Neural Networks for Approximating the Cost and Production Functions

Tsionas, Efthymios G. and Michaelides, Panayotis G. and Vouldis, Angelos

2008

Neural Networks for Approximating the Cost and Production Functions

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

Panayotis G. Michaelides and Angelos T. Vouldis
Department of Applied Mathematics and Physical Sciences
National Technical University of Athens
Athens, Greece
pmichael@central.ntua.gr and avouldis@biosim.ntua.gr

***Abstract*— Most business decisions depend on accurate approximations to the cost and production functions. Traditionally, the estimation of cost and production functions in economics relies on standard specifications which are less than satisfactory in numerous situations. However, instead of fitting the data with a pre-specified model, Artificial Neural Networks let the data itself serve as evidence to support the model's estimation of the underlying process. In this context, the proposed approach combines the strengths of economics, statistics and machine learning research and the paper proposes a global approximation to arbitrary cost and production functions, respectively, given by ANNs. Suggestions on implementation are proposed and empirical application relies on standard techniques. All relevant measures such as scale economies and total factor productivity may be computed routinely.**

***Keywords*— *Neural networks, Econometrics, Production and Cost Functions, RTS, TFP*.**

I. INTRODUCTION

Business decisions often depend on accurate approximations and analyses of the cost and production functions [1]. Commonly used specifications such as the Cobb-Douglas or the Translog are intuitively appealing and computationally straightforward. However, they are often less than satisfactory because they attempt to explain the complex variation in cost or production with a quite simple mathematical function despite the fact the real – world data are much more complicated. As a result their explanatory power is quite low. On the contrary, the nonparametric feature of Artificial Neural Networks (ANNs) makes them quite flexible and attractive in modelling economic phenomena where the theoretical relationship is not known a priory [2].

Consequently, instead of fitting the data with a pre-specified model, ANNs let the data itself serve as evidence to support (or reject) the model's estimation of the underlying process [2]. ANNs have found numerous applications in financial modelling [3]-[9]. However, with the exception of very few papers ([1], [10]) no systematic research on pure economic modelling using ANNs has been done.

This paper focuses on scholars and researchers in applied mathematics and attempts to combine tools from the statistical community with neural network technology. It proposes new flexible cost and production functions, respectively, which are based on ANNs allowing for multiple outputs. Contrary to widely used local approximations like the Translog [11], the generalized Leontief [12] or the symmetric McFadden form [13] the proposed flexible functions are global approximations to the unknown functions. The Fourier flexible form [14], [15] is also a global approximation but it requires an excessive number of parameters. The neural functions provide a better approximation using considerably less parameters [16].

II. ELEMENTS OF NEURAL NETWORKS

Neural networks are "data-driven, self-adaptive nonlinear methods that do not require specific assumptions about the underlying model" [2]. By combining simple units with multiple intermediate nodes, ANNs can approximate any smooth nonlinearity [17]. As demonstrated in Hornik et al. [17], [18], they have the ability to approximate arbitrarily well a large class of functions while keeping the number of free parameters to a minimum.

In mathematical terms, ANNs are collections of transfer functions that relate an output variable Y to certain input variables $X' = [X_1, \ldots, X_n]$. The input variables are combined linearly to form m intermediate variables $Z_1, \ldots, Z_m$ where

$$Z_i = X'\beta_i, \quad i = 1, \ldots, m \tag{1}$$

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

$$Y = \sum_{i=1}^{m} \alpha_i \phi(Z_i) \tag{2}$$

where ϕ is an activation function, the α_i 's are parameters and m is the number of intermediate nodes [19]. For various activation functions see, for instance, [19].

III. THE COST FUNCTION

In economics, the cost function is a function of input prices and output quantity and its value expresses the cost of producing that output given the input prices. Let $p \in R^n$ denote a price vector corresponding to n factors of production, and $y \in R_+^J$ the output vector. The neural cost function has the form:

$$\ln C(p,y) = \alpha_0 + \sum_{k=1}^{m} \alpha_k \phi(\ln p \cdot \beta_k + \ln y \cdot \gamma_k) + \ln p \cdot \theta \tag{3}$$

where $C(p,y)$ is the cost function, $a_k \in R, \beta_k \in R^n, \gamma_k \in R^J$ and $\theta \in R^n$ are parameters, and m is the number of intermediate nodes. For vectors a and b, $a \cdot b$ denotes the inner product.

Factor share equations are derived by (3) via formal differentiation with respect to prices using Shephard's lemma [20]:

$$w_i(p,y) = \frac{\partial \ln C(p,y)}{\partial \ln p_i} = \sum_{k=1}^{m} \alpha_k \beta_{ki} \phi'(\ln p \cdot \beta_k + \ln y \cdot \gamma_k) + \theta_i,$$

$$i = 1, \dots, n \tag{4}$$

In order for (3) to represent a proper cost function, $C(p,y)$ must be concave in p, which is expressed by the condition that the Hessian matrix $D^2 C(p)$ is negative semidefinite for every $p \in R_+^n$. Concavity is, traditionally, not imposed *a priori* but checked *a posteriori*.

A. *Returns to Scale*

In econometric studies, returns to scale describe what happens as the scale of production increases. Returns to scale refers to a technical property of production 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) [21].

The neural cost function does not place a priori restrictions on the behavior of returns to scale like other functional forms. It is known that if $\frac{\partial \ln C(p,y)}{\partial \ln y} < 1 \ (\geq 1)$ the production technology is characterized by increasing (non-increasing) returns to scale. For the neural cost function:

$$RTS = \sum_{i=1}^{J} \frac{\partial \ln C(p,y)}{\partial \ln y_i} = \sum_{i=1}^{J} \sum_{k=1}^{n} \alpha_k \gamma_{ki} \phi'(\ln p \cdot \beta_k + \ln y \cdot \gamma_k) \tag{5}$$

B. *Total Factor Productivity*

In economics, growth in total-factor productivity (TFP) represents output growth not accounted for by the growth in inputs [22] and presumably changes over time. It is traditionally used as a proxy for technical change.

If we modify (3) to include time (t) as an index of technical change, we have:

$$\ln C(p,y) = \alpha_0 + \sum_{k=1}^{m} \alpha_k \phi(\ln p \cdot \beta_k + \ln y \cdot \gamma_k + \delta_k t) + \ln p \cdot \theta \tag{6}$$

Therefore:

$$\frac{\partial \ln C(p,y)}{\partial t} = \sum_{k=1}^{m} \alpha_k \delta_k \phi'(\ln p \cdot \beta_k + \ln y \cdot \gamma_k + \delta_k t) \tag{7}$$

By definition, total factor productivity measure is given by $TFP = \frac{\partial \ln y}{\partial t}$. Since: $TFP = \frac{\partial \ln C(p,y) / \partial t}{\partial \ln C(p,y) / \partial \ln y}$ it follows that:

$$TFP = \frac{\sum_{k=1}^{m} \alpha_k \delta_k \phi'(\ln p \cdot \beta_k + \gamma_k \ln y + \delta_k t)}{\sum_{k=1}^{m} \alpha_k \gamma_k \phi'(\ln p \cdot \beta_k + \gamma_k \ln y + \delta_k t)} \tag{8}$$

Apparently, TFP as derived from the neural cost function is a weighted average of coefficients $\frac{\delta_k}{\gamma_k}$. The weights are normalized first-order derivatives of the activation functions at the different nodes of the neural network.

C. *Model Building*

Empirical estimation is based on the cost function and the system of share equations. The system is highly nonlinear in the parameters. Although the system is nonlinear in terms of the parameters β_k and γ_k the neural cost function's global approximation properties do not depend on this nonlinearity. As has been shown in [16], one may select the nonlinear parameters by a random search procedure, fix their values at the outcome of the random search, and estimate the linear parameters by the usual econometric methods. This will not affect the global approximation properties of the network. The weights are estimated and refit from scratch instead of being updated from previous data with a learning algorithm [18]. A modification of the Stinchcombe and White [16], procedure has to be followed here, because we have a system of equations instead of a single equation. The procedure is as follows:

Step 1: Let $\beta_k^{(i)}$ and $\gamma_k^{(i)}$ ($k = 1,..,m$) be drawn from a uniform distribution.

Step 2: Given these parameters, estimate α_k ($k = 1,..,m$) and θ by least squares applied to the cost function:

$$\ln C(p_t, y_t) = \alpha_o + \sum_{k=1}^{m} \alpha_k \phi(\ln p_t \cdot \beta_k + \gamma_k \ln y_t) + \ln p_t \cdot \theta + v_t,\ t = 1,..,T \tag{9}$$

where T denotes the number of observations, p_t the vector of factor prices of date t, and y_t the output level of date t.

Step 3: Compute the residual sum of squares $SSR^{(i)} \equiv SSR(\beta^{(i)}, \gamma^{(i)})$. Repeat for $i = 1,..,I$ and select the values $\overline{\beta}$ and $\overline{\gamma}$ that yield the minimum value of $SSR^{(i)}$.

Step 4: Estimate the following system of equations:

$$\ln C(p_t, y_t) = \alpha_o + \sum_{k=1}^{m} \alpha_k \phi(\ln p_t \cdot \overline{\beta}_k + \overline{\gamma}_k \ln y_t) + \ln p_t \cdot \theta + e_{o,t} \tag{10a}$$

$$w_{it} = \sum_{k=1}^{m} \alpha_k \overline{\beta}_{ki} \phi'(\ln p_t \cdot \overline{\beta}_k + \overline{\gamma}_k \ln y_t) + \theta_i + e_{i,t},\ i = 1,..,n-1 \tag{10b}$$

where $e_t \equiv [e_{0t}, e_{1t}, ..., e_{n-1,t}]'$ is a vector random variable, distributed as i.i.d. $N(0, \Sigma)$ where Σ is a covariance matrix. System (10a) and (10b) is linear in the parameters $[\alpha, \theta] \in R^{n+m}$ and can be estimated using standard, iterative seemingly unrelated regressions equations technique (SURE) [23]. This is feasible even for extremely large systems.

IV. The Production Function

Let $x \in R^n$ denote an input vector corresponding to n factors of production, and $Y \in R_+^J$ the output vector. The neural production function, for each output, has the form:

$$\ln Y_i(x) = a_{0i} + \sum_{k=1}^{m_i} \alpha_{ki} \phi_i(\ln x \cdot \beta_{ki}) + \ln x \cdot \theta_i$$
$$i = 1,...,J-1 \tag{11}$$

where $Y_i(x)$ is the production function of output *i*, $a_{ki} \in R, \beta_{ki} \in R^n, \theta_i \in R^n$ are parameters and m_i is the number of intermediate nodes. For the last output J the equation governing its production process has the following form:

$$\ln Y_J(x) = a_{0J} + \sum_{k=1}^{m_J} \alpha_{kJ} \phi_J(\ln x \cdot \beta_{kJ}) + \ln Y \cdot \gamma + \ln x \cdot \xi \tag{12}$$

where $\gamma \in R^J, \xi \in R^n$ are parameters, and m_J is the number of intermediate nodes for output J.

In addition, for (11) to represent a proper production function $Y_i(x)$ must be increasing in *x* and $Y_J(x)$ decreasing in Y. Also, quasi-concavity of $Y_i(x)$ and $Y_J(x)$ is implied by economic theory. These assumptions are not imposed *a priori* but rather checked *a posteriori*. Finally, $Y_J(x)$ must be homogeneous of degree one, a fact which places parametric restrictions on the production function. More precisely, homogeneity of degree one implies:

$$\sum_{j=1}^{J} \gamma_j = 0 \tag{13}$$

A. Returns to Scale

As we have seen, returns to scale (RTS) describe what happens as the scale of production increases. The neural production function does not place a priori restrictions on the behavior of returns to scale. It is know that typically the RTS are equal to the sum of the output elasticities of the various inputs. Let ε^j denote the elasticity of output with respect to factor x ^{j}:

$$\varepsilon^j = \frac{\partial \mathrm{Y}(x)}{\partial x_j} \cdot \frac{x_j}{Y(x)} = \frac{\partial \ln Y(x)}{\partial \ln x_j}, \quad j = 1,...,n \tag{14}$$

where $x \in R^n$ denotes the input vector corresponding to n factors of production.

Therefore, for the neural production function RTS for each output are equal to:

$$RTS^i = \sum_{j=1}^{n} \frac{\partial \ln Y_i(x)}{\partial \ln x_j}, \quad i = 1,...,J-1 \tag{15}$$

Consequently:

$$RTS^i = \sum_{j=1}^{n}\sum_{k=1}^{m_i}\beta_{kj} a_{ki}\phi_i'(\ln x\cdot\beta_{ki}) + \sum_{q=1}^{n}\theta_q,\ i=1,...,J-1,\quad j=1,...,n \tag{16}$$

For the last output J, we have:

$$RTS^J = \sum_{j=1}^{n}\sum_{k=1}^{m_J}\beta_{jkJ} a_{kJ}\phi_J'(\ln x\cdot\beta_{kJ}) + \sum_{i=1}^{J-1}\gamma_i(\sum_{j=1}^{n}\sum_{k=1}^{m_i}\beta_{kj} a_{ki}\phi_i'(\ln x\cdot\beta_{ki})) + \sum_{q=1}^{n}\xi_q \tag{17}$$

B. Total Factor Productivity

If we modify (11) to include time (t) as an index of technical change, we have:

$$\ln Y_{it}(x) = a_{0i} + \sum_{k=1}^{m_i}\alpha_{ki}\phi_i(\ln x\cdot\beta_{ki} + \delta_{ki}t) + \ln x\cdot\theta_i$$

$$i = 1,\ldots,J-1 \tag{18}$$

By definition Total Factor Productivity (TFP) measure, for each output, is given by:

$$TFP_{it} = \frac{\partial \ln Y_{it}(x)}{\partial t} \tag{19}$$

Therefore, it follows that:

$$TFP_{it} = \sum_{k=1}^{m_i}\delta_{ki} a_{ki}\phi_i'(\ln x\cdot\beta_{ki} + \delta_{ki}t),\quad i=1,...,J-1 \tag{20}$$

For the last output J, we have:

$$TFP_{Jt} = \sum_{k=1}^{m_J}\delta_{kJ} a_{kJ}\phi_J'(\ln x\cdot\beta_{kJ} + \delta_{kJ}t) +$$

$$\sum_{i=1}^{J-1}\gamma_i(\sum_{k=1}^{m_i}\delta_{ki}\alpha_{ki}\phi_i'(\ln x\cdot\beta_{ki} + \delta_{ki}t)) \tag{21}$$

We can see that TFP depends on time and inputs.

C. Model Building

Similarly to the cost function, estimation is based on the system of production functions (11) – (12). The system is highly nonlinear in the parameters. The procedure is, practically, the same as earlier:

Step 1: Let $\beta_k^{(i)}$ be drawn from a uniform distribution.

Step 2: Given these parameters, estimate $\alpha_k^{(i)}$, $\gamma_k^{(i)}$, $\theta^{(i)}$ and $\xi^{(i)}$ by means of the system:

$$\ln Y_{it}(x_t) = a_{0i} + \sum_{k=1}^{m_i}\alpha_{ki}\phi_i(\ln x_t\cdot\beta_{ki}) + \ln x_t\cdot\theta_i + e_{i,t} \tag{22a}$$

$$\ln Y_J(x_t) = a_{0J} + \sum_{k=1}^{m_J}\alpha_{kJ}\phi_J(\ln x_t\cdot\beta_{kJ}) + \ln y_t\cdot\gamma$$
$$+ \ln x_t\cdot\xi + e_{J,t}\quad i = 1,\ldots,J-1 \tag{22b}$$

where x_t denotes the vector of inputs of date t, y_t the output levels of date t, $e_t \equiv [e_{0t}, e_{1t}, ..., e_{J,t}]'$ is a vector random variable, distributed as i.i.d. $N(0,\Sigma)$, Σ is a covariance matrix. The system of equations (22a) and (22b) is linear in the parameters $\alpha_k^{(i)}$, $\gamma_k^{(i)}$, $\theta^{(i)}$ and $\xi^{(i)}$ and can be estimated using standard, iterative SURE. 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 $\overline{\beta}$ that yield the minimum value of $\det\Sigma^{(i)}$.

Step 4 : For $\overline{\beta}$ that yield the minimum value of $\det\Sigma^{(i)}$ re-estimate the system and keep the estimated values for parameters $\alpha_k^{(i)}$, $\gamma_k^{(i)}$, $\theta^{(i)}$ and $\xi^{(i)}$.

V. 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 [2]. Consequently, the number of nodes m could be selected using one of the following methods: (a) the ${R_{adj}}^2$ criterion, (b) Schwartz's criterion [24] or (c) Akaike's criterion [25].

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, ${R_{adj}}^2$ is a modification of R^2 that adjusts for the number of explanatory terms in a model, i.e. the number of independent variables and the number of data points. According to this very popular criterion in model selection one should select the number of nodes that maximizes the ${R_{adj}}^2$. When ${R_{adj}}^2$ finds a global maximum one should stop adding explanatory terms [18].

According to the Bayesian Information Criterion or the so-called Schwartz's criterion [25], one should select the number of nodes that minimizes the BIC which is defined as:

$$BIC = -2\ln(L) + k\ln(n) \tag{23}$$

where n is the number of observations, k is the number of free parameters to be estimated and L is the maximized value of the likelihood function for the estimated model. The BIC minimizing model keeps a balance between bias and variance, in that additional complexity must be justified by a correspondingly large improvement in fit. BIC has been shown to be statistically consistent [18].

According to Akaike [26], one should determine the number of nodes that minimizes the AIC criterion defined as:

$$AIC = 2k - 2\ln(L) \tag{24}$$

where k is the number of free parameters to be estimated and L is the maximized value of the likelihood function for the estimated model. The *AIC* rewards the goodness of fit but also includes a penalty that is an increasing function of the number of parameters.

Finally, it should be noted that the algorithm for randomly drawing parameters from a hyper-rectangle to estimate the cost and production functions shall be refined by means of more sophisticated optimization techniques in case of very large dimensional problems.

VI. Empirical Results

A. Data and Variables

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. The dataset 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.

B. Results for the Cost Function

We followed the procedure described earlier and estimated the parameters $[\alpha, \theta] \in R^{n+m}$. However, the desirable number of nodes m also has to be selected using one of the methods described earlier. ${R_{adj}}^2$ criterion is depicted in Fig. 1 whereas Schwartz's (1978) and Akaike's (1973) criteria are depicted in Fig. 2.

It is clear that the BIC finds a global minimum for m=7 while the Akaike criterion, which punishes less strictly the increase in the number of nodes, finds also other local minimums for greater numbers of nodes. However, even for the Akaike criterion m=7 is the global minimum. Also, the R^2 and ${R_{adj}}^2$ find a global maximum for m=7 nodes. So, for an ANN with m=7 modes and activation function $f(x) = (1+e^{-x})^{-1}$ the estimated coefficients α, θ are statistically significant for almost all of the estimated coefficients.

Next, the Returns to Scale are computed through equation (5) and are found to follow a Gaussian-like distribution around unity (1). This result implies, roughly speaking, constant returns to scale and can be characterized as expected (see Figs. 3-4) because, as is well known, as a result of the optimization principle the production function for the firm will generally exhibit constant returns to scale.

The factor shares of the five (5) inputs were calculated and were found to range between 0 and 1, as expected.

Subsequently, the issue of concavity is investigated. As it has already been mentioned, the concavity condition can be checked by calculating the eigenvalues of the Hessian matrix for each observation and examining if they are all negative. It was confirmed that the vast majority of eigenvalues are negative implying that the cost function is, practically, globally concave with respect to prices, a result which is consistent with economic theory [21]. For each observation there were five eigenvalues equal to the dimension of the Hessian matrix.

More precisely, for each observation, the four greater (in absolute value) eigenvalues were negative. Also, the lower eigenvalues for each observation have generally a much greater absolute value than its most positive eigenvalue. In total, approximately 90% of all eigenvalues were found to be negative. Any devation from this rule can be attributed to omitted variables, measurement errors, and inefficiency. A failure of the proposed functional form to comply with this assumption would imply empirical findings non-consistent with neo-classical economic theory. However, not all cost functions proposed, so far, in the empirical literature satisfy this assumption, despite it being dictated by economic theory.

Finally, in Fig. 5, the histogram of all TFP values (%) is depicted. We see that TFP is negative on the average with a longer tail to the left indicating the prevalence of negative technical progress for the organizations of the US Banking sector in the 1989-2000 time span.

B. Results for the Production Function

The estimation procedure described earlier was used to estimate the parameters $[a, \theta, \gamma, \xi] \in R^{J(n+1)+\sum_{i=1}^{J} m_i - 1}$. However, a choice has to be made regarding the number of nodes of the neural network. The system R^2_{wide} had a maximum for

$m_i = 3$ nodes (Fig. 6). Consequently, for the rest of our analysis of production functions we set $m_i = 3$ $(i = 1,...,J)$. As it can be inferred from the value of the R^2_{wide}, the neural network production function provides a very good approximation to the actual production function. Also, almost all of the estimated coefficients of the production functions were statistically significant.

Next, the RTS are also calculated and the results are shown in Fig. 7.

The histogram of the TFP values is depicted in Fig. 8.

Finally, the hypothesis that $Y_J(x)$ is increasing in *x*, decreasing in $Y_i(x)$, for $i = 1,...,J-1$, $i \neq j$ and the quasi-concavity of $Y_i(x)$ and $Y_J(x)$ were checked *ex post* and were found to be, in general terms, consistent with economic theory.

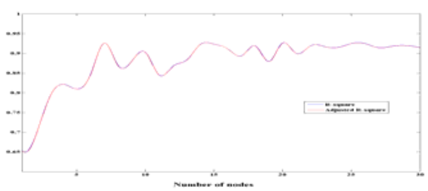

Figure 1. R^2 and $R_{adj}{}^2$ and the number of number of nodes

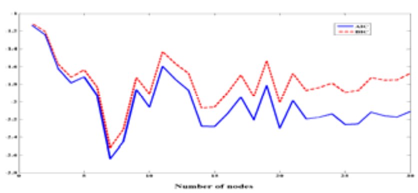

Figure 2. Akaike's Information Criterion, Bayesian Information Criterion and the number of nodes

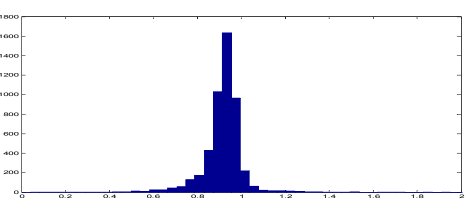

Figure 3. Histogram of RTS (Unconstrained regression)

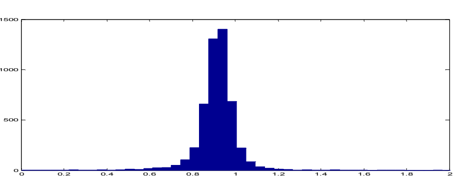

Figure 4. Histogram of RTS (Constrained regression)

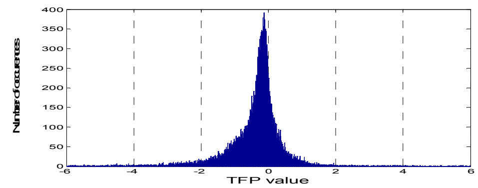

Figure 5. Histogram of TFP values

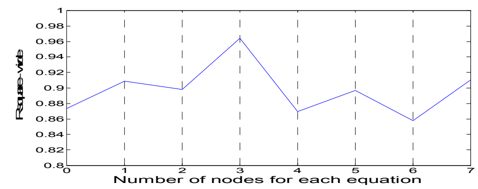

Figure 6. $\tilde{R}^2$ and the number of nodes

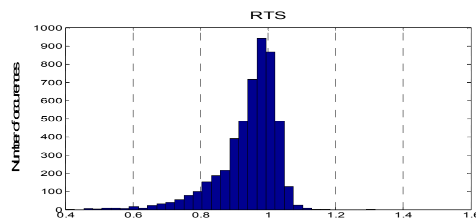

Figure 7. Histogram of RTS for the *J*th output

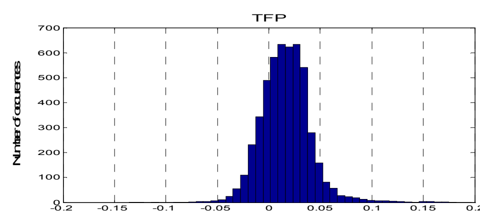

Figure 8. Histogram of TFP for the *J*th output

VII. CONCLUSIONS

Commonly used production and cost functions usually estimated by means of linearized multifactor models are known to be less than satisfactory in numerous situations. However, ANNs let the data itself serve as evidence to support the model's estimation of the underlying process. In this context, the proposed procedure attempted to combine the strengths of economics, statistics and machine learning research. The paper proposed a global approximation to arbitrary cost and production functions, respectively, given by ANN specifications. All relevant measures such as scale economies and total factor productivity were computed routinely. The empirical application referred to a large panel data set consisting of all U.S. commercial banks that report to the Federal Reserve banks over the time period 1989-2000. The results of the empirical implementation were consistent with conventional economic theory.

REFERENCES

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

[2] 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, pp. 26–34, 1999.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

[20] Shephard, R., *Cost and Production Functions*, Princeton, NJ: Princeton University Press, 1953.

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

[22] Hornstein, A. and Krusell, P., “Can Technology Improvements Cause Productivity Slowdowns?”, *NBER Macroeconomics Annual 11*, MIT Press: Cambridge MA, pp. 209-259, 1996.

[23] Greene, W., *Econometric Analysis*, New Jersey: Prentice Hall, 2003.

[24] Schwartz, G., “Estimating the dimension of a model”, *Annals of Statistics* 6 (2), pp. 461-464, 1978.

[25] Akaike, H., “A new look at the statistical model identification”, *IEEE Trans. Automatic Control,* Vol. 19, No. 6, pp. 716–723, 1974.