



PREDICTION OF MAJOR TRENDS OF TRANSPORTATION DEVELOPMENT

Mindaugas Mazūra

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

E-mail: Mindaugas.Mazura@ti.vtu.lt.

Received 2001 12 28; accepted 2002 03 04

Abstract. The importance 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) is demonstrated. Methods for predicting the development of transportation based on multidimensional regression and correlation analysis and realizing mathematical models for finding linear and non-linear multidimensional regression equations as well as a mathematical model for choosing linear and non-linear regression equations, more accurately approximating the empirical data, are presented.

The techniques aimed to obtain and apply the linear correlation coefficient and correlative relationship in determining the forecast accuracy is also given. The efficiency of methods, determining the linear correlation coefficient and correlative relationship, used in achieving higher accuracy of forecasts is shown.

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

1. Introduction

National transportation development program, compiled in 1993 and approved by the Lithuanian government in 1994 embraced the period up to 2010. It played an important role in the reconstruction and development of Lithuanian transportation system. This program was also approved by other countries. However, the European and other states usually revise their transportation strategies every four or five years. The same need is found in the Republic of Lithuania. A number of objective factors exist conditioning the necessity to make corrections in the national program of transportation development. Primarily, it is the policy pursued by our country to integrate into the European Community (EC) as well as new trends of transportation development both in Europe and other countries (i.e. multimodal transport development, increasing requirements raised to transport for environment protection, the growth of transportation services market, etc.).

Two out of 17 units of the national transportation development programs concern the forecasting of transport infrastructure development. They are as follows:

1. Current situation in the transportation sector and trends of its development;
2. The development of road, railroad, water and air transport (when integrating into the EC transportation network).

This makes it necessary to compile one of the most important sections of a new development program associated with the description of the present state of all means of transport and trends of their development. The latter

should include the forecasts for transporting freight and passengers by road, railroad, water and air transport in 2005, 2010 and 2015.

To predict this the application of two – dimensional regression and correlation analyses [1–3] is not sufficient. In this case transportation development should be considered in the context of the development of national economy and transport as its particular branch. Major factors as well as their interrelation and effect on the development of this or that means of transport should be evaluated. General indices characterizing the economic state of the country may be considered gross output, national income, volume of output, etc. Therefore, the selection of suitable regression equations to predict the economic characteristics of transportation (i.e. the volume of freight and passengers transit, freight and passengers turnover, etc. with respect to both transport system and its particular branch) requires the application of multidimensional regression and correlation analyses [4–8].

2. Methods of Using Multidimensional Regression and Correlation Analyses in Transportation

Suppose that the value y_1 of an economic index considered corresponds to a set of particular characteristics (x_p, z_p, \dots, t_p) of the variables x, z, \dots, t , while the value y_2 corresponds to the n – th set (x_n, z_n, \dots, t_n) . Then, the relationship between the variables x_p, z_p, \dots, t_p and the considered value y_1 may be functional or correlative. If y is any economic index of transportation, when x, z, \dots, t

are gross national output, national income, volume of output, etc, and time spaces, then having found the analytical expression of the relationship $y = f(x, z, \dots, t)$ between the set (x, z, \dots, t) and an index considered, we may determine the values of y for several periods of time in future. In the simplest case, if we suppose that there exists the relationship between three variables t, x, y (i.e. t – time spaces, x – gross national output and y – freight turnover) and if this relationship is linear then:

$$y' = at + bx + c. \tag{1}$$

If the relationship is not linear, but of the second degree, then:

$$y'' = at^2 + at + cx^2 + dx + e. \tag{2}$$

The problem arises to calculate the parameters a, b, c of the regression equations (1) or the parameters a, b, c, d, e of (2).

This problem may be solved by the least square method implying that such parameters of regression equations should be found that the analytical values of y have minimal deviation from actual statistical data y_i^* values:

$$S = \sum_{i=1}^n (y_i - y_i^*)^2 \rightarrow \min.$$

Substituting the expressions (1) or (2) for y we will get:

$$S_1 = \sum_{i=1}^n (at_i + bx_i + c - y_i^*)^2 \rightarrow \min$$

or

$$S_2 = \sum_{i=1}^n (at_i^2 + bt_i + cx_i^2 + dx_i + e - y_i^*)^2 \rightarrow \min.$$

Differentiating S_1 with respect to a, b, c , or S_2 with respect to a, b, c, d, e as well as equating parts of the derived value to zero (finding the extremum) a system of equalities will be obtained for the linear relationship (3) and for non – linear relationship (4) from the following correlation table:

t	t ₁	t ₂	...	t _n
x	x ₁	x ₂	...	x _n
y	y ₁	y ₂	...	y _n

$$\begin{cases} a \sum_{i=1}^n t_i^2 + b \sum_{i=1}^n t_i x_i + c \sum_{i=1}^n t_i = \sum_{i=1}^n y_i t_i, \\ a \sum_{i=1}^n t_i x_i + b \sum_{i=1}^n x_i^2 + c \sum_{i=1}^n x_i = \sum_{i=1}^n y_i x_i, \\ a \sum_{i=1}^n t_i + b \sum_{i=1}^n x_i + nc = \sum_{i=1}^n y_i, \end{cases} \tag{3}$$

$$\begin{cases} a \sum_{i=1}^n t_i^4 + b \sum_{i=1}^n t_i^3 + c \sum_{i=1}^n t_i^2 x_i^2 + d \sum_{i=1}^n t_i^2 x + \\ + e \sum_{i=1}^n t_i^2 = \sum_{i=1}^n y_i t_i^2, \\ a \sum_{i=1}^n t_i^3 + b \sum_{i=1}^n t_i^2 + c \sum_{i=1}^n t_i x_i^2 + d \sum_{i=1}^n t_i x_i + \\ e \sum_{i=1}^n t_i = \sum_{i=1}^n y_i t_i, \\ a \sum_{i=1}^n t_i^2 x_i^2 + b \sum_{i=1}^n t_i x_i^2 + c \sum_{i=1}^n x_i^4 + d \sum_{i=1}^n x_i^3 + \\ + e \sum_{i=1}^n x_i^2 = \sum_{i=1}^n y_i x_i, \\ a \sum_{i=1}^n t_i^2 x_i + b \sum_{i=1}^n t_i x_i + c \sum_{i=1}^n x_i^3 + d \sum_{i=1}^n x_i^2 + \\ + e \sum_{i=1}^n x_i = \sum_{i=1}^n y_i x_i, \\ a \sum_{i=1}^n t_i^2 + b \sum_{i=1}^n t_i + c \sum_{i=1}^n x_i^2 + d \sum_{i=1}^n x_i + ne = \\ = \sum_{i=1}^n y_i. \end{cases} \tag{4}$$

Solving the system of equations (3) we will calculate the equation (2) coefficients a, b, c, d and e . Basing oneself on the obtained equations (1) or (2) it is possible to predict rather accurately what values of the function will correspond to the values of variables which lie beyond the statistical data, belonging to some time periods in future.

When calculating y values for any $t_{n+1}, t_{n+2}, \dots, t_{n+p}$ we are faced with the problem how to find the values of $x_{n+1}, x_{n+2}, \dots, x_{n+p}$, because we do not know them. This problem may be solved in two ways:

1. To predict these values based on the expertise or heuristic methods;
2. To predict these values using two – dimensional regression equation $x = f(t)$ and available data correlation tables.

When choosing the analytical expression (trend) of the function (in our case, linear (1) and non – linear (2) relations are found which are then compared to establish which one more accurately represents the relation between y^* and $f(x, t)$. In our case two different analytical expressions (1) and (2) of the functions $f(x, t)$ have been found. Statistical data (from the correlation table) are then supplemented with the calculated y'_i and y''_i values.

The following calculations were made:

$$S_1 = \sum_{i=1}^n (y_i - y'_i)^2 \text{ and } S_2 = \sum_{i=1}^n (y_i - y''_i)^2.$$

t	t_1	t_2	...	t_n
x	x_1	x_2	...	x_n
y	y_1	y_2	...	y_n
y'	y'_1	y'_2	...	y'_n
y''	y''_1	y''_2	...	y''_n
$(y - y')^2$	$(y_1 - y'_1)^2$	$(y_2 - y'_2)^2$...	$(y_n - y'_n)^2$
$(y - y'')^2$	$(y_1 - y''_1)^2$	$(y_2 - y''_2)^2$...	$(y_n - y''_n)^2$

If $S_1 < S_2$, then the analytical expression $y' = at + bx + c$ is more accurate, whereas if $S_2 < S_1$, the expression $y'' = at^2 + bt + cx^2 + dx + e$ is more accurate.

However, in some cases, prognostic calculations [7,8] made according to the techniques discussed do not provide the required accuracy. Therefore, some methods of finding correlation coefficient or correlation relationship should be used. When a linear regression equation [4] is used in making predictions, then, by adding a row representing the difference between analytical and empirical y_i values to the correlation table, the correlation coefficient may be found from it. If its value is about zero, this means that the linear regression equation is acceptable for prediction, while if the correlation coefficient equals to 0.5 or is higher, another analytical expression of $y = f(x, t)$ should be chosen. Using modern application packages (Excel, Statgraf, etc.) it is possible to obtain accurate prediction results by means of mathematical simulation. This implies that such analytical expression of the function $y = f(x, t)$ is chosen that the correlation coefficient of the difference between analytical and empirical y values of a new time row approach zero. The same technique may be applied in case of the non – linear regression equation (2) used in prediction except that the correlation showing the relationship between non – linear y and x, t is determined here.

3. Conclusions

1. The application of the suggested techniques reveals that the non – linear regression equation more exactly corresponds to actual statistical data, thereby allowing a more accurate prediction to be made.

2. The suggested prediction methods deal only with the effect of two factors x and t on the resulting index y , however, these techniques may be expanded to include some more factors x, z, \dots, t .

3. It would not be reasonable to use the functions $y=f(x, z, \dots, t)$ with a very large number of parameters in predictions, since trend equations obtained in this way (particularly, with a small number of observations) will show some random deviations rather than the main trend of development.

References

1. Fumero F. and Vercellis C. Synchronized Development of Production, Inventory and Distribution Schedules. *Transportation Science*, 33, 1999, p 330-340.
2. Mazūra M. Forecasting in Transport. Vilnius: Technika, 1995. 90 p (in Lithuanian).
3. Mazūra M. Forecasting Cargo and Passenger Flows. *Transportas*, No 9. Vilnius: Technika, 1994, p 50-58 (in Lithuanian).
4. Yang W.-H., Mathur K. and Ballou R. H., Stochastic Vehicle Routing Problem with Restocking. *Transportation Science*, 34, 2000, p 99-112.
5. Reiman M. I., Rubio R. and Wein L. M., Hevy Traffic Analysis of the Dynamic Stochastic Inventory-Routing Problem. *Transportation Science*, 33, 1999, p 361-380.
6. Yang H., System Optimum, Stochastic User Equilibrium, and Optimal Link Tolls. *Transportation Science* 33, 1999, p 354-360.
7. Science and Arts of Lithuania. Transport 22 book. Baublys A., Griškevičienė D., Žvaliauskas A. Transport Operation Forecast. Vilnius: Lietuvos mokslas, 1999, p 9-61 (in Lithuanian).
8. Mazūra M. Forecasting of Transport Activitie of Using Multidimensional Regression and Correlation Analyses. *Proceedings International Conference „TRANSBALTICA-98“*. Vilnius: Technika, 1998, p 37-40 (in Lithuanian).