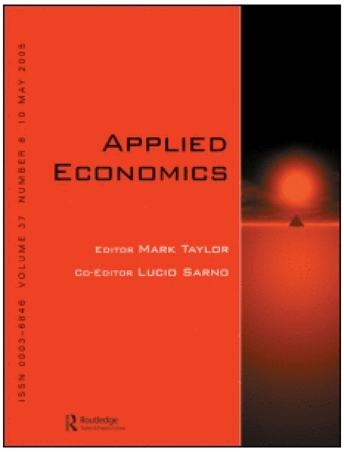

Applied Economics
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713684000

The measurement of technical efficiency: a neural network approach

Daniel Santín [a]; Francisco J. Delgado [b]; Aurelia Valiño [a]
[a] Department of Applied Economics VI, Complutense University of Madrid, [b] Department of Economics, University of Oviedo,

Online Publication Date: 10 April 2004

To cite this Article Santín, Daniel, Delgado, Francisco J. and Valiño, Aurelia(2004)'The measurement of technical efficiency: a neural network approach',Applied Economics,36:6,627 — 635

To link to this Article: DOI: 10.1080/0003684042000217661

URL: http://dx.doi.org/10.1080/0003684042000217661

Applied Economics, 2004, **36**, 627–635

Routledge
Taylor & Francis Group

The measurement of technical efficiency: a neural network approach

DANIEL SANTÍN*†, FRANCISCO J. DELGADO‡ and AURELIA VALIÑO†

†*Department of Applied Economics VI, Complutense University of Madrid* and ‡*Department of Economics, University of Oviedo*

The main purpose of this paper is to provide an introduction to artificial neural networks (ANNs) and to review their applications in efficiency analysis. Finally, a comparison of efficiency techniques in a non-linear production function is carried out. The results suggest that ANNs are a promising alternative to traditional approaches, econometric models and non-parametric methods such as data envelopment analysis, to fit production functions and measure efficiency under non-linear contexts.

I. INTRODUCTION

A wide range of statistical and econometric techniques exists to apply in economics, where the complex reality must be modelled. Artificial neural networks (ANNs) are relatively new techniques that have been applied with success in a variety of disciplines: speech and image recognition, engineering, robotics, meteorology, banking, stock markets, etc.

ANNs have their origins in the study of the complex behaviour of the human brain. McCulloch and Pitts (1943) introduced simple models with binary neurons. Then, Rosenblatt (1958) proposed the multi-layer structure with a learning mechanism based on the work of Hebb (1949), the so-called *perceptron*, and the first neural networks applications began with Widrow (1959).

However, Minsky and Papert (1969) pointed out that a two-layer perceptron was unable to solve the logical XOR (a basic non-linear problem). After a decay in neural networks researching during the 1970s, the work by Rumelhart *et al.* (1986) had an important role in the growth of this technique. They rediscovered the most used learning algorithm, the so-called backpropagation algorithm (BP), together with the use of a three layer perceptron. This neural network was able to deal with non-linear problems.

Although ANNs arose to model the brain, they have been applied when there is no theoretical evidence about the functional form. In this way, ANNs are data-based, not model-based.

The paper is organized as follows. Section II provides an introduction to ANNs. Their advantages and drawbacks are revised too. Section III is dedicated to ANNs in efficiency analysis, where neural networks form a promising analysis tool together with known econometric models such as stochastic frontier analysis (SFA) and non-parametric methods such as data envelopment analysis (DEA). This section concludes with a review of some published papers about ANNs and efficiency. A simulation procedure is carried out in Sections IV and V to compare several efficiency techniques in a non-linear production function context. The final section of the paper offers conclusions and suggests areas for future research.

II. ARTIFICIAL NEURAL NETWORKS: AN OVERVIEW

There has been a vast literature about ANNs, basically in the empirical field, since the middle 1980s. In this section

*Corresponding author. E-mail: dsantin@ccee.ucm.es

Applied Economics ISSN 0003–6846 print/ISSN 1466–4283 online © 2004 Taylor & Francis Ltd
http://www.tandf.co.uk/journals
DOI: 10.1080/0003684042000217661

theoretical background is supplied.[1] ANNs are normally arranged in three layers of neurons, the so-called multilayer structure:

- Input layer: its neurons (also called nodes or processing units) introduce the model inputs.
- Hidden layer(s) (one or more layers): its nodes combine the inputs with weights that are adapted during the learning process.
- Output layer: this layer provides the estimations of the network.

Another breaking point in the neural history was 1989. Several authors published this year that ANNs are universal approximators of functions (Carroll and Dickinson, 1989; Cybenko, 1989; Funahashi, 1989; Hecht-Nielsen, 1989; Hornik *et al.*, 1989; White, 1990). Later, it was demonstrated that ANNs could also approximate their derivates (Hornik *et al.*, 1990). These results justified the forward success reached in applications. Scarselli and Chung (1998) provide an actual and complete review of this property.

Among the different networks, the feedforward neural networks or *multilayer perceptron*[2] (MLP) are the most commonly used. In these networks, the output[3] is function of the linear combination of hidden units activations, each of one is a non-linear function of the weighted sum of inputs. In this way, from:

$$y = f(\mathbf{x}, \theta) + \varepsilon \tag{1}$$

where $\mathbf{x}$ is the vector of explanatory variables, ε the error component (assummed independently and identically distributed, with zero mean and constant variance), $f(\mathbf{x}, \theta) = \hat{y}$ is the unknown function to estimate from the available information, the network consists of:

$$\hat{y} = f(\mathbf{x}, \theta) = F\left(\beta_o + \sum_{j=1}^{m} G(\gamma_j + \sum_{i=1}^{n} x_i\alpha_{ij})\beta_j\right) \tag{2}$$

where:

$\hat{y}$: network output
F: output layer activation function
G: hidden layer activation function
n : number of input units
m: number of hidden units
$\mathbf{x}$: inputs vector (i = 1 . . . n)
θ: weights vector (parameters):
β_o: output bias
γ_j : hidden units biases (j = 1 . . . m)
α_{ij}: weight from input unit i to hidden unit j
β_j : weights from hidden unit j to output

From Equation 2, it can be observed that MLPs are mathematical models often equivalent to traditional models in econometrics such as linear regression, logit, AR models for time series analysis..., but with specific terminology and estimation methods (Cheng and Titterington, 1994). For example, in time series analysis, it is possible to predict the value of a variable y at the moment t, y_t, from past observations, y_{t-1}, y_{t-2}, $y_{t-3}, \ldots$; then the network is a non linear autoregressive model:

$$\hat{y}_t = F\left(\beta_o + \sum_{j=1}^{m} G(\gamma_j + \sum_{i=1}^{n} y_{t-i}\alpha_{ij})\beta_j\right) \tag{3}$$

Figure 1 represents a MLP with three layers and one output. The activation function for output layer is generally linear. The logistic function is used for classification purposes. The non-linear feature is introduced at the hidden transfer function. From the previous universal approximation studies, these transfer functions must have mild regularity conditions: continuous, bounded,

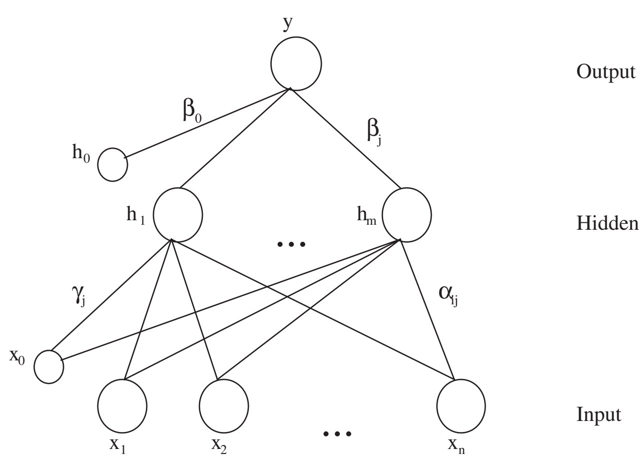

Fig. 1. *Single-output, three layer feed-forward neural network MLP(n,m,1)*

[1] For more details it can be consulted Hertz *et al.* (1991), Bishop (1995) and Ripley (1996). White (1989a) exposes a detailed statistical analysis of the neural learning, BP included. In Cheng and Titterington (1994) ANNs and traditional statistical models are shown together (with discussion). Kuan and White (1994) exhibit ANNs as non-linear models, with an asintothic theory of the neural learning. Zapranis and Refenes (1999) review the model identification and selection with many examples from financial economics. They conclude with an interesting case of study totally developed. Zhang *et al.* (1998) revise the enormous literature about forecasting with ANNs.
[2] Other networks are Radial Basis Functions Networks, relate to cluster and principal component analysis. The Recurrent Networks are extensions of the feed-forward networks, because they incorporate feedbacks, such as the Jordan and Elman networks (Kuan and Liu, 1995).
[3] For simplicity, we consider one output, but it is easy to extend to various outputs.

differentiable and monotonic increasing. The most popular transfer function is sigmoid or logistic,[4] nearly linear in the central part. The transfer functions[5] bound the output to a finite range, [0,1] in the sigmoidal function:

$$G : \Re \to [0,1] \Big| G(a) = \frac{1}{1+e^{-a}}, \quad a \in \Re \tag{4}$$

Augmented single layer networks incorporate direct links between input and output layers with a linear term. Kuan and White (1994) explained that '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 than as alternatives to, the familiar models':

$$\hat{y} = f(\mathbf{x}, \theta) = F\left(\beta_o + \sum_{i=1}^{n} x_i \alpha_i + \sum_{j=1}^{m} G(\gamma_j + \sum_{i=1}^{n} x_i \alpha_{ij})\beta_j\right) \tag{5}$$

From Equation 5 note that if $\beta_j = 0, j = 1 \ldots m$, and if F is linear, the network is a linear model. Hence White (1989a) implemented a neural network test for non linearity. This test is compared with other similar tests by Lee *et al.* (1993).

Architecture selection is one major issue with implications on the empirical results and consists of:

1. Data transformation.
2. Input variables and number, n.
3. Hidden units number, m.
4. Hidden and output activation function.
5. Weight elimination or *pruning*.

All are open questions today and there are many answers to each one. Data transformation is a common issue: [0,1] or [a,b] normalization, detrended and/or deseasonalized data in time series analysis. The hidden units number is determined by a trial-error[6] process considering $m = 1, 2, 3, 4 \ldots$. Finally, it is common to eliminate 'irrelevant' inputs or hidden units (White, 1989b).

Another critical issue in ANNs is the neural learning or model estimation, based upon searching the weights that minimize some cost function such as square error:

$$\min_{\theta \in \Theta} \left[E(y - f(x, \theta))^2 \right] \tag{6}$$

The most popular process is the BP algorithm:

$$\theta(k+1) = \theta(k) + \eta \frac{\partial E}{\partial \theta}(k) \tag{7}$$

$$\theta(k+1) = \theta(k) + \eta \nabla f(x, \theta)[y - f(x, \theta)] \tag{8}$$

BP is an iterative process (k indicates iteration). Parameters are revised from the error function (E) gradient by the learning rate η, constant or variable. The error propagates backwards to correct the weights until some stop criterion – iteration number, error – is reached. BP has been criticized because of slow convergence, local minimum problem and sensitivity to initial values and η. Schiffmann *et al.* (1992) proposed some improvements.[7]

After neural training (training set), new observations (validation and/or test sets) are presented to the network to verify the so-called generalization capability. Here it is relevant the statistical classical bias–variance dilemma (Geman *et al.*, 1992) or overfitting problem.

ANNs have advantages, but logically they also have several drawbacks (Fig. 2). Therefore, ANNs can learn from experience and can generalize, estimate, predict, with few assumptions about data and relationships between variables. Hence, ANNs have an important role when these relationships are unknown (non-parametric method) or non-linear (non-linear method), provided there are enough observations (flexible form and universal approximation property). However, the flexibility can conduct to learn the noise, and data are not very large in economic series. These restrictions promotes the search of parsimonious models. Finally, algorithm convergence and trial and error process are some relevant drawbacks too.

III. ANNs AND EFFICIENCY

Efficiency analysis (Farrell, 1957) is a relevant field in economics. The appropriate use of few resources with the available technology is referred to as technical efficiency. When the technology is not fixed, input combination is searched, and the problem is the so-called allocative efficiency. Fried *et al.* (1993) and Álvarez (2001) are excellent references for a review of the techniques and applications in the measurement of productive efficiency.

In this analysis, a key issue is the frontier function estimation. This estimation can be carried out following three

[4] In networks without hidden layer, the output can be interpreted as 'a posteriori' probabilities relate to discriminant functions. With hidden layer, one can interpret the outputs as conditional probabilities (Bishop, 1995).
[5] Another frequent function is $\tan h$: $\Re \Rightarrow [-1,1] \big| G(a) = (e^a - e^{-a})/(e^a + e^{-a})$. This function differs from sigmoidal (4) in a linear transformation, $\tan h(a/2) = 2\, sigma(a) - 1$, and occasionally it can achieve faster convergence (Bishop, 1995).
[6] Common criteria for model selection are SIC (*Schwartz Information Criterion*) or AIC (*Akaike Information Criterion*).
[7] One alternative consists of adding a term called momentum:

$$\theta(k+1) = \theta(k) + \eta \frac{\partial E}{\partial \theta}(k) + \mu \Delta\theta(k-1)$$

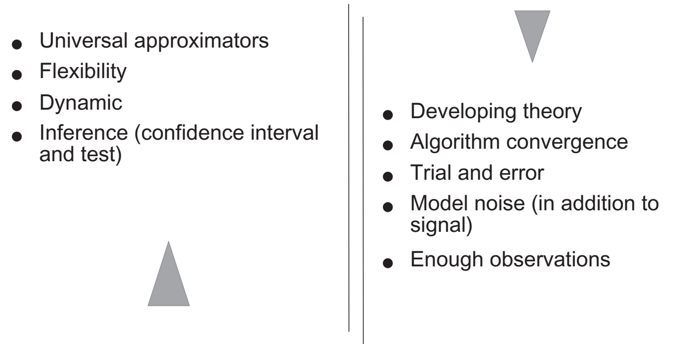

Fig. 2. *ANNs' advantages and drawbacks*

alternatives, parametric and non parametric techniques, and neural networks:

- Parametric methods: a functional form is adopted such as Cobb-Douglas, translog, CES, Leontief generalized. Parametric techniques can be deterministic or stochastic:
  - Deterministic: frontier deviations are explained because of inefficiency.
  - Stochastic: frontier deviations are decomposed into noise – usually semi-normal- and inefficiency components (Aigner *et al.*, 1977).

Estimations can be done by COLS (corrected ordinary least squares), or maximum likelihood. In COLS independent term is corrected by adding the largest positive error from initial OLS.

- Non parametric techniques: no functional form is assumed:[8]
  - Data envelopment analysis (DEA), Charnes *et al.* (1978). A deterministic frontier is formed by enveloping the available data using mathematical programming. Constant/variable returns to scale and input/output combinations convexity are common assumptions.
- Neural networks: from the following general expression:

$$y_i = f(\mathbf{x}_i, \theta) + \varepsilon_i - u_i \tag{9}$$

where $u_i \geq 0$ is technical inefficiency, we can adopt the network (2) to estimate the frontier. Costa and Markellos (1997) proposed two procedures: a) similar way to COLS after neural training; b) by an oversized network until some signal to noise ratio is reached. Then, inefficiency is determined as observation-frontier distance.

ANNs are flexible, non parametric (free-model) and stochastic techniques, and it is theoretically possible to make statistical inference such as interval confidence[9] to inefficiency indexes. However, ANNs have no theoretical studies in efficiency analysis and few applications have been made in this field. Moreover, results are not easily interpretable and many technical resources are needed. As was expected, no technique is superior to the rest, and the nature of the particular problem will determine the most appropriate one. The comparison of efficiency measurement approaches is summarized in Table 1 (partially based on Costa and Markellos, 1997). Table 2 summarizes the principal publications about ANNs and efficiency.

Finally, are ANNs 'efficient' techniques in efficiency analysis today? Clearly, much work remains to be done in this area. At the present time, the answer is uncertain. The future answer to that question will be the result of the balance between costs (knowledge, model complexity, algorithms, economic interpretation, ...) and benefits (better results, decision making, flexibility...).

IV. THE EXPERIMENT

In order to examine the performance of efficiency techniques, let $F(x)$ be the further one input-one output non-linear continuous production function:

$$F(x) = \begin{cases} \left(\dfrac{x}{e}\right)^2 & \text{if } x \in [0, e] \\ \mathrm{Ln}\,(x) & \text{if } x \in [e, e^2] \\ A^* \cos(x - e^2) + 2 - A & \text{if } x \in [e^2, e^2 + \pi], \\ \quad \text{where } A = 0.25 & \\ \mathrm{Ln}\,(x - 2\pi) & \text{if } x \in [e^2 + \pi, 26] \end{cases} \tag{10}$$

Through this production function (see Fig. 3) we introduce all returns to scale possibilities. The first part of Equation 10 presents increasing returns to scale (IRS). The second and fourth sections show decreasing returns to scale (DRS). The third section presents a not common theoretical technology where an increase in one input implies a decrease in one output. According to Costa and Markellos (1997) this phenomenon can be called a 'congested area'.

However, the intention here is to illustrate what occurs with efficiency estimations when the 'traditional linear models' are not the real production functions for the

[8] Another non-parametric technique is FDH, Free Disposal Hull.
[9] Confidence intervals in general neural network framework are proposed and revised by Hwang and Ding (1997), De Veaux *et al.* (1998) and Rivals and Personaz (2000).

Table 1. *Efficiency measurement techniques*

Comparative factor	Econometrics	DEA	ANNs
Assumptions: functional form, data...	Strong	Medium	Low
Flexibility	Low–medium	Medium	High
Theoretical basis	Strong	Strong	Medium
Theoretical studies and applications in efficiency	Yes	Yes	Few
Statistical significance	Yes	Yes	Yes
Interpretability of results	Medium	Low	Medium
Estimation/prediction	High	No	High
Costs: software, estimation time...	Low	Low	High

Table 2. *Summary of publications about ANNs and efficiency*

Joerding *et al*. (1994) Production function	• Theoretical properties imposition about technology – positivity, monotonicity, quasiconcavity. • ANNs similar to Fourier flexible form. • Simultaneous estimation of production function and inputs demand system. • Not possible to impose Constant Returns to Scale in all **x** because linear activations – not universal approximation. Approximation by adding a term to squares sum.
Athnassopoulos and Curram (1996) ANNs vs. DEA	• Simulated data: Cobb–Douglas, 2 inputs, 1 output. Inefficiency distributions: semi-normal and exponential. Result: DEA superior to ANN for measure purpose. ANN similar to DEA ranking units. • Application: bank. Multi-output: 4 inputs, 3 outputs. Ranking, ANN more similar to DEA constant than DEA variable scale returns.
Costa and Markellos (1997) Transport efficiency	• Application: London Underground, time series data, annual 1970–1994, 2 inputs – fleet and workers – and 1 output – kms. • Synthetic sample to frontier estimation – adding noise $N(0, \sigma^2)$ to the inputs. • ANNs' results similar to COLS and DEA; however ANNs offer advantages in decision making, impact of constant vs. variable returns over scale, congestion areas.
Guermat and Hadri (1999) Stochastic frontier functions	• Monte Carlo simulation. • Data from Cobb–Douglas, CES and generalized Leontief. • Functions: ANNs, Cobb–Douglas, translog, CES and Leontief. 2 inputs. • Comparison: mean, maximum and minimum efficiency, standard deviation, correlation between real and estimated efficiencies. • ANNs outperform translog and Cobb–Douglas when translog function is simulated. No differences when Leontief or CES are simulated. • Functional form mis-specification – with ANNs and translog – does not affect mean, maximun and minimum efficiency, but leads to incorrect firm efficiency and ranking.
Santín and Valiño (2000) Education efficiency	• Two-level model: student – production function is estimated by ANNs – and school. • Application: data from 7454 students, 12 inputs. • ANNs superior to econometric approach at frontier estimation.
Fleissig *et al*. (2000) Cost functions	• Comparison: ANNs, Fourier flexible form, AIM – *asymptotically ideal model* – translog and generalized Leontief. • Data: simulated from CES and generalized Box–Cox. • ANNs worse than Fourier – not to impose symmetry and homogeneity like Fourier and AIM. Convergence problems when impose these properties on ANNs.

multi-input and multi-output specification. Here one is thinking in a large group of other non-linear relationships possibilities beyond those outlined in economic theory with a soft and constant curvilinear increasing and decreasing returns to scale into the our production process, not only between one input and one output even between different inputs. Should one consider any chance for the existence of this kind of technology?

Costa and Markellos (1997) found this kind of non-linear relationship in their analysis of the production func-

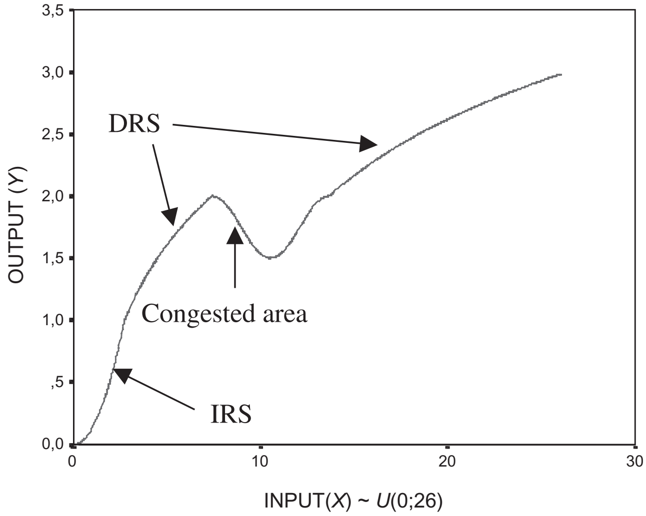

Fig. 3. *The non-linear production function*

tion in the London underground from 1970 to 1994 with an MLP. They showed the existence of a negative slope between inputs (fleet size and workers) and outputs (millions of trains km per year covered by fleet). Baker (2001) concludes in his empirical educational production function analysis with different kinds of neural networks, how substantial performance gains can be achieved for class sizes declining from 14 to 10 students, but also increasing class size (reducing the theoretical input) from 18 to 20 students, meanwhile a linear model only detects a slight downward slope.

Moreover, many educational research articles[10] have found significant coefficients with the 'wrong sign' (e.g. higher per pupil district expenditure or higher teacher education associated with lower student test scores). Eide and Showalter (1998) and Figlio (1999) conclude that traditional restrictive specifications of educational production functions fail to capture potential non-linear effects of school resources. Although they employ more flexible specifications for approximating educational production function like quantile regression and translog function respectively with good results over linear and homothetic relationships, why do not explore the possibility of other non-linear models?

Returning to the experiment, four different scenarios are considered with 50, 100, 200 and 300 decision making units (DMUs). Pseudo-random numbers uniformly distributed across the input space are generated for each scenario:

$$X \sim U(0, 26) \qquad (11)$$

Afterwards, the true output is calculated that is also the true production frontier shown in Fig. 4 and inefficiencies generated through injecting different quantities of noise. Statistical noise is assigned only to the output in the next manner:

$$y^* \sim U(y + ay, y - by) \qquad (12)$$

where y^* will be the observed output, $a = 0.05$ if $b = 0.1$, 0.2, 0.3; and $a = 0.15$ if $b = 0.35$, 0.6, and one measures true technical efficiency (te) as follows:

$$\text{te} = (y^*/y) \qquad (\text{one allows for te} > 1) \qquad (13)$$

For the sake of simplicity, we assume data is free of noise term and all differences between true and observed output are inefficiencies.[11] However, one allows for te > 1 with the aim of representing the existence of outliers.

For each scenario technical efficiency is computed for OLS, COLS with SPSS software, SFA with FRONTIER 4.1 (Coelli, 1996b), DEAcrs and DEAvrs with DEAP 2.1 (Coelli, 1996a) and MLP with S-PLUS software.

Previous to train the MLPs, the data were split into two parts, training and validation sets.[12] Normally, the model is developed on the training set and tested on the validation set. After an exploratory analysis, the study tested how error differences for training and validation patterns was almost identical so it was decided to join in-sample (training set) and out-of-sample (validation set) estimations for computing estimated output. A search was performed from three to eight neurons in one hidden layer with learning coefficient and weight decay fixed with 0.5 and 0.001 values respectively. In order to prevent overfitting, training was stopped when 500 iterations was reached. Neural networks validation sets estimations closer to y^* (MLP Best) were selected for comparisons with remaining techniques.[13]

V. THE RESULTS

Pearson's correlation coefficients[14] were calculated between estimated and true efficiency scores for all techniques over all scenarios (Table 3).

According to results displayed in Table 3, MLP results were best in all cases except one. Note that compared with other techniques, MLP obtains robust estimations with few variations respecting true efficiency over number of DMUs and injected noise. MLP is superior to traditional

[10] See Hanushek (1986) for a survey.
[11] Zhang and Bartels (1998) also assume free of noise data. Nevertheless, one would obtain identical results in this experiment if one decomposes the error term in a normal error variable i.i.d. $u \sim N(0, \delta^2)$ and in a half normal efficiency variable i.i.d. $v \sim N|(0, \delta_v^2)|$.
[12] A typical rule of thumb was chosen on a 80:20 ratio.
[13] A different quite interesting alternative was proposed by Hashem (1993) through combining all trained neural networks according with its performance, i.e. a higher weight in final result for best fitting in validation sets.
[14] We also compute Spearman's rank correlation coefficients with similar results.

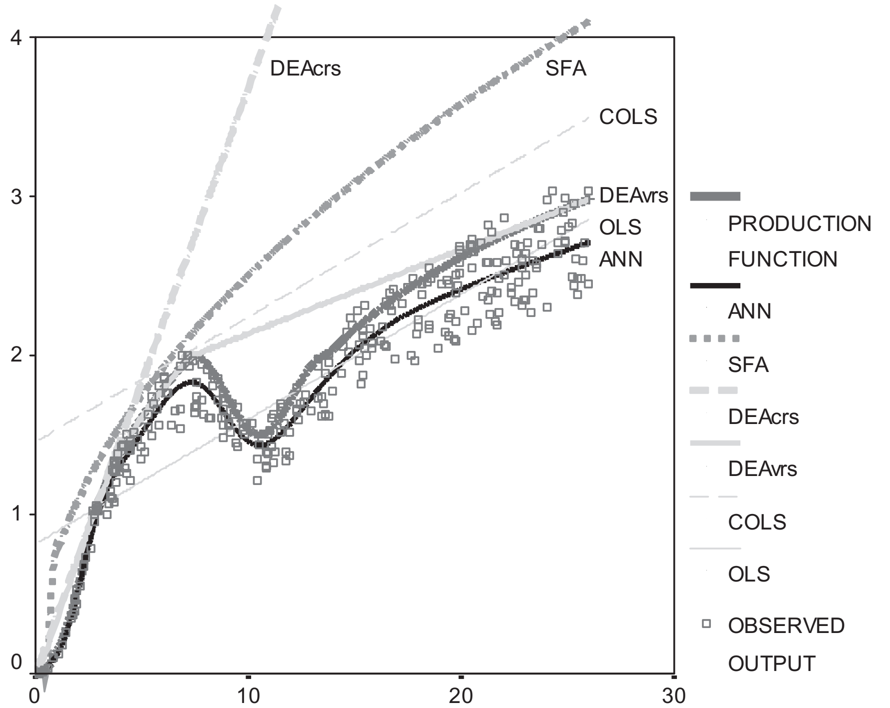

Fig. 4. *Production functions estimated by different techniques*

Table 3. *Pearson's correlation coefficients between estimated and true efficiency scores*

Number of DMUs & percentage of noise injured	Efficiency techniques					
	OLS	COLS	SF	DEAcrs	DEAvrs	MLP_BEST
50 DMUs						
50(15)	0.180	0.104	0.441	0.297	0.431	0.788
50(25)	0.230	0.249	0.294	0.119	0.296	0.838
50(35)	0.464	0.405	0.581	0.419	0.714	0.804
50(50)	0.584	0.575	0.630	0.378	0.798	0.873
50(75)	0.608	0.520	0.443	0.473	0.895	0.887
100 DMUs						
100(15)	0.145	0.146	0.096	0.090	0.183	0.897
100(25)	0.255	0.211	0.239	0.286	0.293	0.751
100(35)	0.297	0.237	0.332	0.357	0.498	0.919
100(50)	0.496	0.490	0.321	0.345	0.661	0.951
100(75)	0.557	0.517	0.474	0.543	0.728	0.855
200 DMUs						
200(15)	0.184	0.205	0.139	0.076	0.249	0.816
200(25)	0.326	0.322	0.258	0.187	0.439	0.961
200(35)	0.377	0.329	0.280	0.348	0.479	0.947
200(50)	0.554	0.557	0.331	0.365	0.686	0.924
200(75)	0.685	0.705	0.337	0.483	0.794	0.934
300 DMUs						
300(15)	0.214	0.248	0.029	0.026	0.302	0.887
300(25)	0.374	0.332	0.388	0.280	0.457	0.935
300(35)	0.447	0.409	0.417	0.316	0.587	0.975
300(50)	0.606	9.697	0.663	0.319	0.736	0.935
300(75)	0.759	0.722	0.804	0.541	0.857	0.973

techniques when underlying technology is under moderate noise together with more DMUs. However, the results show how DEA with variable returns to scale is a little superior to ANN with a lot of efficiency-noise and few DMUs.

In Fig. 4, a particular example for 300 DMUs is illustrated when 25% of uniform noise is injected in true output. After drawing true frontier and all efficiency estimations provided by the different approaches, it is observed how MLP is able to find out the non-linearity contained in data. It is seen that MLP is an average performance technique, although it could be seen that MLP becomes a frontier moving upwards in the curve to the highest residual as is usually done with COLS.

Through Fig. 4, it can also be seen how ANNs are a good tool, as noted by Lee *et al*. (1993), to do an exploratory analysis for searching the existence of non-linear relationships between inputs and outputs before applying a conventional approach and avoiding possible functional form misspecifications. Moreover, this possibility increases exponentially as long as one augments the number of inputs, outputs and contextual variables implied in the production process.

VI. CONCLUSIONS

The results of our simulations confirm that MLP can be used as an alternative tool to econometric and DEA based-techniques for measuring technical efficiency. Another conclusion is that no methodology is always the optimal one for all situations. The benefits of the MLP are its high flexibility and its freedom of a priori assumptions when estimating a noisy non-linear model that allows one to prevent functional form misspecifications and to test if there exists an underlying structure in the available data. Although it is believed that ANNs can be a potential alternative for measuring technical efficiency and can outperform other techniques' results when the production process is unknown, it seems to be reasonably more applied and comparative research. On one hand, although ANNs are increasingly common in a broad variety of domains in economics, there is still a lack of both theoretical and empirical work in efficiency analysis. On the other hand, here one only concentrates on the MLP approach but there are many neural models. Further research should explore the abilities and drawbacks of others ANNs approaches like Bayesian Neural Networks or Generalized Regression Neural Networks versus backpropagation in measuring efficiency through Monte Carlo experiments.

ACKNOWLEDGEMENTS

We thank Antonio Álvarez and José Manuel González-Páramo for useful comments and suggestions. A preliminary version of this paper can be found in the Efficiency Series Papers (09/2001) issued by the Permanent Seminar on Efficiency and Productivity at the Department of Economics, University of Oviedo (Spain).

REFERENCES

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

Álvarez, A. (2001) *La medición de la eficiencia y la productividad*, Ed. Pirámide.

Athnassopoulos, A. and Curram, S. (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**, 1000–16.

Baker, B. D. (2001) Can flexible non-linear modelling tell us anything new about educational productivity?, *Economics of Education Review*, **20**, 81–92.

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

Carrol, S. and Dickinson, B. (1989) Construction of neural networks using the rado transform, *IEEE International Conference on Neural Networks*, **1**, 607–11.

Charnes, A., Cooper, W. W. and Rhodes, E. (1978) Measuring the efficiency of decision making units, *European Journal of Operational Research*, **2**, 429–44.

Cheng, B. and Titterington, D. M. (1994) Neural networks: a review from a statistical perspective, *Statistical Science*, **9**(1), 2–54.

Coelli, T. (1996a) *A Guide to DEAP Version 2.1: A Data Envelopment Analysis Program*, Centre for Efficiency and Productivity Analysis (CEPA), Working Paper 96/08.

Coelli, T. (1996b) *A Guide to FRONTIER Version 4.1: A Computer Program for Stochastic Frontier Production and Cost Function Estimation*, Centre for Efficiency and Productivity Analysis (CEPA), Working Paper 96/07.

Costa, A. and Markellos, R. N. (1997) Evaluating public transport efficiency with neural network models, *Transportation Research C*, **5**(5), 301–12.

Cybenko, G. (1989) Approximation by superpositions of a sigmoidal function, *Mathematics of Control, Signals and Systems*, **2**, 303–14.

De Veaux, R. D., Schumi, J., Schweinsberg, J. and Ungar, L. H. (1998) Prediction intervals for neural networks via nonlinear regression, *Technometrics*, **40**(4), 273–82.

Eide, E. and Showalter, M. H. (1998) The effect of school quality on student performance: a quantile regression approach, *Economics Letters*, **58**, 345–50.

Farë, R., Grosskopf, S. and Lovell, C. A. K. (1985) *The Measurement of Efficiency of Production*, Kluwer, Boston.

Farrell, M. J. (1957) The measurement of productive efficiency, *Journal of the Royal Statistical Society*, **120**, 253–81.

Figlio, D. N. (1999) Functional form and the estimated effects of school resources, *Economics of Education Review*, **18**, 241–52.

Fleissig, A. R., Kastens, T. and Terrell, D. (2000) Evaluating the semi-nonparametric Fourier, AIM, and neural networks cost functions, *Economics Letters*, **68**(3), 235–44.

Fried, H. O., Lovell, C. A. and Schmidt, S. S. (1993) *The Measurement of Productive Efficiency*, Oxford University Press, Oxford.

Funahashi, K. (1989) On the approximate realization of continuous mappings by neural networks, *Neural Networks*, **2**, 183–92.

Geman, S., Bienenstock, E. and Doursat, R. (1992) Neural networks and the bias/variance dilemma, *Neural Computation*, **4**, 1–58.

Guermat, C. and Hadri, K. (1999) *Backpropagation Neural Network Vs Translog Model in Stochastic Frontiers: a Monte Carlo Comparison*, Discussion Paper 99/16, University of Exeter.

Hanushek, E. (1986) The economics of schooling, *Journal of Economic Literature*, **24**(3), 1141–71.

Hashem, S. (1993) *Optimal Linear Combinations of Neural Networks*, Doctoral Dissertation, University of Purdue.

Hebb, D. O. (1949) *The Organization of Behavior,* Science Editions, New York.

Hecht-Nielsen, R. (1989) Kolmogorov's mapping neural network existence theorem, *International Joint Conference on Neural Networks*, **3**, 11–14.

Hertz, J., Krogh, A. and Palmer, R. G. (1991) *Introduction to the Theory of Neural Computation*, Addison-Wesley, Redwood City, CA.

Hornik, K., Stinchcombe, M. and White, H. (1989) Multilayer feedforward networks are universal approximators, *Neural Networks*, **2**, 359–66.

Hornik, K., Stinchcombe, M. and White, H. (1990) Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks, *Neural Networks*, **3**, 551–60.

Hwang, J. T. G. and Ding, A. A. (1997) Prediction intervals for artificial neural networks, *Journal of the American Statistical Association*, **92**(438), 748–57.

Joerding, W., Li, Y., Hu, S. and Meador, J. (1994) Approximating production technologies with feedforward neural networks, in *Advances in Artificial Intelligence in Economics, Finance and Management* (Eds) J. D. Johnson and A. B. Whinston, **1**, 35–42.

Kuan, C. M. and White, H. (1994) Artificial neural networks: an econometric perspective, *Econometric Reviews*, **13**, 1–91.

Kuan, C. M. and Liu, T. (1995) Forecasting exchange rates using feedforward and recurrent neural networks, *Journal of Applied Econometrics*, **10**, 347–64.

Lee, T.-H., White, H. and Granger, C. W. J. (1993) Testing for neglected nonlinearity in time series models. A comparison of neural network methods and alternative test, *Journal of Econometrics*, **56**, 269–90.

McCulloch, W. S. and Pitts, W. (1943) A logical calculus of the ideas immanent in nervous activity, *Bulletin of Mathematical Biophysics*, **5**, 115–33.

Minsky, M. and Papert, S. (1969) *Perceptrons: An introduction to Computational Geometry*, The MIT Press, Cambridge, MA.

Ripley, B. D. (1996) *Pattern Recognition and Neural Networks*, Cambridge University Press, Cambridge.

Rivals, I. and Personnaz, L. (2000) Construction of confidence intervals for neural networks based on least squares estimation, *Neural Networks*, **13**, 463–84.

Rosenblatt, F. (1958) The perceptron: a probabilistic model for information storage and organization in the brain, *Psychological Reviews*, **62**, 386–408.

Rumelhart, D., Hinton, G. and Williams, R. (1986) Learning internal representations by error propagation, in *Parallel Distributed Processing: Explorations in the Microstructures of Cognition* (Eds) D. Rumelhart and J. McClelland, **1**, 318–62.

Santín, D. and Valiño, A. (2000) Artificial neural networks for measuring technical efficiency in schools with a two-level model: an alternative approach, II Oviedo Workshop on Efficiency and Productivity, Oviedo.

Santín, D. and Valiño, A. (2001) Comparing performance of efficiency techniques in non-linear production functions, 7th European Workshop On Efficiency And Productivity Analysis, Oviedo.

Scarselli, F. and Chung, A. (1998) Universal approximation using feedforward neural networks: a survey of some existing methods, and some new results, *Neural Networks*, **11**(1), 15–37.

Schiffmann, W., Joost, M. and Werner, R. (1992) *Optimization of the Backpropagation Algorithm for Training Multilayer Perceptrons*, Technical Report 16/92, Institute of Physics, University of Koblenz.

White, H. (1988) Economic prediction using neural networks: the case of IBM daily stock returns, proceedings of the IEEE International Conference on Neural Networks, San Diego, 451–8.

White, H. (1989a) Learning in artificial neural networks: a statistical perspective, *Neural Computation*, **1**, 425–64.

White, H. (1989b) Some asymptotic results for learning in single hidden-layer feedforward network models, *Journal of the American Statistical Association*, **84**(408), 1003–13.

White, H. (1990) Connectionist nonparametric regression: multilayer feedforward networks can learn arbitrary mappings, *Neural Networks*, **3**, 535–49.

Widrow, B. (1959) Adaptative sampled-data systems, a statistical theory of adaptation, *1959 IRE WESCON Convention Record*, part 4, Institute of Radio Engineers, New York.

Zapranis, A. and Refenes, A.-P. (1999) *Principles of Neural Model Identification, Selection and Adequacy. With Applications to Financial Econometrics*, Springer, Heidelberg.

Zhang, Y. and Bartels, R. (1998) The effect of sample size on the mean efficiency in DEA with an application to electricity distribution in Australia, Sweden and New Zealand, *Journal of Productivity Analysis*, **9**, 187–204.

Zhang, G., Patuwo, B. E. and Hu, M. Y. (1998) Forecasting with artificial neural networks: the state of the art, *International Journal of Forecasting*, **14**, 35–62.