Prognosis of Rail Transport Demand in The Slovak Republic

Research of Input Parameters

Jozef Danis

Faculty of Operation and Economics of Transport and Communications, University of Zilina
Zilina, Slovakia
jozef.danis@fpedas.uniza.sk

Anna Dolinayova

Faculty of Operation and Economics of Transport and Communications, University of Zilina
Zilina, Slovakia
anna.dolinayova@fpedas.uniza.sk

Juraj Camaj

Faculty of Operation and Economics of Transport and Communications, University of Zilina
Zilina, Slovakia
juraj.camaj@fpedas.uniza.sk

Abstract— **Nowadays modern management cannot do without mathematical and statistical tools. Managers need a variety of market information, but they must have inside information about the company that will make their decisions relevant. This paper describes the possibility of creating a regression model of transport requirements in railway transport, depending on other economic indicators. It deals with the problem of multicollinearity among input parameters, which are needed for the creation of a prognosis.**

Keywords— regression model, multicollinearity, demand management, railway transport, transport requirements, economic indicators

I. INTRODUCTION

Globalization of economics requires rail company managers to make many decisions. A lot of information on the future development of monitored parameters is needed for each decision. Managers have to make estimates about transport performances at the creation of operation plans because these influence sales, individual cost items, utilization of rolling stocks and railway infrastructure and so on. They can estimate dominant trends and seasonal effects or identify interdependencies between variables and the estimates of necessary data or correct their own human judgment from the amount of information available to them for a certain period of time.

II. 1. INPUT DATA FOR THE CREATION OF A REGRESSIVE MODEL FOR RAILWAY TRANSPORT DEMAND

Demand is an abstract concept reflecting what individuals would like to consume under different scenarios. It is generally considered that the quantity demanded of a commodity is influenced by its price, the price of other goods, tastes and the level of income [1]. For a prognosis of rail transport demand it is necessary to research some macro- and micro-economic indicators.

When a model of rail transport demand (expressed in rail transport by transport performances statistically) is created by statistical methods it is necessary to have access to internal and external data. Internal data are usually available from company software (for example economic software, information systems and the like) [2]. Thus processed information can be used to create a regression model very simply.

Nowadays rail companies monitor a lot of basic indicators:

- number of passengers;
- transport performances in passenger or ton kilometers;
- average transport distance by type of travel in rail passenger transport or commodities in rail freight transport;
- transport performances in train kilometers;
- transport performances in gross ton kilometers;
- fulfillment of train traffic diagram;
- change of internal and external own prices;
- expenses and revenues;
- quantity and utilization of wagons, locomotives, electric or diesel motor units etc.

Rail companies sometimes carry out marketing surveys to obtain relevant data about the trip purpose, quality of services and so on.

In the case of external information, which is needed to create a prognosis for rail transport performances, a problem may occur not only in collecting it, but in its informative value and the possibility of using it in mathematical/statistical models [3].

The most important data problems include:

- absence data for a sufficiently long period of time (e.g. some outputs have stopped monitored, because they were very small, but customers have started to use this service again);

- real data comparability is not guaranteed (e.g. monitoring of macroeconomic data depends on the system of national accounts; evidence of internal data can be changed when the company purchases or changes software, etc.);

- spatial data comparability is not ensured (for example, if the administrative organization of the state changes, the data, which were evidenced before, cannot be transferred to new regions).

If a prognosis for rail transport demand for the whole territory is created, problems can "only" be related to subject comparability [4]. In the case of the creation of a prognosis for regional rail transport demand, data are missing or are not exact.

III. 2. MULTIPLE REGRESSIVE MODEL

Rail transport demand is influenced by several indicators simultaneously. Therefore it is appropriate to use multiple regressive model for this prognosis. The general form for this multiple linear regression model is:

$$y_t = b_0 + b_1 x_{1t} + b_2 x_{2t} + ... + b_k x_{kt} \tag{1}$$

where:

y_t = the values of the dependent parameter

x_t = the values of the independent parameters

b_0 = the values of the dependent parameter, if the values of the independent parameters are equal to 0

$b_1, b_2, ...b_k$ = partial regression coefficients

Since it is an additive model, estimation of partial regression coefficients can be realized by the least-squares method (LSM) by minimizing the function [5]:

$$S^2 = \sum \left(y_t - \hat{y}_t \right)^2 =$$
$$= \sum \left(y_t - \left(b_0 + b_1 x_{1t} + b_2 x_{2t} + ... + b_k x_{kt} \right) \right)^2 \to \min \tag{2}$$

where:

$\hat{y}_t$ = the estimated dependent parameter

other indicators are explained in formula 1.

When we make the first partial derivative of the function S^2 by $b_0, b_1, ... b_k$ we get k normal equations:

$$\sum_{i=1}^n y_i = nb_0 + b_1 \sum_{i=1}^n x_{1ti} + b_2 \sum_{i=1}^n x_{2ti} + b_3 \sum_{i=1}^n x_{3ti} ... + b_k \sum_{i=1}^n x_{kti}$$

$$\sum_{i=1}^n y_i x_{1ti} = b_0 \sum_{i=1}^n x_{1ti} + b_1 \sum_{i=1}^n (x_{1ti})^2 + b_2 \sum_{i=1}^n x_{2ti} x_{1ti} + b_3 \sum_{i=1}^n x_{3ti} x_{1ti} + ... + b_k \sum_{i=1}^n x_{kti} x_{1ti}$$

$$\sum_{i=1}^n y_i x_{2ti} = b_0 \sum_{i=1}^n x_{2ti} + b_1 \sum_{i=1}^n x_{1ti} x_{2ti} + b_2 \sum_{i=1}^n (x_{2ti})^2 + b_3 \sum_{i=1}^n x_{3ti} x_{2ti} + ... + b_k \sum_{i=1}^n x_{kti} x_{2ti} \tag{3}$$

$$\vdots$$

$$\sum_{i=1}^n y_i x_{kti} = b_0 \sum_{i=1}^n x_{kti} + b_1 \sum_{i=1}^n x_{1ti} x_{2ti} + b_2 \sum_{i=1}^n x_{2ti} x_{kti} + b_3 \sum_{i=1}^n x_{3ti} x_{kti} + ... + b_k \sum_{i=1}^n (x_{kti})^2$$

Before using the multiple linear regression model to create a prognosis for rail transport demand, the intensity of the dependence between the independent variable model parameters must be detected, i.e. a multicollinearity test has to be performed.

The main causes of multicollinearity can be described as follows:

- in the model there are a lot of independent parameters that describe the same factors (for example, average monthly wage and net income per household member);

- the existing dependence amongst independent parameters (for example, the development of road freight transport and GDP);

- the character of the model (for example, using the polynomial function of higher degree or physical limitation in the data model;

- the inappropriate distribution of data points (for example, the small variability of independent indicators).

There are a number of ways to measure the significance of multicollinearity. Analysis of scattering parameters is the most commonly used method (RFVk):

$$RFV_k = \frac{1}{1 - r_k^2} \tag{4}$$

where:

r_k^2 = multiple coefficient of independent Xk with other independent parameters

$X_1, X_2, ..., X_{k-1}, X_{k+1}, ..., X_P$ = other independent parameters

If a group of independent parameters are not correlated, scattering parameter RFVk is equal to 1. The model can be used to create a prognosis if RFVk is lower than 5. If the RFVk is higher than 5, the regression model is characterized by multicollinearity [6].

The problem of multicollinearity can be solved in different ways as follows [7]:

- the omission of one of the independent parameters – the simplest solution. If there is strong statistical dependence between two parameters, one of them can be left out;

- the use of a priori information – information about the change of one independent parameter (b1) in response to the change of another independent parameter (b2) if the other independent parameter changes in terms of the unit. The multiple regression model then has one less independent parameter; other parameters are estimated using the LSM and the independent parameter b1 from identified a priori information. The problem is obtaining adequate a priori information; the source can be previous empirical analysis;

- the combination of cross-sectional data and time series data – additional information is obtained from other types of data;

- the transformation of independent parameters – for example, delay of the original one-period model, e.g. using the model in the shape of the first differences;

- obtaining additional data – multicollinearity problem is a problem of choice. In practice, it is often not possible to obtain additional data, especially if they are annual data;

- methods of distorted estimation parameters – LSM is replaced by another method for the estimation of parameters, for example the ridge regression method, the principal components method, the method of general inversion and so on.

IV. DEVELOPMENT OF FACTORS THAT INFLUENCE RAIL TRANSPORT DEMAND

The main factors that influence rail transport demand are [8]:

- GDP;

- average monthly wage;

- ownership of a car;

- price of rail transport services etc.

The development of these factors from 2000 to 2015 is shown in the following table.

TABLE I.   THE DEVELOPMENT OF THE FACTORS INFLUENCING RAIL TRANSPORT PASSENGER DEMAND

Year	GDP per inhabitant	average monthly wage	number of passenger cars	average price	traffic performance of passenger car	unemploy-ment rate	traffic performance of road public transport	traffic performance of rail passenger transport
	€	€	number	€/pas. km	million passenger km	%	million passenger km	million passenger km
	x_1	x_2	x_3	x_4	x_5	x_6	x_7	y
2000	5 269.0	33.9	1 274 244	0.0351	23 929	17.9	8 435	2 870
2001	6 070.0	380.5	1 292 843	0.0351	24 056	19.2	8 253	2 805
2002	6 670.0	433.7	1 326 891	0.0351	24 978	18.5	8 236	2 682
2003	7 310.0	436.3	1 356 185	0.0440	25 224	17.4	7 757	2 316
2004	8 100.0	485.9	1 197 030	0.0439	25 332	18.1	7 882	2 228
2005	9 110.0	557.9	1 303 704	0.0439	25 824	16.2	7 525	2 166
2006	10 110.0	622.8	1 333 749	0.0439	26 342	13.3	7 665	2 194
2007	11 520.0	668.7	1 433 926	0.0439	25 994	11.0	7 596	2 148
2008	12 290.0	723.0	1 544 888	0.0452	26 395	9.6	6 446	2 279
2009	11 920.0	744.5	1 589 044	0.0463	26 538	12.1	4 538	2 247
2010	12 350.0	769.0	1 669 065	0.0463	26 876	14.4	4 436	2 309
2011	12 840.0	786.0	1 749 271	0.0463	26 887	13.2	4 611	2 431
2012	13 230.0	805.0	1 824 190	0.0536	26 935	13.6	4 584	2 459
2013	13 570.0	824.0	1 879 759	0.0536	27 155	14.1	4 338	2 485
2014	13 970	858	1 949 055	0.0536	27 251	12.8	4 495	2 583
2015	14 440.0	883	2 027 023	0.0536		11.5	4 590	3 450

[a.] Source: Processed data from the Statistical Yearbook and Transport Authority [9], [10], [11]

We researched factors that are available in material and documents of the Statistical Office of the Slovak Republic or company material available on their web pages.

V. Results and discussion

The research into multicollinearity significance was carried out by calculation of the scattering parameter RFVk for the data shown in Table 2. Scattering parameter RFVk was calculated by formula 4.

TABLE II.   The research into multicollinearity significance

parameters	correlation	RFV_k	parameters	correlation	RFV_k
x_1x_2	0.99573	234.42991	x_3x_4	0.84703	6.53723
x_1x_3	0.89193	9.25338	x_3x_5	0.84288	6.36456
x_1x_4	0.88155	8.44258	x_3x_6	-0.58505	2.40992
x_1x_5	0.96155	26.00797	x_3x_7	-0.90991	11.09987
x_1x_6	-0.83154	5.93599	x_4x_5	0.88076	8.38633
x_1x_7	-0.88420	8.63528	x_4x_6	-0.59777	2.48612
x_2x_3	0.90841	10.91862	x_4x_7	-0.79921	4.98042
x_2x_4	0.87436	7.95953	x_5x_6	-0.73930	3.83581
x_2x_5	0.96896	32.22114	x_5x_7	-0.86796	7.57325
x_2x_6	-0.78413	4.63243	x_6x_7	0.56537	2.30081
x_2x_7	-0.91094	11.22826			

[b.] Source: Authors

As can be seen in Table 2, the coefficient RFVk has a high level in many pair combinations. It is caused by the dependence of macroeconomic indicators on each other (for example, in the Slovak Republic there is strong dependence between GDP and average monthly earnings.

Scattering parameter RFVk is high in most of the studied parameters. When we increased the number of parameters (data have been investigated since 1995) the dependence among parameters was higher. It follows that the regression model can be created (in our study factors) with three independent variants (every variant has three independent parameters), namely:

1. GDP per inhabitant, number of passenger cars, average price of rail passenger transport;

2. number of passenger cars, average price of rail passenger transport, unemployment rate, and

3. average price of rail passenger transport, unemployment rate, traffic performance of road public transport.

Because the traffic performance of road public transport and the traffic performance of rail passenger transport have decreased and the reason for this decrease is an increase in the standard of living, the best model for creating a prognosis for rail transport demand is variant 2 or 3. This selection is usable only in the researched conditions. If the conditions change, new research will need to be carried out.

VI. Conclusion

Mathematical/statistical models are one of the tools that facilitate manager decisions. However, it is not always possible to establish a model that would include all the factors that affect the decision-making process, therefore the creation of the model would be very difficult. One of the solutions is to create multiple models with partial indicators. Even if a model were created with high validity, it should not be forgotten that the passengers decide on the choice of transport means. Their preferences may be different in the future than they were in the past. It is very important to know the reason for choosing transport means. Nowadays the creation of a reliable prognosis requires a combination of mathematical/statistical and expert methods.

Acknowledgment

References

[1] Button K. Transport Economics. 3rd ed., Massachusetts. Edward Elgar Publishing Limited, 2010.

[2] E. Nedeliakova, J. Sekulova, I. Nedeliak, and P. Majercak, "Services in Railway Passenger Transport and its Evaluation by Dynamic Models," in Proceedings: 18th International Conference on Transport Means Kaunas, Lithuania, pp. 227-230, October 2014.

[3] K. Teplicka, M. Daubner, and E. Augustinova, "Analysis of Causal Relationships between Selected Factors in Process of Performance Management in Industrial Companies in Slovakia", in Ekonomicky casopis, Slovak Academic Press Ltd., vol. 63, Issue 5, pp. 504-523, 2015

[4] J. P. Skrinjar, B. Abramovic, and N. Brnjak, "The Use of Game Theory in Urban Transport Planning, " in Tehnicki vjesnik-Technical Gazette, UNIV OSIJEK, vol. 22, Issue 6, pp. 1617-1621, December 2015

[5] Hindls R., Hronová S., Seger, J., and Fischer, J. Statistic for Economists, 8th ed., Professional Publishing Praha, 2007.

[6] Gozora V. Crisis Management, SPU Nitra, 2000.

[7] Hatrák, M. Econometrics, 1st ed., Iura Edition Bratislava, 2007.

[8] M. Kendra. "Integration of individual car transport and public passenger transport in cities," in Proceedings: 1st international conference on engineering and applied sciences optimization, Greece, pp. 1582-1592, June 2014

[9] The maximum prices for domestic rail passenger services. Available at: https://www.slov-lex.sk/pravne-predpisy/SK/ZZ/2005/654/20090101.html

[10] Statistical Yearbook SR 2010, 1st ed., VEDA SAV, Bratislava 2010.

[11] Statistical Yearbook SR 2015, 1st ed., VEDA SAV, Bratislava 2015.