

---

## How to Calculate the Share of a Land Plot in the Value of Office and Retail Real Estate: Regression Analysis to Increase the Cadastral Valuation Reliability

---

Submitted 06/06/20, 1st revision 20/07/20, 2nd revision 18/08/20, accepted 15/09/20

Vladislav Sutyagin<sup>1</sup>, Yana Radyukova<sup>2</sup>, Irina Zinovieva<sup>3</sup>, Svetlana Zenchenko<sup>4</sup>, Valentina Smagina<sup>5</sup>, Nikolay Pakhomov<sup>6</sup>

**Abstract:**

**Purpose:** In 2014 the Tax Code was updated. Currently, it also deals with the cadastral valuation of capital construction projects. However, there are problems associated with the cadastral valuation reliability. One of the key problems is how to evaluate a capital construction object and exclude the cost of the land plot.

**Design / Methodology / Approach:** This study proposes an alternative approach to calculating the value of a land plot by analyzing the share of a land plot in the value of a single property. In the article, the authors consider that a land plot is included in the value of a single property due to a number of various pricing factors. The method of regression analysis was used as a methodological basis.

**Findings:** Based on the results of the regression analysis, the authors gave recommendations on how to use the proposed factor models and made certain conclusions.

**Practical Implications:** The results of this study can be used in cadastral valuation and in the process of challenging the cadastral value.

**Originality / Value:** The presented study is an alternative way of calculating the value of land in a single property. It is useful since there is little information on how to apply the traditional methods of comparative sales.

**Keywords:** Cadastral value, reliability of cadastral valuation, share of a land plot in a single real estate object.

**JEL:** D46, H21.

**Paper type:** Research article.

---

<sup>1</sup>Corresponding author, Tambov State University named after G. R. Derzhavin, Tambov, Russian Federation, [sutyagin.vladislav@yandex.ru](mailto:sutyagin.vladislav@yandex.ru);

<sup>2</sup>Same as in 1, [radyukova68@mail.ru](mailto:radyukova68@mail.ru);

<sup>3</sup>Voronezh State University of Forestry and Technologies named after G.F. Morozov, Voronezh, Russian Federation, [zinovirs@mail.ru](mailto:zinovirs@mail.ru);

<sup>4</sup>North Caucasus Federal University, Stavropol, Russian Federation, [zen\\_sveta@mail.ru](mailto:zen_sveta@mail.ru);

<sup>5</sup>Tambov State University named after G.R. Derzhavin, Tambov, Russian Federation, [troveo@mail.ru](mailto:troveo@mail.ru);

<sup>6</sup>Same as in 5, [dissovet@tsu.tmb.ru](mailto:dissovet@tsu.tmb.ru);

## **1. Introduction**

In August 2019, it was 20 years since the institution of Russian cadastral valuation was established. For quite a long time, this institution regulated only land legal relations. However, in accordance with the federal law of 04.10.2014 N 284-FZ "On Amendments to Articles 12 and 85 of Part One and Part Two of the Tax Code of the Russian Federation and the dead law of the Russian Federation "On Taxes on Property of Individuals" the institution of cadastral valuation started to perform new functions. Currently, the objects of capital construction are also considered as objects of taxation. Due to the above-mentioned fact, there are a number of problems associated with cadastral valuation reliability (as well as the problems associated with the estimates on the basis of which the cadastral value is disputed).

One of the serious problems associated with cadastral valuation (and market valuation in the process of challenging) is related to land value. The fact is that the cadastral (as well as the market) value is based on market data. Moreover, it widely uses the methods of comparative analysis, which implies that the value of a particular property depends on the prices of similar objects. However, such objects are sold on the market under the conditions of civil and land law, which implies the alienation of a building (or part of it) and the rights to the land on which it is located.

However, when it comes to taxation, the land plot and the object of capital construction are artificially separated and become objects of taxation. Thus, there is a need to deduct the cost of land from the cost of single real estate objects sold on the market. It might seem that this is a methodological question which does not require attention: the appraiser determines the value of the land and deducts its cost from the cost of a single property. However, there are a number of nuances that should be taken into consideration.

First of all, in the process of assessment of the cadastral value, a large volume of data needs to be taken into account, which makes it hard to apply this approach. But even if the purpose of the assessment is to challenge the cadastral value (individual assessments of a specific property), it is still problematic to use this approach.

Secondly, if there are several buildings located on the same plot of land, it is a challenging task to divide this plot of land. The situation gets even more complicated if there are specific structures (including underground structures), located on the same plot of land.

Thirdly, in some cases buildings are located on unregistered plots (for example, many residential buildings constructed in the time of the USSR).

Fourthly, sometimes it is unclear which share of plot is included in the cost of the building. The authors of this article express their own vision and refer to their previous

publications (Sutyagin, 2015). However, if the cost of land is not deducted from the cost of the premises, it makes the taxable base and the tax higher.

Summing up, we can conclude that the method of comparative analysis implying the deduction from the value of a single property cannot be widely applied.

An alternative and better way for calculating the value of land in a single property is based on an assessment of the share of a land plot in the value of a single property (the so-called *distribution method*, widely used in the assessment of built-up areas). According to this method, the market value of a land plot is calculated as:

$$MV_l = D_l * MV_{eol} \quad (1)$$

$MV_l$  stands for the market value of the land plot;  $D_l$  stands for the share of the value of the land plot in a single real estate object ("share of the land plot");  $MV_{eol}$  stands for the market value of a single property.

## 2. Research Methods

As for the general theoretical basis of the study, the researchers used a logical method, as well as the methods of analysis and induction. In addition to the general scientific methods, the authors also used statistical methods, such as sample observation, grouping (clustering), structural analysis, and correlation-regression analysis.

Using the method of *sample observation*, the researchers analyzed the evaluation of real estate objects carried out by Tambov appraisal firms within the period from 2012 to 2019. At the preliminary stage 198 objects were selected. In the course of selective observation, the authors used the Smirnov-Grubbs test to exclude atypical objects. The final sample consisted of 183 objects. At the stage of *grouping*, the objects underwent clustering, and three groups were formed: "Industrial and warehouse real estate", "Office and retail real estate", "Residential real estate".

In this article the authors analyze the results related to the cluster "Office and retail real estate". The sample results are presented on the official website «*Otsenka+*» (OTSENKA, 2020). Carrying out the *structural analysis*, the authors studied the value of specific elements (land plot and capital construction object) of each single real estate object included in the sample. Then the relative indicator (the share of the land plot) was calculated. Within the framework of the *correlation and regression analysis*, the authors determined the factors influencing the cost of capital structures and land plots. The researchers analyzed and generalized the main pricing factors. As for the cluster "Office and retail real estate", the authors took into account and analyzed the following factors: "Floor/number of floors"; "Technical condition"; "Interior"; "Location"; "Separate entrance"; "Transport accessibility"; "Construction material". Since the identified factors are mainly qualitative, the authors converted the data into a digital format.

The authors built factor models showing the dependence of the share of a land plot included in the value of a single property (function) on the values of the main pricing factors (function arguments). Within the framework of this study, the researchers compiled the functions on the basis of 3, 4, 5, 6 and 7 factors. The alignment was based on the linear function. Further, the functions were analyzed (the authors used the indicators of the correlation coefficients, determination, average approximation error and Fisher's F-test). Finally, the researchers gave specific recommendations on how to use these functions.

As for the informational basis of the study, the authors analyzed the archival base of evaluation reports (from 2012 to 2019) made by appraisal firms and experts in Tambov region (*LLC Otsenka +, LLC Novator, LLC Expert Bureau No. 1, LLC "Audit-Invest", LLC "Console", LLC "Expert-Otsenka*).

Carrying out the analysis, the researchers used the empirical data for the period 2012 to 2019. A relative indicator was applied (the ratio of the value of a land plot to the value of a single property), in which the influence of the inflationary factor is excluded.

### **3. Research Results**

Many Russian researchers deal with the problem of assessment of the share of a land plot (Sutyagin, 2015; Demidova and Gorelikov, 2015; Gladkikh and Kuznetsova, 2015; Leifer *et al.*, 2003., Leifer and Kashnikova, 2006; Morozova, 2015, Topal and Baybarina, 2014, Sutyagin and Cherkashnev, 2012). As a rule, researchers make an attempt to assess the average value of the share of a land plot included in the value of a single object. They usually conduct their studies analyzing the region's database.

Yet, Russian most researchers have not carried out detailed studies, since it is impossible to collect and generalize initial data without the help from state bodies (for example, Russian State Register). However, it would significantly affect the quality and objectivity of the cadastral estimates.

It should be noted that the studies differ from each other significantly both in terms of the scope of research and the degree of detail. Even if we analyze a cluster, the share of a land plot may differ considerably due to the regional specifics of pricing. At the same time, the average value of the share of a land plot does not always correlate to the actual conditions for assessing a particular property. And due to the confidence interval, it is problematic to calculate a specific value (Leifer, 2017; Sutyagin, 2017).

Within the framework of this study, an alternative approach to calculating the share of a land plot is proposed. The authors suggest using a regression model, which implies the construction of a mathematical dependence of the share of a land plot on specific factors. For practical purposes, it is advisable to use a method of multiple regressions, since the cost often depends on several factors.

In this case, the regression analysis *algorithm* includes the following stages:

1. formulation of the initial economic hypotheses;
2. selection and analysis of the initial sample. Gross error diagnostics;
3. a specific function is selected to carry out the alignment;
4. independent variables (cost factors) are identified and if they are non-quantitative, they are digitized;
5. the regression equation is derived and the indicators characterizing its quality are analyzed;
6. numeric labels are assigned to cost factors and optimized (if necessary);
7. the analysis of the economic sense of the regression and indicators characterizing the quality of the regression is carried out.

*Formulation of the initial economic hypotheses.* In contrast to purely mathematical and statistical problems, the use of correlation and regression analysis is based on the formulation of initial economic hypotheses, including:

1. A hypothesis about the nature of the relationship between the cost factors and the share of the land plot;
2. Hypothesis about the number of independent variables in regression;
3. Hypothesis about the economic nature of the digitization of non-quantitative regression factors.

According to the first hypothesis, the selected function used in alignment, should correlate to the factors determining the share of the land plot. In this case, the economic factor should dominate over the mathematical factor. Secondly, linear alignments are used to describe the economic processes. Thus, to get the final result, it is necessary just to sum up the economic factors. This means that the choice of a linear function is preferable. In this regard, the authors used a linear function.

Selecting a function, researchers should take into account the third important aspect - the intercept of the regression. From the economic point of view, the intercept determines the share of the land plot with zero arguments. In the majority of cases the value of the intercept is negative, which has no economic sense.

From the statistical point of view, statistical indicators of the quality of the regression are taken into account (for example, the correlation coefficient and / or determination). It should be noted that the correlation and determination coefficients are much higher in the model without an intercept. Thus, in this article the authors suggest building a function without an intercept.

According to the hypothesis about the number of independent variables in the regression, their number should correspond to the number of independent economic factors. In practice, it is advisable to take into account 3-7 factors (Anisimova *et al.*, 2004). If fewer variables are used, a significant factor might be overlooked.

On the other hand, it is not advisable to use more variables due to the mutual conjugation (multicollinearity) of variables. Secondly, if more factors are taken into account, a new regression should be built since the sample has to be enlarged (the ratio of the number of objects in the sample to the number of variables should be at least 5-8) (Anisimova *et al.*, 2004).

Finally, it is necessary to focus on the economic nature of digitization of non-quantitative regression factors. All factors (with the exception of "floor/number of floors") are qualitative and need to be digitized, which requires characteristics ranking. *Selection and analysis of the initial sample. Gross error diagnostics.* The authors selected 198 objects for the period from 2012 to 2019. It is an acceptable interval, since the influence of inflationary processes (in the analyzed relative indicator) is eliminated.

Selecting the initial sample, the researchers carried out an analysis to detect gross errors ("statistical outliers") in it. The "standard deviation rule", Grubbs' test for outliers, the Smirnov-Grubbs test, and T. Moore criteria can be used as knockout criteria. This study was based on the Smirnov-Grubbs test (t).

The preliminary sample is ranked, the mean value ( $\bar{x}$ ) and standard deviation ( $\sigma_c$ ) are calculated. And then the maximum ( $x_{max}$ ) and minimum ( $x_{min}$ ) values of the sample are analyzed to detect a gross error. Thus,  $t_v$  is calculated:

$$t = \frac{|x_i - \bar{x}|}{\sigma} \tag{2}$$

The calculated value is compared with the tabular  $t_\alpha$ . If  $t_v > t_\alpha$ , the result  $x_i$  is considered a gross error and discarded. Thus, 183 objects remained (the full list of objects is posted on the "Otsenka +" portal (OTSENKA, 2020)). The results of the analysis are shown in Table 1.

**Table 1.** *Analysis of the sample and gross error detection*

N	Indicator	Quality score	Critical value
1	$\bar{x}$	0,95	
2	$\sigma_c$	0,069	
3	$n$	183	
4	$x_{max}$	0,342	
5	$t_{max}$	3,164	3.290
6	$x_{min}$	0,005	
7	$t_{min}$	1,677	3.290

**Source:** *Own study.*

*Selection of the type of function to carry out the alignment.* The alignment was carried out on the basis of a linear function:

$$f(x) = b + a_1 * x_1 + a_2 * x_2 + \dots + a_n * x_n + \varepsilon \quad (3)$$

$f(x)$  stands for the resulting feature;  $x_1, x_2, x_n$  - independent variables;  $a_1, a_2, a_n$  - coefficients for independent variables (the degree of sensitivity of the resulting feature to the independent variable);  $b$  - the intercept in the regression;  $\varepsilon$  - the influence of unaccounted factors.

The adequacy of the resulting regression equation is assessed (statistical significance) (for example, the average approximation error is assessed and Fisher's F-test is applied). The regression equation can be used for practical purposes.

*Identifying the independent variables - cost factors.* To select the independent variables, the authors analyzed the typical pricing factors that are taken into account when the value of various types of real estate is assessed. The following factor arguments were analyzed: "Floor / number of floors"; "Technical condition"; "Interior"; "Location"; "Separate entrance"; "Transport accessibility"; "Construction material".

The authors described the economic sense of the factors (Table 2). It should be noted that conditionally all factors can be divided into factors of "land" (determining the value of land) and factors of "capital construction objects". The former increase the value and share of the land plot; while the latter (for example, good interior or technical condition)) increase the value of the capital construction object.

In fact, none of the above-mentioned factors are quantitative and they need to be assigned numeric labels.

It should also be noted that the factors correspond to the economic hypothesis and build a logical hierarchy. In the table below, the share of the land plot increases from top to bottom. Almost all the factors are qualitative and are assigned numeric labels (1, 2, 3) (Table 2):

**Table 3.** *Numeric labels assigned to cost factors*

Cost factor	Cost factor characteristics
Floor/number of floors	Accepted values. The ratio of the number of floors in the assessment object to the total number of floors in the building is calculated Accepted values: from 0 to 1.
Technical condition	Accepted values and numeric labels: • Good – 1; Satisfactory - 2; Unsatisfactory - 3.
Interior	The factor: qualitative (characteristics of the capital construction object) Accepted values: • Comfort - 1; Economy class -2 ; Renovation is needed - 3.
Location	The factor: qualitative (characteristics of the plot of land) Accepted values:

	<ul style="list-style-type: none"> <li>Rural settlements -1; Other cities in the region - 2; Center of the subject of the Russian Federation - 3.</li> </ul>
Separate entrance	The factor:: qualitative (characteristics of the capital construction object) Accepted values: <ul style="list-style-type: none"> <li>Yes - 1; No - 2.</li> </ul>
Transport accessibility	The factor: qualitative (characteristics of the plot of land) Accepted values: <ul style="list-style-type: none"> <li>Driveways - 1; Red line (Pedestrian area) - 2.</li> </ul>
Construction material	The factor: qualitative (characteristics of the capital construction object) Accepted values: <ul style="list-style-type: none"> <li>Brick / concrete - 1; Wood - 2; Metal / sandwich panels / prefabricated structures - 3.</li> </ul>

Source: Own study.

*Derivation of a regression equation and analysis of the regression quality indicators.*  
The authors of the research derived 10 regression equations with different numbers of factors (from 7 to 3). It should be noted that the fewer the number of factors, the higher the quality of the regression (in accordance with the above-mentioned economic hypotheses).

**Table 4.** Regression equations and regression quality indicators

Number and list of factors	Regression equation	Regression quality indicators $r$ – multiple correlation coefficient; $r^2$ – coefficient of determination; $\bar{A}$ – average approximation error; $F_V$ – calculated F-Test statistic (Fisher Value); $F_t$ – critical value of F-Test statistic
<b>1. 7 factors:</b> <ul style="list-style-type: none"> <li>Floor/number of floors (x1);</li> <li>Technical condition (x2);</li> <li>Interior (x3);</li> <li>Location (x4);</li> <li>Separate entrance (x5);</li> <li>Transport accessibility (x6);</li> <li>Construction material (x7).</li> </ul>	$y = 0,139x_1 + 0,006x_2 + 0,016x_3 - 0,009x_4 - 0,015x_5 + 0,005x_6 + 0,012x_7$	$r = 0,960; r^2=0,922; \bar{A} = 32,1\%; F_V = 298.87; F_t = 2.11$
<b>2. 6 factors:</b> <ul style="list-style-type: none"> <li>Floor/number of floors (x1);</li> <li>Technical condition (x2);</li> <li>Interior (x3);</li> <li>Separate entrance (x4);</li> <li>Transport accessibility (x5);</li> <li>Construction material (x6).</li> </ul>	$y = 0,136x_1 + 0,003x_2 + 0,016x_3 - 0,026x_4 + 0,002x_5 + 0,010x_6$	$r = 0,960; r^2=0,921; \bar{A} = 58,4\%; F_V = 344.07; F_t = 2.22$



<p><b>3. 5 factors:</b></p> <ul style="list-style-type: none"> <li>• Floor/number of floors (x1);</li> <li>• Technical condition (x2);</li> <li>• Interior (x3);</li> <li>• Transport accessibility (x4);</li> <li>• Construction material (x5).</li> </ul>	$y = 0,140x_1 + 0,002x_2 + 0,010x_3 - 0,007x_4 + 0,007x_5$	$r = 0,959; r^2=0,919; \bar{A} = 29,3\%; F_v = 398.16; F_t = 2.38$
<p><b>4. 4 factors:</b></p> <ul style="list-style-type: none"> <li>• Floor/number of floors (x1);</li> <li>• Technical condition (x2);</li> <li>• Interior (x3);</li> <li>• Construction material (x4).</li> </ul>	$y = 0,138x_1 - 0,0001x_2 + 0,008x_3 + 0,005x_4$	$r = 0,958; r^2=0,917; \bar{A} = 48,0\%; F_v = 494.49; F_t = 2.62$
<p><b>5. 4 factors:</b></p> <ul style="list-style-type: none"> <li>• Floor/number of floors (x1);</li> <li>• Interior (x2);</li> <li>• Transport accessibility (x3);</li> <li>• Construction material (x4).</li> </ul>	$y = 0,141x_1 + 0,011x_2 - 0,007x_3 + 0,007x_4$	$r = 0,958; r^2=0,918; \bar{A} = 51,5\%; F_v = 500.11; F_t = 2.62$
<p><b>6. 4 factors:</b></p> <ul style="list-style-type: none"> <li>• Floor/number of floors (x1);</li> <li>• Technical condition (x2);</li> <li>• Interior (x3);</li> <li>• Transport accessibility (x4).</li> </ul>	$y = 0,144x_1 + 0,001x_2 + 0,011x_3 - 0,005x_4$	$r = 0,958; r^2=0,917; \bar{A} = 27,6\%; F_v = 494.89; F_t = 2.62$
<p><b>7. 3 factors:</b></p> <ul style="list-style-type: none"> <li>• Floor/number of floors (x1);</li> <li>• Technical condition (x2);</li> <li>• Construction material (x3).</li> </ul>	$y = 0,138x_1 + 0,006x_2 + 0,008x_3$	$r = 0,957; r^2=0,915; \bar{A} = 52,4\%; F_v = 659.81; F_t = 3,01$
<p><b>8. 3 factors:</b></p> <ul style="list-style-type: none"> <li>• Floor/number of floors (x1);</li> <li>• Technical condition (x2);</li> <li>• Interior (x3).</li> </ul>	$y = 0,141x_1 + 0,0001x_2 + 0,009x_3$	$r = 0,957; r^2=0,917; \bar{A} = 35,0\%; F_v = 659.81; F_t = 3,01$
<p><b>9. 3 factors:</b></p> <ul style="list-style-type: none"> <li>• Floor/number of floors (x1);</li> <li>• Interior (x2);</li> <li>• Transport accessibility (x3).</li> </ul>	$y = 0,145x_1 + 0,012x_2 - 0,004x_3$	$r = 0,958; r^2=0,917; \bar{A} = 35,9\%; F_v = 663.25; F_t = 3,01$

<p><b>10. 3 factors:</b></p> <ul style="list-style-type: none"> <li>• Technical condition (x1);</li> <li>• Interior (x2);</li> <li>• Transport accessibility (x3).</li> </ul>	$y = 0,035x_1 + 0,013x_2 + 0,025x_3$	$r = 0,875; r^2=0,766; \bar{A} = 82,7\%; F_v = 196,71; F_t = 3,01$
---	--------------------------------------	--

*Source: Own study.*

#### 4. Discussion

According to the table, all regression equations have a high level of statistical significance, which is confirmed by the Fisher's F-test. Moreover, the correlation coefficient shows that the share of the land plot is functionally dependent on cost factors, since the latter is more than 0.95. Further, according to the analysis of the coefficient of determination, the result (the share of the land plot) mostly depends (more than 91%) on these factors (with the exception of the last function). Thereby, the influence of the unaccounted factors is minimal.

At the same time, predict functions have a limit, which results in the high average error of approximation. Only 5-factor models as well as some 4-factor and 3-factor models are reliable. It should be noted that a number of regression equations have negative arguments, which violates the formulated economic hypotheses. The factor grows but the share of the land plot gets lower (which is actually not true). From the point of view of the formulated economic hypotheses, models No. 7, 8, 10 are the most adequate.

The quality of the regression can be improved by means of *optimization procedures* (the space between numeric values needs to be removed). *Sequential recoding* or *optimization procedures* are the best ways to do this (Andreev, 2003; 2004). The quality of the regression depends on the space between numeric values. It is necessary choose the appropriate method to make necessary changes. The following indicators are most often used as optimization criteria:

- Correlation coefficient (in this case, multiple correlation coefficient) and its maximization;
- The coefficient of determination and its maximization;
- Average error of approximation and its minimization.

Applying the method of *sequential recoding*, an expert labels the numeric values. After that the regression analysis procedures are repeated. However, the process of labeling seems rather subjective. To make it more objective, it is advisable to use the following expert rating scale (Table 5).

Since the first two coefficients are quite high, the indicator characterizing the average error of approximation is chosen as the optimization criterion. The results of the optimized functions are shown in Table 6.

**Table 5. Ordinal expert rating scale (Andreev D.M., 2003a)**

Cost factor characteristics	Scale value
There is no factor domination	1
The factor is slightly dominant	2-4
The factor is noticeably dominant	5-6
The factor is extremely dominant	7-8
The factor is absolutely dominant	9

Source: Own study.

**Table 6. Regression equations and regression quality indicators**

Number and list of factors	Numeric labels	Regression equation	Regression quality indicators
<b>7 factors:</b> Floor/number of floors (x1);	0 - 1	$y = 0,143x_1 + 0,003x_2 + 0,009x_3 - 0,010x_4 - 0,002x_5 + 0,003x_6 + 0,012x_7$	$r = 0,961;$ $r^2=0,923;$ $\bar{A} = 30,2\%;$ $F_v = 301,83;$ $F_t = 2,11$
Technical condition (x2);	Good-1; Satisfactory-3; Unsatisfactory-4;		
Interior (x3);	Comfort - 1; Economy Class - 2; Renovation is needed - 5		
Location (x4);	Rural settlements -1; Other cities in the region - 2; Center of the subject of the Russian Federation - 3.		
Separate entrance (x5);	Yes -1;No -7		
Transport accessibility (x6);	Driveways - 1; Red line (Pedestrian area) - 2		
Construction material (x7)	Brick / concrete - 1; Wood - 2; Metal / sandwich panels / prefabricated structures - 3.		
<b>6 factors:</b> Floor/number of floors (x1)	0 - 1	$y = 0,138x_1 + 0,002x_2 + 0,010x_3 - 0,013x_4 - 0,001x_5 + 0,010x_6$	$r = 0,960;$ $r^2=0,922;$ $\bar{A} = 32,7\%;$ $F_v = 348,01;$ $F_t = 2,22$
Technical condition (x2);	Good -1; Satisfactory -3; Unsatisfactory -5;		
Interior (x3);	Comfort - 1; Economy Class - 2; Renovation is needed - 5		
Separate entrance (x4);	Yes -1;No - 3		
Transport accessibility (x5);	Driveways - 1; Red line (Pedestrian area) - 2		
Construction material (x6).	Brick / concrete - 1; Wood - 2; Metal / sandwich panels / prefabricated structures - 3.		
<b>5 factors:</b> Floor/number of floors (x1);	0 - 1	$y = 0,139x_1 + 0,003x_2 + 0,010x_3 - 0,006x_4 + 0,004x_5$	$r = 0,959;$ $r^2=0,919;$ $\bar{A} = 28,3\%;$ $F_v = 404,86;$ $F_t = 2,38$
Technical condition (x2);	Good-1; Satisfactory -2; Unsatisfactory -3;		
Interior (x3);	Comfort - 1; Economy Class - 2; Renovation is needed - 3;		

*How to Calculate the Share of a Land Plot in the Value of Office and Retail Real Estate:  
Regression Analysis to Increase the Cadastral Valuation Reliability*

678

Transport accessibility (x4);	Driveways - 1; Red line (Pedestrian area) - 2		
Construction material (x5).	Brick / concrete - 1; Wood - 3; Metal / sandwich panels / prefabricated structures - 7.		
<b>4 factors:</b> Floor/number of floors (x1);	0 - 1	$y = 0,140x_1 + 0,001x_2 + 0,004x_3 + 0,004x_4$	$r = 0,959;$ $r^2=0,919;$ $\bar{A} = 47,2\%;$ $F_v = 508,28;$ $F_t = 2,62$
Technical condition (x2);	Good -1; Satisfactory -3; Unsatisfactory -4;		
Interior (x3);	Comfort - 1; Economy Class - 3; Renovation is needed - 7;		
Construction material (x4)	Brick / concrete - 1; Wood - 2; Metal / sandwich panels / prefabricated structures - 3.		
<b>4 factors:</b> Floor/number of floors (x1);	0 - 1	$y = 0,138x_1 + 0,003x_2 + 0,0002x_3 + 0,003x_4$	$r = 0,959;$ $r^2=0,919;$ $\bar{A} = 33,2\%;$ $F_v = 511,12;$ $F_t = 2,62$
Interior (x2);	Comfot - 1; Economy Class - 7; Renovation is needed - 9;		
Transport accessibility (x3);	Driveways - 1; Red line (Pedestrian area) - 9		
Construction material (x4)	Brick / concrete - 1; Wood - 7; Metal / sandwich panels / prefabricated structures - 9.		
<b>4 factors:</b> Floor/number of floors (x1);	0 - 1	$y = 0,143x_1 + 0,002x_2 + 0,010x_3 - 0,004x_4$	$r = 0,958;$ $r^2=0,917;$ $\bar{A} = 27,5\%;$ $F_v = 496,66;$ $F_t = 2,62$
Technical condition (x2);	Good -1; Satisfactory -3; Unsatisfactory -5;		
Interior (x3);	Comfort - 1; Economy Class - 2; Renovation is needed - 3;		
Transport accessibility (x4).	Driveways - 1; Red line (Pedestrian area) - 2		
<b>3 factors:</b> Floor/number of floors (x1);	0 - 1	$y = 0,138x_1 + 0,006x_2 + 0,003x_3$	$r = 0,958;$ $r^2=0,918;$ $\bar{A} = 39,5\%;$ $F_v = 668,75;$ $F_t = 3,01$
Technical condition (x2);	Good -1; Satisfactory -3; Unsatisfactory -5;		
Construction material (x3).	Brick / concrete - 1; Wood - 7; Metal / sandwich panels / prefabricated structures - 9.		
<b>3 factors:</b> Floor/number of floors (x1);	0 - 1	$y = 0,141x_1 + 0,0001x_2 + 0,009x_3$	$r = 0,957;$ $r^2=0,917;$ $\bar{A} = 35,0\%;$ $F_v = 659,81;$ $F_t = 3,01$
Technical condition (x2);	Good -1; Satisfactory -2; Unsatisfactory -3;		
Interior (x3)	Comfot - 1; Economy Class - 2; Renovation is needed - 3;		
<b>3 factors:</b>	0 - 1		

Floor/number of floors (x1);		$y = 0,142x_1 + 0,003x_2 + 0,0004x_3$	$r = 0,958;$ $r^2=0,918;$ $\bar{A} = 28,7\%;$ $F_v = 668,89;$ $F_t = 3,01$
Interior (x2);	Comfort - 1; Economy Class - 7; Renovation is needed - 9;		
Transport accessibility (x3).	Driveways - 1; Red line (Pedestrian area) - 9		
<b>3 factors:</b> Technical condition (x1);	Good - 1; Satisfactory - 2; Unsatisfactory - 3	$y = 0,041x_1 + 0,017x_2 + 0,004x_3$	$r = 0,872;$ $r^2=0,761;$ $\bar{A} = 70,2\%;$ $F_v = 191,25;$ $F_t = 3,01$
Interior (x2);	Comfort - 1; Economy Class - 2; Renovation is needed - 3;		
Transport accessibility (x3).	Driveways - 1; Red line (Pedestrian area) - 7		

**Source:** Own study.

Optimization procedures are the way to minimize the average approximation error. In a number of cases (formulas No. 4, 5, 9) the functions confirm the formulated economic hypotheses (models No. 4, 5, 7, 8, 9, 10).

The second method that is based on *optimization procedures* implies that numeric values should be labeled iteratively, which increases the accuracy of the regression model. This approach requires the use of spreadsheets (Sivets and Levykina, 2003; Anisimova, 2003).

However, the analysis showed that this approach was unsuccessful. Although the accuracy of the regression model increased, the economic sense was lost. Moreover, the order of the numeric labels often changed. In this regard, the results of the optimization procedures are not presented in the article. The authors also made an attempt to sum up the main ideas related to the economic sense of the derived regressions and gave some basic recommendations on how to use the models.

Firstly, predict functions have a limit, which results in a high average error of approximation. In this regard, these models should be used only if it is impossible to apply basic approaches to valuation of land.

Secondly, only models № 4, 5, 7, 8, 9, 10 can be used since they confirm the formulated economic hypotheses.

Thirdly, 5-, 6- and 7-factor models can be applied to evaluate office and retail real estate, since a large list of cost factors is taken into account.

Finally, 4- and 3-factor models should be used in accordance with the recommendations given in Table 7.

**Table 7. Recommendations on how to use the models**

Number and list of factors	Regression equation	Recommendations
<b>4 factors:</b> Floor/number of floors (x1); Technical condition (x2); Interior (x3); Construction material (x4)	$y = 0,140x_1 + 0,001x_2 + 0,004x_3 + 0,004x_4$	It is suitable in case transport accessibility is not important (for example, if objects are located inside factory areas or there are a lot of auxiliary areas)
<b>4 factors:</b> Floor/number of floors (x1); Interior (x2); Transport accessibility (x3); Construction material (x4)	$y = 0,138x_1 + 0,003x_2 + 0,0002x_3 + 0,003x_4$	It is suitable for recently constructed buildings, including shopping centers and business centers. Not recommended for evaluating objects which are in unsatisfactory condition
<b>4 factors:</b> Floor/number of floors (x1); Technical condition (x2); Interior (x3); Transport accessibility (x4).	$y = 0,143x_1 + 0,002x_2 + 0,010x_3 - 0,004x_4$	It is suitable for all types of office and retail real estate and land plots they are located on. Not recommended for evaluating non-standard constructions made with unusual building materials
<b>3 factors:</b> Floor/number of floors (x1); Technical condition (x2); Construction material (x3).	$y = 0,138x_1 + 0,006x_2 + 0,003x_3$	It is suitable in case interior is not important or if there are a lot of auxiliary areas
<b>3 factors:</b> Floor/number of floors (x1); Technical condition (x2); Interior (x3)	$y = 0,141x_1 + 0,0001x_2 + 0,009x_3$	It is suitable in case transport accessibility is not important (for example, if objects are located inside factory areas or there are a lot of auxiliary areas)
<b>3 factors:</b> Floor/number of floors (x1); Interior (x2); Transport accessibility (x3).	$y = 0,142x_1 + 0,003x_2 + 0,0004x_3$	It is suitable for evaluating modern shopping centers and business centers
<b>3 factors:</b> Technical condition (x1); Interior (x2); Construction material (x3).	$y = 0,041x_1 + 0,017x_2 + 0,004x_3$	It is suitable for evaluating objects located in rural areas

**Source:** *Own study.*

This list of recommendations is not complete. In fact, each specific case should be considered separately.

**References:**

Andreev, D.M. 2003. Optimization model for assigning scores for pricing factors. Evaluation issues, 2, 15-19.

Andreev, D.M. 2004. Applicability of optimization methods for digitizing qualitative factor features in regression models of cost estimation. Evaluation issues, 3, 62-64 (In Russian).

Anisimova, I.N., Barinov, N.P., Gribovsky, S.V. 2004. Accounting for heterogeneous pricing factors in multidimensional regression models for real estate valuation. Evaluation issues, 2.

- Anisimova, I.N. 2003. Application of regression analysis methods to assess market value in MS Excel. Bulletin of the Khakass state. University of N.F. Katanov, Ser. 1: Computer Science, 5, 4-18.
- Demidova, P.M., Gorelikov, V.G. 2015. The influence of the share of the value of the land plot of a single property on the market value of the land of horticultural non-profit associations of the Leningrad region. Notes of Mining Institute, 3, 198-202.
- Gladkikh, N.I., Kuznetsova, V.V. 2015. Determining the share of the value of a land plot in the value of a single property for industrial and warehouse use in Chelyabinsk. Property value: assessment and management: proceedings of the Seventh International Scientific and Methodological Conference, Moscow Financial and Industrial University "Synergy", 71-77.
- Kolesnichenko, E.A., Radyukova, Y.Y., Pakhomov, N.N. 2019. The role and importance of knowledge economy as a platform for formation of industry 4.0. Studies in Systems, Decision and Control, 169, 73-82.
- Leifer, L.A. 2017. Recommendations on the determination and justification of the boundaries of the interval in which the value of the valuation subject may be located (paragraphs 25 and 26 of the Federal Security Service No. 1 and paragraph 30 of the Federal Security Service. 7). Nizhny Novgorod, CJSC PTsFKO
- Leifer, L.A., Kashnikova, Z.A. 2006. A modified allocation method for assessing the market value of land for industrial and warehouse purposes. Property relations in the Russian Federation, 10, 93-96.
- Leifer, L.A., Kashnikova, Z.A., Kuzmin, A.V., Leifer, I.L. 2003. The proportion of land in the total cost of a single property. Evaluation issues, 52-56.
- Morozova, E.V. 2015. On the issue of determining the share of a land plot in the cost of a single real estate property. Evaluation issues, 3, 38-53 (In Russian).
- OTSENKA. 2020. Real estate evaluation. Available at: [http://ocenka-plus.pro/f/povesit\\_to.xlsx](http://ocenka-plus.pro/f/povesit_to.xlsx).
- Radyukova, Ya.Yu., Kolesnichenko, E.A., Bespalov, M.V., Pakhomov, N.N. 2017. Systemic Crisis of the World Economy or the "Bourgeois Mode of Production". European Research Studies Journal, 3B(20), 412-431, DOI: 10.35808/ersj/797.
- Sivets, S.A., Levykina, I.A. 2003. Econometric modeling in real estate valuation. Zaporozhye, Polygraph, 220.
- Sutyagin, V.Yu. 2015. Notes on expert amendments. Socio-economic phenomena and processes, 4(163), 29-36.
- Sutyagin, V.Yu., Cherkashnev, R.Yu. 2012. Methodological support for the assessment of built-up areas. Socio-economic phenomena and processes, 12, 296-312.
- Topal, E.G., Baybarina, P.A. 2014. The share of land in the value of a single property as a function of the utilization of the territory. Evaluation issues, 4, 38-53.