MATRIXES LEAST SQUARES METHOD AND EXAMPLES OF ITS APPLICATION

Vladimir S. Donchenko, Inna M. Nazaraga, Olga V. Tarasova

Abstract: *General framework of Least Square Method (LSM) generalization is represented in the paper. Namely, - generalization on vectors and matrixes case of the observation. Also, some principal examples are represented in the article. These examples illustrate the advantages of LSM in the case under consideration. General algorithm LSM with matrixes observations is proposed and described in step-by-step variant. The examples of method applications in macroeconomics and TV-media illustrate the advantages and capabilities of the method.*

Keywords: *Moore-Penrose pseudo inverse, regression, least squares method, macroeconomic indicators, media indicators, prediction.*

ACM Classification Keywords: *G.3 Probability and statistics, G.1.6. Numerical analysis: Optimization, H.1.m. Models and Principles: miscellaneous.*

Introduction

The Least Squares Method (LSM) is reliable and prevalent means to solve prediction problems in applied research and in econometrics particularly. It is used in the case when the function is represented by its observations $(x_i, y_i), i = \overline{1,N}$. Commonly used statistical form of LSM is called Regression Analysis (RA). It is necessary to say, that RA is only statistical shape for representing the link between the components $x_i, y_i, i = \overline{1,N}$ in observations $(x_i, y_i), i = \overline{1,N}$. So using RA terminology of LSM for solution of function estimating problem, and correspondingly, - prediction problem, is only the form for problem discussing.

It is opportune to note, that the LSM is equivalent to Maximum Likelihood Method for classic normal regression. Linear regression (LA) within RA has the advantage of having a closed form solution of parameter estimation problem and prediction problem. Real valued functions of vector argument are the object of investigation in RA in general and in LA in particular. Such suppositions are due to technical capabilities of technique for solving optimization problems in LSM. This technique is in the essence an investigation of extremum necessary conditions. This remark is entirely true for yet another widely used assumption, namely, full column rank assumption for appropriate matrix, which ensure uniqueness of parameter estimation. It's interesting that another technique: Moore – Penrose pseudo inverse (M-Ppi) ([Moore, 1920; Penrose,1955]) provides a comprehensive study and solution of parameter estimation problem.

Below in the article the results developing M-Ppi technique are presented. These ones enable operation with matrices as with real valued vectors and in optimization problem of LSM. And, as the consequence, the results enable designing of LSM for observations with matrix components. It is interesting to note, that such results would require the development of a full arsenal M-Ppi conception for objects in matrix Euclidean spaces. But in the case under consideration manage to use M-Ppi results for Euclidean spaces of real valued vectors to solve the problem of LSM estimation for matrixes observations. Correspondent results are also represented below as well as illustration of its applications for predicting in macroeconomics of Ukraine and in estimating of TV audience.

And the remark in conclusion. Obvious advantage of matrixes LSM, besides the explicit closed estimation form, is the fact that matrixes observations preserve relationships between the characteristics of phenomenon under consideration.

Theoretical foundation: the least squares method

The LSM in its classic version – this is a way to "restore" the numeric functions $y = f(x,\theta), x \in \mathrm{X}, \theta \in \Theta$ from parametric function family, when this function is represented by this or that collection observations $(x,y), x \in \mathrm{X}, y \in R^1$. «Restored function» $\hat{y} = \hat{f}(x,\theta) = f(x,\hat{\theta})$ is defined by choosing appropriate $\hat{\theta} \in \Theta$ (estimation of parameter). The value of parameter $\hat{\theta}$ and restored function $\hat{y} = f(x,\hat{\theta})$ I call by its estimation correspondingly.

In the version of the discrete set of observations, a collection of observations (sample) is discrete: $(x_i, y_i): x_i \in \mathrm{X}, y_i \in R^1, i = \overline{1,N}$ and parameter is real valued vector: $\Theta \subseteq R^p : \theta^T = (\theta_0, \theta_2, ..., \theta_{p-1})$.

"Recovery" can be understood in different ways:

Establish the true value of the function when the model observation is $y_i = f(x_i, \theta_0), i = \overline{1,N}, \theta_0 \in \Theta$;

Approximation of the observed values $(x_i, y_i): x_i \in \mathrm{X}, y_i \in R^1, i = \overline{1,N}$ by a function from appropriate parametric family: by the choice of appropriate parameters $\hat{\theta} \in \Theta$. Such choice has to be done in such a way that the function $y = f(x,\hat{\theta})$ were the "best" to conform to the observation $(x_i, y_i), i = \overline{1,N}$.

Two previous versions can be combined in a model of observations, which can be described as a model of observations with errors:

$$y_i = f(x_i, \theta_0) + \varepsilon_i, i = \overline{1,N}, \theta_0 \in \Theta,$$

$\varepsilon_i, i = \overline{1,n}$ interpreted as errors of observations.

Last model of observations in the version, when $\varepsilon_i, i = \overline{1,N}$ - are the values of independent random variables is the subject of statistical theory, called regression analysis.

Problem "restoration of function" within the first model of observations can be reduced to the solution of simultaneous equations

$$\{y_i = f(x_i, \theta), i = \overline{1,N} \quad (1)$$

In the rest two cases the approximation criteria $\Im(\theta)$ are to be determined.

In the method of least squares such criterion is determined by the formula:

$$\Im(\theta) = \sum_{i=1}^{N} (y_i - f(x_i, \theta))^2, \quad (2)$$

Correspondingly, $\hat{\theta} \in \Theta$ defined as a solution of the optimization problem of LSM:

$$Arg\min_{\theta\in\Theta} \Im(\theta) = Arg\min_{\theta\in\Theta} \sum_{i=1}^{n} (y_i - f(x_i, \theta))^2$$

$$\hat{\theta} \in Arg\min_{\theta\in\Theta} \Im(\theta) = Arg\min_{\theta\in\Theta} \sum_{i=1}^{n} (y_i - f(x_i, \theta))^2. \quad (3)$$

It is easily to check, that the $\theta_0 \in \Theta$ in the first model (the system of equations (1)) belongs to the set of optimization solutions:

$$\theta_0 \in \underset{\theta \in \Theta}{Arg\min} \Im(\theta) = \underset{\theta \in \Theta}{Arg\min} \sum_{i=1}^{N} (y_i - f(x_i, \theta))^2 .$$

Thus, the recovery function problem for the function presented by its observations in both of the forms discussed earlier is reduced to solving an optimization problem (3).

Thus, in all cases of the recovery (estimation) problem for the function, presented by its observations $(x_i, y_i), i = \overline{1, N}$, through parameter estimation $\hat{\theta}$: $\hat{f}(x, \theta) = f(x, \hat{\theta})$, $\hat{\theta}$ can be described as a solution of the optimization problem from (3) and called LSM estimation for parameter or function correspondingly:

The widespread use of LSM in solving of restoration problem for the function, presented by its observations, is determined by its very attractive feature. It is closed form solution for the parameter estimation problem. For a family of functions

$$f(x, \theta) = \sum_{j=0}^{p-1} \theta_j f_j(x) , \theta^T = (\theta_0, \theta_2, ..., \theta_{p-1}) ,$$

$f_j(x), j = \overline{0, p-1}$ - known function of vector argument *x*

Under additional assumption *rank X*=*p*.

Closed formed solution in LR for optimization problem (3) \ is determined by formula

$$\hat{\theta} = (X^T X)^{-1} X^T Y, \tag{4}$$

where X – matrix determined by relation

$$X = \left(f_j(x_i)\right), i = \overline{1, N}, j = \overline{1, p-1} ,$$

Y - vector of observed values of the function: $Y^T = (y_1, ..., y_N)$.

Constraint *rank X*=*p*.is technical, determined only by the solution method for the optimization problem (3) and ensure uniqueness Gauss- Markov equation of the extremum necessary conditions for the functional in (2).

Functional $\Im(\theta)$ of LSM for LR turns to the form

$$\Im = \| Y - X\theta \|^2 .$$

Correspondingly, and the optimization problem (3) turns to form of

$$\underset{\theta \in \Theta}{Arg\min} \Im(\theta) = \underset{\theta \in \Theta}{Arg\min} \| Y - X\theta \|^2 . \tag{5}$$

Optimization problem (5) is essential element of pseudo inverse definition X^+ of a matrix *X* by Penrose [Penrose, 1955] (M-Ppi). By this definition pseudo inverse X^+ for $Y \neq 0$ is determined as norm minimal solution of optimization problem (5):

$$X^+ Y = \arg \min_{\hat{\theta} \in \underset{\theta \in \Theta}{Arg\min} \|Y - X\theta\|^2} \| \hat{\theta} \| .$$

This definition is only one from more than ten or more equivalent definitions of M-Ppi. M-Ppi technique enables comprehensive solution of optimization problem (3) in form (5) (see, for example [Кириченко, Донченко, 2005]):

$$\underset{\theta \in \Theta}{Arg\min} \| Y - X\theta \|^2 = X^+ Y + (E_p - XX^+) R^p = \left\{ \theta : \theta = X^+ Y + (E_p - XX^+)\beta, \beta \in R^p \right\} ,$$

with M-Ppi X^+ for matrix X . For classical conditions: under condition $rankX = p$ matrix $X^T X$ invertible

$$X^{+} = (X^{T}X)^{-1}X^{T},\ XX^{+} = E_{p} \Rightarrow E_{p} - XX^{+} = 0,$$
$$\underset{\theta\in\Theta}{Arg\min} \| Y - X\theta \|^{2} = X^{+}Y + (E_{p} - XX^{+})R^{p} = \left\{X^{+}Y\right\} = \left\{(X^{T}X)^{-1}X^{T}Y\right\}$$

Preferential use estimates from (4) and the equation of Gauss - Markov is quite restrictive in applying LSM, while advanced M-Ppi technique, as it mentioned above, enables comprehensive solution of an optimization problem (5). Such preferences of LSM users seems to be the results of habit and the fact of clarity of the source of Gauss - Markov equation as well as the fact, that M-Ppi technique require additional efforts for its mastering and applying.

Actually, directly Penrose [Penrose, 1955] pseudo inverse matrix A^{+} to $m \times n$ matrix A defined as $n \times m$-matrix, which specifies a linear operator $A^{+}: R^{m} \to R^{n}$, whose action for arbitrary $y \in R^{m}, y \neq 0$ is defined by

$$A^{+}y = \underset{x\in \underset{z\in R^{n}}{Arg\min}\|Az-y\|^{2}}{arg\min} \| x \|^{2}. \tag{6}$$

So, by this definition, $A^{+}y$ associated with SLAE (system of linear algebraic equations) $Ax = y$ and defined as smallest norm solution of the optimization problem of best quadratic approximation of the right side of SLAE values of the left side:

$$\underset{x\in R^{n}}{Arg\min} \| y - Ax \|^{2}, A \in R^{m\times n}, y \in R^{m}. \tag{7}$$

The set of all solutions of the optimization problem (7) (see, for example, [Кириченко, Донченко, 2005]) is determined by relation

$$\underset{x\in R^{n}}{Arg\min} \| y - Ax \|^{2} = A^{+}y + (E_{n} - A^{+}A)R^{n} = \left\{x : x = A^{+}y + (E_{n} - A^{+}A)v, v \in R^{n}\right\}. \tag{8}$$

M-Ppi efficiency owes singular valued decomposition (SVD) in its special tensor product form (will be denoted as SVDtp) (see, for example, [Донченко, 2011]): any $m \times n$ matrix A is represented by singularities of two matrixes $A^{T}A, AA^{T}$: by orthonormal collections of eigenvectors $v_{i} \in R^{n}, i = \overline{1,r}, u_{i} \in R^{m}, i = \overline{1,r}$ of $A^{T}A, AA^{T}$ correspondingly and common collection of correspondent nonzero eigenvalues $\lambda_{1}^{2} \geq ... \geq \lambda_{r}^{2} > 0, r = rankA$:

$$A = \sum_{i=1}^{r} \lambda_{i}u_{i}v_{i}^{T},$$
$$u_{i} = \frac{Av_{i}}{\lambda_{i}}, u_{i} = \frac{A^{T}u_{i}}{\lambda_{i}}, i = \overline{1,r}.$$

For another definitions of SVD see, for example, [Алберт, 1977].

M-Ppi definition by SVDtp among more than a dozen other equivalent, is represented by equality:

$$A^{+} = \sum_{i=1}^{r} \lambda_{i}^{-1}u_{i}v_{i}^{T}$$

M-Ppi is even more than just a tool for working with only vector objects. It provides a means for the manipulating with matrixes. Particularly, M-Ppi technique for real valued vectors enable comprehensive solution of optimization problem type of (3) in form (5) for matrix objects:

$$\underset{X\in R^{n\times p}}{Arg\min} \| Y - AX \|_{tr}^{2}, Y \in R^{m\times p}, A \in R^{m\times n}, \tag{9}$$

where the trace norm $\|\cdot\|_{tr}$ generated by trace scalar product:

$$(C,D)_{tr} = \sum_{i,j} c_{ij}d_{ij} = trC^{T}D = sum\ of\ the\ diagonal\ elements\ of\ the\ matrix\ C^{T}D$$

Full solution of the optimization problem (9) is given by the theorem 1 below (for example, [Донченко, 2011]).

Theorem 1. For any $m \times n$ matrix A

$$\underset{X \in R^{n\times p}}{Arg\min} \| Y - AX \|_{tr}^2 = X^+Y + (E_n - A^+A)R^{n\times n} = \left\{ Z : Z = X^+Y + (E_n - A^+A)V, V \in R^{n\times n} \right\}. \quad (10)$$

As in the vector case, the solutions of matrix optimization problem (10) coincide with the set of all solutions of matrix algebraic equations relatively X:

$$AX = Y, A - m \times n, Y - m \times p, X - n \times p,$$

when such solutions exist.

Optimization problems and its solutions (8), (10) for, correspondingly, vector and matrix objects, namely, the problem of the best quadratic approximation of the right part of linear equation by the left one, constitute the basis for the least squares method for vector and matrix of observations. "Vectors" or "matrixes" case for observations (x,y) means, that both its components:, x, y - are simultaneously the vectors or the matrixes correspondingly under supposition that relation between them determined by the components a $m \times n$ matrix A.

Theorem 2. Let the collection of vector pairs $(x_i, y_i) : x_i \in R^n, y_i \in R^m, i = \overline{1,N}$ or matrix pairs $(X_i, Y_i) : X_i \in R^{n\times p}, Y_i \in R^{m\times p}, i = \overline{1,N}$ are an observations of linear operator, defined by $m \times n$- matrix $A : R^{n\times p} \to R^{m\times p}$.

Then the set of LSM estimation of the operator A, is determined by the set of optimization problem solutions

$$\underset{A \in R^{m\times n}}{Arg\min} \Im(A)$$

with

$$\Im(A) = \begin{cases} \sum_{i=1}^{N} (y_i - Ax_i, y_i - Ax_i)^2, \ vector\ observations \\ \sum_{i=1}^{N} (Y_i - AX_i, Y_i - AX_i)_{tr}^2, \ \mathrm{m}atrix\ observations \end{cases}$$

is equivalent to optimization problem of the best quadratic approximation of the right hand part of algebraic equation $X^T A^T = Y^T$ by it left hand part respectively matrix A^T with matrices X, Y defined by the components of the observations accordingly to the relations:

$$X = \begin{cases} (x_1 \vdots ... \vdots x_N) - \ vector\ observation \\ (X_1 \vdots ... \vdots X_N) - matrix\ observation \end{cases}, \quad (11)$$

$$Y = \begin{cases} (y_1 \vdots ... \vdots y_N) - vector\ observation \\ (Y_1 \vdots ... \vdots Y_N) - matrix\ observation \end{cases}. \quad (12)$$

Proof. Indeed, It is easy to verify, that simultaneous equations: vectors $y_i = Ax_i, i = \overline{1,N}$ or matrixes $Y_i = AX_i, i = \overline{1,N}$, - in the observations model, are equivalent to matrix equations correspondingly:

$$(y_1 \vdots ... \vdots y_N) = (Ax_1 \vdots ... \vdots Ax_N) = A(x_1 \vdots ... \vdots x_N),$$
$$(Y_1 \vdots ... \vdots Y_N) = (AX_1 \vdots ... \vdots Ax_N) = A(X_1 \vdots ... \vdots X_N),$$

which follows from the definition of matrix algebra operations

Thus, in the notation (11), (12) observation models for both types of observations are represented by matrix equation $AX = Y$ with known matrixes X, Y and unknown matrix A.

Besides

$$\underset{A \in R^{m\times n}}{Arg\min} \Im(A) = \underset{A \in R^{m\times n}}{Arg\min} \| AX - Y \|_{tr}^2$$

So, equivalently

$$\mathop{Arg\min}_{A\in R^{m\times n}} \Im(A) = \mathop{Arg\min}_{A\in R^{m\times n}} \| AX - Y \|_{tr}^2 = \mathop{Arg\min}_{A^T\in R^{n\times m}} \| Y^T - X^T A^T \|_{tr}^2 , \qquad (13)$$

which proves the theorem.

Theorem 2. The set of all solutions for LSM - estimation of the linear operator by its vectors or matrixes observations is given by the relation:

$$\mathop{Arg\min}_{A\in R^{m\times n}} \Im(A), \Im(A) = \left\{ A : YX^+ + V(E_n - XX^+), V \in R^{m\times n} \right\} , \qquad (14)$$

Proof. The proof follows directly from theorem 1, relation (10), that describes the solution of matrix algebraic equations through obvious changes in notation and subsequent transposition using commutative property for M-Ppi for matrix and its transpose.

General algorithm of LSM with matrix observations

LSM wit matrixes observation for prediction is implemented in the usual way: by estimation of the function (operator) and using of the estimation on the appropriate argument. Observations, necessary for the estimation procedure to be applied, should be constructed on the basis of a data available. It is the first step of the algorithm proposed.

Step 1. Constructing the matrices components of observations. This step is performed on the based on statistical data by its aggregating firstly in vector and then - in matrixes $R_1, R_2, ..., R_k$. Such two – step procedure uses natural elements of phenomenon description. Vector constituents as a rule are a collection of that or those characteristics of phenomenon under consideration which corresponds to fix moments of time. These vectors constituents which correspond some “time window” are aggregated in matrix. Then the “time window” is shifted and new matrix is built, and so on.

Step 2. Constructing the observations. The matrixes $R_1, R_2, ..., R_k$ being built the observation pairs (X_j, Y_j) are built by consequence elements of $R_1, R_2, ..., R_k : (X_j, Y_j) = (R_j, R_{j+1}), j = \overline{1, k-1}$.

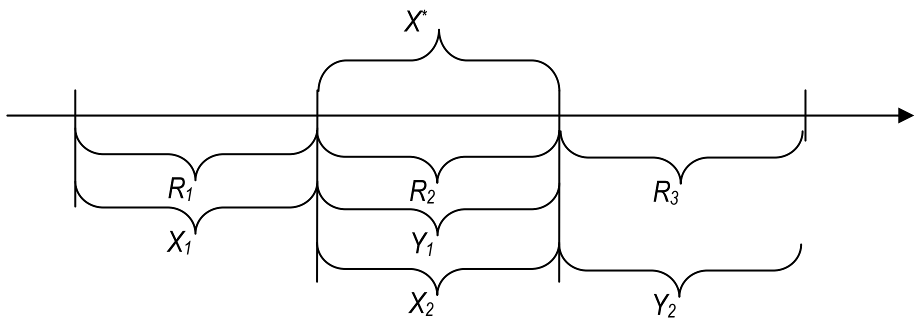

Fig. 1. Aggregated matrixes and observations

Step 3. Parametrization of the model. The relationship between matrixes elements of observations is established by matrix equation $Y = AX$ with matrix *A* as a parameter.

Step 4. LSM – estimation. The essence of this step is constructing the LSM-estimation accordingly to (14) by choosing the one with minimal norm:

$$\hat{A} = YX^+ , \qquad (15)$$

Step 5. Constructing the prediction formula. Prediction problem solution, based on the estimated operator $\hat{A}$ is standard: for any appropriate matrix argument *X** predicted *Y** is defined by relation $Y^* = \hat{A}X^*$.

Step 6. Calculations and the accuracy of prediction. The accuracy of prediction in economic research, as a rule, is estimated by formal criterion of accuracy called "absolute percentage error (*APE*)", defined by the relation $APE = \left| \frac{z_t - \hat{z}_t}{z_t} \right|, t = \overline{1,T}$, where z_t - the actual value of the index at the time t , $\hat{z}_t$ - prognostic value of the index at the time t .

It is generally accepted that the value of *APE* which is less than 10%, corresponds to high prediction accuracy, so, values from 10 to 20% is interpreted as good prediction accuracy, values from 20 to 50% are considered to be satisfactory, more than 50% - unsatisfactory prediction accuracy

There are some examples that illustrate the method below. Some more examples one can find in [Donchenko, Nazaraga, Tarassova, 2013].

Example 1: prediction economic indicators

In this example, the statistical data of the State Statistics Service of Ukraine was used [Ukrstat].

As described in [Харазішвілі, 2007], the regression methods most often used to predict of economic indicators in the normal way. In this example, the theory of matrixes LSM (Sections 1) was used.

In particular, Table 1 - 3 presents the value of gross domestic product (GDP), wages of employees (WE), final consumption expenditure (FCE), exports of goods and services (E) and imports of goods and services (I) for the 2007 – 1 quarter 2013 (1q2013) years (quarterly and annual data at current prices).

Table 1. The value of 5 indicators for 2007 - 2008 years (at current prices; mln.UAH)

	1 quarter 2007	**2 quarter 2007**	**3 quarter 2007**	**4 quarter 2007**	**Total 2007**	**1quarter 2008**	**2 quarter 2008**	**3 quarter 2008**	**4 quarter 2008**	**Total 2008**
GDP	139444	166869	199535	214883	720731	191459	236033	276451	244113	948056
WE	69078	82021	91922	108915	351936	100492	116441	121522	132009	470464
FCE	112494	130245	140935	174907	558581	161565	182154	194262	220921	758902
E	67513	79664	88491	87537	323205	88516	116640	132177	107526	444859
I	-76022	-85992	-93895	-108464	-364373	-110802	-135800	-144433	-129553	-520588

Table 2. The value of 5 indicators for 2009 - 2010 years (at current prices; mln.UAH)

	1 quarter 2009	**2 quarter 2009**	**3 quarter 2009**	**4 quarter 2009**	**Total 2009**	**1 quarter 2010**	**2 quarter 2010**	**3 quarter 2010**	**4 quarter 2010**	**Total 2010**
GDP	189028	214103	250306	259908	913345	217286	256754	301251	307278	1082569
WE	99206	111616	114251	126270	451343	114062	133690	139108	153791	540651
FCE	172426	188041	196074	216285	772826	194511	216027	232397	271295	914230
E	86994	95390	114962	126218	423564	112105	134553	145563	157144	549365
I	-92892	-96846	-116057	-133065	-438860	-114550	-131242	-156102	-179050	-580944

Table 3. The value of 5 indicators for 2011 – 1q2013 years (at current prices; mln.UAH)

	1 quarter 2011	**2 quarter 2011**	**3 quarter 2011**	**4 quarter 2011**	**Total 2011**	**1 quarter 2012**	**2 quarter 2012**	**3 quarter 2012**	**4 quarter 2012**	**Total 2012**	**1 quarter 2013**
GDP	261878	314620	376019	364083	1316600	293493	349212	387620	378564	1408889	301598
WE	135831	155367	158186	178727	628111	158145	180432	179944	199638	718159	165337
FCE	236580	268688	285548	314385	1105201	272970	311851	328173	356607	1269601	291388
E	156545	179626	184258	187524	707953	165810	181413	188467	181657	717347	162250
I	-173046	-187916	-202131	-215935	-779028	-186323	-215091	-214364	-219616	-835394	-180530

The use of the algorithm

1A. Indicators prediction for 2011-2012 years on the basis of 2007-2010 years.

1. Based on table 1 obvious way form a matrix of observations R_1, based on table 2 – matrix R_2, based on table 3 – matrix R_3.

2. Pair of input output matrix data (X_1, Y_1) will have the form (R_1, R_2).

3. Matrix $\hat{A}_1$ (see (15)), obtained from the equation $Y_1 = A_1 X_1$:

$$\hat{A}_1 = \begin{pmatrix} 0,472791 & 4,270925 & -0,44387 & 2,83749 & 4,410361 \\ -0,53049 & 3,940514 & -0,71641 & 2,130856 & 2,325999 \\ -0,79734 & 5,886174 & -0,49137 & 3,70521 & 4,549089 \\ 0,071931 & 0,956233 & 0,215418 & 0,572184 & 0,742387 \\ -0,34424 & 2,097631 & -1,97502 & -0,54289 & -0,95884 \end{pmatrix}$$

4. From the equation $Y^* = \hat{A}X^*$ calculate matrix predictive indicators Y^*, $X^* = X_2$.

$$\begin{pmatrix} 273694,0 & 338005,3 & 333617,9 & 337446,1 & 1282763,3 & 316432,6 & 399450,3 & 357960,9 & 337906,6 & 1411750,4 \\ 136422,5 & 169529,1 & 151972,4 & 164184,7 & 622108,7 & 167281,7 & 217282,3 & 168935,4 & 167030,4 & 720529,8 \\ 248255,2 & 306758,3 & 274580,8 & 292074,4 & 1121668,7 & 296834,3 & 377567,3 & 293639,5 & 294664,7 & 1262705,8 \\ 126419,3 & 145322,3 & 149113,3 & 159464,7 & 580319,6 & 145704,8 & 172400,0 & 172152,0 & 184595,4 & 674852,2 \\ -155678,3 & -169885,3 & -184891,2 & -192705,1 & -703159,9 & -170727,7 & -181818,3 & -200242,7 & -232626,0 & -785414,8 \end{pmatrix}$$

5. For matrices predictive indicators Y^* and actual indicators 2011 - 2012 years Y_2 calculate the matrix APE:

$$\begin{pmatrix} 4,51\% & 7,43\% & 11,28\% & 7,32\% & \mathbf{2,57\%} & 7,82\% & 14,39\% & 7,65\% & 10,74\% & \mathbf{0,20\%} \\ 0,44\% & 9,12\% & 3,93\% & 8,14\% & \mathbf{0,96\%} & 5,78\% & 20,42\% & 6,12\% & 16,33\% & \mathbf{0,33\%} \\ 4,93\% & 14,17\% & 3,84\% & 7,10\% & \mathbf{1,49\%} & 8,74\% & 21,07\% & 10,52\% & 17,37\% & \mathbf{0,54\%} \\ 19,24\% & 19,10\% & 19,07\% & 14,96\% & \mathbf{18,03\%} & 12,13\% & 4,97\% & 8,66\% & 1,62\% & \mathbf{5,92\%} \\ 10,04\% & 9,60\% & 8,53\% & 10,76\% & \mathbf{9,74\%} & 8,37\% & 15,47\% & 6,59\% & 5,92\% & \mathbf{5,98\%} \end{pmatrix}$$

In accordance with the submitted values matrix APE, error prediction GDP in 2011 (as a whole) was 2,57%, WE - 0,96%, FCE - 1,49%, E - 18,03%, I - 9,74%, error prediction GDP in 2012 (as a whole) was 0,20%, WE - 0,33%, FCE - 0,54%, E - 5,92%, I - 5,98%, and excess error 20% for some mentioned quarterly indicators can be explained by the fact that prediction is used, in particular, data the years of crisis 2008 - 2009.

1B. Indicators prediction for 2013 year on the basis of 2011-2012 years.

1. Data of 2011 year (Table 3) form a matrix R_1, data of 2012 year (Table 3) form a matrix R_2, data of 2013 year (Table 3) form a matrix R_3.

2. Pair of input output matrix data (X_1, Y_1) will have the form (R_1, R_2).

3. Matrix $\hat{A}_1$ (see (15)), obtained from the equation $Y_1 = A_1 X_1$:

$$\hat{A}_1 = \begin{pmatrix} 0,636101 & 0,151580 & -0,160623 & 1,242501 & 0,290005 \\ -0,120202 & 0,443779 & 0,343330 & 0,498337 & 0,172744 \\ 0,052472 & 0,519415 & 0,596880 & 0,282772 & -0,018491 \\ -0,033165 & -0,444692 & -0,527290 & 1,253027 & -0,944773 \\ 0,058681 & -0,274726 & 0,110891 & -1,420427 & -0,183491 \end{pmatrix}$$

4. From the equation $Y^* = \hat{A}X^*$ calculate matrix predictive indicators Y^*, $X^* = X_2$.

$$\begin{pmatrix} 318801,99 & 362421,85 & 393132,98 & 375805,98 & 1450162,79 \\ 179064,63 & 198412,63 & 202824,03 & 218113,67 & 798414,96 \\ 310804,97 & 353456,17 & 366941,54 & 391838,84 & 1423041,53 \\ 159803,61 & 174273,50 & 172762,31 & 145740,75 & 652580,17 \\ -197286,77 & -212712,68 & -218667,93 & -210819,92 & -839487,30 \end{pmatrix}$$

5. First column of the matrix of errors APE is calculated from the matrix of the forecast indicators values Y^* and actual data for 1q2013. Thereby, the prediction error of GDP (1q2013) - 5,7%, WE - 8,3%, FCE - 6,66%, E - 1,51%, I - 9,26%.

$$\begin{pmatrix} 5,70\% \\ 8,30\% \\ 6,66\% \\ 1,51\% \\ 9,28\% \end{pmatrix}$$

In general, comparing the results with the values of the relevant indicators the consensus prediction [Me], it can be argued about the competitiveness of the proposed article approach for forecasting macroeconomic indicators.

Example 2: prediction of TV audience performance

Media planning is based on the use of predictive indicators of TV audience. All players of the advertising market depend on the accuracy of TV audience indices predictive. In practice, five basic TV indicators are forecast:

- share of TV channel audience (share of the channel - sc) – this index determines the amount of viewers who watched TV from the total number of viewers at the investigated time period;
- rating of TV channel audience (ratings of the channel - rc) - this index determines the amount of TV audience, it takes into account the duration of watching TV every spectator in the analyzed period of time;
- TotalTV rating (rt) - this index determines the total size of the television audience, it takes into account individual TV time watching by every spectator in the analyzed time period;
- advertising TV audience rating of the channel (an advertisement rating - ra) - this index determines the size of TV advertising audience it takes into account the duration of advertisement viewing by every TV viewers;
- break-factor (bf)- this index determines the proportion of the audience that stays for advertising viewing.

Data description

We used five indicators data for 2013 year by months and three time slots (7:00-13:00, 13:00-19:00, 19:00-25:00).

The result of observations for this period of five television performance (audience share of channel sc, rating of channel rc, TotalTV rating rt, ratings of channel advertising ra and break-factor bf) forms the matrix of monthly observations $r(i) = \begin{pmatrix} r(i)_1 \\ r(i)_2 \\ r(i)_3 \\ r(i)_4 \end{pmatrix}, i = \overline{1,21}$.

Respectively $r(i)_1, i = \overline{1,21}$ — is the row-vector of monthly TotalTV raiting; $r(i)_2, i = \overline{1,21}$ — is the row-vector of monthly data of channel audience share; $r(i)_3, i = \overline{1,21}$ — is the row-vector of monthly data of channel rating; $r(i)_4, i = \overline{1,21}$ — is the row-vector of monthly advertising raiting data, $r(i)_5, i = \overline{1,21}$ — is the row-vector of monthly break-factor data. These monthly data vectors for a given period (in our case – 7 months) naturally organizing the matrix of observations $R(i), i = \overline{1{,}21}$.

Table 4. TV data perfomance by months and time slots

Period	Time slot	rt	sc	rc	ra	bf
Jan.2013	07:00 - 13:00	11,3	8,1	0,9	0,8	0,8
	13:00 - 19:00	18,9	9,1	1,7	1,3	0,8
	19:00 - 25:00	28,2	8,7	2,5	1,8	0,7
Feb.2013	07:00 - 13:00	10,9	7,4	0,8	0,6	0,8
	13:00 - 19:00	17,2	7,7	1,3	0,9	0,7
	19:00 - 25:00	27,5	8,4	2,3	1,6	0,7
Mar.2013	07:00 - 13:00	11,5	8,0	0,9	0,7	0,8
	13:00 - 19:00	17,6	8,6	1,5	1,0	0,7
	19:00 - 25:00	27,7	9,4	2,6	1,9	0,7
Apr.2013	07:00 - 13:00	9,4	8,2	0,8	0,6	0,8
	13:00 - 19:00	13,4	10,1	1,4	1,0	0,8
	19:00 - 25:00	24,7	9,9	2,4	1,8	0,7
May.2013	07:00 - 13:00	9,1	8,4	0,8	0,6	0,8
	13:00 - 19:00	13,0	10,1	1,3	1,0	0,7
	19:00 - 25:00	22,5	9,6	2,1	1,6	0,7
Jun.2013	07:00 - 13:00	8,3	7,7	0,6	0,5	0,8
	13:00 - 19:00	12,4	8,8	1,1	0,8	0,7
	19:00 - 25:00	21,6	9,4	2,0	1,4	0,7
Jul.2013	07:00 - 13:00	8,2	7,2	0,6	0,5	0,8
	13:00 - 19:00	11,9	7,9	0,9	0,7	0,7
	19:00 - 25:00	20,6	9,1	1,9	1,3	0,7

To apply the theory of pseudo inverse we use the signs of Sections 1 and construct from the observations matrix the matrix pairs of input and output data of our model. We grouped the observational data matrix $R(i), i = \overline{1{,}21}$ in the matrix R_1, R_2, R_3 as follows

$$R_1 = (r(1) \vdots \ldots \vdots r(3)),$$
$$R_2 = (r(4) \vdots \ldots \vdots r(6)),$$
$$R_3 = (r(7) \vdots \ldots \vdots r(9)),$$
$$\ldots\ldots\ldots\ldots\ldots\ldots$$
$$R_7 = (r(19) \vdots \ldots \vdots r(21)).$$

Then the matrix pair (X_1, Y_1), on which evaluation matrix of the model parameters $\hat{A}$ will be calculated from the matrix equation $Y = AX$, is as follows $(X_1, Y_1) = (R_1, R_2)$. The matrix pair (X_2, Y_2) is used to construct the forecast indicators matrix Y^* from the matrix equation $Y^* = \hat{A}X_2$ and accuracy estimation of prediction Y^* by the criterion of accuracy $APE = \left|\dfrac{Y_2 - Y^*}{Y_2}\right|$, where $(X_2, Y_2) = (R_2, R_3)$. Then

$(X_2, Y_2) = (R_2, R_3) \Rightarrow (X_3, Y_3) = (R_3, R_4)$

$(X_3, Y_3) = (R_3, R_4) \Rightarrow (X_4, Y_4) = (R_4, R_5)$

$(X_4, Y_4) = (R_4, R_5) \Rightarrow (X_5, Y_5) = (R_5, R_6)$ and so on.

The use of the algorithm

1. We constructed the matrix of monthly observations $R_1,..,R_7$ on the basis of the matrix of observations $r(i), i = \overline{1,21}$ (five basic monthly indicators in 2013(Table 4)) by grouping data.

2. A pair of input-output data matrices (X_1, Y_1) takes the form (R_1, R_2).

3. Then the matrix of estimates of the model parameters $\hat{A}$, that obtained from the equation $Y = AX \Rightarrow Y_1 = AX_1 \Rightarrow$

$$\hat{A} = \begin{pmatrix} 1.406 & -0.784 & 0.349 \\ 1.288 & -1.126 & 0.846 \\ 0.577 & -0.889 & 1.341 \end{pmatrix}$$

4. From the equation $Y^* = \hat{A}X^*$ calculate predictive indicators matrix Y^*, $X^* = X_2$.

$$\begin{vmatrix} 11,55 & 7,29 & 0,91 & 0,74 & 0,81 \\ 18,07 & 7,95 & 1,51 & 1,14 & 0,82 \\ 27,98 & 8,66 & 2,38 & 1,70 & 0,77 \end{vmatrix}$$

5. A comparison of the predictive indicators matrix Y^* and actual performance matrix Y_2 gives a matrix of errors APE:

$$\begin{vmatrix} 0,2\% & 9,2\% & 1,0\% & 1,6\% & 2,0\% \\ 2,4\% & 7,8\% & 0,3\% & 8,2\% & 15,3\% \\ 1,0\% & 8,3\% & 9,1\% & 9,3\% & 7,1\% \end{vmatrix}$$

Similarly, we have continued calculation and got a table of forecast values for five TV media indicators. In the same way we got the error matrix table APE (Table 5).

Table 5. TV data perfomance forecast фтв by month and time slot

predictive indicators matrix	Period	Time slot	rt	sc	rc	ra	bf	APE(rt)	APE(sc)	APE(rc)	APE(ra)	APE(bf)
Y_3^*	Mar.2013	07:00 - 13:00	11,6	7,3	0,9	0,7	0,8	0,2%	9,2%	1,0%	1,6%	2,0%
		13:00 - 19:00	18,1	8,0	1,5	1,1	0,8	2,4%	7,8%	0,3%	8,2%	15,3%
		19:00 - 25:00	28,0	8,7	2,4	1,7	0,8	1,0%	8,3%	9,1%	9,3%	7,1%
Y_4^*	Apr.2013	07:00 - 13:00	12,2	8,5	1,0	0,8	0,9	22,7%	3,8%	21,7%	23,3%	8,0%
		13:00 - 19:00	18,2	9,3	1,6	1,2	0,9	26,2%	8,6%	14,2%	13,5%	11,8%
		19:00 - 25:00	27,9	10,0	2,7	2,1	1,0	11,5%	0,7%	9,3%	14,9%	24,8%
	May.2013	07:00 - 13:00	7,8	7,9	0,5	0,5	0,9	15,7%	5,9%	42,5%	27,9%	11,4%
		13:00 - 19:00	12,0	9,5	0,9	0,7	1,1	8,7%	7,2%	51,0%	34,5%	30,5%
		19:00 - 25:00	23,6	9,0	2,1	1,6	0,9	4,9%	6,6%	1,0%	3,0%	18,0%
Y_5^*	Jun.2013	07:00 - 13:00	8,9	8,7	0,8	0,6	0,8	5,7%	10,8%	17,2%	16,3%	2,1%
		13:00 - 19:00	12,7	10,2	1,3	1,0	0,7	2,8%	13,0%	15,7%	16,3%	0,4%
		19:00 - 25:00	20,5	9,3	1,9	1,4	0,8	5,4%	0,2%	5,6%	1,7%	7,1%
Y_6^*	Jul.2013	07:00 - 13:00	7,6	7,3	0,6	0,5	0,8	7,6%	1,1%	5,5%	5,8%	2,3%
		13:00 - 19:00	11,6	8,2	1,0	0,7	0,8	2,8%	3,5%	3,6%	9,3%	9,0%
		19:00 - 25:00	20,7	9,2	1,9	1,4	0,7	0,3%	1,8%	2,4%	1,9%	2,4%

Y_7^*

As the *APE* errors table shows, the average annual forecast indicators error for Mar-Jul.2013 is: 5.9% – for TV channel audience share; 13.3% – for TV channel audience rating; 7.9% – for TotalTV rating; 12.5% – for TV channel advertising rating; 10.1% – for break-factor. The average prediction accuracy for all five indicators is acceptable for monthly year forecasts. However, exceeding the 10% threshold accuracy in some months is critical and shows the necessity of expert correction.

Conclusion

In the article case of matrix of observations for the arguments and values of the renewable function of the linear relationship between the components of observation has been considered.

Based on the matrixes least squares method, approach to prediction of indicators was proposed.

Testing approach with the use of statistical data of the economic and media indicators was made.

Results of prediction with available statistics were compared. The proposed approach for finding predictive values indicators is competitive.

Bibliography

[Donchenko, Nazaraga, Tarasova, 2013] Vladimir S. Donchenko, Inna M. Nazaraga, Olga V. Tarasova. Vectors and matrixes least square method: foundation and application examples.// International Journal “Information Theories & Applications”. - 2013.- Vol.20.- Number 4.-P.311 -322.

[Me] Official web-site of the Ministry of Economic Development and Trade of Ukraine [Electronic resource] – Mode of access: http://www.me.gov.ua/ control/uk/ publish/category/main?cat_id=73499. – Title from the screen.

[Moore,1920] Moore E.H. On the reciprocal of the general algebraic matrix // Bulletin of the American Mathematical Society. – 26, 1920. – P.394 -395 .

[Penrose, 1955] Penrose R. A generalized inverse for matrices // Proceedings of the Cambridge Philosophical Society 51, 1955. – P.406-413.

[Ukrstat] Official web-site of the State Statistics Service of Ukraine [Electronic resource] Mode of access: http://www.ukrstat.gov.ua. Title from the screen.

[Ukrstat] Офіційний сайт Державного комітету статистики України [Електронний ресурс] – Режим доступу: http://www.ukrstat.gov.ua. – Назва з екрану.

[Алберт, 1977] Алберт А. Регрессия, псевдоинверсия, рекуррентное оценивание.– М.: Наука.–1977.– 305 с.

[Донченко, 2011] Владимир Донченко. Евклидовы пространства числовых векторов и матриц: конструткивные методы описания базовых структур и их использование.// International Journal “Information technologies & Knowledge”.- 2011.- Vol. 5.- Number 3. P.203 - 216.

[Кириченко, Донченко, 2005] Кириченко М.Ф., Донченко В.С. Задача термінального спостереження динамічної системи: множинність розв’язків та оптимізація//Журнал обчислювальної та прикладної математики. – 2005. –№5– С.63-78.

[Кириченко, Донченко, 2007] Кириченко Н.Ф., Донченко. В.С. Псевдообращение в задачах кластеризации// Киб. и СА.- №4, 2007– С.98-122.

[Харазішвілі, 2007] Харазішвілі Ю.М. Теоретичні основи системного моделювання соціально-економічного розвитку України: Моногр. К.: ПоліграфКонсалтинг, 2007. – 324 с.

Authors' Information

Vladimir S. Donchenko - Professor, Kyiv Taras Shevchenko National University, Faculty of Cybernetics, Ukraine, e-mail: voldon@unicyb.kiev.ua;

Inna M. Nazaraga – Younger Researcher; Kyiv Taras Shevchenko National University, Faculty of Cybernetics, Ukraine, e-mail: inna_na@ukr.net;

Olga V. Tarasova – postgraduate student; Kyiv Taras Shevchenko National University, Faculty of Cybernetics, Ukraine, e-mail: olga_ta@bigmir.net.