Faculteit der Economische Wetenschappen en Econometrie

Serie Research Memoranda

Spatial Model Patterns in European Freight Transport Networks: Result of Neurocomputing and Logit Models

Aura Reggiani
Peter Nijkamp
Lucia Nobilio

Research Memorandum 1997-29

vrije Universiteit amsterdam

Spatial Modal Patterns In European Freight Transport Networks: Results of Neurocomputing and Logit Models

Aura Reggiani
Peter Nijkamp
Lucia Nobilio

Research Memorandum 1997 - 29

SPATIAL MODAL PATTERNS IN EUROPEAN FREIGHT TRANSPORT NETWORKS: RESULTS OF NEUROCOMPUTING AND LOGIT MODELS

Aura Reggiani*, Peter Nijkamp** and Lucia Nobilio*

* Department of Economics, Universith di Bologna, Piazza Scaravilli, 2
40 126 Bologna, Italy - e-mail: areggiani@ecn0 1 .economia.unibo.it

** Department of Regional Economics, Free University, De Boelelaan 1105
108 1 HV - Amsterdam, The Netherlands - e-mail: pnijkamp@econ.vu.nl

Abstract

The present paper aims to analyse interregional freight transport movements in Europe with a view on forecasting future patterns of transport flows using simple economic **scenarios.**

In view of the high dimension of our data-base on transport flows, two different approaches are compared, viz. the logit model and the neural network model. Logit models are well-known in the literature; however, applications of logit analysis to large samples are more rare. Neural networks are nowadays receiving a considerable attention as a new approach that is able to capture major patterns of flows, on the basis of fuzzy and incomplete information. In this context an assessment of this method on the basis of a large amount of data is an interesting research endeavour.

The paper will essentially deal with a European research experiment, oriented towards both calibration/learning procedures and spatial forecasting, in order to compare the two above methodologies as well as to investigate the potential/limitations of the two above mentioned intrinsically different, but nevertheless related assessment methods. The application field is the assessment of European (mainly Transalpine) modal freight flows. The first results in this framework highlight the fact that the two models adopted, although methodologically of a different nature, are both able to provide a reasonable spatial mapping of the interregional transport flows under consideration.

1. Recent Trends in the European Freight Transport Developments

Europe is nowadays in a state of flux. After the completion of the European market and with the widening of Europe towards easterly direction, mobility in general has shown a steady increase in Europe. In particular, cross-border transport has been at a rising edge with annual growth rates exceeding 10 percent, a process reinforced by the current globalisation trends. The integration of former segmented markets - and the related liberalisation in the European space - has led to drastic changes in both goods and passenger transport all over Europe.

In recent years, the European Commission has recognised this profound restructuring phenomenon, an observation which can also be found in the Maastricht Treaty. European networks are seen as the backbone of integration forces, while changes (i. e. improvements) in the morphology of the networks are expected to generate (positive) system-wide impacts. Clearly, the emphasis on the potential of these networks for competitiveness and cohesion provokes various questions on the relative efficiency and substitutability of the different modes of this network. This issue is particularly important, as the competition between different modes and the social acceptability of modal choices are not only determined by the direct operational costs, but also by environmental externalities. An interesting new situation is created by Swiss initiatives to reduce Transalpine trucking through limitations on load volumes and the supply of new tunnel infrastructures.

It is thus no surprise that there is an increasing interest in the issue of intermodal competition and complementarity. For surface transport in Europe, especially the competitive position of rail vis-a-vis road is at stake. This holds increasingly also for commodity transport. It needs to be added however, that the analysis of freight transport in Europe is fraught with many difficulties, as freight is not a homogeneous commodity, but is composed of an extremely diversified set of goods with specific haulage requirements and logistic needs. This means that a commodity-specific approach is necessary to analyse in depth implications of changes in network configurations. This approach will also be adopted in the present paper, with a particular view on Transalpine movements.

Therefore, the aim of the present paper **is** to investigate freight flow patterns in Europe from a multiregional perspective, by looking into the modal choice for these goods from the viewpoint of freight costs and transport time. In this paper, two competing models, viz. a discrete choice model and a neural network model, wil be employed to map out the spatial flow patterns in an explanatory context. This offers also a possibility to compare the relative performance of those models. A selection of Italian/Greek regions (in order to highlight the Alpine crossing movements) will be used **to** test the predictive power of the models concerned. Next, a sensitivity analysis will be carried out in order to investigate the expected consequences of a rise in transport costs (e.g., as a consequence of a European environmental tax on freight costs) as well in transport time. The analysis will also illustrate the need for a reliable and up-to-date European data base on European freight flows.

2. Models for Spatial Choice Analysis

Market requirements, cost considerations, the morphology of the European transport network and capacity limits lead to complicated modal choices. The present paper aims to analyse inter-regional freight transport movements in Europe as well as to forecast resulting spatio-temporal flow patterns on the basis of new transport-economic scenarios. For this purpose, a modal split analysis will be carried out by means of two statistical models, namely the logit model and the neural network model (see also Reggiani *et al.,* 1997). A binary logit model will be discussed in Section 2.1, while a feedforward neural network model will be presented in Section 2.2. Our empirical analysis - illustrated in Section 3 - will use both types of models.

2.1 The Logit Model

A first approach to modelling freight transport flows is to use an explanatory, behaviourally-based choice model in an origin-destination setting. In this context, widely adopted approach for modal split analysis is the logit model (see e.g. Ben-Akiva and Lerman, 1985). Recent experiments using logit models / spatial interaction models in order to map out the freight transport in Europe have been carried out by Tavasszy (1996), who showed the suitability of logit models also for the goods transport sector (where data are more ‘fuzzy’ and incomplete compared to the passenger sector). Logit models are discrete choice models, which are used for modeling a choice from a set of mutually exclusive and exhaustive alternatives. It is assumed that the decision-maker chooses the alternative with the highest (stocastic) utility among the set of alternatives. The utility of an alternative is determined by a utility function, which consists of independent attributes of the alternative concerned and the relevant parameters. By maximizing the stochastic utility, the highest choice probability is then the probability that the alternative has the highest utility among all relevant alternatives (see e.g. Ben-Akiva and Lerman, 1985, Cramer, 1991, and McFadden, 1977). Since in our case two discrete choices - rail and road - will be considered, a binary logit model is adopted. Our binary logit choice model for modal split choice of a transport mode m (m = 1, 2) vis-a-vis the complementary transport mode between two regions i and j has the following formulation:

$$P_{ij}^{m} = \frac{\exp(U_{ij}^{m})}{\sum_{m} \exp(U_{ij}^{m})} \tag{1}$$

where

$$U_{ij}^{m} = \sum \beta_{z} X_{z.i}^{m}; \qquad z=1,...n \tag{2}$$

and where:

- m = the mode of transport (m = train or truck);
- P_{ij}^{m} = the probability of choosing the mode m from region i to region j ($i \neq j$);
- U_{ij}^{m} = the utility connected with the rail mode m on the link ij;
- $X_{z,ij}^{m}$ = the vector of attributes for mode *m* in the utility function for the link ij (in our case cost and time);
- β_z = the vector of parameters related to the vector of attributes time and cost.

The binary logit model has become in the past decades a standard analytical tool in discrete choice modelling. The results of this logit model for some empirical cases on European freight transport will be given in Section 3. In particular, given the large amount of data concerning the road mode, the calibration and forecasting analysis related to model (1) and (2) will be illustrated - in the following Section 3 - with reference to the latter mode. In the same vein also some policy scenario experiments will be presented later on.

2.2 The Neural Network Approach

In recent years, a great deal of interest has arisen in neuro-computing, in particular neural network analysis (see, e.g., Anderson and Rosenfeld, 1990, and Rumelhart and McClelland, 1986). Neural network (NN) analysis has in recent years become a popular analysis tool (see, for a review, Himanen *et al.,* 1997). NNs replicates human brain functions and are thus considered as 'intelligent', since they learn and generalize by examples (see, e.g., Reggiani *et al.,* 1997). NNs have been widely applied to the area of transport engineering, in particular in relation to traffic control problems and accidents (see Himanen *et al.,* 1997). However, so far only a few experiments exist in the field of transport economics or transport route / mode / destination choice (see, e.g. Nijkamp *et al.,* 1996, and Schintler and Olurotimi, 1997). Our experiments aim to explore also this novel research direction by comparing NN results with those of a logit approach.

In a way analogus to most applications of NNs, in this study a two-layer feedforward, totally connected NN will be used in order to analyse the freight transport modal split problem in Europe. The methodological structure of the main steps related to the application of a feedforward NN is described in Reggiani and Tritapepe (1997), wich consists of three stages: i) definition of network architecture; ii) learning phase; iii) forecasting phase. It is necessary to define the right architecture of the network, i.e. the number of units on the relevant levels. Usually, the input and output units depend on the number of input and output variables which define the problem. In our application one possible NN architecture contains 4 input units which correspond to the attributes time and cost related to each transport mode (rail and road) and one output unit corresponding to the probability of choosing one mode[1] (e.g., the rail mode). In the past years we have witnessed an increasing acceptance of NN models in social science research, including transportation science. Section 3 will offer empirical results obtained by applying an NN model to European freight flow data, while scenario results will be presented in Section 4.

3. Empirical Results from European Freight Flow Analysis

After the concise presentation of the methodology, in this section the experiments with the logit and the NN approach (see Subsection 2.1 and 2.2) will be presented and discussed. In Subsection 3.1 a concise description of the data set will be given. The experiments carried out by means of the logit approach and the NN approach will be presented in Subsections 3.2 and 3.3, respectively. Then the two approaches will be mutually compared in Subsection 3.4.

3.1 Description of Available Data

Comprehensive detailed data on European freight flows are rare. The data set[2] contains the freight flows and the attributes related to links between 108 European regions for the year 1986. The attributes considered here are '*time*' and '*cost*' between each link (ij) for each transport mode, so that the data set pertains to variables related to each link (ij). Furthermore, the flow distribution in the matrices concerned refers to one particular kind of goods, viz. food. Clearly, other sectors might have been chosen as well.

After screening and elimination our data set contains finally 4,409 observations on interregional freight flows in Europe. This data set is next randomly subdivided into three sub-sets:

- a *training set* containing 2,992 observations, i.e. about 68% of the data-set;
- a *cross-validation set* containing 447 observations, i.e. about 10% of the data-set;
- a *test set* containing 970 observations, i.e. about 22% of the data-set.

[1] The choice probability of the other mode is just the complement.
[2] The data set has been kindly provided by NEA Transport Research and Training, Rijswijk. (see, for a detailed map of the European regions, Reggiani et *al.*,1997).

On the basis of these data we will now carry out our experiments.

3.2 Applications Based on a Logit Approach

In our approach a binary logit model is used to analyse the modal split problem between road and rail in relation to inter-regional food transportation between 108 regions in Europe. In Subsection 3.2.1, the calibration results and an evaluation of the logit model will be presented. Then, results from a spatial forecasting of the calibrated logit model will be presented and evaluated in Subsection 3.2.2.

3.2. I Estimation of the binary logit model

In our empirical application, the logit model is estimated in order to assess the unknown parameters in the utility function. For this purpose, the learning data set, which is the training set combined with the cross-validation set, has been used. Concerning the logit model structure, two distinct cases are considered; in case A only the **cost** attribute is used for estimating the parameters, while **in** case B **cost** and **time** are considered as attributes. In particular, the logit model has been calibrated by using the LIMDEP software. The estimated parameters resulting from the calibration stage are presented in Tables 1 and 2 for Case A and Case B[3], respectively.

Next, the goodness-of-fit of the model has been evaluated using two statistical indicators: the likelihood-ratio (ρ^2) and the t-test. The related results are also presented in Tables 1 and 2.

Table 1 about here

Table 2 about here

The t-test indicates that the two parameters are significantly different from zero in both cases (see again Table 1 and Table 2). Also the value of ρ^2 indicates that the calibrated logit models are performing reasonably well for the two cases. However, the calculated ρ^2 for Case B

[3] We will then denote as Case A (L), Case B (L) the respective cases A and B implemented by logit models, while Case A (NN) and Case B (NN) will indicate the respective cases A and B estimated by NN models.

(L) is better than that for Case A (L), which suggests that Case B (L) - with inclusion of more attributes in the utility function - performs better than Case A (L).

3.2.2 Statistical performance of the binary logit model

Next the binary logit model, estimated in the previous subsection, is used to make freight transport forecasts on the basis of various transport economic scenarios. For this predictive purpose, both the data set used in the calibration stage and the test set which is not used in the calibration stage, are employed.

In our analysis of the spatial forecasting performance of the binary logit model, the statistical indicators ρ^2, ARV, R^2, MSE, RMSE, EPMA have been adopted. The definition of these indicators can be found in Annex 1. These indicators have been used - individually or jointly considered - for examining the statistical/econometric merits by varying the combination of attributes in the utility function. In particular, they have been calculated for both the calibration and the test set. They will also be used subsequently to explore the performance of NN models. The probabilities of train and truck have been used in calculating the statistical indicators. The results are presented in Tables 3 and 4 for case A (L) and Tables 5 and 6 for case B (L), successively.

Table 3 about here

Table 4 about here

Table 5 about here

Table 6 about here

Concerning the 'optimal' values of the above indicators it should be noted that the ARV, MSE and RMSE measures should ideally approach zero; EPMA suggest an extremely good forecast when his value is less than 10%, and a good forecast when his value is ranging from 10% to 20% (see Annex 1), while the R^2 measure should approach one. Regarding the general results presented in Tables 3-6, the binary logit model appears to have an insufficient predictive ability, when only one attribute is considered (see, e.g., case A (L) for both the calibration and the test set in Tables 3 and 4, respectively). On the contrary, the logit model performs quite better for case B (L) - related to the two attributes (cost and time) - for both the calibration and the test set (see, e.g., again the values of ARV and R^2 in Tables 5 and 6). Consequently, the parameters emerging from Case B (L) will be employed for the forecasting analysis carried out - on the basis of policy scenarios - in the next Section 4.

3.3 Exneriments bv means of a Neural Network Approach

After the applications of the logit model, now the modal split problem will be analysed by means of a more recently developed statistical model, viz., the feedforward NN model (see Subsection 2.2).

It has already been mentioned that the whole data set contains 4,409 observations (examples or patterns). The following general considerations apply to the experiment undertaken here:

- Both Case A and Case B - analogously to logit analysis - are trained (they have been named Case A (NN) and Case B (NN), respectively; see footnote 3).
- The training for the neural net model (and the calibration for the logit model) has been carried out by using *the training set.*
- The performance measure has been evaluated by using the *test set* (spatial forecasting).

Concerning the number of hidden units, they have empirically been defined by taking into account the number of observations in the data set as well as by carrying out a large number of experiments. In regard to the parameters defining the neural architecture, they have been determined after several empirical experiments. Finally, the parameters of the NNs are set as follows:

- number of hidden units: 6
- learning rate a = 0.9
- momentum factor 1 = 0
- training tolerance = 0.1
- initial weight values: randomly between [-0. l;O. 1]

It should be noted that by using a feedforward NN it is necessary to cope with the over-fitting problem. Consequently, in the experiments the cross-validating technique (by using the cross-validation subset) has been used in order to avoid such a problem (for details on the overfitting problem and the cross-validating technique, see e.g. Fischer and Gopal, 1994, and Reggiani and Tritapepe, 1997).

The results related to the above mentioned experiment will now be presented. In general, by using a statistical model for forecasting, the first step is to evaluate the predictive quality of the model, i.e. to determine how well the model learned to approximate the unknown input-output function for arbitrary values of input units, while the final aim of our work is to evaluate the freight transport movements in Europe in order to forecast spatio-temporal patterns **on** the basis of new transport economic scenarios. The present section will particularly analyse this first research stage, i.e. the spatial forecasting of the model adopted. The predictive quality will be evaluated **-** by means of several performance measures **-** by using the test set which had been set apart and not yet used for the calibration (learning) phase, as mentioned above.

The predictive performance of an NN can be judged by means several statistical indicators like ARV, R2, MSE, RMSE, EPMA, as previously indicated for binary logit model. The results of the statistics are displayed in the following Tables 7 and 8.

Table 7 about here

Table 8 about here

It is evident from the above tables that the ARV and R^2 indicators, emerging from NNs, give a better result than the ARV and R^2 indicators emerging from the logit analysis, while the MSE, RMSE and EPMA values are slightly better for logit models with respect to NNs.

3.4 Comparison of the Logit and Neural Network Approach

After the presentation of the above results, we will, in this subsection, compare and evaluate the spatial forecasting performance of the two alternative approaches. First, the values of goodness-of-fit indicators for both the two models and for each case A and B are compared - with reference to the test set - by means of histograms in Figures l-5.

Figure 1 about here

Figure 2 about here

Figure 3 about here

Figure 4 about here

Figure 5 about here

It is noteworthy that Figures l-5 show in particular the higher performance of NNs - with respect to logit models - in Case A (adoption of only one attribute in the utility function). This results confirms previous findings in the framework of different data bases (see Nijkamp *et al*, 1996), underlining the goodness-of-fit of NNs in the presence of uncertain and incomplete data (viz., by using only one attribute). Concerning Case B, dealing with the assumption of two attributes in the utility function, both approaches display good values to the assessment of freight flows.

Finally, in order to better evaluate the performance of logit and NN models, an extrapolation of estimated data againist the real data will be shown (see Tables 9-12).

In our empirical investigation, we have focused our attention in particular, on the Transalpine area. Considering that the Alpine chain separates more or less Europe from Geece and Italy, an extrapolation of these data has been carried out (see Tables 9 and 10). More precisely, Table 9 illustrates the estimated/real flows for the outflows, from Greece and Italy, towards Northen and Western Europe (without considering Spain and Portugal), while Table 10 displays the estimated values for the inflows from Europe to Italy and Greece. It should be noted that the values illustrated in Tables 9-10 emerge from an estimation process carried out on a data set constituted by the sum of the learning and test set for both the two approaches (NNs and logit models).

Table 9 about here

Table 10 about here

It is evident from Tables 10 and 11 that a NN model performs overall slightly better than the logit model (see the values of the relative prediction error for each link as well as the related mean values). Surprisingly, despite a weaker behavioural basis in NN models, the results seems to be fairly precise.

4. Policy Scenario Experiments

The above estimation procedures were based on time and financial sacrifices in the European transport sector. It is noteworthy however, that freight transport causes also high social costs (environmental pollution, accidents, etc.), which might be charged to the transportation sector. We will now investigate the consequences of varying the transportation time/costs for freight flows by including some level of social costs. A sensitivity analysis of the previous results based on some economic scenarios will now be carried out in this section by using again both the binary logit model and the NN model. Two policy scenarios based on different external time/cost assignments will be used; they will concisely be discussed here. Later on, we will present the results related to the sensitivity analysis for the logit and the NN approach.

Nowadays, because of severe problems on the road transport network (for example, pollution, congestion), governments are trying to reduce the road usage by imposing policy measures that serve to increase the cost of road usage (see Verhoef, 1996) in order to induce a shift to other modes. An example of a Pigouvian policy for coping with environmental externalities is the recently increased tax on fuel in the Netherlands. In so doing, the usage of the road transport network is made less attractive than other transport networks, so that a modal/shift is encouraged.

Since it is very difficult to assess the social costs of freight transport in Europe, two scenarios are developed and considered for a sensitivity analysis on the transport costs. Generally we assume that a uniform European tax policy for freight transport is adopted and that the cost attribute related to the road mode is increased by 10 % for all links (ij). In Scenario 1 we assume - on the road mode - only an increase of transport time of 10%, due to congestion problems.

Scenario 2 is a slight variation of Scenario 1, since here it is assumed that each attribute (time/cost) is increased by 10%. In particular, Scenarios 1 and 2 have been implemented for the Alpine sector (see Tables 11 and 12).

The results from the above sensitivity analysis can be highlighted by the relative prediction error, which is defined as the (relative) difference between the predicted flow and the real flow as a percentage of the real flow. In this context, the results in our tables indicate that the binary logit model is relatively more sensitive to changes in the time/cost attribute than the NN model.

It is interesting to note that in the neural network case, and particularly in the case of inflows from Europe to Greece/Italy, the model shows -in the mean value- a slight increase of flows, despite the time/cost increase. This result may be plausible by taking into account the increasing amount of interaction among regional flows as a result of increased efficiency. It would certainly be relevant to compare these results with more updated data in order to better evaluate the 'forecasting' analysis of the two models, in particular since we have used -as a starting point- a set related to the year 1986.

In general, however, the above results may be considered to be plausible, in the absence of updated data that would be able to test our hypothesis on an increase in the time or cost indicator, given the good performance from the calibration / test phase. Moreover, these results may offer a 'range of plausible values' to policy actors who aim to evaluate the impact of time or cost changes on freight flows, given the intrinsic limits of both adopted models for freight transport analysis.

It is noteworthy that the large amount of freight flow data at an aggregate level hampers a behavioural micro perspective inherent in logit models. A further limitation consist of the type of architecture adopted in NN models, which seems critical for the validity of the results. Consequently, the results of our model may be used as a benchmark for the results of other models (for example, genetic algoritms), by offering a more 'flexible'output to policy actors.

Table 11 about here

Table 12 about here

5. Concluding Remarks

The European integration, the regulation policy and new infrastructure policies call for a proper insight into European freight flow developments. This paper has aimed to depict transport flows of commodities in an inter-regional European setting. Based on an extensive data set, various estimates of the impacts of costs on transport movements have been made. The test results show that both the logit and the NN approach are giving fairly plausible results. In general, NN models seem to perform slightly better.

After this exploratory comparative study of two modelling approaches, it is certainly opportune to investigate more thoroughly the differences in backgrounds of these two research paradigms. It is well known that the logit model is a particular spatial interaction model that has its roots in social behaviour of actors, albeit with some limitations like the well known IIA (Independence from Irrelevant Alternatives) assumption. The NN model is based on similarily of learning experiments and has certainly a behavioural adjustment potential, but is less easily interpretable from social science motives, even though recent results show a compatibility between feedforward NNs and binary logit models (see Schintler and Olurotimi, 1997), feedforward NNs and spatial interaction models (see Fischer and Gopal, 1994) and feedforward NNs and logistic regression models (see Schumacher et al., 1996). Given its predictive ability, more research is needed to better investigate the behavioural roots of NN models, while also extentions towards genetic algorithms may be explored.

Acknowledgments

The first author gratefully acknoledges the Italian CNR Project PFT2 n°96.00098.PF74 as well as the MPI project 60%.

References

Anderson, J. A. and E. Rosenfeld (eds.), **Neurocomputing: Foundations and Research,** MIT Press, Cambridge, Massachusetts, 1990.

Ben-Akiva, M. and S. R. Lerman, **Discrete Choice Analysis: Theory and Application to Travel Demand,** MIT Press, Cambridge, Massachusetts, 1985.

Cramer, J. S., **The LOGIT Model: an Introduction for Economists,** Routledge, Chapman and Hall Inc., New York, 1991.

Fischer, M. M. and S. Gopal, Artificial Neural Networks: A New Approach to Modelling Interregional Telecommumicanon Flows, **Journal of Regional Science,** Vol. 34, N. 4, 1994, pp. 503-527.

Himanen, V., Nijkamp, P. and A. Reggiani (eds.), **Neural Networks in Transport,** Avebury, Aldershot, 1997 (forthcoming).

McFadden, D., Econometric Models of Probabilistic Choice, in Mansky, C. F. and D. McFadden (eds), **Structural Analysis of Discrete Data with Econometric Applications,** MIT Press, Cambridge, Massachusetts, 1977.

Nijkamp, P., Reggiani, A. and T. Tritapepe, Modelling Inter-Urban Transport Flows in Italy: A Comparison between Neural Network Approach and Logit Analysis, **Transportation Research** C, vol. 4, 1996, pp. 323 - 338.

Reggiani, A. and T. Tritapepe, Neural Networks and Logit Models to Commuters' Mobility in the Metropolitan Area of Milano, in Himanen, V., Nijkamp, P. and A. Reggiani (eds.), **Neural Networks in Transport,** Avebury, Aldershot, 1997 (forthcoming).

Reggiani A., P. Nijkamp and W.-F. Tsang, European Freight Transport Analysis Using Neural Networks and Logit Models, **in** A. Reggiani (ed.), **Accessibility, Trade and Location Behaviour,** Avebury, Aldershot, 1997 (forthcoming).

Reggiani, A., R. Romanelli, T. Tritapepe and P. Nijkamp, Neural Networks: An Overview and Applications in the Space Economy, in Himanen, V., Nijkamp, P. and A. Reggiani (eds.), **Neural Networks in Transport,** Avebury, Aldershot, 1997 (forthcoming).

Rumelhart, D. E. and J. L. McClelland (eds.), Parallel Distributed Processing, MIT Press, Cambridge, Massachusetts, 1986.

Scardovi, I. and P. Monari, **Variabilità Statistica Bidimensionale,** CLUEB, Bologna, 1988.

Schintler, L. A. and 0. Olurotimi, Neural Networks as Adaptive Logit Models, in Himanen, V., Nijkamp, P. and A. Reggiani (eds.), **Neural Networks in Transport,** Avebury, Aldershot, 1997 (forthcoming).

Schumacher, M., Roßner, R. and W. Vach, Neural Networks and Logistic Regression, **Computational Statistics & Data Analysis,** Vol. 21, 1996, pp. 661-682.

Tavasszy, L. A., **Modelling European Freight Transport Flows,** Ph.D. Thesis, Delft University of Technology, Delft, 1996.

Verhoef, E. T., **The Economics of Regulating Road Transport,** Edward Elgar, Aldershot, 1996.

Wegener, M. and E. Graf, Estimating Multiattribute Spatial Choice Models, User's Manual, University of Dortmund, 1982.

Figures and Tables

Table 1: Results for case A (L) related to the logit model

Case A

Number of observations: 3,349

Attributes: cost

Variables	Coefficient	St. Error	t-ratio	ρ^2
cost	-0.091482	0.1308E-04	-699.40	0.6538

Table 2: Results for case B (L) related to the logit model

Case B

Number of observations: 3,349

Attributes: cost, time

Variables	Coefficient	St. Error	t-ratio	ρ^2
cost	-0.064926	0.1535E-04	-4229.7 1	0.71023
time	-0.095 153	0.35 14E-05	-2707.83	

Table 3:Case A (L); n° of observations: 3,439

ARV	R^2	MSE	RMSE	EPMA
0.96	0.45	0.19	0.44	36%

Table 4:Test Case A (L); n° of observations: 970

ARV	R^2	MSE	RMSE	EPMA
0.91	0.48	0.19	0.43	35%

Table 5: Case B (L); n° of observations: 3,439

ARV	R^2	MSE	RMSE	EPMA
0.35	0.73	0.07	0.26	17%

Table 6: Test Case B (L); n° of observations: 970

ARV	R^2	MSE	RMSE	EPMA
0.20	0.83	0.04	0.20	15%

Table 7: The values of goodness-of-fit indicators for the NN model: Case A (NN)

Case A (NN)
N° of observations: 970 (test set)
Attributes: cost

ARV	R^2	MSE	RMSE	EPMA
0.16	0.94	0.07	0.26	20%

Table 8: The values of goodness-of-fit indicators for the NN model: Case B (NN)

Case B (NN)
N° of observations: 970 (test set)
Attributes: cost, time

ARV	R^2	MSE	RMSE	EPMA
0.17	0.92	0.07	0.16	18%

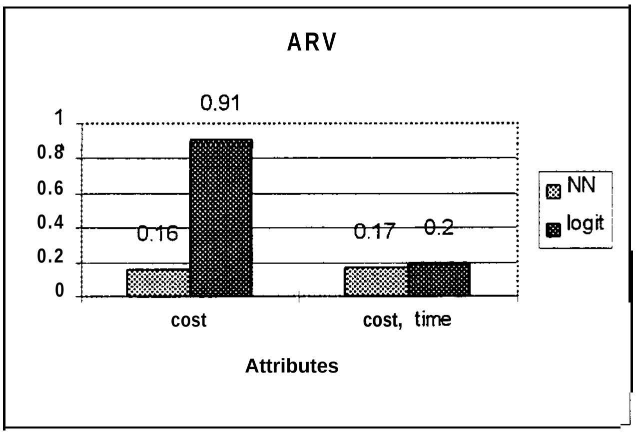

Figure 1: The value of the indicator ARV for cases A and B

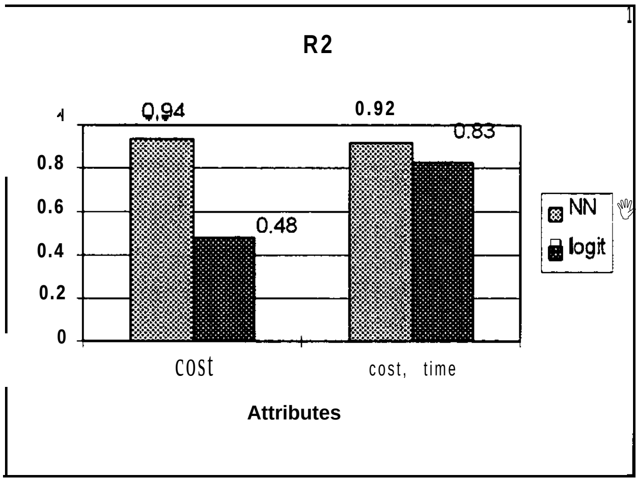

Figure 2: The value of the indicator R^2 for cases A and B

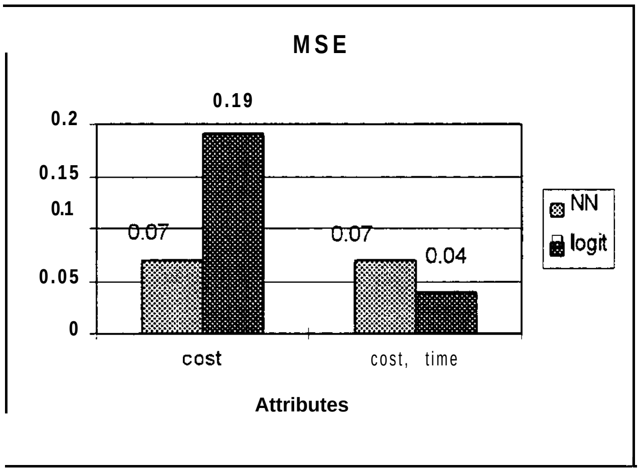

Figure 3: The value of the indicator MSE for cases A and B

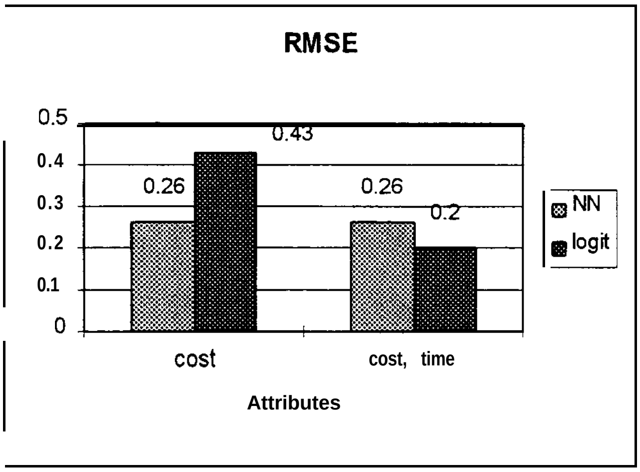

Figure 4: The value of the indicator RMSE for cases A and B

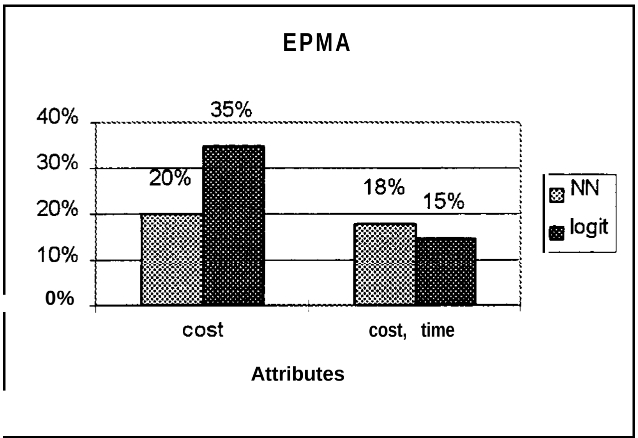

Figure 5: The value of the indicator EPMA for cases A and B

Table 9: Transalpine food transport flows by road fi-om Italian plus Greek regions to the rest of Europe' (flows in tons; year: 1986)

REGIONS	Transalpine real flows	Logit results	NN results	rel. pred. err. Logit (%)	rel. pred. err. NN (%)
Thessaloniki	19764	16122	19274	-18.43	-2.48
Athens	25965	29120	28287	12.15	8.94
Patras	22569	13082	22478	-42.04	-0.40
Heraklion	18622	17711	19104	-4.89	2.59
Turin	820281	724980	780102	-11.62	-4.90
Milan	3980845	3137979	3835833	-21.17	-3.64
Venice	922574	648524	864446	-29.70	-6.30
Bologna	7213650	5821584	6442638	-19.30	-10.69
Florence	1143048	1055770	1044481	-7.64	-8.62
Ancona	1035352	992532	931062	-4.14	-10.07
Pescara	683626	628053	624495	-8.13	-8.65
Rome	351976	313587	318727	-10.91	-9.45
Naples	1258182	1167890	1155432	-7.18	-8.17
Bari	2442992	2086402	2279118	-14.60	-6.71
Reggio C.	222407	211238	209597	-5.02	-5.76
Palermo	703347	614024	668120	-12.70	-5.01
Cagliari	48357	48196	46894	-0.33	-3.03
M*				-12.10	-4.84
MA**				13..53	6.20

1) in order to highlight the Alpine crossing movements, data related to Spanish and Portuguese regions have been eliminated
* M = mean value of the variations from the real data
** MA = mean value of the absolute variations from the real data

Table 10: Transalpine food transport flows by road from the rest of Europe' to Italian plus Greek regions (flows in tons; year: 1986)

REGIONS	Transalpine real flows	Logit results	NN results	rel. pred. err. Logit (%)	rel. pred. err. NN (%)
Thessaloniki	44380	38636	43297	-12.94	-2.44
Athens	52047	43557	51038	-16.31	-1.94
Patras	53626	46130	52145	-13.98	-2.76
Heraklion	56930	53420	56916	-6.17	-0.02
Turin	259075	379966	398615	46.66	53.86
Milan	414190	350049	432237	-15.49	4.36
Venice	53795	40932	56748	-23.91	5.49
Bologna	365183	355578	377438	-2.63	3.36
Florence	178632	157254	185540	-11.97	3.87
Ancona	43653	42143	43540	-3.46	-0.26
Pescara	119774	113282	115746	-5.42	-3.36
Rome	35705	31264	34076	-12.44	-4.56
Naples	183553	188948	194825	2.94	6.14
Bari	105824	93432	99806	-11.71	-5.69
Reggio C.	29960	29558	28841	-1.34	-3.73
Palermo	126464	114747	124608	-9.27	-1.47
Cagliari	64435	57372	64633	-10.96	0.31
M*				-6.38	3.01
MA**				12.21	6.10

Table 11: Results of a sensitivity analysis by means of scenarios (Transalpine food transport flows from Italian plus Greek regions to the rest of Europe')

REGIONS	Real Flows	Scenario 1 (time + 10%)		Scenario 2 ((time/cost)+ 10%)		Scenario 1 rel. pred. err.		Scenario 2 rel. pred. err.	
		LOGIT	NN	LOGIT	NN	Logit (%)	NN (%)	Logit (%)	NN (%)
Thessaloniki	19764	6649	19324	16283	19488	-66.36	-2.23	-17.61	-1.40
Athens	25965	6813	28431	26213	28501	-73.76	9.50	0.96	9.77
Patras	22569	8713	22101	16846	22155	-61.39	-2.07	-25.36	-1.83
Heraklion	18622	12340	19055	12086	19052	-33.73	2.33	-35.10	2.31
Turin	820281	774691	793751	400811	796122	-5.56	-3.23	-51.14	-2.95
Milan	3980845	3880482	3874579	1808444	3887546	-2.52	-2.67	-54.57	-2.34
Venice	922574	886303	830692	392430	831592	-3.93	-9.96	-57.46	-9.86
Bologna	7213650	7057064	6247481	3258554	6257957	-2.17	-13.39	-54.83	-13.25
Florence	1143048	1102063	1014918	559767	1011246	-3.59	-11.21	-51.03	-11.53
Ancona	1035352	991756	892805	518466	893153	-4.21	-13.77	-49.92	-13.73
Pescara	683626	642061	616435	339372	615614	-6.08	-9.83	-50.36	-9.95
Rome	351976	329765	304467	172024	301735	-6.31	-13.50	-51.13	-14.27
Naples	1258182	1190396	1135565	624124	1129908	-5.39	-9.75	-50.39	-10.20
Bari	2442992	2385679	2202194	1120275	2186937	-2.35	-9.86	-54.14	-10.48
Reggio C.	222407	205187	202594	113536	201097	-7.74	-8.91	-48.95	-9.58
Palermo	703347	567622	646903	389092	645495	-19.30	-8.03	-44.68	-8.23
Cagliari	48357	48921	45347	24706	45661	1.17	-6.22	-48.91	-5.58
M*						-17.84	-6.64	-43.80	-6.65
MA**						17.97	8.03	43.91	8.07

Table 12: Results of the sensitivity analysis by means of scenarios (Transalpine food transport flows from to the rest of Europe' to Italian plus Greek regions)

REGIONS	Real Flows	Scenario 1 (time + 10%)		Scenario 2 ((time/cost)+ 10%)		Scenario 1 rel. pred. err.		Scenario 2 rel. pred. err.	
		LOGIT	NN	LOGIT	NN	Logit (%)	NN (%)	Logit (%)	NN (%)
Thessaloniki	44380	22564	42904	29176	43115	-49.16	-3.33	-34.26	-2.85
Athens	52047	16050	50565	39341	50693	-69.16	-2.85	-24.41	-2.60
Patras	53626	25265	52340	35481	52603	-52.89	-2.40	-33.84	-1.91
Heraklion	56930	50317	57730	30332	57824	-11.62	1.41	-46.72	1.57
Turin	259075	343890	401170	220704	401352	32.74	54.85	-14.81	54.92
Milan	414190	386103	432723	208729	432999	-6.78	4.47	-49.61	4.54
Venice	53795	47947	56775	26790	56851	-10.87	5.54	-50.20	5.68
Bologna	365183	365225	378716	192796	379190	0.01	3.71	-47.21	3.84
Florence	178632	164387	181798	101032	182357	-7.97	1.77	-43.44	2.09
Ancona	43653	42401	43617	22771	43574	-2.87	-0.08	-47.84	-0.18
Pescara	119774	107433	116100	64002	116021	-10.30	-3.07	-46.56	-3.13
Rome	35705	30840	34466	18261	34464	-13.63	-3.47	-48.86	-3.48
Naples	183553	140320	197702	122312	197946	-23.55	7.71	-33.36	7.84
Bari	105824	99633	101972	52125	102277	-5.85	-3.64	-50.74	-3.35
Reggio C.	29960	21101	28566	20258	28608	-29.57	-4.65	-32.38	-4.51
Palermo	126464	90232	126808	75681	127036	-28.65	0.27	-40.16	0.45
Cagliari	64435	64478	64336	31155	64503	0.07	-0.15	-51.65	0.11
M*						-17.06	3.30	-40.94	3.47
MA**						20.92	6.08	40.94	6.06

ANNEX 1

The Ro Squared Coefficient (ρ^2)

The statistical indicator ρ^2 is defined as:

$$\rho^2 = 1 - (\lambda_{(0)} / \lambda_{(\beta)}) \tag{a.1}$$

where $\lambda_{(0)}$ = the value of the log likelihood function when all weights are zero and $\lambda_{(\beta)}$ = the value of the log likelihood function at its maximum (see Ben-Akiva and Lerman, 1985).

The Average Relative Variance (ARV)

The statistical indicator ARV is defined as:

$$ARV = \frac{\sum (y - \hat{y})^2}{\sum (y - \bar{y})^2} \tag{a.2}$$

where Y= the observed transport flow using car, $\hat{y}$ = the transport flow by truck, predicted by the adopted model, and $\bar{y}$ = the average of the observed transport flow by truck (see Fischer and Gopal, **1994).**

The Correlation Coefficient (R^2)

The statistical indicator R^2 is defined as:

$$R^2 = \frac{\sum (\bar{y} - \hat{y})^2}{\sum (y - \bar{y})^2} \tag{a.3}$$

where the variables are defined in equation (a.2)

The Mean Squared Error (MSE)

The MSE indicator is applied for all estimates and is indipendent of the underlying methodology. In fact it is specified as the squared difference between the observed values (y_i) and the predicted values $(\hat{y}_i)$ (see, e.g., Scardovi and Monari, 1988):

$$MSE = \frac{1}{n} \sum_{i=1}^{n} \left(y_i - \hat{y}_i \right)^2 \tag{a.4}$$

This indicator denotes a good performance of the calibrated model when its value is approaching zero.

The Root Mean Squared Error (RMSE)

This indicator is obtained as the root of the MSE indicator (see equation (a. 1)):

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2} \qquad (a.5)$$

The Mean Absolute Percentage Error (MAPE or EPMA)

The analytical expression of this indicator is the following:

$$\mathrm{EPA44} = \frac{\sum_{i=1}^{n}\left|\mathrm{Yi} - \hat{y}_i\right|}{n}\% \qquad (a.6)$$

According to some authors (see e.g. Wegener and Graef, 1982), the EPMA values may be interpreted as following:

- EPMA < 10%; $\Rightarrow$ extremely good forecast
- 10% < EPMA < 20%; $\Rightarrow$ good forecast
- 20% < EPMA < 50%; $\Rightarrow$ reasonable forecast
- EPMA > 50%. $\Rightarrow$ bad forecast