VU Research Portal

Modelling Inter-Urban Transport Flows in Italy: A Comparison between Neural Network Analysis and Logit Analysis

Nijkamp, P.; Reggiani, A.; Tritapepe, T.

published in

Transportation Research. Part C
1996

document version

Publisher's PDF, also known as Version of record

[Link to publication in VU Research Portal](#)

citation for published version (APA)

Nijkamp, P., Reggiani, A., & Tritapepe, T. (1996). Modelling Inter-Urban Transport Flows in Italy: A Comparison between Neural Network Analysis and Logit Analysis. *Transportation Research. Part C*, 4(6), 323-338.

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

E-mail address:

vuresearchportal.ub@vu.nl

MODELLING INTER-URBAN TRANSPORT FLOWS IN ITALY: A COMPARISON BETWEEN NEURAL NETWORK ANALYSIS AND LOGIT ANALYSIS

PETER NIJKAMP,¹ AURA REGGIANI² and TOMMASO TRITAPEPE^{1,2}

¹Department of Economics, Free University, De Boelelaan 1105, 1081 HV, Amsterdam, The Netherlands

²Department of Economics, University of Bologna, Piazza Scaravilli 2, 40126, Bologna, Italy

Abstract—In the present paper a modal split problem is analysed by means of two competing statistical models, the traditional logit model and the new technique for information processing, viz. the feedforward neural network model. This study aims to explore the modal split between rail and road transport modes in Italy in relation to the introduction of a new technological innovation, the new High-Speed Train. The paper is sub-divided into two major parts. The first part offers some general considerations on the use of neural networks in the light of the increasing number of empirical applications in the specific area of transport economics. The second part describes the Italian case study by using the two above mentioned statistical models. The results highlights the fact that the two adopted models, although methodologically different, are both able to provide a reasonable spatial forecasting of the phenomenon studied. In particular, the neural network model turns out to have a slightly better performance, even though there are still critical problems inherent in its application. Copyright © 1996 Elsevier Science Ltd

1. INTRODUCTION

Competition effects between distinct transport modes are usually analysed by means of discrete choice models, e.g. a logit model. Their general purpose is to assess the choice behaviour based on a utility function related to each transport mode in relation to explanatory attributes (e.g. time, cost, distance) associated with that mode. Usually, such a model is calibrated on the basis of a data set which contains discrete real-world observations. Clearly, there are also alternative approaches, such as spatial interaction models and neural network models. In the present paper we use both logit models and neural network models to evaluate the impact of the High-Speed Train in Italy, with a specific focus on the competition between road and rail transport modes.

The theoretical foundation of the logit approach is based on the theory of utility maximization incorporating a random factor for the representation of stochastic choice or decision elements (see Ben Akiva and Lerman, 1985). The (feedforward) Neural Network (NN) approach is based on a more recent statistical model which does not presuppose the use of mathematical rules to explain the behaviour of the problem under consideration, and therefore, it may be able to better describe decision events suggested by learning theory. In other words, the modal choice decision process can be conceived of as a NN where the decision maker's choice is a serial result of the network's adaptation. Therefore, in this paper, the results obtained from the logit model are compared to the results of the feedforward NN model.

In Section 2, we concisely describe the functions of a NN (Section 2.1), the problem of the sample size and the proper approximation of NNs (Section 2.2), and finally also new advances and research directions in this framework (Section 2.3). In Section 3, the empirical problem related to modal split choice in Italy between road and rail are analysed by means of both the logit model and the NN model, and the results are compared. Finally, Section 4 contains some concluding remarks.

2. THE USE OF FEEDFORWARD NNs IN TRANSPORT APPLICATIONS: GENERAL OBSERVATIONS

In the past decade a great deal of attention has been paid in the scientific literature to NNs as alternative models of information processing. NNs have successfully been used in different branches of learning processes such as artificial intelligence, operational research, biology, psychology. Increasingly, geographers, regional scientists and transportation analysts have recognized the analytical potential of this new approach.

In this section the feasibility, i.e. the features and emerging problems in relation to applications of NNs in regional and transportation science, are analysed¹. Rather than describing the theoretical aspects of NNs here², we discuss the basic elements of NNs in order to offer further insights for future research, mainly in light of the advances in empirical applications.

In Section 2.1, the functions of NNs are investigated in order to examine the type of data which can be handled in an NN context. An important issue of NNs is related to the size of the data set. Since in the specific area of transport economics, it is rather difficult to use large data sets (for computational reasons), this dilemma of sample size has great relevance in this context. It is analysed in more detail in Section 2.2. Finally, Section 2.3 briefly describes new advances and research directions in this framework.

2.1. Functions of NNs

The term "neural networks" is rather general, as it may cover various typologies/models. However, a main distinction can be made between NNs with and without supervisor.

This difference is based on the underlying different training process; in fact, the supervised training implies the knowledge of input/output data in order to find, during the learning phase, the weights³ of the network which minimize the error function of the target outputs and the network outputs. Although different training algorithms⁴ exist, the most utilized one is the back-propagation algorithm. As far as the unsupervised training in NNs is concerned, these networks do not need the target outputs; they modify the weights by means of competitive learning algorithms, in response to the input data.

The aim of the present section is to investigate whether and how it is possible to use NNs in regional and transportation science. In particular, considerable attention is given to a feedforward⁵ NN with a supervised training algorithm, an approach which is widely adopted in NN applications (see, for a review of NNs in transportation science, Dougherty, 1995; Himanen *et al.*, 1996; Reggiani *et al.*, 1995). It is important to underline the general characteristics of NNs in order to better grasp their real potential, their context sensitivity and, finally, the type of data suitable for application. Three analytical purposes—corresponding to different features of feedforward NNs—are distinguished here (see also Fig. 1):

1. *Explanatory*: although several efforts have been made to interpret and open the black-box characterizing NN methodologies, so far insufficient success has been achieved. For example, socioeconomic or spatial behavioural explanations have thus far only been reflected in the value of the weights of the connections between the units.

¹An interesting analysis of the potential role of NNs in geographic information processing can be found in Fischer's work (Fischer, 1994). As far as the applications of NNs in transport economics and regional science are concerned, an overview is contained in Reggiani *et al.*, 1995.

²For a theoretical presentation on NNs, see, among others, Dougherty, 1995 and Reggiani and Tritapepe, 1996.

³The weights are the values (adaptive coefficients) assigned to the links between the units of the network. The aim of the training phase is to find the values of the weights which will produce a reasonable output in response to input signals.

⁴The aim of a training algorithm is to minimize the error function by adjusting the value of the weights.

⁵The term *feedforward* refers to the transmission of the information only from the input units towards the output units.

Fig. 1. The features of a feedforward neural network (increase in gray shade denotes a rise in relevance of the features).

2. *Exploratory*: NNs have at least two main features supporting exploratory data analysis:
 - NNs do not *a priori* require algorithms or rules' development; this feature may be very useful in cases of a large quantity of data (e.g. in a GIS context; see Fischer, 1994) in which the knowledge of the exact statistical model for explaining the phenomenon examined is lacking.
 - NNs can learn and then forecast even on the basis of incomplete, noisy and fuzzy information. In this context, it should be noted that NNs are well suited for solving pattern recognition problems. NNs have been used in various areas for such a task, and in general rather good performances have been obtained. Also in transport systems several applications—with reference to traffic control/pattern recognition—can be found (see among the others Reggiani *et al.*, 1995; Ritchie and Cheu, 1993).
3. *Forecasting*: NNs can be useful as predictive models, mainly taking into account the two features previously described (under point (2)). In this context, one question arises concerning the use of micro-variables versus macro-variables. From a computational/modelling viewpoint, NNs are able to handle any kind of data. Usually, in the transportation field, most of the statistical models are theoretically based on micro-variables describing individual behaviour. However, the data are often aggregated under the assumption that classes of individuals behave in the same way. Therefore, the statistical models often reflect the limits associated with this assumption. NNs seem to be able to overcome this limit by capturing stochastic elements neglected in the previous assumption.

In Fig. 1 the main NN features, illustrated in the previous three points, are summarized.

It is certainly interesting to complement the previous considerations with empirical experiments. At present, only a few experiments have already been carried out (see e.g. Fischer and Gopal, 1994) and in these works some practical problems inherent in using NNs have been pointed out. Our experiments, which will be illustrated in Section 3, are a further contribution to this field by investigating the modal split in Italy between rail and road transport. This may offer useful information for a subsequent analysis of the competitive effects resulting from the introduction of the High Speed Train in Italy. In the next section, we consider the problem of the size of the data set.

2.2. Sample size and approximation of NNs

2.2.1. Introductory remarks. An important question in NN analysis concerns the sample size. Unfortunately, there is no simple or general answer to the question how large a data set should be in order to generate a good approximation. If we denote the size of the training-set by n and the number of the weights of the NN by s , then “practical experience in econometrics suggests that it is probably safe to recommend taking n to be on the order of magnitude of $s \times 100$, but this is no more than an educated guess. Rules suggesting that $s \times 10$ may be adequate (see e.g. Baum and Haussler, 1989) are relevant to classification problems in which the relation between X and Y is non-probabilistic” (see White, 1992).

From an analytical point of view, it is only possible to find a precise answer by specifying the exact distribution of the weights w_n for a given n ; but this is an intractable problem. By using standard statistical tools, including the law of large numbers and the central limit theorem, it is possible to demonstrate that the probability law of w_n collapses to a particular well-defined set, becoming more and more concentrated around this set as n increases (property of ‘consistency’) (see White, 1992). It is however, possible to obtain some limited information on the behaviour of w_n for fixed n by means of Monte Carlo experiments. An obvious limitation is then that the results pertain only to the environment in which the experiments are carried out.

It is also possible to assess an empirical distribution for w_n that does not depend on large n approximations nor on artificial data generating assumptions, by using a bootstrap technique (see Section 3.2). Some indicators useful for evaluating the performance of statistical models are now presented in Section 2.2.2, while the problem of the size of the data-set, pointing at the concept of variability of the results, are dealt with in Section 2.2.3.

2.2.2. Statistical indicators. The results of NN models have to be evaluated by statistical tools. The Mean Squared Error (MSE) is a very common indicator and can be represented as follows:

$$MSE = \frac{\sum_{n=1}^N (y_n - \hat{y}_n)^2}{N} \quad n = 1, \dots, N \quad (2.1)$$

where y_n denotes the expected values and $\hat{y}_n$ denotes the values calculated from the model by using a data sample with N observations.

A useful indicator, often used in the literature on NNs (see e.g. Fischer and Gopal, 1994), is the Average Relative Variance (ARV(N)). It is defined as follows:

$$ARV(N) = \frac{\sum_{n \in N} (y_n - \hat{y}_n)^2}{\sum_{n \in N} (y_n - \bar{y})^2} \quad n = 1, \dots, N \quad (2.2)$$

where y_n denotes the expected value, $\hat{y}_n$ denotes the value calculated from the model and $\bar{y}$ is the average of the expected values belonging to the set of data N . By assuming that all events are ‘equiprobable’, becomes:

$$ARV(N) = \frac{1}{\sigma^2} \frac{1}{M_N} \sum_{n \in N} (y_n - \hat{y}_n)^2 \quad (2.3)$$

or

$$ARV(N) = \frac{MSE}{\sigma^2} \quad (2.4)$$

In the expressions (2.3) and (2.4), M_N is the number of observations belonging to the set N (i.e. using NN one can fix an arbitrary value of N (epoch)) and σ^2 is the variance of the expected values of the entire data set (this is a proxy, since the variance of each epoch would have been considered).

In order to compare two methodologically different models, the use of the ARV indicator is mainly useful for two reasons (in relation to the non-linear regression nature of the NN experiments):

1. Relationship (2.3)/(2.4) ensures an independence—both quantitatively and qualitatively—of the error estimated from the selected samples;
2. such an indicator will approach zero, the more exact the estimate is.

However, it is necessary to specify properly, whether the performance measure is calculated by using absolute values of the predicted flows [as in the case of spatial interaction models (see Fischer and Gopal, 1994)] or probabilistic flows [as in the case of logit models (see Reggiani and Tritapepe, 1996)]. In fact, in the second case the value of the indicator is much higher than in the first case.

In the next section we further deal with the problem of the size of the data set in a more interpretative sense.

2.2.3. Impact of the data set on the performance of a feedforward NN. In this section the need for using a large data set in a feedforward NN environment is briefly discussed from a methodological viewpoint. Suppose that the aim is to compare two statistical models, e.g. the logit model and the NN model. In order to compare two predictive models, it is necessary to use a training-set for calibrating and a test-set for testing⁶ the models⁷. As far as the training-set is concerned, it is important to stress the relevance of the size n (the number of observations of the set) and of the quality of the observations, since the performance of the two models depends on these two factors. In particular we investigate here why a small value of n is problematic in using a feedforward NN.

The first consideration is that, in general, the variability of the results increases if the data-set is small; for example, these results can be evaluated by using the MSE indicator (defined in the previous section) calculated for the test-set. It has already been mentioned that it is not possible to find an analytical formulation for the distribution of the weights w_{ij} ; however, it is possible to find an empirical distribution by using the bootstrap technique. Analogously, though it is not possible to find an analytical distribution of the results, it is however possible to use the bootstrap method in order to obtain the empirical distribution of the errors. Consequently, if we evaluate the empirical distribution of the error indicators (by means of the MSE indicator evaluated by using the test-set) for both the logit and the NN models, we can obtain the curves represented in Fig. 2, where, on the axes, the indicator (MSE) and the frequency (f) are denoted.

It is clear that we can not infer the mean (first moment) of the distributions⁸ without exactly calculating these distributions. On the contrary, we may expect a difference

Fig. 2. An example of variability of the results.

⁶The phase of "testing the models" means to evaluate the performances of the models under consideration by presenting to each model new observations (examples/patterns) for which the expected values are known. It is then possible to calculate an error indicator by comparing the predicted values and the expected values.

⁷The two data sets are the same for both models.

⁸The value of the mean depends on the performance of the two models.

Fig. 3. Scheme of the variability problem.

between the variabilities of the two distributions. In principle, we could state that the difference of the variabilities is high, if the size n of the training-set is small.

Before explaining in detail the reasons for such variability, it is important to stress its relevance. In Fig. 2 two different points on the NN curve, A and B, have to be emphasized; each point corresponds to the error (evaluated for the test-set) of an experiment carried out by randomly varying the observations as indicated in the bootstrap technique (see below). The figure shows that the errors can be very different (for example, A or B), while by using statistical models (e.g. the logit model) we usually find similar results (i.e. a small variability).

As illustrated in Fig. 3, it is possible to identify two different causes of variability:

1. variability because the limited data set is not a representative mapping of the real world;
2. variability resulting from the particular model adopted.

Regarding the first point, if the data set is large, it is evident that the variability is small and independent of the particular model. It should also be noted that the models present an intrinsic sensitivity to this variability since such a variability depends on the degrees of freedom of the model. For example, the logit model is less sensitive than the feedforward NN model to this kind of variability, due to the small number of parameters involved (i.e. weights).

Concerning point 2, it is noteworthy that error variability is inherent in particular models, such as NN models. In the feedforward NN model it mainly depends on two factors:

1. the parameters defining the NN;
2. the use of a Cross Validating Technique for solving the overfitting problem.

The first factor is related to the definition of the architecture of a feedforward NN; all parameters (e.g. the learning rate, the momentum factor, the epoch size, initial weight values) and the number of hidden units influence the variability of the results. The second factor concerns the overfitting problem. For an explanation of this characteristic problem in NN model estimation, see e.g. Reggiani and Tritapepe (1996). Two different techniques exist in order to overcome the overfitting problem: the pruning technique (based on eliminating the weights of the network whose values are close to zero; i.e. the weights which do not, or hardly, contribute to the transmission of information) and the cross validating technique. The latter method is often used, but it increases the variability of the results. This technique is based on subdividing the whole data set into three subsets: a training set for training the NN, a cross-validation set for determining the stopping point in order to avoid overfitting, and, finally, a test-set which is set apart and never used in the estimation stage (see Fischer and Gopal, 1994). By using the latter technique, the problem is that the results depend on changes in the choice of the three subsets; and such a variability increases if the data set is small.

In this section the general problem of the variability in results has been discussed in the light of applied NN analysis. In this respect, two main issues are important:

1. to know how large n should be in order to ensure 'good' predictions;
2. to know how large the variability of the results is.

From the considerations mentioned in Section 2.2, it seems that the bootstrap technique is the most suitable analysis for making proper statistical inferences. This technique has recently become more popular, since it requires modern high computer power to simplify the calculations of traditional statistical theory (see Efron and Tibshirani, 1993). "Bootstrapping involves resampling the data with replacement many, many times in order to generate an empirical estimate of the entire sampling distribution of a statistic" (see Mooney and Duval, 1993). The central idea of the bootstrap technique is that by resampling, with replacement of the data set, one may obtain many new data sets (resamples) and then therefore one may use the model at hand for each data set. In such a way, an empirical distribution of the results is available. As far as the NN analysis is concerned, one may refer here to the work by Weigend and LeBaron (1994) in which the authors use the bootstrap method for visualizing the distribution of the results of an NN by showing the variability due to the network parameters (case a) above and the variability due to the choice of the three subsets (case b) above. They show that the first variability (case a) is relatively small compared to the second variability (case b).

2.3. *Software and new advances in NNs*

It is virtually impossible to consider all the different NN models and their associated learning techniques in a few pages; consequently, in light of future research developments, we concentrate here only on hidden layer feedforward networks with a back-propagation learning algorithm⁹, since this is the most popular and utilized network in almost all applications. In Fig. 4 some promising new research lines for feedforward NNs are shown.

The first research direction concerns the investigation of new algorithms, different from the back-propagation algorithm, in order to make the learning phase faster. In this respect, the work of De Groot and Wurtz, 1994 should be noted, which deals with the algorithm: polytope, conjugate gradient, quasi-Newton and Levenberg-Marquardt. The reduction of time in the learning phase is a very important issue, since the choice of the parameters defining the architecture of the network leads to easier results. Furthermore, by using the new algorithms mentioned above, the number of parameters often decreases. At present however, no commercial or easily accessible software containing these new tools seems to exist.

⁹This paradigm was invented from the PDP (Parallel Distributed Processing) Research Group at MIT (see Rumelhart *et al.*, 1986a,b) and others (see Le Cun, 1986; Parker, 1985); for a historical view, see Anderson and Rosenfeld, 1990.

Fig. 4. Scheme of the new research directions for feedforward neural networks.

The second research direction concerns the investigation of techniques which are able to avoid the overfitting problem. Various software programmes adopting some of these techniques already exist; however, no software contains all these tools.

Finally, as far as the third research line is concerned, i.e. how to control the variability of the results in a feedforward NN, the bootstrap technique is a very useful method as mentioned already. This method requires a modern powerful computer; to our knowledge, nowadays no commercial software containing the bootstrap technique applied to NNs exists.

In the next section we evaluate the potential of NN tools *vis-a-vis* logit analysis by means of an empirical case study.

3. THE EMPIRICAL APPLICATION

The aim of the present section is to explore the modal split between rail and road transport modes¹⁰ in Italy in view of the future introduction of a specific new technological innovation in transport, notably the new High-Speed Train. For this purpose a modal split analysis has been carried out by means of two statistical models which are methodologically different (see Reggiani *et al.*, 1995), viz. the logit model and the feed-forward NN model. In our case, the modal split problem refers to the links between 67 Italian areas (corresponding to a subdivision of the whole territory into provinces or aggregations of provinces). In Fig. 5 a scheme of the modal split problem under consideration is shown, while the next sections will present the details of the empirical application.

¹⁰In the year 1987 air transport was not so developed in Italy, then the related matrix contained several zero cells. Obviously it may be useful in the future to add the air dimension, given the fast growth of this alternative choice.

Fig. 5. Scheme of a modal split problem between 67 areas.

3.1. The basic elements

3.1.1. The data. The data set¹¹ contains the flows and the attributes related to each link between 67 Italian areas which correspond to the whole of Italy (except the Island of Sardinia). Most of the areas correspond to the provincial territory and some others are an aggregation of two or three provincial territories (see Figs 6 and 7).

The attributes considered are 'distance', 'time' and 'cost' between each link [i-j] with reference to each transport mode. In particular each observation of the data-set contains variables related to each link [i-j]. Concerning the flow-distribution, the related matrices refer to aggregated data¹² (hence we cannot take into account specific classes of individuals, e.g. related to age, profession, etc.).

Since 67 areas have been considered, the data set should ideally contain 4489 observations according to the previous definitions of observations. However, the data set contains, finally, 1396 observations because of the following considerations (by analysing the data set):

1. the intra-area flows are zero;
2. the flow matrices are symmetric;
3. only the links where the flows and the attributes (of both road and train) are different from zero have been considered.

Before describing in detail the experiments and the related results (see Section 3.2), the statistical indicator utilized for evaluating the results are presented in the next section.

3.1.2. The performance measures. The modal split problem is analysed by means of two different statistical models. Therefore, in order to evaluate the performance of these two models, the statistical indicator ARV widely used in the NN literature as well as in classical models (see e.g. Fischer and Gopal, 1994; Reggiani and Tritapepe, 1996) has been utilized. The ARV indicator was already defined in Section 2.3.

In the present work all experiments are evaluated by using the absolute values of the predicted flows according to the following expression:

$$T_{ij}^{\text{train}} = p_{ij}^{\text{train}} * T_{ij} \quad (3.1)$$

where T_{ij} is the total flow related to link (ij), T_{ij}^{train} is the total train flow related to link (ij) and P_{ij}^{train} is the probability of choosing the mode train with reference to link (ij). Furthermore, the performance power is evaluated by means of residual analysis (see Section 3.2.3).

¹¹The data-set has been kindly provided by the Italian State Railways ("Ferrovie dello Stato") and it refers to Census data (1987).

¹²Then, the estimated probability refers to the choice of all the travelers on a link.

Fig. 6. The subdivision of Italian land in 67 zones.

3.2. The experiments

The experiments were carried out by using two statistical models: the logit model (see Section 3.2.1) and the neural net model (see Section 3.2.2). The whole data set contain 1396 observations (examples or patterns). The following considerations concern all experiments carried out:

1. The data set has been randomly subdivided into three sub-sets:
 - (a) a training-set containing 698 observations, i.e. 50% of the data set;
 - (b) a cross-validation set containing 349 observations, i.e. 25% of the data set;
 - (c) a test-set containing 349 observations, i.e. 25% of the data set.

THE PROVINCES					
2	VERCELLI	21	GORIZIA	41	PESARO
3	ASTI	22	TRIESTE	42	LIVORNO
4	NOVARA	23	VENEZIA	44	GROSSETO
5	ALESSANDRIA	24	PADOVA	46	ROMA
6	PAVIA	25	ROVIGO	47	VITERBO
7	MILANO	26	IMPERIA	49	CHIETI
8	VARESE	27	SAVONA	50	FROSINONE
10	COMO	28	GENOVA	51	LATINA
11	BRESCIA	29	PARMA	54	FOGGIA
12	PIACENZA	30	REGGIO EMILIA	55	NAPOLI
14	TRENTO	31	MODENA	57	POTENZA
15	BOLZANO	32	FERRARA	59	BRINDISI
16	VERONA	33	BOLOGNA	60	TARANTO
17	VICENZA	35	MASSA	61	LECCE
19	PORDENONE	36	LA SPEZIA	63	REGGIO CAL.
20	UDINE	39	SIENA	64	MESSINA
AGGREGATION OF TWO PROVINCES					
9	BERGAMO Sondrio	38	FIRENZE Pistoia	56	SALERNO Avellino
13	CREMONA Mantova	40	AREZZO Perugia	58	BARI Matera
18	TREVISIO Belluno	48	PESCARA Teramo	62	COSENZA Catanzaro
34	FORLI' Ravenna	52	CASERTA Isernia	67	SIRACUSA Ragusa
37	PISA Lucca	53	BENEVENTO Campobasso		
AGGREGATION OF THREE PROVINCES					
1	TORINO Aosta Cunco	45	TERNI Rieti L'Aquila	66	CATANIA Enna Caltanissetta
43	ANCONA Macerata Ascoli Piceno	65	PALERMO Trapani Agrigento		

Fig. 7. The 67 Italian zones.

- The training for the neural net model (and the calibration for the logit model) was carried out by using the same training-set.
- The comparison of the two models (and therefore the performance measure) was evaluated by using the same test-set.
- The variables V_j ($j=1, \dots, 9$) were transformed (in values between [0–1]) by means of the following functions:

$$V_j^f = \exp(-0.005 * V_j) \quad (3.2)$$

$$V_j^t = \exp(-0.0002 * V_j) \quad (3.3)$$

In particular, eqn (3.2) (identified by the index f) was adopted for transforming the values of the attributes, and eqn (3.3) (identified by the index t) for transforming the flow values. The variables are defined as follows:

- TD_{ij}^f : transformed train distance (according to eqn (3.2)) for link (ij);
- TC_{ij}^f : transformed train cost (according to eqn (3.2)) for link (ij);
- TT_{ij}^f : transformed train time (according to eqn (3.2)) for link (ij);
- RD_{ij}^f : transformed road distance (according to eqn (3.2)) for link (ij);

- (e) RC_{ij}^f : transformed road cost (according to eqn (3.2)) for link (ij);
- (f) RT_{ij}^f : transformed road time (according to eqn (3.2)) for link (ij);
- (g) T_{ij}^f : transformed total passenger flow (according to eqn (3.3)) for link (ij);
- (h) $T_{ij}^{train(f)}$: transformed rail mode flow (according to eqn (3.3)) for link (ij);
- (i) p_{ij}^{train} : rail mode probability for link (ij) (see eqn (3.1)).

3.2.1. The logit approach. The logit model is well known in the transport economics literature (see e.g. Ben Akiva and Lerman, 1985) and widely used in modal split analysis. It estimates the probability of choosing a transport mode by maximizing a utility function related to each mode. The general form is:

$$p_j = \frac{e^{u_j}}{\sum_{i=1}^n e^{u_i}} \quad i, j = 1, \dots, n \quad (3.4)$$

where p_j is the probability of choosing mode j , n is the number of possible choices and u_j is the utility function related to mode j , which is usually assumed linear in the attributes x_k , by means of the parameters β_k , as follows:

$$u_j = \sum_k^K \beta_k x_k + \delta_j \quad k = 1, \dots, K \quad (3.5)$$

where δ_j is the random component. In this application we assume $j=1,2$ ($n=2$), i.e. only two choices are possible: road or train. The related attributes x_k are six in number ($k=1, \dots, 6$); more precisely, they are the following:

$$TD_{ij}^f, TC_{ij}^f, RC_{ij}^f, RD_{ij}^f, RT_{ij}^f$$

which have already been previously defined.

The estimation of eqn (3.4) and eqn (3.5) on the area under investigation was carried out here by means of the procedure *Blomit* in the STATA¹³ software.

3.2.2. The feedforward neural net model. Following the majority of applications on NNs, in this study a two-layer feedforward, totally connected (connection between consecutive layers) NN is used in order to analyse the modal split problem. The methodological structure of the main phases related to the application of a feedforward NN is described in Reggiani and Tritapepe (1996). Concisely, it consists of three stages: (a) definition of network architecture; (b) learning phase; (c) forecasting phase. First, it is then necessary to define the right architecture of the network, i.e. the number of units on the levels. Usually, the input and output units depend on the number of input and output variables which define the problem. In our application, aiming to compare the neural net approach with the logit model, one possible NN architecture contains six input units which correspond to the attributes of the logit model (previously described) and one output unit corresponding to the probability of choosing one mode¹⁴ (e.g. the train mode; see the case NA in Fig. 8).

However, due to the versatility of NNs, it is possible to identify further architectures (in order to solve the same problem) by investigating the related performances:

Case NB: this architecture consists of seven input units which correspond to the six attributes and to the total flow related to link (i,j), with one output unit which corresponds to the total flow on the train mode.

¹³Such a procedure allows the calculation of logit estimates respecting aggregate data, adopting a maximum likelihood technique and applying the Newton–Raphson algorithm as the convergence procedure.

¹⁴The probability of the other mode is complementary.

Fig. 8. The architecture of the experiments—related to the neural net and logit model—for the modal split problem under investigation (six attributes).

Case NC: this last architecture is formed by six input units which correspond to the 6 attributes, and two output units which show both the total flow and the flow on the train mode related to link (i,j) .

The performance measure of each case is discussed in the next section.

Concerning the number of hidden units, they were empirically 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 were set out after several empirical experiments. Finally, the parameters of the NNs result as follows:

- number of hidden units: 3
- learning rate $\alpha = 0.5$

- momentum factor $\lambda = 0.2$
- epoch size: 1
- initial weight values: randomly between $[-0.1; +0.1]$

By using a feedforward NN it is necessary to cope with the overfitting problem. Consequently, in all three cases NA, NB and NC, the cross-validating technique (by using the cross-validation subset; see Section 2.2) 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; Reggiani and Tritapepe, 1995).

3.2.3. The results. In this section the results related to the above experiments are shown. 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 values. Since the final aim of our work is to evaluate the impact of the High-Speed Train in Italy, the logit model and the neural net model have to be calibrated (trained) in order to forecast next the new probabilities of using the transport modes, on the basis of new transport scenarios. The present section analyses this first research stage, i.e. the spatial forecasting of the two models adopted. The comparison of predictive quality is based on the performance measure, by using the test-set which had been set apart and never used for the calibration (learning) phase, as mentioned above.

The prediction performance is evaluated by means of the statistical indicator ARV (defined in Section 2.3) as follows (see Table 1).

Three considerations emerge from the results shown in Table 1:

1. The performance of the NN model in case NA is slightly better than the performance of the logit model.
2. Case NA corresponds to the best neural architecture. In this case the input and the output variables are the same as in the logit model.
3. The error in case NC is much higher. This result is rather intuitive, since in this case the NN forecasts not only the train flow but also the total flow departing from each origin.

An interesting representation can now be given, illustrating the errors of the whole test-set by means of residual analysis. In particular, Fig. 9 shows the following two results concerning the probability related to the train mode for both the feedforward neural net model (case NA) and the logit model (n is the number of observations of the test-set and P_n^{obs} denotes the real rail mode probability):

1. absolute residuals $(P_n^{obs} - P_n^{neur})$ and $(P_n^{obs} - P_n^{logit})$;
2. relative residuals $(P_n^{obs} - P_n^{neur})/P_n^{obs}$ and $(P_n^{obs} - P_n^{logit})/P_n^{obs}$

The residuals are ordered in the graphs according to the size of the P_n^{obs} rail mode probability.

By observing the graphs (a1) and (a2) in Fig. 9, it should be noted that in the predictions of the NNs more oscillations exist and that the difference in the performance of the models corresponds mainly to the observations whose P_n^{obs} (real rail mode probability) has medium values (around the value of 0.5).

Finally, by observing the graphs (b1) and (b2) in Fig. 9, significant differences between the models do not appear. In both models relative residuals have absolute high values corresponding to the small values of probabilities of choosing rail (because of the definition of the relative residuals).

Table 1. The results measured by means of the ARV indicator

$ARV_{NN}(NA) = 0.095$	$ARV_{Logit} = 0.1029$
$ARV_{NN}(NB) = 0.1341$	
$ARV_{NN}(NC) = 0.7109$

Fig. 9. Residuals of the neural network (case NA) and the logit model predictions.

4. CONCLUSIONS

In this paper, the logit model and the feedforward NN model have been compared by analysing a modal split problem. Due to versatility of NNs three different neural architectures have been considered. However, in only one case (case NA) the NN performs better than the logit model and, further, also in this case, the difference between the values of the statistical indicators is small. By comparing such a difference, it is necessary to conclude that both models provide a very good performance, probably because of the structure of the data set.

We may thus conclude that NN may play an important role in transportation analysis (as has been illustrated in Section 2); however, at this time, such a new technology has some apparent problems. Using a feedforward NN, at least three problems exist:

1. the choice of all parameters defining the neural architecture; consequently, this leads to a "time-consuming problem" and to a "variability of the results problem" (see Section 2.2.2).
2. the necessity of a procedure to take into account the "overfitting problem". In particular, by using the cross-validating technique it is not exactly clear when to stop the learning procedure, so that, by choosing the subsets, the "variability of the results problem" returns (see Section 2.2.2).
3. NNs are not able to provide an economic significance of the problem under investigation.

The first two points pinpoint the "variability of the results problem"; therefore we propose to use a technique like the bootstrap method in order to visualize a distribution

of the results. Because of these reasons, it seems that the NN model should be used with caution and only in cases where the performance is much better than the traditional discrete choice model. Concerning point 3, it should be noted that some research has been already carried out in this context, precisely in macroeconomics (see e.g. Sargent, 1993) and in chaos theory (see e.g. Freeman, 1994). Concerning the logit model, our experiments reconfirm its validity and usefulness in the area of transport economics.

Acknowledgements—The second author gratefully acknowledges the CNR project PFT2 no. 96.00098.PF74. The third author acknowledges a Human Capital and Mobility Research Grant of the EC (Strategic Infrastructure Networks in Europe, Contract no. ERBCHBGCT920162), as well as for the good facilities to carry out the study at the host institution (Economic and Social Institute, Department of Economics, Free University, Amsterdam). All the authors wish to thank three anonymous referees for the helpful comments.

REFERENCES

- Anderson, J. A. and Rosenfeld, E. (eds) (1990) *Neurocomputing: Foundations of Research*, MIT, Cambridge, MA.
- Baum, E. and Haussler, D. (1989) What size net gives valid generalization? *Neural Computation*, 1, 151–160.
- Ben Akiya M. and Lerman S. R. (1985) *Discrete Choice Analysis Theory and Application to Travel Demand*, MIT Press, Cambridge, MA.
- DeGroot, C. and Wurtz, D. (1994) Plain backpropagation and advanced optimization algorithms: a comparative study. *Neurocomputing*, 6, 153–161.
- Dougherty, M. (1995) A review of neural networks applied to transport. *Transportation Research C*, 3, 247–260.
- Efron, B. and Tibshirani, R. J. (1993) *An Introduction to the Bootstrap*, Chapman and Hall.
- Fischer, M. M. (1994) Expert systems and artificial neural networks for spatial analysis and modelling. *Geographical Systems*, 1, 221–235.
- Fischer, M. M. and Gopal, S. (1994) Artificial neural networks: a new approach to modelling interregional telecommunication flows. *Journal of Regional Science*, 34, 503–527.
- Freeman, W. J. (1994) Neural networks and chaos. *Journal of Theoretical Biology*, 17, 13–18.
- Himanen, V., Nijkamp, P. and Reggiani, A. (eds) (1996) *Neural Networks in Transport*. Avebury, Aldershot (forthcoming).
- Le Cun, Y. (1986) Learning processes in an asymmetric threshold network. *Disordered Systems and Biological Organization* (Bienenstock, E., Fogelman Souli, F. and Weisbuch, G., eds) Springer, Berlin.
- Mooney, C. Z. and Duval, R. D. (1993) Bootstrapping: a nonparametric approach to statistical inference. *Sage, University Papers*, Sage Publications, Inc.
- Parker, D. (1985) Learning logic. *Technical Report TR-87*, Center for Computational Research in Economics and Management Science, MIT, Cambridge, MA.
- Reggiani, A., Romanelli, R., Tritapepe, T. and Nijkamp, P. (1995) Neural networks: an overview and applications in the space-economy, Discussion Paper TI 95-40 Tinbergen Institute, Amsterdam. *Neural Networks in Transport Systems* (Himanen, V., Nijkamp, P. and Reggiani, A., eds), Avebury, Aldershot (forthcoming).
- Reggiani, A. and Tritapepe, T. (1996) Neural network and logit models to commuters' mobility in the metropolitan area of Milan. *Neural Networks in Transport Systems* (Himanen, V., Nijkamp, P. and Reggiani, A., eds), Avebury, Aldershot (forthcoming).
- Ritchie, S. G. and Cheu, R. L. (1993) Simulation of freeway incident detection using artificial neural networks. *Transportation Research C*, 1, 203–217.
- Rumelhart, D. E., Hinton, G. E. and Williams, R. J. (1986a) Learning internal representations by error propagation. *Parallel Distributed Processing: Explorations in the Microstructures of Cognition* (Rumelhart D.E. and McClelland J.L., Eds.) Vol. 1, pp. 318–362. MIT Press, Cambridge, MA.
- Rumelhart, D. E., Hinton, G. E. and Williams, R. J. (1986b) Learning representations by back-propagating errors. *Nature*, 323, 533–536.
- Sargent, T. J. (1993) *Bounded Rationality in Macroeconomics*. Oxford University Press, Oxford.
- Weigend, A. S. and LeBaron, B. (1994) Evaluating neural networks predictors by bootstrapping. Paper submitted to ICONIP'94-Seoul.
- White, H. (1992) *Artificial Neural Networks: Approximation and Learning*. Blackwell, Oxford.