CONTRASTING NEURAL NETS
WITH REGRESSION IN PREDICTING
PERFORMANCE IN THE TRANSPORTATION INDUSTRY

by

Katherine A. Duliba
Information Systems Department
Leonard N. Stern School of Business
New York University
90 Trinity Place
New York, New York 10006

October, 1990

Abstract

The purpose of this paper is to compare and contrast traditional regression models with a neural network model, in order to predict performance in the transportation industry. No regression model has emerged as obviously superior in previous work conducted on predicting transportation performance. Therefore, a neural network model was investigated as an alternative to regression. It was found that a neural net model outperformed the corresponding random effects specification, but did not perform as well as the fixed effects specification.

1. Introduction

Neural nets have been applied to two different categories of problems: recognition problems and generalization problems. Recognition problems include visual applications, such as learning to recognize particular words, and speak them [10]. Generalization problems include classification and prediction problems. Dutta and Shekhar [4] predict bond ratings using a neural network model, and contrast it with a regression model. They found that neural nets outperformed regression analysis. However, there are a few issues associated with this comparison that they did not investigate. For example, they contrasted a neural net with an additive regression model. Yet, if a neural net is capable of extracting nonlinearities from a dataset, perhaps a multiplicative regression should have been compared to the neural net. A multiplicative regression may outperform an additive regression if nonlinearities exist. In addition, it appears they used a random effects model for the regression, and not a fixed effects model, and compared that to the neural net. An interesting question which arises is whether it is possible for a neural net to outperform both types of regressions. These issues regarding the functional forms of models will be investigated. Perhaps it is inappropriate to compare a neural net with an additive model. Is it possible for a neural net to capture the fixed effects contained within a dataset? This study investigates these issues using a real-world example, that of predicting transportation performance.

Transportation performance has been predicted using different models and mathematical specifications. For example, some researchers [2] specify multiplicative regressions, while others [3] specify additive regressions. In addition, some [2] specify a fixed effects model to predict performance, while others [3] specify a random effects

model. All of these models have met with varying degrees of success, that is, they more or less accurately predict performance, and the parameters of the model are statistically significant. However, from this work it is not clear what is the correct specification of the model. Is the transportation data linear, or does it contain nonlinearities? Are there fixed effects associated with the data, or is a random effects model applicable? The purpose of this paper is to train a neural net using transportation data, and compare it to additive, multiplicative, fixed effects, and random effects specifications in a regression analysis. To date, it is not clear what kind of data is best modeled by a neural net. This study attempts to investigate the conditions under which a neural net model is more applicable than a regression model, and vice versa.

The plan of the paper is as follows. Section 2 outlines some previous work in modeling performance in the transportation industry. Section 3 reviews neural network models, and compares them to regression models. Section 4 discusses the method used to estimate the regression parameters and train the neural net. Section 5 describes the results, and Section 6 provides some concluding thoughts.

2. Modeling in the Transportation Industry

This section describes some of the work which has been done regarding modeling performance in the transportation industry. First, some measures of performance are outlined, followed by a discussion of some factors which might account for the differences in performance. The section concludes by reviewing different specifications of the models.

2.1. Transportation Performance

Performance in the transportation industry has been measured from a few different aspects: financial results, productivity, and output to input ratios. Financial results include operating revenue, operating expense, and operating profit or loss. Operating revenue consists of revenue obtained from the sale of tickets. Operating expense includes primarily variable costs, such as salaries and fuel, but it also includes depreciation and amortization of equipment.

Productivity in the transportation industry is frequently measured using "Percent Capacity Utilized" [3]. This measure indicates how full the particular transport equipment is. Take a specific example in the train industry. If there are 400 seats available on the train, and 300 are occupied by passengers, the percent capacity utilized is 75%. A higher utilization indicates higher productivity.

Various output to input ratios have been calculated to determine transportation productivity. Some of these ratios are simple, some are complex. Simple ratios, for example, include available capacity/employee. An example of a complex output to input ratio is total factor productivity. Total factor productivity has been calculated using a combination of a few different indexes of input (labor, equipment, fuel, and materials) and one index of output (capacity utilized) in order to arrive at a total factor productivity index [2]. It is argued that the sophisticated ratio provides a more accurate picture of the real productivity than the simple ratios.

2.2. Analysis of Differences in Performance

Some studies have examined factors which attempt to account for the differences in transportation performance. These factors include average distance and output, defined

below.

Average distance has been used as a factor contributing to productivity, measured as percent capacity utilized or output to input ratios, in a number of studies [2,3]. Average distance is the average length between the origin and destination. It is considered to be a leading indicator of structure. Structure refers to whether a particular transportation mode consists primarily of long trips or short trips. Structure is important because it has been found that increased average distance contributes to higher productivity.

Output has been measured both as available capacity, and as used capacity, and has been considered to be a factor leading to productivity. Available capacity is the total number of seats available, while used capacity consists of the total number of passengers. These output measures are sometimes weighted by the length of the trip. For example, if there are 75 passengers travelling 100 miles, the output would be 75 x 100 = 7500.

2.3. Previous Model Specifications

Productivity in the transportation industry has been investigated using a variety of mathematical specifications. For example, one study [2] predicts performance from output and average distance. It calculated total factor productivity to measure performance. The authors use a multiplicative, fixed effects specification for their model, with a fixed effect (dummy variable) for each year and each firm (See Table 1 for coefficients):

$$\ln\text{TFP}_{it} = \alpha + \sum_{t=1972}^{1976} d_t + \sum_{i=1}^{10} d_i + \beta_{OP}\ln OP_{it} + \beta_{AD}\ln AD_{it} + \beta_{PCU}\ln PCU_{it}$$

TFP	= Total Factor Productivity
d_t	= Dummy variable for time
d_i	= Dummy variable for firm
OP	= Output
AD	= Average Distance
PCU	= Percent Capacity Utilized

In contrast, another study [3] estimates transportation productivity using both additive and multiplicative random effects specifications. No dummy variables for either firm or year are included in the model. The authors predict percent capacity utilized from average distance, the number of passengers, and the number of firms servicing a particular market:

$$PCU = 0.588 - 2.11*10^{-5}AD + 7.62*10^{-7}N - 7.06*10^{-2}F$$

$$PCU = 0.257 - 0.019\ln AD + 0.073\ln N - 1.46\ln F$$

PCU	=	Percent Capacity Utilized
AD	=	Average Distance
N	=	Average Number of passengers in market
F	=	Number of Firms in market

In sum, different researchers specify different functional forms when modeling transportation performance. It is not clear which one, if any, is the right model. The next section discusses now neural nets can be used to model transportation performance.

3.   Neural Nets

This section briefly reviews the neural net model, how learning occurs in a neural net, and contrasts it to regression models (and how they can be used as an alternative to

regression models.)

3.1. The Neural Net Model

A neural net is a model consisting of units, also called neurons, activation levels associated with each neuron, and connections between the neurons [9]. The neurons are elementary processing elements. There are three possible types of neurons: input neurons, output neurons, and hidden layer neurons. The input neurons receive input from sources external to the neural net. The output neurons send signals outside of the neural net. The hidden layer neurons act as feature detectors, corresponding to certain regularities or generalizations in the data. What these generalizations are does not have to be known prior to specifying the model; the neural net determines them. With the hidden layer, the neural net can perform more powerful classifications than it can without the hidden layer. In particular, a three-layer neural net (not counting the input layer) can separate arbitrarily complex data, whereas a two-layer neural net can separate data existing in a convex open or closed region, and a single-layer neural net can separate data which falls on opposite sides of some hyperplane [8].

The activation level refers to the strength of the neuron. It is determined by a linear or nonlinear function, associated with each neuron, which transforms the inputs to that neuron into a single output [9].

There are two types of connections between neurons: excitatory connections and inhibitory connections. Excitatory connections are positive weights between the neurons, while inhibitory connections are negative weights between the neurons. There are two different types of networks, based on two different ways of connecting the neurons. First, if the neurons in one layer are connected to neurons in the next layer, with no

distribution of the data when training a neural net. It is possible that the neural net model is general enough to be useful for problems where the assumptions of regression are not met.

4. Method

4.1. Variables

This study uses percent capacity utilized to measure transportation performance. Percent capacity utilized is widely used within the industry as an indicator of performance. This statistic is closely monitored; changes in just a few percentage points can cause a firm to move from operating at a profit, to operating at a loss. If percent capacity utilized drops below a certain level, the firm can react, such as offering special disco nts for a limited time. Financial results, such as operating profits, were not used as it is beyond the scope of this study to find all the factors which influence operating profits, and include them in the model. Some of the factors would be beyond the firm's control, such as increases in the price of fuel, or general economic conditions. Finally, total factor productivity was not used because the transportation industry does not appear to calculate or monitor this statistic, whereas percent capacity utilized is frequently mentioned in trade publications.

Based on the review of previous work, three variables were selected to predict transportation performance: capacity utilized, average distance, and advertising. Capacity utilized is a measure of output. It has been used to calculate the output index. One previous study [3] used the average number of passengers, N, to measure output. However, N does not take into account the different distances that different passengers

travel. Therefore, capacity utilized weighted by the distance traveled is used to measure output. This measure includes more information relating to output than simply the number of passengers. Capacity utilized is related to percent capacity utilized in that, if everything else remained constant, the more capacity utilized associated with a transportation firm, the greater the percent capacity utilized.

Average distance has been used in previous work to predict performance [2,3]. One of the first studies [3] found that as the average distance increased, the percent capacity utilized decreased. The authors explain this result by noting that the break-even percent capacity utilized is lower on long distances than short distances, and therefore long distance trips can provide a profit even though they have fewer passengers than short distance trips. However, later studies [1,3] have found a positive relationship between average distance and percent capacity utilized. The later studies explain that the difference in the results may be due to data from a peculiar year, 1969, which was used to estimate the initial model. That particular year was a recession year in which percent capacity utilized averaged about 50%. Since then, it has averaged between 55% and 62%. The later study [1] estimates their model twice, using two different years of data (1976 and 1981), and obtains a positive relationship between average distance and percent capacity utilized each time. A different study [2] estimates a similar model using a total of five years of data (1972 to 1976), and obtains a positive relationship as well. Therefore, the result obtained by the initial study may be due to an anomaly in the data. It is hypothesized that average distance will be positively related to percent capacity utilized.

Advertising is hypothesized to affect performance. If a particular firm in the

transportation industry advertises more effectively than another firm, it is hypothesized that more people will use that firm, resulting in a higher load factor.

In addition to these three variables, a dummy variable for strikes was added. Operations during strikes are substantially different from nonstrike operations: percent capacity utilized often remains reasonable, but the capacity utilized drops by at least 50%.

4.2. Data

The data for these variables were obtained from a highly proprietary transportation database. One segment of the transportation industry was chosen, with data collected on all of the firms competing in that particular segment. As a result, ten firms were selected. Each firm consists of quarterly observations between 1973 and 1987, resulting in 60 observations per firm, resulting in a total of 600 observations. Some of the firms have undergone mergers. Mergers were dealt with by dropping the observations from the point at which the merger occurred. Consequently, 26 observations were omitted, leaving a total of 573. Out of this 573, 73 were randomly selected to form the test sample. Both the neural net and the regression were empirically tested using the test sample, in order to increase the comparability between them. The remaining 500 observations formed the training sample, used to train the neural net and estimate the regression parameters.

4.3. Regression Model

The regression was estimated using LIMDEP, a software package developed at New York University, capable of estimating many types of regression models. It can generate coefficients for both fixed effects and random effects models. Four regressions

were specified: an additive model and a multiplicative model, with fixed and random effects specified for each of them. Capacity utilized, average distance, advertising and strikes were used to predict percent capacity utilized.

4.4.   Neural Net Model

The neural net was trained using NeuralWorks Explorer, by NeuralWare, Inc. This package allows the user to specify the number of neurons, the number of layers, the training algorithm, and the sequence in which the data is selected to train the net. The four input neurons consist of capacity utilized, average distance, advertising, and strikes, the same variables as the independent variables in the regression. The one output neuron is percent capacity utilized. The single hidden layer consists of ten neurons (See Figure 1). Ten neurons were chosen to determine if the neural net could distinguish each of the ten firms, picking up a fixed effect in the data. Each of the neurons in the input layer was connected to the ten neurons in the hidden layer, and to the output neuron. The ten neurons in the hidden layer were connected to the single output neuron. The rationale for this connection design is to investigate whether the neurons in the hidden layer are capable of picking up any fixed effects in the data. No feedback connections were built in the neural net. The net was trained using the cumulative delta rule.

5.   Results

This section describes the results of the study. The results of the regression are presented first, followed by the neural net results. Table 2 summarizes the performance of the regression and the neural net.

In general, ignoring the additive/multiplicative specification for the moment, the fixed effects specifications (R^2 = 31.9%, 23.8%) performed substantially better than the random effects specifications (R^2 = -25.0%, -36.6%). These estimates of R^2 were obtained empirically from the test sample consisting of 73 observations. The R^2 being less than zero indicates that the random effects specification accounts for none of the variance in the dependent variable; it would be more accurate simply to predict the average of the dependent variable's observations. It is possible to obtain an R^2 less than zero because the R^2 was calculated empirically using the test sample. As a result, zero is no longer the lower bound, as it would be when calculating the R^2 from the training sample which was used to estimate the regression parameters. The R^2 was calculated empirically to increase the comparability between the regressions and the neural net.

The fixed effects specification is chosen as the correct one on the following basis. The ten firms which were selected form an exhaustive set of the firms competing in that particular market. In other words, the population is being sampled 100%. Therefore, a fixed effects model is appropriate for the regression, because the firms were not chosen at random (Hays, 1988).

In addition to choosing between a fixed effects specification and a random effects specification, the choice must also be made whether to choose an additive or a multiplicative model. To help in choosing between them, a j-test is performed [5] as follows. First, the additive model is estimated. The percent capacity utilized values predicted from this model are used as an independent variable in the multiplicative model. If this term is significant, the data is then displaying additive effects. Next, the multiplicative model is estimated. The values predicted from this model are used as an

independent variable in the additive model. If this term is significant, then the data is displaying multiplicative effects. It is possible for none, one, or both of these terms to be significant. In this particular dataset, both the additive and multiplicative terms are significant. This implies that the data has both additive and multiplicative components. The j-test is inconclusive as to which specification is superior. Therefore, the resulting equations predicting percent capacity utilized (PCU) are:

$$
\begin{aligned}
\text{PCU} = \quad & 0.0025*\text{CU} + 0.013*\text{Avg\_Dist} - 0.00026*\text{Adv} \\
& + 6.25*\text{Strike} + 37.58*\text{Firm1} + 45.48*\text{Firm2} \\
& + 38.52*\text{Firm3} + 42.98*\text{Firm4} + 35.20*\text{Firm5} \\
& + 49.13*\text{Firm6} + 41.22*\text{Firm7} + 33.87*\text{Firm8} \\
& + 51.15*\text{Firm9} + 47.13*\text{Firm10}
\end{aligned}
$$

$t_{CU}, t_{Avg\_Dist}, t_{Adv} < 0.00001$
$t_{Strike} < 0.05$

$$
\begin{aligned}
\text{PCU} = \quad & 0.19*\ln(\text{CU}) - 0.10*\ln(\text{Avg\_Dist}) \\
& - 0.024*\ln(\text{Adv}) + 0.19*\ln(\text{Strike}) \\
& + 3.35*\text{Firm1} + 3.43*\text{Firm2} + 3.24*\text{Firm3} \\
& + 3.33*\text{Firm4} + 3.30*\text{Firm5} + 3.48*\text{Firm6} \\
& + 3.39*\text{Firm7} + 3.30*\text{Firm8} + 3.46*\text{Firm 9} \\
& + 3.49*\text{Firm10}
\end{aligned}
$$

$t_{\ln(CU)}, t_{\ln(Strike)} < 0.001$
$t_{\ln(Avg\_Dist)} < 0.01, t_{\ln(Adv)} < 0.05$

The results of these models are consistent with the earlier models. Capacity utilized (CU) is positively related to percent capacity utilized, having coefficients of 0.0025 and 0.19. Average distance is positively related to percent capacity utilized, confirming the results obtained by the later studies.

These regression models are compared to the neural net on the basis of R-Square. R-Square is calculated, in general, as $R^2 = 1 - \text{var } e_t / \text{var } dv_t$ [10]. The R^2 of the neural net is calculated to be 19%. The neural net performs better than the random effects

specifications, but poorer than the fixed effects ones. The fixed effects regression may outperform the neural net because the additional knowledge associated with each firm is explicitly incorporated into the regression. Future research could examine the possibility of storing explicit knowledge in a neural net.

In addition to examining overall performance, the net was examined in more detail (See Figure 2). The weights connecting the input neurons to the output neuron are larger than the weights connecting the hidden layer neurons to the output neuron. In addition, the variability of the weights connecting the input neurons to the output neuron is substantially larger than the variability of the weights connecting the hidden layer neurons to the output neuron. It appears that the input neurons have the greatest effect in predicting the output neuron, with the hidden layer neurons modifying the output to a small degree. It is possible that the input-output connection is capturing the random effects of the model, and that the input-hidden-output connection is capturing the fixed effects of the model. Neural nets seem to have the potential to incorporate both a random effects component and a fixed effects component into a single model, unlike regression, which requires that one specification or the other is chosen.

6.   Conclusion

In summary, it appears that this transportation dataset provides an example in which a neural net model outperforms a random effects specification, but not a fixed effects one. A neural net may be more applicable in modeling data consisting of random effects than data consisting of fixed effects. This should be investigated further to determine if this result can be generalized. In addition, further research could be

conducted to determine if it is possible for a neural net to legitimately incorporate fixed effects in the hidden layer, with the random effects corresponding to the connection between the input layer and the output layer.

Table 1: Regression Coefficients

α	0.045	d_{Firm1}	-0.024
β_{OP}	0.377	d_{Firm2}	0.297
β_{AD}	0.028	d_{Firm3}	0.394
β_{PCU}	0.241	d_{Firm4}	-0.195
d_{1972}	-0.049	d_{Firm5}	0.409
d_{1973}	-0.068	d_{Firm6}	-0.130
d_{1974}	-0.061	d_{Firm7}	-0.123
d_{1975}	-0.048	d_{Firm8}	0.465
d_{1976}	-0.023	d_{Firm9}	0.283
		d_{Firm10}	-0.135

Table 2: Performance Comparison (R^2)

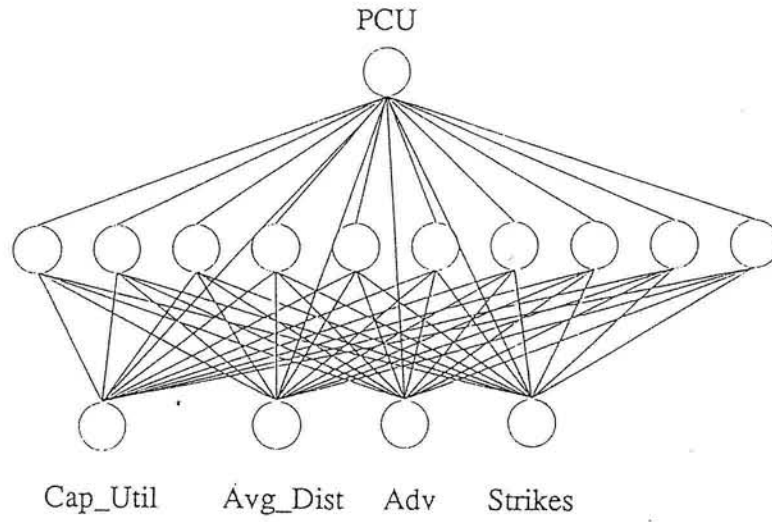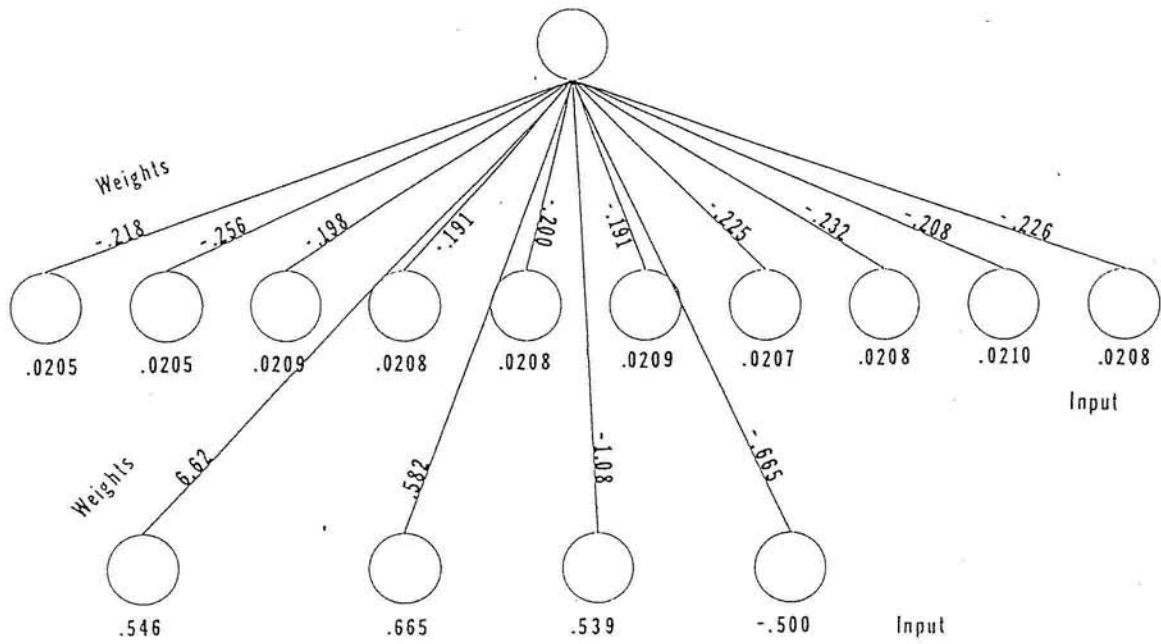| Specification	Additive	Multiplicative
Fixed Effects	31.9%	23.8%
Random Effects	-25.0%	-36.6%

Neural Net -- 19.0%

Figure 1: Neural Net Design

Figure 2: Neural Net

References

[1] E. Bailey and D. Kaplan, *Deregulating the Airlines*, MIT Press, Cambridge, Mass., 1985.

[2] T. Cowing and R. Stevenson, *Productivity Measurement in Regulated Industries*, Academic Press, New York, 1981.

[3] G. Douglas and J. Miller, *Economic Regulation of Domestic Air Transport: Theory and Policy*, The Brookings Institution, Washington, D.C., 1974.

[4] S. Dutta and S. Shekhar, "Bond Rating: A Non-Conservative Application of Neural Networks," *ICNN 2*, 1988, pp. 443-450.

[5] W. Greene, *Econometric Analysis*, MacMillan, New York, 1990.

[6] W. L. Hays, *Statistics*, Holt, Rinehart and Winston, Inc., New York, 1988.

[7] G. E. Hinton, "Connectionist Learning Procedures," *Artificial Intelligence*, September, 1989, pp. 185-234.

[8] R. P. Lippmann, "An Introduction to Computing with Neural Nets," *IEEE \SSP Magazine*, April, 1987, pp. 4-22.

[9] D. E. Rumelhart, G. E. Hinton, and J. L. McClelland, "A General Framework for Parallel Distributed Processing," in D. E. Rumelhart and J. L. McClelland (eds.) *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, Volume 1, MIT Press, Cambridge, Mass., 1986, pp. 45-76.

[10] T. J. Senjowski and C. Rosenberg, "NETtalk: a parallel network that learns to read aloud," *Technical Report 86/01*, Dept. of Electrical Engineering and Computer Science, John Hopkins University, 1986.

[11] H. White, "Economic Prediction Using Neural Networks: The Case of IBM Daily Stock Returns," *ICNN 2*, 1988, pp. 451-458.