



Deposited via The University of Leeds.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/2541/>

Article:

Clark, S.D., Grant-Muller, S.M. and Chen, H. (2002) Using Non-Parametric Tests to Evaluate Traffic Forecasting Performance. *Journal of Transportation and Statistics*, 5 (1). pp. 47-56.

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



White Rose
university consortium
Universities of Leeds, Sheffield & York

White Rose Research Online

<http://eprints.whiterose.ac.uk/>

ITS

[Institute of Transport Studies](#)

University of Leeds

This is a publisher produced version of a paper published in the Journal of Transportation and Statistics. The papers from this journal are in the public domain and may be used and reprinted without special permission. For more information please visit their [website](#).

White Rose Repository URL for this paper:
<http://eprints.whiterose.ac.uk/2541/>

Published paper

Clark, S.D.; Grant-Muller, S.M.; Chen, H. (2002) *Using Non-Parametric Tests to Evaluate Traffic Forecasting Performance*. Journal of Transportation and Statistics, 5(1), pp.47-56.

Using Nonparametric Tests To Evaluate Traffic Forecasting Performance

STEPHEN D. CLARK*
Morley Town Council

SUSAN M. GRANT-MULLER
HAIBO CHEN
University of Leeds

ABSTRACT

This paper proposes the use of a number of nonparametric comparison methods for evaluating traffic flow forecasting techniques. The advantage to these methods is that they are free of any distributional assumptions and can be legitimately used on small datasets. To demonstrate the applicability of these tests, a number of models for the forecasting of traffic flows are developed. The one-step-ahead forecasts produced are then assessed using nonparametric methods. Consideration is given as to whether a method is universally good or good at reproducing a particular aspect of the original series. That choice will be dictated, to a degree, by the user's purpose for assessing traffic flow.

INTRODUCTION

Many models attempt to predict the behavior of a system. These models may be physical, mathematical, statistical, or simulation representations of the system. Within the transportation field, physical models can be scale models of the geographical area of interest, mathematical models can be queuing models, statistical models can be platoon dispersion models, and simulation models can be meso- or microsimulation models. These models may operate on cross-section data, which represent a snapshot of a system at a particular point in time, or on time series data, which represent the "movement" of a system through time.

If the appropriate model for the system is known, a dataset is used to calibrate the parameters in the model and then the model is applied. If the model is not known, then procedures are necessary to select from a range of models. As part of this selection process, a commonly used procedure is to split the dataset into two portions for training and testing purposes. The training portion, which is usually the larger, is used to calibrate the parameters in the model, and then the testing portion is used to assess the accuracy of the calibrated model in reproducing observed behavior. If the performance of the model with the testing dataset is deemed adequate, then the two datasets are pooled and the model recalibrated. Sometimes there is either insufficient data of acceptable quality to enable this partition to take place or no obvious way of dividing the datasets. In such cases, a with-replacement sampling approach may be adopted to construct the two datasets. To accurately assess without bias a model's goodness-of-fit, the modeler must first determine the values of the calibration parameters and then assess the performance of that model.

This paper, while incorporating forecasting models, is not concerned with a detailed study of the relative merits of these models, but with methods of assessing their ability to produce useable forecasts. In particular, this paper does not concern itself with the accepted iterative procedures of model identification, model estimation, and model diagnosis. It is assumed that these stages have been successfully completed and that the practitioner is now interested in how the model performs.

ASSESSING GOODNESS-OF-FIT

In the modeling processes and for models used for forecasting discussed in this paper, there are two types of discrepancy between the observed and modeled values. Within-sample discrepancies, which are typically

generated during the model-fitting stages, are termed residuals in this paper. The outside-sample discrepancies are those that arise from applying the model to “unseen” data and are termed forecast errors in this paper. It is this latter form of discrepancy that is of most interest to practitioners and is the one considered in this paper.

A primary requirement is that a goodness-of-fit test be dependable. It should also be accurate and consistent in application. The fewer the number of assumptions that accompany the test the better the assessment of goodness-of-fit. Such assumptions may include the distribution of observations or existence of a sufficiently large sample size. Sometimes the test may be robust to departures from these assumptions, but a doubt may still exist over any measure that compromises any of these assumptions.

An additional modeler task is to communicate information to those who have the authority or influence to use it. Unfortunately, such individuals’ expertise often differs from that of the modeler. This places a requirement that the metrics used in assessing goodness-of-fit are readily comprehensible and acceptable to specialists in other fields. Much of the motivation for this paper comes from earlier work by Dadkhah and Zahedi (1986), in which they propose various nonparametric tests to identify models that can predict turning points and directions of change in a time series. They also list a wide range of model evaluation tests in their appendix. In practice, however, not many of these evaluation tests outlined are actually used, because they would prove daunting when communicating results to a nonstatistically aware audience

The commonly used measures are those that involve an averaging of a simple function of the difference between the observed and forecast behavior. One such term is the root mean square error (RMSE):

$$RMSE = \sqrt{\frac{1}{N} \sum_{t=1}^N (f_t - v_t)^2} \quad (1)$$

Where f_t is the forecast at time t , v_t is the observation at time t , and N is the number of observations while another is the mean absolute percentage error (MAPE):

$$MAPE = \frac{1}{N} \sum_{t=1}^N \frac{|f_t - v_t|}{f_t} \quad (2)$$

Both of these statistics have the advantage of being easily comprehended by most practitioners. Some disadvantages of these measures follow.

1. There is no criterion for assessing whether one value of the statistic is acceptable or not. Usually a range of forecasts are produced using either different methods or different datasets and a subjective opinion made as to whether one result is good or not in the context of the other results.
2. The RMSE or MAPE are often used in the model calibration stage to estimate the parameters in a model. Thus, there is the possibility that any calibrated model may be biased in producing estimates that give good performance on that measure but poor performance on other, equally valid, measures of goodness-of-fit.
3. While some forecasting methodologies specify several distributional requirements on the residuals from estimated models, and these requirements can be tested (but see 6 below), it is not usually necessary to place distributional requirements on outside sample forecast errors.
4. These statistics group all the observations together, losing the individual point-to-point relationship that exists. This drawback is particularly serious for time series data where the time element is important but lost in the aggregation.
5. The measures are not especially robust to outliers in the data, in particular the RMSE will exaggerate the impact of any outliers in either the observed or forecast series.
6. If any standard statistical tests are applied to these data, certain assumptions on the distribution of the difference between the modeled and observed values, termed the residuals, are required. These assumption can be (but are seldom) tested, but even when assumptions are found to be

valid, there is still a remaining doubt (Type II errors).

NONPARAMETRIC METHODS OF ASSESSMENT

Nonparametric methods provide an alternative approach to assessing goodness-of-fit and pose certain advantages over parametric or averaging approaches, namely:

1. they do not assume any underlying distribution for the data used in the test,
2. they are able to provide objective methods for assessing whether a result is acceptable,
3. they are applicable with small sample sizes,
4. they can be robust to outliers, and
5. they are more readily comprehensible to specialists in other disciplines.

Three types of nonparametric tests are discussed in this paper. The first set are tests of the location of distributions based on signs, the second on the equality of shape of distributions, and the third on correspondence of distributions. These tests may be applied to the original and forecast data points and/or the original and forecast directions of change in a series.

DATA AND PRACTICAL CONTEXT

The English Highways Agency collects traffic information continuously at one-minute intervals on traffic flow (measured in vehicles), speed (km/hour), headway (seconds), and detector occupancy (percentage) on the M25 motorway (freeway). Detectors are typically located 500 meters apart and there is one in each traffic lane of the carriageway. One of the primary purposes for this infrastructure is to monitor traffic on the motorway with a view to activating a series of speed variable message signs as congestion builds (Maxwell and Beck 1996; Nuttall 1995). Currently, the Highways Agency uses the system in a reactive mode, that is, decisions on whether to activate the message signs are made on the basis of the most recent traffic situation. They are actively investigating whether an anticipatory mode may be more efficient, where traffic conditions are forecast for a short time horizon, typically less than one hour, and action taken to forestall anticipated congestion.

For the purposes of this study the 1-minute, 4-lane traffic flows have been aggregated into 15-minute carriageway flows (expressed as equivalent flows in vehicles per hour) starting at 6 a.m. and continuing until 9 p.m. The data were aggregated to overcome (or diminish) the effect of the few outliers or missing observations present in the one-minute lane measurements. Four sites were chosen for data sources, three are four-lane sections and between junctions, labeled as 4757A, 4762A, 4767A and the remaining site, 4802B, is a three-lane carriageway within a motorway junction site. [Figure 1](#) shows the location of the three between-junction sites. Data were collected for all 4 sites for between 15 and 25 days in each of the months of August, September, and October 1997. This provided 184 days of traffic flows spread over 4 sites and 3 months. [Figure 2](#) gives a typical flow profile for a day at one of the sites.

SUMMARY OF FORECASTING METHODS USED

Many studies have attempted to forecast traffic flows using a variety of techniques. Some have used computerized models of the network that represent the actual movement of traffic. On a simple level, the TRANSYT program (Vincent et al. 1980) contains a technique for predicting future downstream arrivals at signalized links in a traffic network. More complicated approaches involve the computerized simulation of individual vehicles moving through a traffic network (Morin et al. 1996; Algers et al. 1997).

The second group of work has attempted to model traffic flow as a time series of observations. Many well-recognized statistical models can be fitted to historical time series data and then used to produce short-term (usually one-step or two-steps-ahead) forecasts. Moorthy and Ratcliffe (1988) produced time series forecasts for an area of West Sussex, and Smith and Demetsky (1997) demonstrated application of a time series model (among others) to forecast traffic volumes on a freeway in Northern Virginia.

A third, more recent, direction is the use of artificial neural networks that can be trained to recognize complex (nonlinear) patterns in historic traffic flows and identify them in unseen data to produce “typical” follow-on conditions. This research has produced a large number of publications since the late 1980s, and Dougherty (1996) contains a review and extensive bibliography of such applications.¹

In this paper the four forecasting methods used were selected in an earlier study (Clark et al. 1999) to encompass a range of time-series forecasting techniques.

Naive Model

The simplest forecasting technique is to assume that the currently observed level of flow will persist into the next time period:

$$f_{t+1} = v_t \quad (3)$$

Where f_t is the forecast flow at time t ; v_t is the observed flow at time t ;

This technique forms a benchmark that any competent forecasting methodology needs to exceed. No assumptions can be made about the distribution of the residuals or forecast errors from this model.

Long-Term Memory Model

A refinement is to forecast the future level of flow as an average of current and previous levels of flow. This method uses the arithmetic mean of four previous observations.

$$f_{t+1} = 0.25v_t + 0.25v_{t-1} + 0.25v_{t-2} + 0.25v_{t-3} \quad (4)$$

Where f_t is the forecast flow at time t ; and v_t is the observed flow at time t ;

The structure of this model arises from the data format used in this paper, which comprises 15-minute observation periods, that is, a time lag of 4 provides 1 hour of data. Once again, no assumptions can be made about the distribution of the residuals or forecast errors from this model.

ARIMA Model

The next level is to assume a static structure for the period-to-period relationship in the data, but allow the strength of this relationship to vary over time. This may involve fitting a Box-Jenkins ARIMA-type model (Box and Jenkins 1976) to the series. Initial investigations indicate that in order to render the series stationary, a differenced logarithmic transformation is required.

$$(f_t - \mu_v) = \varphi_{1,t}(v_{t-1} - \mu_v) + \varepsilon_t$$

Where f_t is the forecast flow at time t ,

v_t is the observed flow at time t ,

μ_v is the mean of the observed flow,

$\varphi_{1,t}$ is a parameter to be estimated from data to time period t , and

ε_t is a random residual term.

This model is a general formulation of the previous two. Unlike the other models, the procedures used to estimate parameters in this model require certain normality assumptions for the residuals, but no assumptions are possible for forecast errors.

Nonlinear Model

Sometimes the assumption of an essentially linear relationship between two quantities, as in the previous three models, is not valid. In such cases, a nonlinear formulation of the model is required. The structure adopted here is to formulate a back-propagation neural network that relates previous levels of flow to future levels. Once again, it is not possible to explicitly derive a distribution for the residuals or forecast errors from this model.

TESTS

The application of nonparametric tests is well described in the statistical literature, and the reader is directed to these texts if further explanation is required.

Signs Test

One of the features of a series of errors from a well-behaved forecasting model is that it should contain a similar number of positive and negative observations. The assumption underlying this test is that the number of, say, positive errors is shown as a binomial distribution. The parameters of this distribution are the number of trials as $(n-m)$, where n is the number of observations and m is the number of ties (i.e., the original and forecast values are the same) and the probability of success is half. The term success is commonly used when discussing the binomial distribution, but the term has no pejorative meaning here. Once an observed number of positive errors has been found, the two-tailed probability of obtaining this number of positive errors may be calculated. This probability may then be compared to some significance level to determine whether the assumption of an equal number of positive and negative errors is valid.

For a well-behaved forecasting methodology, one would hope to be able to accept the hypothesis that there are a similar number of positive and negative residuals. This ensures that the method does not tend to systematically over- or underpredict.

Wilcoxon Test on Location

When comparing the observed and forecast series, one of these two series should not be overrepresented when considering the magnitude of the values. To test for this, the two series are merged and the observations in the merged series given ranks. The ranks associated with observations from each of the series (original and forecast) are identified and summed. If the two series values are of similar magnitudes, then these two numbers should be similar, and tables are available to test for this. A modified Wilcoxon procedure may also be applied to establish whether the location of the differences between the observed and the forecast series is zero. Here the differences are ranked, and the sum of the ranks of positive differences should be similar to the sum of ranks of negative differences. The degree to which this is the case can be tested against tabulated values.

This test measures whether the location of two distributions are the same. In this case, the two distributions could be either the observed and forecast series or the differences between the observed and forecast series. In both cases, one would hope that the tests revealed that the location of the appropriate series was the same, or zero in the case of the modified Wilcoxon procedure.

Wilcoxon Test on Variance

Rather than test whether the location of two series are similar, this test measures whether the dispersion of two series are similar. Consider the case where one series occupied the lower and upper quartile of the

merged series and the other, the middle two quartiles. Using conventional rankings, these two series would produce similar rank sum statistics and a conclusion that the location of the two series were similar would be made. It is clear, however, that in this extreme case the spread of observations is not the same. To test this, a different ranking method is deployed that spreads the lower ranks toward the ends of the series. The smallest value is given a rank of 1, the largest, 2, the second largest 3, the second lowest 4, the third lowest 5 and this pattern is repeated, moving into the center of the concatenated series. By adopting this ranking scheme, it is clear that in our extreme example the series at the extremes would have a significantly lower rank sum than the other series. This test should only be applied after determining that they have similar centrality locations.

Rank Correlation

This test enables a judgment to be made as to whether the same magnitude of observation is made at each time period. Ideally, the largest forecast is made at the same time the largest magnitude is seen in the original series and so on to the smallest magnitude of the two series. This statistic may be calculated on either the observed series or the differenced series. When applied to the differenced series, the test is focused on whether the magnitude of the changes in both observed and forecast series are seen at the same time.

In an ideal situation, the correlation would be +1. The “worst” case situation applies when there is an opposite relationship and the correlation would then be -1.

Direction of Change

Sometimes it is desirable to know whether a forecast series is generally moving in tandem with the original series. This is the equivalent of asking whether the successive differences in two series are the same. If the number of times that the direction of change for the forecast and observed series agree are counted, then this statistic should follow a binomial distribution. If the yardstick is to perform better than a random toss of a coin, then the probability of success is half. The probability of observing the number of agreements can then be calculated on this hypothesis. Before this test is applied, however, it is necessary to establish whether the occurrence of continuations or changes in direction are independent events through time (Dadkhah and Zehedi 1986). This may be tested for using a $2 \times 2 \chi^2$ contingency test but, like tests on distributional assumptions, this outcome is subject to hypothesis errors and weakens the general utility of this test.

A good forecasting method should pass the test for independence and the number of times the direction of change agrees should be greater than what would be expected through chance.

EVALUATION OF THE FORECASTING METHOD

In this section, the three strands of data, forecasting method, and goodness-of-fit measure are brought together. For each forecasting measure, the performance over all 184 days is summarized in [table 1](#). For the root mean square error, mean absolute percentage error, and rank correlation statistics, the mean and standard deviation (given in parentheses) of the statistic are presented. For the Wilcoxon tests, the number of times a significant difference is found at the 10% and 5% levels are presented. For the direction of change measure, four counts are provided and classified as to whether or not the observed changes are independent events ($p(\chi^2) > 10\%$ or $p(\chi^2) > 5\%$) and if prediction of direction change is better than an even chance ($p(\text{Bin}) < 5\%$ or $p(\text{Bin}) < 10\%$). For this last measure, the best possible performance for an individual day is ($p(\chi^2) > 10\%$ and $p(\text{Bin}) < 5\%$).

The Wilcoxon location test on the differences between successive observations failed to produce any days with significant outcomes and has not been reported in table 1.

An assessment based on the root mean square and absolute percentage error indicators suggests that the nonlinear method performs best, followed by the naive and ARIMA models with the long-term memory model performing worst. This ordering is also preserved to some extent for the rank correlation statistic on

the original and the first differenced series, although the rank correlation between the observed and forecast first differences has proved to be low across all forecasting methods. There is evidence from the test on the number of positive residuals and both the Wilcoxon tests that the distribution of one-step ahead forecasts for the nonlinear model is not in accord with those of the observed series. The naive and ARIMA models perform well at maintaining a similar distribution for the original and the forecast series. In the case of the naive method, this is not surprising since the forecast is the original series, only shifted by one time period. The test that emphasizes the ability of a forecast to predict correctly the direction of change in the original series shows the long-term memory model performs well.

ADDITIONAL DIAGNOSTIC TESTS

As mentioned in the first section of this paper, the focus here is on the evaluation stage of the performance of a forecasting method. It is correct to say that this evaluation should only be conducted once the modeler is satisfied and can demonstrate that the model chosen is appropriate for the data. This should not preclude, however, some form of ongoing model suitability evaluation.

In the earlier iterative model building process, residuals from the modeling are commonly examined to ensure that they adhere to some distributional assumptions. Of particular concern when dealing with time series data is that the residuals should not be autocorrelated and should have a constant variance. These issues are commonly covered in textbooks on econometrics (Maddala 1992; Gujarati 1995). There may be value in checking forecast errors when forecasting techniques are applied to ensure the errors have not acquired any of these features.

In performing these checks, a number of nonparametric techniques are available. As an illustration of this issue, autocorrelation may exist in the model residuals or the forecast errors. To test for first-order autocorrelation, one approach would be to establish whether there were an unreasonable number of runs of positive or negative values in the forecast errors. If there were too few runs, this would indicate positive autocorrelation, while too many runs would indicate negative autocorrelation. A slightly more complex but explicit nonparametric test for serial correlation of higher orders is given in Hoel (1984). Similarly, nonparametric approaches may be adopted to test for nonconstant variance in the forecast errors.

Returning to the example models and data used in the earlier section of this paper, the application of a runs test on the forecast errors shows that the number of days on which significant first-order autocorrelation at the 95% level was detected was low for the naive (9 days), ARIMA (7 days), and nonlinear (13 days) models but extremely high for the long-term memory model (179 days). The very high number of such days for the long-term memory model does not necessarily invalidate it because its parameter values were not estimated using a method that relies on uncorrelated residuals, but the reasons behind this feature would need to be explored.

CONCLUSIONS

Nonparametric tests are rarely used to evaluate the goodness-of-fit for a forecasting model. Given that such tests require fewer assumptions than parametric tests and that they can be correctly used with small samples, this appears to be a serious oversight. Nonparametric tests also allow for tests on the performance of a forecasting methodology without regard to the performance of other methods.

There are a wide variety of forecasting methods and tasks. It is unreasonable to assume that a forecasting methodology that is good at performing one task will necessarily be the best for other tasks. A modeler needs to make a judgment as to what is required from a forecasting method. The task is then to select or devise a goodness-of-fit measure that emphasizes the desirable properties of the forecast. Once the forecasts are known, the modeler is then able to make an objective judgment as to which method is the most appropriate. The nonparametric tests discussed in this paper are able to measure and compare different aspects of the performance of a forecasting method.

For the example given in this paper, each of the forecasting methods has its strengths. The nonlinear and naive methods are good at predicting the original level of the series, via low RMSE, MAPE, and high-rank

correlation statistics. This may be important if it is necessary to predict when the level of flow crosses some form of traffic threshold, initiating the need for outside intervention. The ARIMA method is good at reproducing the distributional aspects of the original series. The long-term memory model is good at predicting the direction of change in a series—an ability that is useful for predicting a turning movement in a series. In the context of transportation, this has particular value in forecasting the beginning or end of a period of traffic volume growth.

ACKNOWLEDGMENTS

The authors would like to thank the English Highways Authority for supplying the traffic data used in this study. They would also like to thank Mr. Stuart Beale of the Agency for his help and advice during the conduct of the research, of which this paper forms a part. The views expressed in this paper represent those of the authors only and should not be taken to be those of the Highways Agency or the Department of Transport.

REFERENCES

Algers, S., E. Bernauer, M. Boero, L. Breheret, C. Di-Taranto, M. Dougherty, K. Fox, and J. Gabard. 1997. Review of Microsimulation Models, SMARTTEST Project Deliverable D3. European Commission.

Bishop, C.M. 1995. *Neural Networks for Pattern Recognition*. Oxford, United Kingdom: Clarendon Press.

Box, G.E.P. and Jenkins. 1976. *Time Series Analysis, Forecasting and Control*. San Francisco, CA: Holden-Day.

Clark, S.D., H.C. Chen, and S.M. Grant-Muller. 1999. Artificial Neural Network and Statistical Modeling of Traffic Flows—The Best of Both Worlds. *World Transport Research, Proceedings of the 8th World Conference on Transport Research*, vol. 2. Edited by H. Meersman, E. Van de Voorde, and W. Winkelmann. Oxford, United Kingdom: Elsevier Science, Ltd.

Dadkhah, K.M. and F. Zahedi. 1986. A Nonparametric Approach to Model Evaluation. *Journal of the Operational Research Society* 37(7):696–704.

Dougherty, M.S. 1996. Investigation of Network Performance Prediction: Literature Review, Technical Note 394. Institute for Transport Studies, University of Leeds, Leeds, United Kingdom.

Hoel, P.G. 1984. *Introduction to Mathematical Statistics*, 5th ed. New York, NY: Wiley and Sons.

Gujarati, D.N. 1995. *Basic Econometrics*, 3rd ed. New York, NY: McGraw-Hill.

Maddala, G.S. 1992. *Introduction to Econometrics*, 2nd ed. Englewood Cliffs, NJ: Prentice Hall.

Maxwell, H.A. and I. Beck. 1996. Traffic Control on the English Motorway Network, *Proceedings of the Eighth International Conference on Road Traffic Monitoring and Control*, Conference Publication No. 422, Apr. 23–25 1996, 136–44.

Moorthy, C.K. and B.G. Ratcliffe. 1988. Short-Term Traffic Forecasting Using Time Series Methods. *Transportation Planning and Technology* 12:45–56.

Morin, J-M., B. Baradel, and J. Bomier. 1996. Online Short-Term Simulation and Forecast of Motorway Traffic Patterns: Field Results Obtained on ASF Network in France. *Proceedings of the Third World Congress on Intelligent Transport Systems*, Orlando, Florida.

Nuttall, I. 1995. Slow, Slow, Quick, Quick, Slow: Taking the “Stop-Start” Out of the London Orbital. *Traffic Technology International*, 1995/Winter, 46–50.

Smith, B.L. and M.J. Demetsky. 1997. Traffic Flow Forecasting: Comparison of Modeling Approaches. *Journal of Transportation Engineering* 123(4): 261–66.

Vincent, R.A., A.I. Mitchell, and D.I. Robertson. 1980. User Guide to TRANSYT Version 8, Laboratory Report 888. Transport Research Laboratory, Crowthorne, Berkshire, United Kingdom.

Address for Correspondence and End Notes

Stephen D. Clark, Morley Town Council, Town Hall, Morley, Leeds, United Kingdom LS27 9DY. Email: http://www.bts.gov/publications/journal_of_transportation_and_statistics/volume_05_number_01/html/paper_03/mailto:stephendclark@ntlworld.co.uk.

¹ For a detailed description of the technical aspects of artificial neural networks, the reader is directed to Bishop (1995).

Find this web page at:

http://www.bts.gov/publications/journal_of_transportation_and_statistics/volume_05_number_01/html/paper_03/index.html