Comparing performance of efficiency techniques in non-linear production functions

Daniel Santín González* and Aurelia Valiño Castro*
Department of Applied Economics
Complutense University of Madrid

Abstract

Non-linear production functions are common in economic theory and in real life, especially in cases with increasing and diminishing returns to scale but there are also contexts where an increase in one input implies a decrease in one output. The aim of this paper is to test how non-linearity affect estimations of technical efficiency obtained by ordinary and corrected least squares (OLS, COLS), data envelopment analysis with constant and variables returns to scale (DEAcrs, DEAvrs), stochastic frontier analysis (SFA) and by multilayer perceptron neural networks with backpropagation (MLP). To do this we will construct a very simple non-linear one input-one output production function and we will obtain different synthetic data with 50, 100, 200 and 300 decision-making units (DMUs). Afterwards we will add up different quantities of noise to the data and finally we will compare real efficiency with estimated values for all techniques named before among the different scenarios. Our results suggest that MLP is a flexible tool to fit production functions and a possible alternative to traditional techniques under non-linear contexts.

Keywords: Non-linear production function, technical efficiency, artificial neural networks.

1. Introduction

In empirical production function analysis our prior knowledge about the technology that relates inputs and outputs is many times quite weak. Taking into account this fact, the two common approaches for measuring efficiency, linear programming methods like data envelopment analysis (DEA) and econometric techniques like multiple regression and in particular stochastic frontier analysis (SFA)[1], assume only a few restrictions when estimating production frontiers. 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. On the other hand, SFA has been modelled through Cobb-Douglas specifications or even through more flexible transcendental logarithmic (translog) form where we include non-linear effects into a linear parametric model.

---

* Corresponding author. E-mail: dsantin@ccee.ucm.es, ecap316@sis.ucm.es. Departamento de Economía Aplicada VI. Facultad de Ciencias Económicas y Empresariales. UCM. Campus de Somosaguas. Pozuelo de Alarcón 28223 Madrid. Spain

[1] See Fried et al. (1993) and Alvarez (2001) for a thorough review of these techniques.

Evolving from neuro-biological insights, artificial neural networks (ANNs) is a relatively young field of research with a rapid expansion in both theory and applications. Neural networks have shown during last decade to be especially useful for mapping problems under noise and uncertainty, when enough data example are available and particularly when inputs and outputs are related in non-linear ways which cannot be easily described in advance in linear equations.

Guermat and Hadri (1999) carried out a Monte Carlo experiment with the purpose of analyse the effects of functional form misspecifications and the performance of neural networks versus translog models for approximating different theoretical production functions like Cobb-Douglas, translog, CES function and Generalized Leontief model. They concluded that the neural network specification is a good alternative to the most commonly used translog model for measuring efficiency.

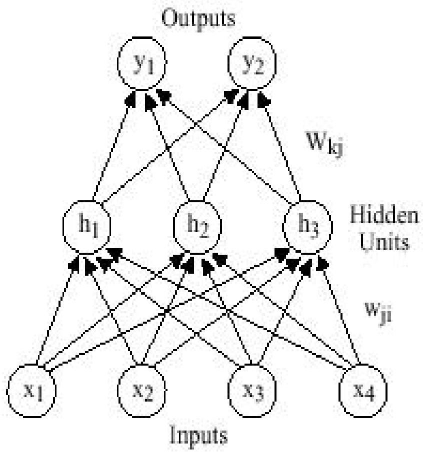
The aim of this paper is to provide additional evidence in order to propose the use of flexible non-linear models free of a priori assumptions, in particular backpropagation neural networks, for estimating technical efficiency in those contexts where we recognize an almost complete ignorance about the technology implied in the production process. To fulfil with this propose we will compare the results obtained by traditional efficiency techniques with neural networks in a non-linear production function.

The rest of the paper is organized as follows. In the next section we will discuss the main characteristics of neural networks techniques. Section 3 describes the simulation model, the experimental design together with a brief discussion of non-linear production functions, the distributions used to generate our synthetic data and how we have altered data to introduce inefficiencies. In section 4 we will present and discuss the obtained results. Finally, section 5 offers concluding remarks and suggests areas for further research.

2. Artificial Neural Networks.

The most commonly applied ANN is the so-called Multilayer Perceptron (MLP). This tool has attracted many researchers from diverse fields including signal processing, medical diagnosis or weather forecasting. Moreover, during last decade MLP have been applied also in economics for forecasting and classification in many areas especially for time series, banking and stock markets. Furthermore, Costa and Markellos (1997) apply MLP for evaluating public transport efficiency with neural networks models and compare the results with regression and DEA models. We can observe a typical MLP architecture in figure 1.

Figure 1: The Multilayer Perceptron Architecture



We can define a MLP like a group of processing elements, known as "neurons", organised in at least three layers, input, hidden(s) and output. These neurons are all connected in one direction by unidirectional communications channels or connections, is the so-called feed-forward network. The target of a MLP is learning to match a vector of inputs (X) to a vector of outputs (Y) through the interactions among neurons (W). This imply learning the function (equation 1):

$$\mathbb{R}^n \xrightarrow{f} \mathbb{R}^m$$

$$f\,(W; X) \sim Y$$

$$y_k = f_0\left[\sum_{i\to k} w_{ik}x_i + \sum_{j\to k} w_{jk}\, f_h\left(\sum_{i\to j} w_{ij}x_i\right)\right] \qquad (1)$$

Through the sample {X(p), Y(p)}, p = 1,2,…,N where $x(p)\in \mathbb{R}^n$ is the input vector and $y(p)\in \mathbb{R}^m$ is the output. Where *i* denotes input, *j* hidden and *k* output layer.

This is carried out by adjusting the weights (W) of given interconnections according to some learning algorithm. MLP employs a supervised learning algorithm popularised by Rumelhart et al. (1986) called backpropagation. The learning is guided by specifying the desired response to the network for each training input pattern through the comparison with the actual output computed by the network in order to adjust the weights. These adjustments have the purpose of minimize some energy function, normally the square difference between the desired and actual outputs (equation 2).

$$E(W) = \sum_p \left\| t^p - f\left(x^p ; w\right) \right\|^2 \qquad (2)$$

The derivatives of the function with respect to the weights (equation 3) are employed to propagate the error backwards through the network from output to hidden layer(s), until it reaches the input layer. Each weight is modified according with its particular contribution to the global error. The performance of the net is measured most of times in terms of the root mean square error. After a number of loops, when the benefits of further optimisation are regarded as small, the training process converges and stops[2].

$$\Delta w_{kj}^{o} = -a \frac{\partial E_p}{\partial w_{kj}^{o}} \qquad (3)$$

MLP is both non-parametric and stochastic and have been identified by statisticians like a powerful non-linear regression method. In fact, there exist a strong theorem provided from different authors Cybenko (1988), Hornik et al. (1989, 90), Funahashi (1989), etc. that showed how MLP with four layers using a non-linear continuous transfer function and with enough number of neurons in the hidden layers are universal approximators of any function and its derivatives to any desired degree of accuracy. Furthermore, if *f* is continuous we can achieve the same result with only a three layer MLP. Kuan and White (1994) probed that linear regression and binary logit and probit models are special cases of neural networks models.

3. 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

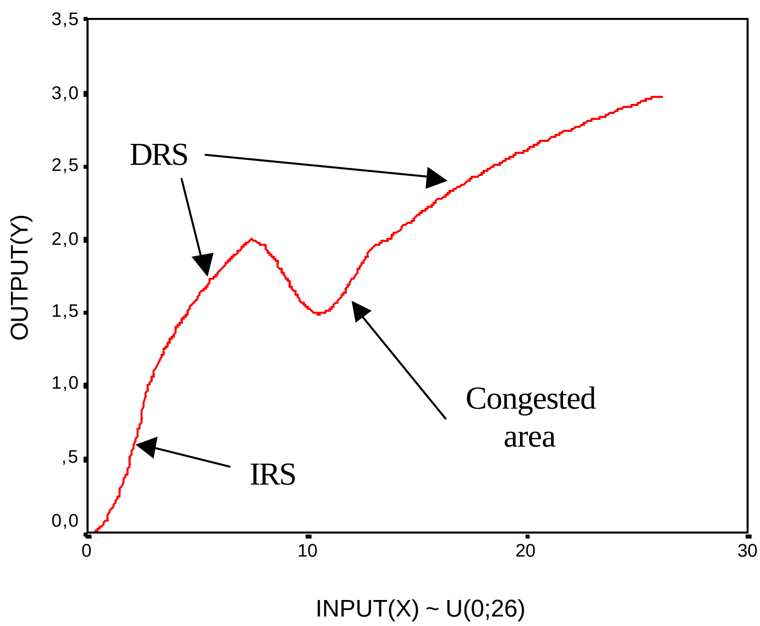
[2] It is not the aim of this paper to do a long discussion about the different architectures and methods for training neural networks. For an excellent review of this technique see Bishop (1995) and Ripley (1996).

$$
F(x) = \begin{cases} \left(\dfrac{x}{e}\right)^2 & \text{if } x \in [0, e] \\ \text{Ln}(x) & \text{if } x \in [e, e^2] \\ A * \text{COS}(x - e^2) + 2 - A & \text{if } x \in [e^2, e^2 + \pi], \text{ where } A = 0{,}25 \\ \text{Ln}(x - 2\pi) & \text{if } x \in [e^2 + \pi, 26] \end{cases} \tag{4}
$$

Through this production function (see figure 1) we introduce all returns to scale possibilities. The first part of (4) is increasing returns to scale (IRS). Second and fourth sections show decreasing returns to scale (DRS). 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) we will call this phenomenon a "congested area".

However, our intention here is to illustrate what occurs with efficiency estimations when our "traditional linear models" are not the real production functions for the multi-input and multi-output specification. Here we are thinking in a large group of others non-linear relationships possibilities beyond those outlined in economic theory with a soft and constant curvilinear increasing and decreasing returns to scale into our production process, not only between one input and one output even between different inputs. Should we consider any chance for the existence of this kind of technology?.

Figure 1: The Non-Linear Production Function



Costa and Markellos (1997) analysing the production function in London underground from 1970 to 1994 with a MLP found this kind of non-linearity. 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) conclude in his empirical educational production function analysis employing 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 our theoretical input) from 18 to 20 students meanwhile a linear model only detects a slight downward slope.

Moreover, many educational research articles[3] 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 others non-linear models?.

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

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

Afterwards, we calculate the true output that is also the true production frontier showed in figure 1 and we generate inefficiencies through injecting different quantities of noise. Statistical noise is assigned only to the output in the next manner.

$$y^* \sim U(y+ay;\ y-by)$$

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 we measure true technical efficiency (te) as follows:

$$te \sim (y^*/y) \qquad\qquad \text{(we allow for te>1)}$$

[3] See Hanushek (1986) for a survey.

For the sake of simplicity we assume data is free of noise term and all differences between true and observed output are inefficiencies[4]. However, we allow for te > 1 with the aim of representing the existence of outliers.

For each scenario we compute technical efficiency for OLS, COLS with SPSS software, SFA with FRONTIER 4.1 by Coelli (1996b), DEAcrs and DEAvrs with DEAP 2.1 by Coelli (1996a) and MLP with S-PLUS software. Previous to train the MLPs we split data in two parts, training and validation sets, we choose a typical rule of thumb on a 80:20 ratio. Normally, the model is developed on the training set and tested on the validation set. After an exploratory analysis we test how error differences for training and validation patterns was almost identical so we decide to join in-sample (training set) and out-of-sample (validation set) estimations for computing estimated output. We did a search 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 we stopped training when 500 iterations was reached. Neural networks validation sets estimations closer to y* (MLP Best) were selected for comparisons with remaining techniques[5].

4. The Results

We calculate Pearson's correlation coefficients[6] between estimated and true efficiency scores for all techniques over all scenarios (Table 1).

According with results displayed in table 1, MLP results best in all cases except one. Note that compared with others techniques MLP obtains robustness estimations with few variations respect true efficiency over number of DMUs and injected noise. MLP is more superior to traditional techniques when underlying technology is under moderate noise together with more DMUs. However, our results show how DEA with variable returns to scale is superior (not significant) to ANN with a lot of efficiency-noise and few DMUs.

---

[4] Zhang and Bartels (1998) also assume free of noise data. Nevertheless we would obtain identical results in this experiment if we decompose the error term in a normal error variable iid $u \sim N(0, \delta^2)$ and in a half normal efficiency variable iid $v \sim N\left|(0, \delta_v^2)\right|$

[5] 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.

[6] We also compute Spearman´s rank correlation coefficients with similar results

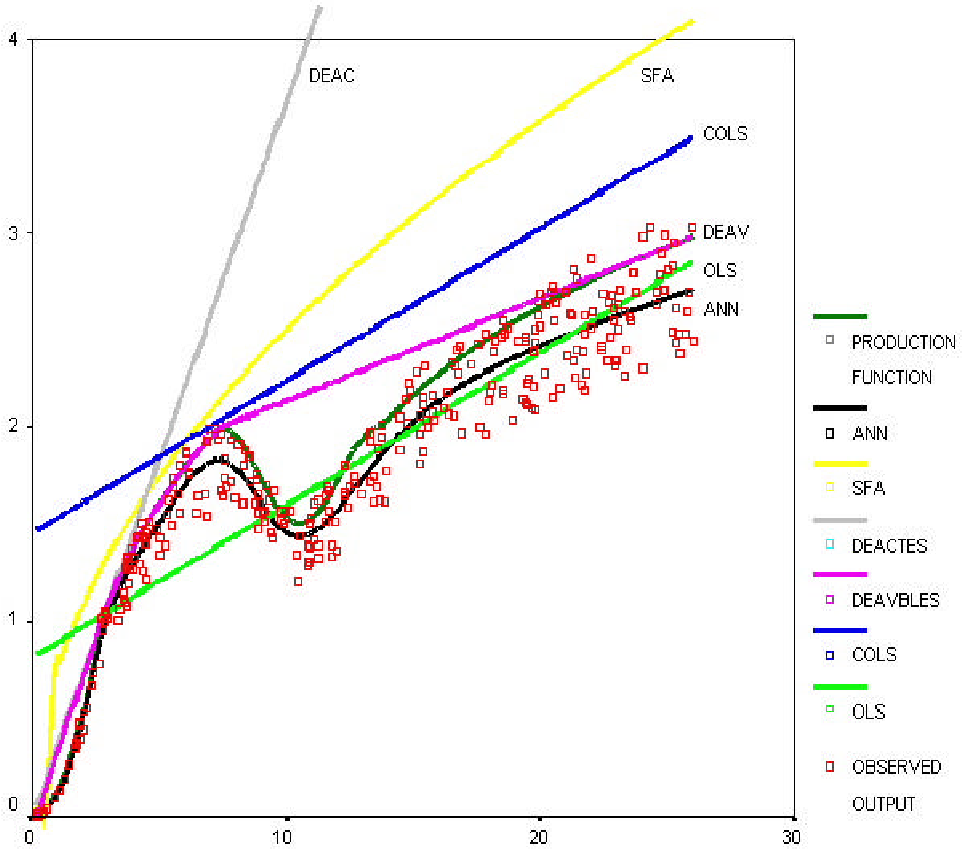
Table 1. Pearson's correlation coefficients between estimated and true efficiency scores for different techniques, number of DMUs and different quantities of injected noise.

Number of DMUs & Percentage of noise injected

	Efficiency Techniques					
50 DMUs	**OLS**	**COLS**	**SF**	**DEACtes**	**DEAVbles**	**MLP_BEST**
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	**OLS**	**COLS**	**SF**	**DEACtes**	**DEAVbles**	**MLP_BEST**
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	**OLS**	**COLS**	**SF**	**DEACtes**	**DEAVbles**	**MLP_BEST**
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	**OLS**	**COLS**	**SF**	**DEACtes**	**DEAVbles**	**MLP_BEST**
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	0,607	0,663	0,319	0,736	0,935
300(75)	0,759	0,722	0,804	0,541	0,857	0,973

In figure 2 we illustrate a particular example for 300 DMUs and when 25% of uniform noise is injected in true output. After drawing true frontier and all efficiency estimations provided by the different approaches we observe how MLP is the only technique able to find out the non-linearity contained in data. We see that MLP is an average performance technique although we could do MLP becomes a frontier moving upwards the curve up to the higher residual as we usually do with COLS. Through figure 2, we can also see how ANNs is a good tool, 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 we augment number of inputs, outputs and contextual variables implied in our production process.

Figure 2: Production functions estimated by different techniques



In Table 2, we make a rough comparison among all techniques that we can apply in an efficiency problem.

Table 2: Comparison of different approaches for estimating an unknown production function

Comparative factor	OLS, COLS	SFA	DEA	MLP
Statistical and functional form assumptions	Strong	Strong	Modest	None
Flexibility	Low	Modest	Modest	High
Theoretical basis, efficiency studies	Strong	Strong	Strong	Weak
Statistical significance	Yes	Yes	Yes	None
Interpretability of results	High	High	Modest	Modest
Projection, generalisation of results	High	High	None	High
Cost of analysis	Low	Low	Low	High
Kind of frontier	Stoch/Deter	Stoch	Determ	Stoch.
Adv. For exploring non-linear relationships	Low	Modest	Modest	High

Source: Own elaboration from Costa and Markellos

Table 2 shows how no single approach appears to be overall superior to the remaining techniques. This fact points out how the efficiency technique should be chosen according with the problem we afford. For those cases where we have enough DMUs, we suspect the possibility of non-linear relationships between variables and we do not have a strong theoretical model about the production technology, ANNs can be an alternative to fit production functions versus traditional techniques.

5. 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 free of a priori assumptions when estimating a noisy non-linear model that allow us to prevent functional forms misspecifications and to test if there exist an underlying structure in the available data.

Although we believe that ANNs can be a potential alternative for measuring technical efficiency and outperform other techniques results when the production process is unknown, it seems reasonable 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 empirical work in efficiency and comparison analysis. On the other hand, here we only concentrate on 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.

References

Álvarez, A. (2001): "La Medición de la Eficiencia y la Productividad". Ed. Pirámide.

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

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

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 y Markellos, R (1997): "Evaluating public transport efficiency with neural networks models". Transportation Research. Part C. Vol. 5C, Nº5. October.

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

Eide, E. y Showalter, M. H. (1998): "The effect of school quality on student performance: A quantile regression approach". Economics Letters 58, pp. 345-350.

Figlio, D. N. (1999): "Functional form and the estimated effects of school resources". Economics of Education Review Vol.18. Pp. 241-252.

Fried, H. O., Lovell, C. A. K. 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 2: 183.

Guermat, C. y Hadri, K. (1999): "Backpropagation Neural Network Versus Translog Model in Stochastic Frontiers: A Monte Carlo Comparison". Discussion Paper in Economics Nº 99/16, University of Exeter.

Hanushek, E. (1986): "The economics of Schooling". Journal of Economic Literature. Vol 24, nº3, pp. 1141-1171.

Hashem, S. (1993): "Optimal linear combinations of neural networks". Doctoral dissertation. University of Purdue.

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

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

Kuan, C. and White, H. (1994): "Artificial neural networks: an econometrics perspective". Econometrics Review, 13 (1): 1-91.

Lee, T., White, H. and Granger, C. W. J. (1993): "Testing for neglected nonlinearity in time series models: a comparison of neural networks methods and alternative tests". Journal of Econometrics, 56 (3): 269-91.

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

Rumelhart, D. E.; McClelland, 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.

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.