Output Distance Functions from a Complexity Perspective: The Neural Network Approach

Tsionas Efthymios
Department of Economics
Athens University of Economics and Business
Athens, Greece
tsionas@aueb.gr

Michaelides Panayotis
School of Applied Mathematics and Physical Sciences
National Technical University of Athens
Athens, Greece
pmichael@central.ntua.gr

Vouldis Angelos, *Member, IEEE*
Department of Economics
University of Athens
Athens, Greece
avouldis@econ.uoa.gr

***Abstract*—The output distance function is a key concept in economics. However, its empirical estimation is less than satisfactory because it often violates properties dictated by economic 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 and imposes all theoretical properties implied by production theory such as monotonicity, curvature, homogeneity for all economically admissible values of outputs and inputs. The model possesses all of the properties thought as desirable in production theory in a way not matched by its competing specification. Fitted to data sets originating in US data for all commercial banks between 1989-2000, the NDF is capable of explaining a very high proportion of the variance of output while keeping the number of parameters to a minimum and satisfying all the theoretical properties dictated by production theory.**

Keywords: Output distance function, neural networks, RTS, TFP

I. INTRODUCTION

In the last decades complexity is becoming increasingly useful as a tool of the physical, biological and social sciences. In this context, modern economic analysis has not remained indifferent to the power of such models to describe complex structures and behaviours. In fact, the theoretical and analytical tools of complexity analysis are regarded as a promising way towards overcoming the problems associated with the traditional approaches to the understanding of economic reality while taking advantage of the available computational techniques. In this context, we propose a new approach to the output distance function which is a key concept in productivity analysis assisting economists in the estimation of a firm's production efficiency.

Output distance functions provide a functional characterization of the production technology with respect to output sets. In addition, they provide measures of technical efficiency and thus they constitute an indispensable tool for conducting productivity analysis for specific industries and firms. Many functional forms in the production context have been proposed for their approximation but perhaps the most important in current use are the Cobb-Douglas specification introduced by Cobb and Douglas [1], and the translog flexible functional form introduced by Christensen et al. [2]. Both of theses models, that are intuitively appealing and computationally straightforward, have been extensively 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. As a result its explanatory power is low.

Furthermore, and even more importantly, a second problem is common in the empirical estimation of the Cobb–Douglas, the translog specification and of other flexible functional forms, namely the violation of properties dictated by economic theory, such as monotonicity, curvature, and homogeneity conditions. Some researchers have developed numerical [3] and Bayesian techniques [4] for imposing curvature conditions that tend to provide encouraging results. However, in production theory it is absolutely necessary to have estimated 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" [5] and still remains "one of the most difficult challenges faced by empirical economists" [6].

In this paper, we propose and estimate a new model which is of comparable generality but which has considerable advantages. The nonparametric feature of Artificial Neural Networks (ANN) makes them quite flexible and attractive in production theory where the theoretical relationship is not known a priory [7]. However, production functions are inappropriate with technologies producing multiple outputs. In

---

Contact author: e-mail: avouldis@econ.uoa.gr

such cases, we have to use distance functions. Our model, which we call the Neural Distance Function (NDF), gives an arbitrary approximation to any production process; it satisfies the properties dictated by production theory globally and not only over the set of inputs and outputs where inferences are drawn; it is flexible with respect to time as an indicator of technical progress; it allows for arbitrary returns to scale; it is simple to estimate; it avoids the need for non linear estimation; it uses fewer parameters than other flexible functional forms; it provides a better approximation when tested over a very large number of data sets than other specifications; it has a functional form which is consistent with known production theory data. 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, it is superior to them.

Contrary to widely used local approximations like the translog [2], the generalized Leontief [8] or the symmetric McFadden form [5] the proposed flexible production function is a global approximation to the unknown function. The Fourier flexible form [9]-[10] is also a global approximation but it requires an excessive number of parameters. The NDF that we propose provides a better approximation using fewer parameters [11]. Moreover, we extend the analysis to the case of technologies with many 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 scale economies, total factor productivity may be computed routinely.

Our methodology of estimating the synaptic weights of the neural network specification is not based on the selection of a training data set. Instead, it is formulated as a four-step algorithm which employs the SUR technique for estimating the coefficients of a system of linear equations and an iterative optimization algorithm for the nonlinear parameters of the NDF. This choice, regarding the estimation method, was made having in mind the characteristics of the data sets which economists usually work with and in particular their size. In most practical applications the distance function has to be estimated from a rather small data set (usually annual/quarterly data for a period of some years) rendering the selection of a training data an unaffordable luxury. In addition, in the context of approximating a distance function, the neural network specification does not aim at providing forecasting or pattern recognition capabilities as it is usually the case in engineering applications. Instead, the nonlinear nature of the neural network specification is used in order to obtain a sufficiently flexible functional form which is capable of approximating an existing distance function while enabling the *ex ante* imposition of the properties dictated by economic theory. It must be noted that using the NDF for forecasting purposes, when a sufficiently large data set is available, is an issue for further research but it lies outside the scope of the present paper.

An application investigating the model's fitting illustrates our technique. Fitted to a very large number of data sets originating in all US commercial banks in the 1989-2000 time span, the NDF is capable of explaining a very high proportion of the variance of output while keeping the number of parameters to a minimum and satisfying all the theoretical properties dictated by production theory. Thus, the NDF which imposes all theoretical properties globally, for all economically admissible values of outputs and inputs provides a very good approximation by using considerably fewer parameters over a very large data set. It should be noted that, so far, no empirical study has imposed all these properties on parametric distance functions [4]. Also, very few studies report the degree to which the estimated functions satisfy these conditions [12]. We believe that the results of this study suggest that the NDF is the appropriate vehicle for testing, expanding and improving conventional production theory.

The first section of the paper introduces the neural output distance function; in the second section the theoretical properties of the NDF, namely monotonicity, curvature and homogeneity, are examined. The third section presents a procedure for the econometric estimation of the NDF. Finally the fourth section summarizes and concludes the paper.

II. OUTPUT DISTANCE FUNCTIONS WITH NEURAL NETWORKS

Artificial Neural Networks (ANN), instead of fitting the data with a pre-specified model, let the data set itself serve as evidence to support the model's approximation of the underlying production technology [7]. In general, ANNs have found numerous applications in financial modelling [13]-[20]. However, with the exception of very few papers [21]-[22] no systematic research on economic modelling using ANN has been done.

Neural networks are "data-driven, self-adaptive nonlinear methods that do not require specific assumptions about the underlying model" [7]. In mathematical terms, ANNs are collections of transfer functions that relate an output variable to certain input variables $X' = [X_1, ..., X_n]$. In the case of a single hidden layer, the input variables are combined linearly to form m intermediate variables $Z_1, ..., Z_m$ where

$$Z_k = X'\beta_k \ \ (k = 1, ..., m) \qquad (1)$$

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

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

where φ is an activation function, the α_k 's are parameters and m is the number of intermediate nodes [23].

By combining simple units with multiple intermediate nodes, the neural network can approximate any smooth nonlinearity [24]. As demonstrated in Hornik et al. [25]-[26], ANNs have the ability to approximate arbitrarily well a large class of functions while keeping the number of free parameters to a minimum.

A. *The Production Technology*

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

$$P = \left\{ (Y, x) : x \quad can \quad produce \quad Y \right\}.$$

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: $L(x) = \left\{ Y : (Y, x) \in P \right\}$. The concept of output distance function is 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 $\Delta(x, Y) = \min\left\{ \mu : Y / \mu \in L(x) \right\}$. The distance function will take a value of unity if *Y* is located at the production frontier for the specific input vector *x*. Generally: $\Delta(x, Y) \leq 1$ 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 and the x s are predetermined and ε is a stochastic term representing inefficiency (with $\varepsilon \in N\left(0, \sigma^2\right)$).

The estimation of distance functions has attracted increasing attention in the literature. This is due to the fact that distance functions can be used to model multi-input multi-output production technologies without aggregating outputs. As is well known, an output distance function describes the degree to which a firm can expand its output vector, given an input vector. Econometric estimation of distance functions is, generally, a complicated issue; see for example the excellent discussion in [27]. The typical problem is that we have $J > 1$ endogenous variables (the Y 's) but only one equation which is the output distance function. One commonly used technique for estimating distance functions has been data envelopment analysis (DEA). A very serious drawback of this approach is rooted in the fact that DEA does not allow for measurement errors [4]. One can attempt to overcome this problem by specifying a functional form for the production surface. Färe et al. [28] specify a translog output distance function and estimate its parameters. 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) approach can also be used for distance function estimation [29]. To deal with the problem of estimation various unsatisfactory approaches have been proposed in the literature.

Parametric distance functions have allowed numerous researchers to measure the distance that each firm has below the production technology [30] –[35]. In theory, one can use GMM by taking prices of inputs and outputs and the inputs themselves as instruments. However, prices are not available in many cases of interest.

Suppose we write the distance function in log terms as follows: $D(\ln x, \ln Y) = e$, where e is an error term. The idea is to express any Y (say Y_J) as a function of the others and obtain: $\ln Y_J = f\left(\ln x, \ln Y_1, ..., \ln Y_{J-1}\right)$. One can estimate this equation by various techniques (e.g. Ordinary Least Squares (OLS)) but the endogeneity issue is not taken into account. In order to account for endogeneity, we must consider the reduced form: $\ln Y_{-J} = g(\ln x)$, where $\ln Y_{-J} = \left[\ln Y_1, ..., \ln Y_{J-1}\right]'$, and *g* is a vector function $g : \mathbb{R}^N \rightarrow \mathbb{R}^{J-1}$. The reduced form expresses all other outputs as functions of the inputs alone. In econometric terms, we have a system of equations as follows:

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

$$\ln Y_{-J} = g\left(\ln x\right) + e_{-J} \tag{3b}$$

where $e_{-J} = \left[e_1, ..., e_{J-1}\right]'$, and $e = \left[e'_{-J}, e_J\right]'$ represents a J-dimensional random vector. The f function can be specified as a translog. Of course, whether or not such approximations can work in any specific application, is an empirical matter. More difficult is the specification of the vector function g. In many instances, g can be assumed to be a linear function. But this is quite arbitrary and it is not expected to approximate the technology with any reasonable accuracy.

At any rate, if the system of equations in (3) is estimated using the SUR, then the resulting estimates are known as LIML. The crucial part is, however, to specify the *g* and *f* functions using a flexible functional form.

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

$$\ln Y_j(x) = a_{0j} + \sum_{k=1}^{m_j} \alpha_{kj} \phi_j (\ln x \cdot \beta_{kj}) + \ln x \cdot \theta_j + \delta_j t \qquad j = 1, \ldots, J-1 \tag{4}$$

where $Y_j(x)$ is the reduced form function of output i , m_j is the number of intermediate nodes and t is a time index.

Thus, the neural output distance function, in general, can be written as:

$$\ln D = a_{0J} + \sum_{k=1}^{m_J} a_{kJ} \phi_J (\ln x \cdot \beta_{kJ}) + \ln Y \cdot \gamma + \ln x \cdot \xi + \delta_J \cdot t \tag{5}$$

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. In general, for vectors a and b, $a \cdot b$ denotes the inner product. Equation (5) represents our specification for the distance function and equations (4) are simply reduced forms.

For the output distance function it would be possible to adopt a more general specification of the form:

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

However, the drawback of this approach is that the output distance function must satisfy certain theoretical properties which are plausible when $\theta_{kJ} = 0$ but not otherwise, for general activation functions. These properties are detailed in the next section.

The next step is to convert equation (5) into an estimable model and this can be accomplished by exploiting the property indicating that output distance functions are homogenous of degree one in outputs. It can easily be checked that that imposing this constraint is equivalent to normalising the output distance function by one of the outputs. See [27]. Thus, an alternative form of (5), imposing homogeneity of degree one in inputs, is

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

where $u = -\ln D$ is a non-negative term such that $0<D\leq 1$, $-\infty < \ln D \leq 0$ that captures the effects of inefficiency. If we follow the stochastic frontier approach proposed by Aigner et al. [36] and Meeusen and van der Broeck [37] we should add a symmetric error term e to capture the effects of white noise. Then, the neural output distance function for empirical estimation takes the form:

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

which is the expression of a typical stochastic frontier model.

In the following sections, we show how this model can be used in order to estimate its parameters in such a way that the estimated functions satisfy the monotonicity, curvature and homogeneity properties dictated by production theory.

III. MONOTONICITY, CURVATURE AND HOMOGENEITY CONDITIONS

A. Theoretical Properties and Restrictions

Output distance functions are typically 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 traditional econometric approaches and are usually imposed at each data point. More precisely, the problem is converted to a non-linear programming problem which is very difficult and complex using conventional econometric method even for sampling techniques which are said to present some advantages when compared to non-sampling techniques [38].

The NDF can thus provide a global approximation to any true production system, whether derived from production theory or not. However, there is a limited number of parameters in (5) and on most data sets that are unlikely to be all well determined. It is, thus, important to use some straightforward procedure for eliminating unnecessary parameters which are not consistent with production theory, 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. It will be possible to impose these restrictions on a single equation basis. These restrictions will ensure the NDF satisfies globally the following monotonicity (i.e. non-increasing in *x* and non-decreasing in *Y*), curvature (i.e. quasi-convex in *x* and convex in *Y*) and homogeneity (i.e. homogeneous in *Y*) properties implied by production theory for all economically admissible values of *x* and *Y*.

1) Monotonicity in outputs

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

$$\gamma_j \geq 0, \; j = 1,...,J \qquad (9)$$

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

$$\frac{\partial \ln D}{\partial \ln Y_j} = \gamma_j .$$

2) Homogeneity in outputs

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

$$\sum_{j=1}^{J} \gamma_j = 1 \qquad (10)$$

As it has been mentioned above (see Section II), imposing this constraint is equivalent to normalising the output distance function by one of the outputs which leads to equation (7). Moreover, given that $\gamma_j \geq 0$ this clearly implies that $\gamma_j \in [0,1]$, $j = 1,...,J-1$, so that $\sum_{j=1}^{J-1} \gamma_j \leq 1$.

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 [39]. In another formulation, the 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 [40]. Given that the typical element of H is zero (since $\frac{\partial \ln D}{\partial \ln Y_j} = \gamma_j$, $j = 1,...,J$) all the principal minors of H are equal to zero. Consequently, H is positive semidefinite and the output distance function is globally convex in outputs.

4) Monotonicity in inputs

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

$$\sum_{k=1}^{m_J} a_{kJ}\beta_{kiJ}\varphi_J'(\ln x \cdot \beta_{kJ}) + \xi_i \leq 0,\ i = 1,...,n \qquad (11)$$

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

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

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 [40]. The necessary condition is typically used to impose quasi convexity in empirical work [4].

This implies that

$$\frac{\partial^2 D}{\partial {x_i}^2} = \frac{\partial\left[\frac{\partial \ln D}{\partial \ln x_i}\right]}{\partial x_i} \cdot \frac{D}{x_i} - \frac{\partial \ln D}{\partial \ln x_i} \cdot \frac{D}{{x_i}^2} \geq 0,\ i = 1,...,n \qquad (12a)$$

where

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

and

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

Thus,

$$\frac{\partial^2 D}{\partial {x_i}^2} = [[\sum_{k=1}^{m_J} a_{kJ}{\beta_{kiJ}}^2\varphi_J''(\ln x \cdot \beta_{kJ})\ ] - [\sum_{k=1}^{m_J} a_{kJ}\beta_{kiJ}\varphi_J'(\ln x \cdot \beta_{kJ}) + \xi_i]]\frac{D}{{x_i}^2} \qquad (12b).$$

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

$$\frac{\partial^2 Y_J}{\partial {x_i}^2} = [\sum_{k=1}^{m_J} a_{kJ}{\beta_{kiJ}}^2\varphi_J''(\ln x \cdot \beta_{kJ})\ ] - [\sum_{k=1}^{m_J} a_{kJ}\beta_{kiJ}\varphi_J'(\ln x \cdot \beta_{kJ}) + \xi_i] \geq 0 \qquad (13)$$

It can easily be checked that for the following values of the parameters the monotonicity, curvature and homogeneity conditions (7) – (11) are satisfied for any economically admissible value of inputs and outputs: $\gamma_j \geq 0$, $\sum_{j=1}^{J}\gamma_j = 1$, $\xi_i \leq 0$, $a_{kJ} \leq 0$, $\beta_{kiJ} \geq 0$, ($j = 1,...,J$, $i = 1,...n$) x≥1 given that $Y_{-J} \geq 0$ and $\varphi'(z) \geq 0$. This implies $\varphi''(z) \leq 0$. Consequently, the NDF satisfies globally all theoretical properties (7) – (11) dictated by production theory. Note that x≥1 implies normalization of inputs in order to yield $\ln x \cdot \beta_{kJ} \geq 0$ and, therefore, $\varphi''(z) \leq 0$ which is necessary for imposing curvature conditions.

It must be emphasized that using these restrictions to impose the theoretical properties dictated by production theory does *not* destroy the flexibility of the NDF since it satisfies all the theoretical properties implied by production theory *for any economically admissible value of the inputs and outputs*. In fact, this is common practise for globally flexible functional specifications such as the symmetric McFadden form.

B. Returns to Scale

The neural production function does not place a priori restrictions on the behavior of returns to scale (RTS) like other functional forms. It is known that if RTS <1 (>1) the production technology is characterized by decreasing (increasing) returns to scale. If RTS=1 we have constant returns to scale.

It is also know that typically the RTS are equal to the sum of the output elasticities of the various inputs. Let denote the elasticity of output with respect to factor x_i :

$$\varepsilon_i = \frac{\partial \mathrm{Y}_J(x)}{\partial x_i} \cdot \frac{x_i}{Y_J(x)} = \frac{\partial \ln Y_J(x)}{\partial \ln x_i} \qquad (14)$$

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

Given that the determinist part of equation (7) is equivalent to the following expression:

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

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

$$RTS = \sum_{i=1}^{N} \frac{\partial \ln Y_J(x)}{\partial \ln x_i} =$$

$$-\frac{1}{\gamma_J}\left\{ \sum_{i=1}^{n} \left[\begin{array}{l} \sum_{k=1}^{m_J} a_{kJ}\beta_{kiJ}\phi_J'(\ln x \cdot \beta_{kJ}) + \\ \sum_{j=1}^{J-1} \gamma_j (\sum_{i=1}^{n}\sum_{k=1}^{m_j} a_{kj}\beta_{kij}\phi_i'(\ln x \cdot \beta_{kj}) + \\ \sum_{i=1}^{N} \theta_{ji}) + \xi_i \end{array} \right] \right\} \tag{15}$$

C. *Total Factor Productivity*

In economics, growth in total-factor productivity (T.F.P.) represents output growth not accounted for by the growth in inputs [41] and presumably changes over time. It is typically used as a proxy for technical change.

By definition, Total Factor Productivity (TFP) measure is given by

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

Thus, the TFP index is given by the relation

$$TFP = -\frac{1}{\gamma_J}\left[\sum_{j=1}^{J-1} \gamma_j \delta_j + \delta_J \right] \tag{17}$$

IV. ECONOMETRIC ESTIMATION

We will outline below a procedure which imposes all the restrictions implied by the theoretical properties for the NDF, globally.

The activation function has the following form

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

Other activation functions include the Gaussian kernel and the threshold logical unit as used in Hopfield networks [42].

Finally, the specification of the translog flexible functional form imposing homogeneity restrictions typically used for constructing output distance functions can be written as [4]

$$-\ln Y_J(x) = a_{0J} + \sum_{j=1}^{J-1} a_j \ln\left(\frac{Y_j}{Y_J}\right)$$

$$+0.5 \sum_{j=1}^{J}\sum_{l=1}^{J} a_{jl} \ln\left(\frac{Y_j}{Y_J}\right)\ln\left(\frac{Y_l}{Y_J}\right)$$

$$+\sum_{i=1}^{n} b_i \ln x_i + 0.5 \sum_{i=1}^{n}\sum_{p=1}^{n} b_{ip} \ln(x_i)\ln(x_p)$$

$$+\sum_{i=1}^{n}\sum_{j=1}^{J-1} g_{ij} \ln(x_i)\ln\left(\frac{Y_j}{Y_J}\right) + u + e \tag{19}$$

In Section V, a comparison will be made between the fitting obtained with the translong and the NDF specifications.

A. *Model Building*

Estimation of the NDF is based on the system of equations (3). The system is highly nonlinear in the parameters. 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 draws 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

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

$$j = 1,\ldots,J-1 \tag{20}$$

$$-\ln Y_{Jt}(x) = a_{0J} + \sum_{k=1}^{m_J} a_{kJ}\phi_J(\ln x \cdot \beta_{kJ}) + \ln\left(\frac{Y_{-J}}{Y_J}\right)\cdot\gamma_{-J}$$

$$+\ln x \cdot \xi + \delta_J \cdot t + e_{J,t} \tag{21}$$

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, independent and identically distributed (*i.i.d.*) as $N(0,\Sigma)$, 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

Seemingly Unrelated Regressions (SUR). This is feasible even for extremely large systems.

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 $\det\Sigma^{(i)}$.

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 α_{0j} ($j = 1,...,J$), α_{kj} ($j = 1,...,J-1$; $k = 1,...,m_j$), θ_j ($j = 1,...,J-1$), δ_j ($j = 1,...,J$).

B. Model Selection

Although it has been demonstrated that ANNs can approximate any nonlinear function with arbitrary accuracy, no accepted guideline exists in choosing the appropriate model for empirical applications. Consequently, the number of nodes *m* could be selected using the generalized R^2, $\tilde{R}^2$ goodness-of-fit criterion [43].

As is well known, R^2 is a statistical measure of how well the estimated line approximates the real data point and a value equal to 1 indicates perfect fit to the data. In this framework, $\tilde{R}^2$ is a modification of R^2 for systems of equations. According to this criterion one should select the number of nodes that maximizes $\tilde{R}^2$. When $\tilde{R}^2$ reaches a maximum one should stop adding explanatory terms.

V. EMPIRICAL APPLICATION

In this section, the NDF will be used to fit a large real data set. The data are taken from the commercial bank and bank holding company database managed by the Federal Reserve Bank of Chicago over the 1989-2000 time span. It is based on the Report of Condition and Income (Call Report) for all U.S. commercial banks that report to the Federal Reserve banks and the FDIC. The output variables are: (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). The input variables are: (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 data set is available to any researcher for purposes of replication upon request.

The estimation procedure described earlier was used to estimate the parameters $\left[a,\theta,\gamma,\xi\right] \in R^{J(n+1)+\sum_{j=1}^{J} m_j - 1}$. However, a choice has to be made regarding the number of nodes of the ANN. For reasons of convenience, we use the same number of nodes for each equation. In case this assumption is dropped, the procedure to be outlined remains unchanged. The $\tilde{R}^2$ value as a function of the number of nodes had a maximum value equal to $m = 1$ nodes (Fig. 1). Also, the determinant of the covariance matrix, $\det\Sigma^{(i)} \equiv \det\Sigma(\beta)$ had a minimum value for nodes (Fig. 2). Consequently, we set the number of nodes equal to $m_i = 1$ nodes ($i = 1,...,J$). In Table 1, the estimated coefficients for the NDF are shown along with their t-values (in parentheses).

Next, the RTS are calculated and are found to follow a Gaussian-like distribution around unity (Fig. 3). This results implies, roughly speaking, constant returns to scale, and can be characterized as expected because as a result of the optimization principle firms will generally exhibit constant returns to scale.

In Table 2, the estimates of the translog output distance function are presented. It can be inferred that the NDF is capable of explaining a very high proportion of the variance of output while using fewer parameters and satisfying all the theoretical properties dictated by production theory compared to the translog popular specification. Next, we compare over 1000 datasets the fitting of the two specifications, namely the NDF and the translog specification which is typically used for estimating the output distance function. In Figs. 4, 5, 6, 7 the values of $\tilde{R}^2$ and R^2 are depicted for these datasets.

Moreover, we checked the fitting of the NDF to data which were generated by the translog output distance function for 1000 datasets (see Figs. 8-9). Apparently, when fitted to a large number of data sets originating in US data for all commercial banks in the 1989-2000 time span, the NDF was found capable of explaining a very high proportion of the variance of output while keeping the number of parameters to a minimum and satisfying all the theoretical properties dictated by production theory in contrast to the translog specification. Also, when fitted to a large number of data sets generated by the translog specification, the NDF was found to capture a very high proportion of the variance of the data output. Consequently, the NDF is capable of capturing almost perfectly the translog specification while satisfying all the properties dictated by production theory globally and keeping the number of estimated parameters to a minimum. In this sense, it is superior to the most widely used competing specification.

VI. SUMMARY AND CONCLUSIONS

In this paper we have introduced a new output distance function, the Neural Distance Function (NDF), which constitutes a global approximation to any arbitrary production technology with many outputs given by a neural network specification. Relevant procedures have been developed and suggestions on implementation have been proposed. Thus, empirical implementation relied on standard techniques. All relevant measures such as returns to scale and total factor productivity have been computed routinely. The model is shown to possess all of the properties thought as desirable in conventional production theory, and to do so in a way not matched by any single competing specification. Fitted to a very large number of data sets originating in US data for all commercial banks in the 1989-2000 time span, the NDF is capable of explaining a very high proportion of the variance of output while keeping the number of parameters to a minimum

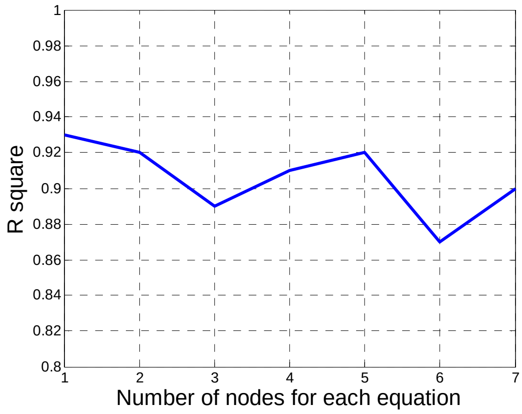

Figure 1: $\tilde{R}^2$ and the number of nodes.

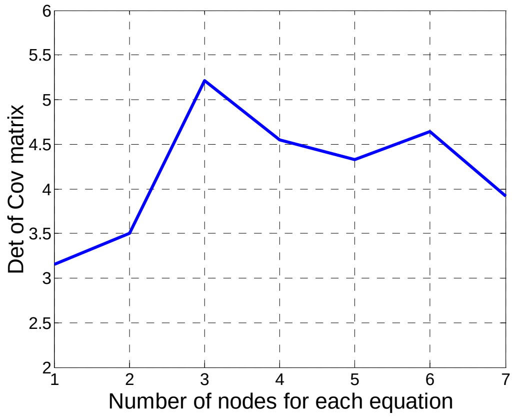

Figure 2: $\det \Sigma(\beta)$ and the number of nodes.

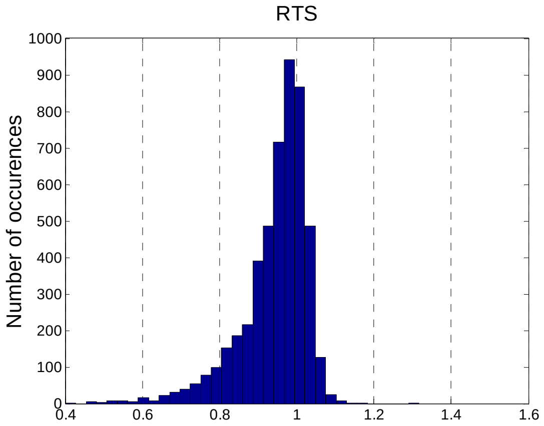

Figure 3: Returns to scale.

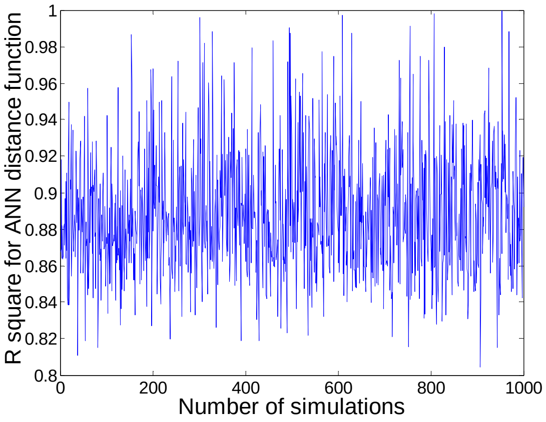

Figure 4: NDF fitting ($\tilde{R}^2$)

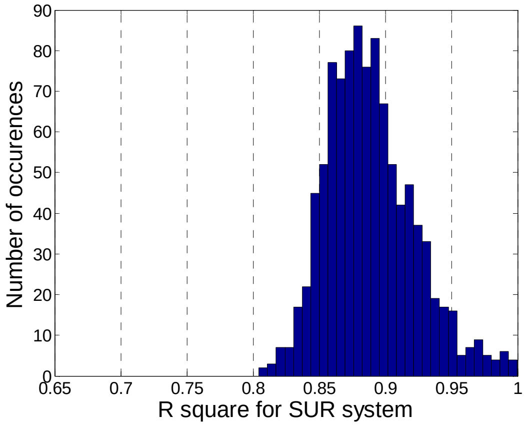

Figure 5: NDF fitting ($\tilde{R}^2$) – frequency

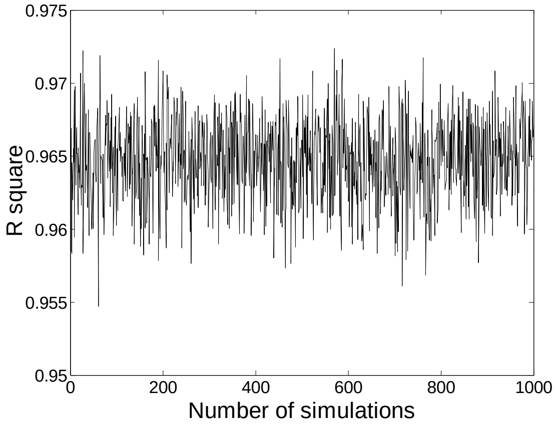

Figure 6: Translog fitting (R^2)

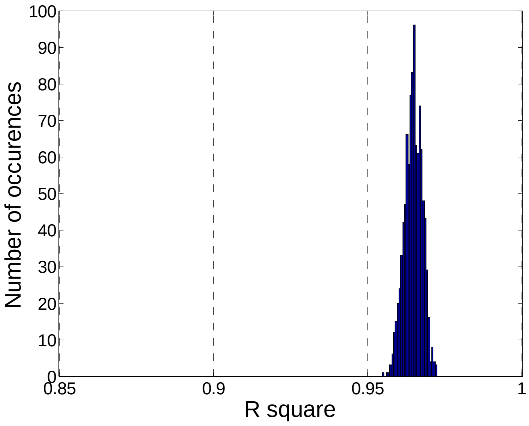

Figure 7: Translog fitting ($\tilde{R}^2$) – frequency

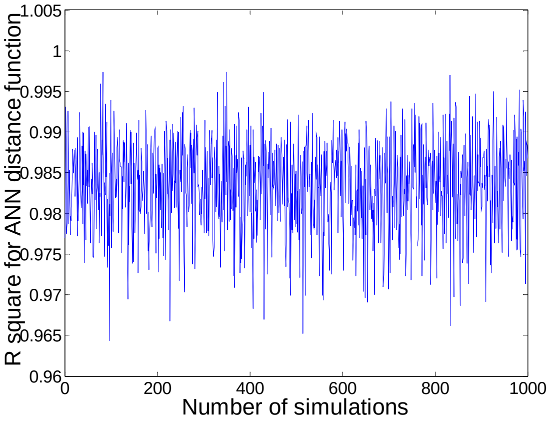

Figure 8: NDF fitting ($\tilde{R}^2$) against translog data

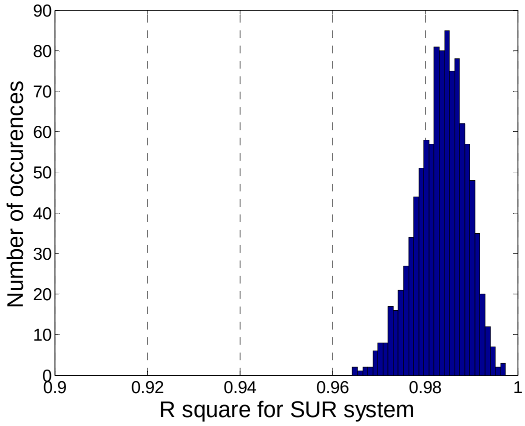

Figure 9: NDF fitting ($\tilde{R}^2$) against translog data - frequency

and satisfying all the theoretical properties dictated by production theory. The empirical estimation of the competing functional forms was less than satisfactory compared to the NDF. We believe that the NDF with its generality, conformity with theory and simplicity of structure is superior to the competing specifications and provides a vehicle for testing, expanding and improving conventional production theory.

Table 1: Neural distance function coefficients (t-statistic in parentheses)

	Equation 1	**Equation 2**	**Equation 3**	**Equation 4**	**Equation 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 (-4.9392)	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)	0.3760 (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					-0.024
$\tilde{R}^2$					0.9214

Table 2: Translog distance function estimate

Coefficient	**Estimate** (t-statistic in parentheses)
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	0.964

REFERENCES

[1] C. W. Cobb and P. H Douglas, “A Theory of Production”, American Economic Review, 18 (Supplement), pp. 139-165, 1928.

[2] L.R.Christensen, D.W. Jorgenson and L.J. Lau, “Conjugate duality and the transcendental production function”, Econometrica, 39: 206-218, 1971.

[3] A.R. Gallant and G.H Golub, “Imposing curvature restrictions on flexible functional forms”, Journal of Econometrics 26, pp. 295–321,1984.

[4] J. C. O’Donnell and J. T Coelli, “A Bayesian Approach to imposing curvature on distance functions”, Journal of Econometrics, Vol. 126, Issue 2, pp. 493-523, 2005.

[5] E. Diewert and J. T. Wales, “Flexible functional forms and global curvature conditions”, Econometrica, 55: 43-68, 1987.

[6] D. Terrell, “Incorporating Monotonicity and Concavity Conditions in Flexible Functional Forms”, Journal of Applied Econometrics, 11, pp. 179-194, 1996.

[7] G.P. Zhang and V.L. Berardi, “Time series forecasting with neural network ensembles: an application for exchange rate prediction”, Journal of the Operational Research Society 52: 652-664, 2001.

[8] E. Diewert, “An application of the Shephard duality theorem: A generalized Leontief production function”, Journal of Political Economy, 79: 481-507, 1971.

[9] A. R. Gallant, “Unbiased determination of production technologies”, J.Econometrics, 20: 285-323, 1982.

[10] A. R. Gallant, “On the bias in flexible functional forms and an essentially unbiased form: The Fourier flexible form”, J.Econometrics, 15: 211-246, 1981.

[11] M. Stinchcombe and H. White, “Using feedforward neural networks to distinguish multivariate populations”, Proceedings of the International Joint Conference on Neural Networks, 1994.

[12] S. Reinhard and G. Thijssen, “Resource Use Efficiency of Dutch Dairy Farms: A Parametric Distance Function Approach”, paper presented at Georgia Productivity Workshop III, Athens Georgia, October 1998.

[13] U. Anders, O. Korn, and C. Schmitt, “Improving the pricing of options: A neural network approach”, J. Forecasting, vol. 17: 369–388, 1998.

[14] Hanke, M., “Neural networks versus Black-Scholes: An empirical comparison of the pricing accuracy of two fundamentally different option pricing methods”, J. Comput. Intell. Finance, Vol. 5: 26–34, 1999.

[15] J. M. Hutchinson, A. W. Lo, and T. Poggio, “A nonparametric approach to pricing and hedging derivative securities via learning networks”, J. Finance, Vol. 49: 851–889, 1994.

[16] J. Y. Campbell, A. W. Lo, and A. C. MacKinley, The Econometrics of Financial Markets. Princeton, NJ: Princeton Univ. Press, 1997.

[17] J. N. Tsitsiklis and B. V. Roy, “Regression Methods for Pricing Complex American-Style Options”, IEEE Trans. Neural Networks, Vol. 12, No. 4: 694-703, 2001.

[18] M. Magdon-Ismail, “The Equivalent Martingale Measure: An Introduction to Pricing Using Expectations”, IEEE Trans. Neural Networks, Vol. 12, No. 4: 684-693, 2001.

[19] B. LeBaron and A. S. Weigend, “A Bootstrap Evaluation of the Effect of Data Splitting on Financial Time Series”, Letters IEEE Trans. Neural Networks, Vol. 9, No. 1: 213-220, 1998.

[20] A. Sfetsos and C. Siriopoulos, “Time Series Forecasting of Averaged Data With Efficient Use of Information”, IEEE Trans. Systems, Man, and Cybernetics – A, Vol. 35, No. 5, 738-745, 2005.

[21] H. White, H. and J. Racine, “Statistical Inference, The Bootstrap, and Neural-Network Modeling with Application to Foreign Exchange Rates”, IEEE Trans. Neural Networks, Vol. 12, No. 4: 657-673, 2001.

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

[23] C. M. Kuan and H. White, “Artificial neural networks: An econometric perspective”, Econometric Reviews, 13: 1-91, 1994.

[24] N. H. Chan and R. C. Genovese, “A Comparison of Linear and Nonlinear Statistical Techniques in Performance Attribution”, IEEE Trans. Neural Networks, Vol. 12, No. 4, July: 922-928, 2001.

[25] K. Hornik, M. Stinchcombe and H. White, “Multilayer feedforward networks are universal approximators”, Neural Networks, 2: 359-366, 1989.

[26] K. Hornik, M. Stinchcombe and H. White, “Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks”, Neural Networks, 3: 551-560, 1990.

[27] S. C. Kumbhakar and C. A. K. Lovell, Stochastic Frontier Analysis, Cambridge University Press, Cambridg, 2000.

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

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

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

[31] T. J. Coelli and S. Perelman, “Technical efficiency of European railways a distance function approach”, Applied Economics 32, pp. 1967–1976, 2000.

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

[33] B. Brummer, T. Glauben and G. Thijssen, “Decomposition of productivity growth using distance functions the case of dairy farms in three European countries”, American Journal of Agricultural Economics 84 (2002), pp. 628–644, 2002.

[34] J. R. Swinton, “At what cost do we reduce pollution? Shadow prices of SO2 emissions”, The Energy Journal 19, pp. 63–83, 1998.

[35] S. Grosskopf, D. Margaritis and V. Valdmanis, “Estimating output substitutability of hospital services a distance function approach”, European Journal of Operational Research 80 (1995), pp. 575–587, 1995.

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

[37] W. Meeusen, W., J. van den Broeck, “Efficiency estimation from cobb-douglas production functions with composed error”, International Economic Review 18, 435–444, 1977.

[38] L. J. Lau, “Testing and imposing monoticity, convexity and quasi-convexity constraints”, in: Fuss, M., McFadden, D. (Eds.), Production Economics: a dual approach to theory and applications, Amsterdam, North Holland, 1978.

[39] A. R. Rao and P. Bhimasankaram, Linear Algebra, McGraw Hill, New Delhi, 1992.

[40] C. Simon and L. Blume, Mathematics for Economists, New York: Norton and Company, 1994.

[41] H. Varian, Microeconomic Analysis, New York: W.W. Norton, 1992.

[42] S. Haykin, Neural Networks, New Jersey: Prentice-Hall, 1999.

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