



Estimation of gravity models by OLS estimation, NLS estimation, Poisson and Neural Network specifications

Erik Bergkvist and Lars Westin

CERUM
Working Paper No. 6:1997
Free Internet Edition
Printed edition can be bought from CERUM

*This document may be copied but not sold.
The contents must not be changed in any way.*

CERUM
Centre for Regional Science
SE-901 87 Umeå
regional.science@cerum.umu.se
Fax +46 90 786 51 21
Phone + 46 90 786 60 79
www.umu.se/cerum

UMEÅ UNIVERSITET
CERUM
Centrum för regionalvetenskap



UMEÅ UNIVERSITY
CERUM
Centre for Regional Science

Estimation of gravity models by
OLS estimation, NLS estimation, Poisson,
and Neural Network specifications

Erik Bergkvist
Department of economics, Umeå University
Lars Westin
Department of Economics and CERUM, Umeå University

November 1997

Financial support has been received from the Swedish Transport and Communications
Research Board (KFB)

Regional Dimensions
Working Paper No. 6 • 1997
ISBN 91-7191-455-2
ISSN 1400-4526

Address: CERUM, Umeå University, S-901 87 Umeå, Sweden. Telefax: +46-90-7865121
Internet adress: <http://www.umu.se/cerum>

Abstract

Four specifications of gravity models for freight flow prediction are compared. The traditional specification with OLS estimation is compared with non-linear least squares (NLS) estimation as well as with a model where data are assumed to be Poisson distributed. These are compared with a semi-parametric neural network model. Data consists of freight flows between Norwegian counties. The attribute describing a node is population while distance gives the friction on links of transportation. Results show that estimation with OLS and NLS is inferior to Poisson and Neural Network specifications. However, the Poisson model, although advantageous compared to OLS, may still be improved upon. The semi-parametric Neural Network does require less of these restrictions to hold and also outperforms the others as a tool for forecast in terms of Root Mean Square Error (RMSE). The NLS model although showed the best performance when estimated on known data.

JEL classification: C45, R41

Keywords: Gravity model, Transportation, Freight flows, Spatial interaction, OLS, Poisson-regression, Non-linear regression, Neural network.

1. INTRODUCTION

In the process of infrastructure investment analysis, a crucial point is the forecast of flows on links after an improvement. In freight flow analysis, the impact on the flows in the network is connected with the size of the investment and the complexity of the network. If the investment has a critical impact on the flows, the change may be substantial and hard to predict by linear models. In the field of regional economics, the gravity model has gained wide acceptance as a reasonable although simple model of flows between nodes in a network, c.f. Haynes, K.E. and A.S. Fotheringham (1984) or Sen, A. and T.E. Smith (1995).

Traditionally the gravity model has been estimated by OLS regression. However, in large networks data often consists of zero flows between some nodes. Zero flows are not easily handled with OLS estimation but may be dealt with in a Poisson model. Whether data are Normally or Poisson distributed is an empirical question. Although interaction data often are close to a Poisson distribution, this may still not be the ideal distribution describing existing and future flows. We also compare the traditional Poisson model with an equally specified non-linear system, but with a different loss function. In this paper, we therefore compare these three models with a fourth more general specification, the feed forward back propagation (BP) neural network (NN). The hypothesis to be tested is if the BP-NN is able to predict flows better than the other three methods.

The use of NN in forecasts of spatial interaction gains increased popularity. Earlier, Nijkamp et al. (1996) compared logit and NN in the case of transport mode choice while telecommunication flows were analysed by Fischer and Sucharita (1994). Hence, the focus has mainly been on the application of existing NN rather than development of NN for spatial interaction analysis. Often, as in this study, the possibilities of “traditional” techniques are compared with the performance of NN. Current knowledge seems to be that NN has a tendency to perform better while the disadvantage of NN is the difficulty to interpret the parameters of the model and to derive it from models of transportation behaviour. Our contribution in this paper is to further explore the possibilities of NN and compare it with OLS and NLS-estimation as well as with a Poisson model. We moreover calculate the numerical elasticities of the NN specification.

The paper is structured as follows. In section two, the gravity model and different ways to specify and estimate it are discussed. Our data is described in section three and results are presented in section four. Conclusions and comments are finally presented in the last section.

2. ESTIMATION OF GRAVITY MODELS

The gravity model as most often formulated becomes (c.f. Sen, A. and T.E. Smith (1995)).

$$X_{rs} = A O_r^a D_s^b \exp(I c_{rs}) \exp(\mathbf{e}_{rs}) \quad (1)$$

Here an error term is added in a way that will make OLS estimation possible, if instead \mathbf{e}_{rs} would be assumed to enter (1) additively we would get Non-Linear least squares (NLS). In both cases \mathbf{e}_{rs} is assumed to be normally distributed and $E(\mathbf{e}_{rs})=0$. In (1), the flow between nodes r and s is a function of the attributes of the nodes given by O_r and D_s while affinity between nodes are given by c_{rs} . Parameters to be estimated are $A, \mathbf{a}, \mathbf{b}$ and I . The model is estimated in a linear form by OLS when logarithms are taken on both sides, rendering.

$$\ln X_{rs} = \ln A + \mathbf{a} \ln O_r + \mathbf{b} \ln D_s + I c_{rs} + \mathbf{e}_{rs} . \quad (1)$$

Apparently, this is impossible when flows (X_{rs}) are zero. If the OLS model should be retained, three methods are commonly used to deal with the problem of zero flows. Aggregation of the network until all flows are positive is the from statistical point of view best solution. However, this may not always be possible due to empirical and infrastructure evaluation reasons. In this case zero flows may either be excluded during estimation or replaced by a very small flow. In the second case information is decreased while in the third case desinformation is increased. Neither is thus very satisfying.

A Poisson formulation of the model has no problems with the zero flows. A Poisson formulation of the gravity model would be

$$E[X_{rs} | O_r, D_s, c_{rs}] = A O_r^a D_s^b \exp(I c_{rs}) \quad (2)$$

The critical question is if the Poisson distribution gives a better performance compared to the normal distribution assumption of $\ln(X_{rs})$ derived from (1).

The new competitor to the regression formulations above, is the set of neural network models. Here we use a back propagation neural network (BP-NN) which consists of two levels. The network in the core has the following structure

$$X_{rs} = \sum_{i=1}^M w_i \frac{e^{z_i} - e^{-z_i}}{e^{z_i} + e^{-z_i}}; \quad z_i = w_{O_i} O_r + w_{D_i} D_s + w_{C_i} C_{rs} \quad (3)$$

Data are the same as in (1) and (2) but here we have two sets of parameters. The weights/parameters w_{D_i} , w_{O_i} , w_{C_i} , and w_i are estimated so that the Root Mean Square Error (RMSE) is minimised, for a given M . The parameter M is on the other hand a free parameter to be set by the researcher after evaluation of performance has been done.

That is, the size of the neural network is not a priori defined since the final form of (3) is given by the trial-and-error process at the BP level. The function $\frac{e^{z_i} - e^{-z_i}}{e^{z_i} + e^{-z_i}}$ is called the transfer function. Here the *TanH* transfer function is used, which besides a logistic (the sigmoid) transfer function is the most commonly used. However, so far the weights have not been found to have a direct intuitive interpretation so it serves little purpose to present and interpret them. For a more detailed description of the BP network, see e.g. Rumelhart et al. (1986).

The estimation of the OLS, NLS and Poisson specifications of the gravity model is straightforward and can not be performed in to many ways. Decisions regarding which variables to include, if a constant should be used and not much more gives one unique optimal solution. In the case of a BP-NN specification, the optimal solution is not given directly. A number of free parameters, especially in the gradient descent learning algorithm, have to be set by the researcher. Each has to be varied until an as optimal value as possible is found. So far, no derived rules or analytical solutions exist as guidance. Due to the nonlinearities and non-monotonous properties of most NN specifications, optimal values change with changes in other free parameters. Illustrations of this may be found in Bergkvist & Suurmond (1996). These properties usually prohibit localisation of the global optimum. Hence, one may just hope to be close “enough”. For our purpose it is however not of main interest to reach the global optima. If we may find that BP-NN has a probability to perform better this is enough for our current purpose. On the other hand, if this not is found the NN specification still could be better than OLS and Poisson regression for this type of problems. We perhaps just haven’t

found the right parameter combination. In this case, the difference between the methods has to be evaluated against the time they take to estimate.

3. THE NORWEGIAN FREIGHT DATA

Our data consists on freight flows in whole tons of general cargo between nineteen Norwegian counties in the year 1988. The flow matrix is produced by TØI. Data characterising the counties, i.e. the nodes of the network, are total population. Population gives information on the size of the counties and potential demand but is of course not a perfect determinant of general cargo. However, population data is usually available and it is in a first step of interest to analyse the performance of different models on simple information. In a second step, it may be interesting to study if any of the specifications is sensitive to specific types of information.

The total number of observations, i.e. flows on links, is 361, which includes flows within counties as well as all zero flows. Friction between counties is measured by distance as the crow flies and is within counties arbitrary set to thirty kilometres.

The observations are randomly divided into two sets of size 122 and 239 which becomes the test and training set respectively. This division into a test and a training set is necessary when the performance of the models in forecasting are evaluated. In the case of NN, the performance on the model estimation set (train set) gives to little information on how it will perform on previously unseen data. The statistical properties of NN are more or less unknown and makes theoretical inference impossible. Hence, the performance on the train set must be controlled against the performance on the test set.

Table 1. Descriptive information of the Norwegian freight flows.

Set	Whole set $N=361$				Train set $N=239$				Test set $N=122$			
Var.	X	O	D	C	X	O	D	c	X	O	D	c
Mean	198.48	219684	219684	778.34	216.93	212833	223798	783.70	162.34	233107	211626	767.84
Std.Dev	745.60	104911	104911	695.54	837.87	100505	107859	692.34	520.70	112263	98815	704.50
Min	0	74654	74654	30	0	74654	74654	30	0	74654	74654	30
Max	6531	451099	451099	2831	6531	451099	451099	2831	3071	451099	451099	2777

One way of understanding this is to resemble it with OLS dummy regression. One may always fit as many dummy variables to the OLS model until a perfect model fit is obtained within the train set, but when one goes out of the sample, model performance may in such cases rapidly decrease. A similar risk is latent in NN since most NN are parameter “intensive” which makes it easy to totally fit, or over fit a data set, without the statistical evaluation possible in e.g. OLS. Descriptive information regarding the three sets is obtained in

Table 1. As can be seen there the differences between the train and test set seem not to be that big. Which indicates that the random division gave an even result.

4. RESULTS FROM ESTIMATION AND FORECASTING

In the evaluation of the three specifications, RMSE is used. However, it is also of interest to see if the models capture different parts of the observations to the same extent. It is also possible that a model performs better on known data than in the forecast. We show residual plots for each model for both the training and testing sets. Figure one and two gives the residuals from the OLS estimation and forecast on the test set.

Figure 1 Residual plot for the training set with OLS estimation. Residuals versus actual flows.

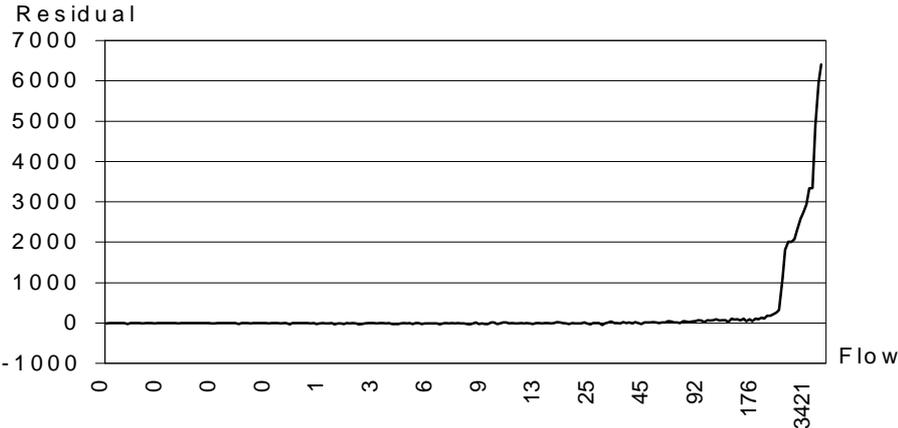
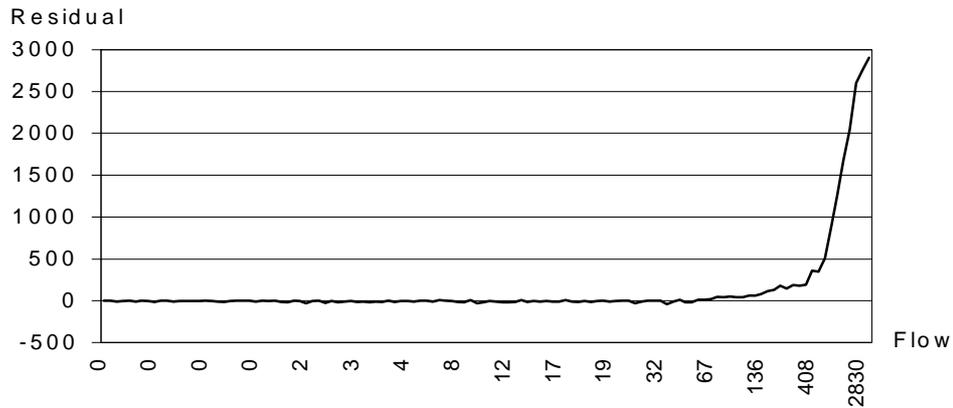


Figure 2 Residual plot for the test set with OLS estimation. Residuals versus actual flows.



The OLS specification performs quite well on smaller flows but has problems with the larger ones. It also performs better on the test set compared to the training set in terms of largest residuals.

The NLS specification gives a somewhat different result as is shown in figures three and four below. The NLS performs better on larger flows. Especially on the test set the specification

Figure 3 Residual plot for the training set with NLS estimation. Residuals versus actual flows.

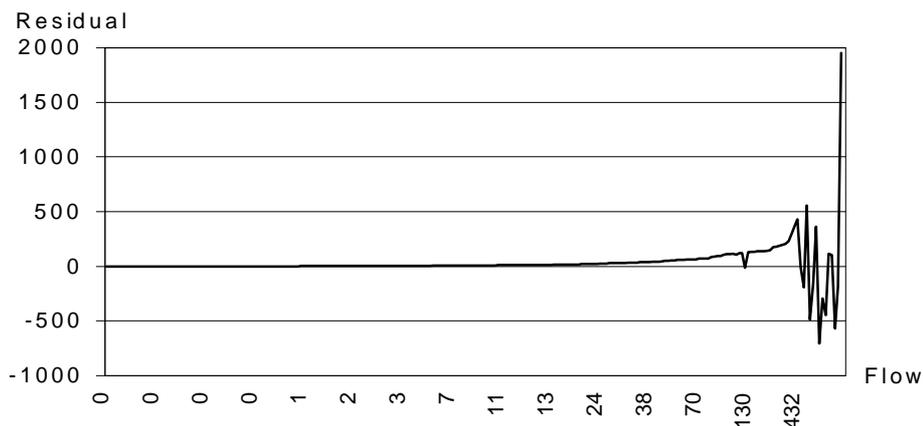
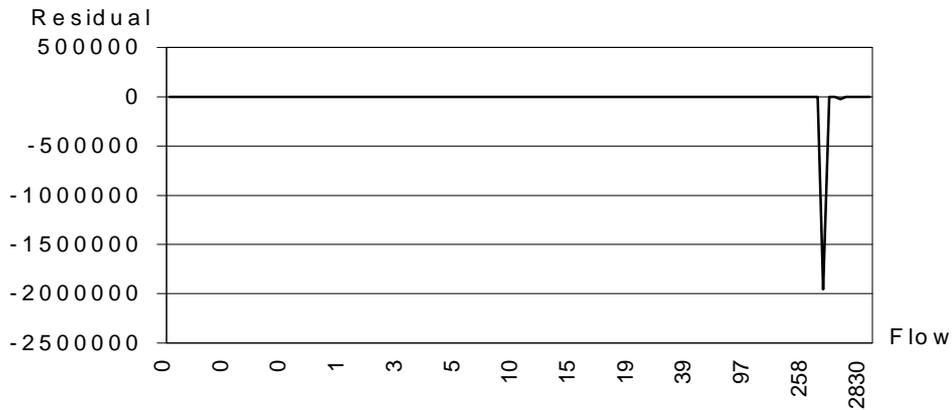


Figure 4 Residual plot for the test set with NLS estimation. Residuals versus actual flows.



seems to be better than the OLS version. However, when it comes to out of sample performance on the test set, the NLS totally misses some of the larger flows. Hence, it performs worse compared to the OLS model in this case.

Figure 5 Residual plot for the train set with a Poisson specification. Residuals versus actual flows.

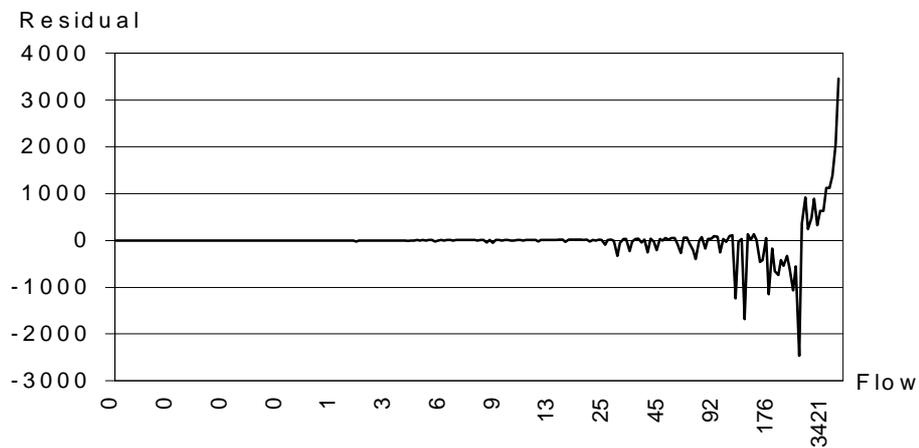
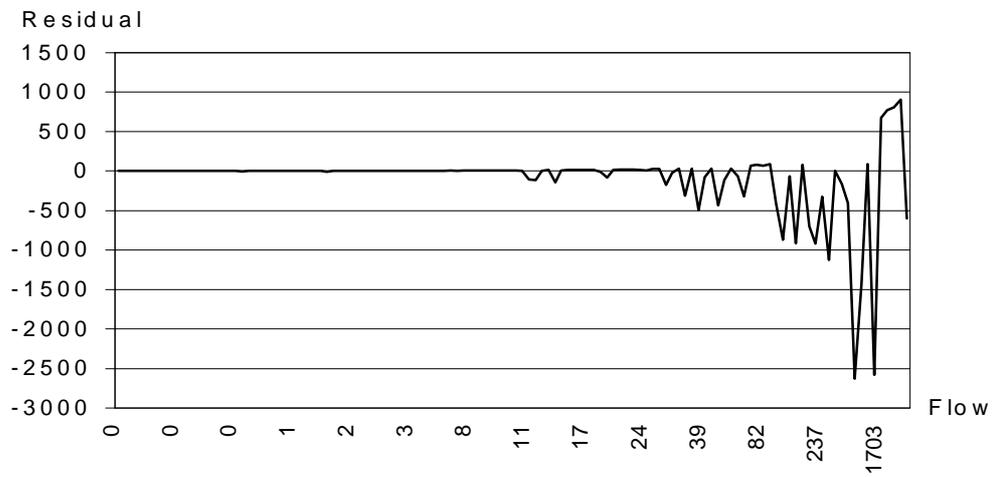


Figure 6 Residual plot for the test set with a Poisson specification. Residuals versus actual flows.



The Poisson model is the one with the most even results between the test set and the training set. Both in terms of RMSE and prediction error versus flow size.

Figure 7 Residual plot for the training set with a BP-NN specification. Residuals versus actual flows.

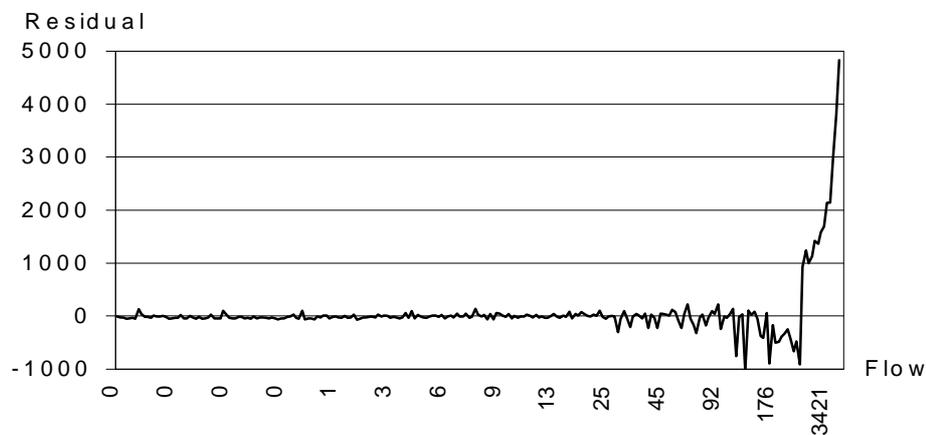
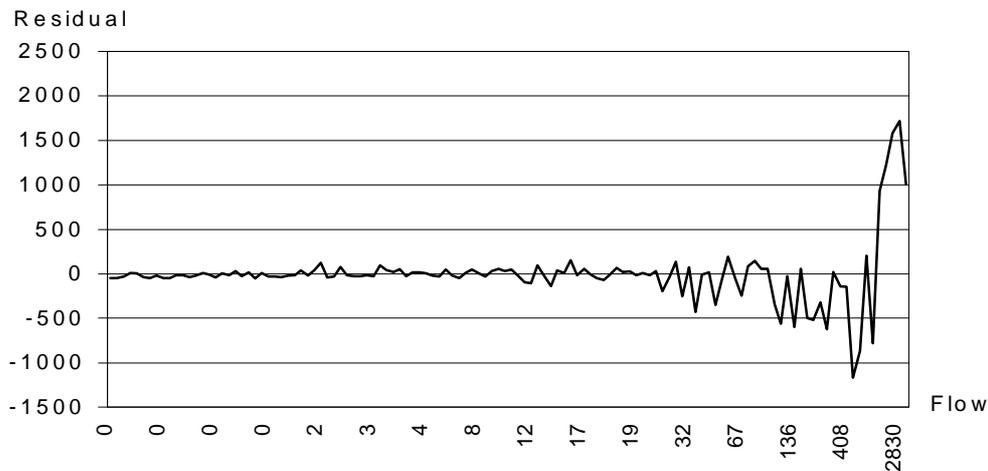


Figure 8 Residual plot for the test set with a BP-NN specification. Residuals versus actual flows.



Due to reasons given in section two above, only unique estimators exist in the OLS and Poisson regressions. In the BP-NN case, numerous results exist. A presentation here of all would not fill any purpose. Hence, only the best result is shown here, others can be found in the appendix. The parameters of the OLS, NLS and Poisson estimations are given in Table 2. No relevant parameters for the BP-NN are obtained from the estimations.

Table 2. Parameters of OLS, NLS and Poisson regressions*.

Parameter	OLS	NLS	Poisson
Constant	-10,30 (-2,11)	3,70	-2,80
Population Origin	0,61 (2,16)	-12,10	0,16
Population Dest.	0,59 (2,14)	13,10	0,73
Km	-0,25 (-8,99)	-0,26	-1,50

*t-values in parenthesis

We have chosen to use the RMSE measure to compare and evaluate our methods. Reasons for this is that it is a well known measure with well known properties. It also has its advantage since three of the estimators have minimisation of the squared residual sum (MSE) as the loss function. Which is not the case for Poisson maximum likelihood (ML) estimation which instead maximizes its loglikelihood function. MSE is defined as

$$MSE = \sum_{i=1}^N (y_i - \hat{y}_i)^2$$

which minimisation is the objective function this also makes the RMSE (4) a reasonable evaluation measure, however not completely fair against the Poisson ML estimator with its different loss function given in equation (5).

$$RMSE = \left(\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 \right)^{\frac{1}{2}} \Leftrightarrow \sqrt{MSE} \quad (4)$$

In table 3 below the results from the estimations are given.

$$\begin{aligned} \max \ln L = \max \sum_i & [-\exp(\ln A + \mathbf{a} \ln O_{ir} + \mathbf{b} \ln D_{is} + \mathbf{I} c_{irs}) \\ & + y_i (\ln A + \mathbf{a} \ln O_{ir} + \mathbf{b} \ln D_{is} + \mathbf{I} c_{irs}) - \ln y_i!] \end{aligned} \quad (5)$$

In the BP-NN case only the best result is given. The rest may be found in appendix. There we also show the intervals in which we have varied the free parameters and which free parameters we have varied.

Table 3. Root Mean Square Error (RMSE) for the different methods.

Data set	Estimator			
	OLS	NLS	Poisson	BP - NN
Train set	838	168	408	574
Test set	520	177194	443	341

Table 3 confirms that the NLS model performed best on the training set while the neural network performed best on the test set. A comparison of the residual figures in figures 3 and 4 confirms that NLS was especially good at large flows in the training set.

Figure 9 Residual plot for the training set. All specifications.

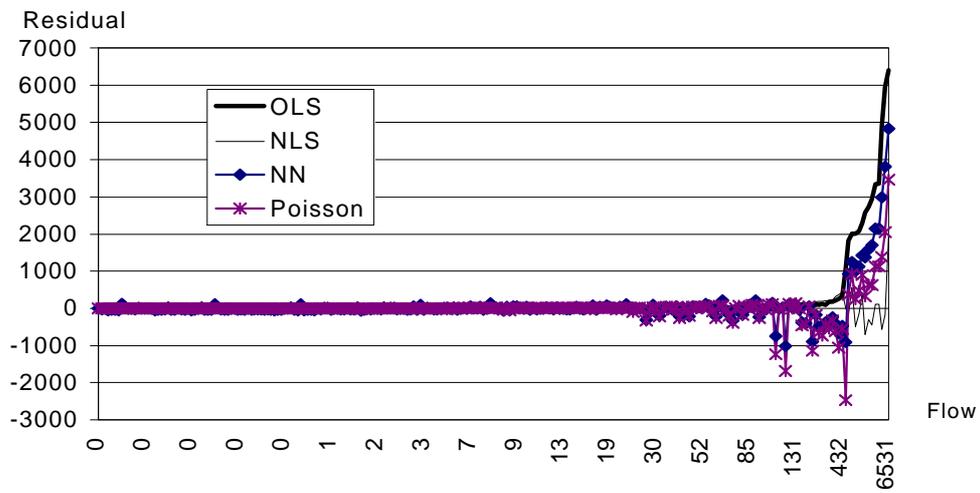
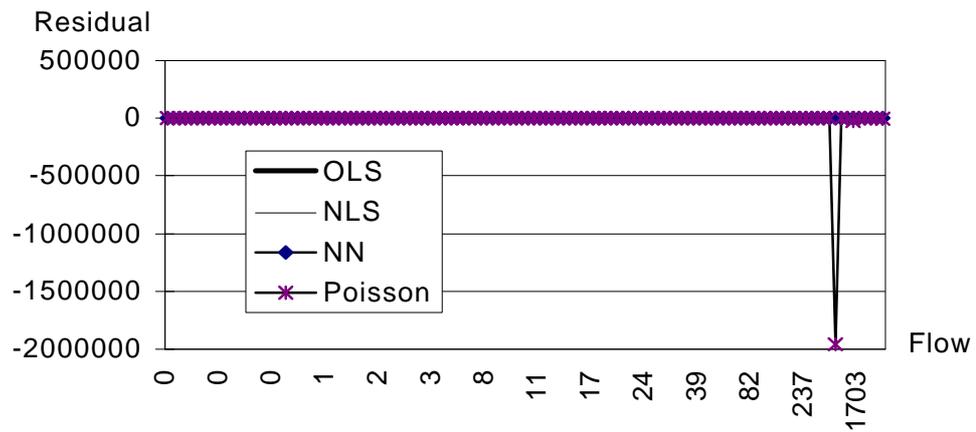


Figure 10 Residual plot for the test set. All specifications.



For the neural network specification, analytical elasticities do not “exist”. It is hard to do and the expressions one may end up with are complicated and cumbersome to use. We therefore have derived these numerically around the mean of every explanatory variable while holding the others constant. The input variable is varied 10 percent up and down from the mean in steps of one percent. From those data a mean elasticity may be calculated.

Table 4 Analytical** and numerical* elasticities.

	OLS**	NLS**	Poisson**	BP-NN*
Population Origin	0,61	-12,10	0,16	0.54
Population Dest.	0,59	13,10	0,73	-1.58
Km	-0,002	-0,26	-1,5	0.002

The positive distance elasticity in the neural network case is the most odd result. The negative impact of population in the NLS and BP-NN cases should be discussed. An obvious reason is that population alone is not the sole determinant of freight flows, access to harbours and such things may also have a strong impact. Which may be treated with dummies. Another reason is the fact that larger regions have a tendency to be more self-sufficient compared to smaller regions. A fact that introduces non-linearities, which may be shown as negative signs on the potential measures.

5. FINAL COMMENTS

In this paper we have compared four ways of estimating and forecasting freight flows. Although the BP-NN model outperformed the three other models in the forecast test, it is however too early to make any definite conclusions. The three regression models may be improved by introduction of dummy variables, a process that would resemble the handicraft associated with the calibration of neural networks. However, the BP-NN would also make use

of the extra information provided by such dummies and would most probably also improve its performance.

Interesting enough, the NLS model performed best in the estimation phase although it completely missed in the forecasting.

REFERENCES

- Bergkvist, E., and Suurmond, R.T. (1996) *Artificial neural networks and statistical approaches to classifying remotely sensed data*, IIASA WP-96-131
- Erlander , S. And N.F. Stewart (1990) *The Gravity Model in Transportation Analysis- Theory and Extensions*. Utrecht: VSP
- Fischer, M.M. and Sucharita, G. (1994) Artificial neural Networks: A new Approach to Modelling Interregional Telecommunication flows. *Journal of Regional Science*, Vol. 34, No. 4, 1994
- Greene, W.H. (1993) *Econometric Analysis*, second ed., Macmillan publishing company
- Haynes, K.E. and A.S. Fotheringham (1984) *Gravity and Spatial Interaction Models*. Beverly Hills: Sage Publ.
- Johansson, B. and L. Westin (1994) Affinities and frictions of trade networks. *The Annals of Regional Science*. Vol 28, pp. 243-261.
- Nijkamp, P. et al. (1996) Modelling Inter-Urban Transport Flows In Italy., TI 96-60/5, Tinbergen Institute, The Netherlands
- Rumelhart, D. E., McClelland, J. L. (editors) (1986). Parallel distributed processing: explorations in the microstructure of cognition. Volume I, Chapter 8. MIT Press.
- Sen, A and T.E. Smith (1995) Gravity Models of Spatial Interaction Behaviour. Berlin: Springe-Verlag.
- SINTEF (1996) *Freight Transportation and Distribution for Manufacturing and Trade in the Barents Region*. Trondheim: STF38 F96607
- TOI (1988) Norwegian domestic trade flows between counties 1988.

APPENDIX

Table A1. Influence on RMSE for test set when the number of iterations and the number of processing elements (M) are changed. tanh, Delta rule, L.Coeff. = 0.3, Momentum = 0.4, L.Coeff. Ratio = 0.5.

PE/Iterations	1	2	3	5	10	20	25	30	35	40
50.000	509	615	520	604	544	565	526	621	481	523
100.000	507	496	502	505	497	496	508	508	504	503

Table A2. Influence on RMSE when the number of iterations and size of momentum term is varied. TanH, Delta rule. No convergence for values > 0.9 . PE=2.

Mom./Iter.	0.05	0.1	0.2	0.4	0.6	0.7	0.8	0.9
50.000	521	620	620	615	607	603	602	615
100.000	500	499	499	496	492	490	491	495

Table A3. Influence on RMSE when number of iterations and size of learning coefficient is varied. TanH, Delta rule, Momentum=0.7, PE=2

L.Coeff./Iter.	0.3	0.6	0.75	0.9	1.2
50.000	615	615	618	682	607
100.000	496	475	471	476	476

Table A4. Influence on RMSE when Epoch size and number of iterations are varied. TanH, Momentum=0.7,L.Coeff.=0.75, PE=2. Norm.Cum. Rule.

Epoch./Iter.	100	50	5	2	1
50.000	-	484	608	614	638
100.000	-	492	486	478	459

Table A5 Influence on RMSE when number of iterations and size of learning coefficient ratio is varied. TanH, Delta rule, Momentum=0.7,L.Coeff.=0.75, PE=2.

It./L.Coeff.R.	25	29	30	31	35	40	50	100
0.1	502	418	421	425	438	447	459	468
0.5	545	349	341	459	545	393	618	471

Table A6. Influence on RMSE when number of iterations and size of learning coefficient ratio is varied. TanH, Delta Rule. Constant L.Coeff. = 0.75. Momentum = 0.7

It./L.Coeff.R.	30	32	35	37	40	50	69
0.1	343	341	346	350	355	376	411

Table A7. Influence on RMSE when number of iterations and size of learning coefficient ratio is varied. TanH, Norm.Cum. rule., Epoch size=1,Constant L.Coeff. = 0.75. Momentum = 0.7

It./L.Coeff.R.	25	29	30	31	32	35	50
0.1	511	373	361	360	361	365	386

REGIONAL DIMENSIONS

Working Papers

- No. 1 Einar Holm och Ulf Wiberg (Red.), *Samhällseffekter av Umeå universitet*, 1995.
- No. 2 Örjan Pettersson, Lars Olof Persson & Ulf Wiberg, *Närbilder av västerbottningar - materiella levnadsvillkor och hälsotillstånd i Västerbottens län*, 1996.
- No. 3 Jeanette Edblad, *The Political Economy of Regional Integration in Developing Countries*, 1996.
- No. 4 Lena Sahlin och Lars Westin, *Prissättning av subventionerad kultur. Vilka är de internationella erfarenheterna?*, 1996.
- No. 5 Lars Westin och Mats Forsman, *Regionerna och finansieringen av infrastrukturen: Exemplet Botniabanan*, 1997.
- No. 6 Erik Bergkvist and Lars Westin, *Estimation of gravity models by OLS estimation, NLS estimation, Poisson, and Neural Network specifications*, 1997.