MULTI-CRITERIA OPTIMIZATION BASED ON THE REGRESSION EQUATION SYSTEMS IDENTIFICATION

A.P. Kotenko[1,2], D.A. Pshenina[2]

[1] Samara National Research University, Samara, Russia
[2]Samara State Technical University, Samara, Russia

Abstract. Consider the problem of multi-criteria optimization with conflicting criteria. An example is the complex chemical production with random parameters. We describe using regression equations dependence of targets from control actions. Investigation of the systems of interdependent regression equations requires description of all possible variants of their properties. The most productive case is exact identification of the system parameters. It allows to find the optimal values of the control parameters for manufacturing of quality assured production.

Basic notation system

Let dependence of optimization criteria y_1, $y_2,\ldots$, y_n from the controlling factors x_1, $x_2,\ldots$, x_m is expressed by a system of linear regressions. In the structural form of the model (SFM) [1,2]

$$\begin{cases} \vec{y}_1 = a_{12}\vec{y}_2 + a_{13}\vec{y}_3 + \ldots + a_{1n}\vec{y}_n + b_{11}\vec{x}_1 + b_{12}\vec{x}_2 + \ldots + b_{1m}\vec{x}_m + \vec{\varepsilon}_1, \\ \vec{y}_2 = a_{21}\vec{y}_1 + a_{23}\vec{y}_3 + \ldots + a_{2n}\vec{y}_n + b_{21}\vec{x}_1 + b_{22}\vec{x}_2 + \ldots + b_{2m}\vec{x}_m + \vec{\varepsilon}_2, \\ \ldots \\ \vec{y}_n = a_{n1}\vec{y}_1 + a_{n2}\vec{y}_2 + \ldots + a_{n,n-1}\vec{y}_{n-1} + b_{n1}\vec{x}_1 + b_{n2}\vec{x}_2 + \ldots + b_{nm}\vec{x}_m + \vec{\varepsilon}_n, \end{cases} \quad (1)$$

let us associate endogenous variables $\vec{y}_i = \left(y_{i1}, y_{i2},\ldots, y_{iN}\right)^T \in \mathfrak{R}^N$, $i \in \overline{1,n}$; and exogenous variables $\vec{x}_j = \left(x_{j1}, x_{j2},\ldots, x_{jN}\right)^T \in \mathfrak{R}^N$, $j \in \overline{1,m}$.

It corresponds to the reduced form of the model (RFM) [1,2]

$$\tilde{y}_1 = \beta_{11}\vec{x}_1 + \beta_{12}\vec{x}_2 + \ldots + \beta_{1m}\vec{x}_m, \quad \tilde{y}_2 = \beta_{21}\vec{x}_1 + \beta_{22}\vec{x}_2 + \ldots + \beta_{2m}\vec{x}_m, \quad \ldots$$

$$\ldots, \quad \tilde{y}_n = \beta_{n1}\vec{x}_1 + \beta_{n2}\vec{x}_2 + \ldots + \beta_{nm}\vec{x}_m \tag{2}$$

with regression values $\tilde{y}_i = \left(\tilde{y}_{i1}, \tilde{y}_{i2}, \ldots, \tilde{y}_{iN}\right)^T \in \mathfrak{R}^N$ and the reduced coefficients β_{ij}, found by the Ordinary Least Squares (OLS) method.

At the same time vectors of the regression values of endogenous variables are obtained by orthogonal projection Pr of vectors of observations of endogenous variables to the linear span [1] $L(\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_m) \subseteq \mathfrak{R}^N$ of dimension

$$\dim L(\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_m) \le m$$

of observed values vectors of the predefined variables

$$\tilde{y}_i = \mathrm{Pr}\left(\vec{y}_i; L(\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_m)\right) \in L(\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_m).$$

Since the structure coefficients are functions of the solutions of the reduced system of linear algebraic equations (SLAE), the solvability of SFM identification problem (1) is determined by the ranks of the system's matrix and the extended matrix of the system.

Accordingly, there may be the following cases:
- one solution (accurate identifiability of the system's equation by indirect OLS);
- no solution (superidentifiable equation when two-step OLS is applicable);
- infinite many solutions (unidentifiable equation of the SFM system).

Replacing the sample values of the endogenous variables of SFM (1) by the regression values from PFM (2), we obtain in the linear span $L(\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_m)$ of system (1) predefined variables' sample values the following independent linear equations, which enable us to identify the structural factors:

$$\tilde{y}_1 = a_{12}\tilde{y}_2 + a_{13}\tilde{y}_3 + \ldots + a_{1n}\tilde{y}_n + b_{11}\vec{x}_1 + b_{12}\vec{x}_2 + \ldots + b_{1m}\vec{x}_m,$$
$$\tilde{y}_2 = a_{21}\tilde{y}_1 + a_{23}\tilde{y}_3 + \ldots + a_{2n}\tilde{y}_n + b_{21}\vec{x}_1 + b_{22}\vec{x}_2 + \ldots + b_{2m}\vec{x}_m, \tag{3}$$
$$\ldots$$
$$\tilde{y}_n = a_{n1}\tilde{y}_1 + a_{n2}\tilde{y}_2 + \ldots + a_{n,n-1}\tilde{y}_{n-1} + b_{n1}\vec{x}_1 + b_{n2}\vec{x}_2 + \ldots + b_{nm}\vec{x}_m.$$

From (1)-(3), we obtain the system of m linear equations to identify structural coefficients of i-th equation:

$$\tilde{y}_i = \sum_{k=1}^{m} \beta_{ik}\vec{x}_k = \sum_{j=1; j\neq i}^{n} a_{ij}\tilde{y}_j + \sum_{k=1}^{m} b_{ik}\vec{x}_k = \sum_{j=1; j\neq i}^{n} a_{ij}\left(\sum_{k=1}^{m} \beta_{jk}\vec{x}_k\right) + \sum_{k=1}^{m} b_{ik}\vec{x}_k \Leftrightarrow$$

$$\Leftrightarrow \beta_{i1} = b_{i1} + \sum_{j=1; j\neq i}^{n} \beta_{j1}a_{ij}, \quad \beta_{i2} = b_{i2} + \sum_{j=1; j\neq i}^{n} \beta_{j2}a_{ij}, \quad \ldots, \quad \beta_{im} = b_{im} + \sum_{j=1; j\neq i}^{n} \beta_{jm}a_{ij} \tag{4}$$

Let us consider the linear span L by combining multiple regression vector of endogenous variables values from the right hand side of the i-th equation of the system (4) and the set of vectors of sampled values of predefined variables.

Identifiability of structure coefficients of the i-th equation is determined by a combination of the following factors: the fact of its left hand side belonging (not belonging) to the linear span L of the SFM's right hand side vectors and by their linear (in)dependence.

In this case the two-step OLS result coincides with indirect OLS result if the considered system (1) equation can be exactly identified.

Setting up a problem

The content of bitumen in the physical volume of asphalt is only 5-7%, but it has the most significant influence on the quality and durability of the road surface. Bitumen production is one of the most energy-intensive. Energy costs for production of oxidized bitumen consist of steam, fuel and electricity consumptions. [3]

Residual bitumen production technology is based on the concentration of heavy petroleum residues of tar vacuum distillation. Tars characteristics affect the quality of road asphalts, in turn, these characteristics can be influenced by changing the composition of the hydrocarbon composition of the feedstock.

The main factors of the residual bitumen production process are the depth of the vacuum, steam flow and distillation temperature. Physical and chemical properties of bitumen depend on the hydrocarbon composition of raw materials. Thus, there is an inverse problem as follows: to obtain the desired properties of the product we need to determine the characteristics of raw materials composition and production process parameters.

Mathematical model

For this purpose, it proposed to use the methods of mathematical modeling, in particularly, to set up the system of regression equations where the input parameters are the parameters of the approved standards, and the results of calculations are the characteristics of the bitumen.

Input variables are as follows:

x_1 – penetration depth of the needle at 25^0C,

x_2 – penetration depth of the needle at 0^0C,

x_3 – extensibility at 25^0C,

x_4 – dynamic viscosity at 60^0C,

x_5 – softening temperature change after heating,

x_6 – extensibility after warming up.

Calculations will enable us to determine the following values:

y_1 – sulfur in %%,

y_2 – paraffins and naphtha of tar in %%,

y_3 – heating temperature,

y_4 – air consumption,

y_5 – warming-up duration,

y_6 – oil tar in %%,

y_7 – VU80 tar index,
y_8 – light aromatics of tar in %%.
65 observations have been processed. [4]
Setting up each of the regression equation was carried out in the frameworks of a posteriori approach, implying the inclusion of all feasible variables and sequential exclusion of insignificant ones.
At the first step let us set up regression equations with six variables:
y_1=4,05+0,014x_1–0,08x_2+0,003x_3–0,005x_4–0,11x_5–0,0014x_6, R^2=0,745399.
Analysis of the Student's coefficients revealed the following significant variables:
y_1=5,17–0,075x_2–0,0054x_4–0,117x_5, R^2=0,72951.
Similarly,
y_2=2,61+0,17x_1+0,27x_2–0,043x_3+0,002x_4+0,89x_5–0,011x_6, R^2=0,759902; $\Rightarrow$
y_2=3,34+0,23x_1–0,05x_3+1,35x_5, R^2=0,74412;
y_3=251,98–0,16x_1–0,58x_2–0,076x_3+0,001x_4+4,07x_5–0,03x_6, R^2=0,219015; $\Rightarrow$
y_3=222,17+3,69x_5, R^2=0,191535;
y_4=7,15–0,02x_1–0,02x_2–0,003x_3+0,0004x_4–0,02x_5–0,006x_6, R^2=0,234903; $\Rightarrow$
y_4=5,24–0,002x_6, R^2=0,05344;
y_5=–6,73+0,13x_1+0,13x_2–0,01x_3+0,0075x_4+0,38x_5–0,0075x_6, R^2=0,743188; $\Rightarrow$
y_5=–10,73+0,18x_1+0,0064x_4+0,764x_5, R^2=0,705288;
y_6=11,72+0,0096x_1–0,16x_2–0,009x_3–0,0008x_4+0,18x_5+0,009x_6, R^2=0,121453; $\Rightarrow$
y_6=12,12–0,16x_2+0,18x_5, R^2=0,061457;
y_7=543,24–4,75x_1-2,35x_2–0,51x_3–0,1x_4–12,16x_5+0,94x_6, R^2=0,677759; $\Rightarrow$
y_7=321,05–4,59x_1+1,024x_6, R^2=0,61394;
y_8=13,21–0,066x_1+0,1x_2+0,0079x_3–0,024x_4+0,597x_5–0,0088x_6, R^2=0,502146; $\Rightarrow$
y_8=14,96–0,0235x_4, R^2=0,393982.
Let us set up a system of equations out of significant regression equations:
y_1=5,17–0,075x_2–0,0054x_4–0,117x_5;
y_2=3,34+0,23x_1–0,05x_3+1,35x_5;
y_3=222,17+3,69x_5;
y_4=5,24–0,002x_6;
y_5=–10,73+0,18x_1+0,0064x_4+0,764x_5;
y_7=321,05–4,59x_1+1,024x_6;
y_8=14,96–0,0235x_4.
The parameters provided by the RF Standard are considered to be the given variables. They are as close as possible to Euro-requirements establishing quality standards for EU road bitumen:
x_1=60; x_2=13; x_3=80; x_4=320; x_5=3; x_6=40.
Hence,
y_1=1,104; y_2=21,85; y_3=11,07; y_4=–0,08; y_5=15,14; y_7=–234,4; y_8=7,52.
The results show the statistical unreliability of production results when production process is carried out in accordance with current technologies under the specified Standard values.

Forecasts confidence intervals

On the basis of 10 samples there were calculated correlations (Table 1), indicating a strong linear relationship between the following parameters [5]:

X –softening temperature,

Y_1 – dynamic viscosity at 60^0C,

Y_2 – needle penetration depth at 25^0C,

Y_3 – extensibility after warming up.

This allows to find linear regressions among all pairs of the indices.

As a general regressor there was chosen a softening temperature X, for which the measurement can be made in the most accurate way.

Table 1. Linear correlation coefficient matrix

	X	Y_1	Y_2	Y_3
X	1			
Y_1	0,983569	1		
Y_2	–0,977430	0,975450	1	
Y_3	–0,972410	0,961180	0,988793	1

Significant regressions are obtained:

$Y_1=-2038,63+45,29X$, $R^2=0,9674$; $Y_3=868,37–15,97X$, $R^2=0,9456$.

The high values of determination coefficients indicate the adequacy of the proposed linear models to experimental data. The equations' parameters are conservatively significant according to Fisher's test at the significance level of α=0,01–0,05–0,10.

The link between Y_2 and X has not been studied, as it is co-directional to the link between Y_3 and X. There were found the forecast confidence intervals for Y_{regr1} (Table 2) and Y_{regr3} (Table 3).

Regressor X values and Y_1 and Y_3 forecast confidence intervals boundaries satisfying Standard ($X \geq 49$, $Y_1 \geq 250$, $Y_3 \geq 80$) are marked in the Tables 2,3 by italics. Thus, in Table 2 only the rows with $T \geq 50$ allow the values of Y_1 to satisfy the Standard, while in Table 3 these are only the rows with $T \leq 50,4$.

However, at significance level of $\alpha \geq 0,05$ there is only the line with $T=50$ in both tables, which is suitable. Y_3 index observed value is sufficiently shifted to the left boundaries of confidence intervals (both at α=0,05 and at α=0,01), which do not meet the Standard.

Table 2. Forecast confidence intervals for dynamic viscosity Y_1 according to X softening point regression

Variables		$\alpha=0{,}01$		$\alpha=0{,}05$		$\alpha=0{,}10$	
X	Y_1	$Y_{regr1}-\Delta$	$Y_{regr1}+\Delta$	$Y_{regr1}-\Delta$	$Y_{regr1}+\Delta$	$Y_{regr1}-\Delta$	$Y_{regr1}+\Delta$
48	138	79,68	191,18	97,11	173,74	104,53	166,32
48,5	159	104,14	212,00	121,01	195,14	128,18	187,96
49	171	128,20	233,23	144,63	216,81	151,62	209,82
49,3	180	142,43	246,18	158,66	229,96	165,56	223,05
50	230	175,00	277,02	190,95	261,07	197,74	254,28
50,4	241	193,19	295,07	209,12	279,14	215,90	272,36
51	300	219,91	322,70	235,99	306,62	242,82	299,79
51,8	316	254,51	360,57	271,10	343,99	278,15	336,93
52	320	262,98	370,21	279,75	353,45	286,89	346,31
53	341	304,42	419,36	322,40	401,39	330,04	393,74

Table 3. Forecasts confidence intervals for extensibility after warming up Y_3 according to X softening point regression

Variables		$\alpha=0{,}01$		$\alpha=0{,}05$		$\alpha=0{,}10$	
X	Y_3	$Y_{regr3}-\Delta$	$Y_{regr3}+\Delta$	$Y_{regr3}-\Delta$	$Y_{regr3}+\Delta$	$Y_{regr3}-\Delta$	$Y_{regr3}+\Delta$
48	110	75,94	127,34	83,98	119,30	87,40	115,88
48,5	100	68,79	118,51	76,57	110,74	79,88	107,43
49	80	61,46	109,87	69,03	102,30	72,25	99,08
49,3	79	56,96	104,78	64,44	97,31	67,62	94,12
50	**64**	46,18	93,20	53,53	85,85	56,66	82,72
50,4	55	39,82	86,78	47,17	79,44	50,29	76,31
51	47	30,03	77,41	37,44	70,00	40,59	66,85
51,8	46	16,50	65,38	24,14	57,74	27,39	54,49
52	43	13,03	62,46	20,76	54,73	24,05	51,44
53	25	−4,72	48,26	3,57	39,98	7,09	36,45

Conclusion

Standard boundaries are contradictory for indexes X, Y_1 and Y_3, which means the necessity of technological upgrade of production process if it is aimed at European bitumen standards.

This example illustrates the possibility of using identifiable systems of regression equations. They help to define the boundaries of control factors of production to ensure product quality at a given conflicting criteria and random perturbations.

References

1. Kotenko AP. Geometry of Systems of Econometric Equations. Control of organizational and economic systems. Samara State Aerospace University, 2012; 9: 35-41. [In Russian]
2. Kotenko AP, Bukarenko MB. Geometry of Systems of Linear Regression Equations. News of Samara Science Center or Russian Academy of Science, 2013; Vol. 15; 6(3): 820-823. [In Russian]
3. Kotenko AP, Kuznetsova OA. The Use of Multivariate Regression Analysis for Optimizing of the Production of Bitumen of Standardized Specifications. Modern Information Technology and IT-Education, Proc. of 10th Conf., Moscow State University, 2015: 356-359. [In Russian]
4. Dokuchaev AV, Kotenko AA. Software for Multi-criteria Optimization of Production with the Help of Systems of Regression Equations. Rospatent. Certificate of registration No 2016612606. [In Russian]
5. Tukilina PM, Melnikov VN, Tyschenko VA, Ermakov VV, Pimenov AA. The Use of Multivariate Data Analysis Method in the Development of Production Technology of High-quality Road Bitumen. Chemistry and Technology of Fuels and Oils, 2015; 5: 18-23. [In Russian]