Mathematical model for identifying the leading geopolitical actor by the principal component analysis

Yu A Pichugin¹, O A Malafeyev², I V Zaitseva²,³, A I Voskoboev⁴ and V V Shabaev⁴

¹Saint-Petersburg State University of Aerospace Instrumentation Saint-Petersburg, 67, Bolshaya Morskaya str., Saint Petersburg, 190000, Russia
²Saint-Petersburg State University, 7/9 Universitetskaya nab., St. Petersburg, 199034, Russia
³Stavropol State Agrarian University, Zootekhnichesky Lane, 12, Stavropol, 355017, Russia
⁴Rostov Law Institute of the Ministry of Internal Affairs of the Russian Federation, Eremenko street, 83, Rostov-on-don, 344015, Russia

E-mail yury-pichugin@mail.ru, o.malafeev@spbu.ru, d.kolesov@spbu.ru, irina.zaitseva.stv@yandex.ru, alexandrivanych@mail.ru, vadim.shabaev.88@bk.ru

Abstract. In this paper, an approach is developed that allows one to solve an applied problem of identifying hidden factors of geopolitical influence. It uses the principal component analysis. The solution of the problem is based on the principal components and the method of informative selection of the components of the response of the linear regression model. A numerical example is given on the basis of data on the cost of armament of a number of leading countries.

1. Introduction
Solving geopolitical problems is an urgent task of modern international life. Differences in opinions and assessments on geopolitical issues are sometimes a sign of the lack of objectivity in these assessments. Such objectivity can be achieved through the analysis and use of models of multidimensional mathematical statistics, such as the regression model for sample principal components and the model of informative selection and ordering of the response components. The principal component analysis has the beginning of the problem of K. Pearson [1] on finding a line that delivers a minimum sum of squares of distances from this line to a given finite set of points. Generalization of this problem has provided a way to solve a number of statistical problems. This paper develops an approach to solve the applied problem of identifying hidden factors of geopolitical influence. The principal component method is used in conjunction with the estimation of Shannon information quantity. An example of the construction and application of such models based on data on the costs of arms of the leading countries of the world is given below.

2. Model of hidden factors of geopolitical influence based on the principal components analysis
The set of main geopolitical actors \{A_i, i=1, 2, \ldots, m\} is considered. Let us call the geopolitical financial component of the i-th actor for the j-th unit of time (for example, a year) the sum of the
financial costs of armaments, participation in armed conflicts, including the provision of military assistance, and the provision of any financial assistance to other countries, including lending. Thus, \( y_{ij} \) is financial resources spent on direct or potential actions (development and production of new types of weapons) in the external geopolitical environment.

Let \( Y \) be the \( m \times n \)-matrix of values of geopolitical financial components \( y_{ij} \) over the observation period \( n \) years (\( j = 1, 2, \ldots, n \)). The vector \( y_j \) is the \( j \)-th column of the matrix \( Y \). Suppose that the vectors \( y_j (j = 1, 2, \ldots, n) \) are realizations of the random vector \( y \). Also suppose that \( y \) is distributed according to a multidimensional Gaussian distribution with parameters \( \theta \) and \( V_y \), i.e., \( y \sim N(\theta_y, V_y) \).

Calculate estimates of distribution parameters

\[
\hat{\theta}_y = \frac{1}{n} \sum_{j=1}^{n} y_j, \quad \hat{V}_y = \frac{1}{n-1} \sum_{j=1}^{n} (y_j - \hat{\theta}_y)(y_j - \hat{\theta}_y)^T.
\]  

(1)

Here \( T \) is a transposition operator. Next, we calculate the orthogonal matrix \( \hat{P} \) such that the equality

\[
\hat{P}^T \hat{V}_y \hat{P} = \Lambda = \text{diag}(\hat{\lambda}_1, \hat{\lambda}_2, \ldots, \hat{\lambda}_m)
\]  

(2)

holds, where \( \hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \ldots \geq \hat{\lambda}_m \).

In this situation, generally speaking, the estimate \( \hat{V}_y \) is not equal to matrix \( V_y \) and, therefore, the matrix \( \hat{P} \) does not necessarily coincide with a matrix \( P \) such that

\[
P^T V_y P = \Lambda = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_m).
\]

Define a vector of dimension \( k \) (\( k < m \)):

\[
z = P_{(k)} (y - \hat{\theta}_y),
\]

where the matrix \( P_{(k)} \) contains only the first \( k \) columns of the matrix \( \hat{P} \). Strict selection of the number \( k \) is associated with testing the statistical hypothesis \( H: \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_k \geq \lambda_{k+1} = \lambda_{k+2} = \ldots = \lambda_m \).

However, there are more simple ways to determine \( k \) [2]. Consider a linear regression model

\[
y = \theta_y + P_{(k)} z + \varepsilon.
\]  

(3)

Model (3) is an approximation of a hypothetical model

\[
y = \theta_y + P_{(k)} z + \varepsilon.
\]  

(4)

The components of the vector \( z \) in model (4) are called the principal components. In model (3), the components of the vector \( z \) are called the sample principal components. Here we assume that model (3) is a fairly accurate approximation of model (4). The components of the vector \( z \) are mutually independent. The estimation of the matrix of mutual covariances of the components \( z \) is

\[
\hat{V}_z = \hat{\Lambda}_{(k)} = \text{diag}(\hat{\lambda}_1, \hat{\lambda}_2, \ldots, \hat{\lambda}_k).
\]

Here, the components of the vector \( z \) are interpreted as hidden geopolitical factors.

We assume that in model (3) and in model (4), the error vector is distributed over the Gaussian multidimensional distribution

\[
\varepsilon \sim N(0, \sigma^2 I),
\]  

(5)

where \( 0 \) is the zero vector, and \( I \) is an identity matrix of the corresponding dimension. Assumption (5) means that \( E(\varepsilon_i) = 0, \quad \text{var}(\varepsilon_i) = \sigma^2, \quad (i = 1,2,\ldots,m) \) and \( \text{cov}(\varepsilon_i, \varepsilon_j) = 0 \quad (i \neq j) \), where \( E, \text{var} \) and \( \text{cov} \) are operators of mathematical expectation, variance and covariance, respectively.
3. Model of informative selection and informative ordering of geopolitical actors according to their geopolitical influence

Change the notation of the linear regression model (3) as follows

\[ y = Fz + \varepsilon. \]  

where \( y \) is the vector of deviations of the geopolitical financial components from the average for the observed period (n years, see above) values.

Let \( q \) be a subset of numbers in the list of geopolitical actors, i.e. \( q \subseteq \{1, 2, \ldots, m\} \). Then \( y_q \) is a vector of dimension \( \text{card} q \) (\( \text{card} q \) is the number of elements in the subset \( q \)), which contains components \( y \) only with numbers from the set \( q \), and \( F_q \) is a matrix that contains only rows with numbers from the set \( q \). Using the vector \( y_q \), we can calculate the OLS-estimate (OLS - Ordinary Least Squares) of the vector of hidden geopolitical factors \( \hat{z} = (F_q^T F_q)^{-1} F_q^T y_q \).

The covariance matrix of the OLS-estimate \( \hat{z} \) is equal to \( V_{\hat{z}} = \sigma^2 (F_q^T F_q)^{-1} \), and the inverse of this matrix is the Fisher information matrix \( M_q = \sigma^{-2} F_q^T F_q = V_{\hat{z}}^{-1} \).

For the selected subset \( q \) (see above), the Shannon information quantity contained in the components of the vector \( y_q \) relative to the components \( z \) is equal to (see [3, 4])

\[ I_s(y_q,z) = \frac{1}{2} \log_a \det(I + M_q \Lambda_{(k)}) . \]  

Here the choice of the base of the logarithm \( a \) is the choice of the scale of units of information. It can be shown that an increase in the amount of information (an increase in the number of elements \( q \)) leads to an increase in the accuracy of estimating the vector of hidden geopolitical factors. This means an increase in the accuracy of reproducing the geopolitical picture (see model (6)).

As \( \sigma^2 \) of formula (7), one should use the biased estimate \( \hat{\sigma}^2 = m^{-1} (\text{tr} V_y - \text{tr} \Lambda_{(k)}) \) (tr is a trace of the matrix), since the calculation of information is based on biased estimation and, in the case of a Gaussian distribution, has a geometric interpretation. Increasing \( \text{card} q \) incrementally from zero to \( m \) and, maximizing the value of \( I_s(y_q,z) \) at each step, we obtain a sequence of information values \( I_j \) \((j=1, 2, \ldots, m)\). It is convenient to consider these values in relative units (excluding the dependence on the choice of the base of the logarithm \( a \), see (7)), that is \( I_j = (I_j/I_s(y,z))100\% \), where \( I_s(y,z) = I_s(y_q,z) \) for \( q = \{1, 2, \ldots, m\} \) there is all the information quantity.

As a result, we get a sequence of geopolitical actors \( A_{(j)} \) in descending order of their geopolitical significance. The value of \( \varphi(j) \) is determined by the maximization of \( I_s(y_q,z) \) at the \( j \)-th selection step (see above).

4. Numerical example

Table 1 presents data on the expenditures on armaments of a number of leading countries of the world for the period 2006–2018. Data on the issuance of foreign loans is difficult to find in open sources, so we are forced to consider only the power component of the geopolitical influence.

Table 1. (a) Data on the expenditures on armaments of a number of leading countries of the world for the period 2006–2018 (billions of dollars).

| Countries | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 |
|-----------|------|------|------|------|------|------|------|
|           |      |      |      |      |      |      |      |
Table 1. (b) Data on the expenditures on armaments of a number of leading countries of the world for the period 2006–2018 (billions of dollars).

| Countries          | 2013  | 2014  | 2015  | 2016  | 2017  | 2018  |
|--------------------|-------|-------|-------|-------|-------|-------|
| USA                | 639.7 | 609.9 | 596   | 611   | 610.0 | 714   |
| China              | 177.9 | 199.7 | 214.8 | 215   | 228.0 | 175   |
| Saudi Arabia       | 67    | 80.8  | 87.2  | 63.7  | 69.4  | 56    |
| Russia             | 90.4  | 85.7  | 67.4  | 69.2  | 66.3  | 46    |
| India              | 47.4  | 50.9  | 51.3  | 55.9  | 63.9  | 63.9  |
| France             | 62.4  | 63.6  | 50.9  | 50.4  | 57.8  | 40    |
| Britain            | 56.9  | 59.2  | 55.5  | 55.5  | 47.2  | 60    |
| Japan              | 49    | 45.9  | 40.9  | 50.4  | 45.4  | 46.1  |
| Germany            | 45.9  | 46.1  | 39.4  | 40.3  | 44.3  | 44.8  |
| Korean republic    | 34.4  | 37.3  | 36.4  | 36.8  | 39.2  | 39.2  |
| Brazil             | 32.9  | 32.7  | 24.6  | 26.2  | 29.3  | 29.3  |
| Italy              | 33.9  | 31.6  | 23.8  | 27.5  | 29.2  | 37.7  |
| Australia          | 24.8  | 25.8  | 23.6  | 24.5  | 27.5  | 26.3  |
| Israel             | 16.9  | 18.1  | 16.1  | 19    | 19.6  | 20    |

By successively calculating the basis of principal components (matrix columns $F = \hat{P}_{(k)}$, $k = 2$), the following was found. The variance of the first two principal components ($\hat{\lambda}_1$ and $\hat{\lambda}_2$) is 96% of the total variance of 58% and 38%, respectively. This means that geopolitical dynamics is determined by two hidden geopolitical factors.

Figure 1 presents the growth of the amount of information regarding geopolitical factors with consistent informative selection. From figure 1 it is clear that China, the USA and Saudi Arabia deliver almost all the information about geopolitical dynamics. Consequently, it is these three countries that have the greatest influence on geopolitical development in the main regions of the world as a whole.
Figure 1. Growth in the amount of information ($I\%$) on geopolitical factors in order of informative selection of actors.

Figure 2 presents graphically the basis vectors $F_1$ and $F_2$ of geopolitical factors $z_1(a)$ and $z_2(b)$.

Figure 2 shows that in the first factor, China has the greatest value (see figure 2a), and in the second factor, the United States has the greatest value (see figure 2b). Therefore, it is natural to call the factors Chinese and American, respectively.

Figure 3 (see below) shows a temporary change in the geopolitical factors $z_1$ and $z_2$. The Chinese factor ($z_1$) shows relatively stable growth, and the American factor ($z_2$) shows a fluctuation with a period of $\approx 11$ years. The Chinese factor reveals only a slight temporary decline in synchronization with the American factor (see figure 3, 2013-2015).
5. Conclusions

Based on the above, we can conclude that the statistical model constructed on the basis of the principal components method, as demonstrated by the numerical example, is an effective tool for analyzing geopolitical dynamics. The model made it possible to identify the leading geopolitical factors in importance and to consider their dynamic interaction, as well as to identify the countries that have the greatest impact on geopolitical development in the main regions of the world as a whole. There is no doubt that the use of static models makes it possible to make the most objective assessments and conclusions about the geopolitical dynamics. Other aspects of mathematical simulating geopolitical processes are considered in the papers of the authors (see [5-27]).

Acknowledgment

The work is partially supported by the RFBR grant # 18-01-00796.

References

[1] Pearson K 1901 On lines and planes of closest fit to systems of points in space Philosophical Magazine pp 559–72

[2] Pichugin Yu 2018 Notes on using the principal components in the mathematical simulation St. Petersburg Polytechnical University Journal: Physics and Mathematics 11(3) 74–89

[3] Pichugin Yu 2002 Notes on the selection of data related multiple linear regression Industrial Laboratory. Diagnostics of materials 5 61-2

[4] Pichugin Yu and Malafeyev O and Rylow D and Zaitseva I 2017 A statistical method for corrupt agents detection International Conference of Numerical Analysis and Applied Mathematics ICNAAM 2017 1978 100014

[5] Kriulina E and Tarasenko N and Miroshnitchenko N and Zaitseva I and Dedyukhina I 2016 Environmental Management in Agriculture: Problems and Solutions Research Journal of Pharmaceutical, Biological and Chemical Sciences 7(3) 1908-12

[6] Zaitseva I and Popova M 2013 Technique to study the employment potential of the region: economic-mathematical aspect World Applied Sciences Journal 22(1) 22-5

[7] Zaitseva I, Kriulina E, Ermakova A, Shevchenko E and Vorokhobina Y 2016 Application of Factor Analysis to Study the Labour Capacity of Stavropol Krai Research Journal of Pharmaceutical, Biological and Chemical Sciences 7(4) 2183-6

[8] Zaitseva I, Popova M, Ermakova A, Bogdanova S and Rezenkov D 2016 Determination Prospects Of Development Labor Potential In Agriculture Stavropol Territory Based On Assessment His Condition Research Journal of Pharmaceutical, Biological and Chemical Sciences 7(3) 2592-5

[9] Zaitseva I, Malafeyev O, Strekopytov S, Ermakova A and Shlaev D 2018 Game-theoretical model of labour force training Journal of Theoretical and Applied Information Technology

[10] Zaitseva I, Malafeyev O, Strekopytov S, Bondarenko G and Lovyannikov D 2018 Mathematic model of regional economy development by the final result of labor resources AIP
Conference Proceedings

[11] Zaitseva I, Ermakova A, Shlaev D, Shevchenko E and Lugovskoy S 2017 Workforce planning redistribution of the region's results Research Journal of Pharmaceutical, Biological and Chemical Sciences 8(1) 1862-6

[12] Kostyukov K, Zaitseva I, Bondarenko G, Svechinskaya T and Nechayeva S 2016 Workforce Planning as An Element of Control System Research Journal of Pharmaceutical, Biological and Chemical Sciences 7(6) 2315–9

[13] Zaitseva I, Malafeyev O, Poddubnaya N, Vanina A and Novikova E 2019 Solving a dynamic assignment problem in the socio-economic system Journal of Physics: Conference Series

[14] Zaitseva I, Malafeyev O, Marenchuk Y, Kolesov D and Bogdanova S 2019 Competitive Mechanism for the Distribution of Labor Resources in the Transport Objective Journal of Physics: Conference Series

[15] Malafeyev O, Lakhina J, Zaitseva I, Redinskikh N, Smirnova T and Smirnov N 2019 A mathematical model of production facilities location Journal of Physics: Conference Series

[16] Malafeyev O, Redinskikh N, Nemnyugin S, Kolesin I and Zaitseva I 2018 The optimization problem of preventive equipment repair planning AIP Conference Proceedings 1978 100013

[17] Malafeyev O, Awasthi A, Zaitseva I, Rezenkov D and Bogdanova S 2018 A dynamic model of functioning of a bank AIP Conference Proceedings, American Institute of Physics

[18] Malafeyev O, Rylow D, Zaitseva I, Popova M and Novozhilova L 2017 Game-theoretic model of dispersed material drying process AIP Conference Proceedings

[19] Malafeyev O, Saifullina D, Ivaniukovich G, Marakhov V and Zaytseva I 2017 The model of multi-agent interaction in a transportation problem with a corruption component AIP Conference Proceedings, American Institute of Physics

[20] Malafeyev O, Farvazov K, Zenovich O, Zaitseva I et al 2018 Geopolitical model of investment power station construction project implementation AIP Conference Proceedings

[21] Zaitseva I, Malafeyev O, Kolesin I, Ermakova A and Shlaev D 2018 Modeling of the labour force redistribution in investment projects with account of their delay" International Conference on Power, Control, Signals and Instrumentation Engineering pp 68-70

[22] Pichugin Y, Malafeyev O, Rylow D and Zaitseva I 2018 A statistical method for corrupt agents detection AIP Conference Proceedings

[23] Malafeyev O, Rylow D, Zaitseva I, Ermakova A and Shlaev D 2018 Multistage voting model with alternative elimination AIP Conference Proceedings

[24] Malafeyev O, Awasthi A, Kambekar K and Kupinskaya A 2019 Random walks and Market Efficiency in Chinese and Indian equity markets Statistics, Optimization and Information Computing 7(1) 1-25

[25] Kolokoltsov V N and Malafeyev O A 2018 Corruption and botnet defense: a mean field game approach International Journal of Game Theory 1-23

[26] Malafeyev O and Redinskikh N 2017 Quality estimation of the geopolitical actor development strategy 2017 Constructive Nonsmooth Analysis and Related Topics, Dedicated to the Memory of V F Demyanov Available from http://dx.doi.org/10.1109/CNSA.2017.7973986

[27] Kolokoltsov V N and Malafeyev O A 2017 Mean-Field-Game Model of Corruption Dynamic Games and Applications 7 (1) 34-47