Monograph:
Dietrich, M. (2005) Using simple neural networks to analyse firm activity. Working Paper. Department of Economics, University of Sheffield ISSN 1749-8368

Sheffield Economic Research Paper Series 2005014

Sheffield Economic Research Paper Series

SERP Number: 2005014

Michael Dietrich

Using simple neural networks to analyse firm activity

July 2005

Department of Economics
University of Sheffield
9 Mappin Street
Sheffield
S1 4DT
United Kingdom
www.shef.ac.uk/economics

1. Introduction

Characteristically, in economics, the analysis of firm activity is based on a production function that defines a deterministic relationship between factor inputs and firm output. The analysis of the firm as an organisation takes a somewhat different approach. For instance, behavioural economics (for example Simon, 1955; March and Simon, 1958; Cyert and March, 1963), transaction cost theory (Williamson, 1975, 1985) and capabilities approaches (for example Foss and Loasby, 1998; Foss, 2005) emphasise that economic agents have inevitably incomplete information and knowledge and are at most boundedly or limitedly rational. The implication here is that while general principles governing intra-firm interaction can be specified, detailed organisational processes inside the firm are, for practical academic purposes, effectively unobservable. Hence, the usual analytical tools designed to analyse firm behaviour, based on production functions and optimising principles with full information, are in practice an oversimplification of firm activity (Loasby, 1999).

This problem is not unique to economics. For example, the functioning of the human brain can be understood in general terms as inputs of sensory data generating bio-chemical reactions. These reactions activate various interconnected neurons that in turn lead to (in principle) measurable ‘output’. But, as with the firm, the way in which this general model of the brain is understood and applied in specific circumstances is too complex to be specifiable in terms of a simple functional relationship. For this reason neural scientists have developed neural network analysis to model the interactions inside the brain that are otherwise too complex to be modelled (see, for example, Bishop, 1995). This technique can therefore be usefully applied to the analogous idea of the firm and (potentially) model its functioning assuming complex internal interactions. It is not, of course, original to claim that firm organisation is the brain of the firm; previous work based on this principle is Beer (1972). But this paper is an attempt in this tradition using a neural network framework. Outside of economics, for example in business and finance, neural network analysis is not uncommon (see, for example, Altman et al, 1994; Wilson et al, 1995; Chiang et al, 1996; Wong et al, 1995; Jasic and Wood, 2005). Non-firm applications of neural network analysis within economics are becoming increasingly important, for example: Binner et al (2005) on inflation; Papadas and Hutchinson in input-output analysis; Johnes (2000) on macroeconomic modelling; Franses and Homelen (1998) and Plasmans et al (1998) on exchange rate modelling. But previous published work in economics that has applied a neural network framework to the firm (for example, Delgado et al, 2004) has not developed the analysis using actual data as is done in this paper.

The rest of the discussion is set out as follows. In the next section the key principles of neural network modelling, and how this might be applied to the firm, are set out. In addition basic estimations are reported using a sample of 248 firms in UK SIC34. For comparative purposes the estimation results are compared to Cobb-Douglas and trans-log equivalents. It is shown that neural network frameworks not only provide better estimates but also have superior predictive ability. Following this, in section three, the various estimates are analysed in terms of returns to scale characteristics. It is shown that the most the most complex neural network developed here has an intuitive economic interpretation. In section four a predictive evolutionary model of the firm is developed using the estimated neural network. It is shown that this model has superior predictive capabilities than earlier reported models. In the final substantive section five, simulation results are reported using the neural network based model set out in section four. Finally brief conclusions are drawn.

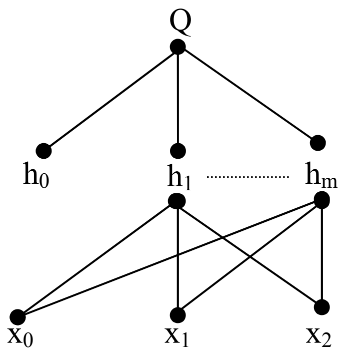
2. Neural network modelling of the firm

A basic neural network (see Bishop, 1995) suggests a two-stage input-output framework rather than a single stage traditional production function. Between inputs and outputs an unobservable hidden layer (**H**) exists:

Traditional production function: $\mathbf{X} \rightarrow Q$

Neural network: $\mathbf{X} \rightarrow \mathbf{H} \rightarrow Q$.

For current purposes, this unobservable hidden layer is assumed to define organisational functioning. With two inputs (x_1 and x_2), the functioning of this network can be set out as follows:



The various factor inputs, along with an input bias (x_0), do not have a direct impact on output, but are instead inputs into m hidden units. These hidden units, along with a hidden unit bias (h_0), determine firm output. This standard formulation is a feed forward network as no feedback effects are modelled.

To model a feed forward network it is assumed that the various hidden units have a separable impact on output. In addition the significance of the various links in the network is defined by a system of weights. In general terms with n inputs, m hidden units, and a single output, we can specify a feed forward neural network as follows (Delgado et al, 2004):

$$Q = F\left[\beta_0 + \sum_{j=1}^{m} G\left(\chi_j + \sum_{i=1}^{n} x_i \alpha_{ij}\right)\beta_j\right]$$

β_0 = the output bias

γ_j = hidden unit biases (j = 1, …, m)

α_{ij} = weights from input unit i to hidden unit j

β_j = weights from hidden unit j to output.

If we assume functional form F is linear in logs, i.e. the analogue of a Cobb-Douglas formulation, we can define in the two input case

$$q = \beta_0 + \beta_1 g_1 + \beta_2 g_2$$

where: $q = \ln(Q)$; $g_j = G\left(\chi_j + \sum_{i=1}^{n} x_i \alpha_{ij}\right)$.

In this case we can interpret $\beta_1/(\beta_1+\beta_2)$ as the 'share' of g_1 in q, and equivalently for g_2. The relevance of this will become clear from later discussion.

There are two common activation functions that define G (see Bishop, 1995):

1. A threshold relationship in which hidden unit j becomes activated following input signals of sufficient intensity.
2. A continuous logistic relationship.

In terms of economic modelling of the firm it is not intuitively obvious why a threshold function should be the appropriate form for the relationship between factor inputs and hidden unit activity.[1] Instead a standard logistic relationship is used here:

$$q = \beta_0 + \sum_{j=1}^{m} \beta_j \frac{1}{1 + \exp(-z_j)} \qquad [1]$$

where, in the two input case, $z_j = \gamma_j + \alpha_{1j}\ln(x_1) + \alpha_{2j}\ln(x_2)$. Logged inputs are used here to maintain comparability with a Cobb-Douglas relationship. The logistic relationship has the desirable property that it is locally (approximately) linear over its middle range and hence collapses back to an approximate Cobb-Douglas relationship. But over extreme z_j it is obviously non-linear.

Formulation [1] can, in principle, be estimated using non-linear regression methods. In addition to qualifications made below on the use of non-linear least squares estimation, we must recognise that any non-linear estimation is potentially sensitive to the algorithm starting point (Curry and Morgan, 1997) because of possible convergence to local solutions (Athanassopoulos and Curram, 1996). For illustrative purposes the estimation is carried out on a sample of 248 UK firms from SIC34 i.e. the automobiles sector. This sample is the complete set of firms available from the FAME database for which SIC34 is the main activity of the firms and covers the full range from small to very large firms. A 2 digit level of aggregation is used to internalise much firm diversification. Data for 1995 and 2000 is used, covering the following: labour (number of employees), capital (net assets) and firm sales. Capital and sales for 2000 are rebased to 1995 levels using the GDP deflator.

---

[1] A 'psychological' model of the firm might use a threshold signal. In this case each hidden unit might describe a different firm 'mental map' with switching between such maps depending on exogenous shocks of sufficient intensity. The necessary shock is defined by the threshold signal. In a managerial context these 'mental maps' might define dominant firm culture. While such modelling is potentially interesting it is left for future work.

To indicate the reliability of a neural network framework the following method is adopted:

1. Use 1995 data to estimate Cobb-Douglas, trans-log and neural network production functions.
2. Use the regression output in (1) and 2000 values of L and K to predict firm sales for that year.
3. Calculate the root mean squared deviation and Theil's inequality coefficient of actual from predicted sales. These measures are taken as indicators of predictive accuracy.

With step (1) we can view a generalised neural network as an arbitrarily complete specification of a data set (see White 1989). It follows that we can view Cobb-Douglas and trans-log functions as, respectively, first and second order approximations of this underlying fully specified relationship. We might therefore expect an actual, empirically derived, neural network to be a more accurate modelling device. This greater accuracy is, indeed, found below. But a possible criticism of a neural network framework is that this greater modelling accuracy may be based on tracking stochastic deviations from underlying systematic economic relationships rather than simply the economic relationships themselves. It therefore may follow that greater modelling accuracy need not imply greater predictive ability if the latter is based on systematic rather than stochastic factors. For this reason steps (2) and (3) are important in indicating the reliability of a neural network framework. We find that a neural network framework has greater predictive ability. Detailed comments about how reliability is assessed are made below.

The main complexity for regression modelling is that standard OLS is a potentially inappropriate tool for production function estimation because an efficiency bound can, in general, be assumed to exist. For this reason two sets of regression results are reported below: non-linear OLS estimates and those derived using a stochastic frontier model. For the latter the total error term is divided into two elements: a random part (v) and the degree of technical inefficiency (u). The random element has the usual characteristics of being independently $N(0, \sigma_v^2)$ distributed over the observations. The inefficiency element is assumed independently half-normal $N^+(0, \sigma_u^2)$ distributed. In addition, to control for non-constant variance in the inefficiency element, because of heteroskedastistic firm scale effects, σ_u^2 is modelled as a linear function of ln(K)-ln(L).

Stochastic frontier models are straightforward to estimate for linear specifications; for instance the STATA package estimates them as standard. But for non-linear formulations, such as a sigmoid based neural network, estimation is not possible using available estimation packages. To undertake neural network frontier estimation the procedure adopted here involves removing the efficiency effects from the dependent variable of the neural network, following which non-linear least squares can be used to estimate the model as if all firms were operating at optimal efficiency. The efficiency effects used in the frontier neural network are those estimated in the trans-log frontier regression. The logic here is that the trans-log formulation represents a 2nd order approximation of the underlying production process rather than the 1st order approximation of a Cobb-Douglas function. This 2nd order approximation produces greater modelling accuracy as indicated below. The resulting regression models are as follows.

OLS Cobb-Douglas:

$$\ln(R) = a_0^{CD} + a_1^{CD}\ln(L) + a_2^{CD}\ln(K) + e^{CD}$$

Frontier Cobb-Douglas:

$$\ln(R) = b_0^{CD} + b_1^{CD}\ln(L) + b_2^{CD}\ln(K) + v^{CD} - u^{CD}$$

OLS trans log:

$$\ln(R) = a_0^{TL} + a_1^{TL}\ln(L) + a_2^{TL}\ln(K) + a_3^{TL}\ln(L)^2 + a_4^{TL}\ln(K)^2 + a_5^{TL}\ln(L)\ln(K) + e^{TL}$$

Frontier trans log:

$$\ln(R) = b_0^{TL} + b_1^{TL}\ln(L) + b_2^{TL}\ln(K) + b_3^{TL}\ln(L)^2 + b_4^{TL}\ln(K)^2 + b_5^{TL}\ln(L)\ln(K) + v^{TL} - u^{TL}$$

OLS neural network:

$$\ln(R) = a_0^{NN} + \sum_{j=1}^{m} a_{1j}^{NN}\frac{1}{1+\exp-\left[a_{2j}^{NN} + a_{3j}^{NN}\ln(L) + a_{4j}^{NN}\ln(K)\right]}$$

Frontier neural network:

$$\ln(R) + u^{TL} = b_0^{NN} + \sum_{j=1}^{m} b_{1j}^{NN}\frac{1}{1+\exp-\left[b_{2j}^{NN} + b_{3j}^{NN}\ln(L) + b_{4j}^{NN}\ln(K)\right]}$$

The appropriate number of hidden units in the neural network is estimated using standard test statistics. In practice, the maximum number of units that can be effectively estimated is two; this has an intuitive economic interpretation presented below. Separate results are presented for networks with one and two hidden units. For trans-log formulations the regressions are reported after removal of insignificant parameters. Results are reported in tables 1-4.

Table 1a: 1995 OLS Cobb-Douglas regression

	estimated coefficient	t statistic
a_0^{CD}	3.926	16.60
a_1^{CD}	0.797	9.99
a_2^{CD}	0.200	4.23
R^2	0.855	
S.E. of regression	0.674	
Log likelihood	-252.65	

Table 1b: 1995 Frontier Cobb-Douglas regression

		estimated coefficient	z statistic
	b_0^{CD}	4.091	18.84
	b_1^{CD}	0.681	8.63
	b_2^{CD}	0.297	4.77
Wald $chi^2(2)$		1427.77	
σ_v		0.617	
$\ln\sigma_v^2$	cons	-0.965	-6.76
$\ln\sigma_u^2$	cons	-3.324	-3.01
	ln(K)-ln(L)	0.523	2.21
Log likelihood		-251.85	

Table 2a: 1995 OLS Trans-log regression

	estimated coefficient	t statistic
a_0^{TL}	5.843	18.04
a_1^{TL}	0.385	3.49
a_3^{TL}	0.041	3.64
a_4^{TL}	0.009	3.78
R2	0.870	
S.E. of regression	0.640	
Log likelihood	-239.25	

Table 2b: 1995 Frontier Trans-Log regression

		estimated coefficient	z statistic
	b_0^{TL}	6.084	15.67
	b_1^{TL}	0.420	3.49
	b_3^{TL}	0.036	3.09
	b_4^{TL}	0.010	3.64
Wald $chi^2(2)$		1635.82	
σ_v			
$\ln\sigma_v^2$	cons	-2.403	-6.25
	ln(K)-ln(L)	0.430	4.12
$\ln\sigma_u^2$	cons	-1.789	-2.39
Log likelihood		-228.80	

Table 3a: 1995 OLS Neural Network: one hidden unit

		estimated coefficient	t statistic
	a_0^{NN}	6.215	6.09
	a_{11}^{NN}	11.417	4.67
	a_{21}^{NN}	-3.367	-4.282
	a_{31}^{NN}	0.378	3.82
	a_{41}^{NN}	0.061	2.63
R^2		0.875	
S.E. of regression		0.628	
Log likelihood		-234.15	

Table 3b: 1995 Frontier Neural Network: one hidden unit

		estimated coefficient	t statistic
	b_0^{NN}	6.544	7.17
	b_{11}^{NN}	11.717	5.04
	b_{21}^{NN}	-3.348	-4.84
	b_{31}^{NN}	0.366	4.21
	b_{41}^{NN}	0.062	2.91
R^2		0.906	
S.E. of regression		0.531	
Log likelihood		-192.42	

Table 4a: 1995 OLS Neural Network: two hidden units

	estimated coefficient	t statistic
a_0^{NN}	6.945	18.39
a_{11}^{NN}	4.111	2.25
a_{21}^{NN}	-11.362	-2.37
a_{31}^{NN}	0.577	2.01
a_{41}^{NN}	0.557	1.55
a_{12}^{NN}	4.588	2.59
a_{22}^{NN}	-4.474	-3.99
a_{32}^{NN}	0.840	2.64
a_{42}^{NN}	0.035	0.55
R^2	0.880	
S.E. of regression	0.621	
Log likelihood	-229.27	

Table 4b: 1995 Frontier Neural Network: two hidden units

	estimated coefficient	t statistic
b_0^{NN}	7.244	19.95
b_{11}^{NN}	3.963	2.11
b_{21}^{NN}	-11.351	-2.44
b_{31}^{NN}	0.551	2.08
b_{41}^{NN}	0.560	1.59
b_{12}^{NN}	4.935	2.63
b_{22}^{NN}	-4.240	-4.57
b_{32}^{NN}	0.755	2.85
b_{42}^{NN}	0.044	0.89
R^2	0.911	
S.E. of regression	0.524	
Log likelihood	-186.98	

Table 5: Predictive accuracy of 2000 sales forecasts

	OLS CD	OLS TL	OLS NN1	OLS NN2	Frontier CD	Frontier TL	Frontier NN1	Frontier NN2
RMS	0.645	0.616	0.613	0.606	0.628	0.584	0.580	0.571
Theil's U	0.0315	0.0301	0.0299	0.0296	0.0307	0.0285	0.0283	0.0278
U^M	0.00101	0.00076	0.00103	0.00115	0.00070	0.00110	0.00114	0.00107
U^S	0.00091	0.00030	0.00004	0.00006	0.00064	0.00026	0.00002	0.00005
U^C	0.02968	0.02912	0.02895	0.02845	0.02946	0.02723	0.02722	0.02681

Note: U^M, U^S and U^C need not sum to U because of rounding errors.

Table 5 reports the root mean squared error (RMS) and Theil's inequality coefficient (U) of 2000 sales forecasts using the eight different models reported above. A clear ranking of forecasting accuracy is indicated, with the neural network models being more accurate predictors that trans-log and Cobb-Douglas formulations. In addition, the frontier models predict better than the OLS equivalents. For this reason the rest of the discussion will be based on use of the frontier models. The table also presents the standard decomposition of Theil's U into proportions representing the bias (U^M), the variance (U^S) and the covariance (U^C). It is apparent that the neural network models are particularly effective in tracking the variance of the forecasts.

3. Interpretation of results: returns to scale

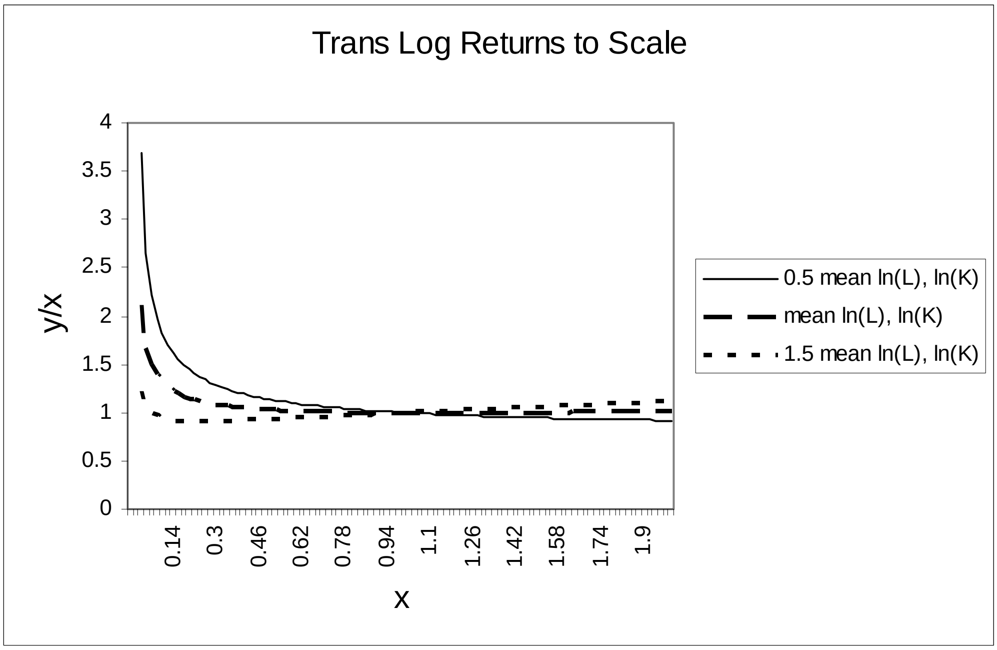
This section of the paper presents a first interpretation of the results in terms of a standard returns to scale analysis. For a production function Q = F(L, K) we can define returns to scale in the standard manner by the ratio y/x, where

$$yQ = F(xL, xK),$$

with y/x greater (less) than one indicating increasing (diminishing) returns to scale.

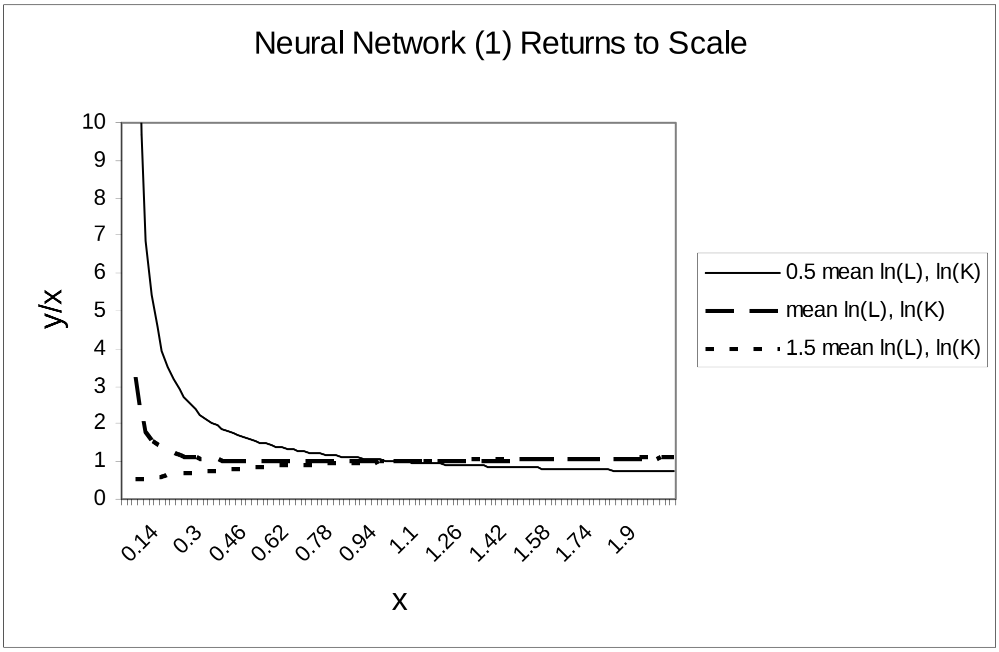
The Cobb-Douglas regression results indicate scale effects that are not significantly different from constant returns. With a trans-log function returns to scale are obviously endogenous to L, K and x. For frontier parameter estimates as given above, returns to scale characteristics are set out in figure 1. With small x, increasing returns to scale become more important with smaller firm size, a result that has an intuitive appeal. Firms of all size appear to converge on approximate constant returns i.e. the Cobb-Douglas solution.

Figure 1



For a neural network returns to scale are also endogenous to L, K and x. With a single hidden unit, and parameter estimates as given above for the frontier model, returns to scale characteristics are set out in figure 2. As with the trans-log function there is convergence on approximate constant returns. But with small size, i.e. 0.5 mean ln(L), ln(K), there are greater increasing returns than with a trans-log function, an effect that results from the non-linearity in the sigmoid relationship.

Figure 2



A neural network with two hidden units allows a potentially more revealing analysis of firm characteristics and behaviour. As mentioned above, because the impacts of the two hidden units are separable, and as we have assumed a linear relationship in logs between hidden units and output, we can define the following relative shares for the hidden units:

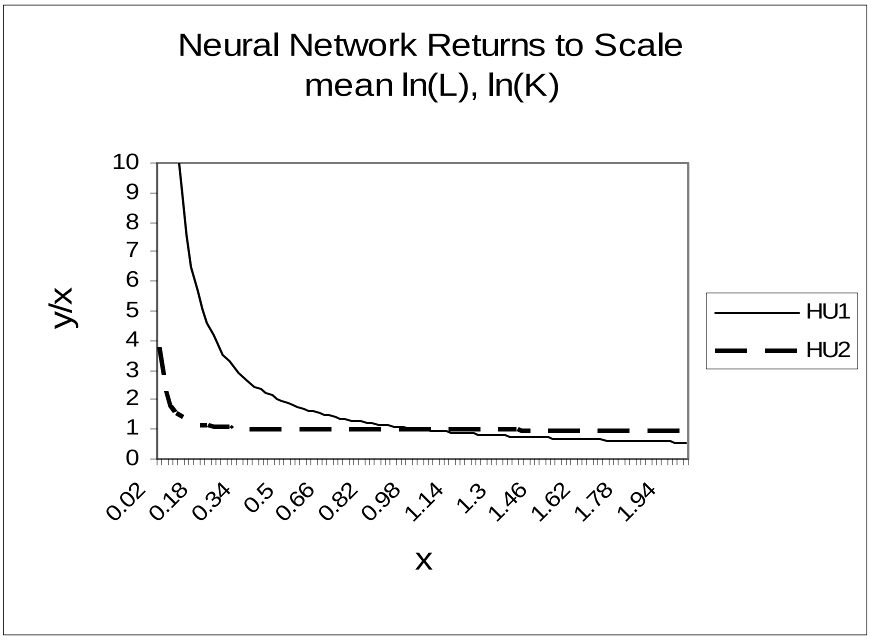
$$s_1 = b_{11}^{NN} / (b_{11}^{NN} + b_{12}^{NN}), \qquad s_2 = b_{12}^{NN} / (b_{11}^{NN} + b_{12}^{NN}).$$

With the parameter estimates reported above

$$s_1 = 0.45, \qquad s_2 = 0.55.$$

This implies that we can interpret hidden unit one as occurring, on average for all firms, 45 per cent of the time and hidden unit two, on average, 55 per cent of the time.

Figure 3



Using this logic we can define separate returns to scale characteristics for the two hidden units. Results for mean ln(L), ln(K) are set out in figure 3. Mapping the returns to scale characteristics exhibited here into implied unit costs, it can be suggested that, for an average firm, hidden unit 2 displays constant unit costs, except for very small x, whereas hidden unit 1, for an average firm, exhibits falling unit costs up to x=1 followed by slightly increasing unit costs. Interpreting the returns to scale characteristics in this manner suggests that hidden unit 2 defines long-run costs whereas hidden unit 1 defines the short-run cost structure.[2] The overall cost structure is consistent with that used in standard undergraduate economics teaching. Long-run average costs initially fall followed by constant returns. Short-run unit costs rise increasingly steeply at outputs lower than expectations and less steeply at outputs

---

[2] This conclusion is reinforced by the fact that the simulated outputs (not shown here) used to derive the returns to scale ratio y/x are consistently greater with hidden unit 2 compared to the equivalent hidden unit 1 position; a finding we would expect if HU1 and HU2 define, respectively, short-run and long-run functioning. A minor conceptual point is, therefore, that describing the ratio x/y for hidden unit 1 as defining returns to scale is incorrect as it appears to describe short-run characteristics.

greater than expectations. The results suggested here imply that, for 1995, SIC34 firms are away from their long-run average cost curve, on average, 45 per cent of the time, whereas output expectations are correct 55 per cent of the time.

4. Predicting firm sales using a two hidden unit network

Interpreting the two hidden unit neural network in the manner just presented reveals an important potential issue when attempting to predict firm sales based on observed input use. Even with unchanged technology, and implied cost structure, prediction is only accurate to the extent that the estimated proportion of time average output expectations are correct is unchanged. This perspective suggests a potentially useful development in the way in which firm output predictions are generated. If the short-run, long-run interpretation of the two hidden unit case presented here is correct, a two hidden unit neural network suggests a possible modelling of changes in s_1 and s_2, and hence in principle greater predictive accuracy.

The modelling of changes in s_1 and s_2 is based on the simple principle that a firm operating away from its long-run cost curve in one operating period will plan, in the next period, to reduce the implied disequilibrium. This modelling requires individual firm estimates of the hidden unit shares rather than the average estimates reported above. Using 1995 parameter estimates, these firm specific shares are generated in the following way:

1. Define firm specific values for s_1 and s_2 as, respectively, s_{k1} and $s_{k2} = 1-s_{k1}$ (k=1, …, 248).
2. For imputed values of s_{k1} (0, 0.1, …, 0.9, 1) and hence implied values of s_{k2} (1, 0.9, … 0.1, 0) define firm specific estimates of b_{11}^{NN} and b_{12}^{NN} :

$b^{NN}_{k11} = s_{k1}(3.963/0.45)$

and $b^{NN}_{k12} = s_{k2}(4.935/0.55)$.

3. Using b^{NN}_{k11} and b^{NN}_{k12} compute predicted 1995 firm sales for each value of s_{k1} and s_{k2} and the root mean squared deviation of actual from predicted 1995 sales.
4. The firm specific values of s_{k1} and s_{k2} are taken as the values that minimise the root mean squared deviation for each firm.

A crude check of the accuracy of the procedure set out in (1)-(4) is to compare the mean of the resulting firm specific estimates of s_{k1} with the earlier reported estimated value of s_1. Both estimates are 0.45.

Having derived estimates of s_{k1} and s_{k2} using (1)-(4), changes in these firm specific parameters over the interval 1995-2000 are modelled as responding to (a) planned movements towards long-run efficiency potential defined by hidden unit two and (b) reactions to external pressures. The planned change is modelled as a standard partial adjustment mechanism:

$$\Delta s^p_{k2} = d(s^*_{k2} - s_{k2}),$$

where: Δs^p_{k2} = the planned change in the second hidden unit share for the k'th firm i.e. the planned change towards long-run potential; s^*_{k2} = the desired s_{k2} for the k'th firm; d = the standard partial adjustment parameter, assumed the same for all firms.

Reactions to external pressures for each firm are modelled using the ratio of firm growth in output over the interval 1995-2000 to average growth in output for all firms i.e. g_k/g_{av}. This ratio is used to define a simple firm specific multiplier (m_k) that is used to model external pressures for each firm:

$m_k = e_1 + [g_k/g_{av}]$^e_2

with the parameters e_1 and e_2 assumed the same for all firms.

The updated value of the share parameter for the second hidden unit (denoted $s^u{}_{k2}$) is modelled by combining planned changes and external pressures as follows:

$$s^u{}_{k2} = m_k(s_{k2} + \Delta s^p{}_{k2}),$$

$$= m_k[ds^*{}_{k2} + (1\text{-}d)s_{k2}]$$

The logic here is there is a planned change in the share parameter, defined by the term in square brackets, the external pressures then impact on this planned change. Simulations are undertaken to determine the values of $s^*{}_{k2}$, d, e_1 and e_2 for this adjustment process. In each case predicted 2000 firm sales use 2000 factor input levels but two hidden unit frontier parameter estimates based on 1995 regression results for all parameters except b_{11}^{NN} and b_{12}^{NN}. The adjustment process parameter values are chosen to minimise the root mean squared deviation of actual from predicted 2000 firm output. Results are reported in table 6.

Table 6: Simulated adjustment parameter results

	$s^*{}_{i2}$	d	e_1	e_2	root mean sq dev
adjustment model	0.98	0.56	-0.30	0.05	0.509
planned adjustment	0.51	0.69			0.530
external pressures			-0.08	0.01	0.637
unchanged shares					0.674

The first row of results in table 6 shows parameter estimates and RMS for the full adjustment model. In addition, and for comparative purposes, three other sets of results are shown: (a) the impact of planned changes alone; (b) the impact of external pressures alone; and (c) imputed firm specific but unchanged 1995 hidden unit shares. It will be recalled that earlier results (reported in table 5) indicated a root mean squared deviation for 2000 sales forecasts, with common and unchanged hidden unit shares, of 0.571 for the two hidden unit frontier neural network model. The simulation results reported here indicate that with firm specific but unchanged unit shares, predictive accuracy deteriorates in terms of root mean squared deviation. A model with external pressures alone also offers no predictive superiority over the earlier two hidden unit model. With planned adjustment alone there is an improvement in predictive accuracy: the root mean squared deviation is 0.530 compared to 0.571 earlier. But with planned changes alone the parameter estimates indicate that firms have an objective of achieving only 51 per cent hidden unit two activity. Intuitively this would seem to be overly small. Combining external pressures and planned adjustment in unit shares further improves predictive capability, indicated by a reduction in root mean squared deviation from the earlier 0.571 to 0.509. In addition, the simulated parameter values are more plausible, indicating that firms have an objective of operating at 98 per cent hidden unit two activity. But over the five year interval 1995-2000 only 56 per cent of adjustment to desired objectives was achieved. An implication of the improvement in predictive accuracy using this adjustment model is to support the earlier conclusion that hidden units one and two describe respectively short-run and long-run activity.

5. Simulating firm and industry evolution

The modelling presented in previous sections can be used to analyse the implied evolutionary characteristics of SIC34 firms. The parameter estimates for the frontier two hidden unit neural network, along with adjustment in hidden unit shares predicted by the adjustment model developed in the previous section, can be used to predict firm sales, with values for one period endogenising the adjustment process for the next period. Periods 0 and 1 for the simulations use actual 1995 and 2000 data. Hence, each iteration can be viewed as representing a five year period.

For simulations reported below, firm specific input use grows (or declines) assuming no input supply constraints exist. Two sets of simulations are undertaken that assume (1) constant, firm specific, capital:labour ratios, and (2) capital:labour ratios adjusting in each period if this increases profitability. The capital:labour ratio adjustment process in (2) allows up to a doubling or halving in each period, with choice being based on the largest profit derivable. These two sets of simulations can be viewed as being based on two different firm types. In (1) firms are assumed to operate with rigid and unchanged routines, fundamental capabilities, or standard operating procedures (depending on the particular view of the firm). In (2) firms are assumed to be long-run optimisers.

The main technical issue with simulating firm evolution is defining and endogenising firm viability. When a firm ceases to be viable exit takes place. This possibility of firm exit occurs when firm losses cannot be rectified by reduction in input use. Hence firm viability requires an ability to avoid losses at some input use and output level. To operationalise this view of viability a zero profit condition for each firm must be used. Such a condition is derived assuming an unknown distribution of firm profitability across the actual (not simulated) population of firms, but that a subset of the actual firms are operating at zero profit. These marginal, zero profit, firms are assumed to define a lower bound for the population. Hence a zero profit condition can be derived by estimating this lower bound. To derive this bound we can recognise that for any one firm using labour (L) and capital (K) inputs

$$\pi = R - p_L L - p_k K$$

i.e. $$R/L = \pi/L + p_L + p_K(K/L),$$

where: π = firm profit; R = firm revenue; p_L and p_K = input prices.

It follows that for a zero profit, marginal firm

$$R/L = p_L + p_K(K/L)$$

and for a positive profit, non-marginal firm

$$R/L > p_L + p_K(K/L).$$

It is assumed that p_L and p_K are the same for all marginal firms and that labour and capital prices are no lower for non-marginal than marginal firms. Using these assumptions a scatter diagram of (R/L) against (K/L) can be used to derive the lower bound, marginal firms.

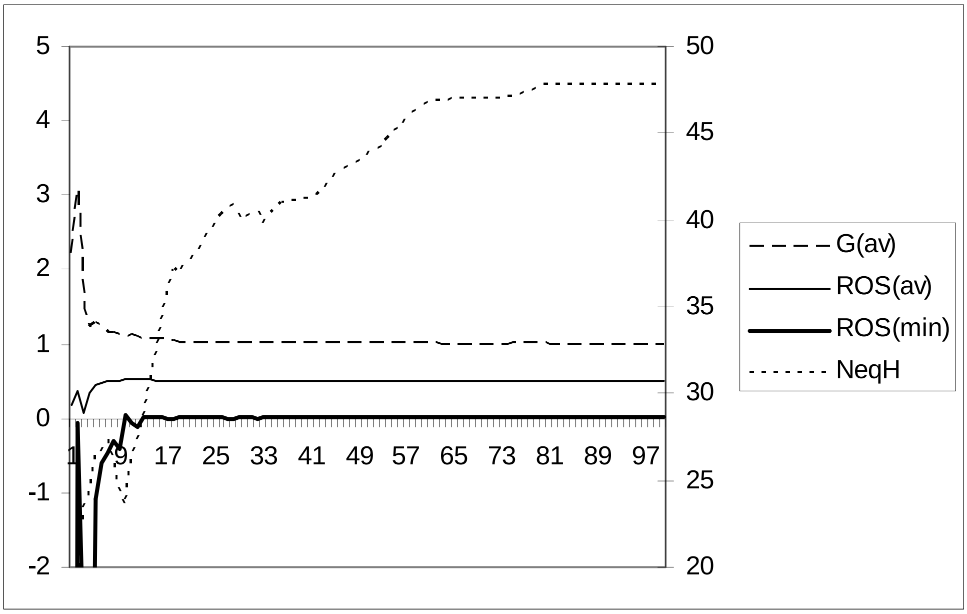
To derive the lower bound the following procedure is used:

1. Locate the observation with minimum average labour product: $(R/L)_1$ and the associated $(K/L)_1$
2. From the point $[(R/L)_1, (K/L)_1]$ determine the slope of the rays to all other points $[(R/L)_k, (K/L)_k]$ (k = 2, 3, …., 258).
3. Rank the absolute values of the slopes derived in (2) in ascending order. The B-firm lower bound is then defined by the first B observations in this ranked series.
4. Formulate the regression equation for the B observations

   $$R/L = f_0 + f_1(K/L).$$

The regression equation in (4) defines the B-firm lower bound estimates of p_L and p_K as respectively f_0 and f_1. Three different values for B were initially used: 10, 20, 30. Simulation results for B=20 and B=30 resulted in the majority of firms becoming non-viable in all simulations. This unlikely result is taken to indicate that B=20 and B=30 defines a bound that is too wide. For this reason, a 10 firm lower bound is used below.[3] For simulation purposes, the lower bound estimates of p_L and p_K are used to impute firm profitability. For any firm, in any period, in which the calculated profitability is negative factor inputs that generate zero profit are calculated, assuming a constant or variable K/L (as relevant). If this resulting K/L is negative the firm is deemed non-viable and hence exit occurs.

Figure 4: Basic simulation, constant K/L



[3] An alternative method to identify the lower bound can involve using steps (1)-(3) in the text, but then using the B-firm lower bound to define a B-firm dummy variable that is 1 for the B marginal firms. Using this dummy variable a regression with two 'regimes' can be defined involving marginal and non-marginal firms. These two methods give effectively the same result for the bound. But the method defined in (1)-(4) is preferred because omitted variable bias will be, potentially, present to the extent that factor prices are not the same for non-marginal firms and that non-marginal firms have different positive profits.

The basic simulation with constant K/L is shown in figure 4. The dotted line (NeqH) shows the numbers equivalent Herfindahl index. This index is measured on the right hand scale whereas the other three series in figure 4 are measured on the left hand scale. The starting point for the numbers equivalent index, the actual value for the year 2000, is just under 20. This actual figure for the number of equivalently sized firms compares with 248 actually existing firms in 2000. Apart from two temporary falls, NeqH increases until a steady-state is achieved with 47 equivalently sized firms, i.e. the 'natural' evolution of the sector is for it to become less concentrated. This lower concentration occurs even though there are fewer actual firms because of the exit of those that are non-viable. This characteristic is based on the simulated growth of smaller compared to larger firms. In addition, it is perhaps relevant to point out that steady-state concentration levels are only achieved after (approximately) 80 periods. Given the five year periodicity of the data, 80 periods implies 400 years. This indicates that in real time concentration levels will effectively not equilibriate.

The G(av) curve in figure 4 reports the average growth in turnover for all firms, excluding those that have exited. It can seen that this reaches a steady-state with G(av) = 1. This steady state is effectively achieved after approximately 20 periods. The implied 100 years indicates that in real time average profits are not in equilibrium. The initial increase in G(av) is caused by the early exit of non-viable, loss making firms. This early non-viability is indicated by the negative minimum return on sales, i.e. ROS(min) shown as a bolded line. This minimum return on sales is seen to stabilise at a level that is effectively zero. If the entry of new firms is based on the profitability of marginal existing firms, it follows that the long-run zero level for ROS(min) is consistent with no new firm entry. The average return on sales is shown as ROS(av). It can seen that this equilibriates to 0.5 after approximately 15 periods, equivalent to a real time 75 years.

Figure 5: Basic simulation, variable K/L

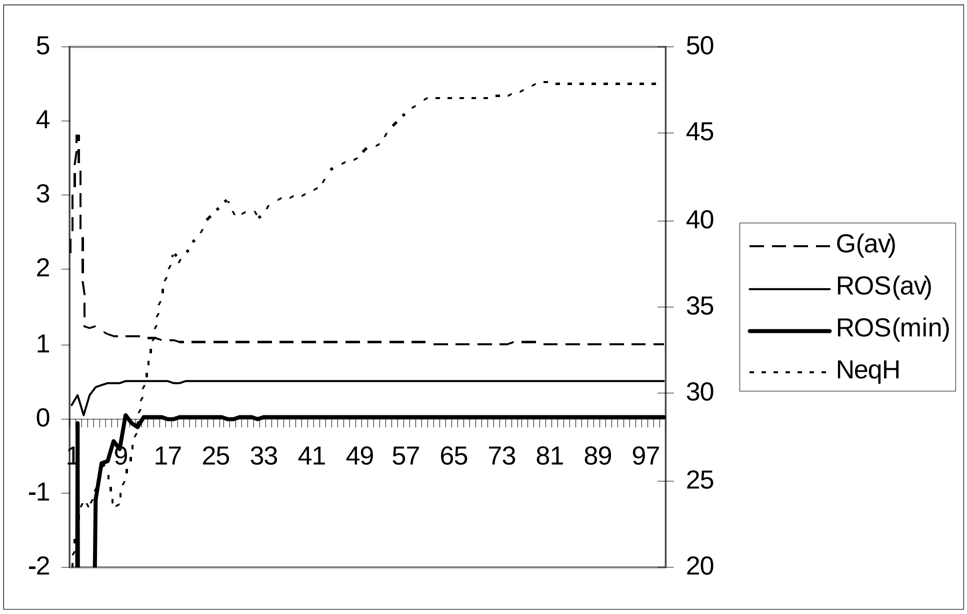


Figure 5 shows the equivalent simulation to that reported in figure 4 but with firms able to vary K/L. With respect to the numbers equivalent Herfindahl index it can be seen that in the early simulated periods the sector is more concentrated than in figure 4. This would appear to suggest that the adjustment in K/L offers a relative advantage to larger rather than smaller firms. But the equilibrium level of NeqH in figure 5 is effectively the same as in figure 4, indicating that the early advantage to larger firms is temporary. In addition equilibrium concentration is achieved in the same time scale with constant and varying K/L. Allowing firms to be long-rum optimisers has the perhaps logical result in figure 5, that in the early disequilibrium periods average growth is higher than in figure 4. But average and minimum return on sales appear to be minimally affected by varying K/L.

To test the stability of these results two additional sets of simulations are undertaken. The first allows the log of firm sales in each simulated period to be increased by 1.05 i.e. an exogenous, and continuing sales expansion is introduced. The second change imposes an equivalent sales decline of 0.95 for each firm in each simulated period. Figures 6 and 7 show the effect of the exogenous expansion in sales for simulations based on constant and varying K/L.

Figure 6: Sales expansion by a factor of 1.05, constant K/L

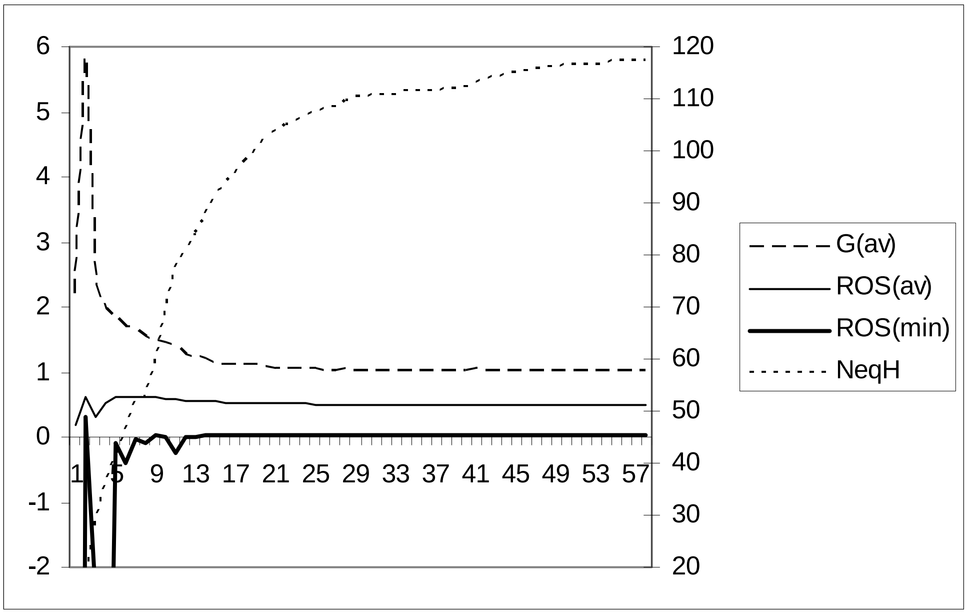
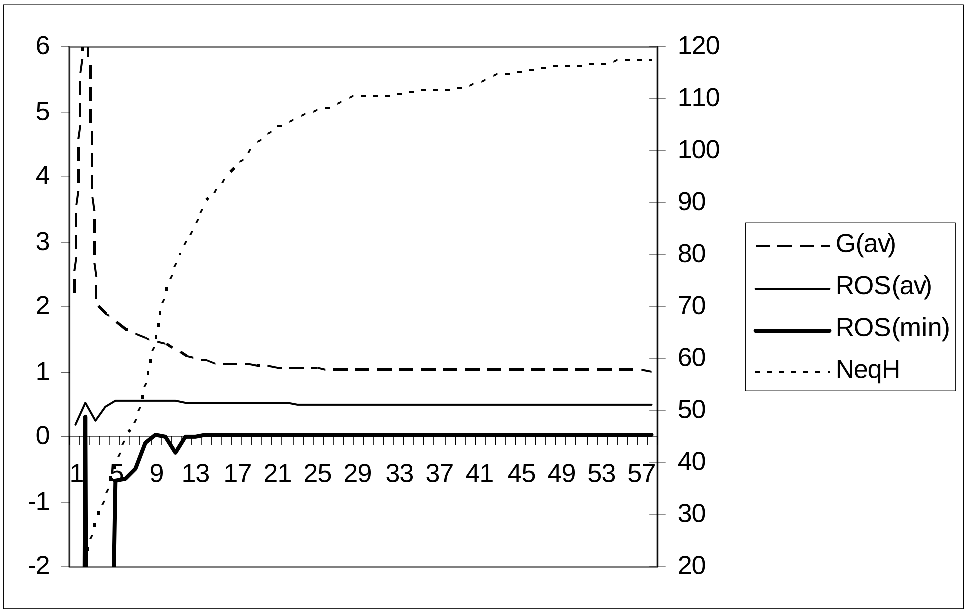


Figure 7: Sales expansion by a factor of 1.05, variable K/L



Comparing figures 6 and 4 (note the different number of iterations reported in these two cases) perhaps the most dramatic change is the greater level of deconcentration that is introduced when an exogenous sales expansion is imposed. The equilibrium level of NeqH in figure 6 is above 116 equivalently sized firms. Because of the non-linearities embodied in the neural network framework the sales expansion is to the relative disadvantage of larger firms. In addition, equilibrium concentration is achieved after approximately 55 periods in figure 6, compared to the earlier approximately 80 periods. But 55 periods is equivalent to 275 years, hence the substantive economic conclusion is unchanged. It is also apparent in the comparison of figures 6 and 4 that average growth and average return on sales are, logically, greater with the sales expansion in the early disequilibrium periods. But the minimum return on sales in figure 6 still stabilises at (effectively) zero. The reason for this is that the sales expansion reduces firm exit with the result that the marginal firm is still earning only normal profit. Using earlier arguments, this suggests that a sales expansion will not lead to the entry of new firms.

Comparing figures 6 and 7, one differences is that the initial disequilibrium growth is (logically) greater in figure 7. But in both cases equilibrium average growth is achieved only after the equivalent of greater than 125 years. In addition, the initial average return on sales is lower in figure 7 than in figure 6. This average return on sales effect is produced by relatively inefficient, and hence low profit, firms having greater opportunities to generate efficiency gains by changing capital:labour ratios. An implication here is that a population of optimising firms need not have higher average profitability than a population of non-optimising firms. For the latter population greater firm exit will occur.

Figure 8: Sales contraction by a factor of 0.95, constant K/L.

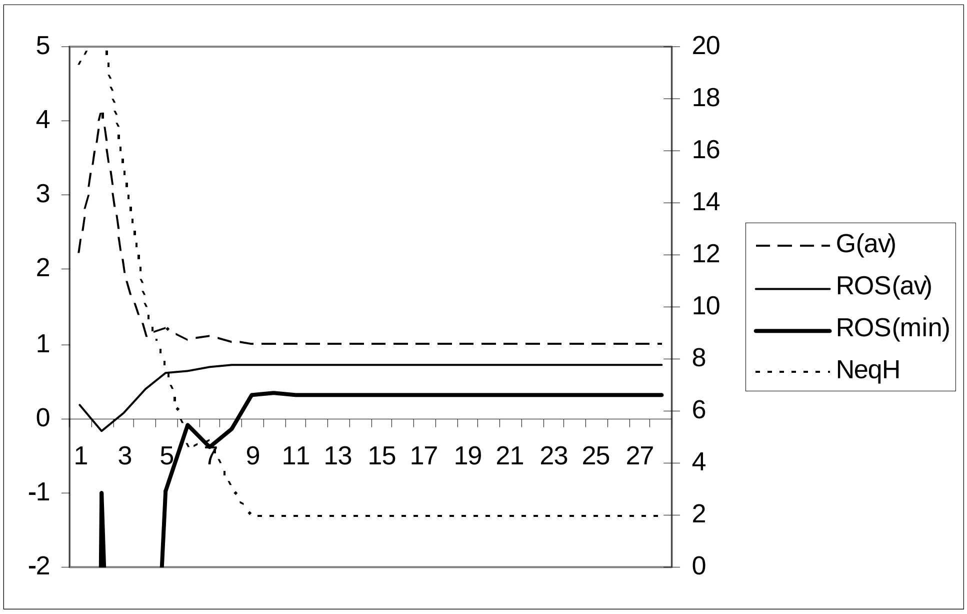
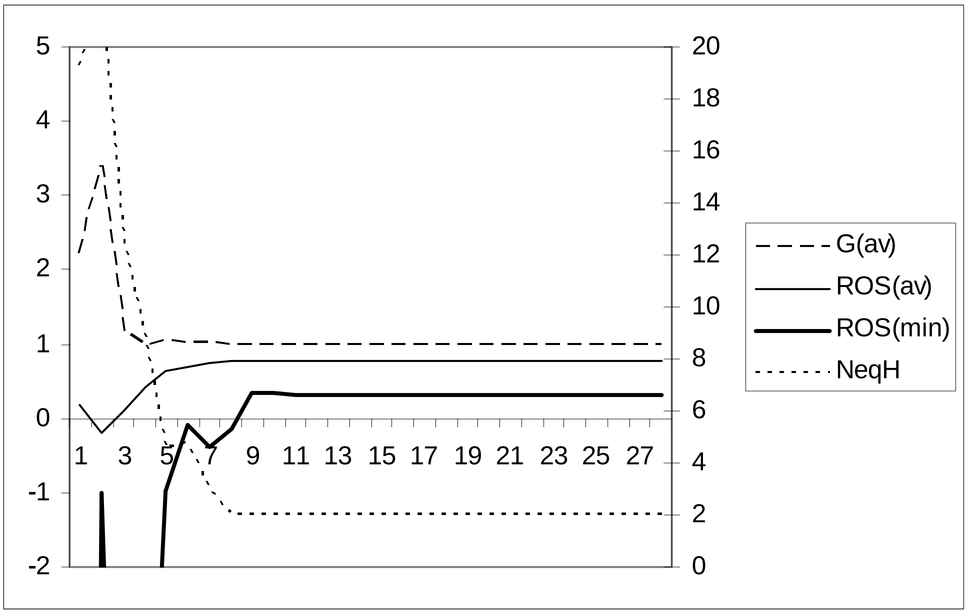


Figure 9: Sales contraction by a factor of 0.95, variable K/L



It is apparent from figures 8 and 9 that a sales contraction has a significant concentrating effect. In both figures the number of equivalently sized firms falls from less than 20 in 2000 (i.e. period 1) to an equilibrium level of (effectively) 2 firms i.e. the 'natural' evolution produced by contraction is to become more monopolised. Furthermore equilibrium concentration is achieved after 8 periods. This is faster than earlier results but is still equivalent to 40 years. In addition to this point, another apparent difference, when figures 8 and 9 are compared to earlier simulations, concerns predicted return on sales. Both the average and minimum ROS is greater with a sales contraction. The reason for this is that the sales contraction produces more disequilibrium firm exit because of non-viability with the result that the marginal firm has greater equilibrium profits. In turn this characteristic of the marginal firm increases the average equilibrium profitability of the sector as a whole. It follows that if new firm entry is determined by the characteristics of the marginal firm, sales contraction may eventually promote entry because of the shake-out of relatively inefficient firms. Comparison of figures 8 and 9 indicates that the only real difference between populations of sub-optimising compared to long-run optimising firms is that the speed of adjustment in the average growth rate appears faster in figure 9 compared to figure 8. The implication here is that optimisation has a dynamic but not an equilibrium effect.

6. Conclusion

This paper has shown that neural networks can be an effective tool for the analysis of the firm. Both one and two hidden unit networks provide better frameworks than those traditionally used in economics in terms of both estimation and prediction. In addition a two hidden unit network appears to have an intuitive economic interpretation in terms of short-run and long-run decisions. This characteristic allows a further modelling development based on adjustment to efficient, hidden unit two, behaviour. This more evolutionary model was investigated here using simulation techniques. It was shown that the neural network framework is consistent with different evolutionary paths depending on exogenous market growth or decline. In addition the model is shown to predict that steady states, while being theoretically achievable, only occur after arguably excessive iterations. In effect, given reasonable economic interpretations of real time, the neural network model is shown to be consistent with a disequilibrium perspective.

These insights indicate a comparative advantage to the use of neural networks in the analysis of the firm. But the 'simple' analysis referred to in the title perhaps needs some comment. In this context the following points would seem to be pertinent conclusions. First, the introduction to the paper made reference to approaches to the firm that emphasise the importance of organisation. The hidden units in a neural network were assumed to describe these organisational factors. But the data used here was restricted to two productive inputs. In principle it is possible to separate costs into costs of production and other non-production costs. This can be mapped into an analysis in which both production-based and organisational factors of production can exist; see Dietrich (2003) for a non-neural network framework based on these principles. Secondly, the simulation results indicate that even with variable K/L, and in equilibrium, firms' profitability does not converge. The underlying profitability dispersion is based on 'x' inefficiencies, i.e. labour and capital productivities, defined by the inefficiency vector u. More sophisticated modelling might allow for best practice productivities to diffuse through the population of firms. This would appear to require more than the simple feed forward neural network used here. Finally, while some of the above discussion has an intuitive economic logic it might be the case that different sectors display different characteristics. These comments, and the discussion in the paper, indicate a rich vein of research potential.

References

Altman EI, Marco G, Varetto FC (1994), "Corporate distress diagnosis: comparisons using linear discriminant analysis and neural networks (the Italian experience)", Journal of Banking and Finance, 18, 505-529.

Athanassopoulos AD, Curram SP (1996), "A comparison of data envelopment analysis and artificial neural networks as tolls for assessing the efficiency of decision making units", Journal of the Operational Research Society, 47, 1000-16.

Binner JM, Bissoondeeal RK, Elger T, Gazely AM, Mullineux AW (2005), "A comparison of linear forecasting models and neural networks: an application to Euro inflation and Euro Divisia", Applied Economics, 37(6), 665-680.

Beer S (1972), Brain of the firm: the managerial cybernetics of organization, London: Allen Lane.

Bishop CM (1995), Neural networks for pattern recognition, Oxford, Clarendon Press.

Chiang WC, Urban TL, Baldridge GW (1996), "A neural network approach to mutual fund asset value forecasting", OMEGA, 24, 205-16.

Curry B, Morgan P (1997), "Neural networks: a need for caution", OMEGA, 25, 123-33.

Cyert RM and March JG (1963), A Behavioral Theory of the Firm, Englewood Cliffs NJ, Prentice Hall.

Delgado FJ, Valiño A, Santín D (2004), "The measurement of technical efficiency: a neural network approach", Applied Economics, 36(6), 627-635.

Dietrich M (2003), "The Importance of Management and Transaction Costs for Large UK Firms', Applied Economics, 35, 1317-29.

Foss NJ (2005), Strategy, Economic Organization and the Knowledge Economy: the coordination of firms and resources, Oxford, Oxford University Press.

Foss NJ and Loasby BJ (eds) (1998), Economic Organization, Capabailities and Co-ordination, London, Routledge.

Franses PH, Homelen Pv (1998), "On forecasting exchange rates using neural networks", Applied Financial Economics, 8 (6), 589-596.

Jasic T, Wood D (2004), "The profitability of daily stock market indices trades based on neural network predictions: case study for the S&P 500, the DAX, the TOPIX and the FTSE in the period 1965-1999", Applied Financial Economics, 14 (4), 285-297.

Johnes G (2000), Up Around the Bend: linear and nonlinear models of the UK economy compared", International Review of Applied Economics, 14 (4), 485-493.

Loasby BJ (1999), Knowledge, Institutions and Evolution in Economics, London, Routledge.

March JG and Simon HA (1958), Organizations, London, John Wiley & Sons.

Papadas CT, Hutchinson WG (2002), "Neural network forecasts of input–output technology", Applied Economics, 34 (13), 1607-1615.

Plasmans J, Verkooijen W, Daniels H (1998), "Estimating structural exchange rate models by artificial neural networks", Applied Financial Economics, 8 (5), 541-551.

Simon HA (1955), "A behavioural model of rational choice", Quarterly Journal of Economics, 69, 99-118.

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

Williamson OE (1975), Markets and Hierarchies: analysis and anti-trust implications; a study in the economics of internal organization, New York, Free Press.

Williamson OE (1985), The Economic Institutions of Capitalism: firms, markets, relational contracting, London, Macmillan.

Wilson N, Chong KS, Peel MJ (1995), "Neural network simulation and the prediction of corporate outcomes", International Journal of the Economics of Business, 2, 31-50.

Wong BK, Bodnovich TAE, Selvi Y (1995), "A bibliography of neural network business applications research: 1988-1994", Expert Systems, 12, 253-261.