



## THE RELIABILITY OF FORECASTS IN TRANSPORTATION

Mindaugas Mazūra<sup>1</sup>, Olga Fadina<sup>2</sup>

Vilnius Gediminas Technical University, Plytinės g. 27, LT-2040, Vilnius, Lithuania

Received 2002 06 27; accepted 2002 10 01

**Abstract.** Major problems of forecasting the economic characteristics of transportation (i.e. the amount of freight and passengers carried, the turnover rate of freight and passengers, etc. in transportation as a whole and in particular areas using various transport facilities) are demonstrated. Methods for predicting the development of transportation based on multidimensional regression and correlation analysis and realizing mathematical models for choosing linear and non-linear regression equations, more accurately approximating the empirical data, are presented.

The research conducted has demonstrated that the most reliable forecasts may be made when the methods of choosing the proper non – linear regression equation described in Section 2 of the present paper are used.

**Keywords:** reliability, forecasting, multidimensional regression, correlative analysis, mathematical models, approximation, empirical data, correlative relationship, linear correlation coefficient.

### 1. Introduction

Forecasting of transportation is one of the major goals of a multistage complex program of transport development in Lithuania. It is one of the most complicated problems, being, however, paid too little attention in special literature. Unlike the problem of planning it has been analysed only by a few researchers. Therefore in the present paper an emphasis is placed on the urgent problems of transportation associated with making forecasts and their reliability.

The problem of forecasting may be formulated as follows: let  $Y = \{y(t_i)\}$  be one – dimensional time series of statistical observations  $t_{i+1} - t_i = const$  at the particular time moments, where  $i = 1, \dots, n$ . The problem is to find the moments  $t_j$ , where  $j = n + 1, \dots, n + T$ . Moreover, the inertia of the investigated process is assumed as keeping a general trend in time.

A level of the time series at the moment  $t$  may be determined by summing up two components:

$$y = f(\Theta, t) + \varepsilon(t), \quad (1)$$

where  $f(\Theta, t)$  is a certain constant time function (trend);  $\varepsilon(t)$  – a function assessing the effect of stochastic factors (agents);

$\Theta$  – the parameter vector to be determined.

In formula (1)  $f(\Theta, t)$  describes the main trends of the process analysed, assessing the effect of major factors on this process. This implies that if  $f(\Theta, t)$  a constant component of the process could be determined then

$\varepsilon(t)$  – the remaining component, would be random, thus proving that the process is stochastic.

In order to determine  $f(\Theta, t)$ , the regression and correlation analysis is applied.

Usually a researcher analyses a certain phenomenon quantitatively describing the value  $y$ , varying due to some factors which in turn are quantitatively determined by the values deliberately changed by the researcher seeking to cause the variation of the analysed value  $y$ .

Let us suppose that the particular value  $y_1$  of the analysed value  $y$  corresponds to the particular set of values  $(x_1, z_1, \dots, t_1)$  of the variables  $x, z, \dots, t$ , while the value  $y_2$  corresponds to another set  $(x_2, z_2, \dots, t_2)$ , etc. It means that there is a certain relationship between the variables  $x, z, \dots, t$  and the analysed value  $y$ . It may be functional relationship or correlation.

Correlational analysis allows us to determine the extent of the dependance of the value  $y$  on the variables  $x, z, \dots, t$ . Based on correlations as well as linear and non – linear correlational analysis it is possible to determine which of the factors characterized by the values  $x, z, \dots, t$ , are relevant, which have but small effect on the value  $y$  and which are completely irrelevant to this value. Clearly observed correlations allow us to make an assumption that, in this case not only correlations, but functional relationships might be found. The latter may be described by the regression equations. Having obtained an analytical expression  $y = f(x, z, \dots, t)$  for a functional relationship between a set  $(x, z, \dots, t)$  and the analysed value  $y$  in this way, it can be extrapolated, i.e.  $y$  values related to the values of the variables  $x, z, \dots, t$ , lying outside the interval of values used in the tests, may be determined.

<sup>1</sup> E-mail: Mindaugas.Mazura@ti.vtu.lt

<sup>2</sup> E-mail: Olga@lineka.lt

Extrapolation is the basis of forecasting made for the particular phenomena under varying conditions.

The analytical expression of functional relationship based on regression analysis may be sought, if the type of function could be established first. If this is not feasible, the relationship may be simulated based on the principle of simplicity of a mathematical model. Such problems make the basis for the application of mathematical simulation techniques involving the approximation of functional relationship by polynomials and probability approach.

**2. The Establishment of Linear and Non – linear Correlation between four Variables of the Regression Equation**

The application of multidimensional regression and correlation analysis for a simple case of establishing the relationship between three variables has been described [1].

However, a comparative analysis of the forecasts and statistical data revealed that the above forecasts obtained by non – linear three variables or second – order regression equations were not always reliable [2–7].

If there exists a relationship between  $t, x, z$  and  $y$ , for example,  $t$  – time periods,  $x$  – gross national product,  $z$  – volume of production or national income, with  $y$  being freight (passengers) turnover, then, when the relationship is linear, we get:

$$y' = at + bx + cz + d \tag{2}$$

while in case of a non – linear (second – order) relationship, we obtain:

$$y'' = at^2 + bt + cx^2 + dx + ez^2 + fz + g \tag{3}$$

The problem is to calculate the parameters  $a, b, c$  and  $d$  of the regression equations (2), and the parameters  $a, b, c, d, e, f$  and  $g$  of (3). This problem can be solved in a similar way that was used in [1] for the case of regression equations. Only for finding the equation (2) a system of four equations should be solved.

Let  $t = (t_1, t_2, \dots, t_n)$ ,  $x = (x_1, x_2, \dots, x_n)$ . Then the scalar product of the vector  $(t, x) = t_1x_1 + t_2x_2 + \dots + t_nx_n$ . Thus a system of equations would be as follows:

$$\begin{cases} a(t, t) + b(t, x) + c(t, z) = (t, y), \\ a(x, t) + b(x, x) + c(x, z) = (x, y), \\ a(z, t) + b(z, x) + c(z, z) = (z, y), \\ \sum_{i=1}^n (at_i + bx_i + cz_i + d) = \sum_{i=1}^n y_i, \end{cases} \tag{4}$$

while in case of seeking the equation (3), a system of seven equations should be solved. Let us write down  $at_i + b = t_i'$ ,  $cx_i + f = z_i'$ ,  $ez_i + g = y_i'$ .

Then the expression of the system will be:

$$\begin{cases} \sum_{i=1}^n t_i^3 t_i' + t_i^2 (x_i x_i' + z_i z_i' + y_i') = 0, \\ \sum_{i=1}^n t_i^2 t_i' + t_i (x_i x_i' + z_i z_i' + y_i') = 0, \\ \sum_{i=1}^n x_i^3 x_i' + x_i^2 (t_i t_i' + z_i z_i' + y_i') = 0, \\ \sum_{i=1}^n x_i^2 x_i' + x_i (t_i t_i' + z_i z_i' + y_i') = 0, \\ \sum_{i=1}^n z_i^3 z_i' + z_i^2 (t_i t_i' + x_i x_i' + y_i') = 0, \\ \sum_{i=1}^n z_i^2 z_i' + z_i (t_i t_i' + x_i x_i' + y_i') = 0, \\ \sum_{i=1}^n (bt_i + dx_i + fz_i + y_i') = 0. \end{cases} \tag{5}$$

Having solved the system (4), we will get the coefficients  $a, b, c$  and  $d$ , of the equation (2) while the solution of a system of equations (5) will allow us to obtain the coefficients  $a, b, c, d, e, f$  and  $g$  of the equation (3). Based on the above equations (2) or (3) the sufficiently reliable forecasts may be made as to what values of the function will correspond to the values of variables lying outside the statistical data.

**3. Reliability of Forecasts in Transportation**

The analysis of the forecasts of transportation development in 1999, 2000, 2005, 2010 and 2015 contained in [2] which were obtained by the methods described in [1] and their comparison with actual statistical data for 1999 and 2000 allowed us to summarize the data and make some conclusions.

**4. Conclusions**

1. The forecasting error for the time series containing the statistical data for the years 1992 – 1998 which remained unchanged ranged from 1 % to 10 %. This error is very small, therefore, the forecasts are quite reliable.

2. The errors of forecasting in the time series of statistical data for the years 1992 – 1998 which changed because of the calculation methods used by Eurostat range from 25 % to 50 %. These errors are very great, therefore the forecasts can not be reliable.

3. The effect of the gross national product and time factors [1] is best reflected in the time series referring to freight volume and turnover. For example:

The forecast of freight turnover in air transportation for 1999–2015 (mln. ton – km), with account of time and gross national product variations

92	93	94	95	96	97	98	99	00	05	10	15
Statistical data							Forecast				
6.0	5.0	8.0	4.0	3.0	3.5	3.3	4.04	4.15	4.32	5.39	6.54
6.6	5.0	8.0	4.0	3.0	4	3	3.5	4			

The following regression equation is used in the forecast:

$$y = \left(-7.288E-08\right)t_i^2 + \left(-0.2691823\right)t_i + \left(3.3E-09\right)x_i^2 + \left(5.02E-16\right)x_i + 2.32$$

where  $t_i$  is the time factor of transportation development trends, while  $x_i$  – the factor associated with gross national product,  $-7.288E-08 = -7.288 \cdot 10^{-08}$ , for all the equations given below.

The forecast of freight turnover in railroad transportation for 1999–2015 (mln. ton – km), with account of time and gross national product variations

92	93	94	95	96	97	98	99	00	05	10	15
Statistical data							Forecast				
11337	11030	7996	7220	8103	8622	8265	9099	9316	9777	11717	13798
11337	11030	7996	7220	8103	8622	8265	7849	8919			

The following regression equation is used in the forecast:

$$y = \left(-6.485E-07\right)t_i^2 + \left(-408.841\right)t_i + \left(5.46E-06\right)x_i^2 + \left(3.8E-13\right)x_i + 6076.19$$

The forecast of freight transportation by air for 1999 – 2015 (thous. tons), with account of time and gross national product variations

92	93	94	95	96	97	98	99	00	05	10	15
Statistical data							Forecast				
3.3	2.8	4.4	2.6	1.9	2.8	2.8	3.30	3.43	3.85	4.98	6.18
3.3	2.8	4.4	2.6	1.9	2.8	2.8	3.0	3.3			

The following regression equation is used in the forecast:

$$y = \left(-7.27E-08\right)t_i^2 + \left(-0.159715\right)t_i + \left(2.648E-09\right)x_i^2 + \left(5.03E-16\right)x_i + 1.628$$

The forecast of freight transportation by sea for 1999 – 2015 (mln. tons), with account of time and gross national product variations

92	93	94	95	96	97	98	99	00	05	10	15
Statistical data							Forecast				
5.1	5.5	4.9	5.8	4.7	4.5	4.2	4.96	5.02	5.20	5.64	6.37
5.1	5.5	4.9	5.8	4.7	4.5	4.2	4.3	4.5			

The following regression equation is used in the forecast:

$$y = \left(-7.287E-08\right)t_i^2 + \left(-0.251872\right)t_i + \left(2.641E-09\right)x_i^2 + \left(5.02E-16\right)x_i + 3.773$$

The forecast of freight transportation by water for 1999 – 2015 (mln. tons), with account of time and gross national product variations

92	93	94	95	96	97	98	99	00	05	10	15
Statistical data							Forecast				
6.5	6.2	5.5	6.3	5.3	5.2	5.4	5.4	5.5	5.8	6.5	7.6
6.5	6.2	5.5	6.3	5.3	5.2	5.5	5.1	5.4			

The following regression equation is used in the forecast:

$$y = \left(-7.292E-08\right)t_i^2 + \left(-0.285237\right)t_i + \left(3.3E-09\right)x_i^2 + \left(5.01E-16\right)x_i + 3.74$$

4. Time series referring to the forecast of the amounts of passengers and their turnover reflects the effect of gross national product and time factors only to small extent. For example:

The forecast of passenger transportation by Lithuanian public transport for 1999 – 2015 (mln. of passengers), with account of time and gross national product variations

92	93	94	95	96	97	98	99	00	05	10	15
Statistical data							Forecast				
997.1	816.5	788.5	694.7	593.1	532.3	514.8	591.5	622.5	710.7	957.5	1219
997.1	816.5	788.5	694.7	609.4	551.4	516.2	471.9	383.2			

The following regression equation is used in the forecast:

$$y = \left(-1.223E-08\right)t_i^2 + \left(-36.07968\right)t_i + \left(5.854E-07\right)x_i^2 + \left(-1.8E-16\right)x_i + 228.665$$

The forecast of passenger transportation by public road transport for 1999 – 2015

(mln. of passengers), with account of time and gross national product variations

92	93	94	95	96	97	98	99	00	05	10	15
Statistical data							Forecast				
973.0	789.9	768.1	676.0	578.0	519.4	502.2	560.8	586.4	653.4	863.7	1087
973.0	789.9	768.1	678.2	593.5	537.1	502.1	458.3	372.7			

The following regression equation is used in the forecast:

$$y = \left(-1.209E-07\right)t_i^2 + \left(-35.09\right)t_i + \left(5.286E-07\right)x_i^2 + \left(-1.9E-16\right)x_i + 245.792.$$

The forecast of passenger transportation by bus for 1999–2015 (mln. of passengers),

with account of time and gross national product variations

92	93	94	95	96	97	98	99	00	05	10	15
Statistical data							Forecast				
646.9	509.1	481.5	405.7	361.6	346.8	316.3	350.2	386.9	426.4	554.7	691.5
646.9	509.1	481.5	405.7	361.6	346.8	316.3	273.5	213.3			

The following regression equation is used in the forecast:

$$y = \left(-1.032E-07\right)t_i^2 + \left(-22.238668\right)t_i + \left(3.28E-07\right)x_i^2 + \left(-3.1E-16\right)x_i + 178.155.$$

The evaluation of the effect of the gross national product and time factors is most relevant to forecasting the above parameters.

5. The research conducted has demonstrated that the most reliable forecasts may be made when the methods of choosing the proper non – linear regression equation described in Section 2 of the present paper are used.

## References

1. Mazūra M. Prediction of Major Trends of Transportation Development. *Transport*, 2002, Vol XVII, No 2, p 57-59.
2. Baublys A.; Griškevičienė D.; Žvaliauskas A. Prognostication of Transport Activity Vilnius. Lithuania Academy of Science, Book 22. Lithuania science, 1999, p 9-61.
3. Yang W.-H.; Mathur K.; R. H. Ballou. Stochastic Vehicle Routing Problem with Restocking. *Transportation Science*, 2000, 34, p 99-112.
4. Fumero F. and Vercellis C. Synchronized Development of Production, Inventory and Distribution Schedules. *Transportation Science*, 1999, 33, p 330-340.
5. Reiman M. I.; Rubio R. and Wein L. M. Heavy Traffic Analysis of the Dynamic Stochastic Inventory-Routing Problem. *Transportation Science*, 1999, 33, p 361-380.
6. Miller-Hooks E. D. and Mahmassani H. S. Least Expected Time Paths in Stochastic Time - Varying Transportation Networks. *Transportation Science*, 2000, 34, p 198-215.
7. Yang H. System Optimum, Stochastic User Equilibrium, and Optimal Link Tolls. *Transportation Science*, 1999, 33, p 354-360.