Journal of the Operational Research Society (2009) 60, 1511–1517

www.palgrave-journals.com/jors/

The hyperbolic-oriented efficiency measure as a remedy to infeasibility of super efficiency models

AL Johnson[1]* and LF McGinnis[2]

[1]*Texas A&M University, Texas, USA and* [2]*Georgia Institute of Technology, Atlanta, GA, USA*

The infeasibility problem in traditional super efficiency models has been well established. A generalization of traditional input- or output-oriented super efficiency models, the directional distance function also suffers from infeasibility and related problems. The hyperbolic-oriented efficiency measure provides an alternative to the input-oriented, output-oriented, and directional efficiency measures in super efficiency models and it has the distinct advantage of eliminating the infeasibility problem for positive input/output data. We also show that using a hyperbolic orientation in a super efficiency model allows us to find feasible solutions for certain cases when the requirement for all data to be positive is relaxed. Further we demonstrate the hyperbolic orientated super efficiency method in an outlier detection application. Together, these results establish the use of the hyperbolic orientation in super efficiency analysis as a realistic alternative in practice.

Journal of the Operational Research Society (2009) **60,** 1511–1517. doi:10.1057/jors.2009.71
Published online 29 July 2009

Keywords: data envelopment analysis; hyperbolic distance function; super efficiency; outlier

Introduction

Performance measurement is an important issue in any enterprise. The ability to distinguish between top performing units and poorer performing units is important for a variety of reasons. In a public environment or a large company, performance measurement allows resources to be allocated to the units that are the most productive. In a competitive environment, it allows poor performers to understand the quality of their performance and to apply benchmarking techniques to guide them toward improvement. However, many industries or units operate in multi-input/multi-output environment. To understand performance, the set of relevant inputs and outputs needs to be considered simultaneously.

The study of performance measurement has an important starting point in Koopmans (1951), where he developed his definition of efficiency:

> A possible point in the commodity space is called efficient whenever an increase in one of its coordinates (the net output of one good) can be achieved only at the cost of a decrease in some other coordinate (the net output of another good).

The term *commodity space* is more commonly referred to as the production possibility set (PPS), meaning a set of all points, representing input (vector) and output (vector) pairs such that the input can be used to produce the output. Thus, a technically inefficient producer could produce its output with less of at least one input, or could use its inputs to produce more of at least one output. Based on these ideas, Shephard (1953) developed the input distance function in which he finds the equiproportional reduction of inputs for which the production of a given output set is still feasible. Later, Shephard (1970) developed the output-distance function in which input levels are fixed and the equiproportional increase of outputs is identified. These two measures represent two common measures of efficiency. Farrell (1957) introduced an efficiency measurement independent of Shephard, which is the basis for the technique referred to as data envelopment analysis (DEA) (Charnes *et al*, 1978). The decision to take an input or an output orientation when using the distance function or DEA has been a critical decision typically made based on an argument of whether the enterprise is cost minimizing (input orientation) or revenue maximizing (output orientation) (see, eg, Färe and Primont, 1995). Chambers *et al* (1998) introduced the directional distance function (DDF) as a generalization of input- and output-oriented distance functions, that allows for estimates of efficiency in a specified directional.

Färe *et al* (1985) introduced the hyperbolic distance function as an alternative to selecting either an input orientation or an output orientation. This measure is a simultaneous equiproportionate expansion of output and contraction of inputs. Because it requires a non-linear optimization problem, rather than the linear optimization problem associated with

*Correspondence: AL Johnson, Department of Industrial and Systems Engineering, Texas A&M University, 237K Zachry Engineering Center, College Station, Texas 77843-3131, USA.
E-mail: ajohnson@tamu.edu*

distance functions and DEA, this measure has been slow to gain popularity. Research in the area has been limited. Färe *et al* (1994) further explored the topic, proving a variety of properties of the hyperbolic distance function. It should be noted, the hyperbolic distance function is a special case of the generalized distance function developed by Chavas and Cox (1999). Färe *et al* (2002) related the measure to return to the dollar. Cuesta and Zofío (2005) used the hyperbolic distance function in a parametric context to extend a translog production function to a multi-input/multi-output production environment.

As with other distance functions, hyperbolic distance functions also can be applied in super efficiency models to measure the efficiency of an observation outside the PPS defined by the remaining observations. Two important applications of super efficiency models are outlier detection and calculating the Malmquist Productivity Index.

Johnson and McGinnis (2008) developed an outlier detection model that calculates both the input-oriented and the output-oriented super efficiency measures. An observation is flagged for further investigation if both the input-oriented and the output-oriented measures exceed a specified critical value. However, because these measures are calculated using linear programming and can return infeasible solutions (as discussed later), an observation may be identified as an outlier based on only one orientation or might be flagged because both orientations return infeasible solutions. Thus, the infeasibility of the input or output orientation causes the outlier detection method to make decisions based on incomplete information.

As noted by Zofío and Lovell (2001), using the hyperbolic distance function allowed them to calculate Malmquist Productivity Index for their data set. If it could be shown that the hyperbolic-oriented measure could always be calculated, particularly in the cases when an input-oriented or output-oriented distance functions could not, then the hyperbolic-oriented measure would provide analysts with a more reliable method to calculate the Malmquist Productivity Index.

A hyperbolic-oriented super efficiency measure can be developed by applying the hyperbolic-oriented function to evaluate the efficiency of an observation relative to a reference set that does not include the observation under evaluation. This paper will show that hyperbolic-oriented super efficiency models have advantages over standard super efficiency models for two reasons. First, a sufficient condition for feasibility is for all observation data to be positive. Second, while it is still possible to have infeasible solutions when zeros are allowed in the data domain, the conditions for infeasibility are more limited for the hyperbolic-oriented measure than for standard DEA-based super efficiency models.

The remainder of the paper is structured as follows: we first provide descriptions of the standard super efficiency models and the super efficiency hyperbolic-oriented model; next we show that feasible solutions are possible for a super efficiency hyperbolic-oriented model in cases when standard super efficiency models cannot provide feasible solutions; we then describe the use of the hyperbolic-oriented super efficiency measure in an outlier detection application; and we conclude with some general observations.

Standard super efficiency models and a super efficiency hyperbolic-oriented model

Super efficiency models measure the efficiency of an observation outside the PPS defined by the remaining observations. These models are a special case of DEA models. Typically input-oriented DEA efficiency estimates are in the range [0, 1] and output-oriented DEA efficiency estimates are in the range [1, ∞]. However, for super efficiency models, the range of [0, ∞] is possible for either orientation. The super efficiency model was first referenced in Banker *et al* (1989) as an outlier detection method developed in a separate working paper by the same lead author. The entire model later appeared in Anderson and Petersen (1993) as a method for developing a full ranking of observations.

Since then, the method has been used in a variety of situations. Charnes *et al* (1996) and Zhu (1996) used the method to study the sensitivity of the efficiency classification (see also Seiford and Zhu, 1998). Färe *et al* (1994) used these models to measure productivity and technology change. Thrall (1996) used them to identify extreme efficient observations, and Wilson (1995) and Johnson and McGinnis (2008) imbed the model in computational methodologies to find outliers.

It has been noted that under various conditions, the standard super efficiency model, taking an input or an output orientation relative to a variable returns to scale (VRS) frontier, may not be solvable and is said to have an infeasible solution. This has been noted by Thrall (1996), and Zhu (1996) elaborated by identifying certain zero patterns appearing in the data domain that cause infeasibility of the super efficiency model with or without the returns to scale assumption. Seiford and Zhu (1999) provided the most comprehensive discussion of this topic, defining exhaustively the conditions under which either an output- or an input-oriented super efficiency model would not have a feasible solution. In contrast to the work reported here, Zhu (1996) and Seiford and Zhu (1999) only studied the infeasibility of the super efficiency models, where they assumed that all input and output observation data are positively valued.

The notation used defines n observations and indexing the observations ($j = 1, 2, \ldots, n$), a vector of inputs, x_j, and a vector of outputs y_j, and let x_{kj} be the kth input in the set P of inputs and let y_{ij} be the ith output in the set Q of outputs. Under the assumption of VRS, two super efficiency DEA models can be expressed as shown in Figure 1.

Note that the super efficiency model differs from a standard VRS DEA model in that the observation under evaluation is excluded from the reference set.

Output-Oriented	Input-Oriented
$\min \Theta$	$\min \theta$
s.t. $\sum_{j=1, j\neq 0}^{n} \lambda_j x_j \leq x_0$	s.t. $\sum_{j=1, j\neq 0}^{n} \lambda_j x_j \leq \theta x_0$
$\sum_{j=1, j\neq 0}^{n} \lambda_j y_j \geq \Theta y_0$	$\sum_{j=1, j\neq 0}^{n} \lambda_j y_j \geq y_0$
$\sum_{j=1, j\neq 0}^{n} \lambda_j = 1$	$\sum_{j=1, j\neq 0}^{n} \lambda_j = 1$
$\Theta, \lambda_j \geq 0, j \neq 0$	$\theta, \lambda_j \geq 0, j \neq 0$

Figure 1 Standard output-oriented and input-oriented super efficiency models.

Using the same notation, a super efficiency hyperbolic-oriented model can be expressed as shown in Equation 1.1.

$$\begin{aligned} \min_{\theta, \lambda} \quad & \theta \\ \text{s.t.} \quad & \sum_{j=1, j\neq 0}^{n} \lambda_j x_j \leqslant \theta x_0 \quad \text{(Set of Input Constraints)} \\ & \sum_{j=1, j\neq 0}^{n} \lambda_j y_j \geqslant 1/\theta y_0 \quad \text{(Set of Output Constraints)} \\ & \sum_{j=1, j\neq 0}^{n} \lambda_j = 1 \quad \text{(Convexity Constraint)} \\ & \theta > 0, \ \lambda_j \geqslant 0, \ j \neq 0 \end{aligned} \tag{1.1}$$

Further, this model will be slightly adapted by substituting $z = 1/\theta$ in the set of output constraints and adding a single additional non-linear constraint, $z \geqslant 1/\theta$. This adaptation is in a similar spirit as the reformulation suggested by Färe *et al* (1985). The new formulation is:

$$\begin{aligned} \min_{\theta, \lambda, z} \quad & \theta \\ \text{s.t.} \quad & \sum_{j=1, j\neq 0}^{n} \lambda_j x_j \leqslant \theta x_0 \quad (1) \\ & \sum_{j=1, j\neq 0}^{n} \lambda_j y_j \geqslant z y_0 \quad (2) \\ & \sum_{j=1, j\neq 0}^{n} \lambda_j = 1 \quad (3) \\ & z \geqslant \frac{1}{\theta} \quad (4) \\ & \theta > 0, \ \lambda_j \geqslant 0, \ j \neq 0 \end{aligned} \tag{1.2}$$

Note the use of $z \geqslant 1/\theta$ is a relaxation of (1.1) and increases the size of the set of feasible decision variables. This slight modification has the benefit that the feasible set is the intersection of a finite number of convex sets and the condition

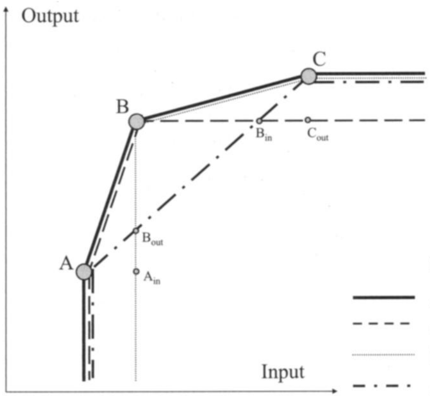

Figure 2 Super efficiency illustrated in two dimensions. (This figure is adapted from Seiford and Zhu, 1999).

$z \geqslant 1/\theta$. For values of z and θ greater than zero, $z \geqslant 1/\theta$ is also convex, thus the solution space for (1.2) is a convex set. Since the objective function is linear and the solution space convex, a local optimal solution will be globally optimal. However, because (1.2) is a relaxation, the equivalence of the optimal value of the two formulations (1.1) and (1.2) may not be obvious. This will be shown in Theorem 1.

Theorem 1 *The optimal solution to* (1.1), θ^*, *and* (1.2), $\bar{\theta}$, *are equal.*

Proof See Appendix. □

The infeasibility problem in two dimensions is illustrated in Figure 2. Here, for a standard DEA model, the efficient frontier is constructed using points A, B, and C. The frontier used to measure the super efficiency of A uses only B and C to construct the frontier. The super efficiency for B can be calculated by either orientation (for now we leave aside the DDF); however, the super efficiency can only be calculated from an input orientation for A and from an output orientation for C.

Feasibility of hyperbolic-oriented efficiency measure

In their Theorems 2 and 7, Seiford and Zhu (1999) recognized that infeasibility occurs when the input constraints in an output-oriented model or the output constraints in an input-oriented model are not satisfiable. Further, Zhu (1996) shows several results regarding infeasibility related to input- and output-oriented measures, but does not discuss the simultaneous expansion of inputs and contraction of outputs. We now show for the hyperbolic orientation that both sets of constraints are always satisfiable and thus infeasibility is not an issue when the hyperbolic orientation is used.

Theorem 2 *When the values for all inputs in set P and all outputs in set Q are positive for all observations, the hyperbolic-oriented super efficiency model under a VRS production frontier always has a feasible solution.*

Proof For any λ satisfying the convexity constraint, the input constraints can be satisfied by selecting $\theta \geqslant U(P)$ where

$$U(P) = \max_{k \in P, j=1\ldots n} \frac{x_{kj}}{x_{k0}} \tag{2.1}$$

Clearly, for any k, $U(P)x_{k0}$ is larger than x_{kj}, for all j, and thus larger than any convex combination of the x_{kj}. By a similar argument, for any λ satisfying the convexity constraint, the output constraints can be satisfied by selecting $\theta \geqslant 1/L(Q)$ where

$$L(Q) = \min_{i \in Q, j=1\ldots n} \frac{y_{ij}}{y_{i0}} \tag{2.2}$$

It is possible to find a θ satisfying both conditions simultaneously, from

$$\theta = \max(U(P), 1/L(Q)) \tag{2.3}$$

This means a feasible value of θ always exists. □

As long as all input and output values are positive, a hyperbolic-oriented super efficiency model always has a feasible solution. Even if this assumption is relaxed, the hyperbolic-oriented super efficiency model may have a feasible solution.

Relaxing the assumption that all input data are positive, allowing zero values for some inputs. Theorem 3 states the conditions under which hyperbolic-oriented super efficiency model has a feasible solution.

Theorem 3 *Assume that all output values are positive. Let* $P' \subseteq P$, $\ni \forall k \in P'$, $x_{k0} = 0$. A necessary and sufficient condition for the hyperbolic-oriented super efficiency model to have a feasible solution is that there is at least one observation, say s, in the reference set, such that $x_{ks} = 0$, $\forall k \in P'$.

Proof *Sufficiency*: A feasible solution can be found by setting $\lambda_s = 1$ to satisfy the input constraints corresponding to P'; and applying the proof of Theorem 1 to determine a sufficiently large θ.
Necessity: Suppose there is a feasible solution and $\nexists x_s \ni x_{ks} = 0\ \forall k \in P'$, that is, $\forall x_s$, $\exists k \in P' \ni x_{ks} > 0$. In order to satisfy the kth input constraint, we must have $\lambda_s = 0$. However, if $\lambda_s = 0$ for all s, then $\sum_j \lambda_j = 1$ cannot be satisfied. Thus the initial assumption is contradicted, and if there is a feasible solution, there must be at least one $x_s \ni x_{ks} = 0\ \forall k \in P'$. □

Further, now consider the case when all inputs are positive but outputs are allowed to take on zero values. Theorem 4 states the conditions under which the hyperbolic-oriented super efficiency model has a feasible solution.

Theorem 4 *Assume that all input values are positive. Let* $Q' \subseteq Q$, $\ni \forall i \in Q'$, $y_{i0} \neq 0$. *A necessary and sufficient condition for the hyperbolic-oriented super efficiency model to have a feasible solution is that for each non-zero output* i *there is at least one observation, say* s*, in the reference set, such that* $y_{is} \neq 0$ *indicating the* i*th output of observation* s *is not zero.*

Proof *Sufficiency*: A feasible solution can be found by setting $\lambda_{s_i} = 1/t$ for all s_i associated with each t, where t is the number of outputs for which $y_{j0} \neq 0$, to satisfy the output constraints corresponding to Q', and applying the proof of Theorem 1 to determine a sufficiently large θ.
Necessity: Suppose there is a feasible solution and $\nexists y_{s_j} \ni y_{js_j} \neq 0$, that is, $\forall y_s$, $y_{is_i} = 0$. In order to satisfy the ith output constraint, we must have $\lambda_s = 0$. However, if $\lambda_s = 0$ for all s, then $\sum_j \lambda_j = 1$ cannot be satisfied. Thus, the initial assumption is contradicted, and if there is a feasible solution, for every $i \in Q'$ there must be at least one $y_{s_i} \ni y_{is_i} > 0$. □

It should be noted that a common definition of *peer group*, that is, that all inputs are used by all members of the group, leads to the scenario of Theorem 4 where inputs are all positive; however, some outputs can take on zero output values. This was recently observed in an online analysis of warehouses called iDEAs described in Johnson *et al* (2009). In this case, all warehouses used labor, space, and capital, but different warehouses produced different mixes of outputs.

Theorem 3 addresses the case when the positivity restriction on input data is relaxed, and Theorem 4 addresses the case when the positivity restriction on the output data is relaxed. When the positivity restriction is relaxed on both the input data and the output data, the hyperbolic-oriented super efficiency model still can be solved except for two cases: (1) $y_i > 0$ for only one observation in the data set; or (2) there is an observation with a unique input usage set (the set of x for which $x_k > 0$).

Note relative efficiency assessment requires the careful matching of input–output models and a set of data to be analyzed. There are several common problems in relative efficiency analysis, such as lack of data to accurately estimate a nonparametric production frontier, improper specification of the input–output model, and heterogeneous observations in the data set. The infeasibility of the hyperbolic-oriented super efficiency model may be an indication that one of these common problems is of concern for a given model and data set. For example if a single observation produces a unique output that none of the other observations produce this may be an indication of any of the common problems identified. Further if an observation uses a unique input set no other observation uses, this may indicate a lack of data in a particular orthant or it may indicate a heterogeneous observation.

The directional-oriented efficiency measure

The hyperbolic-oriented efficiency measure is not the only efficiency measure that simultaneously contracts inputs and expands outputs. An alternative efficiency measure not traditionally used in outlier detection or super efficiency models

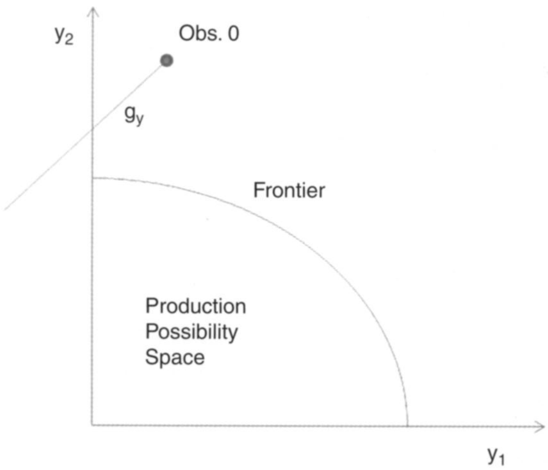

Figure 3 Infeasibility of the directional distance function. *Source*: This figure is adapted from Briec and Kerstens (2009).

is the directional-oriented efficiency measures based on the DDF. The directional-oriented efficiency measures also suffer from the infeasibility problem. Briec and Kerstens (2009) have investigated the infeasibility problem in the general DDF approach. Notably, an infeasible direction is shown to always exist when the number of output dimensions is greater than or equal to 2 and the output direction vector is non-zero, meaning all components of the g_y are non-zero. Further Figure 3 in Briec and Kerstens demonstrates that even if all data are positive and all components of the g vector are positive, it still is possible to have infeasibility.

We specify the super efficiency DDF as the LP:

$$\begin{aligned}
&\max_{\theta,\lambda_j} \quad \theta \\
&\text{s.t.} \quad y_{i0} + \theta g_{i0} \leqslant \sum_{\substack{j=1\\ j\neq 0}}^{N} \lambda_j y_{ij} \qquad \forall i = 1, \ldots, Q \\
&\qquad x_{k0} - \theta g_{k0} \geqslant \sum_{\substack{j=1\\ j\neq 0}}^{N} \lambda_j x_{kj} \qquad \forall k = 1, \ldots, P \\
&\qquad \sum_{\substack{j=1\\ j\neq 0}}^{N} \lambda_j = 1 \\
&\qquad \lambda_j \geqslant 0 \qquad \forall j = 1, \ldots, N.
\end{aligned} \tag{2.4}$$

However, recognizing that the DDF is a generalization of the output-oriented distance function, if the directions, g_x and g_y are selected without consideration for the specific data set under analysis, the DDF model may not identify a benchmark in the PPS. This can be formalized in a theorem.

Theorem 5 *When measuring the super efficiency of an observation outside of the PPS, if g_{kj} and g_{ij} are selected such that $g_{i0} > -y_{i0}/[\min_k(x_{k0} - \min_j x_{kj}/g_{k0})]$ for at least one $i \in 1, \ldots, Q$ then the super efficiency DDF will identify a benchmark not in the PPS.*

Proof See Appendix. □

Identifying a benchmark outside of the PPS does not cause infeasibility for the DDF because of its additive formulation. We can see that Theorem 5 is closely tied to the infeasibility of the output-oriented efficiency measure. Namely if g_{kj} is selected to be too large relative to g_{ij} then the projection ray for observation j will intersect the $y_i = 0$ hyperplane before intersecting the production frontier. In the output-oriented model, this is the infeasibility problem; however, in the additive formulation it is the inability to identify a benchmark in the PPS. In addition the DDF can suffer from the problem of identifying benchmarks outside the PPS when all data are positive and the directional vectors are positive.

Theorems 3, 4, and 5 shows that the hyperbolic-oriented super efficiency measure can be calculated in some cases where the input, output, or directional-oriented super efficiency measure cannot. The hyperbolic distance function avoids the infeasibility with an exception for the case that $y_{ij} > 0$ for only one observation in the data set or there is an observation with a unique input usage set (the set of components of the x vector for which $x_k > 0$). Further it does not require the determination of directional vectors dependent on the data set to ensure an appropriate benchmark.

Application

In this section, we demonstrate the use of the hyperbolic-oriented measure in super efficiency models used for outlier detection. The results are compared to the input- and the output-oriented super efficiency models. It will be shown that the hyperbolic-oriented efficiency measure always has a feasible solution, while both the input- and the output-oriented measures have infeasible results for at least one observation in the data set.

The classic Banker and Morey (1986) data set for pharmacies in the state of Iowa is used. This data set consists of 69 observations, each with two outputs (number of prescriptions and dollar levels of sales for prescriptions) and three inputs (labor costs, other operating costs, and average value of inventory). For more information about the data set, see Banker and Morey (1986).

Automated outlier detection methods flag unusual observations for further inspection. When super efficiency models are used for outlier detection, a critical value must be specified. For a further discussion of how to specify this value, see Johnson and McGinnis (2008). Using an input-oriented model (output-oriented model), observations with super efficiency estimates higher (lower) than the critical level are flagged as possible outliers. Simar (2003) suggests that an observation should require both a significant expansion of inputs and a significant contraction of outputs to be flagged as a possible

Table 1 The input-, output-, and hyperbolic-oriented super efficiency estimates for a subset of the 69 pharmacies.

Observation	*Orientation*		
	Input	*Output*	*Hyperbolic*
1	1.02	0.97	1.01
2	0.86	1.20	0.92
3	0.72	1.49	0.83
4	1.19	0.72	1.12
5	1.56	***	1.36
6	1.25	***	1.37
7	1.19	***	1.14
8	1.10	***	1.48
⋮			
17	***	0.81	1.23
⋮			
68	0.55	1.86	0.73
69	2.56	***	3.89

*** indicates the model was infeasible.

outlier. Using traditional orientations, this suggestion makes it necessary to calculate both an input-oriented and an output-oriented super efficiency estimate to make this determination.

The super efficiency estimates are calculated for each observation in the Banker-Morey data set using the input, output, and hyperbolic orientations. In the interest of space, only a subset of the results is shown in Table 1.

Note that for five observations, the output-oriented model was infeasible, and for one observation, the input-oriented model was infeasible. This indicates that for six observations, there were insufficient data to determine if these observations should be flagged as possible outliers using Simar's criteria. However, for all observations, the hyperbolic distance function gives information about the distance from the frontier. Automated outlier detection techniques allow researchers to explore data sets and identify suspicious observations quickly. When infeasibility arises, it becomes necessary to inspect not only the observations flagged as possible outliers but also the observations with infeasible super efficiency values. In this example, if the critical value was 1.1, nine observations would have been flagged as possible outliers. Six additional observations would have to be inspected because infeasible results were returned. This is a 67% increase in the inspection process. In this data set, the use of the input and output orientations would significantly increase the work required in the inspection process, and for larger data sets, this increase in inspection could be prohibitive.

Conclusion

We studied the use of the hyperbolic-oriented super efficiency measure and its benefits relative to the more traditional input-, output-, or directional-oriented super efficiency measures. The hyperbolic-oriented efficiency measure has been slow to gain popularity, in part because of its increased computational burden. It requires solving a non-linear program rather than a linear program.

Super efficiency has been used for sensitivity analysis, productivity and technology change (such as the Malmquist Productivity Index), and outlier detection. The benefits of using the hyperbolic-oriented super efficiency measure could be realized for each of these applications. Our results indicate it is still possible to have infeasible hyperbolic-oriented measures when the observation under consideration has values of zero for a set of inputs for which no observation in the reference set has zeros for the same set inputs or in the case that $y_{ij} > 0$ for only one observation in the data set. However, these are rather weak conditions, and if these are not satisfied, the analyst may question if the necessary assumptions for relative performance assessment are met. The hyperbolic-oriented efficiency measure has two benefits: it can be calculated for cases when the input-, output-, and directional-oriented measures are not feasible, and it allows the comparison of a broader group of observations by allowing zeros as input and output values in some cases. These results make the hyperbolic-oriented efficiency measure a desirable option.

Acknowledgements—The authors acknowledge the support for this work by the National Science Foundation under grant 0400187, and the Book Industry Study Group (BISG), and the contributions of the anonymous referees in improving the paper.

References

Anderson P and Petersen NC (1993). A procedure for ranking efficient units in data envelopment analysis. *Mngt Sci* **39**: 1261–1264.

Banker RD, Das S and Datar SM (1989). Analysis of cost variances for management control in hospitals. *Research in Governmental and Nonprofit Accounting* 5. Greenwich, JAI Press, pp. 269–291.

Banker RD and Morey RC (1986). The use of categorical variables in data envelopment analysis. *Mngt Sci* **32**: 1613–1627.

Briec W and Kerstens K (2009). Infeasibility and directional distance functions with application to the determinateness of the Luenberger Productivity Indicator. *J Optimiz Theory App* **141**: 55–73.

Chambers RG, Chang Y and Färe R (1998). Profit, directional distance functions, and Nerlovian efficiency. *J Optimiz Theory App* **98**: 351–364.

Charnes A, Cooper WW and Rhodes E (1978). Measuring the efficiency of decision making units. *Eur J Opl Res* **2**: 429–444.

Charnes A, Rousseau JJ and Semple JH (1996). Sensitivity and stability of efficiency classifications in data envelopment analysis. *J Prod Anal* **7**: 5–18.

Chavas J-P and Cox TL (1999). A generalized distance function and the analysis of production efficiency. *Southern Econ J* **66**: 294–318.

Cuesta RA and Zofío JL (2005). Hyperbolic efficiency and parametric distance functions: With application to Spanish savings banks. *J Prod Anal* **24**: 31–48.

Färe R and Primont D (1995). *Multioutput Production and Duality: Theory and Application*. Kluwer Academic Publishers: Boston.

Färe R, Grosskopf S and Lovell CAK (1985). *The Measurement of Efficiency of Production*. Kluwer-Nijhoff Pub: Boston, MA, USA, Distributors for North America, Kluwer Academic Publishers.

Färe R, Grosskopf S and Lovell CAK (1994). *Production Frontiers*. Cambridge University Press: Cambridge, UK; New York, NY, USA.

Färe R, Grosskopf S and Zaim O (2002). Hyperbolic efficiency and return to the dollar. *Eur J Opl Res* **136**: 671–679.

Farrell MJ (1957). The measurement of productive efficiency. *J R Stat Soc Ser A—G* **120**: 253–281.

Johnson AL and McGinnis LF (2008). Outlier detection in two-stage semiparametric DEA models. *Eur J Opl Res* **187**: 629–635.

Johnson AL, Chen W-C and McGinnis LF (2009). Internet-based benchmarking for warehouse operations, under review at *Computers in Industry*.

Koopmans TC (ed.) (1951). An analysis of production as an efficient combination of activities. In: *Activity Analysis of Production and Allocation*. Wiley: New York, Monograph No. 13.

Seiford LM and Zhu J (1998). Sensitivity analysis of DEA models for simultaneous changes in all the data. *J Opl Res Soc* **49**: 1060–1071.

Seiford LM and Zhu J (1999). Infeasibility of super-efficiency data envelopment analysis models. *Infor* **37**: 174–187.

Shephard RW (1953). *Cost and Production Functions*. Princeton University Press: New Jersey.

Shephard RW (1970). *Theory of Cost and Production Functions*. Princeton University Press: Princeton, NJ.

Simar L (2003). Detecting outliers in frontier models: A simple approach. *J Prod Anal* **20**: 391–424.

Thrall RM (1996). Duality, classification and slacks in DEA. *Ann Oper Res* **66**: 109–138.

Wilson PW (1995). Detecting influential observations in data envelopment analysis. *J Prod Anal* **6**: 27–45.

Zhu J (1996). Robustness of the efficient DMUs in data envelopment analysis. *Eur J Opl Res* **90**: 451–460.

Zofío JL and Lovell CAK (2001). Graph efficiency and productivity measures: an application to US agriculture. *Appl Econ* **33**: 1433–1442.

Appendix

Proof of Theorem 1

The optimal solution to (1.1) notated (θ^*, λ^*) along with $z^* = 1/\theta^*$ is a feasible solution to (1.2) because (1.2) is a relaxation of (1.1), thus $\theta^* \geqslant \bar{\theta}$. Consider the optimal solution to (1.2) notated $(\bar{\theta}, \bar{\lambda}, \bar{z})$, taking only $(\bar{\theta}, \bar{\lambda})$ is a feasible solution to (1.1). This can be seen by noting the input constraints, convexity constraint, and restrictions on the variables are identical thus any solution to (1.2) should also satisfy these constraints for (1.1). Thus it is only necessary to show that the output constraint in (1.1) is not violated. Recognizing y_0 is non-negative and $\bar{z}$ must be positive because $\bar{\theta} > 0$, then constraint (2) and (4) of (1.2) can be combined, $\sum_{\substack{j=1 \\ j\neq 0}}^{n} \bar{\lambda}_j y_j \geqslant \bar{z} y_0 \geqslant y_0(1/\bar{\theta})$. Regardless of the value of $\bar{z}$, the $\bar{\theta}$ value must satisfy the output constraint of (1.1) indicating the optimal solution to (1.2) is a feasible solution to (1.1) and hence $\theta^* \leqslant \bar{\theta}$. Finally combining $\theta^* \leqslant \theta$ and $\theta^* \geqslant \theta$ we have $\theta^* = \theta$. □

Proof of Theorem 5

The input constraints are all satisfied when $\theta_0 \leqslant \frac{x_{k0} - \sum_{j=1, j\neq 0}^{n} \lambda_j x_{kj}}{g_{k0}}$, $k=1, \ldots, P$. This implies, with an unknown data distribution, this condition could not possibly hold unless contracting input level x_{k0} made it larger than some observation in the data set. Referencing the smallest input level for each input $\theta_0 \leqslant \frac{x_{k0} - \min_j x_{kj}}{g_{k0}}$ and all input constraints must be satisfied, thus the constraint can be rewritten as $\theta_0 = \min_k \left(\frac{x_{k0} - \min_j x_{kj}}{g_{k0}}\right)$. Substituting this equation into the left hand-side of the output constraint we have $y_{i0} + \left[\min_k \left(\frac{x_{k0} - \min_j x_{kj}}{g_{k0}}\right)\right] g_{i0} \geqslant 0$, $\forall i = 1, \ldots, Q$. Solving for g_{i0}, recognizing if this condition does not hold, we have a nonsensical benchmark, we see the result.

Received August 2007;
accepted May 2009 after five revisions