ISSN 1750-4171

DEPARTMENT OF ECONOMICS

DISCUSSION PAPER SERIES

Neural Network Based Models for Efficiency Frontier Analysis: An Application to East Asian Economies' Growth Decomposition

Hailin Liao, Bin Wang and Tom Weyman-Jones

WP 2007 - 24

Dept Economics
Loughborough University
Loughborough
LE11 3TU United Kingdom
Tel: + 44 (0) 1509 222701
Fax: + 44 (0) 1509 223910
http://www.lboro.ac.uk/departments/ec

Neural Network Based Models for Efficiency Frontier Analysis: An Application to East Asian Economies' Growth Decomposition

Hailin Liao[^], Bin Wang, Tom Weyman-Jones

Department of Economics
Loughborough University
Leicestershire
LE11 3TU

Abstract

There has been a long tradition in business and economics to use frontier analysis to assess a production unit's performance. The first attempt utilized the data envelopment analysis (DEA) which is based on a piecewise linear and mathematical programming approach, whilst the other employed the parametric approach to estimate the stochastic frontier functions. Both approaches have their advantages as well as limitations. This paper sets out to use an alternative approach, i.e. artificial neural networks (ANNs) for measuring efficiency and productivity growth for seven East Asian economies at manufacturing level, for the period 1963 to 1998, and the relevant comparisons are carried out between DEA and ANN, and stochastic frontier analysis (SFA) and ANN in order to test the ANNs' ability to assess the performance of production units. The results suggest that ANNs are a promising alternative to traditional approaches, to approximate production functions more accurately and measure efficiency and productivity under non-linear contexts, with minimum assumptions.

KEY WORDS: Total Factor Productivity; Neural Networks, Stochastic Frontier Analysis, DEA, East Asian Economies.

JEL Classification: D24, C45, O47, O53, L60

.

[^] Corresponding author. Address for correspondence: Economics Department, Loughborough University, Leicestershire, UK, LE11 3TU. Tel: +44 (0) 1509 222727, Fax: +44 (0) 1509 223910

Neural Network Based Models for Efficiency Frontier Analysis:
An Application to East Asian Economies' Growth Decomposition

1. Introduction

The economic development in the Asia-Pacific region over the last four decades is frequently considered to be an 'economic miracle' (World Bank, 1993) in terms that one East Asian country after another has taken off from a stagnant state to achieve an annual growth rate of 10% or even more, which have received great attention from economists. The central analytical and policy question raised by these extraordinary economic performances is what the causes of these fast growth rates are, and there is no general agreement on it. Indeed, there is a continuing controversy in which accumulationists are on one side and assimulationists on the other side. That is, the debate over how much of output growth is due to technological change versus growth in inputs such as the accumulation of physical and/or human capital. Assimilationists, on one side, are persuaded that answer to growth lies in the use of more efficient technology (World Bank, 1993; Sarel, 1996; 1997; Nelson and Pack, 1999); whilst the observations based on growth convergence regressions with prior economic and social conditions do not seem to have warranted such rapid growth to be sustained (Krugman, 1994; Young, 1992; 1994a; 1994b; 1995; Kim & Lau, 1994; 1996; etc).

In the above studies, there is an implicit assumption that the economies are producing along the production possibility frontier with full technical efficiency. With the exception of Kim and Lau (1994) and Chang and Luh (1999), these studies adopted the conventional growth accounting approach and estimated total factor productivity (TFP) growth without distinguishing between its two components: technical progress (TP) and technical efficiency change (TEC). Rather, TP is synonymously considered to be the unique source of TFP growth. Failure to take account of inefficiency and TEC may produce misleading and biased TFP estimates: high rates of TP can coexist with deteriorating technical efficiency, and relatively low rates of TP can also coexist with improving technical efficiency (Nishimizu and Page, 1982); and different policy implications result from different sources of variation in TFP. For this reason, some attempts have been done to apply frontier approach to estimate the TFP of sample countries. The first attempt utilized the data envelopment analysis (DEA) which is based on a piecewise linear and mathematical programming approach, whilst the other employed the parametric approach to estimate the stochastic frontier functions (SFA). Both approaches have their advantages as well as limitations. For instance, they differ in how restriction is imposed on the specification of the best practice frontier and the assumptions on random error and inefficiency, and so on (Coelli *et al.*, 1998). More recent empirical work, for example, includes Fare *et al.* (1994), who employed a nonparametric approach; Leung (1998), who also estimated Malmquist Index for Singapore's manufacturing sectors; and Wu (2000), who presents an economy-level study on China using an econometric model.

This paper sets out to use an alternative approach, i.e. artificial neural networks (ANNs) to estimate the frontier for efficiency measurement, and the relevant comparisons are carried out between DEA and ANN, and SFA and ANN in order to test the ANNs' ability to assess the performance of production units.

The past two decades have been seeing the rise of important applications of neural networks in finance, business, education, marketing, engineering, forecasting and related fields[1] due to its associated memory characteristics and generalisation capability with minimum assumptions (Stern, 1996). In some cases, this approach has also been explored by Athanassopoulos and Curram (1996), Costa and Markellos (1997), Wang (2003), Santin et al. (2004) and Delgade (2005) in handling efficiency measurement problem, but rarely in an international framework with whole countries/industry sectors as units of observation. The majority of the above references has reported that efficiency scores derived from neural networks, if not better than those from traditional approach, at least provide an additional/alternative direction to this problem.

The rest of the paper unfolds as follows. In the second section we give a brief review of NN mechanisms in order to facilitate the constructed frontier estimation. The third section concerns the design of the two comparison experiments in an application of NN in estimating annual 'technical' efficiency measures for several East Asian economies at manufacturing level from 1963 to 1998. And the analysis and discussions of results obtained, with comparisons to those obtained by DEA and SFA, are followed. The last section concludes the paper and offers the possible directions for future research.

2. Analytical Methodology

2.1 Neural Network Basics

In brief, Neural networks (NNs) are sorts of computer-based systems trying to mimic the functioning of the human brain by emulating a network of interconnected neurons. A neural network 'learns' relationships between input and output variables by repeatedly presenting the related input data and changing the internal structure of the network to derive the best possible fit (Athanassopoulos *et al.*, 1996). As Bishop (1995) argued, NNs can be treated as statistical tool, in terms of being 'trained' to solve certain problems or identify specific patterns, more complex than but not principally different from classical approaches. For instance, in their simplest form, NNs are similar to a linear regression,

$$\log(Y_{it}) = \beta'(\log(X_{it})) + \varepsilon_{it} + \mu_{it} \tag{1}$$

The explained variable Y is the 'response', while the explaining variables X are the covariates. In the NN literature, this function-diagram is a single unit 'perceptron': each node processes several inputs and results in several outputs, y, serving as inputs to other nodes of the network, until we reach the final stage where we get the modelled output.

However, an NN is more complex than this plain linear model and one of the most important aspects of an NN is the ability to learn from past patterns to predict new ones – the links between the units (neurons) are not rigid but can be modified through the learning processes generated by the network's interaction with the outside world. We can specify a feed-forward neural network with n inputs X, m hidden units and a single output Y as follows (Santin *et al*,, 2004),

$$Y = F[\beta_0 + \sum_{j=1}^{m} G(\gamma_j + \sum_{i=1}^{n} X_i \alpha_{ij})\beta_j] \tag{2}$$

where β_0: output bias

γ_j: hidden units biases (j=1, …, m)

α_{ij}: weight from input unit i to hidden unit j

β_j: weights from hidden unit j to output

The activation function G(u) in the hidden layer, typically a bounded and monotonic function, can take a specified form such as the logistic form g(u)=1/[1+exp(-u)], which produces a simple output such as 0 or 1 (Fleissig et al. 2001). Other functions, such as Tan-sigmoid transfer function with values within the 0 and 1 range, might serve as well (Bishop, 1995). The activation function for output layer F(x,w) might be a linear function of the inputs through the weights (w) plus a constant.

There are several differences between simple statistical techniques and NNs. The first one is that the relationship between inputs and outputs of each node is not linear, since it combines weights with non-linear functions. Traditional linear based techniques are required to separate the data past the linear dimension they were operating in, while NNs explore potentially 'hidden' correlations among the predictive variables in the hidden layers, which are then entered as additional explanatory variables in the nonlinear function. The down side is the lack of transparency, which makes it impossible to describe them in terms of equations with parameter values. However, if we are looking for a mechanism to represent a complex, multi-dimensional, non-intuitive process, the estimated parameters are of second interest. This is typical in the case of efficiency analysis.

Secondly, the NNs attempt to approximate any function between the original observable inputs and the final modelled outputs by minimizing the error, the gaps between the predicted values and the actual values, and the 'back-propagation rule' (BP) – a supervised learning algorithm, is used to compute the gradient of the error which is passed to the optimization procedure (Rumelhart *et al.*, 1986). In the process of model setting-up, learning is achieved by altering the values of weighted connections between neurons to bring the output of the network closer to the desired target value, and an error is calculated in each iteration which is then back-propagated to the network and used to adjust the weights. The overall aim here is to lower the Root Mean Square Error (RMSE)[2] for the training data by using the steepest gradient method on the error surface. In this sense, NN makes no assumptions about the statistical properties of the data and the functional form of the underlying efficiency model, however, a production frontier based on NN models is deterministic in nature. In other words, for a given set of weights and inputs, the NN produces a set of outputs depending on weights w. Modifying the weights results in a new set of outputs. The distance between the actual value and the NN predicted value, at a given stage, triggers or not a change in weights depending on the size of the mismatch. If the error produced by the change is high enough, the NN 'learns' by adjusting the weights. This process is called the steepest 'gradient descent', which implies processing each observation in the training set a number of times to reach the highest level of convergence.

Before the implementation of the NNs to study the relationship between the inputs and outputs and therefore the efficiency, we need to address two important questions: (1) what is the appropriate NN architecture for a particular data set; and (2) how robust the NN performance is in efficiency analysis in terms of sampling variability. For the first question, there are no definite rules to follow since the choice of architecture also depends on the classification objective. Many factors such as hidden layers, hidden nodes, data normalization and training methodology, etc can affect the performance of NNs, and the best network is typically chosen through experiments. We will expand each of these steps in the next section. For the second question, we employ cross-validation approach to investigate the robustness of the NNs in efficiency analysis.

2.2 Architecture of a Neural Network

A NN is typically composed of several layers of many computing elements called nodes, and NNs are characterized by the network architecture – the number of layers, the number of nodes in the hidden layer and how the nodes are connected. In a popular form of NN called the multi-layer perceptron (MLP), all nodes and layers are arranged in a feed forward manner[3]. There is no general rule to establish the optimal degree of network complexity, but one hidden layer is commonly used. A simple illustration of a neural network with one hidden layer can be shown in Figure 1.

The size of the hidden layer is not easy to determine a priori. Although there are several rules suggested for determining the number of hidden nodes, such as using n/2, n, (n+1) and (2n+1) when n is the number of input nodes, none of them works well for all situations. Determining the appropriate number of hidden nodes usually involves lengthy experimentation since this parameter is problem and/or data dependent[4]. In empirics, we can see the effect of hidden nodes on the performance of NN by using different levels of hidden nodes, and a brute force trial-and-error approach is used in searching for the best architecture. Networks with a single layer of hidden units are first considered, and the number of units varies from 1 to 10, or even more, then ensembles of n networks are trained. After training and validation, we can plot the results of median error as a function of the number of hidden units (for out-of-sample performance)[5]. As expected, the median error decreases as the number of hidden units in the hidden layer increases, and the optimal number of hidden nodes is chosen when the error starts increasing.

In this paper, we use BP network, the most extensively used multiple-layers networks with a back-propagation least mean square error learning algorithm. Figure 2 shows a BP network with two-layer tangent sigmoid/pure linear (tansig/purelin) transfer function. The input vector p1 has two input elements. These inputs post-multiply the 4-row, 2-column input weight matrix (IW), with 4 neurons in hidden layer and 2 elements in input vector. A constant 1 enters as an input and is multiplied by a vector bias b1 with 4 elements. The net input to the tansig transfer function[6] is n1, a 4x1 vector. The sum of the bias b1 and the product of IW and p1 is passed to the tansig transfer function to get the neuron's output a1, which in this case is a vector with 4 elements. Then this output from hidden layer works as the input to output layer multiplying the layer weight matrix (LW). By passing the transfer function of output layer in a linear form[7], we can obtain the output a3, that is, y. After the structure of NN is determined, this NN can be trained so as to pursue the optimal network model with the best weight and bias for every neuron by giving the input data and target data. The training process is repeatedly adjusting the initially random weight of every neuron until minimizing the error between the estimated output and the target output. The factors to affect the training effective are the learning rate parameter and the training method.

3. Empirical Experiments

3.1 Experiments Designed

A key issue in this study is the frontier function estimation, which traditionally can be estimated following non-parametric and parametric techniques[8] (Murillo-Zamorano, 2004):

- DEA, which is the most commonly used mathematical programming method in frontier/efficiency analysis[9] and considered only in this study, makes no assumption on functional form and forms a deterministic frontier by enveloping the available data. The efficiency measures are distance to an empirical production frontier and the values are calculated on the basis of standard Pareto efficiency.
- The parametric approaches assume functional form for the underlying production, such as Cobb-Douglas, CES, translog etc, and estimated the frontier either by COLS or maximum likelihood.

In the (non-linear) production process, apart from the increasing and decreasing returns to scale, it is possible to have a negative slope between inputs and outputs. In empirics, Costa and Markellos (1997) found the so-called 'congested area' in the London underground from 1970 to 1994 in terms of fleet size and workers (inputs) and millions of trains km per year covered by fleet (outputs). Some researchers in educational production function analysis found that traditional restrictive specifications fail to capture potential non-linear effects of school resources (Baker, 2001). In our preliminary studies in several East Asian economies at manufacturing level, we also found quite similar pattern between inputs and outputs. Of course, we can assume monotonicity and concavity for the underlying production technology before the estimation, as it does in DEA. Or some regularity properties can be imposed in order to satisfy the regulatory conditions before the translog functional forms are employed, as we usually do in SFA. What we try to address in this study is that we don't have to make/impose such assumptions when using NN technique to approximate the underlying production functions.

We therefore have two sets of experiments to carry out in this study. The first one is to make comparison with results from traditional-deterministic DEA (hereafter referred to as DEA-NN), focusing on the ability of the two methods to disentangle efficient and inefficient production-units and therefore to give useful managerial insights concerning the performance of individual production-units. The following steps are involved. After the NN production function is estimated, the frontier can be formed as when using the COLS – we shift the estimated production curve upwards, in order to correct the downward bias in the estimated curve, by the magnitude of the largest positive error. Therefore, all corrected residuals are non-positive and at least one is zero. This approach is very similar to that of DEA in the sense that at least one unit is 100% efficient and the frontier will be deterministic. That is, a best practice unit can be established amongst a set of observed units and the units that are inefficient can be identified when compared to the benchmark. The corresponding efficiency measures can then be calculated on the basis of the distances observed, that is, between the obtained production frontier and the realised outputs. The traditional-deterministic DEA efficiency analysis was conducted on the raw data by assuming first constant and then variable returns to scale.

The second experiment is to make comparison with results from conventional SFA (hereafter referred to as SFA-NN). Stochastic frontiers are straightforward to estimate by OLS or maximum likelihood estimation so as to decompose the error term into a stochastic and an inefficiency term. But with non-linear/non-parametric formulations, estimation is not possible using available estimation packages. We could, however, estimate the frontier directly when the inefficiency effect is removed from the dependent variable given the reasonable noise level. The starting values for these inefficiency effects to be removed are those estimated from traditional parametric approach assuming that they are accurate to some degree. These estimated inefficiency effects can then be used to adjust the real/observed outputs. After iterations, the adjusted outputs should be the desired target value for NN training, which results in a NN production frontier rather than production curve obtained from DEA-NN.

3.2 Data

NN-based models are applied to set out comparative annual technical efficiency measures and productivity growth for seven East Asian economies at manufacturing level. The panel data of 28 manufacturing sectors' annual time-series in seven East Asian economies, namely, Hong Kong, Indonesia, Korea, Malaysia, Philippines, Singapore and Taiwan, during 1963-1998 are used[10]. The sectors and their SIC classification numbers are listed in Table 1.

Measurements of sectoral productivity growth rates require data on output, capital and labor input in this study. The raw data series are value-added[11], fixed capital formation and employment measured in numbers. Both series of value added and fixed capital formation are measured in local currency unit at current prices, so GDP deflators[12] from the IFS database and WDI are applied to convert these series into constant price based on year 1990. The standard perpetual inventory method (PIM) is used here to construct the capital stock under a uniform 4% depreciation rate[13] with 1963 as the benchmark, i.e.

$$K_{i,t+1} = K_{i,t} + I_{i,t+1} - \delta * K_{i,t} \tag{3}$$

where $K_{i,t}$ is capital stock of sector i at period t, $I_{i,t}$ is capital formation/investment and δ is depreciation rate. Following Pham, Park & Ha (2002) and Young (1995), the initial capital stock series is initialized by assuming that the growth rate of investment in the first five years of the national accounts investment series is representative of the growth of investment prior to the beginning of the series. That is,

$$K_{i,0} = \sum_{t=0}^{\infty} I_{i,-t-1}(1-\delta)^t = \sum_{t=0}^{\infty} I_{i,0}(1+g_i)^{-t-1}(1-\delta)^t = I_{i,0}/(g_i+\delta) \tag{4}$$

where $I_{i,0}$ is the first year investment data, g_i is the average growth in the first five years of investment series and δ is the depreciation rate. Here, we implicitly assume that no net capital stock exists before 1963 for all countries in question. Past studies have shown that given positive rates of depreciation and a sufficiently long investment series, the PIM is insensitive to the level of capital used to initialize the series. The number of workers employed in each industry was used for labour input, which is not adjusted for changing quality or skill composition due to lack of consistent data[14].

4. Results and discussions

4.1 Neural Network Training Procedure

Pre-processing and Normalization of data. As a data-driven technique, high-quality training data[15] to the NN implementation cannot be overstressed, since the quality and structure of the input data largely determine the success or failure of an NN implementation (Bansal *et al.*, 1993[16]). Data in this study (both inputs and outputs) have a tendency to generate extreme outliers. This is especially a challenge when the data are heavily influenced by measurement error in the case of agricultural and manufacturing applications in developing countries. When calibrating a neural network, outliers can overshadow all other numbers in an input set. To limit the effect of outliers, data are transformed by squeezing them between -1 and 1. After squeezing, the histogram is much more bell-shaped or normally distributed[17]. However, the fact that DEA property of unit invariance is similar to property of scale pre-processing required by NNs validates the rationale to implement a comparison between DEA results and NN results.

We follow Wang's (1996) approach[18] to make the data transformation in this study as follows:

For independent variables: $x_{jNor} = \dfrac{x_j - x_{j\min}}{x_{j\max} - x_{j\min}}$; and

For dependent variable: $y_{tNor} = \dfrac{(1-0.1)(y_t - y_{t\min})}{y_{t\max} - y_{t\min}} + 0.1$. Therefore, output value will be in the range of [0.1, 1][19].

Grouping Technique. This involves two aspects. The first one is the use of cross-validation to avoid over-training[20]. The idea is: the available data set is subdivided into a training set for giving in-sample results which however are not a good indication of the generalization ability of the networks, and a second set for testing which is only used after training the network. The test set is used to compute out-of-sample results that indicate how good the predictions will be on any (comparable) data set. To prevent the model from 'over-training', the cross-validation sample serves for checking that the error in this second sample does not increase while the NN trains itself through multiple simulations on the training sample. In this sense, we can set the stopping criterion to lower the MSE of training and testing simultaneously[21].

Another aspect is the network ensembles. The current standard is not to train a single NN, but instead an 'ensemble' of them, each of them using a slightly different training set generated by naïve bootstrapping (Efron and Tibshirani, 1993). After training and validating each of the networks, an outcome to a given input set is found by choosing the single best network[22].

Our dataset for every economy is a panel including 28 manufacture industries from 1963 to 1998 with one output and two inputs. Troutt et al. (1995) suggested that training data for nonparametric models should be at least 10 times the number of input variables, we thus randomly select 23 observations from the sample each year, leaving the rest of 5 observations for validation. Then we train the ensemble of 50 networks each year[23] in order to find the single best network in terms of minimum MSE. [24]

4.2 Results Analysis

Estimated NN parameters. Table 2 presents the basic parameters of the best estimated neural networks for the case of Korea[25]. As it can been seen, the estimated neural network incorporates 10 log-sigmoid hidden units at the most, and the Levenberg-Marquardt algorithm[26] is employed for the training. 2-10-1 stands for 2 neurons in the input layer, 10 neurons in the hidden layer and only 1 neuron in the output layer. The convergence criteria used for training are a RMSE less than or equal to 0.0001 or a maximum of 1000 iterations[27]. Based on the best estimated neural networks, the industry-sector efficiencies can be calculated.

Comparison of estimated efficiency. As mentioned in the Introduction section, the frontier approach to decompose the productivity growth, in a macroeconomic context, into components due to efficiency change and technical change provides a platform for a straightforward and simple economic interpretation for output growth in an economy: a country can be less inefficient therefore operates within the frontier, and “catch up” to the frontier over time represented as efficiency gains, or the frontier itself can shift over time indicating technical progress. On the other hand, different policy can be drawn from different sources of variation in productivity growth as high rates of TP can coexist with deteriorating technical efficiency, relatively low rates of TP can also coexist with improving technical efficiency (Nishimizu and Page, 1982). Bear this in mind, we will focus on the estimation of a production frontier and the derivation of the components of output and productivity growth in this paper with an aim to addressing the following questions: which countries/manufacturing sectors are making most efficient use of their inputs? Is economic growth driven by removing inefficiencies and moving closer to the production frontier? Or is it driven by movements of or along the frontier itself?

Firstly, a comparative study is carried out for illustrating the potentials of the NNs in the estimation of underlying production frontiers, and the main results are summarized in table 3 along with the standard DEA (both constant-returns-to-scale and variable-returns-to-scale) and conventional SFA technical efficiency scores. A general pattern appears to be that the estimated efficiency measures reveal substantial production inefficiencies among the manufacturing industries in the sample economies under consideration, which rejects the hypothesis that the production is operated on the frontier with full efficiency assumption, as did in the traditional growth accounting exercises.

A number of comments can be made. (1) There are several similarities between DEA and NN in terms of efficiency measurement: both are in non-parametric fashion for the underlying production technology, therefore no predetermined assumption on functional form is made; both allow the use of multiple inputs and multiple outputs; and both techniques are non-linear by nature (Delagdo, 2005). In terms of the best-practice frontier, it is constructed in DEA by enveloping the available data in the sample. For DEA-NN, we calculate the error between the estimated and target output for each country in ever year when data are available; then we shift the estimated output curve up with the amount of maximum positive error and get the estimated frontier. Delgado (2005) proposed a new approach in order to reduce the effect of the largest positive error, by shifting the estimated curve by a mean of the largest positive errors, for instance, 5%. He found that the neural network model performed differently to the standardised one in terms of rank, standard deviations, efficient units, etc, and therefore provided alternative results from this approach. However, no justification for the choice of percentage involved was provided. Whether it would be arbitrary is not known, and we only present the results with a shift by the maximum error. Despite of the similarities of non-linear and non-parametric features, we can see that NN can have the possibility to have more efficient units on the frontier (for instance, in the case of SFA-NN) and have slightly higher efficiency scores in both DEA-NN and SFA-NN comparing to those from DEA and classical SFA. As we mentioned before, NNs have the flexibility to solve complex non-linear functions in a semi-parametric fashion where the main information or 'knowledge' lies implicitly in the data. and can fit the data as perfectly as possible. These results are generally in consistency with those in Wang's (2003) study[28].

(2) After obtaining the frontier, the (in)efficiency is then determined as observation-frontier distance in both DEA-NN and SFA-NN. Although there exist common trends in the efficiencies estimates through different approaches, several differences in the quantitative measures are clearly evidenced. First, we expect that the correction by the largest positive error is sensitive to outliers, as in DEA, but to a less extent in that the neural networks assist model building based on the entire data set rather than some extreme data points from which uncertainty information has been lost. In this sense, it is surprising to find out that DEA-NN might have a little bit higher efficiency score. In some cases, the efficiency scores are larger than 1 in DEA-NN and SFA-NN model, which is not allowable in DEA context, due to the fact that the statistical and probabilistic properties are embedded in NNs.

One possible explanation might be that the monotonicity property is not considered when building up the NN structure whilst DEA model implicitly assumes monotonicyty of inputs and outputs, and therefore does not impose any specific form of the production function. In SFA model, we posterior test these conditions, and 75% data points satisfy these conditions. We mentioned before that the functional form of NN depends on the training data for learning the nonlinear monotonic forecasting by using a least square error minimizing approach. The training data may have to satisfy the monotonicity property for using NN. If not, the resulting model may produce forecasts that are not consistent with the assumption of monotonicity[29], and therefore different with those from DEA and SFA. For instance, both Costa and Markellos (1997) and Santin *et al*. (2004) mentioned the 'congested area' where an increase in one input implies a decrease in one output. Both traditional DEA[30] and SFA cannot consider any chance for the existence of this kind of production technology whilst NN can as NN can be universal approximates of any function and its derivatives to any desired degree of accuracy.

When training data violates the property of monotonicity, some data screening method (e.g. DEA-based approach for selecting the input data, Pendharkar and Rodger, 2003[31]) may be required with an aim to create a training sub sample that does not violate the monotonicity property. However, in the case of a small-size data set, such as the case in our study (only 28 sectors available for every year), such a data screening may not be feasible due to very small size of resulting sub sample. A follow-up research will deal with this issue.

Growth decomposition

In order to compare and check the robustness of neural network frontier approach in TFP estimates, table 4 presents estimates of average TP rates, changes in TE and TFP growth rates along with actual output growth rates for seven East Asian economies over the period 1963 to 1998 for different methods employed. The sectoral TFP growth is not calculated as a residual but is obtained by summing changes in TE, TP and scale efficiency changes (SEC) (in the case of conventional SFA). The general pattern reveals that the trend of TFP growth is accounting for a larger and larger proportion of output growth. Several implications can be drawn. First of all, results are expected to be different among various methods in terms of the magnitude of productivity measurement and its various components. On one hand, TFP growth estimates under deterministic DEA and DEA-NN approach are likely to be higher than those from SFA and SFA-NN methods which are stochastic in nature. The possible explanation might be that any measurement error and/or deviation of the data point are considered as technical inefficiency in the DEA and DEA-NN approach.

On the other hand, in spite of these differences, all estimates show substantial production inefficiency among the manufacturing industries, and positive TECs indicate that the steady trend for improved TE is observed throughout the sample period and TECs might be the key point for the TFP growth in these economies. Due to the backwardness of the economies, the further away is the economy from the frontier, the bigger the technical progress that can be achieved, and vice verse. In case that the magnitudes of TEC are smaller compared to those of TP, the role of knowledge-based factors was a major engine of growth coming via direct imports of capital goods and technology for their current prevalence of assembly production in the manufacturing sectors, which is the case for the second-tier NIEs, such as Indonesia, Malaysia and the Philippines. However, the potential in efficiency improvement is declining though still positive which can be considered to be evidence of catching-up to the frontier, and has been almost exhausted in the 1990s for all economies in question. In this sense, economic growth in the future will mainly rely on innovation, that is, technological progress.

In terms of productivity growth rates for individual economy, with the exception of the Philippines in the conventional SFA estimation, one can see that the impressive TFP growth rates exist where Korea is ranked first, followed by Singapore[32]. The result is generally consistent with that of Sarel (1996, 1997), Hobday (1995), Collins and Bosworth (1997), Drysdale and Huang (1997) and Chang and Luh (1999), that productivity growth is also a main source of output growth, and far more optimistic than the findings of Young (1992, 1995) and Kim and Lau (1994, 1995). For example, Young finds that Singapore's average TFP growth is minus 1.1 per cent (TP and TFP growth are the same since no inefficiency is assumed, Young 1995, p658); and consequently that all its output growth is exclusively due to capital accumulation. Although we find that the trend of TFP growth slightly decreased since the 1970s, the same conclusion as Mahadevan and Kalirajan (2000), our results show different pattern in that the overwhelming improvement in TP leads to the TFP growth. While Leung (1997) suggests that it would be plausible for Singapore's manufacturing industries to have an average annual TFP growth rate between 2-3 per cent during the 1980s and 1990s, his striking finding of no-link between a learning-by-doing effect and TFP growth is quite close to this study's result that negligible technical efficiency improvement is found. For comparison purpose, we calculate the average productivity growth rate of Singapore for the period 1983 to 1993, the one used in Leung (1998). The resulting TFP[33] has an average 4.72% per year, which is quite consistent with Leung's 4.6%. This also confirms Coelli *et al*'s (2005) finding that productivity series from DEA method is more volatile and much sensitive to year-to-year change, and the change in data series' beginnings and ends can influence the measures obtained.

In contrast to many previous studies, the general conclusion from table 4 is that TFP growth is important. With the exception of the Philippines, the percentage contribution of TFP to output growth in six economies has not been negligible, though efficiency gains and TP do not play a similar role in TFP growth. Among four Newly Industrialized Economies, that is, Hong Kong, Korea, Singapore and Taiwan, the TFP growth rates are quite close to each other. It is worth noting that the majority of Hong Kong manufacturing sectors report a negative growth in the 1990s, and this decline in the share of GDP and the considerable shrinkage in the overall manufacturing sector over time are in line with the fact that Hong Kong gradually and significantly switches its manufacturing operations to the mainland China following China's 'open door' policy since the 1980s. As a result, the output growth over the sample period is nearly zero with still positive TFP growth. Taiwan performs rather well in technology adoption, we conjecture this might be because Taiwan adopts more efficient technologies transferred from the industrial countries through foreign investment into its relatively smaller size local firms, especially in the labour- and capital-intensive sectors. While technical progress is identified as one of the major sources of its TFP growth for Singapore, Korea gains both from technological progress and efficiency. This supports the fact that Korean manufacturing can upgrade its production technology possibly through imported technology and R&D and/or technological diffusion, and at the same time master new technologies quickly which might be a result of sufficient investment in education and on-the-job-training, learning-by-doing effect, and so on.

We should also examine the relative performance at manufacturing sector level given the fact that industry-level data is available and used. Towards this end, table 5 presents the results for the three sub-categories of traditional, basic and high-tech industries[34]. The division is based on what we think is a reasonable interpretation of their input usage. From table 5, we can see that the three sub-sectors developed quite differently with strong growth of production in the high-tech sector and moderate growth in the basic sector, but relatively weak growth in the traditional industries. There are, however, notable differences among economies in their productivity growth and growth decomposition in the three sub-categories. First-tier NIEs, that is Korea, Singapore, Hong Kong and Taiwan, have their fastest growing TFPs in the high-tech sector, indicating that the high-tech industry is more exposed to the international market and multinational investment in each economy. Although some did suffer from the deterioration of TPs in some years, this can be offset by the enhanced performance in TEC. For instance, Korea, Singapore and Taiwan all reported TECs in the range of 2-4% in the machinery and electronics sector. This scenario is opposite to the case in the basic sector, where TP dominates the almost negative TECs.

With the exception of Malaysia and the Philippines, all economies show a very strong performance in TP in the chemicals, non-metallic mineral products and metal products sector. Among them, Indonesia has the most impressive record having recorded TP growth rate at 5-6% per annum. In the traditional sector, one might expect that the possibility of an "indigenously generated improvement" in technology (Kim and Lau, 1994) would be limited, and be less effective due to their strong labour-intensity. The result, however, support the expectation that in all second-tier NIEs, that is, Malaysia, Indonesia and the Philippines, TEC is the largest contributor to growth in the traditional sub-sector mainly due to the advantage of backwardness. One possible explanation is that since the production of such low-end products can be gradually relocated to somewhere else with lower cost advantage, according to product-cycle theory, then one way to sustain the higher labour cost is to upgrade either TEC or production technology to survive competitively in international markets.

5. Conclusions and future work

This paper presents a neural network study to efficiency frontier and efficiency analysis. The main aim is to make minimum assumption with maximum flexibility of the underlying production process. The results are comparable to the normal DEA and SFA results on the whole, suggesting that NNs can be used as an alternative tool to both econometric and DEA-based techniques for measuring technical efficiency when a strong theoretical model about the production technology is not firmed and minimum assumptions are imposed.

The neural network structure used in this paper is a single layer of weighted, hidden units. Hidden units or nonlinear functions contribute to a neural network's ability to replicate nonlinear relationships between inputs and outputs. Future research with neural networks in the efficiency analysis is suggested, and the possible directions include the imposition of monotonicity and concavity assumption, and so on.

In addition, we used a traditional back-propagation NN algorithm in our experiments without considering other algorithms. Several approaches have been tried in the literature, for instance, second-order gradient search and other gradient-free methods (Curry and Morgan, 1997) that improve the performance of connectionist models over the traditional back-propagation models using the steepest-decent search method. We believe that the performance of current NNs can be improved further by considering other learning algorithms.

References

Altman, E, G. Macro and F. Varetto (1994) Corporate distress diagnosis: comparisons using linear discriminant analysis and neural networks. *Journal of Banking and Finance*, 18, pp.505-529.

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, pp.1000-1016.

Baker, B.D. and C.E. Richards (2000) A comparison of conventional linear regression and neural network methods for forecasting educational spending. *Economics of Education Review*, 19(2), pp.405-415.

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

Bansal, A, R.J. Kauffman and R.R. Weitz (1993) Comparing the modelling performance of regression and neural networks as data quality varies: a business value approach. *Journal of Management Information Systems, 10(1)*, pp.11-32.

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

Caves, D.W., L.R. Christensen and W.E. Diewert (1982) The economic theory of index numbers and the measurement of input, output and productivity, *Econometrica,* 50, pp. 1393-1414.

Chang, C. and Y. Luh (1999) Efficiency change and growth in productivity: the Asian growth experience. *Journal of Asian Economics,* 10, pp551-570.

Collins, S. and B. Bosworth (1996) Economic growth in East Asia: accumulation versus assimilation. *Brookings Papers in Economic Activity,* 2, pp135-203.

Costa, A. and R.N. Markellos (1997) Evaluating public transport efficiency with neural network models. *Transportation Research*, C5(5), pp.301-312.

Coelli, T.J., D.S. Rao, C.J. O'Donnell and G.E. Battese (2005) *An Introduction to Efficiency and Productivity Analysis*. Second edition. Springer..

Curry, B. and P. Morgan (1997) Neural networks: a need for caution. *Omega: An International Journal of Management Science*, 25 (1), pp.123-133.

Delgade, F. (2005) Measuring efficiency with neural networks. An application to the public sector. *Economics Bulletin*, 3(15), pp1-10.

Drysdale, P. and Y. Huang (1997) Technological catch-up and economic growth in East Asia and the Pacific. *The Economic Record,* 73, pp201-211.

Efron B. and R. Tibshirani (1993) *An Introduction to the Bootstrap*. Chapman & Hall, London.

Fare, R, S. Grosskopf and M. Norris (1997) Productivity growth, technical progress, and efficiency change in industrialized countries: reply, *American Economic Review*, 87, pp.1040-1043.

Fare, R., S. Grosskopf, M. Norris and Z. Zhang (1994) Productivity growth, technical progress and efficiency change in industrialized countries. *American Economic Review,* 84(1), pp66-83.

Fleissig, A., T. Kastens and D. Terrell (2000) Evaluating the semi-nonparametric fourier, aim, and neural networks cost functions. *Economics Letters*, 68(3), pp.235-244.

Hertel, T, A. Nin-Pratt, A. Rae and S. Ehui (1999) Productivity growth and catching-up: implications for China's trade in Livestock products, memo, presented at *the Second Annual conference in global Economic Analysis*, Denmark, 1999.

Hobday, M. (1995) *Innovation in East Asia. The Challenge to Japan.* Cheltenham: Edward Elgar.

Kaefer, F., C.M. Heilman and S.D. Ramenofsky (2005) A neural network application to consumer classification to improve the timing of direct marketing activities. *Computer & Operations Research*, 32(10), pp.2595-2615.

Kalirajan, K.P, M.B. Obwona and S. Zhao (1996) A decomposition of total factor productivity growth: the case of Chinese agricultural growth before and after reforms. *American Journal of Agricultural Economics*, 78, pp331-338.

Kim, J. and L. Lau (1994) The source of economic growth of the East Asian newly industrialized countries. *Journal of the Japanese and International Economies,* 8, pp235-271.

Kim, J. and L. Lau (1996) The source of Asian Pacific economic growth. *Canadian Journal of Economics,* 29, supplement, pp448-454.

Krugman, P. (1994) The myth of Asia's miracle. *Foreign Affairs*, 73(6), pp62-78.

Kumbhakar, S.C. and C. A. Lovell (2000) *Stochastic Frontier Analysis*. New York: Cambridge University Press, pp279-309.

Leung, H.M. (1997) Total factor productivity growth in Singapore manufacturing industries. *Applied Economics Letters*, 4(8), pp525-528.

Leung, H.M. (1998) Productivity of Singapore's manufacturing sector: an industry level non-parametric study. *Asia Pacific Journal of Management,* 15, pp19-31.

Liao, H, M. Holmes, T. Weyman-Jones and D. Llewellyn (2007), Productivity growth of East Asia economies' manufacturing: a decomposition analysis. *Journal of Development Studies*, 43(4), pp649-674.

Mahadevan, R. and K. Kalirajan (2000) Singapore's manufacturing sector's TFP growth: a decomposition analysis, *Journal of Comparative Economics,* 28, pp828-839.

Murillo-Zamorano, L. (2004) Economic efficiency and frontier techniques. *Journal of Economic Surveys*, 18(1), pp.33-77.

Nehru, V. and A. Dhareshwar (1993) A new database on physical capital stock: sources, methodology and results. *Revista de Analistsis Economico*, 8(1), 37-59.

Nelson, R.P. and H. Pack (1999) The Asian miracle and modern growth theory. *Economic Journal*, 109(457), pp416-436.

Nishimizu, M. and J.M. Page (1982) Total factor productivity growth, technological progress and technical efficiency change: dimensions of productivity change in Yugoslavia, 1965-78. *Economic Journal*, 92, pp920-936.

Pendharkar, P. (2005) A threshold-varying artificial neural network approach for classification and its application to bankruptcy prediction problem. *Computer & Operations Research,* 32(10), pp.2561-2582.

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

Pham, H, G. Park and D. Ha (2002) More than meets the eyes: a new approach to detect sectoral technical change where aggregate growth accounting may not. *Working Paper Series WP02-02*, Dept. of Economics, University of Missouri-Columbia.

Rumelhart, D., G. Hinton and R. Williams (1986) Learning internal representations by error propagation. In *Parallell Distributed Processing: Explorations in the Microstructures of Cognition 1* (ed. Rumelhart D. and J. McClelland), pp.318-362. MIT Press: Cambridge.

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

Sarel, M. (1997) Growth and productivity in ASEAN countries. *IMF Working Paper*, WP/97/97, International monetary Fund.

Sarel, M. (1996) Growth in East Asia: what we can and what we cannot infer. *Economic Issues*, No. 1, IMF, September 1996.

Schreyer, P. (2001) The OECD productivity manual: a guide to the measurement of industry-level and aggregate productivity. *OECD International Productivity Monitor,* No.2, spring 2001, pp37-51.

Stein, R. (1993) Preprocessing data for neural networks. *AI Expert*, pp.32-37.

Stern, H.S. (1996) Neural networks in applied statistics. *Technometrics*, 38(3), pp.205-216.

Troutt, M., A. Rai and A. Zhang (1995) The potential use of DEA for credit applicant acceptance systems. *Computers & Operations Research*, 4, pp.405-408.

Wang, S. (2003) Adaptive non-parametric efficiency frontier analysis: a neural-network-based model. *Computers & Operations Research*, 30, pp.279-295.

Wang, S. (1996) Nonparametric econometric modelling: a neural network approach. *European Journal of Operational Research*, 89(3), pp.581-592.

World Bank (1993) *The East Asian Miracle: Economic growth and Public Policy*. New York: Oxford University Press.

Wu, Y. (2000) Is China's economic growth sustainable? A productivity analysis. *China Economic Review,* 11, pp278-296.

Young, A. (1992) A tale of two cities: factor accumulation and technical change in Hong Kong and Singapore. In: *NBER Macroeconomics Annual 1992* (ed. Blanchard O. and Fisher S.), pp13-54. Cambridge MA, MIT Press.

Young, A. (1994a) Accumulation, exports and growth in the high performing Asian economies. A comment. *Carnegie-Rochester Conference Series on Public Policy,* 40, pp237-250.

Young, A. (1994b) Lessons from the East Asian NICs: a contrarian view. *European Economic Review,* XXXVIII, pp964-973.

Young, A. (1995) The tyranny of numbers: confronting the statistical realities of the East Asian growth experience. *Quarterly Journal of Economics,* CVV, pp641-680.

Figure 1 A three-layer back-propagation neural network

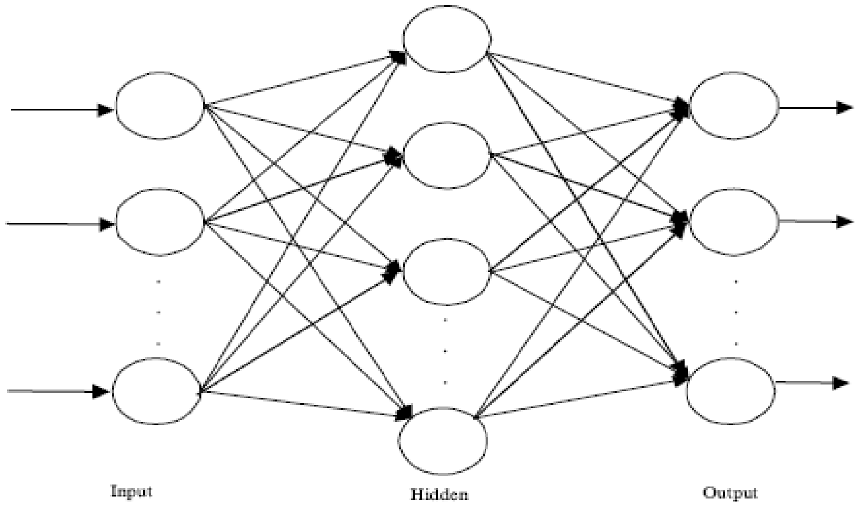

Figure 2 A BP network with two-layer tansig/purelin transfer function

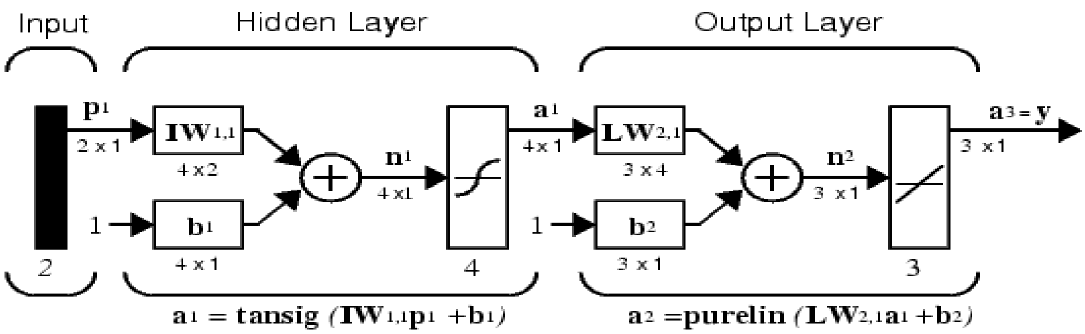

Table 1 Manufacturing Sectors

311-Food Products	342-Printing & Publishing	371-Iron & Steel
313-Beverages	351-Industrial Chemicals	372-Non-ferrous Metal
314-Tobacco	352-Other chemical	381-Fabricated Metal Products
321-Textiles	353-Petroleum refineries	382-Machinery, except Electric
322-Wearing Apparel	354-Misc. Petroleum & Coal	383-Machinery, Electric
323-Leather Products	355-Rubber Products	384-Transport Equipment
324-Footwear	356-Plastic Products	385-Professional & Scientific Equipment
331-Wood Products	361-Pottery, China, Earthenware	390-Other Manufactured Products
332-Furniture	362-Glass and Products	
341-Paper & Products	369-Other Non-metallic Mineral	

Table 2 Estimated neural network parameters

	Results (DEA-NN)	Results (SFA-NN)
Data pre-processing	[0.1, 1]	[0.1, 1]
Network architecture	2-10-1	2-10-1
Activation function:		
hidden/output	Tan-Sigmoid - linear	Tan-Sigmoid - linear
Algorithm	Levenberg-Marquardt	Levenberg-Marquardt
Epochs (max)	1000	1000
R^2	0.9594	0.9893
Mean square error	0.0017	5.09E-04

Table 3 Efficiency main results (Korea)

	DEA crs	DEA vrs	DEA-NN	SFA-NN	SFA
Mean^	0.3256	0.7577	0.7874	0.1961	0.1670
Min	0.0240	0.0390	0.1706	0.0231	0.0290
Max	1.0000	1.0000	1.0000	1.0067	0.9861

^ Weighted means, using shares of manufacturing sector value added as weights.

CRS: constant returns to scale

VRS: variable returns to scale

Table 4 Sources of economic growth

Economies	Methods	TE^	TEC	TP	SEC	TFP	Output growth
Hong Kong	DEA	0.3788/0.8831	-0.00354	0.0536	---	0.0182	-0.0040
	DEA-NN	0.9314	-0.0018	0.0327	---	0.0309	
	SFA-NN	0.3879	0.0009	0.0342	---	0.0351	
	SFA	0.4715	0	0.0215	0.0008	0.0223	
Singapore	DEA	0.5985/0.7966	-0.0237	0.0491	---	0.0254	0.0979
	DEA-NN	0.8671	0.0162	0.0152	---	0.0314	
	SFA-NN	0.5037	0.0017	0.0335	---	0.0352	
	SFA	0.4315	0	0.0410	-0.0029	0.0382	
Korea	DEA	0.3256/0.7577	0.0105	0.0369	---	0.0474	0.1107
	DEA-NN	0.7874	0.0234	0.0350	---	0.0584	
	SFA-NN	0.1961	0.0142	0.0360	---	0.0502	
	SFA	0.1670	0.0152	0.0341	-0.0033	0.0460	
Taiwan	DEA	0.7701/0.8632	0.0140	0.0151	---	0.0291	0.1100
	DEA-NN	0.7826	0.0075	0.0313	---	0.0388	
	SFA-NN	0.632	0.0006	0.0410	---	0.0416	
	SFA	0.5352	0	0.0356	-0.0052	0.0304	
Malaysia	DEA	0.5431/0.8201	0.0210	0.0018	---	0.0228	0.1194
	DEA-NN	0.8511	0.0125	0.0038	---	0.0163	
	SFA-NN	0.3310	0.0150	0.0052	---	0.0202	
	SFA	0.2937	0.0173	0.0156	-0.0073	0.0256	
Indonesia	DEA	0.4114/0.7672	0.0075	0.0603	---	0.0678	0.1813
	DEA-NN	0.8708	0.0211	0.0594	---	0.0805	
	SFA-NN	0.5361	0.0017	0.0260	---	0.0277	
	SFA	0.4916	0	0.0221	0.0097	0.0318	
Philippines	DEA	0.4260/0.7076	0.0189	0.0001	---	0.0190	0.0525
	DEA-NN	0.7925	0.0096	0.0094	---	0.0190	
	SFA-NN	0.5437	0.0006	0.0035	---	0.0041	
	SFA	0.4098	-0.0041	0.0027	-0.0014	-0.0028	

Notes: ^ We present the results for DEA constant returns to scale and DEA variable returns to scale.

Table 5 Comparison of sources of growth decomposition in industrial sub-sectors

		Traditional sector				Basic sector				High-tech sector			
Economies	Methods	TEC	TP	TFP	Output growth	TEC	TP	TFP	Output growth	TEC	TP	TFP	Output growth
Hong Kong	DEA-NN	-0.0049	0.0318	0.0269	-0.0233	0.0003	0.0312	0.0315	-0.0140	-0.0008	0.0461	0.0453	0.0186
	SFA-NN	-0.0015	0.0400	0.0385		0.0023	0.0175	0.0198		0.0002	0.0309	0.0311	
Singapore	DEA-NN	-0.0055	0.0314	0.0259	0.0813	-0.0040	0.0284	0.0244	0.0916	0.0267	0.0293	0.0560	0.01545
	SFA-NN	-0.0008	0.0099	0.0091		-0.0008	0.0248	0.0240		0.0031	0.0403	0.0434	
Korea	DEA-NN	0.0562	0.0209	0.0771	0.0935	0.0068	0.0345	0.0413	0.1158	0.0242	0.0486	0.0728	0.1535
	SFA-NN	0.0155	0.0260	0.0415		0.0138	0.0310	0.0448		0.0169	0.0450	0.0619	
Taiwan	DEA-NN	0.0088	0.0315	0.0315	0.0875	0.0051	0.0289	0.0340	0.1051	0.0120	0.00662	0.0782	0.1600
	SFA-NN	-0.0051	0.0744	0.0693		-0.0013	0.0070	0.0057		0.0045	0.0147	0.0192	
Malaysia	DEA-NN	0.0212	0.0130	0.0342	0.0988	0.0219	0.0075	0.0294	0.1048	0.0197	0.0174	0.0371	0.1643
	SFA-NN	0.0201	0.0036	0.0237		0.0125	0.0058	0.0183		0.0192	0.0063	0.0255	
Indonesia	DEA-NN	0.0279	0.0579	0.0858	0.1859	0.0334	0.0592	0.0926	0.1495	0.0237	0.0424	0.0661	0.2138
	SFA-NN	-0.0099	0.0265	0.0166		-0.0123	0.0268	0.0145		0.0005	0.0289	0.0294	
Philippines	DEA-NN	0.0067	0.01036	0.0171	0.0349	0.0075	0.0124	0.0199	0.0555	0.0100	0.0124	0.0244	0.0665
	SFA-NN	0.0017	0.0018	0.0035		0.0022	0.0031	0.0053		-0.0068	0.0036	-0.0032	

Appendix

A.1 DEA-Malmquist Approach to efficiency measurement and TFP Decomposition

The Malmquist productivity index, introduced by Caves, Christensen and Diewert (1982) (hereinafter refer to 'CCD') and developed by Fare et al (1994) as a way of decomposing productivity measurement is constructed from distance functions, which are reciprocals of Farrell efficiency measures. Fare's approach have a number of desirable properties, for example, easy to computation, no need for either a specific function form or a behavioral assumption such as cost minimization or profit maximization, no requirement on price information, etc. (Hertel, et al 1999). The output distance function is defined at t as

$$D^t(x^t, y^t) = \inf\{\theta : (x^t, y^t/\theta) \in S^t\} \qquad \text{(A.1)}$$

where S^t is the technology set, x^t and y^t are vector of inputs and output, respectively. An output-oriented period t Malmquist productivity index defined by CCD is given by

$$M_{CCD}^{S}(x^{t+1}, y^{t+1}, x^t, y^t) = \frac{D^S(x^{t+1}, y^{t+1})}{D^S(x^t, y^t)} \text{ , s=t, t+1, t=1,…, T-1} \qquad \text{(A.2)}$$

where $D^t(.)$ is the output distance function relative to the technology set for period t. Following Fare et al. (1994), the *adjacent* Malmquist index is defined as the geometric mean of two CCD Malmquist productivity indexes and can be further decomposed as

$$
\begin{aligned}
{}^tM^{t+1}(x^{t+1}, y^{t+1}, x^t, y^t) &= [M_{CCD}^{t} * M_{CCD}^{t+1}]^{1/2} \\
&= \left[\left(\frac{D^t(x^{t+1}, y^{t+1})}{D^t(x^t, y^t)}\right)\left(\frac{D^{t+1}(x^{t+1}, y^{t+1})}{D^{t+1}(x^t, y^t)}\right)\right]^{1/2} \\
&= \frac{D^{t+1}(x^{t+1}, y^{t+1})}{D^t(x^t, y^t)} * \left[\frac{D^t(x^{t+1}, y^{t+1})}{D^{t+1}(x^{t+1}, y^{t+1})}\frac{D^t(x^t, y^t)}{D^{t+1}(x^t, y^t)}\right]^{1/2} \\
&= \text{Efficiency change(TEC) * Technical change(TP)} \qquad \text{(A.3)}
\end{aligned}
$$

Note that, the geometric mean inside the brackets captures the shift in technology between the two periods evaluated at t and (t+1) - technical change, and the first part of equation (A.3) is the ratio of Farrell technical efficiency between two successive period, measuring the capacity to improve technical efficiency from time t to t+1 – a measure of catch-up. Following Fare et al. (1994) and Fare, Grosskopf & Norris (1997), TE can be further decomposed, under the condition of constant-return-to-scale (CRS) technology, into pure efficiency change component calculated relative to variable-return-to-scale (VRS) technology, and scale component capturing changes in the deviation between the VRS and CRS technology. That is,

$$TEC = \frac{D^{t+1}(x^{t+1}, y^{t+1})}{D^{t}(x^{t}, y^{t})} = \frac{D_{VRS}^{t+1}(x^{t+1}, y^{t+1})}{D_{VRS}^{t}(x^{t}, y^{t})} * \frac{\dfrac{D_{CRS}^{t+1}(x^{t+1}, y^{t+1})}{D_{VRS}^{t+1}(x^{t+1}, y^{t+1})}}{\dfrac{D_{CRS}^{t}(x^{t}, y^{t})}{D_{VRS}^{t}(x^{t}, y^{t})}}$$

= Pure Efficiency Change * Scale Efficiency Change (A.4)

Following Fare et al. (1994), the reference technology in period t can be constructed as

$$S^{t} = \Bigg\{ (x^{t}, y^{t}) : y_{m}^{t} \leq \sum_{i=1}^{I} \lambda^{i,t} y_{m}^{i.t}, m = 1,...M,$$

$$\sum_{i=1}^{I} \lambda^{k,t} x_{n}^{i,t} \leq x_{n}^{t}, n = 1,..., N$$

$$\lambda^{i,t} \geq 0, i = 1,..., I \ \Big\} \quad \text{(A.5)}$$

where $\lambda^{i,t}$ is an intensity variable indicating at what intensity production unit i may be employed in production. Then, non-parametric linear-programming techniques are employed to solve for the distance functions that make up the Malmquist index: $D^{t}(x^{i,t},y^{i,t})$, $D^{t+1}(x^{i,t+1},y^{i,t+1})$, $D^{t}_{VRS}(x^{i,t},y^{i,t})$, $D^{t+1}_{VRS}(x^{i,t+1},y^{i,t+1})$, $D^{t+1}(x^{i,t},y^{i,t})$, $D^{t+1}(x^{i,t},y^{i,t})$, for every two consecutive year for each industry in an economy. A total of N(4T-2) linear programming problems must be solved for each economy over the sample period, where N is the number of industry.

A.2 Stochastic Approach to efficiency measurement and TFP Decomposition

We define this so called 'best practice' function f(.) as,

$$y_{it}^{F} = f(x_{it}, t) \quad \text{(A.6)}$$

where y_{it}^F is the potential output level on the frontier at time t for production unit i, given technology f(.), and x_{it} is a vector of inputs. Take logs and totally differentiate (A.6) with respect to time to get

$$\dot{y}_{it}^{F} = \frac{d\ln f(x_{it},t)}{dt} = \frac{\partial \ln f(x_{it},t)}{\partial t} + \sum_{j} \frac{\partial \ln f(x_{it},t)}{\partial x_{jt}} \frac{dx_{jt}}{dt}$$

$$= TP + \sum_{j} e_{jt} \dot{x}_{jt} \tag{A.7}$$

where, variables with a dot over them represent growth rates, and the first term on the right-hand side is the output elasticity of frontier output with respect to time, defined as TP, the second term measures the input growth weighted by output elasticities with respect to input j, $e_j = \frac{\partial \ln f}{\partial \ln x_j} = \frac{\partial \ln f}{\partial x_j}\frac{\partial x_j}{\partial \ln x_j} = \frac{\partial \ln f}{\partial x_j} x_j$. Note that, the conventional conceptualization of TFP growth can be defined as output growth unexplained by input growth[xxxv], i.e.

$$\dot{TFP} = \dot{y}_{it}^{F} - \sum_{j} \frac{w_{jt}x_{jt}}{c} \dot{x}_{jt} \tag{A.8}$$

where, w_{jt} is the price of j-th input and c is the total costs. Combining equation (A.7) and (A.8), one can get

$$\dot{TFP} = TP + \sum_{j} (e_{jt} - \frac{w_{jt}x_{jt}}{c}) \dot{x}_{jt} \tag{A.9}$$

Under the assumption of perfect competition and constant returns of scale, the output elasticities with respect to input j is equal to input share in the total production cost, therefore, TP is the only source of TFP growth. In case of unavailability of input price information, we follow Kumbhakar & Lovell (2000) by assuming $\frac{w_j x_j}{c} = \frac{e_j}{e}$, and the decomposition in equation (A.9) simplifies to [xxxvi]

$$\dot{TFP} = TP + (e-1)\sum_{j} (\frac{e_{jt}}{e}) \dot{x}_{jt} \ .$$

In the spirit of Nishimizu & Page (1982) and further frontier analysis, any observed output y_{it} using x_{it} for inputs can be expressed as,

$$y_{it} = y_{it}^F \exp(-u_{it}) = f(x_{it}, t)\exp(-u_{it}) \qquad \text{(A.10)}$$

where $(-u_{it})$ is a term of output-based technical inefficiency corresponding to observed output y_{it}. The derivative of the logarithm of (A.10) with respect to time yields

$$\dot{y}_{it} = \frac{d \ln f(x_{it}, t)}{dt} - \frac{du_{it}}{dt} = TP + (e-1)\sum_j \left(\frac{e_{jt}}{e}\right)\dot{x}_{jt} - \frac{du_{it}}{dt} \qquad \text{(A.11)}$$

From equation (A.11), TFP growth consists of three components: technical change (innovation and shifts in the frontier technology), technical efficiency change (catching-up) and returns to scale (SEC). That is,

$$T\dot{F}P = TP - \frac{du_{it}}{dt} + (e-1)\sum_j \left(\frac{e_{jt}}{e}\right)\dot{x}_{jt} \qquad \text{(A.12)}$$

This decomposition of TFP growth is useful in distinguishing innovation or adoption of new technology by 'best practice' production units from the diffusion of technology. Coexistence of a high rate of TP and a low rate of change in technical efficiency may reflect the failures in achieving technological mastery or diffusion (Kalirajan, Obwona & Zhao, 1996). However, Nishimizu and Page (1982, p926) ignored the presence of measurement error (v_{it}) in estimating the parameters of the translog approximation to equation (A.10) by using a deterministic frontier. In this study, we are going to estimate equation (A.10) allowing for v_{it} , a symmetric component capturing random variation across production unit and random shocks that are external to its control, into the composed error term with an attempt to distinguish the effects of statistical noise from those of inefficiency so as to obtain consistent and efficient estimates.

Table 1A Sub-categories in Matched Manufacturing Sectors

Categories	Combination of Manufacturing Sectors
	Total Manufacturing
Traditional sector	311/3/4-Food , Beverages and Tobacco Products
Traditional sector	321-Textiles
Traditional sector	322-Wearing Apparel
Traditional sector	323/4-Leather Products and Footwear
Traditional sector	331/2-Wood Products and Furniture
Traditional sector	341/2-Paper Products, Printing and Publishing
Basic sector	351/2/3/4/5/6-Chemicals, Petroleum, Coal, Rubble and Plastic Products
Basic sector	361/2/9-Non-Metallic Mineral Products
Basic sector	371/2/381-Basic and Fabricated Metal products
High-tech sector	382/4-Machinery and Transport Equipment
High-tech sector	383-Electric Machinery and Transport Equipment
High-tech sector	385/390-Other Manufacturing Industries

[1] To name a few, in the fields of bankruptcy assessment (Altman *et al.*, 1994; Pendharkar, 2005); forecasting education spending and productivity (Baker and Richard, 2000; Baker, 2001); customer classification in terms of marketing activities (Kaefer *et al.*, 2005).

[2] A lower value of RMSE meant that the NN forecasting model fit was good. Some researchers argued the reliability of using RMSE or MSE (mean square error) as the evaluation method and suggested several alternatively different accuracy criteria such as mean absolute percentage error (MAPE) and median absolute percentage error (MdAPE). We are still working on it.

[3] Each of the neurons in the input layer is connected to each neuron in the hidden layer. The input layer provides the external input to the network and the input variable value may be a ratio-level value or a categorical value and may be any of several input variables. The hidden layer receives inputs from the input layer or another hidden layer and provides the input to the output layer. The output layer receives the inputs from the hidden layer and then produces the output.

[4] It depends on (1) the complexity of the problem at hand. More hidden nodes are called for in complex problems. (2) the objective of classification. For instance, if the objective is to classify a given set of objects as well as possible, then a larger network may be desirable. On the other hand, if the network is to be used to predict the classification of unseen objects, then a larger network is not necessarily better.

[5] Costa and Markellos (1997) and Santin (2004) use SIC (Schwartz Information Criterion) or AIC (Akaike Information Criterion) as common criteria for model selection. Results are available upon request.

[6] Tansig (tangent sigmoid) is a transfer function with output between -1 and 1.

[7] Output of this transfer function is linear.

[8] Please also see the appendix for detailed discussion of these two methods employed in this study for the purpose of comparison.

[9] Another commonly used non-parametric techniques are free disposal hull (FDH) and goal programming (Fried et al. 1993).

[10] The raw data are mainly drawn from the *UNIDO Industrial Statistics Database* and the International Monetary Fund publication, *International Financial Statistics*. All are available from the Macro-Economic Time Series Data at MIMAS. In addition, price indices are available from the *World Development Indicator* (WDI) CD-Rom and some from countries' national account.

[11] The advantage of using value added is that it accounts for differences in vertical integration, and accommodates quality differences between products as price premiums for quality are translated into higher value added. In addition, as pointed out by Schreyer (2001), the aggregate value-added growth of productivity is a simple weighted average of value added growth in individual industries, which makes this measure comparable across different levels of aggregation. On the other hand, Kim (2000) argued that calculation of TFP changes based on a value-added approach produces a much higher (in absolute term) estimate than one based on gross output since the former requires a more restrictive assumption capital and labour input are separable from intermediate input, that make value-added not an immediately plausible measure of output like gross output. However, if gross output is used then it is necessary to consider all inputs, including material and bought-in services, in forming an explanation. In this study, considering that net output or value added effectively deducts these inputs from both sides of the equation and therefore simplifies the estimation procedure, we will follow the value-added approach.

[12] Since neither the industry-specific GDP deflators, industry-specific producers' price index (PPI) given in the National Accounts nor gross domestic fixed capital formation deflator, can be obtained for some countries/areas, the overall GDP deflator had to be used.

[13] On the one hand, official depreciation derived from the implicit deflator of gross fixed investment on the basis of historical prices would underestimate real depreciation if prices have risen; on the other hand, accounting depreciation tends to overestimate the true depreciation for tax-saving purpose. Although these two factors offset each other to some extent, the real depreciation figures tend to be overestimated. Many works, such as Nehru & Dhareshwar (1993), Collins & Bosworth (1996), have chosen a much lower depreciation rate (4% per year) than the depreciation rate estimated from deflating official nominal depreciation.

[14] Leung (1998) found that the treatment to take account of changing quality of the labour force owing particularly to education did not make any substantial difference to the estimated TFP's, though he used non-parametric approach in calculating industry-level TFP in manufacturing sector for Singapore.

[15] By high-quality, Pendharkar and Rodger (2003) suggested a need to find a balance between the following two requirements – the so-called 'noise-saturation dilemma: (1) the data should adequately represent the fundamental features the network must detect in order to obtain correct outputs; and (2) the training set should provide sufficient variation to allow generalization and discourage 'memorization'.

[16] They carried out a comparative study on the forecasting performance of NNs to linear regression under circumstances of varying data accuracy, and concluded that NN-based forecasts were more robust as the data accuracy degraded.

[17] Stein (1993) argued that, in general, the NN performed better when the input data was normally distributed.

[18] There are also other approaches to pre-process the data, for instance, with a hyperbolic tangent, the ratio of each variable to the maximum figure, and so on. The main aim of normalization is to make the data reasonably distributed around the range of -1 to 1, or 0 to 1.

[19] Athanassopoulos et al (1996) scaled the input and output data to the range 0.2-0.8, allowing 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. We follow this idea, though the range is 0.1-1.

[20] One of NNs advantages is that they can replicate closely any function, including complex nonlinear functions, it is therefore possible to replicate/mimic any kind of outputs. However, the results would become too dependent on the training sample, but might not do well with new observations, hence over-trained in that the model can lose the ability to generalize and instead memorize a pattern which leads to less accurate results when applied to out-of-sample. In order to avoid 'over-training', the error on the validation set is monitored. Training errors decrease, as they should, but validation errors at some point start to increase, indicating that the model is starting to over-fit. When validation errors start to increase, training is stopped and the network that yields the lowest validation error is chosen.

[21] Costa and Markellos (1997) suggested an alternative approach to build up the optimal network topology by over-parametrising the MLP with 50 neurons in the hidden layer. Under this approach, the training is halted at the point when the signal-to-noise ratio reaches a level that satisfies some a priori assumption about the level of statistical noise in the data, for instance, 10% noise in the data. We have also performed this experiment and results are qualitatively similar to those described in the main text.

[22] Two other suggestions on this issue have also been made in literature apart from keeping the single best network. The first one is the simple equally averaging while the other is a weighting scheme which puts higher weight on the best networks. However, the choice of weight is entirely arbitrary, and we use the single best network in the paper instead.

[23] We also tried the ensemble of 500 networks each year, and the result is not qualitatively different from that with 50 networks.

[24] To pursue a balance between the quality of NN and the reasonable duration of training, the NN topology with minimum (testing) MSE is selected as the optimal one with relevant weight and bias.

[25] We have quite similar tables for other economies in question, and only present the results for Korea here for the reason of space-saving. Results for other economies are available from the author upon request.

[26] This algorithm was designed to approach second-order training speed without having to compute the Hessian matrix. When the performance function has the form of a sum of squares, as is typical in training feedforward networks, the Hessian matrix can be approximated as $H=J^TJ$ and the gradient can be computed as $g=J^Te$, where J is the Jacobian matrix that contains first derivatives of the network errors with respect to the weights and biases, and e is a vector of network errors. The Jacobian matrix can be computed through a standard back-propagation technique that is much less complex than computing the Hessian matrix.

[27] The software can produce the convergence characteristics of the proposed NN model which shows the MSE history (decreasing curve) over iterations in the training process.

[28] See Wong (2003) table 2.

[29] Liao et al. (2007) reported that while both the Cobb-Douglas and translog formulations provide positive production elasticities showing no or little violations of monotonicity, the Cobb-Douglas formulation is the only one to fulfil the regularity condition of concavity.

[30] As pointed out by one referee, input congestion can be dealt with by decomposing the technical efficiency scores calculated from a CRS DEA into congestion inefficiency, scale inefficiency and 'pure' technical efficiency. This can be done by assuming weak disposability in inputs/outputs, and then introducing a new parameter in the input restriction, followed by changing the inequalities in the input restriction to equalities (Coelli et al., 2005, p195).

[31] They argued that probability of a monotonic forecasting function learnt by an NN decreases when cases that have lower DEA-based efficiency are used in the training data, and if the original training data sample size is not large – 10 times the number of inputs, then DEA-based screening may not be a viable option.

[32] In the case of Indonesia, results are not consistent among various methods in terms of TFP and its components.

[33] Not reported here in detail but available upon request.

[34] The classification is based on the average capital labour ratios in each manufacturing sector and usual practice. Traditional sector, like food and beverage, textiles and clothing, wood, is labour intensive, while basic sector, including chemical industries, metal works is capital intensive. Other industries, including manufacture of scientific equipment and which are excluded from 382, 383 and 384 according to UNIDO Industrial Statistics Database's classification, are categorised as knowledge intensive. Please see table 1A in the Appendix.

[xxxv] Due to the lack of data on input prices, the output elasticity with respect to input j is equal to input share in the total production cost under the assumption of perfect competition.

[xxxvi] Returns to scale can be defined as $RTS=\sum e_j$