European Journal of Operational Research 206 (2010) 456–469

Contents lists available at ScienceDirect

European Journal of Operational Research

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

Computational Intelligence and Information Management

Globally flexible functional forms: The neural distance function ☆

Panayotis G. Michaelides [a,*], Angelos T. Vouldis [b,c,1], Efthymios G. Tsionas [d]

[a] Laboratory of Theoretical and Applied Economics, Department of Humanities, Social Sciences and Law, School of Applied Mathematics and Physics, National Technical University of Athens, Zografou Campus, 157.80 Athens, Greece
[b] UADPhilecon, University of Athens, Greece
[c] Financial Stability Department, Bank of Greece, 3 Amerikis Street, 105.64, Athens, Greece
[d] Department of Economics, Athens University of Economics and Business, 76 Patission Street, 104.34 Athens, Greece

ARTICLE INFO

Article history:
Received 30 September 2008
Accepted 10 February 2010
Available online 14 February 2010

Keywords:
Output distance function
Translog
Technical efficiency
ANN
LIML

ABSTRACT

The output distance function is a key concept in economics. However, its empirical estimation often violates properties dictated by neoclassical production theory. In this paper, we introduce the neural distance function (NDF) which constitutes a global approximation to any arbitrary production technology with multiple outputs given by a neural network (NN) specification. The NDF imposes all theoretical properties such as monotonicity, curvature and homogeneity, for all economically admissible values of outputs and inputs. Fitted to a large data set for all US commercial banks (1989–2000), the NDF explains a very high proportion of the variance of output while keeping the number of parameters to a minimum and satisfying the relevant theoretical properties. All measures such as total factor productivity (TFP) and technical efficiency (TE) are computed routinely. Next, the NDF is compared with the Translog popular specification and is found to provide very satisfactory results as it possesses the properties thought as desirable in neoclassical production theory in a way not matched by its competing specification.

1. Introduction

Productivity and efficiency in converting inputs into goods and services have typically been key issues in economics and management. In fact, productivity and efficiency analyses have traditionally been indispensable tools for evaluating firms' performance. In this context, reliable measures of productivity and efficiency are of great interest because they can assist in addressing important issues. For this reason, the measurement of productivity and efficiency is a very important task for engineers, managers and decision makers.

Evaluating technical efficiency may imply production (or cost) function estimation, as a preliminary step. The analysis of production has always been a central problem in economics. Despite the fact that the production function is a purely technical relation whereas the cost function embodies an economic relation, there is nevertheless a very close relationship between them in the sense that "the cost function contains essentially the same information that the production function contains" (Varian, 1992, p. 81). This is referred to as the *duality* of production and cost. It means that, given values for the parameters of the cost function, we may obtain the parameters of the production function; economists have established this correspondence for any production technology (Stewart, 2005, p. 479, 491).

In other words, the problem of cost minimization is *dual* to the primal problem of profit maximization. This *duality* was first studied systematically by Shephard (1953) and Diewert (1971) and it has since been found that it has much broader applicability and it is extremely important for the empirical implementation of the theory (see, for instance, Yair, 1996). In this context, we know that any monotonic, homogeneous, concave function of prices is a cost function for some well-behaved technology. Hence, it is necessary to find a functional form with the required properties. In general, a parametric form has to be chosen (Stewart, 2005). This process, which is – in general – not an easy task, often constitutes the first step towards technical efficiency evaluation.

On the other hand, *output distance functions* provide a functional characterization of the production technology with respect to output sets and, in this context, they provide *directly* measures of efficiency and productivity. Many functional forms have been proposed for their approximation but perhaps the most important in current use are the Cobb-Douglas specification introduced by Cobb and Douglas (1928), and the Translog flexible functional form introduced by Christensen et al. (1971). Both of theses models, that are intuitively appealing and computationally straightforward,

☆ We would like to thank the Editors and three anonymous Referees for their constructive comments.

* Corresponding author. Tel.: +302 107721624.
E-mail address: pmichael@central.ntua.gr (P.G. Michaelides).

[1] The views expressed in this article do not necessarily represent the views of the Bank of Greece.

0377-2217/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.ejor.2010.02.013

have been estimated and have, in addition, been used to test various restrictions of production theory. However, they both have certain drawbacks. For instance, the Cobb-Douglas specification is often less than satisfactory because it attempts to explain the complex variation in data with a quite simple mathematical function despite the fact the real-world data are much more complicated.

Furthermore, and even more importantly, a second problem is common in the empirical estimation of the Cobb–Douglas, the Translog specification and other flexible functional forms, namely the violation of properties dictated by neoclassical economic theory, such as monotonicity, curvature, and homogeneity conditions. Some researchers have developed numerical (Gallant and Golub, 1984) and Bayesian techniques (O'Donnell and Coelli, 2005) for imposing curvature conditions that tend to provide encouraging results. However, in production theory it is absolutely necessary to estimate functional forms that satisfy *globally* the curvature conditions dictated by economic theory. This has been "one of the most vexing problems applied economists have encountered in estimating flexible functional forms" (Diewert and Wales, 1987) and remains "one of the most difficult challenges faced by empirical economists" (Terrell, 1996). After all "Ultimately, the biggest challenge for researchers remains the issue of the appropriate production function specification to represent the underlying process technology" (Vaneman and Triantis, 2007).

So far, artificial neural networks (ANNs) have found relatively limited applications in efficiency studies mainly because they are a-theoretical, not easily interpretable, complex, and demanding with respect to technical resources. Of course, one has to balance between the costs of using them and the apparent benefits, such as increased flexibility, better approximations and predictions, more accurate decision making, etc. (Santin et al., 2004). The non-parametric feature of ANNs makes them quite flexible and attractive in production theory where the theoretical relationship is not known *a priori* (Zhang and Berardi, 2001). In this context, recently Santin (2008) showed that ANNs outperform traditional approaches in detecting the underlying shape of the production function from observed data. To demonstrate this, a Monte-Carlo experiment was carried out on a simulated smooth production technology. However, production functions are inappropriate with technologies producing multiple outputs. In such cases, we have to use distance functions which are appropriate for modelling multi-input multi-output production technologies without aggregating outputs.

In this paper, we propose and estimate a new model which has considerable advantages. Our model, which we call the neural distance function (NDF) and is based on ANNs, has the following advantages when compared to the widely adopted specifications of the production technology such as the Translog, etc.: (a) It gives an approximation to any arbitrary production process; (b) it is flexible with respect to time; (c) it allows for arbitrary returns to scale; (d) it is simple to estimate; (e) it avoids the need for nonlinear estimation; (f) it uses fewer estimated parameters than other globally flexible functional forms (parsimonious model); (g) it provides a very good fit to real-world data; (h) it has a functional form which is consistent with neoclassical production theory data; (i) it is based on a system of equations which deals with the endogeneity issue in estimation, and (j) it satisfies the properties dictated by production theory *globally* and not only over the set of inputs and outputs where inferences are drawn. Although some of these desirable properties are also possessed by one or other of the known specifications, neither possesses all of them simultaneously. In this sense, the NDF is superior to them.

Contrary to widely used local approximations like the Translog (Christensen et al., 1971), the generalized Leontief (Diewert, 1971) or the symmetric McFadden form (Diewert and Wales, 1987), the NDF is a global approximation to the unknown function. The Fourier flexible form (Gallant, 1982) is also a global approximation but it requires an excessive number of parameters. The NDF provides a very good approximation using fewer estimated parameters than other globally flexible functional forms. Moreover, we extend the analysis to the case of technologies with multiple outputs and introduce the method of limited information maximum likelihood (LIML) in this context to solve the econometric problems associated with the estimation of such distance functions. All relevant measures such as returns to scale (RTS) and total factor productivity (TFP) may be computed routinely.

Our estimation methodology is formulated as a four-step algorithm which employs the seemingly unrelated regression (SUR) equations technique for estimating the coefficients of a system of linear equations and an iterative optimization algorithm for the nonlinear parameters of the NDF. Also, it should be made clear that in the context of approximating a distance function, the neural network (NN) specification does not aim at providing forecasting or pattern recognition capabilities as is usually the case, for instance, in engineering applications. Instead, the nonlinear nature of ANNs specification is used to obtain a sufficiently flexible functional form which is capable of approximating an existing distance function while enabling the *a priori* imposition of the properties (i.e. by construction) dictated by neoclassical economic theory. After all "In reality, organizational and/or system efficiency performance is dynamic, nonlinear in its relationship to key production drivers, and a function of multiple/complex interactions" (Vaneman and Triantis, 2007).

An application investigating the model's performance illustrates our technique. Fitted to a very large number of samples based on a large data set on all US commercial banks in the 1989–2000 time span, the NDF explains a very high proportion of the variance of output while keeping the number of estimated parameters to a minimum, and satisfies theoretical properties dictated by production theory. However, in order to assess whether the NDF provides satisfactory results, we refer to the widely used Translog specification that is based on a conventional parametric approach. Hence, we compare the Translog with the NDF specification and NDF is found to provide a very good approximation by using fewer estimated parameters, while satisfying the theoretical properties dictated by production theory, globally. It should be noted that, so far, no empirical study has imposed all these theoretical properties dictated by neoclassical production theory on parametric distance functions (O'Donnell and Coelli, 2005). Also, very few studies report the degree to which the estimated functions satisfy these conditions (Reinhard and Thijssen, 1998). We believe that the results of this study suggest that the NDF is an appropriate vehicle for testing, expanding and improving conventional production theory.

The paper is organised as follows: Section 2 introduces the NDF. In Section 3 the theoretical properties of the NDF are examined. Section 4 provides the framework for measuring productive efficiency and productivity change. The procedure for the empirical estimation of the NDF is set out in Section 5, while Section 6 presents the empirical results and compares them with the Translog specification. Finally, Section 7 concludes.

2. Output distance functions with neural networks

2.1. Elements of neural networks

Instead of fitting the data with a pre-specified model, artificial neural networks (ANNs) let the data set itself serve as evidence to support the model's approximation of the underlying production technology. ANNs have found applications in financial modelling (e.g. Hutchinson et al., 1994; Anders et al., 1998; Hanke, 1999; Tsitsiklis and Roy, 2001), in technical efficiency estimation (e.g.

Athanassopoulos and Curram, 1996; Costa and Markellos, 1997; Fleissig et al., 2000) and in economic modelling (e.g. Sfetsos and Siriopoulos, 2005; White and Racine, 2001; Hansen and Nelson, 2003) without making use of an output distance function satisfying the theoretical properties dictated by production theory.

Formally, neural networks (NNs) are data-driven and self-adaptive, nonlinear methods that do not require specific assumptions about the underlying model (Zhang and Berardi, 2001). In mathematical terms, ANNs are collections of transfer functions that relate an output variable Y to certain input variables $X' = [X_1, \dots, X_n]$. The input variables are combined linearly to form m intermediate variables $Z_1, \dots, Z_m$ where

$$Z_k = X'\beta_k \quad (k = 1, \dots, m) \tag{1}$$

and $\beta_k \in R^n$ are parameter vectors. The intermediate variables are combined nonlinearly to produce Y:

$$Y = \sum_{k=1}^{m} \alpha_k \phi(Z_k) \tag{2}$$

where ϕ is an activation function, the α_k 's are parameters and m is the number of intermediate nodes (Kuan and White, 1994).

By combining simple units with intermediate nodes, the NN can approximate any smooth nonlinearity (Chan and Genovese, 2001). As demonstrated in Hornik et al. (1989, 1990), ANNs provide good approximations to a large class of arbitrary functions while keeping the number of parameters to a minimum. In other words, they are universal approximators of functions. Also, they can approximate their derivatives, a fact which justifies their success in empirical applications (Hornik et al., 1990).

2.2. Output distance functions

Let $x \in \mathbb{R}^N_+$ denote an input vector corresponding to N factors of production, and $Y \in \mathbb{R}^J_+$ the output vector when J outputs are produced. The production technology can be described by the production set i.e. the set of feasible input–output vectors defined as

$$P = \{(Y, x) : \ x \text{ can produce } Y\}.$$

The output sets of such a production technology are defined as the sets of output vectors which can be produced for a given input vector: $L(x) = \{Y : (Y, x) \in P\}$. Output distance functions are introduced in order to measure the distance of a production process from the production frontier.

More specifically, an output distance function can be defined as $\Delta(x, Y) = \min\{\mu : Y/\mu \in L(x)\}$. It measures how close a particular level of output is to the maximum attainable level of output that could be obtained from the same level of inputs. In other words, it represents how close a particular output vector is to the production frontier given a particular input vector. The output distance function takes a value of unity if Y is located on the production frontier (i.e. the maximum attainable output) for the specific input vector x. This implies $\Delta(x, Y) \leqslant 1$ because of the presence of inefficiency. In other words, inefficiency leads to a discrepancy between the actual output and the production frontier. We adopt a setup consistent with revenue maximization so that production technology can be described by a distance function of the form $\Delta(x, Y) = 1 - \varepsilon$ where the Y s are endogenous, the xs are predetermined and ε is a non-negative stochastic term representing inefficiency (Kumbhakar and Lovell, 2000).

Applied research usually focuses on the case with multiple inputs and a single output. While this might be a realistic approach for a unit that produces only one output or service, a more flexible model is required when the unit produces more than one type of product or service (Vaneman and Triantis, 2007). In this context, the proposed output distance function allows for multi-output approaches without having to aggregate outputs.

In general, econometric estimation of distance functions is a relatively complicated issue; see, for example, the excellent discussion in Kumbhakar and Lovell (2000). In brief, the estimation can be carried out based on two major alternatives: parametric and non-parametric. In parametric methods, a functional form is adopted and the relevant techniques can be deterministic or stochastic. In non-parametric methods, no functional form is assumed and usually a deterministic frontier is formed.

In this context, one commonly used technique for estimating distance functions has been the non-parametric technique of data envelopment analysis (DEA) that has found countless applications. DEA uses linear programming techniques to estimate a piece-wise frontier that envelops the observations and requires no specific functional form for the production function. Despite some advantages, a very serious drawback of this approach is rooted in the fact that DEA does not allow for measurement errors and cannot discriminate between inefficiency and noise, and tends to produce overestimated inefficiency measures (O'Donnell and Coelli, 2005). Of course, some researchers have employed bootstrap techniques to introduce a stochastic element into linear programming techniques like DEA (see, among others, Efron, 1982; Efron and Tibshirani, 1993; Ferrier and Hirschberg, 1997; Simar and Wilson, 1998; Lothgren, 1998). However, this approach suffers from several problems (e.g. Simar and Willson, 1999). For an excellent review of the major methodological developments in DEA since the appearance of the seminal work of Charnes et al. (1978), see Cook and Seiford (2009).

One can attempt to overcome these problems by using a parametric technique, i.e. by specifying a functional form for the production surface. This approach estimates a frontier where all deviations from the frontier are due to inefficiency, which leads to biased estimates of the shape and position of the frontier. The stochastic Frontier analysis (SFA) can also be used for distance function estimation. The SFA requires a functional form in order to estimate the frontier and is based on the idea that the data are contaminated with measurement errors and other noise.

In general, various specialized computational and estimation techniques have been developed, and numerous applications have been described in the literature, while linkages to other fields, especially statistics and econometrics, have been sought allowing researchers to measure the distance that each firm has below the frontier (see, among others, Färe et al., 1993; Grosskopf et al., 1995; Coelli and Perelman, 1996; Coelli and Perelman, 1999, 2000; Tsionas, 2000; Fuentes et al., 2001; Brummer et al., 2002; Tsionas, 2002; Tsionas, 2003; Tsionas and Kumbhakar, 2004; Kumbhakar and Tsionas, 2005; Kumbhakar and Tsionas, 2006).

2.3. NDF specification

In this paper, we express any given output (say Y_J) as a function of the others and obtain: $\ln Y_J = f(\ln x, \ln Y_1, \dots, \ln Y_{J-1})$. One can estimate this equation by various techniques, e.g. Ordinary Least Squares (OLS) but the endogeneity issue is not taken into account (Kumbhakar and Lovell, 2000). In order to account for endogeneity, we must consider the reduced form $\ln Y_{-J} = g(\ln x)$, where $\ln Y_{-J} = [\ln Y_1, \dots, \ln Y_{J-1}]'$, and g is a vector function $g : \mathbb{R}^N \to \mathbb{R}^{J-1}$. The reduced form expresses all other outputs as functions of the inputs alone. In technical terms, we have a system of equations as follows:

$$\ln Y_J = f(\ln x, \ln Y_1, \dots, \ln Y_{J-1}) + e_J \tag{3}$$

$$\ln Y_{-J} = g(\ln x) + e_{-J} \tag{4}$$

where $e_{-J} = [e_1, \dots, e_{J-1}]'$, and $e = [e'_{-J}, e_J]'$ represents a J-dimensional random vector. The f function could be specified as a Translog. In many cases, g can be assumed to be a linear function. But these are arbitrary choices and they are not expected to approximate the f and g specifications with any reasonable accuracy. Of course, whether or not such approximations are satisfactory in any specific application is an empirical matter.

At any rate, if the system of Eqs. (3) and (4) is estimated using the seemingly unrelated regression (SUR) technique, then the resulting estimates is known as limited information maximum likelihood (LIML). The crucial part is, however, to specify the g and f functions using a globally flexible functional form. After all "Ultimately, the biggest challenge for researchers remains the issue of the appropriate ... specification to represent the underlying process technology" (Hoopes and Triantis, 2001).

The proposed neural reduced form function, for each output, is given by:

$$\ln Y_j(x) = a_{0j} + \sum_{k=1}^{m_j} a_{kj}\phi_j(\ln x \cdot \beta_{kj}) + \ln x \cdot \theta_j + \delta_j t \quad j = 1, \ldots, J-1 \tag{5}$$

where $Y_j(x)$ is the reduced form function of j-th output, m_j is the number of intermediate nodes, t is a time index and $\delta_j,\ a_{kj} \in \mathbb{R},\ \beta_{kj} \in \mathbb{R}^N,\ \theta_j \in \mathbb{R}^N$ are parameters. In general, for vectors a and b, $a \cdot b$ denotes the inner product.

Thus, given Eq. (5) the output distance function can be written as (Kumbhakar and Lovell, 2000):

$$\ln D = a_{0J} + \sum_{k=1}^{m_J} a_{kJ}\varphi_J(\ln x \cdot \beta_{kJ}) + \ln Y \cdot \gamma + \ln x \cdot \xi + \delta_J t \tag{6}$$

where $\delta_J, a_{kJ} \in \mathbb{R},\ \beta_{kJ} \in \mathbb{R}^N,\ \gamma \in \mathbb{R}^{J-1}$ are parameters, m_J is the number of intermediate nodes for output J and t is a time index. Eq. (6), which represents our proposed specification for the distance function, and Eq. (5) are reduced forms.

The next step is to convert Eq. (6) into an estimable model and this can be accomplished by exploiting the property that output distance functions are homogenous of degree one in outputs. It can easily be demonstrated that imposing this constraint is equivalent to normalizing the output distance function by one of the outputs (Kumbhakar and Lovell, 2000). Thus, an alternative form of (6), imposing homogeneity of degree one in inputs, is:

$$\begin{aligned}-\ln Y_J(x) = a_{0J} &+ \sum_{k=1}^{m_J} a_{kJ}\varphi(\ln x \cdot \beta_{kJ}) + \ln\left(\frac{Y_{-J}}{Y_J}\right) \cdot \gamma_{-J} \\ &+ \ln x \cdot \xi + \delta_J t + u\end{aligned} \tag{7}$$

where $u = -\ln D$ is a non-negative term such that $0 < D \leqslant 1,\ -\infty < \ln D \leqslant 0$ that captures the effects of inefficiency (Kumbhakar and Lovell, 2000). If we follow the stochastic frontier approach proposed by Aigner et al. (1977) and Meeusen and van den Broeck (1977) we should add a symmetric error term e to capture the effects of white noise. Then, the neural output distance function for estimation takes the form:

$$\begin{aligned}-\ln Y_J(x) = a_{0J} &+ \sum_{k=1}^{m_J} a_{kJ}\varphi(\ln x \cdot \beta_{kJ}) + \ln\left(\frac{Y_{-J}}{Y_J}\right) \cdot \gamma_{-J} \\ &+ \ln x \cdot \xi + \delta_J t + u + e\end{aligned} \tag{8}$$

which is the expression of a stochastic frontier model.

The well-known idea behind the so-called "composed error" specification is that production is subject to two disturbances (Kumbhakar and Lovell, 2000): (a) The positive disturbance u expresses the fact that each firm's output lies on or below its frontier, where any deviation is regarded as the result of factors controllable by the firm (e.g. the capability of the producer, the defective products, etc.). (b) The frontier itself may vary randomly over time for the same firm and consequently the frontier is stochastic, with random disturbance e, which expresses events beyond the control of the firm (e.g. observation and measurement errors, climate, luck, exogenous shocks, etc.).

2.4. Returns to scale

As is known, Returns to Scale (RTS) describe what happens as the scale of production increases. RTS refers to a technical property of the production process that examines changes in output subsequent to a proportional change in all inputs. If output increases by the same proportional change then there are constant returns to scale (CRTS). If output increases by less than that proportional change, there are decreasing returns to scale (DRS). If output increases by more than that proportion, there are increasing returns to scale (IRS). In other words, if RTS < 1 (> 1) the production technology is characterized by DRS (IRS). If RTS=1 we have CRTS.

The NDF does not place a priori restrictions on the behavior of RTS like other functional forms. As we know, the RTS are equal to the *sum* of the output elasticities of the various inputs. Let ε_i denote the elasticity of output Y_J with respect to factor x_i:

$$\varepsilon_i = \frac{\partial \Upsilon_J(x)}{\partial x_i} \cdot \frac{x_i}{Y_J(x)} = \frac{\partial \ln Y_J(x)}{\partial \ln x_i}, \quad i = 1, \ldots, N \tag{9}$$

where $x \in \mathbb{R}^N$ denotes the input vector corresponding to N factors of production.

Given that the determinist part of Eq. (8) is equivalent to the following expression:

$$\ln Y_J(x) = -\frac{1}{\gamma_J}\left[a_{0J} + \sum_{k=1}^{m_J} a_{kJ}\varphi(\ln x \cdot \beta_{kJ}) + \ln Y_{-J} \cdot \gamma_{-J} + \ln x \cdot \xi + \delta_J t\right],$$

where $\gamma_J = 1 - \sum_{j=1}^{J-1}\gamma_j$, the RTS index, which depends on inputs, is equal to

$$\begin{aligned}RTS &= \sum_{i=1}^{N} \frac{\partial \ln Y_J(x)}{\partial \ln x_i} \\ &= -\frac{1}{\gamma_J}\left\{\sum_{i=1}^{N}\left[\sum_{k=1}^{m_J} a_{kJ}\beta_{kiJ}\varphi_J'(\ln x \cdot \beta_{kJ})\right.\right. \\ &\quad \left.\left. + \sum_{j=1}^{J-1}\gamma_j\left(\sum_{i=1}^{N}\sum_{k=1}^{m_j} a_{kj}\beta_{kij}\varphi_i'(\ln x \cdot \beta_{kj}) + \sum_{i=1}^{N}\theta_{ji}\right) + \xi_i\right]\right\}\end{aligned} \tag{10}$$

In the following section, we demonstrate how to estimate the parameters of this model in such a way so that the estimated function satisfies the properties dictated by production theory.

3. Monotonicity, curvature and homogeneity conditions

According to neoclassical production theory, output distance functions must be non-decreasing, convex and homogenous of degree one in outputs, and non-increasing and quasi-convex in inputs. Typically, the homogeneity condition can be easily imposed. The monotonicity and curvature constraints are difficult to impose using conventional econometric approaches and are usually imposed at each data point. More precisely, the problem is converted to a nonlinear programming problem which is very difficult and complex even for sampling techniques (Lau, 1978).

The NDF can provide a global approximation to any true production system, whether derived from production theory or not. It is, thus, important to use some procedure for eliminating parameters which are *not* consistent with neoclassical production theory, of course without consequences for the properties of the model. In the NDF this can be done by placing whatever restrictions on parameters are thought to be empirically or theoretically plausible to ensure that NDF satisfies globally the following properties implied by production theory for all admissible values of x and Y.

3.1. Monotonicity in outputs

Output distance functions are non-decreasing in outputs. This implies

$$\gamma_j \geqslant 0, \quad j = 1, \dots, J \tag{11}$$

given that $\frac{\partial \ln D(x)}{\partial \ln Y_j} = \frac{\partial D}{\partial Y_j} \frac{Y_j}{D}$, $\frac{Y_j}{D} > 0$ and $\frac{\partial \ln D}{\partial \ln Y_j} = \gamma_j$.

3.2. Homogeneity in outputs

Output distance functions are homogenous of degree one in outputs. From Euler's theorem this implies

$$\sum_{j=1}^{J} \gamma_j = 1 \tag{12}$$

It has been mentioned above and discussed extensively in Kumbhakar and Lovell (2000) that imposing this constraint is equivalent to normalizing the output distance function by one of the outputs which leads to Eq. (7). Moreover, given that $\gamma_j \geqslant 0$ this clearly implies that $\gamma_j \in [0, 1]$, $j = 1, \dots, J-1$, so that $\sum_{j=1}^{J-1} \gamma_j \leqslant 1$.

3.3. Convexity in outputs

Output distance functions are convex in outputs. Thus, the output distance function will be convex in Y over the non-negative orthant if and only if the Hessian matrix (H) is positive semidefinite (Rao and Bhimasankaram, 1992). In another formulation, the

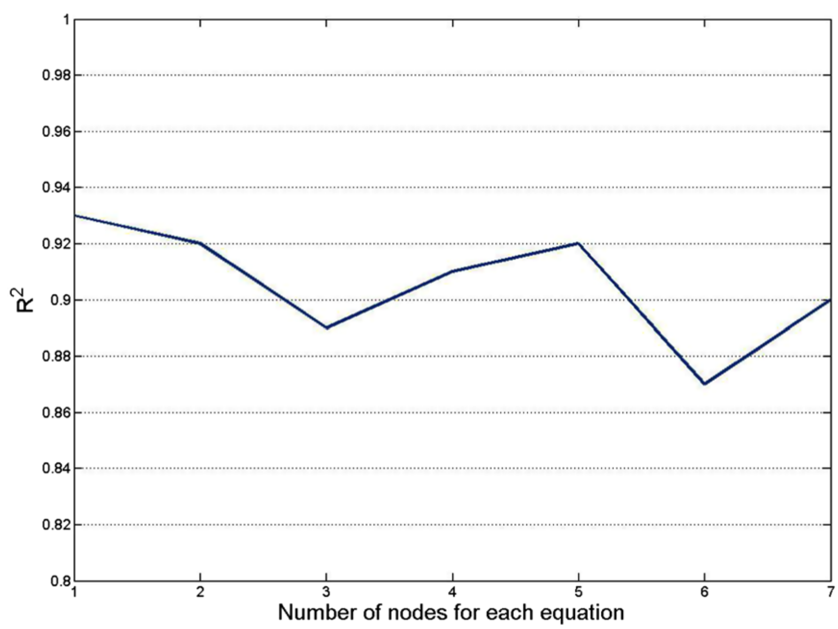

Fig. 1a. $\widetilde{R}^2$ and the number of nodes.

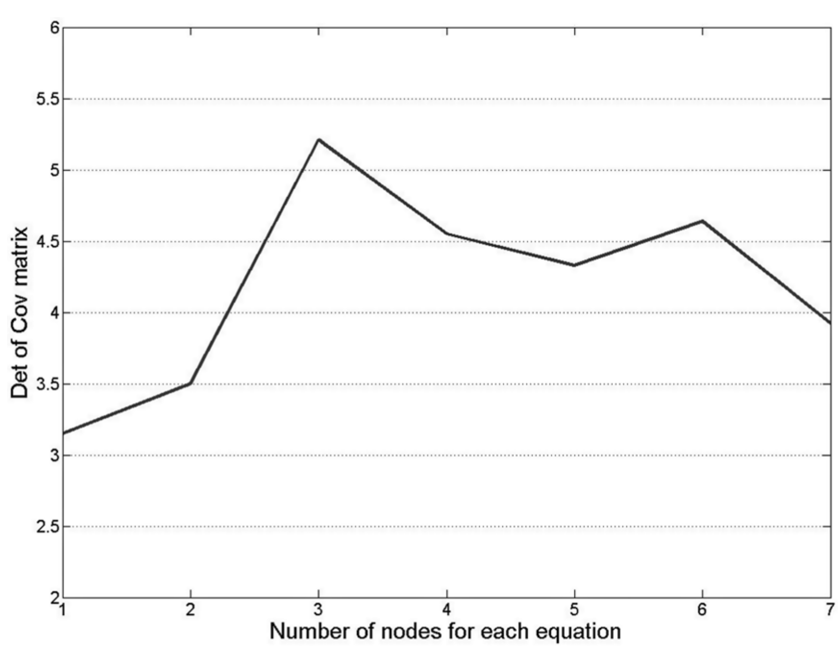

Fig. 1b. $\det \Sigma(\beta)$ and the number of nodes.

output distance function will be convex in Y if and only if all the principal minors of H are non-negative, i.e. positive or zero (Simon and Blume, 1994). Given that the typical element of H is zero (since $\frac{\partial \ln D}{\partial \ln Y_j} = \gamma_j, j = 1, \ldots, J$) all the principal minors of H are equal to zero, and thus, the output distance function is globally convex in outputs.

3.4. Monotonicity in inputs

Output distance functions are non-increasing in inputs. This implies

$$\sum_{k=1}^{m_J} a_{kJ} \beta_{kiJ} \phi_J'(\ln x \cdot \beta_{kJ}) + \xi_i \leqslant 0, \quad i = 1, \ldots, N \tag{13}$$

given that $\frac{\partial \ln D}{\partial \ln x_i} = \frac{\partial D}{\partial x_i} \frac{x_i}{D}$, $\frac{x_i}{D} > 0$ and $\frac{\partial \ln D}{\partial \ln x_i} = \sum_{k=1}^{m_J} a_{kJ} \beta_{kiJ} \phi_J'(\ln x \cdot \beta_{kJ}) + \xi_i$, $i = 1, \ldots, N$

3.5. Quasi-convexity in inputs

Output distance functions are quasi-convex in inputs. For the output distance function to be quasi – convex in inputs over the non-negative orthant it is necessary that the diagonal terms of the H are non-negative (Simon and Blume, 1994). The necessary condition is typically used to impose quasi-convexity in empirical work (O'Donnell and Coelli, 2005).

This implies:

$$\frac{\partial^2 D}{\partial x_i^2} = \frac{\partial \left[\frac{\partial \ln D}{\partial \ln x_i}\right]}{\partial x_i} \cdot \frac{D}{x_i} - \frac{\partial \ln D}{\partial \ln x_i} \cdot \frac{D}{x_i^2} \geqslant 0, \quad i = 1, \ldots, n \tag{14}$$

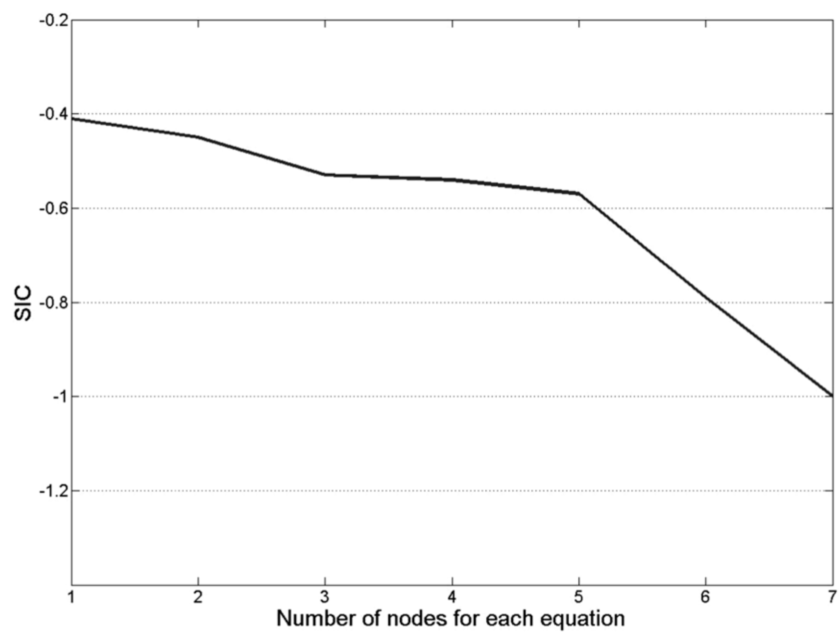

Fig. 1c. SIC and the number of nodes.

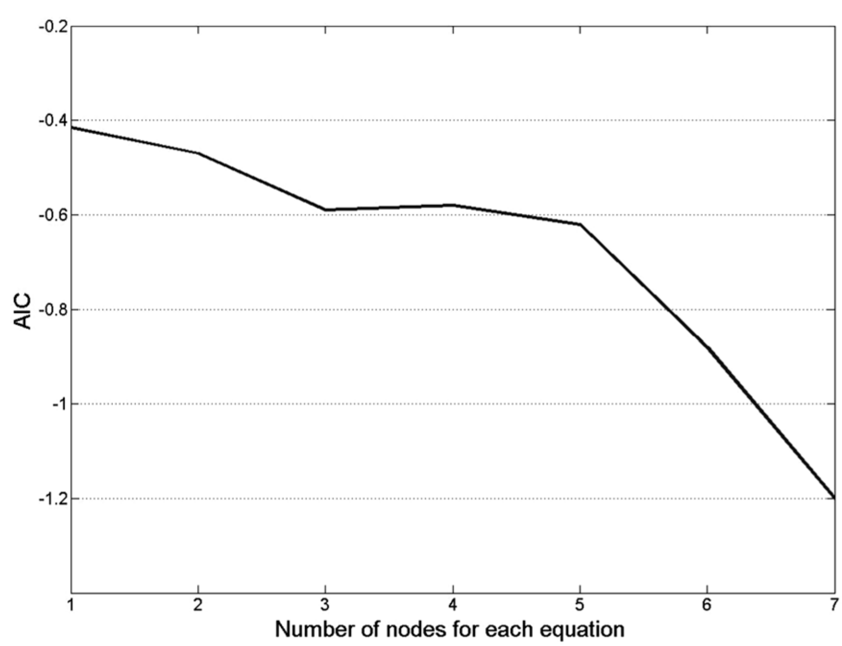

Fig. 1d. AIC and the number of nodes.

where

$$\frac{\partial \ln D}{\partial \ln x_i} = \sum_{k=1}^{m_J} a_{kJ}\beta_{kiJ}\phi_J'(\ln x \cdot \beta_{kJ}) + \xi_i$$

and

$$\frac{\partial^2 \ln D}{\partial \ln x_i^2} = \sum_{k=1}^{m_J} a_{kJ}\beta_{kiJ}^2\phi_J''(\ln x \cdot \beta_{kJ}).$$

Thus

$$\frac{\partial^2 D}{\partial x_i^2} = \left[\left[\sum_{k=1}^{m_J} a_{kJ}\beta_{kiJ}^2\phi_J''(\ln x \cdot \beta_{kJ})\right] - \left[\sum_{k=1}^{m_J} a_{kJ}\beta_{kiJ}\phi_J'(\ln x \cdot \beta_{kJ}) + \xi_i\right]\right]\frac{D}{x_i^2}.$$

Consequently, given that $\frac{D}{x_i^2} \geqslant 0$ quasi-convexity in inputs implies:

$$\frac{\partial^2 Y_J}{\partial x_i^2} = \left[\sum_{k=1}^{m_J} a_{kJ}\beta_{kiJ}^2\phi_J''(\ln x \cdot \beta_{kJ})\right] - \left[\sum_{k=1}^{m_J} a_{kJ}\beta_{kiJ}\phi_J'(\ln x \cdot \beta_{kJ}) + \xi_i\right] \geqslant 0 \quad (15)$$

It can easily be checked that for the following values of the parameters the monotonicity, curvature and homogeneity conditions are satisfied for any economically admissible value of inputs and outputs: $\gamma_j \geqslant 0,\ \sum_{j=1}^{J}\gamma_j = 1,\ \xi_i \leqslant 0,\ a_{kJ} \leqslant 0,\ \beta_{kiJ} \geqslant 0,\ (j = 1,\ldots,J,\ i = 1,\ldots N,\ k = 1,\ldots m_J),\ x \geqslant 1$ given that $Y_{-J} \geqslant 0$ and $\phi'(z) \geqslant 0$ (see Section 5, below). Note that $x \geqslant 1$ (due to relevant normalization of inputs) implies $z = \ln x \cdot \beta_{kJ} \geqslant 0$ and this, in turn, implies $\phi''(z) \leqslant 0$, which is necessary for imposing curvature conditions. Consequently, the NDF satisfies globally the theoretical properties

Table 1
NDF estimate (*t*-statistic in parenthesis).

	Eq. (1)	Eq. (2)	Eq. (3)	Eq. (4)	Eq. (5)
α_{0j}	−1.8912e+003 (−7.5727)	−2.7530e+003 (−6.5075)	−2.6869e+003 (−4.9340)	−9.5372e+003 (−3.0105)	244.1422 (1.2728e+003)
α_{1j}	−1.3827e+003 (−1.9392)	4.7029e+003 (5.7915)	−2.8208e+000 (−0.2561)	4.0856e+000 (0.2189)	−91.4680
θ_1	6.8168e−001 (18.5266)	1.8380e−001 (6.8105)	5.3472e−001 (17.6708)	6.8604e−001 (12.3940)	
θ_2	1.3952e−001 (7.6215)	2.5828e−001 (20.4001)	1.7385e−001 (9.9441)	4.4883e−002 (1.7364)	
θ_3	1.8516e−001 (13.4405)	3.4136e−001 (35.8067)	5.0566e−001 (49.9996)	5.1222e−001 (22.4912)	
θ_4	3.1641e−001 (17.0946)	1.8969e−001 (13.1835)	4.8856e−001 (30.0945)	3.6191e−001 (9.5329)	
θ_5	3.4325e−001 (13.2020)	6.5678e−001 (37.3008)	3.1908e−001 (16.0690)	7.0019e−002 (1.9185)	
δ_j	−3.7286e−002 (−8.3650)	3.8454e−003 (1.2408)	−5.2149e−002 (−9.1056)	−7.7006e−002 (−12.2702)	0.0376 (1.2730e+001)
$b_{1,1}$	0.2061	3.2316	2.7973	1.8110	8.4104
$b_{1,2}$	7.6340	4.0690	6.0888	0.6973	3.7725
$b_{1,3}$	2.0920	7.4919	7.3135	1.8855	7.9100
$b_{1,4}$	0.0649	9.2854	5.8001	9.5824	5.0608
$b_{1,5}$	3.7765	4.8800	0.4460	5.7002	2.6403
ξ_1					−2.5862
ξ_2					−3.3956
ξ_3					−0.2770
ξ_4					−0.7193
ξ_5					−0.2834
γ_1					0.1573
γ_2					0.2289
γ_3					0.2376
γ_4					0.2058
TFP					−0.0256
$\widetilde{R}^2$					0.9214

Note: Eqs. (1)–(4) above refer to Eq. (20) for j = 1, …,4. Also, Eq. (5) refers to Eq. (21) for J = 5.

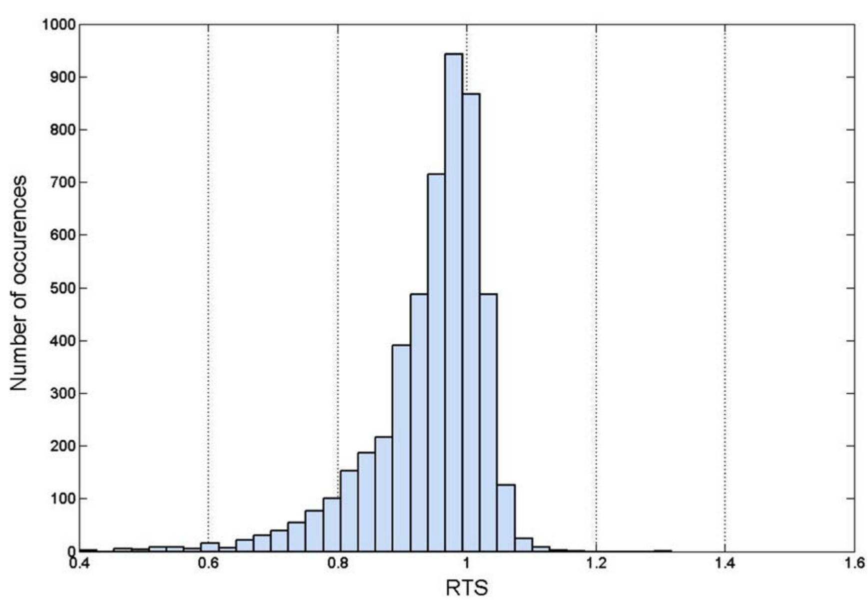

Fig. 2. Returns to scale.

dictated by neoclassical production theory. At this point, it should be stressed that using these restrictions to impose the theoretical properties does not destroy the globally flexible character of the NDF since it satisfies the theoretical properties implied by neoclassical production theory for any admissible value of the inputs and outputs.

In a similar vein, one could easily try imposing *a priori* the following conditions on the system of equations and estimate it easily, as described below (see Section 5): $\alpha_{kj} \geqslant 0,\ \beta_{kj} \geqslant 0,\ \theta_j \geqslant 0\ (j = 1, \ldots, J-1; k = 1, \ldots, m_j)$. These conditions imply that the reduced-form equations which express the respective production functions (i.e. a conventional Cobb-Douglas plus additive NN–based terms) are well-behaved (i.e. monotonic and quasiconcave, apart from being non-negative, finite, continuous, bounded and twice differentiable). However, as we know, the problem with this approach is that there are numerous situations in empirical applications where these assumptions should be relaxed because they are not plausible (Varian, 1992). In addition, one might choose to use an alternative specification other than a well-behaved (neoclassical) production function, based on the structure of the productive system under investigation (Vaneman and Triantis, 2007).

4. Total factor productivity (TFP) and technical efficiency (TE)

4.1. Total factor productivity

In production theory, growth in TFP represents output growth not accounted for by the growth in inputs (Varian, 1992). TFP explains changes in productivity, i.e. the production of more output with a given level of inputs.

By definition, TFP is given by:

$$TFP = \frac{\partial \ln Y_J(x)}{\partial t} \tag{16a}$$

Thus, TFP is given by the relation:

$$TFP = -\frac{1}{\gamma_J}\left[\sum_{j=1}^{J-1} \gamma_j \delta_j + \delta_J\right] \tag{16b}$$

4.2. Technical efficiency

Most theoretical results are based on the assumption that the firm is efficient in the production of goods and services. However, if a firm is not efficient and its inefficiency persists over time it cannot survive in the long run, under reasonably competitive conditions or contestable markets (Tsionas, 2001). In this context, reliable measures of technical efficiency are of great interest.

In this approach, the typical assumption about Eq. (8) is that e are *iid* $(0, \sigma^2)$ and uncorrelated with the regressors. No distributional assumption is needed on the u_i . It is only required that $u \geqslant 0$, which represents inefficiency (Kumbhakar and Lovell, 2000). This model is attractive because of its simplicity. Of course, technical efficiency (TE) could be estimated in numerous ways, where distributional assumptions on the two error components have to be made. For instance, the conventional assumption is that $u \sim N^+(0, \sigma^2)$ which is typically employed in empirical works. However, other distributional assumptions on the one-sided error component are employed (e.g. exponential, normal, truncated normal, gamma, etc.), but less frequently because of their increased computational complexity (Kumbhakar and Lovell, 2000).

In the typical approach of measuring efficiency by means of an output distance function, the residuals $\hat{u}_{i^*}$ are non-negative, with at least one being zero, and are used to provide consistent estimates of technical efficiency of each firm (Kumbhakar and Lovell, 2000). For any given year, the technical efficiency (TE_i) of firm i is equal to

$$D_i = TE_i = \exp(-\hat{u}_{i^*}) \tag{17a}$$

where the residuals are corrected so that: $-\hat{u}_{i^*} = \hat{u}_{i-}\max_i\{\hat{u}_i\}$

For relatively small $\hat{u}_{i^*}$ we have that

$$\hat{u}_{i^*} \approx 1 - \exp(-\hat{u}_{i^*}) = 1 - D_i \tag{17b}$$

So, $\hat{u}_{i^*}$ is often used as a measure of technical inefficiency. In the output distance function approach this analysis is applied to the residuals of Eq. (8), implying a simple change of sign from $+u_i$ to $-u_i$ (Kumbhakar and Lovell, 2000). Consequently, the estimate of the intercept of Eq. (8) is adjusted, so that the function no longer passes through the centre of the observed points but bounds them from above and the distance measure for the ith firm is then calcu-

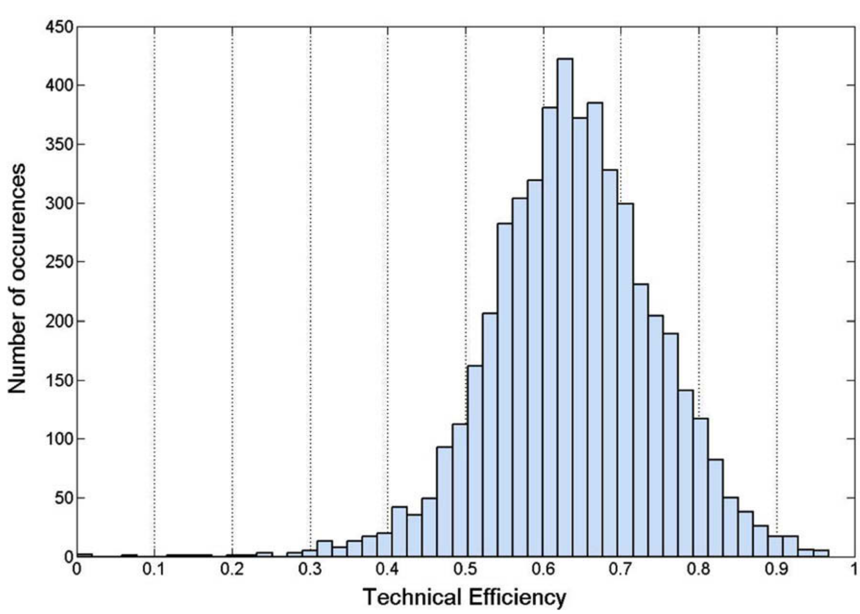

Fig. 3a. Technical efficiency (NDF specification).

lated as the exponent of the (corrected) residual as shown above (Coelli and Perelman, 2000).

Finally, given the measures of TE and TFP change, a measure of technical change may be computed routinely (Färe et al., 1994).

5. NDF implementation

We will outline below a procedure which imposes all the restrictions implied by the theoretical properties for the NDF, globally. From the universal approximation studies (e.g. Hornik et al., 1989, 1990), typical transfer functions must be continuous, bounded, differentiable and monotonic increasing. Probably, the most popular activation function is the so-called sigmoidal:

$$\phi(z) = \frac{1}{1+\exp(-z)}, \quad z \in \mathbb{R} \tag{18}$$

This activation function is nearly linear in the central part and bounds the output to a finite range [0, 1]. We chose this approach because "given the popularity of linear models in econometrics, this form is particularly appealing, as it suggests that ANN models can be viewed as extensions of, rather as alternatives to, the familiar models" (Kuan and White, 1994). For other activation functions see Bishop (1995). However, in general the empirical results are robust, regardless of the activation function used (Haykin, 1999).

The specification of the Translog flexible functional form imposing only homogeneity restrictions typically used for estimating output distance functions is written as (O'Donnell and Coelli, 2005):

$$\begin{aligned} -\ln Y_J(x) = {} & a_{0J} + \sum_{j=1}^{J-1} a_j \ln\left(\frac{Y_j}{Y_J}\right) + 0.5 \sum_{j=1}^{J}\sum_{l=1}^{J} a_{jl} \ln\left(\frac{Y_j}{Y_J}\right)\ln\left(\frac{Y_l}{Y_J}\right) \\ & + \sum_{i=1}^{N} b_i \ln x_i + 0.5 \sum_{i=1}^{N}\sum_{p=1}^{N} b_{ip} \ln(x_i)\ln(x_p) \\ & \times \sum_{i=1}^{N}\sum_{j=1}^{J-1} g_{ij} \ln(x_i) \ln\left(\frac{Y_j}{Y_J}\right) + u + e \end{aligned} \tag{19}$$

In the next sections, we will estimate and compare the NDF model with its competing specification, i.e. the Translog, in order to demonstrate the advantages of the proposed approach.

5.1. Model building

Our approach is based on a four-step algorithm which employs the SUR equations technique for estimating the coefficients of a system of linear equations and an iterative optimization algorithm for the nonlinear parameters of the NDF.

The estimation of the NDF is based on the system of Eq. (3). The procedure is as follows:

Step 1: Let $\beta \in \mathbb{R}_+^{\sum_{j=1}^{J} m_J}$, $\alpha_{kJ} \in \mathbb{R}_-$, $(k = 1, \ldots, m_J)$, $\gamma_j \in [0,1]$, $j = 1, \ldots, J-1$ and $\xi \in \mathbb{R}_-^N$ be drawn from a uniform distribution in their respective domains.

Step 2: Then, estimate α_{0j} $(j = 1, \ldots, J)$, α_{kj} $(j = 1, \ldots, J-1;\ k = 1, \ldots, m_j)$, θ_j $(j = 1, \ldots, J-1)$, δ_j $(j = 1, \ldots, J)$, by means of the system:

$$\ln Y_{jt}(x_t) = a_{0j} + \sum_{k=1}^{m_i} a_{kj}\phi_j(\ln x_t \cdot \beta_{kj}) + \ln x_t \cdot \theta_j + \delta_j t + e_{j,t}, \quad j = 1, \ldots, J-1 \tag{20}$$

$$-\ln Y_{Jt}(x) = a_{0J} + \sum_{k=1}^{m_J} a_{kJ}\varphi_J(\ln x \cdot \beta_{kJ}) + \ln\left(\frac{Y_{-J}}{Y_J}\right) \cdot \gamma_{-J} + \ln x \cdot \xi + \delta_J t + e_{J,t} \tag{21}$$

where x_t denotes the vector of inputs of date t, Y_t the output levels of date t, $e_t \equiv [e_{0t}, e_{1t}, \ldots, e_{J,t}]'$ is a vector random variable, distributed as *iid* $N(0, \Sigma)$, Σ is a covariance matrix. System (20) and (21) is linear in the parameters α_{0j} $(j = 1, \ldots, J)$, α_{kj} $(j = 1, \ldots, J-1;\ k = 1, \ldots, m_j)$, θ_j $(j = 1, \ldots, J-1)$, δ_j $(j = 1, \ldots, J)$ and can be estimated using standard, iterative SUR (e.g. Berndt, 1991).

Step 3: Compute the determinant of the covariance matrix $\det \Sigma^{(i)} \equiv \det \Sigma(\beta)$. Repeat for $i = 1, .., I$ and select the values $\bar{\beta}$, $\bar{\alpha}_J$, $\bar{\gamma}$, $\bar{\xi}$ that yield the minimum value of $\det \Sigma^{(i)}$.

Table 2
Translog distance function estimate (*t*-statistic in parenthesis).

Coefficient	Estimate (*t*-statistic in parenthesis)
α_{0J}	−4.0331e+000 (−2.7738)
α_1	1.1257e−001 (0.6354)
α_2	2.6107e−005 (0.2362)
α_3	3.4184e−001 (1.3878)
α_4	2.1999e−001 (1.8564)
α_{11}	1.4516e−002 (0.0533)
α_{12}	4.0390e+000 (0.1619)
α_{13}	6.4814e+000 (0.0287)
α_{14}	1.9590e+000 (0.1237)
α_{22}	2.0980e−001 (12.5170)
α_{23}	9.0713e+000 (0.1504)
α_{24}	3.9014e+000 (0.1988)
α_{33}	1.7418e−001 (6.3534)
α_{34}	−2.9363e+000 (−0.3498)
α_{44}	2.4631e−002 (5.9700)
b_1	−1.0497e+000 (−2.0157)
b_2	−1.6685e−001 (−0.7868)
b_3	−4.9481e−001 (−3.4197)
b_4	6.2985e−002 (0.2714)
b_5	8.3200e−001 (2.8263)
b_{11}	−1.9089e−001 (−1.9958)
b_{12}	−1.8821e+000 (−0.3303)
b_{13}	−5.3301e+000 (−1.2554)
b_{14}	1.0973e+000 (−0.0139)
b_{15}	4.2537e−001 (1.1129)
b_{22}	−1.3537e−002 (−0.7121)
b_{23}	−1.1385e+000 (0.0229)
b_{24}	8.0624e−001 (0.1458)
b_{25}	−1.5112e+000 (−0.3118)
b_{33}	−3.8866e−002 (−4.2965)
b_{34}	4.8146e−003 (0.1547)
b_{35}	−1.1325e+000 (−0.1594)
b_{44}	−5.7691e−003 (−0.3820)
b_{45}	1.5131e−001 (−0.1056)
b_{51}	1.1348e−002 (−1.1129)
b_{52}	1.4940e+000 (0.3118)
b_{53}	1.2799e+000 (0.1594)
b_{54}	−1.7011e−001 (−0.1056)
b_{55}	−2.6090e−001 (−7.4207)
g_{11}	−3.5243e−002 (−1.3051)
g_{12}	−1.6693e−002 (−0.5960)
g_{13}	−5.6832e−003 (−0.1320)
g_{14}	−1.8570e−002 (−1.0745)
g_{21}	8.4353e−003 (0.6405)
g_{22}	2.8952e−002 (2.1414)
g_{23}	1.2682e−002 (0.6373)
g_{24}	1.2119e−002 (1.4523)
g_{31}	3.2160e−002 (3.4249)
g_{32}	−2.1630e−002 (−2.2863)
g_{33}	−8.3395e−003 (−0.7138)
g_{34}	1.6552e−003 (0.2523)
g_{41}	1.7520e−002 (1.2458)
g_{42}	−1.7010e−002 (−1.2012)
g_{43}	−4.4508e−002 (−2.3028)
g_{44}	4.9699e−003 (0.5626)
g_{51}	−1.6166e−002 (−0.8389)
g_{52}	1.2334e−002 (0.6382)
g_{53}	−1.6445e−003 (−0.0417)
g_{54}	−2.1408e−003 (−0.1542)
R^2	0.9640

Note: The estimation results refer directly to Eq. (19).

Step 4: For $\bar{\beta},\ \bar{\alpha}_J,\ \bar{\gamma},\ \bar{\xi}$ that yield the minimum value of $\det \Sigma^{(i)}$, re-estimate the system and keep the estimated values for parameters $\alpha_{0j}\ (j=1,\ldots,J),\ \alpha_{kj}\ (j=1,\ldots,J-1;\quad k=1,\ldots,m_j),\ \theta_j\ (j=1,\ldots,J-1),\ \delta_j\ (j=1,\ldots,J)$.

5.2. Model selection

Although ANNs can approximate any smooth nonlinearity, no widely accepted guideline exists in choosing the appropriate model for empirical applications. The number of nodes m is determined by a trial-error process. Alternatively, the number of nodes m could be selected using the generalized R^2, $\widetilde{R}^2$ goodness-of-fit criterion. As is well known, R^2 is a statistical measure of how well the estimated model approximates the real data points and a value equal to 1 indicates perfect fit to the data. In this framework, $\widetilde{R}^2$ is a modification of R^2 for systems of equations. According to this criterion one should select the number of nodes that maximizes $\widetilde{R}^2$. When $\widetilde{R}^2$ reaches a maximum, one should stop adding explanatory terms.

Other criteria for model selection include the determinant of the covariance matrix, the SIC (*Schwartz Information Criterion*) (Schwartz, 1978) and the AIC (*Akaike Information Criterion*) (Akaike, 1974). According to these criteria one should select the number of nodes that minimizes their value. In other words, when the determinant of the covariance matrix or the SIC or the AIC, respectively, reaches a minimum, one should stop adding explanatory terms.

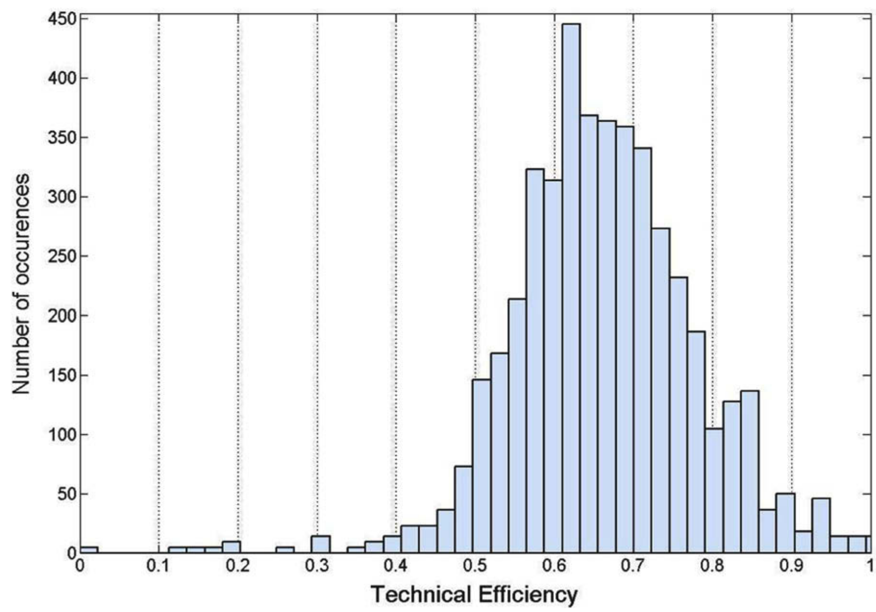

Fig. 3b. Technical efficiency (Translog specification).

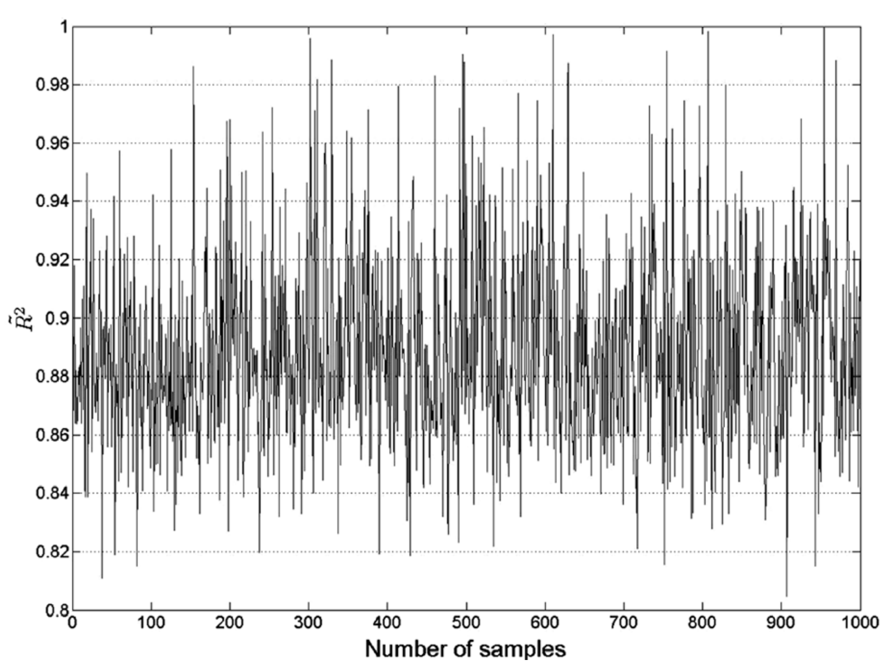

Fig. 4. NDF fitting ($\widetilde{R}^2$).

6. Result analysis and comparison

6.1. Empirical analysis

We believe that it is important to compare the NDF with the most popular competing specification, namely the Translog functional form. In this context, the NDF will be used to fit a large real-world data set and it will then be compared to the Translog specification. The data set comes from the commercial bank and bank holding company database managed by the Federal Reserve Bank of Chicago in the 1989–2000 time span. It is based on the report of condition and income (call report) for all US commercial banks that report to the Federal Reserve banks and the FDIC. The implementation of the output distance function in real organizational settings, where the system has multiple inputs and outputs, is relevant here. Analytically, there are five (5) output variables, namely: (1) instalment loans (to individuals for personal/household expenses), (2) real estate loans, (3) business loans, (4) federal funds sold and securities purchased under agreements to resell, and (5) other assets (assets that cannot be properly included in any other asset items in the balance sheet). Also, there are five (5) input variables, namely: (1) labor, (2) capital, (3) purchased funds, (4) interest-bearing deposits in total transaction accounts and (5) interest-bearing deposits in total non-transaction accounts.

The estimation procedure described earlier was used to estimate the parameters $[a, \theta, \gamma, \xi] \in R^{J(N+1)+\sum_{j=1}^{J} m_j - 1}$. A choice has to

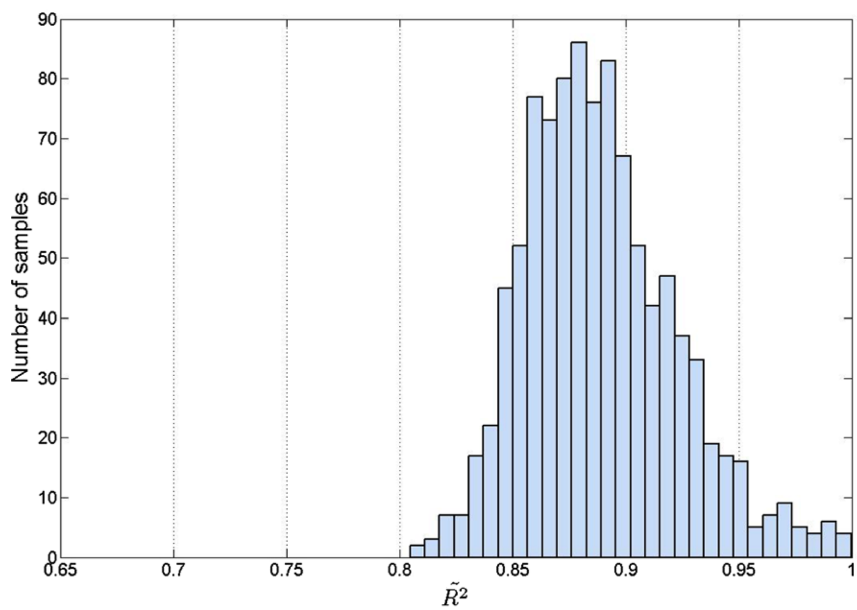

Fig. 5. NDF fitting ($\tilde{R}^2$) – histogram.

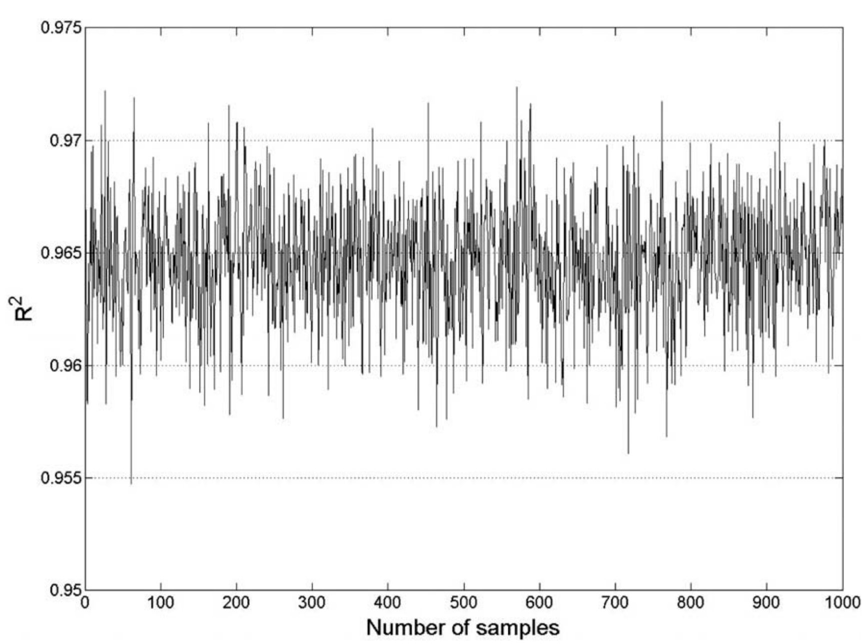

Fig. 6. Translog fitting (R^2).

be made regarding the number of nodes of the ANN. For reasons of convenience, we use the same number of nodes for each equation. The $\widetilde{R}^2$ criterion had a maximum value for $m = 1$ node (Fig. 1a). Also, the determinant of the covariance matrix $\det \Sigma^{(i)} \equiv \det \Sigma(\beta)$ had a minimum value for $m_i = 1$ node (Fig. 1b). Schwartz's (SIC) (Fig. 1c) and Akaike's (AIC) (Fig. 1d) criteria led to similar results. Hence, we set the number of nodes equal to $m_i = 1$ node $(i = 1,] \ldots, J)$. In Table 1, the estimated coefficients for the NDF are shown along with their t-values in parentheses. We confirm that they all take values that are consistent with neoclassical economic theory, and the great majority of the estimated coefficients are highly significant.

Next, the RTS are calculated (Fig. 2) and are found to follow a Gaussian-like distribution around unity (1). This results implies, roughly speaking, CRTS, and is consistent with neoclassical economic theory according to which, as a result of the basic optimization principle, firms generally exhibit CRTS. Finally, TE is calculated by means of the methodology described above and is depicted in Fig. 3a.

6.2. Comparison with the Translog model

In Table 2, the estimates of the competing Translog output distance function are illustrated. Fig. 3b illustrates the TE results calculated routinely with the aid of the standard Translog specification, and are found to be relatively close to the ones calculated by means of the NDF. Next, we draw 1000 random samples from the original data set in order to compare the fitting of the

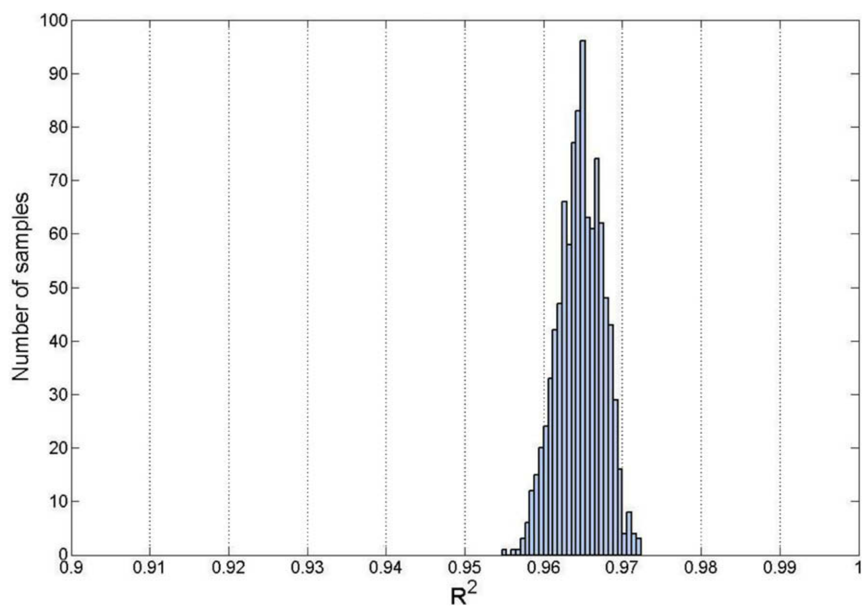

Fig. 7. Translog fitting (R^2) – histogram.

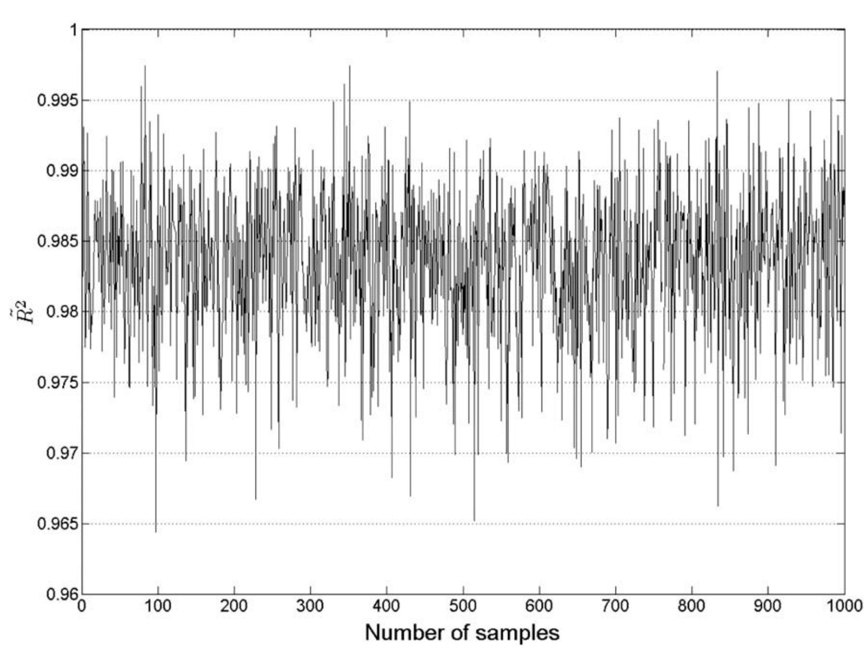

Fig. 8. NDF fitting $(\widetilde{R}^2)$ to Translog generated values.

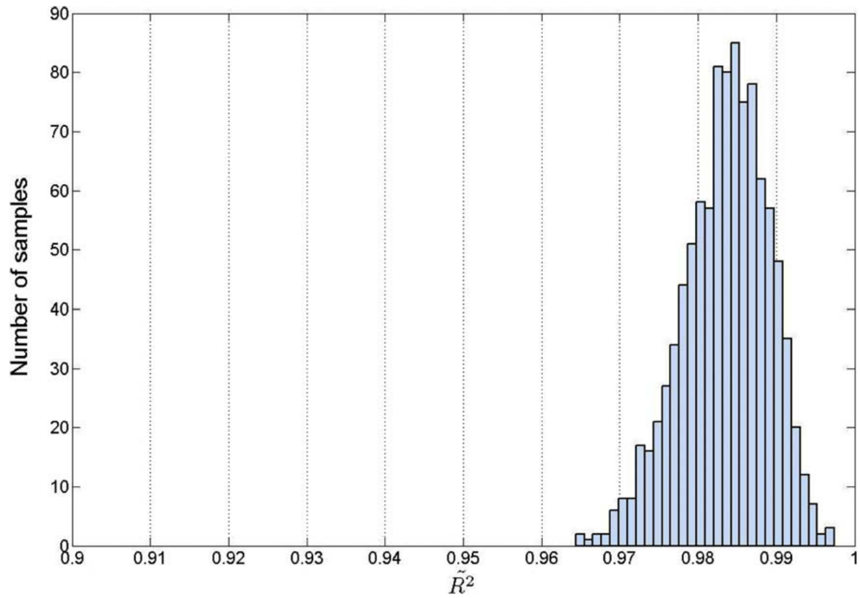

Fig. 9. NDF fitting ($\widetilde{R}^2$) to Translog generated values – histogram.

two specifications, namely the NDF (Figs. 4 and 5) and the Translog (Figs. 6 and 7), which is typically used for estimating output distance functions.

From Figs. 4 and 5 it can be inferred that the proposed model (NDF) explains a very high proportion (around 90%) of the output variance while using fewer estimated parameters – the great majority of which are highly significant – (see Tables 1 and 2, respectively). Also, the NDF satisfies the theoretical properties dictated by neoclassical production theory globally in contrast to the widely adopted Translog specification which achieves a slightly superior fit, as can be inferred from Fig. 6 and 7, by using even more estimated parameters – the majority of which are *not* significant, but fails to satisfy the properties dictated by production theory.

Moreover, we checked the fitting of the proposed NDF to distance function values which were generated by the Translog specification for 1000 samples (Figs. 8 and 9). From Figs. 8 and 9 it can be inferred that the NDF captures (almost) entirely the output distance function values (approximately 98.5%) generated by the popular Translog specification.

To sum up, the proposed model (NDF) satisfies the properties dictated by neoclassical production theory globally and not only over the set of inputs and outputs where inferences are drawn; it gives an approximation to any arbitrary production process; it is flexible with respect to time; it allows for arbitrary returns to scale; it is simple to estimate; it avoids the need for nonlinear estimation; it uses fewer parameters than other globally flexible functional form (parsimonious model); it provides a very good fit to real-world data; it provides an excellent fit to output distance function values generated by its competing Translog specification, it deals with the endogeneity issue in estimation, and, most important, it has a functional form which is consistent with neoclassical production theory. Although, as already stressed, some of these desirable properties are also possessed by one or other of the known specifications, none possesses all of them simultaneously.

7. Conclusion and future research

The results of our research confirm that NNs can be used as an alternative tool for estimating output distance functions and measuring technical efficiency and total factor productivity. More specifically, although several parametric specifications for the output distance function are in use, none satisfies the properties thought as desirable by neoclassical production theory globally. In this paper, we have introduced a new output distance function, the NDF which constitutes a global approximation to any arbitrary production technology with multiple outputs given by an ANN specification. Relevant procedures have been developed and suggestions on implementation proposed. Our empirical implementation relied on standard techniques and all relevant measures were computed routinely.

The proposed model was shown to possess the theoretical properties desired by conventional production theory, in a way not matched by the Translog competing specification. Although some of these desirable theoretical properties are also possessed by one or other of the known specifications, none possesses all of them simultaneously. Moreover, fitted to a very large data set on US commercial banks in the 1989–2000 time span, the NDF explained a very high proportion of the output variance while keeping the number of parameters to a minimum, dealing with the endogeneity issue and satisfying all the theoretical properties dictated by neoclassical production theory globally, in contrast to the Translog popular competing specification. Also, the NDF captured (almost) entirely the output distance function values generated by the Translog specification. We believe that the NDF with its generality, conformity with theory and simplicity of structure is superior to competing specifications and provides a vehicle for testing, expanding and improving conventional production theory.

Of course, there are still several issues that could serve as good examples for future investigation. For instance, an application of the proposed approach to firms and production units operating in other industries and sectors and its extension to include other measures poses an interesting challenge. Also, one could make an attempt to identify the causal factors that are associated with productivity and efficiency performance and incorporate them into the model. Moreover, one could extend the model to account for spillovers across firms. Finally, further research could explore other NN architectures such as Bayesian neural networks, etc. No doubt, future and more extended research on the subject would be of great interest.

References

Aigner, D.J., Lovell, C.A.K., Schmidt, P., 1977. Formulation and estimation of stochastic frontier production function models. Journal of Econometrics 6, 21–37.

Akaike, H., 1974. A new look at the statistical model identification. IEEE Transactions on Automatic Control 19 (6), 716–723.

Anders, U., Korn, O., Schmitt, C., 1998. Improving the pricing of options: A neural network approach. Journal of Forecasting 17, 369–388.

Athanassopoulos, A.D., Curram, S.P., 1996. A comparison of data envelopment analysis and artificial neural networks as tools for assessing the efficiency of decision-making units. Journal of the Operational Research Society 47 (8), 1000–1016.

Berndt, E.R., 1991. The Practice of Econometrics. Addison-Wesley.

Bishop, C.M., 1995. Neural Networks for Pattern Recognition. Clarendon Press, Oxford.

Brummer, B., Glauben, T., Thijssen, G., 2002. Decomposition of productivity growth using distance functions the case of dairy farms in three European countries. American Journal of Agricultural Economics 84, 628–644.

Chan, N.H., Genovese, R.C., 2001. A comparison of linear and nonlinear statistical techniques in performance attribution. IEEE Transactions on Neural Networks 12 (4), 922–928.

Charnes, A., Cooper, W.W., Rhodes, E.L., 1978. Measuring the efficiency of decision making units. European Journal of Operational Research 2, 429–444.

Christensen, L.R., Jorgenson, D.W., Lau, L.J., 1971. Conjugate duality and the transcendental production function. Econometrica 39, 206–218.

Cobb, C.W., Douglas, P.H., 1928. A theory of production. American Economic Review 18 (Supplement), 139–165.

Coelli, T.J., Perelman, S., 1996. Efficiency Measurement, Multi-output Technologies and Distance Functions: With Application to European Railways. CREPP Working Paper 96/05, University of Liege.

Coelli, T.J., Perelman, S., 1999. A comparison of parametric and non-parametric distance functions with application to European railways. European Journal of Operational Research 117, 326–339.

Coelli, T.J., Perelman, S., 2000. Technical efficiency of European railways a distance function approach. Applied Economics 32, 1967–1976.

Cook, W., Seiford, L., 2009. Thirty years on data envelopment analysis (DEA). European Journal of Operational Research 192 (1), 1–17.

Costa, A., Markellos, R.N., 1997. Evaluating public transport efficiency with neural networks models. Transportation Research, Part C 5 (5), 301–312.

Diewert, E., 1971. An application of the Shephard duality theorem: A generalized Leontief production function. Journal of Political Economy 79, 481–507.

Diewert, E., Wales, J.T., 1987. Flexible functional forms and global curvature conditions. Econometrica 55, 43–68.

Efron, B., 1982. The Jacknife, The Bootstrap and Other Resampling Plans, Society for Industrial and Applied Mathematics, Philadelphia.

Efron, B., Tibshirani, R.J., 1993. An Introduction to the Bootstrap. Chapman and Hall, London.

Färe, R., Grosskopf, S., Lovell, C.A.K., Yaisawarng, S., 1993. Derivation of shadow prices for undesirable outputs a distance function approach. Review of Economics and Statistics 75, 374–380.

Färe, R., Grosskopf, S., Norris, M., Xhang, Z., 1994. Productivity growth, technical progress, and efficiency change in industrialized countries. American Economic Review 84, 66–83.

Ferrier, G.D., Hirschberg, J.G., 1997. Bootstrapping confidence intervals for linear programming efficiency scores: With an illustration using Italian bank data. Journal of Productivity Analysis 8 (1), 19–33.

Fleissig, A., Kastens, T., Terrell, D., 2000. Evaluating the semi-nonparametric fourier, aim, and neural networks cost functions. Economics Letters 68, 235–244.

Fuentes, H.J., Grifell-Tatjé, E., Perelman, S., 2001. A parametric distance function approach for malmquist productivity index estimation. Journal of Productivity Analysis 15, 79–94.

Gallant, A.R., 1982. Unbiased determination of production technologies. Journal of Econometrics 20, 285–323.

Gallant, A.R., Golub, G.H., 1984. Imposing curvature restrictions on flexible functional forms. Journal of Econometrics 26, 295–321.

Grosskopf, S., Margaritis, D., Valdmanis, V., 1995. Estimating output substitutability of hospital services a distance function approach. European Journal of Operational Research 80, 575–587.

Hanke, M., 1999. Neural networks versus Black–Scholes: An empirical comparison of the pricing accuracy of two fundamentally different option pricing methods. Journal on Computational Intelligence Finance 5, 26–34.

Hansen, J.V., Nelson, R.D., 2003. Forecasting and recombining time-series components by using neural networks. Journal of the Operational Research Society 54, 307–317.

Haykin, S., 1999. Neural Networks. Prentice-Hall, New Jersey.

Hoopes, B.J., Triantis, K.P., 2001. Efficiency performance, control charts, and process improvement: Complementary measurement and evaluation. IEEE Transactions on Engineering Management 48 (2), 239–253.

Hornik, K., Stinchcombe, M., White, H., 1989. Multilayer feedforward networks are universal approximators. Neural Networks 2, 359–366.

Hornik, K., Stinchcombe, M., White, H., 1990. Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks. Neural Networks 3, 551–560.

Hutchinson, J.M., Lo, A.W., Poggio, T., 1994. A nonparametric approach to pricing and hedging derivative securities via learning networks. Journal of Finance 49, 851–889.

Kuan, C.M., White, H., 1994. Artificial neural networks: An econometric perspective. Econometric Reviews 13, 1–91.

Kumbhakar, S.C., Lovell, C.A.K., 2000. Stochastic Frontier Analysis. Cambridge University Press, Cambridge.

Kumbhakar, S.C., Tsionas, E.G., 2005. The joint measurement of technical and allocative inefficiencies: An application of bayesian inference in nonlinear random-effects models. Journal of the American Statistical Association 100, 736–747.

Kumbhakar, S.C., Tsionas, E.G., 2006. Estimation of stochastic Frontier production functions with input-oriented technical efficiency. Journal of Econometrics 133 (1), 69–71.

Lau, L.J., 1978. Testing and imposing monoticity, convexity and quasi-convexity constraints. In: Fuss, M., McFadden, D. (Eds.), Production Economics: A Dual Approach to Theory and Applications. North Holland, Amsterdam.

Lothgren, M., 1998. How to Bootstrap DEA Estimators: A Monte Carlo Comparison, WP in Economics and Finance, No. 223, Stockholm School of Economics.

Meeusen, W., van den Broeck, J., 1977. Efficiency estimation from Cobb–Douglas production functions with composed error. International Economic Review 18, 435–444.

O'Donnell, J.C., Coelli, J.T., 2005. A Bayesian approach to imposing curvature on distance functions. Journal of Econometrics 126 (2), 493–523.

Rao, A.R., Bhimasankaram, P., 1992. Linear Algebra. McGraw Hill, New Delhi.

Reinhard, S., Thijssen, G., 1998. Resource Use Efficiency of Dutch Dairy Farms: A Parametric Distance Function Approach. Paper Presented at Georgia Productivity Workshop III, Athens Georgia, October.

Santin, D., 2008. On the approximation of production functions: A comparison of artificial neural networks frontiers and efficiency techniques. Applied Economics Letters 15 (8), 597–600.

Santin, D., Delgado, F., Valino, A., 2004. The measurement of technical efficiency: A neural network approach. Applied Economics 36, 627–635.

Schwartz, G., 1978. Estimating the dimension of a model. Annals of Statistics 6 (2), 461–464.

Sfetsos, A., Siriopoulos, C., 2005. Time series forecasting of averaged data with efficient use of information. IEEE Transactions on Systems, Man, and Cybernetics – A 35 (5), 738–745.

Shephard, R., 1953. Cost and Production Functions. Princeton University Press, Princeton, NJ.

Simar, L., Willson, P., 1999. Some problems with the Ferrier/Hirschberg bootstrap idea. Journal of Productivity Analysis 11 (1), 67–80.

Simar, L., Wilson, P., 1998. Sensitivity analysis of efficiency scores: How to bootstrap in nonparametric Frontier models. Management Science 44 (11), 49–61.

Simon, C., Blume, L., 1994. Mathematics for Economists. Norton and Company, New York.

Stewart, K., 2005. Applied Econometrics. Thomson-Brooks/Cole, Belmont.

Terrell, D., 1996. Incorporating monotonicity and concavity conditions in flexible functional forms. Journal of Applied Econometrics 11, 179–194.

Tsionas, E.G., 2000. Full likelihood inference in normal-gamma stochastic Frontier models. Journal of Productivity Analysis 13, 183–205.

Tsionas, E.G., 2001. An introduction to efficiency measurement using Bayesian stochastic Frontier models. Global Business and Economics Review 3 (2), 287–311.

Tsionas, E.G., 2002. Stochastic Frontier models with random coefficients. Journal of Applied Econometrics 17, 127–147.

Tsionas, E.G., 2003. Combining DEA and stochastic Frontier models: An empirical Bayes approach. European Journal of Operational Research 147 (3), 499–510.

Tsionas, E.G., Kumbhakar, S.C., 2004. Markov switching stochastic Frontier model. Econometrics Journal 7 (2), 398–426.

Tsitsiklis, J.N., Roy, B.V., 2001. Regression methods for pricing complex American-style options. IEEE Transactions on Neural Networks 12 (4), 694–703.

Vaneman, W.K., Triantis, K., 2007. Evaluating the productive efficiency of dynamical systems. IEEE Transactions on Engineering Management 54 (3), 600–612.

Varian, H., 1992. Microeconomic Analysis. W.W. Norton, New York.

White, H., Racine, J., 2001. Statistical inference, the bootstrap, and neural-network modeling with application to foreign exchange rates. IEEE Transactions on Neural Networks 12 (4), 657–673.

Yair, M., 1996. Production function estimation: Reviving the primal. Econometrica 64 (2), 431–438.

Zhang, G.P., Berardi, V.L., 2001. Time series forecasting with neural network ensembles: An application for exchange rate prediction. Journal of the Operational Research Society 52, 652–664.