A Comparison of Data Envelopment Analysis and Artificial Neural Networks as Tools for Assessing the Efficiency of Decision Making Units

Antreas D. Athanassopoulos; Stephen P. Curram

The Journal of the Operational Research Society, Vol. 47, No. 8 (Aug., 1996), 1000-1016.

Stable URL:
http://links.jstor.org/sici?sici=0160-5682%28199608%2947%3A8%3C1000%3AACODEA%3E2.0.CO%3B2-9

The Journal of the Operational Research Society is currently published by Operational Research Society.

Journal of the Operational Research Society (1996) 47, 1000–1016 0160-5682/96 $12.00

A Comparison of Data Envelopment Analysis and Artificial Neural Networks as Tools for Assessing the Efficiency of Decision Making Units

ANTREAS D. ATHANASSOPOULOS AND STEPHEN P. CURRAM
University of Warwick

This paper is concerned with the comparison of two popular non-parametric methodologies—data envelopment analysis and artificial neural networks—as tools for assessing performance. Data envelopment analysis has been established since 1978 as a superior alternative to traditional parametric methodologies, such as regression analysis, for assessing performance. Neural networks have recently been proposed as a method for assessing performance. In this paper, we use a simulated production technology of two inputs and one output for testing the success of the two methods for assessing efficiency. The two methods are also compared on their practical use as performance measurement tools on a set of bank branches, having multiple input and output criteria. The results demonstrate that, despite their differences, both methods offer a useful range of information regarding the assessment of performance.

Key words: data envelopment analysis, neural networks, non-parametric methods, technical and scale efficiency, simulation, bank branches

INTRODUCTION

In recent years there has been a widespread tendency in disciplines such as economics, business modelling, operational research, and finance to use mathematical modelling techniques to support decision making. At the same time, however, the limitations of the traditional modelling approaches have been widely recognised, in particular the excessive structure and assumptions that accompany their development. Recent trends show alternative development paths of methods which are systematic in the approach but aim to keep the number of assumptions needed to a minimum. This paper focuses on the use of non-parametric mathematical modelling methods for assessing performance. Two methodologies, namely data envelopment analysis (DEA) and neural networks, have been selected and will be tested for their ability to assess the performance of activity units.

DEA developed in operational research[1] and economic[2] literature as a method for assessing the efficiency of activity units making the minimum possible assumptions regarding the functional form of the underlying production function. DEA is a linear programming based method that has been used extensively for assessing the relative efficiency of activity units of non-profit (e.g. schools[3], hospitals[4] and local authorities[5,6]) and for-profit (e.g. banks[7], restaurants[8], public houses[9], corporate performance[10,11]) organisations. The full technical details of DEA will not be discussed here, but reviews can be found in Boussofiane *et al.*[12] and Fried *et al.*[13].

Neural networks are a family of models which are loosely based on the structure of neurons in the brain. They are made up of simple processing units which are linked by weighted connections to form structures that are able to learn relationships between sets of variables. Initial research on neural networks focused on discovering the way that the brain works, but they have since been exploited for their mathematical properties as signal processors and statistical models. The type of neural network used in this study is the Multilayer Perceptron (MLP) originally developed by Rumelhart *et al.*[14]. It can use both continuous and classification data, being used for non-linear regression[15] and classification[16]. The MLP has been used for assessing the performance of units against an expected performance in the example data set *branches* which has been made available with the 4Thought software package[17].

Correspondence: S. P. Curram, University of Warwick, Warwick Business School, Coventry CV4 7AL, UK

The comparison between DEA and neural networks is pursued at two levels. In the first level, the study will focus on the ability of the two methods to disentangle efficient and inefficient units in a controlled experiment. The experiment is based on simulated data from a known production function with differing levels of inefficiencies and random noise. The use of simulated technologies for comparing the performance of DEA with other techniques is used as a means of ascertaining the advantages and disadvantages of different methodologies. Alongside the simulated study, DEA and neural networks are compared on a multiple-input multiple-output data set of 250 commercial bank branches. This comparison focuses on the ability of the two methods to give useful managerial insights concerning the performance of individual branches.

The rest of the paper unfolds as follows. In the second section we give a brief review of data envelopment analysis and neural network mechanisms in order to facilitate the definition of efficiency measures from the two methods. The third section concerns the design of the simulated production function and the analysis of the results obtained from the two methods. The fourth section contains the results from the comparison of the two methods as tools as assessing the performance of bank branches. In the fifth section the paper is concluded and further research issues are discussed.

DATA ENVELOPMENT ANALYSIS AND NEURAL NETWORKS FOR ASSESSING PERFORMANCE

DEA can be seen as a generalisation of total factor productivity methods in the sense that efficiency is defined as a multivariate ratio of weighted sum of outputs to weighted sums of inputs. Consider a set of $j = \{1, \ldots, n\}$ decision making units (DMUs) with each DMU using x_{ij}, $i = \{1, \ldots, m\}$ input quantities to produce y_{rj}, $r = \{1, \ldots, s\}$ output quantities. The efficiency of unit k under the assumption of constant and variable returns to scale is given by the solution to the optimisation problems in (1) and (2), respectively. The formulation of these models has been proposed by Charnes *et al.*[1] and Banker *et al.*[18].

DEA models for assessing efficiency

Constant Returns to Scale (CRS)	Variable Returns to Scale (VRS)
$\min_{u_r, v_i} \sum_{i=1}^{m} v_i x_{ik}$	$\min_{\alpha_r, \beta_i} \sum_{i=1}^{m} \alpha_i x_{ik} - \omega$
S.t.	S.t.
(1)	(2)
$\sum_{r=1}^{s} u_r y_{rk} = 100$	$\sum_{r=1}^{s} \beta_r y_{rk} = 100$
$\sum_{i=1}^{m} v_i x_{ij} - \sum_{r=1}^{s} u_r y_{rj} \geqslant 0, \forall j$	$\sum_{i=1}^{m} \alpha_i x_{ij} - \sum_{r=1}^{s} \beta_r y_{rj} - \omega \geqslant 0, \forall j$
$v_i, u_r \geqslant \epsilon$	$\alpha_i, \beta_r \geqslant \epsilon$ and ω unrestricted

where

v_i, α_i are weight factors for input i estimated by the two models,
u_r, β_r are weight factors for output r estimated by the two models,
ω is a sign free variable that characterises the variable returns to scale model,
ϵ is a non-archimedian small positive number.

The efficiency of unit k is obtained under constant returns to scale as $E_k^{CRS} = 1/\sum_{i=1}^{m} v_i^* x_{ik}$ and under variable returns to scale as $E_k^{VRS} = 1/\sum_{i=1}^{m} \alpha_i^* x_{ik} - \omega^*$ (where the asterisks indicate optimal values). The managerial and economic implications of selecting either of the two models for assessing performance are discussed in more detail by Athanassopoulos[19]. The assessment of the effi-

ciency of individual DMUs has many managerial by-products regarding the estimation of performance targets for inputs and outputs and the identification of benchmark operating practices. Key issues in the development of a DEA assessment are:

- The comparison of sets of decision making units that perform uniform operations;
- The selection of an input-output set that encompasses the operating process of the DMUs. Note here that there is an implicit causal assumption such that for efficient DMUs higher levels of inputs should lead into higher levels of outputs;
- The selection of an appropriate mathematical model of assessment, notably input or output orientation, constant or variable returns to scale and radial or non-radial projection[19].

Since DEA is a non-parametric deterministic methodology there are no statistical mechanisms for investigating the appropriateness of the assumptions in the three stages alluded to above.

The artificial neural network approach

A neural network 'learns' relationships between input and output variables through being repeatedly shown example data and changing the internal structure of the network to represent the relationships more closely.

The network is made up of simple processor neurons organised into layers with each neuron in a layer having weighted connections (coefficients) to all neurons in neighbouring layers. These weights can take positive or negative values. The input layer and output layer represent the input and output variables of the model and between them lie one or more hidden layers which hold the network's ability to learn non-linear relationships. The greater the number of neurons in the hidden layers, the more able the network is to cope with non-linear relationships.

The data used by the network must be scaled for the network to be effective. In theory the inputs to the network can be any value, however scaling values to the same order of magnitude (generally in the range 0 to 1 or -1 to 1) enables the network to learn relationships quicker and using fewer hidden nodes[20]. In fact, we scaled the input data to the range 0.2–0.8 so as to use a consistent scaling regime for inputs and outputs (see below).

The activation from an output neuron is in the range 0 to 1 due to the effects of a squashing (sigmoid) function so the output values must also be scaled on this range. We used a scaling regime of 0.2–0.8. This allows room for extrapolation of output variables to be outside the range of the original data, as may be required for new examples or sensitivity analysis.

Training involves repeatedly presenting the data to the network. Learning is achieved in the network by altering the values of weighted connections between neurons to bring the output of the network closer to the desired target value. The overall aim is to reduce the Mean Squared-Error (MSE) for the training data. The errors between output and target values are propagated back through the network to attribute them to the weights in the network, these are then altered using the steepest descent method, which aims to reduce the MSE by following the steepest gradient on the error surface. The rate of learning is controlled by gain and momentum parameters. The gain parameter specifies the magnitude of changes to the weights. A small gain term results in slow network learning while a large gain term can miss key features on the error surface leading to oscillation or convergence to local minima. Historical changes to weights are stored as an exponentially smoothed average, or momentum value for each weight, a proportion of this momentum is used in future weight changes so as to smooth learning and reduce oscillation.

Two key disadvantages of the multilayer perceptron are its slow learning speed and its ability to converge to local minima. These are particularly affected by the choice of value for the gain parameter. To help overcome the difficulty in setting the gain parameter we used the adaptations suggested by Vogl *et al.*[21]. Here the gain term is not fixed but is allowed to vary depending on the success of the learning. The gain term is allowed to increase while training results in improvements in the error measure, representing increased confidence in the direction of learning, and is reduced if training leads to an increase in the error measure.

A key aspect in setting up the network is deciding on the size of the hidden layer. The more complex the interactions between the variables, the more neurons are required. If too few hidden neurons are used, the network will fail to learn the richness of relationships, while if too

many are used the data may be overfitted, fitting to individual data points rather than the trend and so reducing the network's ability to generalise. The actual number of neurons required in the hidden layer must be found by trial and error.

An alternative to optimising the number of hidden nodes is to use an independent validation step in the training of the network[22]. Here the network is required to have sufficient nodes to fit the trend but is not streamlined to prevent overfitting. Hoptroff[22] suggests that 10 nodes in one hidden layer are usually sufficient for most forecasting and business problems. More nodes can be used but usually result in slower learning without an improvement in results.

The approach requires that independent validation data are used to test how well the network is able to generalise to unseen data. The validation data are taken out from the training data and should be representative across the range of outcomes. A larger validation set is likely to be more representative; however this does take data away from the training set. It is therefore necessary to strike a balance between the training and validation data set sizes. When large amounts of data are available the selection of validation data can be done using a simple random choice. Where the amount of data is limited it is useful to use a stratification approach. This approach is easiest where there is only one output variable. Here the data is sorted by the value of the output variable, partitioned and one validation data point is chosen at random from each partition. In this way the stratification tries to ensure that validation data is chosen across the range of outcomes.

Each iteration of the training process is as follows. The network is presented with a set of training examples from which weight adjustments are made, then the network is tested using independent validation data to find the ability of the network to fit unseen data. Training is stopped at the iteration where the MSE for the validation data is minimised. This represents the point in training where the network is best able to generalise.

In practice it has been found that the MSE of the validation data can worsen but then improve again. To overcome this, training is usually continued for some time after the optimum point has been reached to ensure that no further reduction in the MSE will occur. In our case, the network weights and biases are saved every time the MSE of the validation set is reduced, and training of the network is stopped if 2000 iterations occur without an improvement. This means that the final saved network represents the optimal training point while the extra iterations offer a large safety margin to ensure that a better stopping point will not have been missed. If the MSE of the validation set does not permanently get worse (i.e. no optimum point in training is reached), it suggests that the network does not contain enough nodes to overfit the data and is unlikely to be able to pick up the full underlying trend. In this situation training is restarted using more nodes in the hidden layer.

The training process results in a model which has a generalised fit through the performance data. The model suggests an expected (average) performance given the inputs from the DMU against which the actual performance is compared. The difference between the expected output given by the network and actual performance indicates the degree to which the DMU is exceeding or falling short of the expected performance. The calculations of the actual performance measures are given below.

Defining DEA and neural network measures of efficiency

Both DEA and neural networks are non-parametric methods in the sense that no assumptions are made concerning the functional form that links the inputs and outputs used to describe an operating process. In DEA a set of weights is assessed for the inputs/outputs of each DMU in order to maximise its relative efficiency subject to the efficiency of the other DMUs in the study. Neural networks are also based on the estimation of sets of weights that link inputs with outputs. In the latter case, however, the weights seek to derive the best possible fit through the observations of the assessed data set.

The neural network efficiency will be determined using the predicted values obtained from the solution to the model. These efficiencies can be obtained either taking the ratio between the observed and predicted values for the inuts and outputs of each DMU as listed in equation (3) or by taking the more extreme transformation listed in (4). Thus the unscaled, E_{rj}^{NNU}, and scaled, E_{rj}^{NNS},

efficiencies for the output r of DMU j can be obtained as:

$$E_{rj}^{NNU} = \frac{y_{rj}}{y_{rj}^{Pr\,e}} \quad \forall r \tag{3}$$

or

$$E_{rj}^{NNS} = \frac{y_{rj}}{y_{rj}^{Pr\,e} + \max_j(Rn_{rj})} \quad \forall r, \tag{4}$$

where

y_{rj} is the observed output of DMU j,
$y_{rj}^{Pr\,e}$ is the predicted output by the solution to the neural network for DMU j,
Rn_j is the residual obtained by the neural network for the DMU j.

Using the formulae in (3) the efficiency estimates are not bounded to be less than or equal to unity and thus there is not sufficient comparability with the corresponding efficiencies obtained from DEA. In fact, using this formula, unity implies that the unit is achieving average performance in this output. The transformation suggested in (4) can be used in order to get efficiencies that do not exceed unity. This will lead, however, into rating as efficient the DMU for each output r with the highest residual value $\max_j Rn_{rj}$. This maximum residual value has implications on the efficiency magnitude of the inefficient DMUs particularly when the data is distorted by random noise and measurement errors. These extreme transformations should only be used in line with diagnostic tests that are available in multivariate analysis in order to detect outliers and influential observations. The problem of extreme observations can be alleviated partially by applying formulae (4) to different segments of the distribution of the dependent variable which leads thereby to the definition of the so-called *thick frontiers*. This issue has not been pursued any further in this paper.

It must be borne in mind that the definition of efficiency has a different basis across DEA and neural network models. DEA efficiency is gauged from empirically defined production functions whilst neural network efficiency is gauged from a 'regression-type' average non-parametric curve fitted on the data. In the next section DEA and neural networks are compared as tools for assessing efficiency using a known production technology. The knowledge of the real efficiency of individual DMUs will be used as a benchmark for assessing the strength of each method as an efficiency assessment tool.

COMPARING DEA WITH NEURAL NETWORKS USING A KNOWN TECHNOLOGY

Bowlin[23] was the first to use simulated data and compare DEA with ordinary regression methods on a study which was later replicated and extended by Thanassoulis[24]. The first systematic simulation study comparing DEA with econometric techniques like the translog production function was done by Banker *et al.*[25]. This work was extended by Gong and Sickles[26] by comparing DEA with panel data estimation methods on an artificial panel data set. A later development on the simulation front was also made by Banker *et al.*[27] where DEA was compared with corrected ordinary least squares (COLS) methods. The most recent experiment in the use of DEA on simulated data is reported by Smith[28] and examines DEA on its ability to handle various types of mispecified technologies. These experiments use different known technologies and also adopt different assumptions regarding the efficiency distributions. The different design adopted by various simulation experiments makes the results of these studies context and assumption specific with limited generalisation flexibility.

Experimental design

The experiment is based on the Banker *et al.*[27] study where DEA was compared with Corrected Ordinary Least Squares. This is based on a known Cobb-Douglas technology with two inputs and

one output. The Banker *et al.*[27] study sought to compare DEA with COLS based on four criteria: technology, inefficiency, distribution of measurement error and sample size, which will also be considered in this study. The experiment is based on a Cobb-Douglas production function which takes the general form in (5).

$$y = \gamma \prod_{i=1}^{m} x_i^{\beta_i}, \tag{5}$$

where x_i are quantities of inputs $i = 1, \ldots, m$ required to produce quantities of output y. The γ factor represents a fixed term of the Cobb-Douglas production technology. The elasticity of output in relation to input x_i is given by β_i and represents the proportional change to output per unit of change to inputs. The sum of the input elasticities $\sum_i \beta_i$ can be used to characterise the presence of increasing ($\sum_i \beta_i > 1$), decreasing ($\sum_i \beta_i < 1$) or constant ($\sum_i \beta_i = 1$) returns to scale respectively.

The production technology was developed using two inputs K and L with output elasticities a_i and b_i, respectively and it was given a piece-wise form by four segments as described in (6).

$$y = \begin{Bmatrix} \gamma_1 K^{a_1} L^{b_1} & 5 \leqslant K \leqslant 10 & 5 \leqslant L \leqslant 10 & \text{Segment 1} \\ \gamma_2 K^{a_1} L^{b_2} & 5 \leqslant K \leqslant 10 & 10 \leqslant L \leqslant 15 & \text{Segment 2} \\ \gamma_3 K^{a_2} L^{b_1} & 10 \leqslant K \leqslant 15 & 5 \leqslant L \leqslant 10 & \text{Segment 3} \\ \gamma_4 K^{a_2} L^{b_2} & 10 \leqslant K \leqslant 15 & 10 \leqslant L \leqslant 15 & \text{Segment 4} \end{Bmatrix} \tag{6}$$

The four segments of the Cobb-Douglas production function were used to develop a concave production technology using combinations of the input elasticity coefficients (a_1, a_2, b_1, b_2) as listed in Table 1. (A non-concave production function was also tested without yielding substantially different results).

TABLE 1. *Cobb-Douglas coefficients of the production technology*

Production technology $a_1 = 0.5$, $a_2 = 0.4$, $b_1 = 0.45$, $b_2 = 0.35$	$\sum_i \beta_i$	Returns to scale
Segment 1	0.95	Decreasing
Segment 2	0.85	Decreasing
Segment 3	0.85	Decreasing
Segment 4	0.75	Decreasing

Note that the concave technology exhibits decreasing returns to scale for all input variables, but with the degree of decreasing returns to scale being larger for larger values of the input (the returns to scale factor $\sum_i \beta_i$ is getting reduced from segment 1 to segment 4).

The value of the constant term γ is used to ensure the continuity of the production function at $K = L = 10$. Therefore, fixing the output value $y = 10$ and using the corresponding production elasticities $a_1 = 0.5$, $a_2 = 0.4$, $b_1 = 0.45$, $b_2 = 0.35$ the following values were obtained for the γ factor:

$$\gamma_1 = 1.122, \gamma_2 = 1.412, \gamma_3 = 1.412, \gamma_4 = 1.778.$$

Thus far, we have shown how output quantities can be generated by the four segments of the Cobb-Douglas production technology. This process yields theoretical output quantities y given by (5), which do not entail any inefficiency or random noise. These output quantities were next distorted by fluctuations attributable to inefficiency (u_j) and also measurement error (v_j) for each DMU $j = 1, \ldots, n$. The generated observed data set would ultimately contain output quantities $\hat{y}_j$ which are given by equation (7).

$$\underbrace{\hat{y}_j}_{\text{Observed}} = \underbrace{y_j}_{\text{Efficient}} \times \underbrace{u_j}_{\text{Inefficiency}} \times \underbrace{v_j}_{\text{Random noise}} \tag{7}$$

The artificial production technology was embedded with known inefficiencies (u_j) using four inefficiency distributions.

1. Half normal, $|Normal(0, 0.2036)|$.
2. Half normal, $|Normal(0, 0.2796)|$, with 25% of the true outputs on the efficient frontier.
3. *Exponential* with mean = 0.1625.
4. *Exponential* with mean = 0.2231, with 25% of the true outputs on the efficient frontier.

Random data was generated according to the above four distributions and then subtracted from 1 to give the efficiency value of each observation. This efficiency value was then multiplied by the value of y which is generated by the Cobb-Douglas production functions in Table 1. These values were multiplied again by the measurement error (random noise) that is embodied on the data in order to obtain the 'observed' output value for each observation. An extra error term (v_j) was introduced in the analysis attributable to random fluctuations in the measurement of individual observations. These types of errors do not correspond to any type of inefficiencies but they can, indeed, distort the efficiency rating of individual activity units if not distinguished during the assessment of efficiency. Note that earlier simulation studies[23,24,25] do not cater for the presence of error terms attributable to productive inefficiency and measurement errors respectively.

The experiment comprises both 'low' and 'high' normally distributed measurement errors. Low measurement errors lead to composite error terms dominated by the inefficiency distribution whereas high measurement errors lead to composite error terms dominated by the measurement error distribution. The measurement errors were generated at the logged (ln) data level. Both low and high measurement errors have the mean value of 0.0 but different variance. The characteristics of the distributions concerning 'low' and 'high' measurement errors are given in Table 2.

TABLE 2. *Distributions of low and high measurement errors*

Low measurement errors	High measurement errors
Mean = 0	Mean = 0
Variance (row) = 1.002	Variance (row) = 1.020
2.5% quantile of row data = 0.91	2.5% quantile of row data = 0.75
97.5% quantile of row data = 1.09	97.5% quantile of row data = 1.33

Following the above methodology, a series of 16 data sets was generated and tested using the artificial neural networks (ANN) and DEA models. The DEA assessment was pursued applying the constant and variable returns to scale models in equations (1) and (2). We have used both models for assessing DEA efficiency since in real applications there is no objective mechanism to assess the validity of constant or variable returns to scale. Thereby, the differences between the estimates obtained in our experiment can be used to derive the bias in the assessed efficiencies that can be attributed to the adoption of particular economic assumptions in the DEA model.

The neural network models used a 2*3*1 architecture. That is, 2 input nodes, a hidden layer with 3 nodes, and 1 output node. Training involved independent validation using 20% of the data separated by a stratification approach (see review of artificial neural networks earlier).

Using the independent validation approach, it would have been possible to use a larger hidden layer but the training time would have likely to have been longer due to the extra calculations which are required per iteration. The use of three hidden nodes was found to be sufficient for all the data sets. That is, a minimum was found for the mean square error of the independent validation set, indicating that the number of nodes in the hidden layer was sufficient to fit the trend and then go on to overfit.

Results and analysis

The comparative study of the differences between the theoretical efficiency model and the alternative methods for assessing efficiency will include three main phases. The first phase concentrates on the absolute deviations between the assessed and the theoretical (true) efficiencies. The second phase focuses on the relative rankings of DMUs obtained from the alternative efficiency assessment tools. Finally, the last phase of the empirical analysis seeks to identify the characteristics of the experimental procedure that have influenced the results obtained.

DEA and ANN were compared using the mean absolute deviations (MAD) between the efficiency assessed by DEA and ANN and the theoretical (true) efficiency ($y_j \times u_j$) respectively. Four MAD differences were defined for each assessed DMU k as follows:

$$\text{MAD}_1 = \frac{1}{n}\sum_{k=1}^{n} |E_k^{\text{True}} - E_k^{CRS}|, \qquad \text{MAD}_2 = \frac{1}{n}\sum_{k=1}^{n} |E_k^{\text{True}} - E_k^{VRS}|,$$

$$\text{MAD}_3 = \frac{1}{n}\sum_{k=1}^{n} |E_k^{\text{True}} - E_k^{NNU}|, \qquad \text{MAD}_4 = \frac{1}{n}\sum_{k=1}^{n} |E_k^{\text{True}} - E_k^{NNS}|.$$

Results concerning the MAD scores of each experiment are listed in Table 3.

The results from Table 3 show a general pattern of superiority of DEA over the ANN. The results, however, are affected to a large extent by the structure of the experiment that was used to generate the data. Under the presence of high random noise and with a large sample size the unstandardised ANN model yielded a MAD value of 0.13 which was lower than the corresponding DEA ones but statistically significant for only the case of the VRS model. The latter was the only case where the ANN gave superior results to that of DEA. A summary of more general observations that emanate from Table 3 is next provided.

TABLE 3. *Mean absolute deviation between the efficiencies*

	% Efficient	Sample	Measurement error			
			0.002		0.02	
			25	100	25	100
Half normal	0%	[1]CRS	$0.03^{3,4}$	$0.05^{3,4}$	0.09^{4}	0.14^{4}
		[2]VRS	0.05^{3}	0.06^{3}	0.09^{4}	0.18^{4}
		[3]NNU	0.18	0.15	0.16^{4}	$0.13^{2,4}$
		[4]NNS	0.06^{3}	0.06	0.22	0.22
	25%	[1]CRS	$0.03^{3,4}$	$0.06^{3,4}$	$0.10^{3,4}$	0.16^{4}
		[2]VRS	0.06^{3}	$0.04^{1,3,4}$	$0.08^{3,4}$	$0.12^{1,3,4}$
		[3]NNU	0.19	0.14	0.22	0.18^{4}
		[4]NNS	0.06^{3}	0.12	0.23	0.26
Exponential	0%	[1]CRS	$0.04^{2,3}$	$0.05^{3,4}$	$0.09^{3,4}$	0.17^{3}
		[2]VRS	0.05^{3}	$0.05^{1,3,4}$	$0.08^{1,3,4}$	$0.13^{1,3,4}$
		[3]NNU	0.15	0.17	0.18	0.18^{4}
		[4]NNS	0.06^{3}	0.10^{3}	0.21	0.24
	25%	[1]CRS	$0.04^{3,4}$	$0.07^{3,4}$	$0.09^{3,4}$	0.18^{4}
		[2]VRS	$0.03^{3,4}$	$0.05^{1,3,4}$	$0.07^{3,4}$	$0.14^{1,4}$
		[3]NNU	0.17	0.15	0.23	0.17^{4}
		[4]NNS	0.15	0.12	0.17	0.26

[1,2,3,4] Denotes significantly smaller MAD via the Wilcoxon matched-pairs signed-rank test at 95% significance level.
CRS: constant returns to scale.
VRS: variable returns to scale.
NNU: neural network score unstandardised.
NNS: neural network score standardised.

- Overall there is no clear winner between the CRS and VRS models in terms of lower MAD scores.

The MAD scores for the CRS and VRS models do not indicate a clear winner. However, under high random noise assumptions the VRS model yields lower MAD scores than the corresponding CRS ones. In the case of high sample sizes these differences are statistically significant. The adoption of two parametric distributions (half-normal and exponential) to incorporate amounts of technical inefficiency within the data do not seem to affect substantially the accuracy of the two DEA models. In a more general context, the adoption of constant or variables returns to scale assumption remains a problem that cannot be addressed retrospectively after the efficiency results have been obtained and it has a very strong bearing on the economic assumptions that follow the assessment of efficiency.

- The comparisons between the unstandardised and the standardised ANN efficiencies did not yield a stable pattern of MAD values across all data sets.

The high or low random error term seems to affect substantially the accuracy of the two ANN models. That is, the standardised ANN efficiencies were more accurate in the case of the low measurement errors whilst the reverse case holds in the case of the high measurement errors. In this latter case, the high measurement errors lead into high residual values which underestimate the true efficiencies when the standardised ANN efficiency is estimated.

The results from the non-parametric statistics revealed some noteworthy cases regarding the significance of the deviations between the assessed and the real efficiency measures. For example, in the case of half-normal inefficiency distribution with 100 sample size and low measurement error the CRS and NNS have very close magnitudes of MAD differences but a statistically significant Wilcoxon test result. Similar arguments hold, for the exponential distribution, 25 sample size, 0% efficient observations and low measurement error, between CRS and VRS efficiency values.

- The ANN models have shown high accuracy (similar to DEA) in ranking the assessed DMU in accord with their true rank order.

Alongside the measurement of the MAD between the assessed and true efficiencies, of similar importance is the question of how accurately are individual DMUs ranked by alternative efficiency methods. Table 4 lists the Spearman rank correlation coefficients between the CRS, VRS and NNU efficiencies and the real efficiency of each DMU. Note that the two ANN models will yield the same relative ranking of DMUs and thus only one of them was kept in the analysis.

The results in Table 4 show varying rank correlation scores depending on the characteristics of the data set and the method adopted for assessing efficiency. On the whole, the rank correlations demonstrated remarkably good results for rankings obtained from the ANN efficiencies compared to the earlier picture of the MAD score differences. This observation is particularly true under the assumption of half-normally distributed inefficiency terms. On the DEA side the two models, namely CRS and VRS, seem to exhibit an interchangeable lead one over the other depending on the data set tested.

As has already been shown in Table 3 the high magnitude of measurement error (0.02) for a part of the experiment affects the accuracy of the efficiency models considerably. In Table 3 this concerned the absolute deviations between the assessed and the true efficiencies. These differences are reinforced here in Table 4 when the relative rankings obtained from the alternative methods are compared. The sample size in the case of the high measurement errors seems also to affect the

TABLE 4. *Spearman rank correlation coefficients* between estimated and real efficiency*

Inefficiency	% Efficient	Sample	Measurement error			
			0.002		0.02	
			25	100	25	100
Halfnormal	0%	CRS	0.89	0.85	**0.81**	0.63
		VRS	0.69	**0.89**	0.57	0.66
		ANN	**0.90**	0.70	0.58	**0.67**
	25%	CRS	0.85	0.90	0.58	**0.80**
		VRS	0.85	0.88	0.57	0.72
		ANN	**0.86**	**0.91**	**0.63**	0.79
Exponential	0%	CRS	0.81	**0.80**	**0.78**	0.59
		VRS	**0.84**	0.79	0.63	0.52
		ANN	0.67	0.72	0.70	**0.64**
	25%	CRS	0.87	**0.87**	0.68	**0.67**
		VRS	**0.89**	0.73	0.65	0.66
		ANN	0.75	0.77	**0.69**	0.59

* All results show significant correlation at the 1% level.
Best performing method in each category of data shown in bold.

relative rankings obtained by different methods. For example, the sample of 25 is not ranked very successfully when 25% of the observations are on the efficient frontier in both the half-normal and exponential inefficiencies.

Adoption of a different distribution of inefficiency has affected more than the ANN rankings and to a lesser extent the DEA rankings. Despite the general reduction in the rank correlations, when the exponential distribution was adopted the VRS results seem to be the least affected.

- A regression analysis of the absolute deviations between the assessed and the true efficiencies verified the relative superiority of the VRS model over all other models of assessment.

The regression included as the dependent variable the absolute deviations (DEV) between the efficiencies obtained by the assessment methods and the true efficiency. The independent variables selected were the sample size (n_{25}, n_{100}) and the measurement error ($\sigma^2_{0.02}$, $\sigma^2_{0.002}$). Dummy {0,1} variables were also introduced representing the selection of half-normal and exponential inefficiency distribution, and the use of CRS, VRS, NNU, NNS methods for assessing efficiency. The interaction between sample size and measurement error was also tested as an independent variable. Results obtained from the regression model are given in Table 5.

TABLE 5. *Absolute deviations regressed on the factors of the experiment*

Explanatory variables	Dependent variable: Absolute deviation between assessed and theoretical efficiency
Inefficiency {0 normal, 1 exponential}	34.43
Measurement error {0.02, 0.002}	279.4
Sample size {25, 100}	−0.0615
VRS {0, 1}	−3.6
CRS {0, 1}	−5.85
NNU {0, 1}	13.82
Inefficiency × VRS	−35.79
Inefficiency × CRS	−33.37
Inefficiency × NNU	−33.26
Inefficiency × NNS	−32.65
Measurement error × CRS	66.94
Measurement error × NNU	−404.28
Measurement error × NNS	302.26
Sample size × VRS	0.079
Sample size × CRS	0.0959
Sample size × NNS	0.07
Measurement error × sample size	2.34
Constant term	6.57

All coefficients different from 0 at the 1% significance level.

The regression analysis results listed in Table 5 shed more light on the results already obtained from the previous analysis. The regression model gave a statistically significant F-test value with a coefficient of determination $R^2 = 30\%$ (which is compatible with the fit obtained from the earlier experiments of Banker *et al.*[27]). The regression model includes a set of explanatory variables directly associated with the design of the theoretical production function. A second set of variables was also used in order to ascertain the impacts of the interactions between the original explanatory variables on the dependent variable.

- The assumption of exponential distribution of the inefficiencies leads to higher efficiency deviations on average.

Within the exponential distribution results, however, the VRS model seems to be the least affected and the NNS the most affected regarding the average efficiency deviations. The latter conclusions are obtained from the interaction terms between the inefficiency and the method for assessing efficiency.

- The measurement error embodied as an extra disturbance term increases the inaccuracy of the efficiency results as it increases in size.

The response of different methods for assessing efficiency to the measurement error terms is, however, variable. The CRS and NNS models are increasingly inaccurate as the measurement error increases. The VRS model did not have a statistically significant interaction with the measurement error factors and finally the efficiency deviations of the NNU model seem to be affected negatively as the measurement error increases. The latter result shows, perhaps, a robustness of the NNU model with regard to the measurement error component that can be prevalent with data sets. On the other hand, the results also show how distorting random noise can be on the assessment of efficiency via deterministic methods like DEA.

- The size of the samples used in the assessment has a positive impact on the efficiency deviations which varies, however, by the type of method adopted for assessing efficiency.

The worse affected method as the size of the sample increases is the CRS, and the least affected the NNS. This result reflects the underlying mechanism of the two approaches. The CRS method allows scale size extrapolations of efficient DMUs to be compared with the inefficient ones, and thus as the sample size increases the effect of scale size differences on the assessed efficiency also increases. In the case of the NNU results the increased sample size has a positive effect on the accuracy of the efficiency estimates, although this was not found statistically significant. When the NNU efficiencies are standardised to obtain the NNS results the increasing sample sizes increase the efficiency deviations as a result of the outlier observations that are used to standardise the assessment. Finally, the interaction between measurement error and sample size has an inevitable positive impact on the deviations between assessed and real efficiencies.

Summing up the previous results regarding the comparison between data envelopment analysis and artificial neural networks one needs to consider a comparative basis of assessment. This basis will include issues regarding the predictive accuracy of the assessed efficiencies and also the relative ranking of the assessed decision making units on the basis of the assessed efficiences. Based on the evidence obtained from the simulated study, data envelopment analysis has a superiority over neural networks for determining the actual level of efficiency. However, one needs also to bear in mind the differences in the way performance is assessed by the two methods and also the different optimisation criteria adopted for measuring efficiency.

Differences between the CRS and VRS efficiency results emphasised the currently unresolved problem of selection of constant or variable returns to scale model. The results have shown that a wrong assumption regarding the effects of economies of scale can distort considerably the assessed efficiencies. DEA, unlike stochastic frontier methods (see Lovell[29]), does not provide information regarding the foundations of variable or constant returns to scale in the operation of decision making units. In other words, the ability to assess efficiency under one or the other assumption is accompanied by any yardsticks that test the appropriateness of these assumptions. Depending on the underlying assumptions of the true efficiency model, the CRS and VRS models have shown varying accuracy when approximating the true efficiency. This is undoubtedly an area of further research interest within the DEA literature.

The high performance of the neural networks with regard to the relative rankings of DMUs is encouraging regarding the method's role as a performance measurement tool. The high absolute deviations regarding the exact efficiency of individual decision making units was due to the tendency to overestimate efficiency by the NNU model (being unbounded) and underestimate efficiency by the NNS model.

DEA AND ANN COMPARED ON EMPIRICAL DATA

The application of DEA and neural networks to a set of commercial bank branches seeks to investigate the ability of the two methods to provide managerially useful information regarding issues of efficiency assessment and target setting. For the management of a multi-unit organisation, the comparative efficiency of its branches is a very important question regarding the performance and viability of the whole organisation. A DEA framework for assisting management on a series of questions regarding the performance of retail outlets has been proposed in the

literature[8,9]. A similar type of analysis is pursued in this paper using as tools of analysis DEA and neural networks.

The data set consists of 250 commercial bank branches using the input-output set in Table 6.

TABLE 6. *Input-output set for assessing market efficiency of bank branches*

INPUTS	• No. of automatic facilities • Trained personnel to sell financial products • No. of counter transactions • Estimated potential market
OUTPUTS	• Loans sales • Liability sales • Investment and insurance policies sold

The inputs listed in Table 6 concern the market characteristics in the surrounding area of individual branches (Potential market) and the resource levels committed by individual outlets to generate sales (automatic facilities, personnel and number of counter transactions).

The output side includes three levels of branch activities. The problem of aggregation of the banking output is a very involved one and outside the scope of this research. The interested reader is referred to previous studies[30,31,32] that discuss the issue more extensively.

On the DEA side the network of bank branches was assessed adopting an output maximisation orientation under constant (CRS) and variable (VRS) returns to scale assumptions. The assessed efficiency and performance targets would characterise the ability of individual branches to utilise their market potential and generate sales of financial products.

A 4*10*3 neural network was used (4 inputs, 10 hidden, 3 outputs) and trained using independent validation with 20% of the data (50 DMUs). Since the data set was large, the validation data was chosen at random from the complete set of DMUs. An unscaled efficiency term (NNU) is used for each output, where an efficiency above 100% shows greater than expected utilisation of inputs, and the converse for efficiencies of less than 100%.

The presentation of the results of DEA and the neural network will be based mainly on the assessment of performance targets and efficiency measures. The latter two measures can easily be obtained by the solution to the DEA model[33]. However, the assessment of targets and efficiencies using the neural network is more complicated. A key issue regarding the neural network assessments is the residual term Rn_j for each DMU j. The results of the neural network treat each output independently in terms of efficiency so that for many DMUs we may face the phenomenon where some outputs would have positive and some negative residuals. Hence a DMU may be deemed efficient in some outputs; but inefficient in others. Results obtained from the two DEA and the neural network models are presented in Table 7.

In the neural network results in Table 7, bank branches with an observed value higher than the corresponding expected value are classified as efficient. The CRS model yields, as expected, the least number of efficient branches while the VRS model yields higher efficiency values. The results obtained by the three efficiency models are not directly comparable as the basis of their definition, namely the 'best practice' frontier, is different. The CRS frontier incorporates scale and technical efficiencies, the VRS frontier focuses on the technical efficiency and finally the neural network efficiency is defined as a non-parametric plane fitted through the average bank branch. The differences between the methods are illustrated in Figure 1.

Figure 1 contains a single input-output production technology. The efficiency of the DMUs in the data set will be assessed with reference to a notional 'best practice frontier'. Alternative fron-

TABLE 7. *Efficient branches for each output (Total no. of branches = 250)*

	Liabilities	Loans	Insurances and Investments	All three outputs
CRS	47	47	47	47
VRS	74	74	74	74
NN[1]	109	117	120	45

[1] Observed value exceeds the expected value.

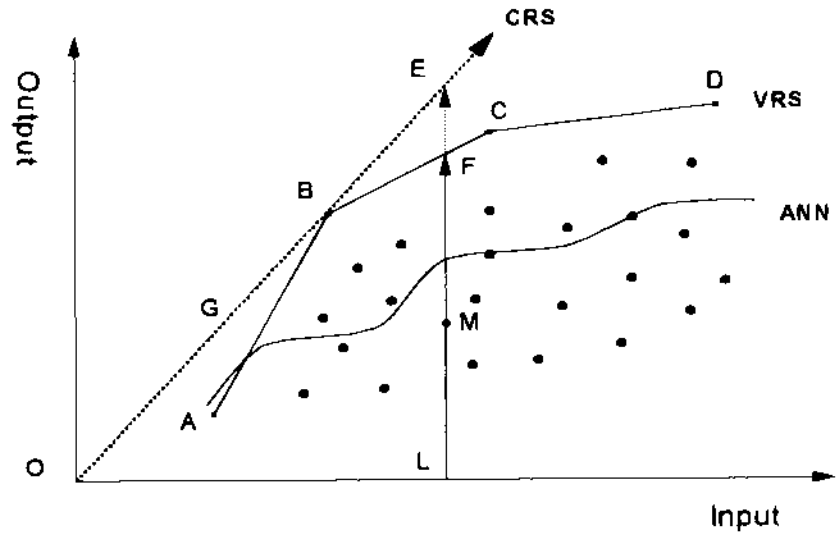

FIG. 1. *Graphical representation of DEA and ANN frontiers.*

tiers can be defined depending on the underlying assumptions adopted. For example, the constant returns to scale (CRS) frontier is defined by DMU B along the conical hull OB from the origin. The output efficiency of DMU M is gauged by the ratio LM/LE. The variable returns to scale (VRS) frontier is defined by DMUs A, B, C, and D and is defined along the convex hull ABCD. Thus, the output efficiency of DMU M is gauged by the ratio LM/LF. The immediate implication of the two DEA frontiers is that the VRS frontier is closer to the actual observations and thereby, the VRS efficiency is always higher or equal to the CRS efficiency.

The ANN best practice surface is given by the best fit curve fitted throughout the data set. The ANN frontier is a central tendency one and it yields classifications concerning above and below average efficient DMUs. It is expected, therefore, that the DEA efficiencies will be lower on average than those obtained by the ANN model. This may not be true, however, for DMUs near the tails of the VRS frontier as, for example, DMU A is VRS efficient but ANN inefficient.

The efficiency results from Table 7 were investigated further in order to identify the extent to which there is concordance in these ratings across the three methods. This investigation has shown that 23 out of 150 branches were efficient in all three methods and another six branches were efficient only according to the VRS and ANN models. In 21 cases branches were found efficient only by the ANN model. The conclusion is that there is no predictable pattern of how branches are assessed by the DEA and the ANN methods which confirms our earlier discussion that was motivated by Figure 1.

Another difference between the DEA and neural network results is that under the neural network assessment a branch may be efficient under particular inputs/outputs but not across all of them. This phenomenon is illustrated in Table 8, which contains a breakdown of the efficiencies of each branch across the three output variables used in the assessment.

The cross tabulation of the neural network efficiencies show that out of the 250 bank branches only 45 were classified as overall efficient branches. These 45 branches are overall benchmarks for the network of outlets included in the study. A number of combinations are also listed with bank branches being rated efficient in one or more outputs and inefficient in the remaining ones. The neural network results, therefore, give output-specific directions of improved efficiency. The latter may prove to be a particularly useful policy for inefficient branches that would be able to focus on the performance of individual outputs that underperform. The DEA targets, on the other hand, are also output specific with the difference that we either get improvement targets for all outputs or that the DMU is overall efficient.

The assessment of the efficiency of individual branches attracts interest from the central organisation's point of view when the issue of overall improvement targets is considered. Results regarding the targets assessed for individual outputs at the global organisational level are reported in Table 9.

TABLE 8. *Neural network efficiency classifications across outputs*

Combinations of efficiency ratings	Branches found ANN efficient in the corresponding output		
	Liabilities (y_1)	Loans (y_2)	Insurances and investments (y_3)
Liabilities	29[2]	—	—
Loans	—	30[2]	—
Insurances and investments	—	—	26[2]
$y_1 \times y_2$	14	14	45[1]
$y_1 \times y_3$	21	45[1]	21
$y_2 \times y_3$	45[1]	28	28
Total efficient out of the sample of 250	109	117	120

[1] Branches found efficient in all outputs.
[2] Efficient branches on a single output.

TABLE 9. *Aggregate target improvements of the three outputs*

Output variable	Constant returns to scale (CRS)	Variable returns to scale (VRS)	Neural networks[1]
Liability sales	126.7%	119.7%	113%
Loans sales	125.2%	119.2%	108%
Insurance policies and investment sales	130.0%	127.0%	117%

[1] If the expected value is lower than the observed the branch is given an efficiency value of 100. If the expected value is higher than the observed the branch is given an efficiency value equal to the ratio of the two.

The assessment of performance targets for the three outputs of banks branches yielded targets that vary in magnitude depending on the method selected for assessing these targets. The most aggressive targets are assessed by the CRS model with a 130% improvement being the maximum expansion factor which was estimated for insurance policies and investment sales. The other two efficiency tools, namely VRS and neural network have yielded less demanding targets which, nevertheless, give estimates of expected aggregate expansion of the outputs of the bank branches.

Targets assessed by the DEA models reflect proportional output improvements according to the efficiency of DMUs adjusted by the output slacks accounted for individual outputs. These output slacks can take substantial values as is the case for the insurance and investment sales of the bank branches in our study. Thus the focus on the target improvements of individual outputs of DMUs reveal more information than the simple reporting of percentage efficiencies. The higher rate of improvement for the insurance and investment policies was also confirmed by the ANN results without, however, matching the same level of expected achievements.

A final step to compare the results obtained from the three alternative methods is to consider the relative rankings of the bank branches obtained from the three models. These rankings, however, are customised for each individual output as the neural network results yield separate rankings. The analysis of the rankings obtained by the efficiency models is summarised, using the rank correlation coefficients, in Table 10.

Rankings obtained from the VRS and CRS models do not differ substantially across the three outputs due to the use of radial efficiency models. That is, all outputs are expanded *pro rata*

TABLE 10. *Spearman Rank correlations for the efficiency of each output*

	Liabilities		Loans		Insurances and investments	
	CRS	NN	CRS	NN	CRS	NN
CRS	—	0.41	—	0.68	—	0.64
VRS	0.86	0.36	0.87	0.53	0.81	0.55

All correlations are statistically significant at the 1% level.

according to the efficiency value of the corresponding branch and any extra improvement for individual outputs is marginal (slack values). The neural network rankings, however, are different and indeed there is variation across the three outputs. The CRS rankings seem to be closer to the neural network rankings whilst there is a fairly high correlation between the VRS and CRS rankings. The differences between the DEA and ANN rankings can partially be explained by the fact that the DEA efficiency results are truncated at the value of 100 whilst those obtained from neural network assessment can take values over 100.

The results show that neural networks can be used in cases with multiple input-output production technologies taking into account, however, the possibility of having output or input specific efficient or inefficient DMUs. On the relative rankings side, the neural network results seem to be more similar with the constant returns to scale and less with the variables returns to scale results. The latter, is a consequence of the implicit assumption of constant returns to scale adopted by the ANN models.

DATA ENVELOPMENT ANALYSIS VS. ARTIFICIAL NEURAL NETWORKS

A comparative framework is proposed drawing upon the methodological differences of the two methods. Most of these differences were verified in the results obtained from the simulated and empirical studies of this paper. The findings that are summarised below can be used as the basis for further comparisons between DEA and neural networks using different experimental assumptions and/or different empirical data sets.

Advantages of Data Envelopment Analysis over Artificial Neural Networks

- The assessment of efficiency is based on the development of a non-parametric extreme and not an average production function.
- For each inefficient activity unit DEA identifies a number of benchmark efficient activity units that are used as comparators.
- Economic assumptions regarding the presence of economies of scale can be incorporated within the assessment of efficiency.
- Efficiency can be decomposed into allocative, technical, congestion and scale terms solving different but similar DEA models.
- The relative importance of individual input and/or output factors can be incorporated within the assessment of efficiency by introducing value judgements.
- DEA results are based on global optimum values as the problem has a linear structure.
- DEA models can also be used to investigate issues concerning efficiency variation and technological progress across activity units and over time (panel data).

Advantages of Artificial Neural Networks over Data Envelopment Analysis

- No causalty commitment is required regarding the positive or negative influence that individual inputs have on the produced outputs. That is, increasing an input can actually have a negative effect on one or more outputs.
- Performance targets are assessed for individual inputs and outputs of each DMU that do not have necessarily the same direction of improvement (i.e. expand or contract).
- The neural network allows the use of both continuous value and classification input variables without the modelling enhancements that are necessary in the corresponding DEA models.
- Neural networks employ validation procedures to test the adequacy of the proposed models for unseen DMUs.

Future research areas in Artificial Neural Networks and Data Envelopment Analysis

- Both DEA and neural networks are deterministic in the sense that they do not account for the presence of random noise within the data set.

- There are no systematic mechanisms to test the adequacy and causality of the input/output specification used for assessing efficiency.
- The neural network mathematical formulation would need to be modified in order to take into account decision maker's preferences regarding the importance of certain inputs/outputs.
- Hypotheses regarding the correctness of assumptions of diseconomies of scale are not available in either DEA or neural networks.
- To facilitate the adoption of either method as an on-going tool for assessing performance more development of dedicated ANN and DEA software[33] needs to be pursued. This development should focus on the speed and performance of software code for obtaining results and also on their ability to present their results comprehensively for non-technical users.

CONCLUSIONS

This paper has sought to investigate the merits of two non-parametric methodologies, namely DEA and neural networks, as tools for assessing efficiency of decision making units. This comparative research used simulated data with theoretically specified efficiencies and also empirical data from bank branches. The intention of this paper is to avoid speculative remarks regarding the superiority or inferiority of one over the other method. This was not even confirmed by either the simulated or the empirical study and thus we should only put forward the idea of complementarity use and not antagonism between alternative methods of assessing performance.

Data envelopment analysis has been extensively tested and applied in the management science and economic fields as a tool for assessing efficiency over the last twenty years. Artificial neural networks are more recent and therefore there is not enough experience regarding their acceptability as a performance measurement tool. Our results have shown a tendency for DEA to perform more satisfactorily as a tool for estimating empirical production functions. Neural networks on the other hand, did rather well as a tool for obtaining relative rankings of DMUs on the basis of their predicted outcomes.

Undoubtedly, neural networks offer many advantages compared with the classical regression methods for assessing performance as they can accommodate multiple input-output production technologies. This advantage should be capitalised further by enhancing the current modelling procedures of neural networks towards integrating more behavioural aspects of the data sets to which they are applied. Examples concern the incorporation of *a priori* value judgments regarding the relative importance of inputs and outputs and the introduction of two error components within the neural network such that one will account for random noise and the other for technical inefficiency of individual DMUs. Finally, future research should explore the possibility of integrated use of neural networks and DEA in order to overcome weaknesses of the two methods and also capitalise on their unique advantages.

Acknowledgement—We would like to acknowledge computational support received from Mr Louis Khoo and the comments made by John Mingers during the preparation of this paper.

REFERENCES

1 A. Charnes, W. Cooper and E. Rhodes (1978) Measuring the efficiency of decision making units. *Eur. J. Opl Res.* **17**, 35–44.
2 R. Fare, S. Grosskopf and K. Lovell (1985) *The Measurement of Efficiency of Production*. Kluwer Nijhoff, Boston.
3. E. Thanassoulis and P. Dunstan (1994) Guiding schools to improved performance using data envelopment analysis: an illustration with data from a local education authority. *J. Opl Res. Soc.* **45**, 1247–1263.
4. S. Grosskopf and V. Valdamanis (1987) Measuring hospital performance: a non-parametric approach. *J. Health Economics* **6**, 87–92.
5. H. Tulkens (1993) On FDH efficiency analysis: some methodological issues and applications to retail banking, courts, and urban transit. *J. Productivity Analysis* **4**, 183–211.
6. E. Thanassoulis, R. G. Dyson and M. Foster (1987) Relative efficiency assessments using data envelopment analysis: an application to data on rates departments. *J. Opl Res. Soc.* **38**, 397–411.
7. G. Ferrier and K. Lovell (1990) Measuring cost efficiency in banking: econometric and linear programming evidence. *J. Econometrics* **46**, 229–245.

8. A. ATHANASSOPOULOS and E. THANASSOULIS (1995) Separating market efficiency from profitability and its implications for planning. *J. Opl Res. Soc.* **46**, 30–45.
9. A. ATHANASSOPOULOS (1995) Performance improvement decision aid systems in retail organisations using data envelopment analysis. *J. Productivity Analysis* **6**, 153–170.
10. A. ATHANASSOPOULOS and E. THANASSOULIS (1995) Assessing the marginal impacts of investments on the performance of organisational units. *Int. J. Production Econ.* **39**, 149–164.
11. P. SMITH (1990) Data envelopment analysis applied to financial statements. *Omega* **18**, 131–138.
12. A. BOUSSOFIANE, R. DYSON and E. THANASSOULIS (1991) Applied data envelopment analysis. *Eur. J. Opl Res.* **51**, 1–15.
13. H. FRIED, K. LOVELL and S. SCHMIDT (1993) *The Measurement of Productive Efficiency: Techniques and Applications.* Oxford University Press, New York.
14. E. RUMELHART, G. E. HINTON and R. J. WILLIAMS (1986) Learning internal representations by error propagation. In *Parallel Distributed Processing* I, (D. E. RUMELHART, J. L. MCCLELLAND and the PDP Research Group, Eds.). pp 318–362. MIT Press, Cambridge, MA.
15. M. SMITH (1993) *Neural Networks for Statistical Modelling.* Van Nostrand Reinhold, New York.
16. S. P. CURRAM and J. MINGERS (1994) Neural networks, decision tree induction and discriminant analysis: an empirical comparison. *J. Opl Res. Soc.* **45**, 440–450.
17. RIGHT INFORMATION SYSTEMS LTD (1994) *4Thought* V2.1.1. Right Information Systems Ltd, London.
18. R. BANKER, A. CHARNES and W. COOPER (1984) Models for the estimation of technical and scale efficiencies in data envelopment analysis. *Mgmt Sci.* **30**, 1078–1092.
19. A. ATHANASSOLPOULOS (1995) The evolution of non-parametric frontier analysis method for assessing performance: a review and recent developments. *Journal of SPOUDAI* **45** (1–2), 13–46.
20. A. HART (1992) Using neural networks for classification tasks—some experiments on datasets and practical advice. *J. Opl Res. Soc.* **43**, 215–226.
21. T. P. VOGL, J. K. MANGIS, A. K. RIGLER, W. T. ZINK and D. L. ALKON (1988) Accelerating the convergence of the back-propagation method. *Biol. Cyb.* **59**, 257–263.
22. R. G. HOPTROFF (1993) The principles and practice of time series forecasting and business modelling using neural networks. *Neural Comput. and Applic.* **1**, 59–66.
23. W. BOWLIN, A. CHARNES, W. COOPER and H. SHERMAN (1985) Data envelopment analysis and regression approaches to efficiency estimation and evaluation. *Ann. Opns Res.* **2**, 113–138.
24. E. THANASSOULIS (1993) A comparison of regression analysis and data envelopment analysis as alternative methods for assessing performance. *J. Opl Res. Soc.* **44**, 1129–1145.
25. R. BANKER, A. CHARNES, W. COOPER and A. MAINDIRATTA (1987) A comparison of data envelopment analysis and translog estimates of production frontiers using simulated observations from a known technology. In *Applications in Modern Production Theory Inefficiency and Productivity.* (A. DOGRAMACI and R. FARE, Eds.). pp 33–55. Kluwer Academic Publishers, Boston.
26. B-H. GONG and R. SICKLES (1992) Finite sample evidence on the performance of stochastic frontiers and data envelopment analysis using panel data. *J. Econometrics* **51**, 259–284.
27. R. BANKER, V. GADH and W. GORR (1993) A Monte Carlo comparison of two production frontier estimation methods: corrected ordinary least squares and data envelopment analysis. *Eur. J. Opl Res.* **67**, 332–343.
28. P. SMITH (1993) Misspecification in data envelopment analysis. Discussion paper 93/15, Department of Economics, University of York.
29. K. LOVELL (1993) Production frontiers and productive efficiency. In *The Measurement of Productive Efficiency: Techniques and Applications.* (H. FRIED, K. LOVELL and S. SCHMIDT, Eds). pp 3–67. Oxford University Press, New York.
30. D. GIOKAS (1990) Bank branch operating efficiency: a comparative application of DEA and the loglinear model. *Omega* **19**, 549–557.
31. L. DRAKE and B. HOWCROFT (1994) Relative efficiency in the branch network of a UK bank: an empirical study. *Omega* **22**, 83–90.
32. A. BERGER (1993) Distribution free estimates of efficiency in the U.S. banking industry and tests of the standard distributional assumptions. *J. Prod. Analysis* **4**, 261–293.
33. Warwick-DEA user's guide (1992) Warwick Business School, Coventry.

Received March 1995; accepted January 1996 after two revisions