ECONOMIC DISCUSSION PAPERS

Efficiency Series Paper 03/2004

On the Approximation of Production Functions: A Comparison of Artificial Neural Networks Frontiers and Efficiency Techniques

Daniel Santín

Departamento de Economía

Universidad de Oviedo

Available online at: www.uniovi.es/economia/edp.htm

UNIVERSIDAD DE OVIEDO

DEPARTAMENTO DE ECONOMÍA

PERMANENT SEMINAR ON EFFICIENCY AND PRODUCTIVITY

ON THE APPROXIMATION OF PRODUCTION FUNCTIONS: A COMPARISON OF ARTIFICIAL NEURAL NETWORKS FRONTIERS AND EFFICIENCY TECHNIQUES♦

Daniel Santín♠

Efficiency Series Papers 03/2004

Abstract: The aim of this paper is to show how artificial neural networks (ANN) is a valid semi-parametric alternative for fitting production functions and measuring technical efficiency. To do this a Monte Carlo experiment is carried out on a simulated smooth production technology for assessing efficiency results of ANN compared with Data Envelopment Analysis (DEA) and Corrected Ordinary Least Squares (COLS). As ANN provides average production function estimations this paper proposes a so-called *thick frontier* strategy to transform average estimations into a productive frontier. Main advantages of ANN are in contexts where the production function is smooth, completely unknown, contains non-linear relationships among variables and the quantity of noise and efficiency in data is moderate. Under this scenario, the results display that an ANN algorithm can detect, better than traditional tools, the underlying shape of the production function from observed data.

Keywords: Neural networks, production function, technical efficiency

---

♦ This paper has benefited from valuable comments and suggestions of participants in both, the I Workshop de Eficiencia y Productividad EFIUCO (Córdoba) and the 8th European Workshop on Efficiency and Productivity Analysis (Oviedo). All remaining errors are my own.

♠ Department of Applied Economics VI, Complutense University of Madrid, Spain.

E-mail: dsantin@ccee.ucm.es.

1. Introduction

The estimation of a production function is a main issue in diverse economics fields that has fundamentally three applications. The first one would be the prediction of production objectives or outputs from a set of inputs in order to efficiently implement a production process. Second, narrowly related with the previous one, is the measurement of the productive efficiency in order to compare the performance of different decision making units (DMUs), like firms or public services producers, in a homogeneous production context. Last, the production function framework is useful for interpreting, in terms of computed partial elasticities, the statistical influence of the input vector over the outputs. All these theoretical objectives try to offer valuable managing information for taking decisions like reallocation of resources, shortage of costs, incentives and so on.

A serious drawback in empirical economics is that most of times the productive technology is unknown and must be estimated. In these cases, it is quite usual in microeconomics textbooks to impose a number of smooth properties about a *well-behaved* production function for a transforming process of a set of inputs into a set of outputs. Some of these typical assumptions are the impossibility to produce some quantity of output without productive factors, monotonicity, positivity, free disposition of inputs and outputs, possibility of constant, decreasing, and increasing returns to scale or twice continuously differentiable among others. Regardless whether or not these theoretical properties are true in real production functions they impose assumptions that facilitate econometric estimations.

Traditional approaches for estimating empirical production functions with the final aim of measuring efficiency can be fundamentally divided in two types. Firstly, econometric approach [see Kumbhakar, *et al*. 2000 for a general review] imposes a well-known parametric production function[1] where the aim is to adjust, based on ordinary least squares or maximum likelihood regression analysis, the model parameters through the empirical data. To do this it exist different estimation strategies provided by a number of authors [Aigner and Chu (1968), Richmond (1974), Aigner *et al*. (1977), Meeusen and Van den Broeck (1977) or Green (1980a, 1980b)]. Secondly, non-parametric approaches like DEA methods [see Fried, *et al*., 1993; Färe, *et al*., 1994 for a review]

[1] Some well-known examples are: Cobb-Douglas, translog, constant elasticity substitution (CES) or generalized Leontief among others.

are more flexible and do not assume any functional form. This approach draws up a linear piecewise convex production frontier through the efficient units detected in the linear mathematical program constructed to solve the problem. Evolving from Farrell (1957) seminal work, DEA was originally proposed by Charnes, *et al.* (1978) imposing constant returns to scale and Banker *et al.* (1984) relaxing this last assumption. Traditional assumptions for DEA models are the convexity of the set of feasible input-output combinations, variables returns to scale and strong disposability of inputs and outputs

Both approaches, parametric and non-parametric, present different limitations derived from its econometric or deterministic nature. On one hand parametric techniques impose a rigid model to the data raising the issue of mis-specification. On the other hand non-parametric approaches are very sensitive to noise and the presence of outliers in data that can severely bias the efficiency measures. In order to overcome these usual problems recent research [Costa et al., 1993; Athanassopoulos et al., 1996; Guermat et al., 1999; Pendharkar, *et al*. 2003; Santin, *et al*., 2004] has proposed ANN algorithms as a third semi-parametric way for measuring efficiency and fitting production functions in different contexts. Evolving from neurobiological insights ANN have shown to be especially useful for fitting problems which are tolerant of some errors, have lot of example data available, but to which hard and fast rules can not easily be applied like in an expert system or in a parametric model.

The objective of this paper is to provide additional evidence about the potential benefits of standard feed-forward neural networks with backpropagation learning algorithm as tool for estimating production functions. To fulfil with this purpose I compare through a Monte Carlo experiment the results obtained by traditional efficiency techniques, DEA and corrected ordinary least squares (COLS), and ANN in a smooth non-linear production function. This numerical technique of calculation permits to analyze multiple settings of a model using different samples of data generated from a probability distribution previously defined

The paper is organized as follows. The next section provides a brief introduction to ANN as a promising tool for the measurement of technical efficiency revising its main statistical advantages and limitations. Section 3 is dedicated to describe and illustrate the Monte Carlo experiment comparing efficiency techniques and ANN results in a non-linear production function. This section also addresses the construction of *thick frontiers* facing up the problem of how to do that ANN average estimations became a production frontier. The final section of the paper offers main conclusions and suggests areas for future research.

2. Artificial Neural Networks

The most commonly used neural network architecture is the Multilayer Perceptron (MLP from now on). We can define a MLP like a group of processing elements, known as neurons, organized in at least three layers, input, hidden(s) and output (figure 1).

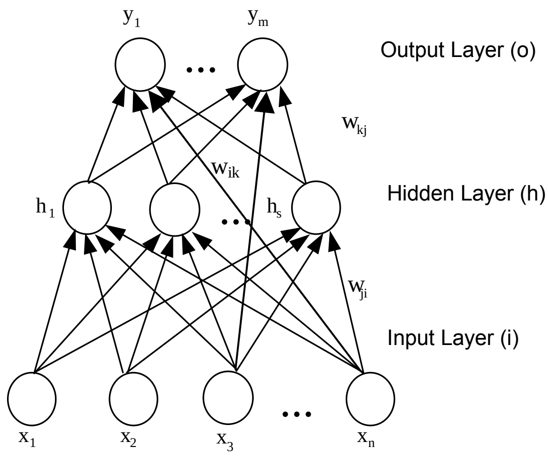

Figure 1

A multi-input-multi-output, three layers *Feed-forward* neural network architecture

As it is shown in figure 1 these neurons are all connected in one direction from input to hidden and from hidden to output by unidirectional connections or weights, is a so called feed-forward neural network. When the neuron receives the weighted information from other neurons all signals are added up and transformed through a squashing or logistic function. With this step a non linear feature is introduced. These transfer functions must have mild regularity conditions: continuous, bounded, differentiable and monotonic increasing. The most popular transfer function is the logistic, nearly linear in the central part. The transfer functions[2] (equation 1) bound the output to a finite range, [0; 1].

$$f:\Re \rightarrow [0,1] \left| f(x)=\frac{1}{1+e^{-x}}, x\in\Re \right. \tag{1}$$

---

[2] Another frequent function to express polarity is *tanh*: $\Re \Rightarrow [-1,1] \left| f(x)=\frac{e^{x}-e^{-x}}{e^{x}+e^{-x}} \right.$

The target of a MLP is learning to match input to output vectors through the interactions among neurons. This implies learning the parameters in function (2)

$$\Re^n \xrightarrow{\phi} \Re^m$$
$$\phi(X;W) = Y$$
$$y_k = f_o\left[\sum_{i\to k} w_{ik} x_i + \sum_{j\to k} w_{jk} f_h\left(\sum_{i\to j} w_{ij} x_i\right)\right] \qquad (2)$$

Through the sample {X(p), Y(p)}, p = 1,2,...,N where $X(p)\in \Re^n$ is the input vector and $Y(p)\in \Re^m$ is the output vector. This is carried out by adjusting the matrix of weights (W) of given interconnections among the neurons according to some learning algorithm. MLP uses a supervised learning algorithm proposed by Rumelhart *et al*. (1986) called *backpropagation* which is the most widely used learning method in empirical applications. This learning is guided by specifying the desired response of the network, the observed output, for each training input pattern and its comparison with the actual output computed by the network in order to adjust the weights (equation 3).

$$\varepsilon_p = \frac{1}{2}\sum_{k=1}^{M} \varepsilon_{pk}^2 \qquad (3)$$

Where ε denotes the error term, p denotes training vector and k denotes output neuron. This error term propagates backwards through the calculation of partial derivatives, (equation 4) from output layer to hidden layer(s), until it reaches the input layer[3].

$$\Delta w_{kj}^o = -\alpha \frac{\partial \varepsilon_p}{\partial w_{kj}^o} \qquad \Delta w_{ji}^h = -\alpha \frac{\partial \varepsilon_p}{\partial w_{ji}^h} \qquad \Delta w_{ki}^i = -\alpha \frac{\partial \varepsilon_p}{\partial w_{ki}^i} \qquad (4)$$

Where α denotes a learning coefficient to guide the strength of weights changes. Each weight is modified according with its partial contribution to the final error. These adjustments have the purpose of minimizing the difference between desired and actual outputs. The performance is measured in terms of some lost function like the root mean square error. After a number of loops, when the benefits of further optimization are regarded as small, the training process converges and stops. The halted can be performed by the analyst or by a specified stop rule as the number of algorithm iterations or reaching a specified minimum error value. As it was said before, ANN can be categorized as a semi-parametric tool because a number of parameters or weights

[3] For a complete revision of the mathematical algorithm and statistical proprieties of artificial neural networks see for example Bishop (1995).

must be computed. Nevertheless, these parameters do not possess robust statistical properties to compute elasticities and testing for significance over the model outputs.

Normally the implementation of the backpropagation algorithm implies to split the sample into three data sets (the so-called early stopping). A training set is used to seek the parameters able to match inputs with outputs. A validation set is used at the same time to control for model complexity and for stopping learning when no gains are obtained over this sample with further optimization (figure 2). Finally, after MLP training, new observations never seen before by the neural net (test set) are presented to the network to obtain an unbiased measure of the so-called generalization capability. Data test is especially important if the final aim for the trained MLP is to predict new data. In all this process it is relevant the classical statistical bias-variance dilemma (Geman et al., 1992) or overfitting problem.

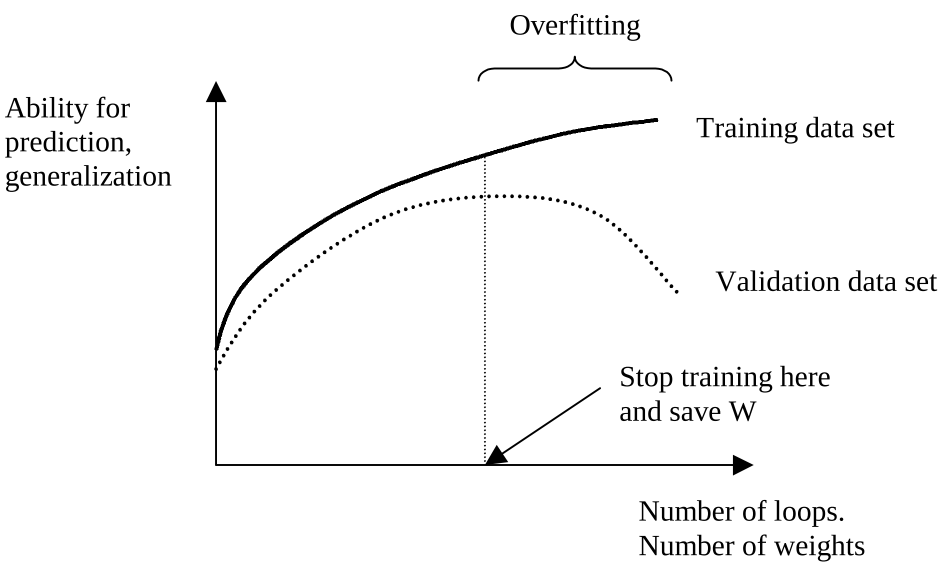

Figure 2: Mechanism to avoid overfitting when training ANN.

The division of the data set into subsamples prevents for severe overfitting. As it is shown in figure 2 the validation data set controls for overtraining. When error improvements over the training set do not imply further tight adjustment over the validation set then training is stopped. There are not easy rules to train neural networks. This means that the researcher does not have *a priori* information about the correct number of parameters for using according with each problem[4]. For this reason

[4] Main parameters are the number of hidden layers, number of neurons in each hidden layer, learning coefficient, transfer functions and momentum vector to introduce some inertia in the weight decay in order to avoid falling trapped in local minima.

nowadays one of the main disadvantages of ANN is that this technique is a high time demanding trial-error algorithm compared with other tools.

On the other hand it has been shown that ANN techniques are universal approximators of functions (Cybenko, 1989; Hornik *et al*., 1989; Funahashi, 1989) and their derivates (Hornik *et al*., 1990). Scarselli and Chung (1998) provide a complete review of this property. MLP is both semi-parametric and stochastic and it has been identified by statisticians like a powerful non-linear regression method. These facts justify the success reached by ANN in multiple applications in an extensive number of science fields. Hill *et al*., (1994) show that the performance of neural networks is at least as powerful as statistical models. Recent results also show how ANN are an alternative approach to generate rules for non-linear [Setiono *et al*., 2002] and linear [Setiono and Thong, 2003] regression models.

3. The Monte Carlo Experiment.

3.1. Experimental Design

In order to examine the performance of efficiency techniques, let G(x) be the further non-linear double-differentiable continuous smooth production function (equation 5):

$$G(x) = 2\left[\sin\left(\left(\frac{x}{2}\right) - \left(\frac{\pi}{2}\right)\right) + 1\right] + 1 \tag{5}$$

where G(*x*) is the output, and *x* is a controllable input. Obviously this is not one of the basic functional forms used in the literature to describe a production process[5]. However this production function fulfils all smooth properties traditionally pointed out in a microeconomics textbook [Mas Colell; *et al*., 1995]. Moreover this production function captures the theoretically increasing and decreasing average product stages. An illustration of this production function is showed in figure 3.

[5] Note that a *sin* production function is only interesting like one of infinite smooth production function alternatives that join together all traditional desirable properties from a microeconomics point of view.

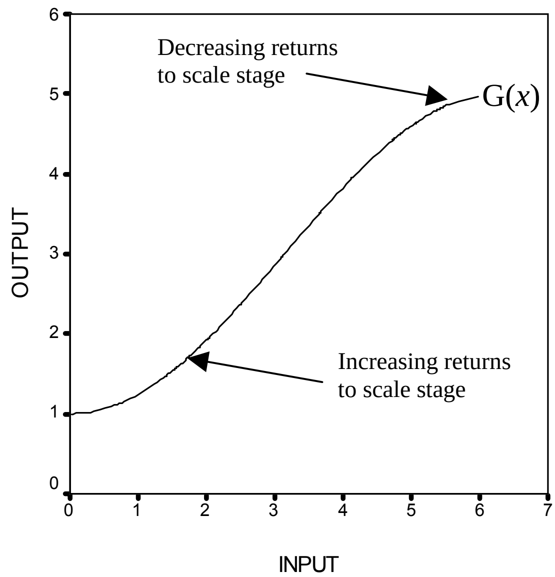

Figure 3: The experimental production function

A number of 50 pseudo-random decision making units (DMU) input data uniformly distributed across the input space (0; 2π] were generated according with X ~ U (0, 2π]. This input vector yields output quantities G(*x*) on the frontier which are free of any inefficiency or random noise. The output depicted above is modified by fluctuations due to both components: inefficiency and random noise. An inefficiency value is calculated for each DMU with a half normal distribution of $\nu \sim |N(0{,}7;0{,}01)|$ and random noise normally distributed $\varepsilon \sim N\ (0;\ 0{,}01)$ is also generated. At the same time, a number of DMUs are allowed to remain efficient belonging to the true frontier. To do this a distribution Bernoulli (0.2) is used to decide what DMUs are 100% efficient. After this process the synthetically generated observed output is obtained. A graphical representation of the cloud of points obtained is showed in figure 4.

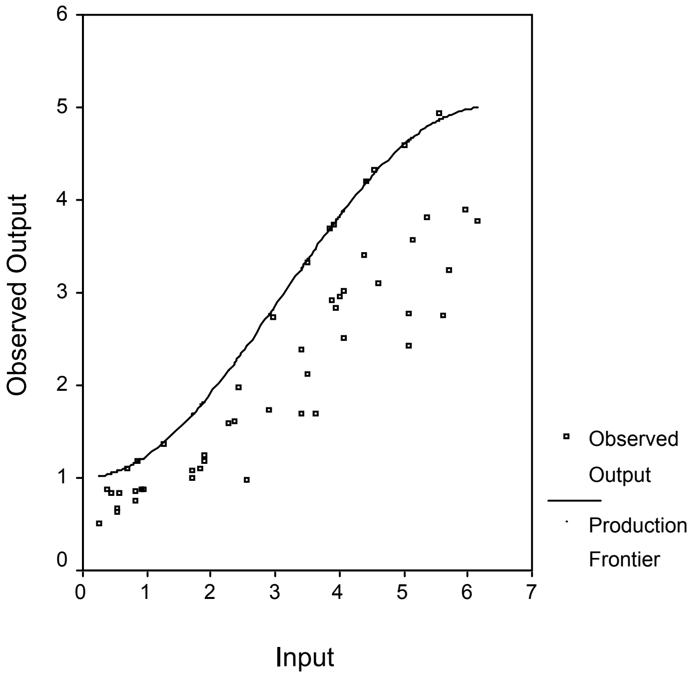

Figure 4: An example of the input-observed output production function problem.

Finally the objective of the experiment is to fit this production function with DEA under variable returns to scale (DEAvrs from now on) using Banker et al. (1984) model[6], COLS[7] [Richmond, 1974; Greene, 1980a; 1980b] and a MLP trained with a backpropagation algorithm.

3.2. The computation of *thick frontiers* with neural networks.

Since MLP average production function estimation is not a frontier a second stage is necessary to bound efficiency scores between one and zero in order to assess efficiency measures. To do this two main strategies are followed based on Athanassopoulos and Curram (1996, pp. 1003-1004) work. The first one consists in adding up the maximum residual term to the average output predicted by the MLP for each DMU. This methodology will be named as MLPMAX and calculated according with equation (6).

$$TE^{MLPMAX} = \frac{y_i}{\hat{y}_i + \max(R_j)} \qquad (6)$$

[6] This is the well-known BCC model.

[7] COLS method was performed under $\mathrm{Ln}(y) = \beta_0 + \beta_1 \mathrm{Ln}(x) + \beta_{11}(\mathrm{Ln}\ x)^2$ in two steps. First, OLS analysis is calculated. Second, the intercept term is shifted upward using the maximum residual term to derive a consistent production frontier.

Where y_i denotes the observed output for DMU i, $\hat{y}_i$ is the predicted output by the MLP for DMU i and R_j is the maximum residual value observed in DMU j. This procedure is illustrated in figure 5 where MLPFIT is an average production function fitted by the backpropagation algorithm and MLPMAX is traced up following equation (6).

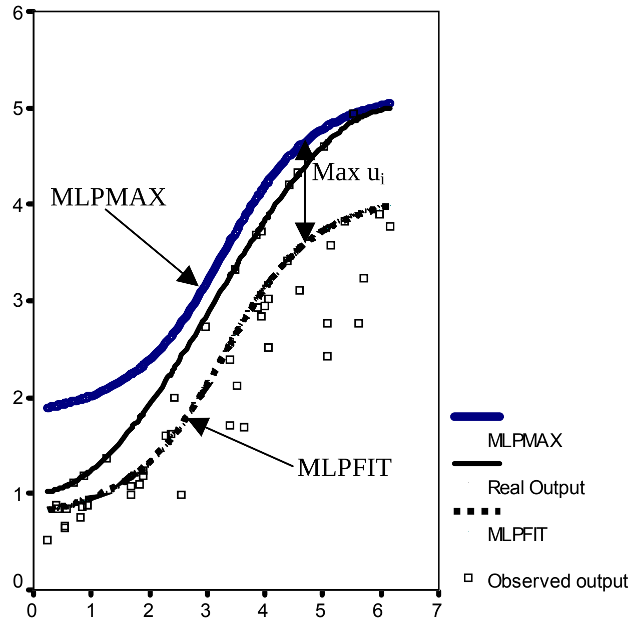

Figure 5

MLP maximum residual correction for drawing up the production frontier.

As it was said before the second methodology was also proposed by Athanassopoulos and Curram (1996) but this idea was not developed in their paper. In order to alleviate extreme maximum residual terms these authors proposed applying (6) to different segments of the distribution of the dependent variables which leads to the concept of the so-called *thick frontiers*. This methodology will be developed in this paper. To fulfill with this purpose the frontier will be drawn up from the least average computed output value to the large one in the following way.

1. Order the N DMUs from the least average fitted output to the large one. DMU $\in$ [1; N] | $DMU_1,\ldots,DMU_N$ where DMU_1 has the least estimated output and DMU_N has the highest one.
2. Following this ranking detect the first positive error ε_1 belonging to DMU i. $\varepsilon_1 = y_i - \hat{y}_i$

3. Once a positive error ε_1 is found, fitted output for all observations with a computed output less than DMU i, from i to 1, are shifted upwards adding up ε_1 to each average computed output.
4. Detect the second positive error ε_2 belonging to DMU j with $\varepsilon_2 > \varepsilon_1$
5. Add up ε_2 to fitted output for all DMUs between DMU j and DMU i
6. Repeat the stages described above until find DMU$_k$ with the highest error $\varepsilon_k > \varepsilon_{k-1} > \varepsilon_2 > \varepsilon_1$
7. Add up ε_k to fitted output for all DMUs between DMU$_k$ and DMU with error ε_{k-1}.
8. The process can finish in these two ways:
   - SMOOTHMLP: Adding up ε_k for remaining DMU from DMU$_k$ to DMU$_N$
   - DEAMLP: If between DMU$_k$ and DMU$_N$ it is found a DMU$_h$ with the biggest real output y_h and this output is smaller or equal that fitted output plus ε_k *i.e.* $y_h \leq \hat{y}_h + \varepsilon_k$ then assign y_h for remaining DMUs from DMU$_h$ to DMU$_N$

When the entire process is over a so defined *thick frontier* is obtained for measuring efficiency. An illustration of both kinds of thick frontiers is showed in figure 6.

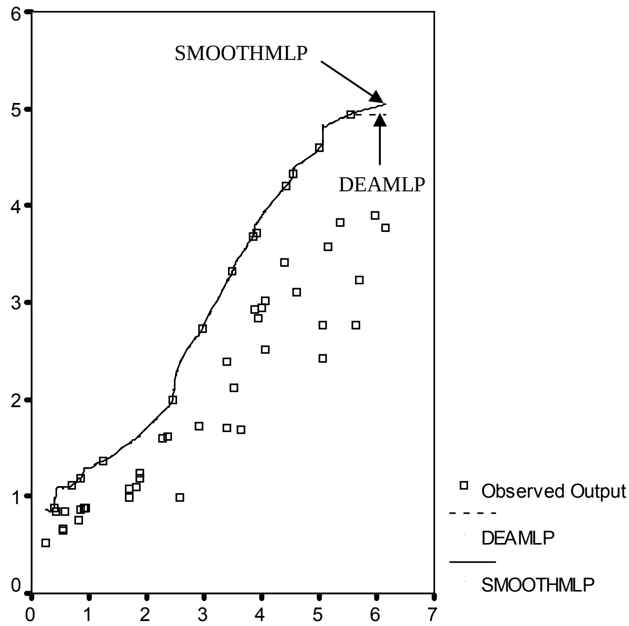

Figure 6: Two kinds of thick frontiers, DEAMLP and SMOOTHMLP.

3.3. Simulation results.

A number of 100 samples with size N=50 were generated in the Monte Carlo experiment to fit production frontiers with each technique. Previous to train the MLP, data was split in two parts, training and validation sets[8]. The model was developed on the training set and tested on the validation set. After an exploratory analysis, it was tested that error differences for training and validation patterns was almost identical. Thus, in-sample (training set) and out-of-sample (validation set) estimations were joined for computing estimated output. A number of five neurons in one hidden layer was selected with learning coefficient and weight decay[9] fixed both at 0.1. A logistic function in all neurons was used as transfer function.

Based on the design considerations named before, the Monte Carlo experiment was conducted to assess efficiency techniques accuracy. Average technical efficiency and Pearson′s correlation coefficient between real and estimated efficiency scores were computed in order to compare the performance of each approach regarding real efficiency. Simulation results are presented in table 1.

Table 1: The Monte-Carlo experiment results after 100 replications.

	Average computed efficiency	Pearson′s correlation coefficient between real and estimated efficiency
Real Efficiency	0,7506 (0,0222)	-
DEAMLP	0,7324 (0,039)	0,9417 (0,06526)
SMOOTHMLP	0,7291 (0,0376)	0,9434 (0,0687)
DEAvrs	0,7004 (0,028)	0,8501 (0,0418)
MLPMAX	0,6032 (0,0424)	0,6417 (0,1295)
COLS	0,5962 (0,0388)	0,6387 (0,1372)

Standard deviation is shown in parenthesis.

The comparative results of the differences between the real efficiency and the various approaches used reveal that ANN thick frontier methodologies obtain both the best results. Thick frontier approaches show superiority results over DEAvrs mainly in terms of correlation with real for average efficiency estimations. However DEAvrs results are

[8] A typical well-known rule of thumb on a 80:20 ratio was used to split the sample into training and validation respectively.

[9] A weight decay term $\sum_{k=1}^{M}\left[y_k - \phi(x_i;W)\right]^2 + \beta\sum W^2$ is introduced over the error to improve the training controlling for overfitting.

substantially better than the so-called MLPMAX approach. Moreover the results displayed in table 1 for the parametric tool (COLS) are very similar to MLPMAX but worse than remaining techniques. The results commented above are consistent for both performance measurements reported in table 1.

Nevertheless differences in terms of superiority between DEAMLP and SMOOTHMLP thick frontiers are not clear. On one hand DEAMLP obtains slightly better results in terms of average efficiency compared with true. On the other hand SMOOTHMLP provides higher accuracy results in terms of Pearson′s correlation coefficient for average efficiency. This result is due to high similarities between both models. Further research is necessary to disentangle the potential benefits and drawbacks of each neural network frontier approach.

No general definitive conclusions can be drawn from this study. However the experiment carried out in this paper shows how ANN thick frontiers type could be a valid alternative to measure technical efficiency with higher accuracy than traditional techniques. The potential benefits of ANN are more evident under non-linear production functions that present a moderate signal to noise ratio. This kind of production technology would allow ANN to find the underlying structure contained in data sample providing better model specification that the other techniques.

To sum up, in empirical production problems a number of different techniques are available for the measurement of technical efficiency. These methodologies can be divided into three subsets: parametric tools (econometric, stochastic frontiers), semi-parametric tools (ANN but also other techniques as kernel regression) and non-parametric techniques (DEA, Free Disposal Hull). The election of the most adequate efficiency technique depends on the problem properties that the researcher affords. Table 2 provides a rough comparison of main potential advantages and disadvantages for these three principal categories of efficiency tools.

Table 2 indicates that the type of problem and the objectives of the study should impose a high restriction over the kind of approach to use. For explicative highly linear problems parametric approaches seems to be the best choice. As long as the researcher suspects or detects non-linearities contained in the production problem a semi-parametric approach will offer better results that its econometric and mathematical programming counterparts. Finally if the input-output problem presents, controlling for outliers, a high proportion of inefficiency and noise and no structure about the production function can be extracted from data sample a non-parametric approach will be preferred.

Table 2

A rough comparison among different strategies for measuring efficiency.

Comparative Factor	Parametric	Non-Parametric	Semi-Parametric
Functional form assumptions	Strong	Modest	None
Flexibility	Low	Modest	High
Theoretical basis	Strong	Strong	Modest
Efficiency Studies	Strong	Strong	Weak
Computation of Elasticities	Yes	None	Modest
Multi-input, multi-output problems	No	Yes	Yes
Projection, generalization, prediction	Modest	None	High
Cost of analysis	Low	Modest	High
Kind of frontier	Stochastic	Deterministic	Deterministic

4. Conclusions

The results obtained in this paper can be summarized as follows. First, ANN is an alternative to traditional techniques for the measurement of technical efficiency. Main relative advantages of ANN are for those problems with non-linear relationships between variables that presents a weak theoretical knowledge about the production technology.

Second, no single approach appears to be overall superior compared with remaining techniques. This fact points out how the efficiency technique should be chosen according with the problem the researcher have to face up to. In any case ANN is always a good tool to do an exploratory analysis to test the existence of non-linear relationships before applying a conventional approach avoiding for possible functional form misspecifications.

Third, the so-called thick frontiers issue developed in this paper through the experiment overcomes traditional methods to draw up the production frontier from ANN average production function. This is a promising alternative for measuring efficiency from semi-parametric tools. However further research is still necessary in order to generalize this result in different scenarios (number of DMUs, signal to noise ratio, average efficiency, heteroscedasticity and so on). This research should also explore the possibilities of integrating several approaches, combining its potential benefits, in order to enhance technical efficiency measurement.

Last, the Monte Carlo experiment carried out in this paper shows how ANN is able to fit a conventional non-linear production function under a moderate level of efficiency and noise. The adjustment reached by ANN overcomes traditional approaches. Another field not developed in this paper and still open for future research is how to decompose the error term into random noise and technical efficiency components. This issue concerns the calculation of ANN stochastic frontiers to obtain better measurements of technical efficiency

References

Aigner, D. J. and S. F. Chu (1968), "On estimating the industry production function", *American Economic Review*, 58, 826-839.

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

Athnassopoulos, A. and S. Curram (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, 1000-1016.

Banker, R. D., Charnes, A. and W. W. Cooper (1984), "Models for the estimation of technical and scale inefficiencies in Data Envelopment Analysis". *Management Science*, 30, 1078-1092.

Banker, R. D., Charnes, A., Cooper, W. W. and A. Maindiratta (1988), *A comparison of DEA and translog estimates of production frontiers using simulated observations from a known technology*. In: Dogramaci, A. y Färe, R. (Eds.) *Applications of Modern Production Theory: Efficiency and Productivity*, Boston, MA, Kluwer.

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

Costa, A. and R.N. Markellos (1997), "Evaluating Public Transport Efficiency with Neural Network Models", *Transportation Research C*, 5 (5), 301-312.

Cybenko, G. (1988), "Continuous valued neural networks with two hidden layers are sufficient", *Technical Report, Dept. of Computer Science*, Tufts University.

Charnes, A., Cooper, W. W. and E. Rhodes (1978), "Measuring the inefficiency of decision making units", *European Journal of Operational Research*, 2, 429-444.

Färe, R., Grosskopf, S. and C. A. K. Lovell (1994). *Production Frontiers*. Cambridge University Press.

Farrell, M.J. (1957), "The measurement of efficiency productive". *Journal of the Royal Statistical Society A*, 120, pp. 253-266.

Fried, H.O., Lovell, C. A. K. and S. S. Schmidt (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: 183-192.

Geman, S., Bienenstock E. and R. Doursat (1992), "Neural Networks and the Bias/Variance Dilemma", *Neural Computation*, 4, 1-58.

Greene, W. H. (1980a), "Maximum likelihood estimation of econometric frontier", *Journal of Econometrics*, 13(1), 27-56.

Greene, W. H. (1980b), "On the estimation of a flexible frontier production model". *Journal of Econometrics*, 13(1), 101-115.

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

Hill, T., Marquez, L., O'Connor, M. and W. Remus (1994), "Artificial neural network models for forecasting and decision making". *International Journal of Forecasting* 10. pp. 5-15.

Hornik, K., Stinchcombe, M. and H. White (1989), "Multilayer Feed-forward Networks are Universal Approximators". *Neural networks* 2, pp. 359-66.

Hornik, K., Stinchcombe, M. and H. White (1990), "Universal Approximation of an Unknown Mapping and its Derivatives using Multilayer Feed-forward Networks", *Neural Networks* 3, pp. 551-60.

Kumbhakar, S. C. and C. A. K. Lovell (2000), *Stochastic Frontier Analysis*. New York. Cambridge University Press.

Mas-Colell, A., Whinston, M., and J. Green (1995), *Microeconomic Theory*, Oxford University Press.

Meeusen, W. and J. Van den Broeck (1977). "Efficiency Estimation from Cobb-Douglas Production Functions with composed error", *International Economic Review*, 18, 435-444.

Pendharkar, P. C. and J. A. Rodger (2003). "Technical efficiency-based selection of learning cases to improve forecasting accuracy of neural networks under monotonicity assumptions". *Decision Support Systems*, 36, pp. 117-136.

Richmond, J. (1974). "Estimating the efficiency of production", *International Economic Review*, 15, 515-521.

Rumelhart, D. E.; Mc Clelland, J.L. and the PDP Research Group (Eds.) (1986): *Parallel Distributed Processing: Explorations on the Microstructure of Cognition*, vol.I. Foundations, MIT Press, Cambridge, MA.

Santin, D., Delgado, F. J. and A. Valiño (2004), The measurement of technical efficiency: a neural network approach. *Applied Economics*, 36(6), 627-635.

Scarselli, F. and A. Chung (1998), "Universal Approximation Using Feedforward Neural Networks: a Survey of Some Existing Methods, and Some New Results", *Neural Networks*, 11 (1), 15-37.

Setiono, R., Leow, W. K. and J. Zurada (2002), "Extraction of rules from artificial neural networks for nonlinear regresión". *IEEE Transactions on Neural Networks* 13(3), 564-577.

Setiono, R. and J. Y. L. Thong (2004). "An approach to generate rules from neural networks for regression problems". *European Journal of Operacional Research*, 155, 239-250.