Evaluation of the linear models’ use in the forecasting investment decisions in the construction sector of the Russian Federation

I N Geraskina and M S Egorova

St. Petersburg State University of Architecture and Civil Engineering, 4, Vtoraja Krasnoarmeiskaya Str., St Petersburg, 190005, Russia

E-mail: Geraskina82@mail.ru

Abstract. A model of multiple linear regression is constructed using the data from the resulting indicators over the past 27 years with the imperative of forecasting the ICC trend and proving the impossibility of its use in predicting the results of innovative and investment decisions in the construction sector of the economy in the Russian Federation, since it does not take into account its following important properties: stochasticity, connections’ nonlinearity, systemic potential, high dynamism of the structure and the desire for the information distortion. The relevance of the regression-differential approach in modeling the investment and construction sector and the support of strategic decision-making is substantiated.

Introduction

The emerging socio-economic crisis will be accompanied by the significant structural dynamics, an increase in the entropy level and chaos in the system, fluctuation waves, leading to the certain world economic relations’ destruction and the emergence of the new ones. In such circumstances, the country’s top leadership in order to maintain the economy and ensure the survival of its subsystems is in need of the development programs and managerial decisions based on the objective forecast data. Of course, the solution of such strategic economic problems requires scientific understanding and development of the economic and mathematical modeling methodological foundations. For the sustainable development, the Russian economy needs the formation of an investment and construction complex (ICC), which is characterized by a significant share in the GDP structure, resource and energy efficiency, technical and technological innovations. However, for today ICC is a self-regulating socio-economic system in a state of sluggish recession, structural imbalance, with a pronounced monopoly, complicated by the project financing of housing construction, which together constrain the investment flows and the necessary production technology modernization.

The purpose

The purpose of the study is to prove the impossibility of using the multiple linear regression models when predicting the ICC trend, since they do not take into account its following main properties: cyclicity and inertia, a high degree of openness and stochasticity, nonlinear parameter relationships and deep processes that form the system potential, as well as a change in the response to the influencing factors’ dynamics, etc.
Methods
Statistical data approximation, regression and phase analysis, system synthesis, linear modeling and visualization of the achieved results in graphical form.

Main part
A comprehensive ICC analysis and the main indicators of its dynamics made it possible to distinguish the following phase variables - indicators that maximally characterize its properties and trend. First of all, this is a statistical indicator “Commissioning of buildings, structures, individual production facilities, houses, social and cultural facilities”, measured in million m²; then “The share of construction in GDP”; as well as the “Share of profitable in the total number of construction organizations”, measured in% [1, 2, 3]. In addition, a spectrum of control parameters that have the maximum effect on the phase variables of the ICC was identified (Table 1) [1].

Table 1. Result indicators.

| Name | Indicator |
|------|-----------|
| Phase variables | |
| Y1   | Commissioning of buildings, structures, individual production facilities, houses, social and cultural facilities, mln m² |
| Y2   | Share of construction in GDP, % |
| Y3   | The profitable share in the total number of construction organizations, % |
| Control parameters | |
| X1   | Investments in fixed assets, million rubles (until 1998 - billion rubles) |
| X2   | The volume of mortgage loans, billion rubles. |
| X3   | The cost of 1 rub. work cop. |
| X4   | The population of the RSFSR / RF, people. |
| X5   | The average monthly salary of workers in construction organizations, thousand rubles. Construction industry production volume index materials |
| X6   | The composite price index for the main types of building materials and works |
| X7   | Per capita income, rubles per month (until 1998 in thousand rubles) |
| X8   | Self-sufficiency of construction sector organizations |
| X9   | The share of expenses on the acquisition of real estate in the cash expenses of the population |
| X10  | Refinancing rate, % |
| X11  | Annual inflation in the Russian Federation, % |
| X12  | Volumes of housing construction, thousand m² |
| X13  | Labor productivity index |
| X14  | Profitability level in construction, % |
| X15  | Average prices 1 m² total area in the primary housing market, rubles |
| X16  | The presence of fixed assets, billion rubles. |
| X17  | Number of operating construction organizations, units |
| X18  | |

In the linear multifactor model (LMM), the final indicator “Commissioning of buildings, structures, individual production facilities, houses, social and cultural facilities” is used, measured in million m². A certain time interval for modeling and forecasting the object of study is divided into the enlarged periods (Tables 2-4), corresponding to the bifurcation moments in the socio-economic development of the USSR and the Russian Federation: 1) 1990 - 1998; 2) 1999 - 2007; 3) 2008 – 2017. When modeling and predicting the complex economic systems’ trends, the multifactor models, where the value of one socio-economic indicator or group is determined by other factors that simultaneously affect, to varying degrees, the outcome, are used [4].

Table 2. The values of the phase variable and the control parameters in 1990 - 1998.
| Indicator | 1990 | 1991 | 1992 | 1993 | 1994 | 1995 | 1996 | 1997 | 1998 |
|-----------|------|------|------|------|------|------|------|------|------|
| Y1        | 61.7 | 49.4 | 41.5 | 41.8 | 39.2 | 41.0 | 34.3 | 32.7 | 40.8 |
| X1        | 249.0| 211.0| 267.0| 27125.0| 108810.0| 266974.0| 375958.0| 408797.0| 407086.0|
| X2        | 85.0 | 84.0 | 83.0 | 80.0 | 81.0 | 82.0 | 90.0 | 90.0 | 94.0 |
| X3        | 147.7| 148.3| 148.7| 148.5| 148.3| 148.1| 148.3| 147.9| 147.7|
| X4        | 0.4  | 0.7  | 8.1  | 78.0 | 283.3| 595.1| 967.4| 1221.0| 1.3  |
| X5        | 102.3| 101.3| 100.1| 95.2 | 87.2 | 92.0 | 74.5 | 95.9 | 94.8 |
| X6        | 115.0| 112.3| 110.5| 131.2| 120.5| 140.1| 105.0| 112.0| 106.8|
| X7        | 0.2  | 0.5  | 47.8 | 542.1| 2476.1| 515.9| 769.5| 940.6| 1010.2|
| X8        | 50.0 | 50.2 | 50.4 | 41.7 | 15.1 | 52.3 | 52.0 | 52.0 | 53.2 |
| X9        | 1.2  | 1.4  | 1.5  | 51.1 | 52.0 | 1.7  | 1.7  | 1.9  | 2.5  |
| X10       | 25.0 | 25.0 | 80.0 | 210.0| 180.0| 160.0| 48.0 | 28.0 | 60.0 |
| X11       | 160.7| 160.4| 2508.8| 840.0| 214.8| 131.6| 21.8 | 11.0 | 84.5 |
| X12       | 61694.5| 49422.6| 41518.3| 41808.2| 39224.3| 41036.4| 34300.9| 32702.6| 30684.6|
| X13       | 101.0| 100.0| 99.6 | 101.5| 100.3| 100.5| 100.6| 101.0| 101.1|
| X14       | 20.1 | 20.2 | 20.4 | 0.2  | 18.6 | 20.2 | 23.3 | 11.6 | 6.5  |
| X15       | -    | -    | -    | -    | -    | 1012.0| 1846.0| 2161.0| 3075.0|
| X16       | 187.1| 189.5| 190.3| 200.0| 210.0| 221.0| 521.0| 434.0| 446.5|
| X17       | 7086.0| 6975.0| 6905.0| 9594.0| 12491.0| 12776.0| 13462.0| 13697.0| 13716.0|

Table 3. The values of the phase variable and control parameters in 1999 – 2007.

| Indicator | 1999 | 2000 | 2001 | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 |
|-----------|------|------|------|------|------|------|------|------|------|
| Y1        | 42.1 | 44.7 | 47.7 | 49.6 | 53.7 | 60   | 66.3 | 75.6 | 98.1 |
| X1        | 67044.0| 116523.0| 150471.0| 176241.0| 218636.0| 286501.0| 361110.0| 473002.0| 671622.0|
| X2        | 1.6  | 8.2  | 15.3 | 50.0 | 100.0| 214.0| 56.3 | 263.6| 564.6|
| X3        | 93.0 | 92.0 | 92.0 | 93.0 | 94.0 | 94.0 | 88.0 | 87.0 | 91.0 |
| X4        | 147.2| 146.6| 146.3| 145.2| 145.0| 144.3| 143.8| 143.2| 142.8|
| X5        | 1.8  | 2.8  | 4.2  | 4.8  | 6.2  | 7.3  | 9.0  | 11.3 | 14.3 |
| X6        | 110.2| 113.1| 105.5| 103.7| 107.0| 106.1| 102.7| 103.1| 103.1|
| X7        | 131.2| 137.4| 145.0| 124.0| 111.2| 112.0| 112.0| 114.9| 121.2|
| X8        | 1658.9| 2281.1| 3062.0| 3947.2| 5170.4| 6410.3| 8088.3| 10154.8| 12540.2|
| X9        | 52.4 | 47.5 | 49.4 | 45.0 | 45.2 | 45.4 | 44.5 | 42.1 | 40.4 |
| X10       | 5.3  | 7.5  | 8.9  | 10.9 | 12.7 | 11.0 | 10.4 | 10.3 | 9.6  |
| X11       | 55.0 | 25.0 | 25.0 | 21.0 | 16.0 | 13.0 | 12.0 | 11.0 | 10.0 |
| X12       | 36.6 | 21.0 | 18.8 | 15.1 | 11.9 | 10.7 | 9.0  | 9.0  | 11.9 |
| X13       | 3201.6| 30295.8| 31703.2| 33832.2| 36449.3| 40410.1| 43559.5| 50552.1| 61221.3|
| X14       | 101.4| 102.0| 101.4| 101.2| 105.3| 106.8| 105.9| 115.8| 112.8|
| X15       | 7.0  | 7.2  | 6.0  | 6.2  | 5.7  | 4.2  | 3.9  | 5.1  | 5.8  |
| X16       | 4000.0| 8254.9| 9724.8| 9014.8| 15275.3| 19660.8| 23771.6| 31473.6| 43883.2|
| X17       | 413.7| 301.5| 337.2| 363.1| 403.3| 431.7| 604.9| 711.3| 992.9 |
| X18       | 130846.0| 129340.0| 118374.0| 112971.0| 1113578.0| 114464.0| 112640.0| 122441.0| 131394|

Table 4. The values of the phase variable and the control parameters in 2008 – 2016.
We apply this approach [5–7] with the goal of constructing an LMM and testing its applied capabilities in the economy. To bring them to a single dimension, the factors and criteria were normalized by the Formula 1:

$$y = \frac{y(t) - \min_y y(t)}{\max_y y(t) - \min_y y(t)} \quad (1)$$

where \( \min_y y(t) \) – is the minimum criterion \( Y \) depending on the year, \( \max_y y(t) \) – is the maximum criterion \( Y \) depending on the year.

The normalized factors and the criterion for the indicated periods are presented in Tables 1 - 3. Next, the factors of the model were analyzed from the standpoint of their mutual correlation in order to exclude the obvious interdependencies. The pair correlation of the data time series was calculated by the Formula 2:

$$\rho = \frac{\sum (x - \bar{x})(y - \bar{y})}{n \sigma_x \sigma_y} \quad (2)$$

where \( x, y \) – are the values of factor and effective indicators, respectively, \( \bar{x}, \bar{y} \) – are the average values of relevant indicators, \( \sigma_x, \sigma_y \) – are the standard deviations (typical deviations of variables \( x \) and \( y \)), verified by the Formula 3:

$$\sigma_x = \sqrt{\frac{\sum (x - \bar{x})^2}{n}}, \quad \sigma_y = \sqrt{\frac{\sum (y - \bar{y})^2}{n}} \quad (3)$$

where \( n \) – defines the number of observations.

LMM results made it possible to identify the factors that have a high cross-correlation:

1. In 1990 - 1997 the factors X9 and X18 correlated with a coefficient of 0.959823. In such a situation, it is logical to exclude X18, because then it showed an already low dependence with the criterion.

2. In 1998 - 2007 the dependencies of the following content were found: a) in the factors X1 and X5, X8, X13, X16 and X17 the high cross-correlation is noted, therefore it is advisable to exclude X1; b) in the factors X5 and X13 exclude the first because X13 in 1990 - 1997 and 2008 - 2016 had the maximum value; c) between X5 and X16 it is more expedient to exclude the latter in order to preserve the more important X17; d) between X8 and X4 in order to minimize the number of factors eliminated from the field of view, it turned out to be logical to remove X6; e) between X12 and X11, X12 is excluded, since in 2008 - 2016 it was of high importance.

3. In 2008 - 2016 X8 was removed between X8 and X9 by the analogy with the period 1998 - 2008. It should be noted that in addition to this, other indicators were excluded, characterized by a low cross-
correlation of factor and criterion: a) in 1990 - 1997 – the model factors $X_7$, $X_{12}$ and $X_{14}$; b) in 1998 - 2007 – the model factor $X_{18}$; c) in 2008 – 2016 – the model factor $X_{10}$ and $X_{11}$.

Built LMM on the “Commissioning of buildings, structures, individual production facilities, houses, objects of socio-cultural purpose” (Formula 4):

$$y_{cal}(t_i) = a_0 + \sum_j a_j x_j(t_i),$$

where $a_0$ – is an independent coefficient, $a_i$ – defines the influence factors $i$-x factors $x_i(t)$ at time (year) $t$ for the criterion value.

It is advisable to determine the LMM coefficients by minimizing the quadratic deviation of the statistical data from the calculated ones according to the Formula 5:

$$S = \sum t(y(t) - y_{cal}(t))^2 \rightarrow \min$$

This will make it possible to identify the LMM regression and to find out which factors in their linear combination have the greatest influence on the response of the system, and which factors have the least influence (Table 5, 6). An analysis of the data made it possible to understand that in the first version of the LMM, $X_{13}$ turned out to be the determining indicator, which is quite obvious, since $K$ is a housing stock that was built and commissioned. In this regard, we will exclude it from the LMM and repeat the calculations.

| Time interval | A0  | A1  | A2  | A3  | A4  | A5  | A6  | A7  | A8  |
|---------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 1990 – 1997   | 0.241 | -0.069 | - | 0.158 | -0.258 | -0.112 | -0.026 | -0.033 | - |
| 1998 – 2007   | 0.070 | - | 0.190 | -0.034 | -0.054 | - | 0.055 | 0.074 | - |
| 2008 – 2016   | 0.019 | 0.136 | 0.164 | 0.142 | 0.070 | 0.003 | 0.088 | -0.282 | - |
| A9            | A10 | A11 | A12 | A13 | A14 | A15 | A16 | A17 | - |
| 1990 – 1997   | -0.118 | -0.183 | 0.022 | - | 0.569 | - | 0.197 | -0.061 | 0.096 |
| 1998 – 2007   | -0.002 | 0.074 | -0.119 | - | 0.314 | -0.079 | -0.044 | - | 0.466 |
| 2008 – 2016   | 0.073 | - | - | 0.175 | 0.300 | 0.037 | -0.157 | -0.041 | 0.061 |

| Time interval | A0  | A1  | A2  | A3  | A4  | A5  | A6  | A7  | A8  |
|---------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 1990 – 1997   | 0.589 | -0.131 | - | 0.336 | -0.567 | -0.368 | -0.044 | - | -0.077 |
| 1998 – 2007   | 0.131 | - | 0.285 | -0.046 | -0.096 | - | 0.056 | 0.076 | - |
| 2008 – 2016   | -0.003 | 0.144 | 0.255 | 0.231 | 0.106 | 0.005 | 0.133 | -0.359 | - |
| A9            | A10 | A11 | A12 | A13 | A14 | A15 | A16 | A17 | A18 |
| 1990 – 1997   | -0.203 | -0.348 | 0.049 | - | - | 0.414 | -0.084 | 0.225 | - |
| 1998 – 2007   | -0.024 | 0.085 | -0.155 | 0.068 | -0.056 | -0.073 | - | 0.615 | - |
| 2008 – 2016   | 0.093 | - | - | 0.261 | 0.056 | -0.244 | -0.053 | 0.101 | 0.009 |

Objective and interesting for top management of ICC are the results [8-11]:

1. The factor $X_2$ (the volume of mortgage loans issued) since 1999 has been steadily significant. The emergence and development of the mortgage lending mechanism after the collapse of the USSR had an undoubtedly positive impact on the growth of housing construction in the structure of the ICC RF, regardless of other economic situations’ dynamics.

2. The influence factor $X_4$ (the country’s population) increased linearly from -0.6 to +0.1, which marked the beginning of changes in public opinion regarding the need and quantity of privatized housing, as well as their income. Since the days of the RSFSR, the bulk of society has sought in the household to improve the living conditions and obtain the ownership of more space, which stimulated the development of the housing market.

3. The factor $X_6$ (the volume of the building materials’ production) to a greater extent influenced the performance of housing after the economic crisis of 2008. This is objective, since most of the building
materials’ production falls primarily on the housing construction and civil infrastructure, and then on other objects.

4. The price dynamics index \( X_7 \), on the contrary, after the 2008 economic crisis acted as a deterrent one in the transformation of the housing market. Indeed, up to this point, most citizens had money and could spend it on improving the housing conditions and purchasing the additional real estate.

5. The fixed assets security effect \( X_9 \) the dynamics of housing construction has decreased and even ceased to slow down growth, which is typical for the period until 1998. This is similar to the conclusion on the factor \( X_6 \), when the industrial facilities were completed according to the plans of the USSR until 1998 by the largest construction organizations. With the optimization of tax legislation in the Russian Federation, the use of subcontracting works was rationalized, which resulted in ignoring the provision level of the general contractor with own fixed assets. A similar trend was observed with the factor \( X_{17} \), which until the early 2000s was steadily of great importance. Since the formation of tender committees and the contract tendering’s entry into force, the priorities in the construction industry have changed and the choice of contractors is based on the results of competitive bidders, which reduce the work cost, terms of contract execution, provide the obligations to improve the construction quality, etc.

6. The factor \( X_{12} \) (inflation), firstly, showed an acceleration of the main statistical indicators of the housing construction, and secondly, in an unstable economic environment, investing the financial resources in private property is not only one of the most common, but also convenient ways to store free cash. We are sure that in this case there is a deviation of the real and official inflation indicators.

7. The factor \( X_{15} \) (profitability level) since the time of the RSFSR, has consistently and continuously exerted a deterrent effect on housing construction. We believe that here there is a result of the “shadow” financial schemes and mechanisms’ implementation, which was quite natural in the ICC (housing, industrial, road, etc.) at this time.

Thus, the analysis of the LMM results makes it possible to explain with some degree of probability the certain events occurring in the ICC. However, to assess the possibility of their use in predicting the impact of organizational, managerial, technical, technological and other innovations on ICC, it is necessary to evaluate the predictive properties of this model using the widespread method of post-forecast for 1, 2, and 3 years (Figures 1-3). The graphic visualization indicates that even for one year the models of the 1st and 2nd segments are difficult to predict, and the 3rd segment model is completely at odds with the 3-year post-forecast.

Summary
The results of evaluating the possibility of obtaining the highest possible achievable linear regression effects indicate the LMM uselessness in predicting the ICC dynamics. LMMs are not able to take into account the inherent properties of the large economic systems due to some factors: a high degree of openness and complexity; inherent cyclicity, fluctuation and hysteresis processes, constant entropy dynamics.
Acknowledgements
This article was prepared as part of the government contract as requested by the Ministry of Science and Higher Education of the Russian Federation on the subject formulated as «Structural changes in economy and society as a result of achieving the target indicators of National projects, which provide opportunities to organize new areas of social and economic activity, including commercial, both in Russia and abroad» (project No. FSSW-2020-0010).

References
[1] Geraskina I N, Zatonskiy A V, Petrov A A 2017 Modeling of the investment and construction trend in Russia International Journal of Civil Engineering and Technology (IJCIET) 8 (10) 1432–1447.
[2] Geraskina I N, Zatonsky A V 2017 Modeling the trend of investment and construction activities of the Russian Federation MGSU Herald 12 (11) (110) 1229–1239.
[3] Federal State Statistics Service. Information on http://www.gks.ru/free_doc/doc_2016/stroit_2016.pdf (access date: 05.02.2020)
[4] Mitsek E B 2020 Econometric modeling of investments in fixed assets of the Russian economy Information on http://dissers.ru/avtoreferati-dissertatsii-ekonomika/2/a45. pdf (access date: 15.02.2020)
[5] Akayev A A, Korotaev A V, Malinetskiy G G, Malkov S Yu 2012 *Modeling and forecasting global, regional and national development* (Prince. LIBROCOM House, Moscow) 488.

[6] Akayev A A, Rumyantseva S Yu, Sarygulov A I, Sokolov V N 2016 *Structural and cyclical processes of economic dynamics* (Publishing house Polytechnic University, Saint-Petersburg) 392.

[7] Akayev A A, Korotaev A V, Malinetskiy G G, Malkov S Yu 2011 *Hierarchical system for modeling global dynamics Projects and risks of the future. Concepts, models, tools, forecasts* (KRASAND, Moscow) 208–231.

[8] Panibratov Y, Latukha M, Lisina P 2019 Developing sustainable competitive advantage of a firm through human resource management practices: A competence-based approach *Global Business and Economics Review* 21 (1) 96-119.

[9] Geraskina I N, Kuligin K N 2019 Methodological aspects of providing balanced innovative development of construction complex of the Russian Federation *IOP Conf. Series: Materials Science and Engineering* 698 077046.

[10] Scott W, Gough S 2003 Sustainable Development and Learning: Framing the Issues (London, New York, Routledge Falmer) 192.

[11] Chainikova G R, Zatonskiy A V, Mitiukov N W, Busygina H L 2018 Development of foreign language lexical competence on the basis of a learner's terminological thesaurus and dictionary *European Journal of Contemporary Education* 7 (1) 51-59.