introduction

In productivity and efficiency analysis, most of theoretical and empirical studies have been based on the Farell–Debreu radial oriented measures (Farrell, 1957, Debreu, 1951, Shephard, 1970). The basic idea was to gauge how much the outputs should be increased proportionally (maximal attainable value), given the level of the inputs used, to reach the efficient frontier. Alternatively, mainly when the outputs are not under the control of the DMUs, like in some service industries, one could analyze how much a firm should reduce its inputs proportionally, given the level of outputs it is producing.

Later, directional distance functions have been introduced (see Chambers, Chung, & Färe, 1996, 1998, Färe & Grosskopf, 2004, Färe, Grosskopf, & Margaritis, 2008) to generalize the radial input and output distance functions. A directional distance function projects the input–output vector onto the technology frontier in a direction given by a vector d.[1] It encompasses indeed both the input and the output oriented radial measures as special cases when some elements of the directional vector d are fixed at zero.

Recently, Simar and Vanhems (2012) have shown that by choosing an appropriate probabilistic formulation of the production process (as initiated by Cazals, Florens, & Simar, 2002), all the known statistical properties of the nonparametric estimators of the radial efficiency scores were easily adapted to the FDH nonparametric estimators of the directional distance functions. They provided also robust versions of these estimators, based on the order-m partial frontiers (Cazals et al., 2002) and order-α quantile frontiers (Daouia & Simar, 2007). Finally, Simar and Vanhems (2012) only sketch how conditional directional distances could be defined in this framework, without providing any information about their computational implementation. Furthermore, Simar et al. (2012) analyze the statistical properties of the DEA estimators of directional distances. Statistical inference for individual directional distances was derived in these papers, and it implies

* Corresponding author at: Institute of Statistics, Biostatistics et Actuarial Sciences, Université Catholique de Louvain, Louvain-la-Neuve, Belgium.

E-mail addresses: daraio@dis.uniroma1.it (C. Daraio), leopold.simar@uclouvain.be (L. Simar).

[1] "Directional distance functions" is the terminology used in Chambers et al. (1996, 1998) and Färe et al. (2008). For sake of simplicity we will rather refer, as in our title, to the shorter "directional distances" terminology.

http://dx.doi.org/10.1016/j.ejor.2014.01.064
0377-2217/

the use of bootstrap methods.

Interestingly, the great flexibility of the directional distances rests in their ability to handle non-discretionary inputs and/or outputs by simply setting at zero any subset of the vector d. The only constraint is that the vector d should not be equal to zero for all its components.

Another important aspect for the practitioner in production analysis is to investigate the impact of environmental-external factors on the production process. Recently, Bădin et al. (2012) have developed a methodology initiated by Daraio and Simar (2005, 2007) for this specific purpose. Their approach uses conditional efficiency measures (see Bădin, Daraio, & Simar, 2014 for a recent survey of available techniques). All these approaches, however, use traditional radial measures.

In this paper we combine the tools recently developed by Simar and Vanhems (2012) and Bădin et al. (2012) by adapting the methodology for detecting the impact of external-environmental variables on the production process to the directional distance framework. Our contribution is thus fourfold.

- First we operationalize, by explicating the algorithms, the computation of directional distance estimates, where Simar and Vanhems (2012) were only mentioning the possibility of extension, without giving any computational details. In particular, we provide a practical procedure for computing FDH estimates of directional distances and their robust versions when some of the elements of the directional vector are zeros (both in inputs and/or in outputs).[2]
- Second, we make explicit the computations for the conditional distance estimates, including their robust versions. By doing this, we particularize to conditional directional distances the procedure for selecting the appropriate bandwidth, suggested by Bădin, Daraio, and Simar (2010).
- Third, we adapt the methodology for measuring the impact of environmental variables implemented so far for radial oriented efficiency scores (Bădin et al., 2012) to the directional distance context. This includes the appropriate second stage regression to explore the effect of external variables on the expected efficiency scores.
- Finally, we provide a test for assessing the significance of the effect of external variables on the expected efficiency scores. This test adapts a bootstrap methodology suggested for traditional nonparametric regression to this particular context. Specifically, we show how a consistent bootstrap test can be implemented by working with order-α quantile frontiers. The procedure is illustrated with some simulated data sets and with a real data set on Mutual Funds. We analyze the role of Market Risk on the mean efficiency in a simple Mean–Variance model.

The paper is organized as follows. Section 2 introduces the basic notation for directional distances and their robust versions. In Section 3 we illustrate how to compute the FDH nonparametric estimators of directional distances when some elements of the direction d are set at zero. Then Section 4 gives all the details for computing conditional directional distances and their robust versions. The significance test for the external factors, based on bootstrap methods, is explained in Section 5. Section 6 illustrates the proposed procedure with some data sets. The bootstrap algorithm (including the double bootstrap) is detailed in Appendix A. Section 7 summarizes the main findings and concludes the paper.

2. Directional distances

2.1. Basic concepts and notations

In production theory (see Shephard, 1970), we consider a set of producing units (hereafter we will use the term "DMU") that produce a set of outputs $Y \in \mathbb{R}^q$ by combining a set of inputs $X \in \mathbb{R}^p$. The technology is characterized by the attainable set T, the set of all the combinations of (x, y) that are technically achievable, defined as:

$$T = \{(x, y) \in \mathbb{R}^p \times \mathbb{R}^q | x \text{ can produce } y\}. \quad (2.1)$$

We know (Cazals et al., 2002) that under the free disposability assumption for the inputs and the outputs,[3] the set can be described as:

$$T = \{(x, y) \in \mathbb{R}^p \times \mathbb{R}^q | H_{XY}(x, y) > 0\}, \quad (2.2)$$

where $H_{XY}(x, y)$ is the probability of observing a unit (X, Y) dominating the production plan (x, y), i.e. $H_{XY}(x, y) = \text{Prob}(X \leqslant x, Y \geqslant y)$.

The efficient boundary of T is of interest and several ways have been proposed in the literature to measure the distance of the unit (x, y) to (from) the efficient frontier. One of the most flexible approaches is based on directional distances introduced by Chambers et al. (1998) (see also Färe & Grosskopf, 2004 & Färe et al., 2008). Given a directional vector for the inputs $d_x \in \mathbb{R}^p_+$ and a direction for the outputs $d_y \in \mathbb{R}^q_+$, a directional distance is defined as:

$$\beta(x, y; d_x, d_y) = \sup\{\beta > 0 | (x - \beta d_x, y + \beta d_y) \in T\}, \quad (2.3)$$

or equivalently, under the free disposability assumption (see Simar & Vanhems, 2012):

$$\beta(x, y; d_x, d_y) = \sup\{\beta > 0 | H_{XY}(x - \beta d_x, y + \beta d_y) > 0\}. \quad (2.4)$$

That is, we measure the distance of unit (x, y) from the efficient frontier in an additive way, and along the path defined by $(-d_x, d_y)$. This way of measuring the distance is very flexible and generalizes the "oriented" radial measures proposed by Debreu (1951) and Farrell (1957), see also Shephard (1970). Certainly, by choosing $d_x = 0$ and $d_y = y$ (or $d_x = x$ and $d_y = 0$), we can recover the traditional Farrell–Debreu output (resp. input) radial distance. The flexibility of this approach rests on the fact that we might have some elements of the vector d_x and/or of the vector d_y that can be set at zero. This is the case when one wants to focus the analysis on distances to the frontier along certain particular *paths* or, for instance, when some inputs or outputs are non-discretionary, or not under the control of the manager. The usefulness of selecting zero elements in the directional vectors is not only to allow to handle input and output orientations but also to handle exogenously fixed inputs or outputs. Banker and Morey (1986) show how to achieve this when using Farrell type efficiency measures. They give the example of units willing to estimate their input savings and having to deal with some inputs that they do not control (like e.g., level of advertising and number of competitors, see Banker and Morey for a detailed discussion). These inputs are thus exogenously fixed and it is not meaningful to reduce them. Banker and Morey (1986) also give an example of exogenously fixed outputs (like e.g. check cashing in transactions in a bank, which is a purely gratis service function, so the banks try to maximize the outputs that are under their own control). We could also consider the case of bad outputs that we do not want to maximize (see Simar & Vanhems, 2012 for a discussion).

An important point to note is that the efficient frontier is uniquely defined by the boundary of the attainable set T (where all the inputs and outputs are involved), but the distance to the fron-

[2] To save space, we limit the presentation to the case of FDH and quantile frontiers. This can be adapted without much difficulty to the order-m partial frontier cases. We summarize in Appendix B the main steps for the order-m cases.

[3] The free disposability we used in this paper is the assumption that if $(x, y) \in T$ then $(\tilde{x}, \tilde{y}) \in T$ for all $\tilde{x} \geqslant x$ and all $\tilde{y} \leqslant y$. It is a minimal assumption generally made on production processes.