УДК 519.711:004.8:681.3

L.V. Sarycheva[1], **Cand. Sc. (Phys.-Math.) Assoc. Prof.,**
M.V. Shamseyeva[2]

1 – State Higher Educational Institution “National Mining University”, Dnipropetrovsk, Ukraine, e-mail: sarycheval@nmu.org.ua
2 – “Malbi” company, Dnipropetrovsk, Ukraine,
e-mail: shamseeva@malbi.dp.ua

“PERMANENT ASSETS PER CAPITA” INDEX FORECASTING ON THE BASE OF REGION STATISTICAL MONITORING

Л.В. Сарычева[1], **канд. физ.-мат. наук, доц.,**
М.В. Шамсеева[2]

1 – Государственное высшее учебное заведение „Национальный горный университет“, г. Днепропетровск, Украина,
e-mail: sarycheval@nmu.org.ua
2 – компания „Malbi“, г. Днепропетровск, Украина,
e-mail: shamseeva@malbi.dp.ua

ПРОГНОЗИРОВАНИЕ ПОКАЗАТЕЛЯ „ОСНОВНЫЕ СРЕДСТВА НА ДУШУ НАСЕЛЕНИЯ“ НА ОСНОВЕ СТАТИСТИЧЕСКИХ ДАННЫХ МОНИТОРИНГА РЕГИОНОВ

The technique of a finding structure of dependence of a macroeconomic indicator (y – “The permanent assets by regions”) on a set of others ($x_1, x_2, \ldots, x_m$ – “Investments into a fixed capital per capita by regions”, “Amount of the small enterprises by regions”, etc.) is offered. It is constructed the model for forecasting of the studied indicator in conditions when it is not known a priori which x_i participate in formation of y. In system ArcGIS the geoinformation analysis of the received model is carried out.

Keywords: *forecast, modeling, GMDH-technique, nonlinear regress analysis, social-economical indices, geoinformation analysis*

The formulation of the problem and the analysis of the last achievements. The urgency of the theme is defined by the fact that the level of social development processes forecasting contributes to the efficiency of planning and management in economy and other spheres. The analysis of publications devoted to forecasting allows telling, that about two hundred methods of forecasting are known. Most of classification schemes divide forecasting methods into three basic classes, they are: methods of extrapolation, expert estimations and modeling. The modeling approach potentially gives the most accurate forecast, but it is the most complex and science intensive. Nowadays the approach based on extrapolation of the time series is the most popular. This approach shall be applied in case parameters are observed in years. If there are not enough data or the data is unsuitable for forecasting owing to incompleteness, drastic jumps of the parameter, etc., then other forecasting approaches are required. One of such approaches is the Group Method of Data Handling (GMDH) [1].

Feature of GMDH algorithms is that a view of basic function, the equations’ class and models’ structure are established in objective way by means of searching the variants on expediently chosen ensemble of criteria. The way of criteria introduction (finding a minimum of their convolution) provides an objective finding of unique model structure of optimum complexity at high noise method stability.

GMDH is the basic method for direct modeling of complex systems on a small number of experimental data. “The mathematical model” is understood as a system of regress equations, serving either for the unitary forecast of the future course of processes in complex system, or for the description of the physical and some other principles operating in object.

Formulation of the problem. Values of social and economic indices of regions monitoring for a small number of years (an example of initial data – tab. 1) are given. It is required to find the dependence structure of one specific index (y – “The permanent assets per capita”) on a set of others ($x_1, x_2, \ldots, x_m$ – “Investments into a fixed capital”, “Amount of the small industrial enterprises”, etc.), to construct model for investigated index forecasting under the conditions when it is not known a priori which x_i is used in formation of Y.

Object examination results are represented as the matrix $X\,[N \times m]$ and the vector $y\,[N \times 1]$. The problem of structural identification has to be solved, using the data of N examinations, i.e. the structure of one-dimensional variable y dependency on the collection of entry variables X has to be determined under conditions that it is not known a priori which factors exactly from the collection of entry variables take part in the output variable Y forming.

Let us assume that the necessary object model can belong to a set G, containing models as follows

$$\hat{y} = f\,(X, \hat{Q}_f\,), \qquad (1)$$

where $\hat{Q}_f$ is a vector of model parameters, evaluated somehow by the examinations data.

So the problem of structural identification reduces to determining the minimum of specified model quality criterion J

$$f^* = \arg\min_{f \in G} J\,(y, f(X, \hat{Q}_f\,)). \qquad (2)$$

Methods of solving the problem may differ at least in the following features:

a) algorithm of models' structures forming (generation) from the set G;

b) parameters of model evaluation methods;

c) quality of J-criterion evaluation itself;

d) organizing of movement of the criterion to minimum.

Let us note that models' parameters evaluation methods, their quality criteria and methods for searching criterion's minimums are generally independent and can be applied in different combinations. That is why a lot of methods for solving the mentioned problem (2) can be suggested.

The method of mathematical model construction uses:

a) GMDH as a method for enumerating models;

b) least-squares and least-modulus methods for evaluating model's parameters;

c) residual sum of squares criteria and "sliding examination" criteria for evaluating quality of resulting models.

Model's structure is chosen in accordance with principles of economy and adequacy.

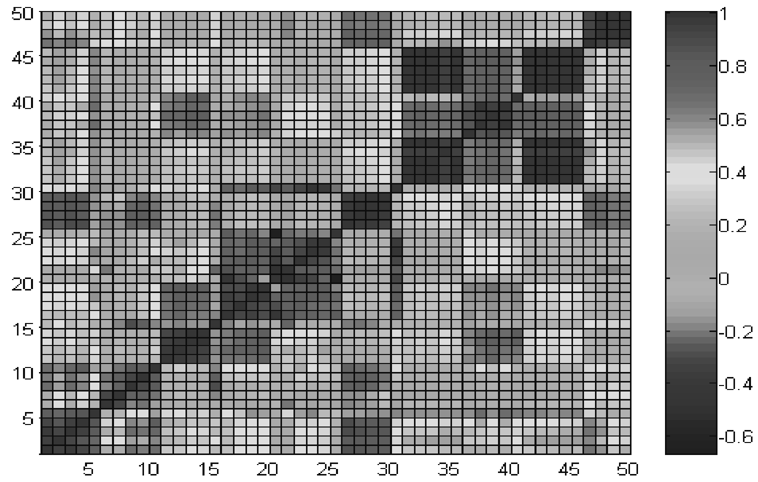
Table 1

An example of the table of regions monitoring indices

№	Region	**The permanent assets by regions** y **(x_{10}),** in the actual prices, million hrn					**Investments into a fixed capital per capita (x_1),** in the actual prices, hrn.					**Retail commodity circulation of the enterprises by regions (x_2)** million hrn					**(x_3)…(x_9)**
		2000	**2001**	**2002**	**2003**	**2004**	**2000**	**2001**	**2002**	**2003**	**2004**	**2000**	**2001**	**2002**	**2003**	**2004**	**…**
1	Dnipropetrovsk	27992	29220	28001	30448	40424	630	975	954	160	259	125	151	199	273	375	…
2	Volnogorsk	35444	39202	41764	40630	45330	1357	1924	2649	422	422	I2,5	10,0	13,0	13,5	12,8	…
3	Dniprodzerzhinsk	32000	37264	37391	39487	41449	368	530	403	655	834	102,	134,	168,	198,	285,	…
4	Joltye Vody	23036	27736	27156	26620	26857	114	171	142	192	411	19,9	13,7	13,2	14,7	19,6	…
5	Krivoy Rog	31206	39019	40456	43699	63700	1102	996	923	156	195	398,	445,	511,	589,	867,	…
6	Marganets	19926	22663	23439	50618	54778	190	206	445	497	355	13,0	16,4	19,1	17,3	19,0	…
7	Nikopol	19153	23752	23120	25260	26378	275	585	736	417	574	70,3	81,7	113,	127,	191,	…
8	Novomoskovsk	8466	12411	12294	13375	14484	132	216	280	597	498	53,3	59,5	72,0	97,2	117,	…
9	Ordzhonikidze	25746	26811	29803	40067	40552	343	742	989	196	231	18,2	18,9	18,2	16,6	19,5	…
10	Pavlograd	21100	25576	28051	27520	49030	327	527	424	828	120	70,2	92,9	130,	182,	238,	…
11	Pervomaisk	27001	28192	30346	33426	11752	1558	1765	1690	318	251	9,3	10,0	10,3	10,5	14,2	…
12	Sinelnikovo	7360	10761	15266	15499	16360	100	243	67	195	747	8,5	7,9	10,1	12,1	17,6	…
13	Ternovka	28292	35361	37158	46129	7611	1888	1846	1761	257	387	8,4	10,3	10,8	13,2	18,1	…
14	Apostolovskiy	13080	38102	9705	10032	10319	28	72	40	315	239	16,9	16,6	16,0	21,1	26,1	…
15	Vasilyevskiy	6958	7071	6675	4642	5275	131	165	353	377	444	8,8	10,6	11,4	9,6	12,4	…
16	Verkhnedneprovskiy	8534	9682	12758	14725	16295	100	213	196	149	837	12,0	13,3	15,5	16,9	27,6	…
17	Dniepropetrovskiy	11834	16090	16778	18077	23063	189	1240	1354	243	379	48,2	70,1	55,2	64,1	53,0	…
18	Krivorozhskiy	9841	9796	7085	8032	9305	146	364	398	716	854	10,9	14,3	12,8	14,2	13,6	…
19	Krinichanskiy	5588	5486	4647	4178	4757	89	132	164	338	430	15,4	16,1	17,5	17,9	22,8	…
20	Magdalinovskiy	10784	10158	10747	10053	11588	2302	2743	392	136	127	16,6	22,4	22,3	23,4	27,5	…
21	Mezhevskoy	6866	6982	6939	5360	5983	70	136	170	198	261	12,8	12,4	11,9	14,6	17,2	…
22	Nikopolskiy	14000	12763	12653	12318	13606	448	636	477	629	821	23,5	28,2	28,2	27,7	32,5	…
23	Novomoskovskiy	12789	15141	11899	10522	11614	194	932	358	468	651	22,4	25,8	26,5	29,6	45,2	…
24	Pavlogradskiy	6869	4570	4446	4094	5125	106	373	137	622	615	15,3	15,7	16,7	14,7	24,6	…
25	Petrikovskiy	5715	6744	4788	3760	4023	192	344	399	544	607	11,6	11,9	11,3	20,2	28,9	…
26	Petropavlovskiy	7433	9094	6553	5644	6197	51	118	155	140	393	11,8	10,2	9,3	8,6	9,7	…
27	Pokrovskiy	11972	12057	12172	11631	11230	243	295	397	200	238	18,5	17,1	18,2	17,5	23,9	…
28	Piatikhatskiy	5090	3873	7840	9142	10992	54	273	112	134	348	16,1	14,9	15,0	16,7	22,2	…
29	Sinelnikovskiy	8203	7603	9029	10171	11646	178	880	1252	105	538	9,0	8,8	8,7	12,4	18,3	…
30	Solonianskiy	5828	7270	9267	6688	8410	146	66	114	94	341	23,3	23,8	23,1	22,2	32,6	…
31	Sofievskiy	10944	10700	6989	6197	6171	150	244	155	219	449	9,0	9,3	10,2	12,7	16,7	…
32	Tomakovskiy	6647	6182	5852	6006	6889	141	174	322	211	417	9,0	8,6	9,1	8,2	11,0	…
33	Tsarichanskiy	8177	7069	6470	7010	7686	86	275	308	323	363	10,2	9,3	9,4	13,9	16,9	…
34	Shirokovskiy	7073	7262	6051	7241	10195	132	419	280	299	263	11,7	13,2	13,8	12,1	9,4	…
35	Yuryevsiy	7987	9683	9002	7403	7748	1038	536	453	433	414	4,5	5,7	5,8	5,5	5,8	…

Forecasting of the index "The permanent assets per capita" on the basis of statistical data of regions monitoring in Dniepropetrovskaya oblast. Initial data for the forecasting representing values of 10 ecology-social-economical (ESE) indices of 35 regions for the five-year period (2000–2004). It has been taken from the guide "Statistic year-book of Dniepropetrovskaya oblast for 2004 year" [2], published by the Central Administrative Board of Statistics in the Dniepropetrovskaya oblast.

At the first stage the prospecting analysis of initial data is carried out. The prospecting analysis included calculation of a pair of factors of correlation, the analysis of the main components and cluster analysis [3]. In the Fig. 1 the matrix of pair factors of initial indices correlation is shown. The original square-rectangular structure is allocated in the matrix. It gives the basis to consider sharp turns of a trajectory of regions development per 2000–2004 were not observed.



No. row	x_i	The name of the index
1–5	x_{10}	The permanent assets by regions (in the actual prices, million UAH) 2000–2004
6–10	x_1	Investments into a fixed capital per capita by regions (in the actual prices, UAH) 2000–2004
10–15	x_2	Retail commodity circulation of the enterprises by regions (million UAH) 2000–2004
16–20	x_3	Amount of the small enterprises by regions 2000–2004
21–25	x_4	Amount of the small industrial enterprises by regions 2000–2004
26–30	x_5	Monthly average nominal wages of hired workers (UAH) 2000–2004
31–35	x_6	Emissions of harmful substances to the atmosphere from stationary sources of pollution (thousands of tons) 2000–2004
36–40	x_7	Emissions of harmful substances to the atmosphere from motor transport (thousands of tons) 2000–2004
41–45	x_8	Emissions of harmful substances in an atmosphere by regions (thousands of tons) 2000–2004
46–50	x_9	Emissions of harmful substances in an atmosphere by regions per km^2 (tons) 2000–2004

Fig. 1. The correlation matrix of indices of regions monitoring in Dniepropetrovskaya oblast for the period of 2000–2004

The results of the principal components analysis are displayed in the Fig. 2. Considering this results all cities are allocated (1–12) as their principal components create separated, far lagged “emissions” behind the general cloud.

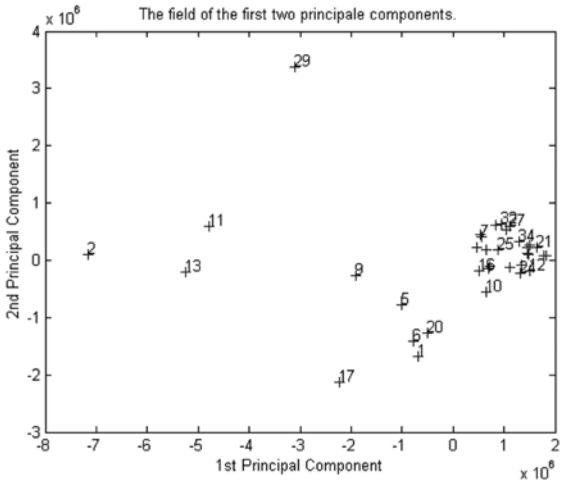


Fig. 2. Results of the principal components analysis of 35 regions in Dniepropetrovskaya oblast by indices $x_{1,\ldots,}\, x_{10}$ for the period of 2000–2004

Cluster analysis results testify that regions with the numbers 5 (Krivoy Rog), 1 (Dniepropetrovsk), 11 (Pervo-maysk) cannot be considered at construction of regression model together with other regions as they form separate classes (Fig. 3).

Considering the results of cluster analysis and the analysis of the principal components, it is possible to draw a conclusion that three regions (1 (Dniepropetrovsk), 5 (Krivoy Rog), 11 (Pervomaysk)) demand separate consideration during the solving of the forecasting task.

On the basis of the results of the prospecting analysis values of 9 ESE-indices (calculated per capita) for 35 regions of the Dniepropetrovskaya oblast for the period of 5 years, 2000–2004 (see interpritation to Fig. 1) are selected as an initial material for forecasting of index x_{10}.

The index x_{10} for the next year is taken as the dependent value Y. The structure of model can be presented in a general view by a dependence (some of $x\,[i,\,1]$, $x\,[i,\,2]\ldots$, $x\,[i,\,10]$ cannot be in final model): $x\,[i{+}1{,}10] = f\,(x\,[i,\,1],\, x\,[i,\,2]\ldots,\, x\,[i,\,10])$, where i is a year of monitoring (2000, ..., 2003).

The corresponding matrix of entrance data has dimension 140 x 11 (140=35 x 4), i.e. monitoring for 35 regions for 4 years, the first column – $x\,[i{+}1{,}10] \equiv y$, other columns – $x\,[i,\,1]$, $x\,[i,\,2]\ldots$, $x\,[i,\,10]$).

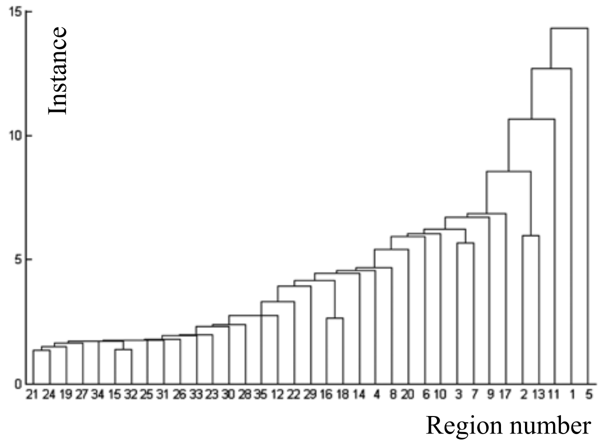


Fig. 3. Results of cluster analysis of 35 regions in Dniepropetrovskaya oblast by indices $x_1, \ldots, x_{10}$ for the period of 2000–2004

Multiserial iterative algorithm GMDH is used for forecasting. The expediency of using GMDH is caused by applying iterative schemes of models complication. Models complication from range to range of selection occurs due to “crossing” the best models of a previous range. Thus the intermediate decision is not unique; checking testing and control samples used for model efficiency; a self-selection of intermediate decisions before reception of the unique final decision occurs.

Synthesized models class has the following look

$$\hat{y} = \sum_{q=1}^{s} \Theta_q \cdot \prod_{j=1}^{m} x_j^{\alpha_{q_j}} , \tag{3}$$

where $\overleftarrow{y}$ is an output variable; s is the number of model members; Θ_q, $q = 1, 2, \ldots, s$ are coefficients; x_j, $j = 1, 2, \ldots, m$ are entry variables; m is the number of entry variables; α_{q_j} is an exponent in which the x_j variable is contained in the q-th member.

We will define the general form of particular descriptions matrix $\hat{Z}^r$ in the following way

$$\hat{Z}^r = [\ \hat{z}_1^r \ \vdots\ \hat{z}_2^r \ \vdots \ldots \vdots\ \hat{z}_{F+2+m+2s}^r\ ] , \tag{4}$$

where $\hat{z}_j^r$, j=1,2,…,F+2+m+2s are vectors (of dimension N×1), particular descriptions; F is the number of the best particular descriptions that are passed from iteration to iteration; s is the number of members in the structure of the (r –1) – the iteration’s best particular description.

The algorithm consists of the following steps:

1. Initial particular descriptions matrix $\hat{\mathbf{Z}}^o$ is selected (it is supposed that s = 0 for this matrix)

$$\hat{Z}^o = [\ O \ \vdots\ o \ \vdots\ I \ \vdots\ X\ ] = [\ O \ \vdots\ \hat{C}^o\ ] , \tag{5}$$

where O is a zero matrix (of dimension N×F); o is a zero vector (of dimension N×1); I is a unit vector (of dimension N×1); $X = [\,x_1 \vdots x_2 \vdots \ldots \vdots x_m\,]$ is a matrix of entry parameters examinations (of dimension N×m).

2. Transformation operator R is found out $\hat{Z}^{r-1} \xrightarrow{R} \hat{Z}^r$. The best F particular decisions are selected from all generated particular decisions by minimum of quadratic residues norm on the checked observation subsampling. The best description parts are ranged and they are used while forming initial F vectors of $\hat{Z}^r$ matrix, last 2s vectors are formed taking into account structure of best particular decisions among F checked.

Models synthesis is starting with stage number $p = 1$ (or with any indicated number p_o). Every stage is an iterative scheme. The initial particular decisions’ matrix of a p number stage is specified by output particular decisions’ matrix of the previous stage. Calculation is over on the stage p^* when the following condition is fulfilled

$$J(\hat{Z}_{F,p^*-1}^{r^*}) - J(\hat{Z}_{F,p^*}^{r^*}) < \delta_p , \tag{6}$$

where $J(\hat{z}_{F,p}^r)$ is a criterion value for the best particular decision of r-th iteration and p-th stage; δ_p is specified number.

Typical features of the algorithm: 1) multistage model computing; 2) model searching in both linear and nonlinear classes by input model variables; 3) methods of extracting of individual members of the best particular decision and based on this spreading of basic argument set; 4) scheme of calculating sliding examination criteria optimal by calculating expenditure for iterative GMDH algorithms arguments; 5) the ability to evaluate model coefficients by both the least square method and the least modulus method [5, 6, 7]. During the model calculating informative index subsets are being selected.

Calculations were carried out on two stages. On the first one all data were considered as training sample “A”, the criterion applied for quality estimation of received models was “sliding examination” $U(s)$ [1], which is not necessarily decreased with increasing of model complexity, and it can have a minimum corresponding to the complexity best model.

At the second stage two samples were considered: training sample “A” (data of monitoring for 2000–2002, the first 105 lines of an entrance matrix data); control sample “C” (data of monitoring for 2003). Five regions listed above have been excluded during the construction of the model at this stage.

The criterion of the residual sum of squares $J(s)$ was applied for quality estimation of received models [1]

$$J(s) = \left\| y - \hat{y}(s)) \right\|^2 , \tag{7}$$

where s – is complexity of model, i.e. number of estimated indices. This criterion allows estimating the complexity of the model as criterion decreases with increasing of complexity of the model.

The nonlinear structure of model was considered at each stage; therefore preliminary data normalization was carried out.

The quality of received models was estimated also by traditional characteristics: mean square deviation of modeling values $\hat{y}$ from observable and the coefficient of multiple determinations that is connected unambiguously with coefficient of multiple correlations.

When choosing the compromise proposal between quality of model and its complexity one should recognize the best of calculated models are the following (tab. 2):

Model *A*

$$y= -0{,}017x_{10}^2+0{,}93x_{10}+ +0{,}084x_1x_5x_8+0{,}053x_4 \quad (8)$$

(by criterion *U* (*s*));

Model *B*

$$y = 0.879\, x_{10} + 21{,}737x_4x_5 \quad (9)$$

(by criterion *J* (*s*)).

Table 2

The results of the best models

Stage	The First	The Second
Model	$y= -0{,}017x_{10}^2+0{,}93x_{10}+ +0{,}084x_1x_5x_8+0{,}053x_4$	$y = 0{,}879x_{10} + +21{,}737x_4x_5$
Criterion $U(s)10^{-3}$	0,321	–
Criterion $J(s)10^{-3}$	–	5,1
Coefficient of multiple determination	$R_A = 0{,}916$	$R_A = 0{,}74$; $R_C = 0{,}89$
Total error	7,62 %	19,7 %
Model errors by years		
2001	6,5 %	16,0 %
2002	7,3%	27,5 %
2003	7,4 %	20,6 %
2004	9,3 %	14,9 %

Such indices as x_2, x_4, x_5, x_8, x_{10} exert the greatest influence on an index “The permanent assets”. For the person making a decision it can be interesting why such important indices as “Amount of the small enterprises” and “Investments into a fixed capital” were not included into the number of the best models.

By means of a method of extreme forecasts it is possible to define borders of falling and growth of the index. The given method allows receiving quite an obvious picture for various variants of events succession, and also gives the information about sensitivity and possible deviations of the model.

Construction of optimistic, pessimistic and expected forecast provides using of the minimal and maximal values of indices.

The optimistic forecast for model *A* looks like that

$$y_{\uparrow}^{A} = 1{,}009 \cdot x_{10,\max} + 0{,}020 \cdot x_{2,\max} \cdot x_{8,\max}\,. \quad (10)$$

The pessimistic forecast for model *A* looks like that

$$y_{\downarrow}^{A} = 1{,}009 \cdot x_{10,\min} + 0{,}020 \cdot x_{2,\min} \cdot x_{8,\min}\,. \quad (11)$$

Optimistic and pessimistic forecasts for model B have the following look respectively:

$$y_{\uparrow}^{B} = 0{,}879 \cdot x_{10,\max} + 21{,}737 \cdot x_{4,\max} \cdot x_{5,\max}\,; \quad (12)$$

$$y_{\downarrow}^{B} = 0{,}879 \cdot x_{10,\min} + 21{,}737 \cdot x_{4,\min} \cdot x_{5,\min}\,. \quad (13)$$

Average value of the forecast

$$\bar{y}^{C} = \frac{Y_{\uparrow}^{C} + Y_{\downarrow}^{C}}{2}, \; C = A, B. \quad (14)$$

The expected forecast

$$y_{expected}^{C} = \frac{y_{\uparrow}^{C} + 4 \cdot \bar{y}^{C} + y_{\downarrow}^{C}}{6}, \; C = A, B. \quad (15)$$

In the Fig. 4 there are presented the comparative characteristic of the forecasting values for model *A* and *B* respectively.

Expected forecast values for model *A* deviate from real by 6.4%, for model *B* the difference makes 11.6%.

As one can see values that have been calculated by model (*A*) have a great disarrangement with real values for the next regions: 10 (Pavlograd), 11 (Pervomaisk), 13 (Ternovka), and 14 (Apostolovskiy). For model *B* with five regions listed above excluded sharp distinctions between actual and the calculated indices are observed only for region 10 (Pavlograd) and 14 (Apostolovskiy).

The divergence corridor of pessimistic and optimistic forecasts can be seen on diagrams. It is interesting that it is much wider for cities, rather than for regions.

Geoinformation analysis of received models was implemented in ArcGIS system (Fig. 6). Rank cartograms of actual index *Y* were computed for model *A* and calculated using classification scheme „standard deviation“ [4].

The residues $|y - \hat{y}|$ have been calculated by the regression model *A*. The total index (2001–2004) of regression model residues was calculated by following formula

$$L = \sum_{j=1}^{k} \sum_{i=1}^{n} \frac{y_j^{(i)} - \hat{y}_j^{(i)}}{y_j^{(i)}}, \quad (10)$$

where $k = 4$ – number of monitoring years, $n = 35$ – number of monitoring regions, $y_j^{(i)}$ – value of the index for *j*-th year in *i*-th region. This index shows the regions lagging most behind average progress trend.

Rank cartogram for *L* using the classification scheme “standard deviation” [4] and geoinformation system ArcGIS (Fig. 7.).

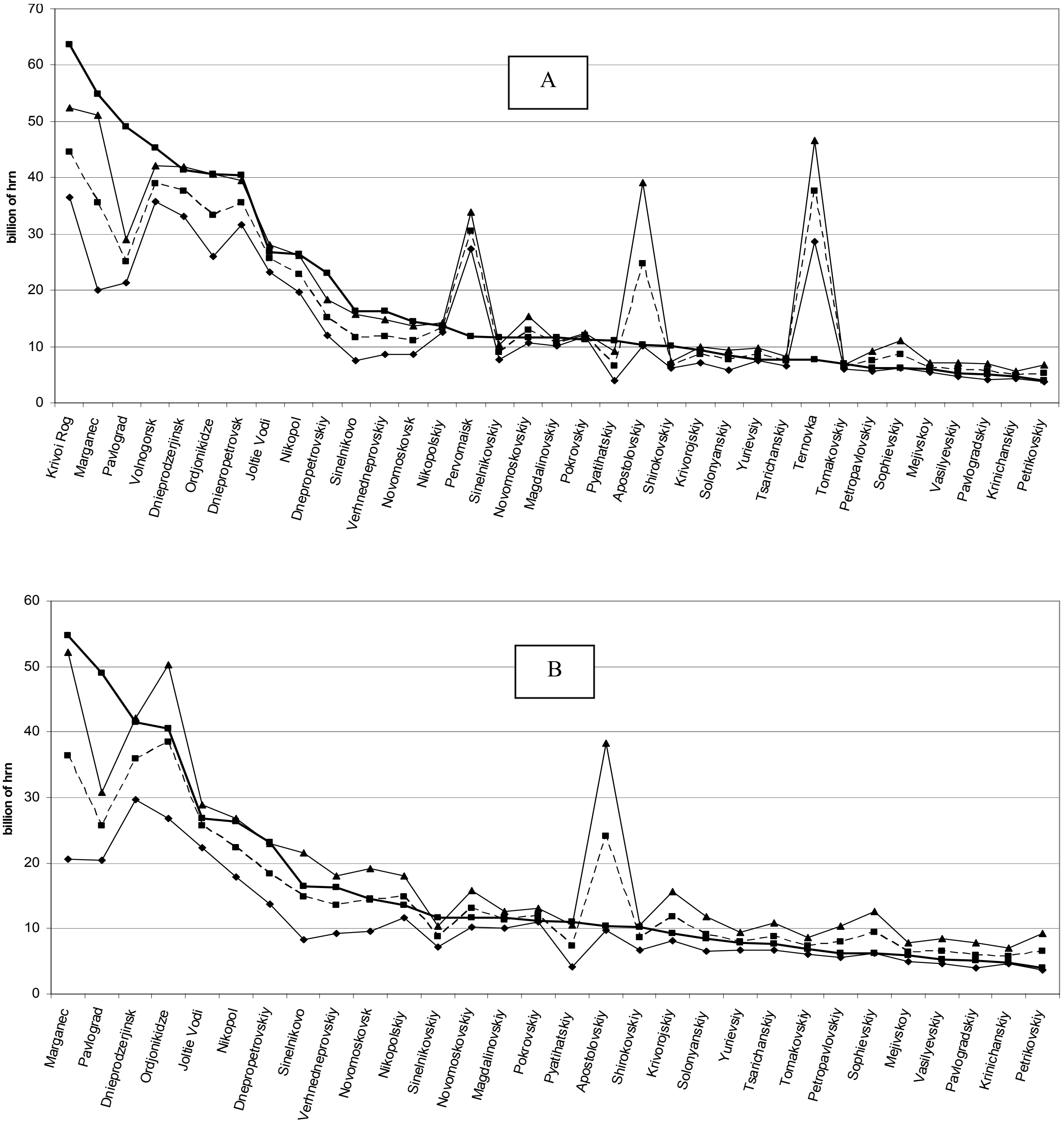
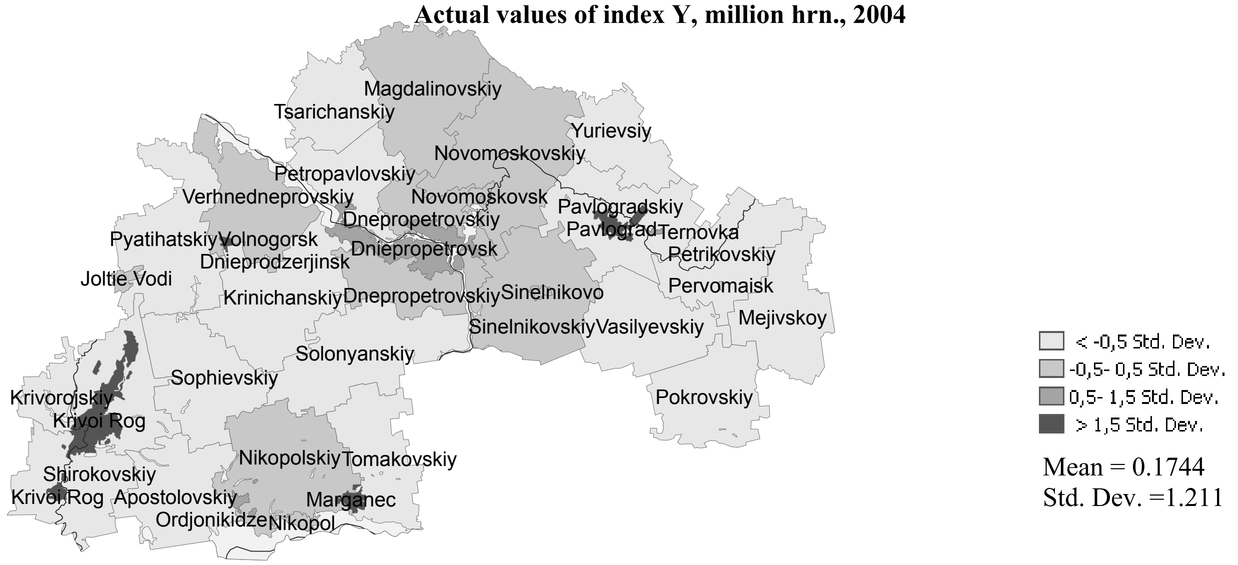


Fig. 4. The comparative characteristic of forecasting values of index Y ("The permanent assets") for 2004 for models A and B

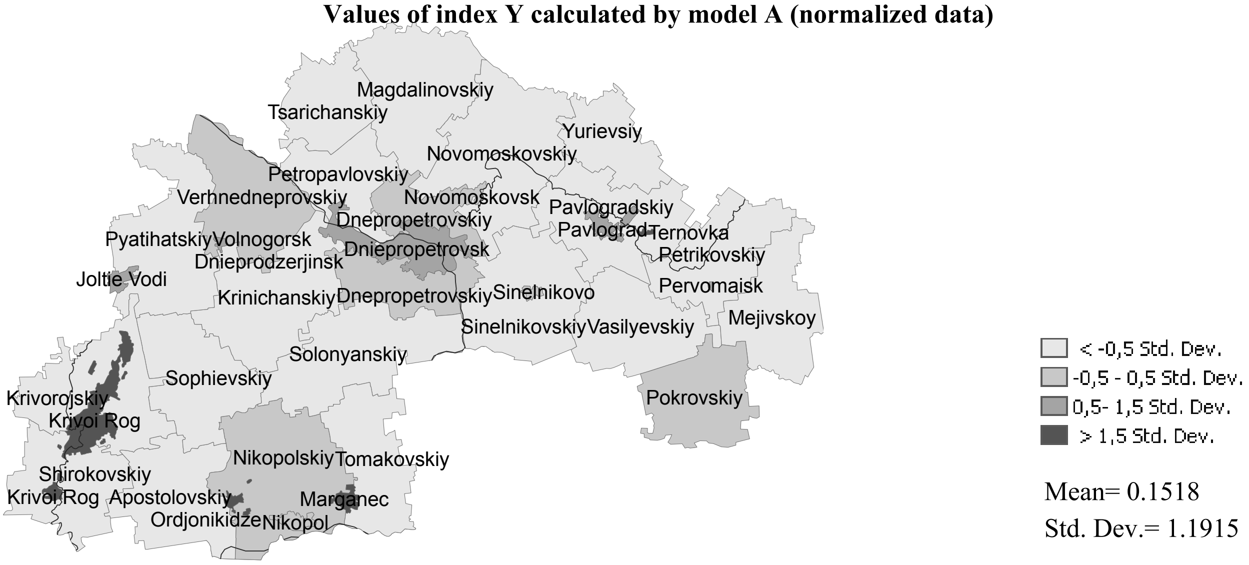
During the analysis of the total residues index of regression model those regions where this parameter sharply differs from average value are standing out (Apostolovskiy, Nikolayevskiy, Tomakovskiy and Pokrovskiy regions, and also Dniepropetrovsk and Zholtye Vody), i.e. the tendency of their development differs from the general tendency inherent in the majority of regions of the Dniepropetrovskaya oblast (it is possible that accounting indices on these regions are not authentic).

Conclusions. Results of modeling allowed to define influence of 10 monitoring indices on "The permanent assets per capita by regions" index. Carried out researches give the information for reflections to the person making a decision: "Why for such cities as 5 (Krivoi Rog), 6 (Marganec), 10 (Pavlograd), 11 (Pervomaisk) and 13 (Ternovka) – cities where there are mines and careers – predicted and actual values especially differ?". Also it is interesting why such an important index as "Investments into a fixed capital" is not used in formation of the "The permanent assets per capita by regions" index.

Actual values of index Y, million hrn., 2004



Values of index Y calculated by model A (normalized data)



Residues of regression model A

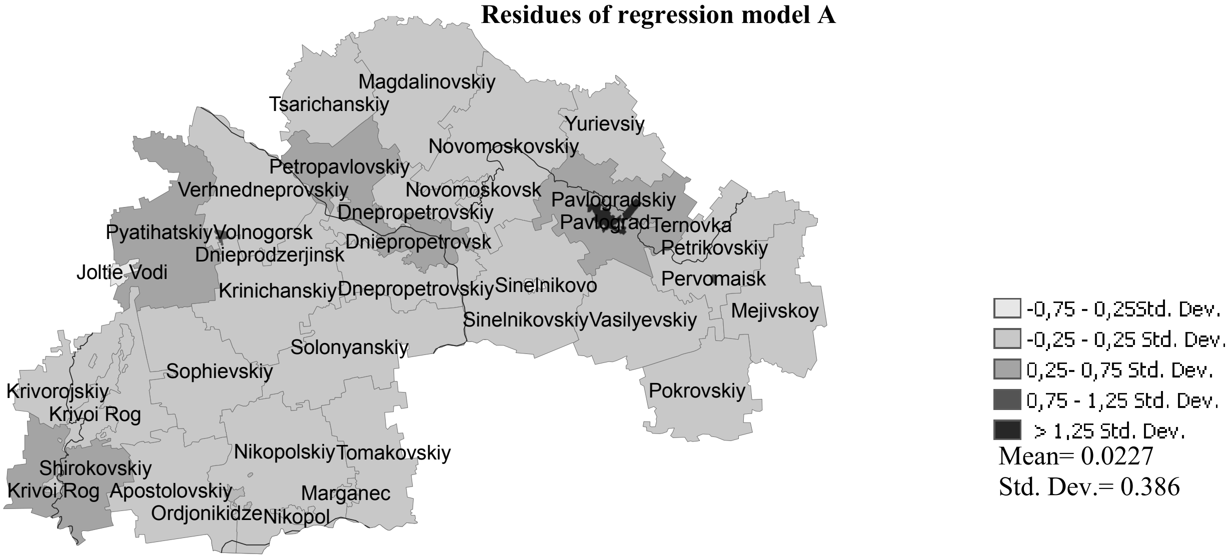


Fig. 5. Rank cartograms of the index "The permanent assets per capita by regions", 2004

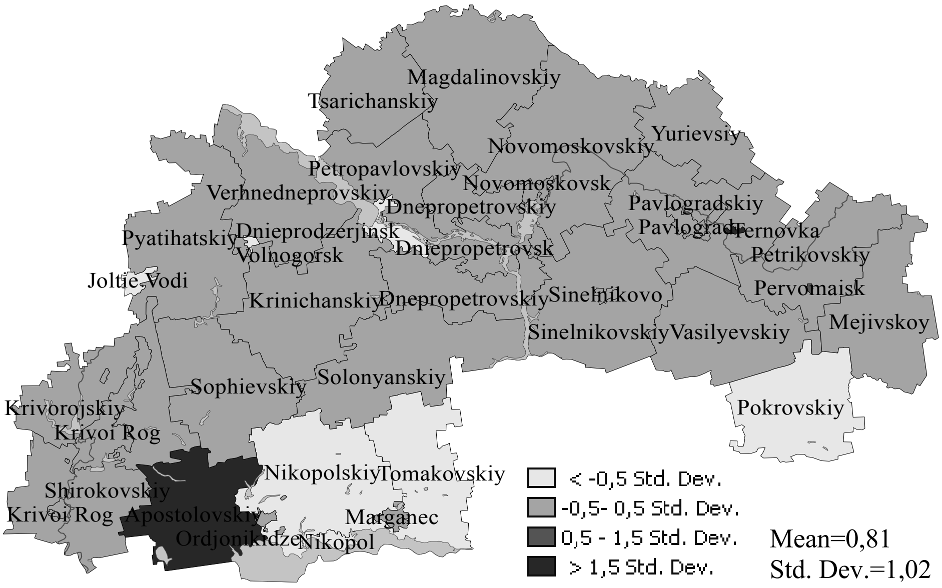


Fig. 6. The total index (2001–2004) of residues of regression model (A)

References

1. Ivanenko O.G. Inductive methods of forecasting and analysis of complex economic systems / Ivanenko O.G., Ivanenko G.O. // Economist. – 1988.–№5. – P. 88–97.
2. Statistic year-book of Dniepropetrovsk area for 2004 year. Dniepropetrovsk: the Head administration of statistics. – 2004. – 570 p.
3. Sarycheva L.V. Computer ecology-social-economic region monitoring. Mathematical support: Monograph / Sarycheva L.V. – Dniepropetrovsk: NGU. – 2003. – 222 p.
4. Sarycheva L.V. Computer ecology-social-economic region monitoring. Geoinformation support: Monograph / Sarycheva L.V. – Dniepropetrovsk: NGU.– 2003. – 174 p.
5. Sarychev A.P. S-Scheme of Sliding Examination for Optimal Set Features Determination in Discriminant Analysis by the Group Method of Data Handling / Sarychev A.P. // System Analysis and Modelling Simulation (SAMS), Taylor & Francis. – 2003. – Vol. 43. – No.10. – P. 1351–1362.
6. Sarycheva L.V. Using GMDH for Modeling Economical Indices of Mine Opening/ Sarycheva L.V. // Systems Analysis and Modeling Simulation (SAMS).– 2003, Vol. 43, №10, P. 1341–1350.
7. Sarycheva L. Using GMDH in Ecological and Socio-economical Monitoring Problems/ Sarycheva L. // Systems Analysis and Modeling Simulation (SAMS). – 2003, Vol.43, №10. – P. 1409–1414.
8. Sarycheva L. Cluster Analysis of Territories by the Totality of Ecological and Socio-Economic Indices / Sarycheva L., Zhuykov A. // IEEE 2001 International Geoscience and Remote Sensing Symposium, UNSW. – P. 663–664.
9. Muller J.-A. Self-organizing Data Mining. Extrating Knowledge from Data/ Muller J.-A., Lemke F. – Hamburg: Libri, 2000. – 250 p.
10. Ivanenko A.G. Selbstorgfnization von Vorherzagemodellen/ Ivanenko A.G., Muller J.-A. – Berlin: VEB Verlag Technik, 1984.– 270 p.

Предложена методика нахождения структуры зависимости макроэкономического показателя (y – „Основные средства на душу населения“) от набора других (x_1, x_2,..., x_m – „Инвестиции в основной капитал“, „Количество малых промышленных предприятий“ и др.). Построена модель для прогнозирования изучаемого показателя в условиях, когда априорно неизвестно, какие именно x_i участвуют в формировании y. В системе ArcGIS осуществлен геоинформационный анализ полученной модели.

Ключевые слова: *прогноз, моделирование, МГУА, нелинейный регрессионный анализ, социально-экономичные показатели, геоинформационный анализ*

Запропоновано методику знаходження структури залежності макроекономічного показника (y – „Основні засоби на душу населення“) від набору інших (x_1, x_2,..., x_m – „Інвестиції в основний капітал“, „Кількість малих промислових підприємств“ та ін.). Побудовано модель для прогнозування досліджуваного показника в умовах, коли апріорі невідомо, які саме x_i беруть участь у формуванні y. У системі ArcGIS здійснено геоінформаційний аналіз отриманої моделі.

Ключові слова: *прогноз, моделювання, МГУА, нелінійний регресійний аналіз, соціально-економічні показники, геоінформаційний аналіз*

Рекомендовано до публікації докт. техн. наук Б.С. Бусигіним. Дата надходження рукопису 14.03.11