VOL. 12, NO. 2, JANUARY 2017 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

MATHEMATICAL MODEL OF STATISTICAL IDENTIFICATION OF CAR TRANSPORT INFORMATIONAL PROVISION

Alexey Skrypnikov, Sergey Dorokhin, V. G. Kozlov and E. V. Chernyshova
Voronezh State University of Engineering Technologies, Voronezh, Russia
E-Mail: alexx1@mail.ru

ABSTRACT

In recent years, development of applied mathematical method sin various spheres of economy included the principles of research borrowed from natural sciences: evolution, selection, and adaptation. These principles allow viewing the systems as self-organizing and self-adapting to changes of external environment. Attraction of principles of development of live organisms in the study of socio-economic systems allowed studying these systems in view of complex dynamics of their functioning in time. As of now, dynamics methods of forecasting and dynamic forecasting are paid a lot of attention in Russia and abroad. Based on the method of statistical identification, which uses the theory of self-organizing systems, the authors build a many-factor model of interconnection between car transport and its system of study. The article views principles and the technology of development of this model and describes stages of its creation. The initial information for the model is given in the form of a range of average annual parameters of functioning of car transport and its informational provision, including the complex of parameters of the system of training (incoming parameters), system of management of car transport, and outgoing parameters.

Keywords: evolution, dynamics, statistical identification, car transport.

INTRODUCTION

Two criteria are set: Criterion of sustainability of model and criterion of correlation. The programme determines their minimum and this finds the sole model of optimal complexity. In number of parameters, mathematical connection of each outgoing parameter with all others is determined. For the purpose of increase of precision and regularity of the forecast, some interpolation nodes are assigned to verification sequence of data. Others form the training sequence.

Solution of the model is based on the selection principle. It is achieved with gradual complication of mathematical description and testing all possible variants of models by the stated criteria. Full description of car transport$\varphi=f(x_1, x_2, x_3, \ldots, x_n)$ is replaced by several lines of private descriptions:

First line:

$Y_1=f(x_1, x_2)$; $Y_2=f(x_1, x_3)$; …; $Y_S=f(x_{n-1}, x_n)$, where $S=C_n^2$.

Second line:

$Z_1=f(y_1, y_2)$; $Z_2=f(y_1, y_3)$; …; $Z_P=f(y_{S-1}, y_n)$, where $P=C_S^2$etc.

The level of full description is doubled with each row. Each private description is the function of only two parameters, so coefficients of private descriptions could be determined by the data of training sequence.

Excluding in term diary variables, it is possible to obtain analog of full description. Consecutive selections in its threshold row pickup a certain number of regular or stable variables.

The level of regular it wise valuated by the value of standard deviation, selected in each row (generation) of variable or for the most precise variable at the test sequence. The criterion of efficiency regularity of car transport with argumentφ_i is calculated by the following formula

$$\Delta^2(1) = \frac{\sum_{i=1}^{N_t}(\varphi_i - \varphi_i^*)^2}{\sum_{i=1}^{N_{\text{пр}}}\varphi_i^2} * 100\%$$

Where $\Delta(I)$ – standard deviation attest sequence, %;
φ_i – forecast in i-th point, $i=1, 2, 3, \ldots, N_t$;
φ_i^*– real values in the same point.

Then, private descriptions for all possible pairs of first row variables are built, which characterize functioning of car transport

$Y_i=a_{0i}+a_{1i}x_j+a_{2i}x_h+a_{3i}x_ix_h+a_{4i}x_i^2+a_5x_h^2$
Where $j=1, 2, 3, \ldots, n-1$; $h=2, 3, \ldots, n$; $i=1, 2, \ldots, C_n^2$;
n – number of variables.

The number of such equations will equal $C_n^2=(18*17)/2=153$. 18 of them are the most regular.

At each following row, 18 most regular equations of regression are selected.

Peculiarity of the forecast, received with the help of statistical identification method, consists in the fact that important variables could be excluded from the model. This shows that the excluded variables could be indirectly expressed through other variables. It is explained by close interconnection of all variables.

The level of sustainability (nonbiased ness) is evaluated by the criterion of sustainability. For its calculation, all parameters are ranked according to the volume of dispersion

$$D^2 = \frac{1}{N}\sum_{i=1}^{N}[\frac{(\varphi_i - \overline{\varphi_\iota})}{\overline{\varphi_\iota}}]^2$$

Parameters with even numbers create the first sequenceR_1, odd – R_2. After each row, selection is expressed by F equations of the following type:
$Y=f(x_i x_j)$; $Z=f(y_i y_j)$; $V=f(z_i z_j)$; $W=f(v_i v_j)$, etc.

Synthesis of equations in each row is performed twice: the first sequence is training one, and the second is verification one ($R_1=N_{tr}$; $R_2=N_{ver}$). Regression equations are denoted $Y^{/}{}_x=f(x_i x_j)$, where $1\leq k\leq F$, then the function of sequences switch positions $R_1=N_{ver}, R_2=N_{tr}$, and regression equations are denoted as $Y^{//}{}_x=f(x_i x_j)$. Each of the regression equations is evaluated by the value of standard deviations, calculated by all points of both sequences

$$\Pi_{sus} = \frac{1}{R_1 + R_2} \sum_{i=1}^{R_1+R_2} (Y_x' - Y_x'')^2$$

Equations with smaller rank $\Pi_{sus1}$are selected. Criterion of sustainability of solutions at the first row is determined as standard value of the indicator of sustainability for Fselected sustainable equations:

$$N_{sus1} = \frac{1}{F}\sum_{i=1}^{F} \Pi_{sus\ i}$$

Sustainability criteria are determined in the same way for other selection rows:

$$N_{sus\ \theta} = \frac{1}{F}\sum_{i=1}^{F} \Pi_{sus\ \theta}$$

Where θ – number of rows.
Selection scheme 18→(18*17)/2→18 (18*17)/2→…
Rule of selection stop $N_{см}$→min

Thus, creation of econometric mode of statistical identification of the system of car transport management consists of several consecutive stages.

First stage is collection of data. The data of functioning of all sub-systems are gathered. Indicators reflecting the main functions of car transport are selected. The initial information is represented by average annual parameters of these subsystems.

Complexity of management of multi-level hierarchical structures leads to necessity for implementing indicators that could be used at all levels of the management system and changing (enlargement and detalization) of certain indicators for simplifying the accounting, analysis, and planning of work of car transport enterprise management.

These stage of creation of econometric model is high-quality and logical analysis of causal connections between the selected indicators. Close connections are established, the block-scheme with all logical connections between elements and parameters is built, and the connections matrix is built. The matrix is divided into blocks and sub-matrices.

The third stage is correlation analysis. The model is built with the least square methods by the data for ten years. After distinguishing the blocks with connections, the initial matrix is found, and level of connection is found (Table 1) between parameters (by the maximum correlation coefficient) in the dependence
y (T) = X_1(T-1); y(T)=X_2(T-2); y(T)=X_3(T-3),
i.e., delay of influence of informational provision on parameters of car transport functioning is found.

Table-1. Matrix of connections in view of delay of influence of managing parameters of car transport on the outgoing parameters.

X_1	**X_2**	**X_3**	**X_4**	**X_5**	**X_6**	**X_7**	**X_8**	**X_9**
0.153	**0.187**	**0.749**	**0.513**	**0.999**	**0.999**	**0.999**	**0.983**	**0.998**
0.673	0.427	0.857	0.653	0.999	0.999	0.998	0.976	0.998
0.614	0.218	0.737	0.667	0.999	0.999	0.998	0.968	0.999
0.667	0.226	0.782	0.919	0.999	0.999	0.999	0.970	0.985
0.647	0.371	0.707	0.841	0.999	0.998	0.998	0.996	0.996
0.527	0.367	0.502	0.787	0.938	0.998	0.999	0.993	0.992
X_{10}	X_{11}	X_{12}	X_{13}	$X_{20,\ г}$	X_{21}	X_{22}	X_{23}	X_{25}
0.984	0.998	0.997	1	0.293	0.908	0.938	0.930	0.991
0.982	0.988	0.994	0.999	0.332	0.776	0.966	0.947	0.989
0.995	0.998	0.994	0.999	0.435	0.647	0.972	0.940	0.931
0.994	0.988	0.984	0.998	0.044	0.406	0.977	0.942	0.936
0.994	0.993	0.997	0.998	0.378	0.165	0.987	0.960	0.999
0.984	0.988	0.985	0.998	0.775	0.885	0.986	0.969	0.997

The process of informational provision functioning supposes that its influence on car transport and effectiveness are calculated as the sum of values of these indicators for the period from the start of functioning. The start of functioning is the moment of the car operator's receiving the information, the end - moral ageing of information. In the first year, only a certain part of

ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

information provides a certain effect - at that, it works not only in this but in the next years. Some types of information come into effect with the largest effectiveness; the others provide maximal effect only several years later – i.e., with certain delay which is called lag. Zero lag of influence means that the maximal effect is achieved in the year of information receipt, one-year lag - a year later, etc.

After finding the delay (Table-2), correlation function is built which determines influence of incoming parameters on the outgoing ones, as well as expanded correlation matrix. Negative delay of influence of one variable on the other shows its strong feedback. Calculation of influence of the viewed parameters of all outgoing parameters is too huge, so let us view only labor efficiency as an example.

Table-2. Delay of influence of managing parameters of car transport on the outgoing parameters.

X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9
1	**1**	**1**	**3**	**0**	**0**	**0**	**0**	**0**
X_{10}	X_{11}	X_{12}	X_{13}	$X_{20,\,\Gamma}$	X_{21}	X_{22}	X_{23}	X_{25}
0	0	0	0	2	1	0	0	0

Lags of influence on efficiency of cert a in parameters are shown in Figure-2. Here X_5 shows close connection between linear dependent values; X_4 means vivid dependence of labor efficiency with the lag that equals 3; X_{21} possesses a zero lag, and the connection between the values is operative; dependence $X_{20,\,\Gamma}$ shows that there are two period of changing of connections between two values: 3 and 5 years.

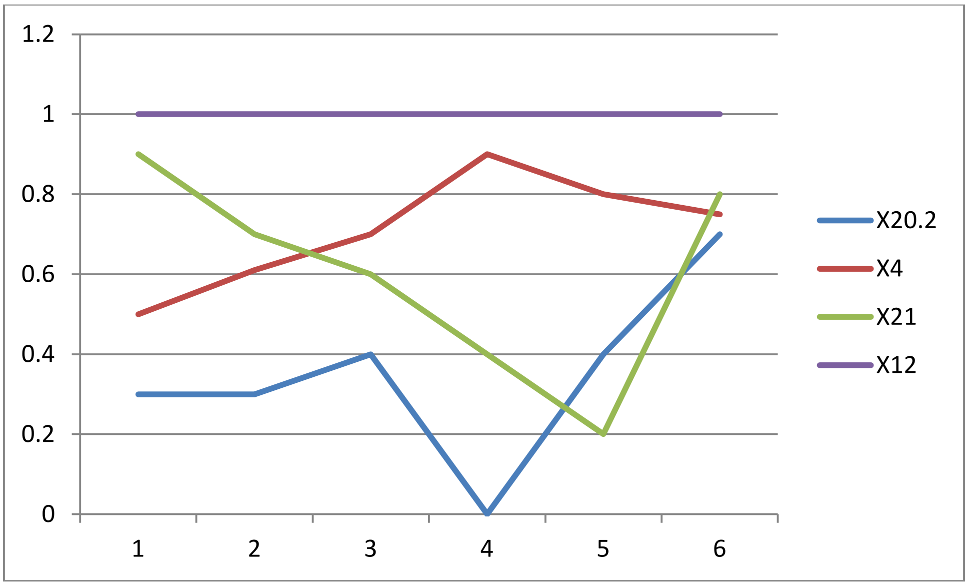

Figure-1. Dynamics of correlation coefficients in view of lags.

The fourth stage if selection of the most representative factors. Closely connected parameters which actively influence the y(T) are chosen from the matrix. At that, the arguments are classified by the time of delay. The selection criterion is coefficient of correlation that equals at least 0.5.

The program selects the variables that influence the labor efficiency in the car transport in the highest degree. The program has selected 18 parameters out of 79 – at that, the others are not excluded from consideration. Their influence is in directly taken into account if factor loads of the main variables.

The fifth stage is factor analysis of variables. All elements of correlation matrix differ by the module from the limit values - i.e., from 0 to 1. Variations of certain attributes could be viewed as a result of influence of several indicators that describe specifics of each attribute.

If there's correlation between attributes, it could be explained differently: either one attribute determines all others, or there's a general factor that is not included into consideration and which influences the correlated attributes. The method of factor analysis has to determine these general factors.

It is impossible to take into account the influence of all parameters– so the method of group accounting of arguments is used, according to which all parameters are divided into three groups. Their parameters, the loads of which are close to 1, are found. The data with noise effect and variable causing this noise are excluded. Five vivid parameters are distinguished in the first group. Parameters with close loads are united and the table of delay is built.

ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

Let us built - in 39-dimensionalenvironment - the vectors that correspond to factor loads that are caused by 39 parameters of functioning of car transport and its informational provision. Then let us draw a plane through them, so that standard deviation of each vector is minimal, and then let us project all vectors on it.

Rotating the axis, let us combine the axis with projection of one of factors on the plane. Thus, the number of changes of environment decreased by one. Having built the plane through 38 vectors and projected the monto the plane, and rotated the axes in the place for combination of the axis with one of vector projections, we receive 37 factors, etc. The described method of main components allows selecting the most important factors from a certain final multitude. The criterion of selection is minimal scatter.

Thus, leaving the unsubstantial factors, we receive five main factors. Out of the received factor loads, the matic direction and volume provide large projection on factor 3 - the factor that takes into account quality and completeness of information provided to employees of car transport. Coefficient of letting to the lime and number of cars provide large projection on factor 2 - factor that takes into account equipping and quality of technical exploitation of transport means. Additional training of drivers provides large projection of the vector on factor 4 - factor of study. And, at last, quantity of KB of information provide large projection on factor 5. Thus, five factors that determine labor efficiency are determined: technical and commercial exploitation, completeness of informational provision, training, and material resources.

The sixth stage - creation of the system of difference-differential equations that describe influence of informational provision on car transport. According to the logical chain, beginning from the lower branch, several parameters are built and connected by ties in pairs. The next level is selected at the test sequence. Also, the level of time forecasting by the program is selected.

Development of the model of interaction of car transport and its informational provision supposes establishment of dependence between the viewed parameters and, as a consequence, the possibility for determination of certain viewed values.

Functioning of car transport and its informational provision are determined by n parameters X_1, X_2, X_3, …, X_n, each of which contains indicators of time row for my ears - i.e., we have the following initial matrix n*m:

$$\begin{vmatrix} X_{11} & X_{12} & X_{13} & \dots & X_{1m} \\ X_{21} & X_{22} & X_{23} & \dots & X_{2m} \\ \dots & \dots & \dots & \dots & \dots \\ X_{n1} & X_{n2} & X_{n3} & \dots & X_{nm} \end{vmatrix}$$

Whilebuildingthemodelwesupposethateachindicat or$I_i$inthemomentj+Idepends on its previous value $X_{ij}$and on all other values of other indicators $X_K$in view of L_{xi}lag. Then, the value in yearT+1 could be expressed as
$X_i(T+1)=f[X_i(T),\ X_1(T-L_{1i}),\ X_2(T-L_{2i}),\ \dots,\ X_{i-1}(T-L_{i-1,\ i}),\ X_{i+1}(T-L_{i+1,\ i}),\ \dots,\ X_n(T-L_{ni})]$ wherei=1, 2, …, n

Using the methods of correlation analysis, let us determined the lags, setting maximal R_{max} and minimal R_{min} values of correlation coefficients. If correlation coefficient $r_{ij}<R_{min}$, random values are not correlated and are excluded from consideration; if $r_{ij}>R_{max}$, they are connected by linear dependence and one of these values is excluded. Let us find lag matrix. For that, let us provide the first dependence on the above equation in the following form:
$X_1(T+1)=f[X_1(T), X_2(T-L_{2i}), X_3(T-L_{3i}), \dots, X_n(T-L_{ni})]$

With the help of the standard program, let us determinecorrelation coefficientsbetween values $X_1$and ist values, shifted by one year, i.e.:

X_{11}	X_{12}	X_{13}	…	$X_{1,\ m-1}$	$X_{1,\ m}$	
	X_{11}	X_{12}	…	$X_{1,\ m-2}$	$X_{1,\ m-1}$	$X_{1,\ m}$

The new find correlation coefficients$r_{12}$with all possible lag values of $L_{21}$from 0 to L_{max}. Values $R_{min}<r_{11}<R_{max}$are considered to the above reasons, and the value $L_{21}$with the largest correlation coefficient is selected.

In order to build function f_i, it is necessary for it to be dependent at least on two arguments.
Two variants are possible here:

a) Variable $X_i(T+I)$ has no argument. Then, the values of variable X_i (T+1), X_i (T+2), … could be obtained by heuristic methods and are set into the computer memory from outside.

b) Variable $X_i(T+1)$ depends only on its previous value $X_i(T)$. Then, the forecast is performed by exponential leveling.

$X_i(T+I)=KX_i(T+1) + (1+K)X_i(T)$, whereK – smoothing coefficient, K=0.1.

Dependence of the function on two arguments shows its connection to other matrix rows. This connection is set during variation of the method of group accounting of arguments. Let us assume that the indicator $X_i(T+I)$ depends on n_1 indicators-arguments – for example, on $X_1(T)$, $X_2(T-2)$, $X_4(T-L_{max})$, …, $X_n(T)$. Let us compile the table of the value of these indicators with lags:

ARPN Journal of Engineering and Applied Sciences

X_{11} X_{12} X_{13}	$X_{14} \dots X_{1,m-1}$ $X_{1,m}$	
X_{11} X_{12}	$X_{13} \dots X_{1,m-2}$ $X_{1,m-1}$	$X_{1,m}$
X_{21}	$X_{22} \dots X_{2,m-3}$ $X_{2,m-2}$	$X_{2,m-1}$ $X_{2,m}$
…	…	…
	$X_{41} \dots X_{4,m-4}$ $X_{4,m-3}$	$X_{4,m-2}$ $X_{4,m-1}$ $X_{4,m}$
…	…	…
X_{11} X_{12}	$X_{13} \dots X_{n,m-2} X_{n,m-1}$	$X_{n,m}$

Here, L_{max}=4. From this table, we take rectangular matrix with the size $(n_1+1)*m_1$, the number of columns of which is $m_1 \geq m - (L_{max}+1)$.

Let us denote indicators $X_1(T+1)$ through Y, and all arguments – throughX_1, X_2, …,X_n. Then, the matrix will have the following form:

$$n1\left\{\begin{array}{l} \quad Y_1 Y_2 Y_3 \;\dots\; Y_{m1} \\ \; X_1 X_2 \quad X_3 \;\dots\; X_{1,m1} \\ \qquad\qquad \dots \\ \underbrace{X \quad X \quad X \quad \dots\; X_{n1,m1}}_{m1} \end{array}\right.$$

On the basis of this data, with the help of the least square method, let us find dependence of Yon the pair of arguments X_1, X_2:

$Y=a_1x_1^2+a_2x_2^2+a_3x_1x_2+a_4x_1+a_5x_2+a_6$

For determining the coefficients a_1 – a_6, which provide the least standard deviation of values Y, seven independent measurements are required, hence the condition:

$m_1 \geq m - L_{max} - 1 \geq 7$

denoting the first function throughY_1, and its values in interpolation nodes (withpairedvaluesX_{11}, X_{21}, X_{12}, X_{22}, etc.) – throughY_{11}, Y_{12}, …, Y_{1m1}, we build dependence on other pairs of arguments (X_1, X_3), (X_2, X_3), etc. Their number equals C_n^2. The values are brought together into matrix. The upper line includes experimental values of value Y:

$$n_2\left\{\begin{vmatrix} Y_1 & Y_2 & \dots & Y_{m1} \\ Y_{11} & Y_{12} & \dots & Y_{1\,m1} \\ \dots & \dots & \dots & \dots \\ Y_{n_2 1} & Y_{n_2 2} & Y_{n_2 3} & Y_{n_2 m1} \end{vmatrix}\right.$$

After that, we find coefficients of correlation between elements of the first and next lines and rank the matrix in the order of reduction of correlation coefficients. Then we exclude $(n_2 - n_1)$ of such lines of the ranked matrix and receive a new matrix with the size n_1*m_1:

$$n_1\left\{\begin{vmatrix} Y_1 & Y_2 & \dots & Y_{m1} \\ Z_{11} & Z_{12} & \dots & Z_{1\,m1} \\ \dots & \dots & \dots & \dots \\ Z_{n_1 1} & Z_{n_1 2} & Z_{n_1 3} & Z_{n_1 m1} \end{vmatrix}\right.$$

Then we find average value of coefficient of correlation for compilation of lines and repeat the above process to r_{cp}until r_{cp}starts reducing. After building first dependence, forecastoftheindicator$X_1$for the next year is performed. Similarly, other functions of the viewed matrix are determined. The model has the following form:

$$X_1(T+1) = f_1(X_1(T-L_{11}), X_2(T-L_{12}), \dots, X_n(T-L_{1n}))$$
$$X_2(T+1) = f_2(X_1(T-L_{21}), X_2(T-L_{22}), \dots, X_n(T-L_{2n}))$$
$$\dots$$
$$X_m(T+1) = f_m(X_1(T-L_{m1}), X_2(T-L_{m2}), \dots, X_n(T-L_{mn}))$$

The developed model is most effective during planning of the sphere eon the whole. However, the main focus during approbation is dome on determination of influence of informational provision on car transport functioning. This is due to the fact that firstly, this aspect of automobilization is the least studies; secondly, informational provision influences the most important parameters of car transport – e.g., labor efficiency; thirdly, specialties author – car transport engineer, and type of activities for improvement of the system of training and management of car transport are combined in this article. This allowed to use the accumulate knowledge and experience rationally.

REFERENCES

SkrypnikovA. V., Umarov M. M., Chernyshova E.V. 2015. Role of the state of forest car roads in provision of comfort and security of road movement in unfavorable seasons. Monthly scientific journal “Actual issues of science, technology, and production”. St. Petersburg. No. 2 (6): 66-68.

Skrypnikov A.V., Chernyshova E.V., Chernov D.A. 2015. Creation of a mathematical model of network system of video surveillance. Integration of global science and technology: imperatives of development: materials of the IV All-Russian scientific and practical conference (March 24, 2015): 2 parts, part 2 - Rostov-on-Don: Prioritet LLC. pp. 62-68.

Skrypnikov A.V., Chernyshova E.V., ShirinkinN.V., Stukalov R.V. 2015. Study of tasks of preparing complex technical provision and a generalized model of their solution. Bulletin of Voronezh State University of Engineering Technologies. - Voronezh: FSBEI HPE“VSUET. No. 4(66): 93-98.