GCPMED 2018

International Scientific Conference "Global Challenges and Prospects of the Modern Economic Development"

REGION CLUSTERING AND MODELING INDICES FOR HOUSING MARKET

Yu.A. Tokarev (a), N.N. Belanova (b)*, O.A. Guzhova (c), G.V. Glukhov (d)
*Corresponding author

(a) Samara State Economic University, Soviet Army Str., 141, 443090, Samara, Russia, e-mail: tokarev_ya@mail.ru
(b) Samara State Economic University, Soviet Army Str., 141, 443090, Samara, Russia, e-mail: bnn371@yandex.ru
(c) Samara State Technical University, Soviet Army Str., 141, 443090, Samara, Russia, e-mail:
guzhova_oksana@inbox.ru
(d) Samara State Technical University, Soviet Army Str., 141, 443090, Samara, Russia, e-mail: kino@sseu.ru

Abstract

This paper demonstrates housing market heterogeneity primarily due to the prices on residential real estate and new housing supply, and the proportion of dilapidated housing. The findings of the study are based on the analysis of means and variation indices on key housing market indicators within regions. The study contributes to the classification of territorial entities of the RF and demonstrates the importance of case specific approach to the state housing policy. The methodology adopted in this paper is based on seven components analysis of 74 territorial entities of the RF. The seven components of the analysis reflect various aspects of the housing market. The results represent five clusters recognized and qualitatively defined. Multiple regression models are generated and critical factors for housing market are identified. Demographic indices are the most influential among others to differentiate territorial housing markets. The next defining factors are natural and geographic ones, e.g. density of population, and living standard factors. The least influential factors include production, infrastructure and social indicators. Clustering approach and due regard to the key factors could be recommended to facilitate the development of the state policy on housing markets. It could increase the efficiency of existing policies and synchronize region development.

Keywords: Housing market, differentiation, grouping, clusters, factors, regional development.

1. Introduction

Housing market is part and parcel of any national economy. It is comprised of various aspects: state social policy, construction economy, demographic processes, etc. The whole cohort of economic researchers has studied market as a complex system. Ge & Wu (2017) demonstrated the properties of housing market and generated a retrieval balanced model. Sedova (2014) modeled housing prices using spatial approach. Antoniucci & Marella (2017) calculated the correlation of polarization on index with city density.

Other authors considered other influential factors, namely unemployment (Gan, Wang, & Zhang, 2018), oil shocks (Killins, Egly, & Escobari, 2017), earthquakes (Cheung, Wetherell, & Whitaker, 2018), monetary policy (Ume, 2018), market trends (Wang, 2014). The studies of Pavlova (2008), Aksibekov (2013), Kokotina (2015), Guzhova & Tokarev (2014) provided in-depth research on various housing markets factors. Economic approaches, however, have failed to address complex statistic research based on comparative regional analysis. There is a lack of correlation regression models which could identify key factors of housing market.

The importance of current research is determined by close attention of the state to the availability and affordability of housing. Since 2006 the national priority project "Available and comfortable housing to the citizens of Russia" has been run. Using market mechanisms the government prioritizes the task of making a purchase of housing available to a considerable amount of citizens including veterans and physically challenged people.

2. Problem Statement

The specific situation in regional housing markets is determined by a considerable differentiation within those regions, namely in their development level, geographical position, and historical development. Regional characteristics of housing in Russia were studied by Kotelnikova (2015), Zhuk & Lozhko (2011) and Chistik (2014). The development gap has to be taken into consideration while making a housing policy. That and the rationale for uneven regional development based on calculations of the housing distribution inequality indices are the issues addressed by Kukhtin & Solovjova (2014).

Priority-driven housing market trends are established in the state program "Providing available and comfortable housing and utilities for the RF citizens". The program contains target indices and development indices without regional specifications. Originally the program was adopted in 2006 but later its target indices were revised. That revision was determined by changing macroeconomic conditions. The indicators of the program implementation also influenced the revision. The analysis showed that in comparison with 2014 the predicted indices were reduced. For example, in 2016-2017 housing supply indices decreased by 11,3% and availability coefficient fell by 23,6%.

Providing homogeneous regional development is one of the most important tasks of the state. That is true for housing market. Selective approach to the development of housing market and due regard to regional specifics in territorial entities of the RF determined clustering approach adopted in this study. Identifying clusters with similar quantitative and qualitative characteristics could allow us to spot the problem regions and generate a framework for action.

3. Research Questions

This study attempts to fulfill the following tasks:

- What are the decisive factors that determine the state housing policy?
- What is the degree of regional differentiation in housing market indices?
- What are the ways to cluster regions according to the level of housing markets development?
- What are the most influential factors for housing market and its sectors?

4. Purpose of the Study

Thus, the aim of this research is to suggest a theoretically sound and statistically relevant background for developing a complex state housing policy that takes into account regional specifics. To reach that target the regions were compared using the average value analysis and variation coefficient. The vivid nonuniformity of the RF regional housing market was grounded. Seven indicators were proposed that reflect different aspects of regional housing market. The most influential factors for the housing market were identified using models.

5. Research Methods

Systematic approach was applied to show the housing market as the complex whole that comprises various elements. The methodology of the study represents a set of methods, namely: Aristotelian method (deduction, induction, reasoning and argumentation); abstract-logical and empirical (observation and experiment). To quantitevely describe the territorial distribution of housing market descriptive statistics was applied (average value indices, variation coefficient, median, and mode).

5.1. Cluster Analysis

To group the regions multidimensionally the cluster analysis was undertaken. It was based on the system of researched indices. While clustering two methods were used:

- Word's hierarchical classification to determine the optimum number of clusters to form the aggregate that is to be divided and
- k-average method to analyze the findings.

5.2. Correlation-regression analysis

Various relatively independent forces influence economy. To study complex socio-economic processes, which housing market belong to, those forces should be taken into account. Therefore correlation-regression methods and regression models generation is applied as a statistical tool to study regions. Many researched has recognized multi steps regression analysis as the most appropriate for assessing the role of interacting factors.

6. Findings

The research studies 6 regional indices from FSSS (Federal State Statistics Service) data: "average primary housing market prices"; "average secondary housing market prices"; "total area of living

premises per capita"; "total area of living premises supply per 1000 residents"; "proportion of slum dwelling"; "average mortgage interest rate".

For consistency in comparison every region was given an average value or a relative value. The first step to analyze the territorial differentiation of housing market indices was variation series design. It allowed us to determine the characteristics and regularities of the aggregate under study. At this stage of the research the pattern and measure of variation values of the criterion were studied. Special approach was applied to calculate variation and distribution pattern.

Table 01. Average values and variation values in the RF regions on housing market indices in 2016

Values	Arithmetical average	Variation coefficient, %	Median
Total area of living premises, m^2/person	25,0	14,5	25,5
Proportion of slum dwelling, %	3,4	92,2	2,4
Total area of living premises supply m^2/ 1000 residents	505,1	46,0	478,5
Average primary housing market prices, rub.	47192,0	36,0	42537,0
Average secondary housing market prices, rub.	49213,3	39,2	42987,0
Average mortgage interest rate, %	12,6	1,9	12,6

Note: Source: Author's calculations based on FSSS data (http://www.gks.ru/)

Table 01 demonstrates the heterogeneity in living premises prices and housing supply. Variation coefficient values also represent the lack of homogeneity in proportion of slum dwelling. Though, living premises supply shows considerable uniformity (14,5 %). There is an evident similarity between median and arithmetical average almost in every index.

On living premises supply index the regions are quite similar to each other. The extremes are 2,4 times different from one another (Moscow region 33,7m/person – Tuva Republic 13,8 m2/person). The difference in housing prices is considerably higher:

- primary housing – 5,3 times (Moscow 155,0 thousand rubles/ m2 – Kalmykia Republic 29,5 thousand rubles/ m2);
- secondary housing – 5,6 times (Moscow 180,9 thousand rubles/ m2 – Orlov region 32,4 thousand rubles/ m2).

Finally, housing supply index is 34,2 times higher in Kaliningrad region (1232 m2) than in Magadan region (36 m2).

Variation series demonstrate quantitative clustering in housing market within the territory of the RF. However, that is only a general picture of index distribution and do not determine its qualitative patterns.

Housing market is a versatile category that could be measured with several criteria. Multidimensional grouping methods of statistics, where clustering analysis belong to.

One of the main tasks of clustering analysis is to generate new classifications in multidimensional space, when there is a need to identify correlations within the aggregate and to regulate its structure. Calculation results determine the number of clusters.

Unfortunately, FSSS lacks the data on Murmansk region, Chukotka Autonomous District, Ingushetia Republic and others. Hence, the data on only 74 territories of the RF were analyzed.

There was a grouping on 7 criteria demonstrating different aspects of regional housing markets:

- Total area of living premises per capita, m^2/person;

- Living premises supply (per 1000 residents);

- Price indices in the primary housing market, %;

- Price indices in the secondary housing market, %;

- Average primary housing market prices (rub./m^2);

- Average secondary housing market prices (rub./m^2);

- Price indices from construction manufacturers (installation and construction work), %.

The aggregate was divided into 5 clusters (Table 02), though Moscow is the only region in the first cluster. This could be explained with its special position in housing market. The prices for a m^2 is by an order of magnitude greater than in other regions of the RF, though housing supply is insufficient in Moscow.

The fourth cluster is represented mainly by the regions with adverse climatic conditions and high prices for living premises. Saint Petersburg stands apart due to its proportion of slum dwelling. That is why the city was singled out into a sub-cluster.

Several regions were organized into clusters according to their low housing price. This situation was observed in in cluster 3 regions. To a greater extend this is true for 3.2 sub-cluster, where there are the regions from the first ten on the lowest housing prices and with an insignificant proportion of slum dwelling, excluding Dagestan Republic.

The second and the fifth clusters include many territorial entities of the RF. Intergroup variations are not very evident in those regions and the indicators here are average. But still there are some variations. Compared to the fifth cluster he regions in the second cluster have lower proportion of slum dwelling, lower prices for primary and secondary housing and less living premises are built. In subcluster 5.2 the prices for primary housing outnumber those for secondary housing and mortgage interest rates are higher.

Table 02. Clustering results of the RF regions according to housing market indices

Clusters	Subclusters	Number of regions	Composition
2	-	20	Samara region, Republic of Tuva, Vologda region, Krasnodar region, Chechen Republic, Kabardino-Balkar Republic, Tula region, Novgorod region, Trans-Baikal region, Tomsk region, Kaliningrad region, Vladimir region, Republic of Khakassia, Ryazan region, Orenburg region, Lipetsk region, Kirov region, Republic of Buryatia, Altai region, Voronezh region
3	3.1	23	Kurgan region, Kemerovo region, Omsk region, Volgograd region, Chuvash Republic, Republic of Mari

			El, Udmurt Republic, Tambov region, Penza region, Republic of Mordovia, Saratov region, Smolensk region, Chelyabinsk region, Orel region, Astrakhan region, Ulyanovsk region, Altai Republic, Kursk region, Kostroma region, Republic of North Ossetia-Alania, Ivanovo region, Jewish Autonomous region, Pskov region
	3.2	6	Republic of Kalmykia, Karachay-Cherkess Republic, Republic of Adygea, Republic of Dagestan, Stavropol territory, Bryansk region
4	4.1	1	Saint-Petersburg
	4.2	5	Primorsky territory, Moscow region, Sakhalin region, Khabarovsk territory
5	5.1	17	Republic of Komi, Sverdlovsk region, Arkhangelsk region, Republic of Sakha (Yakutia), Leningrad region, Kaluga region, Nizhny Novgorod region, Republic of Tatarstan, Kamchatka region, Novosibirsk region, Tver region, Republic of Bashkortostan, Tyumen region, Rostov region, Magadan region, Republic of Karelia
	5.2	7	Amur region, Belgorod region, Krasnoyarsk region, Yaroslavl region, Perm region, Irkutsk region

Sources: Author's calculations based on FSSS data (http://www.gks.ru/)

Correlations among complex socio economic processes including housing market cannot and are not one-dimensional. "Housing market in Russia" is a relatively new concept thus, the thorough study of it and of the factors that could influence it, is of special importance (Sedova, 2014).

To study those factors modeling with multiple regression linear equations was applied. Those equations represent analytical dependences of modeled indices on different factors.

Several regression models were constructed. They represent dependence of housing market indices on socio-economic factors. Data published officially by FSSS were takes as a categorical variable.

List of resultant indices:

Y_1 – «total area of living premises per capita» (m^2/person);

Y_2 – «proportion of slum dwelling» (%);

Y_3 – «total area of living premises supply per 1000 citizens» (m^2/1000 people);

Y_4 – «average primary housing market prices» (rub./m^2);

Y_5 – «average secondary housing market prices» (rub./m^2);

Y_6 – «average mortgage interest rate» (%).

Then regression model of housing (Y_1) is:

$$\hat{Y}_1 = 29{,}170 - 0{,}622X_1 + 0{,}012X_2 \cdot \tag{1}$$

Where, X_1 – birth coefficient, ‰; X_2 – car supply (the number of own automobiles), units/1000 people

Let's interpret the model's parameters.

The increase of 1 point in birth rate leads to the decrease in housing supply by 0,622 m^2. This is arithmetically logical as the appearance of a new family member is followed by the division of the metric area by a bigger number of people.

The growth of car supply by 1 unit (per 1000 people) is followed by the growth in housing supply by 0,012 m^2. The car ownership is an important indicator of adequate living standard and income. Living premises, which are essential goods, belong to costly products.

Multiple determination coefficient equals 0,549, i.e. housing supply variation is determined by variation of model components by 54,9 %.

The following linear model describes the dependence of slum dwelling proportion (Y$_2$) on definite factors:

$$\hat{Y}_2 = -11{,}346 + 0{,}049X_1 + 0{,}528X_2 + 0{,}079X_3. \tag{2}$$

Where, X$_1$ – proportion of mining operations in gross value added (GVA), %; X$_2$ – birth coefficient, ‰; X$_3$ – hospital beds (per 10000 residents).

The extraction specialization of the region influences the housing quality negatively. Mineral reserves are mostly concentrated on the territories with adverse weather conditions. This fact influences the number of residents and causes technical problems in construction. Moreover, rotation workers, who work temporary outside their permanent residence, are a great number of these regions' population. During their employment periods they live in rotation camps, which are equipped according to legally binding minimal requirements to comfort. The growth of mining production in GVA by 1 percentage point (p.p.) is followed by the increase in proportion of slum dwelling by 0,049 p.p. The most vivid examples are Republic of Yakutia (proportion of mining operations 39,5 %, proportion of slum dwelling 16,5 %) and Yamalo-Nenets Autonomous district (61,4 % и 11,8 %, respectively).

Low quality of housing does not deter birth rate. The regions with the highest birth rate (Tyva, Ingushetia, and Dagestan) have more than 10% of dilapidated housing.

The growth in hospital beds supply by 1 unit (per 10000 residents) is followed by 0,079p.p. increase in dilapidated housing proportion. Here the question arises about dualistic character of regional budget spending: usually it goes either in healthcare or in housing. Both of them represent social policy issues. Limited budgets of many RF regions prevent complex decision of social issues.

The factors combined in models explain nearly a half (47,1 %) of regional variation in slum dwelling proportion.

The dependence of housing supply (Y$_3$) from the factors is approximated by the multiple regression model:

$$\hat{Y}_3 = 575{,}577 + 1{,}361X_1 + 0{,}001X_2 - 5{,}379X_3. \tag{3}$$

Where, X$_1$ - migration gain coefficient, ‰; X$_2$ - fixed capital formation per capita, thousand rub./person.; X$_3$ – proportion of unprofitable enterprises, %.

Correlations are quite logical. The rise in migration balance by 1 point per mill is followed by the growth in housing supply by 1,361 м2/1000 people. Positive migration balance increases housing demand (with no reference to housing quality). That, in turn, influences the development pace of construction industry, which is capital intensive and demands solid investments from developers. Hence, investments are seen as the major financial source for construction industry operation, and there is a direct correlation

between investment amount and housing supply. The regression coefficient proves that: every extra square meter of living premises supply demands on average the increase in average per capita investments by 1 million rubles.

Moreover, the substantial amount of overall production is possible under financially sufficient economy. To a certain extent a "proportion of unprofitable enterprises" can be seen as an indicator of that kind of economy. The growth of this indicator by 1 p.p. the amount of housing supply decreases by 5,379 m^2 (per 1000 residents).

All such factors determine variation of the proportion of housing supply by 43,1 % on average within regions.

The forth and the fifth models have the following set of factor variables according to average prices in primary (Y_4) and secondary(Y_5) markets:

$$\hat{Y}_4 = 32480,1 - 565,6X_1 + 7,808X_2 + 0,731X_3 \cdot \qquad (4)$$

$$\hat{Y}_5 = 25348,6 + 8,846X_2 + 1,390X_4 \cdot \qquad (5)$$

Where, X_1 – proportion of agriculture, hunting and forestry in GVA, %; X_2- density of population, person/m^2; X_3 – average income per capita rub/person.; X_4 – consumer expenses per capita, rub./person.

These models are connected only through one indicator, namely "density of population", nevertheless it has a key role in variation of dependence variable. Again, that seems logical as there are different principles of price formation in primary and secondary markets. But it should be noted that FSSS lacks data on some very important indicators, such as number of storeys in a building, structural element of a building (balconies, etc.), and age of building. They are of importance for studying the prices for definite objects, but they have low relevance in analyzing mass phenomenon such as regional housing market. A more detailed analysis based on organized statistical observation is needed.

Below the findings from the models' parameters are represented.

If the proportion of agriculture in GVA rises by 1 p.p., an average primary housing price decrease by 565,6 rub. Agriculture deploys insufficient financial resources. Agricultural regions are comprised of southern territories of the RF European part. Population income here is relatively low. Thus, the housing prices here are not very high either. The highest prices were observed in the northern regions (Yamalo-Nenets Autonomous district, etc.), where climate prevents farming. The proportion of agriculture in GVA is typically not higher than 0,5 %.

The growth of income by 1000 rub. is followed by the increase in average primary housing prices by 731 rubles per a square meter. This factor is of no importance in the secondary market. If people have spare money, they are ready to invest in living premises. On the other hand, consumer expenses indicator implicitly represents the level of wealth in society and is very important for secondary market. If consumer expenses per capita grow by 1000 rub., the prices increase by 1390 rub./m^2 on average. It should be noted that the source of expenses could be represented by both current expenses and savings. The higher the level of well-being the higher is the probability that people will spend their money not only for food and leisure but for capital goods (e.g. housing).

The indicator of "population density" belongs to both models. Some researchers analyzed its influence as geographical position attribution (Antoniucci & Marella, 2017; Sedova, 2014). In this study population density is seen as natural-geographical factor directly correlated with living premises prices.

The higher the degree of density the more acute is a housing problem and the prices are higher. The increase of density by 1 person/km^2 is followed by changes in primary market prices by 7,808 rub., and by 8,846 rub. in the secondary market.

The models are quite relevant. The factors, they contain, explain more than 80% of average prices variations for living premises.

The model on the indicator Y_6 «average mortgage interest rate» (%) was not generated. This could be explained by several reasons. нами построена не была по нескольким причинам. A relatively high level of significance is needed for input data to follow the statistical law. However, in this case a generated model would have a low determination coefficient, hence, interpretation lose any sense.

7. Conclusion

To provide a uniform development of regions is one of the major tasks for government. This is true for housing. To develop trends for living property market the government should follow a selective approach. Regional specifics and differences n territorial development should be considered.

While developing a housing policy the following components should be considered:

7.1. Cluster approach

Seven indices were analyzed to define 5 clusters of the RF regions. There were total area of living premises per capita, m^2/person; living premises supply (per 1000 residents); price indices in the primary housing market, %;

price indices in the secondary housing market, %; average primary housing market prices (rub./m^2); average secondary housing market prices (rub./m^2); and price indices from construction manufacturers (installation and construction work), %. They have quantitative and qualitative parameters., that could allow the government to create and differentiate.

7.2. Factors affecting the real estate market

The factors influencing the residential real estate market were investigated on the basis of modeling in the form of linear equations of multiple regression. Private coefficients of determination were used to assess "pure" contribution of a particular factor. Overall, the results of regression analysis demonstrated that many factors affecting the state of the housing market are stable enough.

Demographical indices are the most critical for territorial differentiation of housing market. They are seen in three of five models. If to take into account that "population density" to a great extent is a demographic one, then every model would have a demographic index.

The next critical factors are natural-geographical ones. Their influence can be traced in two models. Interestingly the group was represented only by a population density index. Its influence was quite substantial, e.g. its share into secondary market average price variation was 53,4 %. The living standard index was found in three models. However, they could not explain even 30 % resultative indices variations.

The least influential factors for housing markets are production-infrastructure and social ones. They were seen only in one model. It could be explained by their indirect influence, e.g. income level

correlates with financial position of a business and industry specifics in the region. Eventually housing market is oriented on residents, i.e. region's population, and the situation on housing market depends firstly on the number, composition, and well-being of the population, and secondly on economic situation.

The results of the study are of importance federal and regional authorities. The government bodies could regulate regional housing markets more efficiently through targeted influence on the most important factors in real life context.

References

Aksibekov, A. (2013). Factors influencing real estate market. *Herald of KazNTU , 3*, 397-401

Antoniucci, V., & Marella, G. (2017). Is social polarization related to urban density? Evidence from the Italianhousing market. *Landscape and Urban Planning, 177*, 340-349. https://dx.doi.org/10.1016/j.landurbplan.2017.08.012.

Cheung, R., Wetherell, D., & Whitaker, S. (2018). Induced earthquakes andhousing markets: Evidence from Oklahoma. *Regional Science and Urban Economics, 69*, 153-166. https://dx.doi.org/10.1016/j.regsciurbeco.2018.01.004.

Chistik, O. (2014). Statistical analysis of dilapidated housing in housing market of Russia. *Herald of SSEU, 1 (111),* 74-78

Gan, L., Wang, P., & Zhang, Q. (2018). Market thickness and the impact of unemployment on housing market outcomes. *Journal of Monetary Economics, 98*, 27-49. https://dx.doi.org/10.1016/j.jmoneco.2018.04.007

Ge, T., & Wu, T. (2017). Urbanization, inequality and property prices: Equilibrium pricing and transaction in the Chinese housing market. *China Economic Review*, *45*, 310-328. https://dx.doi.org/10.1016/j.chieco.2016.04.005

Guzhova, O., & Tokarev, Yu. (2014). Territorial differentiation of housing markets indices in the RF. *Herald of Samara State Economic University, 8* (118), 116-121

Killins, R., Egly, P., & Escobari, D. (2017). The impact of oil shocks on the housing market: Evidence from Canada and U.S. *Journal of Economics and Business, 93,* 15-28. https://dx.doi.org/10.1016/j.jeconbus.2017.07.002]

Kokotina, T. (2015). Statistical analysis of human potential influence on housing market of Volga federal district. *Herald of Mariisk State University. Agriculture. Economics, 2,* 74-78

Kotelnikova, A. (2015). Analysis of regional differentiation of the RF regions according availability of housing. *Herald of SSEU, 7* (105), 42-48

Kukhtin, P., & Solovjova, M. (2014). Housing market differentiation. *Internet journal «ScienceS studies», 6*(25), *1-19.* https://dx.doi.org/10.15862/106EVN614

Pavlova, M. (2008). Factor influencing the housing market. *Herald of SSEU, 11* (49), 83-89

Sedova, E. (2014). Modeling the price of secondary housing market on regional markets: dimensional approach. *OSU Herald, 14* (175), 458-464

Ume, E. (2018). The impact of monetary policy on housing market activity: An assessment using sign restrictions. *Economic Modelling*, *68*, 23-31. https://dx.doi.org/10.1016/j.econmod.2017.04.013]

Wang, Z. (2014). Market sentiment in private housing market. *Habitat International, 44*, 375-385. https://dx.doi.org/10.1016/j.habitatint.2014.08.001

Zhuk, V., & Lozhko V. (2011). Housing issue in solving social-economic problems of Russian regions. *Issues of modern economy, 1,* 243-246