DOI: 10.17323/2587-814X.2025.1.93.107

The method for the land plot value appraisal as part of the single real estate object, based on game theory approach*

Michael B. Laskin
E-mail: laskin.m@iias.spb.su

St. Petersburg Federal Research Center of the Russian Academy of Sciences, St. Petersburg, Russia

Abstract

In mass real estate valuation, in cadastral valuation, there is a problem of splitting the value of a single real estate object into the value of land plot and buildings (improvements) located on it. One of the key information sources for real estate valuation is market data. Such data may contain information on offer prices, as well as actual transaction prices (for example, in mortgage transactions) for the whole object. At the same time, in the accounting policy of enterprises different rates of land and property tax often require separate accounting of the value of land plots and the buildings located on them. The problem of such splitting of a single object's value is the subject of permanent discussions in the valuation community. There are no established methods. This article proposes a method of splitting the value of a single property object based on the approach borrowed from co-operative game theory. A simple game formulation of the problem and its fair solution based on the Shepley value are considered. Simple and well-interpretable computational formulas are obtained, which allow us to split the market value of single objects on large data sets in minimum time. The proposed method is new in the theory and practice of valuation.

* The article is published with the support of the HSE University Partnership Programme

Keywords: single real estate object, Shepley value, multiple linear regression, log-normal price distribution, property value splitting

Citation: Laskin M.B. (2025) The method for the land plot value appraisal as part of the single real estate object, based on game theory approach. *Business Informatics*, vol. 19, no. 1, pp. 93–107. DOI: 10.17323/2587-814X.2025.1.93.107

Introduction

One of the current problems of real estate valuation, especially in cadastral valuation, is the problem of splitting the estimated value of a single real estate object1 (SREO) into shares of the value of the land plot (LP) and improvements on it (buildings, structures, etc.). A detailed description of appraisal practice situations that require such a division is given in [1].

In [1], as one of the possible methods for such splitting, a method based on Shapley value is proposed. It should be noted that the application of the Shepley value in machine learning has received quite a lot of attention by researchers in various applied fields (see, for example, [2–7]). Including for estimating the factors' degree of influence in a linear regression model [8–9]. Article [10] is devoted to the application of a multiple linear regression model and Shepley value in the study of the relationship between land and buildings. Article [10] based on market dataset from the city of Montreal.

The present paper considers an approach to solving the problem of splitting the estimated value of SREO arising from the ideology of applying the Shapley value. In this case, the target variable is the estimated value of the SREO, and the pricing factors[2] are the area of land and the area of improvements on it (buildings and structures).

1. Problem statement

Let's introduce the following notations:

V is the offer (or transaction) price for the SREO;

SB is the area of improvements within the SREO;

SP is the area of the land plot within the SREO.

Suppose, that there are n observations in the form of three-dimensional vectors (V_i, SP_i, SB_i), $i = \overline{1,n}$ and it was possible to construct some estimated functional dependence $Y = f(sp, sb)$, in some sense best reflecting the relationship between the target variable and factors (by Y we will understand the estimation of the SREO value, by sp, sb – fixed values of the area of land and improvements, respectively). We will consider the estimation by the formula $Y = f(sp, sb)$ as a result of the cumulative influence of factors SP, SB at the fix values $SP = sp$, $SB = sb$, and functions $Y_1 = f_1(sp)$ and $Y_2 = f_2(sb)$ – as a result of estimates for each of the factors separately, on the same initial data as $Y = f(sp, sb)$.

It is required to create a model of the dependence of the total value of the object and to distribute the value between the land plot and improvements, applying some fairness criterion.

2. Solution method (construction of the Shepley value)

In our case, only two factors are considered. The area of the LP and the area of improvements. In terms of game theory, it is the simplest cooperative game with

[1] A single real estate object is an object which includes a land plot and buildings located on it.

[2] A linear regression model may involve more factors. This article considers two of these factors, since the problem is to split the value of SREO only between these two factors.

two participants (for game theory see, for example, [11, 12]). The Shepley value in this case represents some optimal value distribution, where the contribution of LPs and improvements to the total value is equal to the average contribution of all possible "coalitions." There are only two possible coalitions (*sp, sb*) and (*sb, sp*). Let us construct a table of the values of the "win function," the calculation of the Shepley value and the shares to be assigned to the value of LP and improvements (*Table 1*).

Table 1.

Calculation of the share of the land component and the share of buildings in the SREO valuation using the Shepley value

Coalitions / factors	*sp*	*sb*
(*sp*, *sb*)	Y_1	$Y - Y_1$
(*sb*, *sp*)	$Y - Y_2$	Y_2
Mean by coalitions (Shepley value)	$\frac{Y + Y_1 - Y_2}{2}$	$\frac{Y - Y_1 + Y_2}{2}$
LP shares/building shares	$\frac{Y + Y_1 - Y_2}{2 \cdot Y}$	$\frac{Y - Y_1 + Y_2}{2 \cdot Y}$

In order to transfer the game approach to the problem of splitting the estimated value of SREO, the following conditions must be fulfilled:

1. Efficiency.

2. Additivity across coalitions.

3. Symmetry.

4. Factors that do not affect the outcome do not participate in the model (in game terminology, the "dummy axiom").

We consider the SREO valuation, i.e., the facility has a non-zero building area and a non-zero land area. Hence, $Y_1 > 0$, $Y_2 > 0$. The inequalities $Y > Y_1$, $Y > Y_2$ ensure the fulfilment of the efficiency and additivity conditions (on coalitions). The fulfilment of the third and fourth conditions is ensured by the problem conditions: there are no identical factors in the training model; there are no factors that do not influence the result.

The fulfilment of conditions $Y_1 > 0$, $Y_2 > 0$, $Y > Y_1$, $Y > Y_2$ ensures the existence of the kernel, hence the existence of the sharing and the existence of the Shepley value as the only fair sharing.

Failure to fulfil the condition $Y > Y_1$ ($Y \leqslant Y_1$) means that the evaluation of SREO only on the first factor "area of LP" gives a value not less than the evaluation on two factors. In this case, the improvements in the SREO have negative or zero value.

Failure to fulfil the condition $Y > Y_2$ ($Y \leqslant Y_2$) (means that the assessment of SREO only by the second factor "area of improvements" gives a value not less than the assessment by two factors. In this case, the LP within the SREO[3] has a negative or zero value.

The calculation of the share of the value of land and improvements within the SREO shown in *Table 1* is independent of the model used to estimate the value of the SREO. However, the choice of model may depend on how the market data are organized. In game theory terms, this means that a separate part of the study is the selection of the characteristic function, i.e. the function that forms the winning rule.

3. Construction of the characteristic function

It should be noted that the choice of the characteristic function significantly affects the result of calculations. In some cases, it is possible to obtain a model of the form

[3] It should be noted that free (or conditionally free) land and land within the SREO are different types of real estate. For free (conditionally free) land it is assumed that there is a turnover in the real estate market. There is no turnover of land plots within the SREO as separate plots.

$$V = a + b \cdot sp + c \cdot sb \quad (1)$$

with satisfactory quality indicators.

In this case, we look for Y_1 and Y_2 in the form $Y_1 = a_1 + b_1 \cdot sp$ and $Y_2 = a_2 + c_2 \cdot sb$.

The share on LP is equal to

$$\frac{Y + Y_1 - Y_2}{2 \cdot Y} = \frac{1}{2} + \frac{a_1 - a_2 + b_1 \cdot sp - c_2 \cdot sb}{2 \cdot (a + b \cdot sp + c \cdot sb)},$$

the share on improvements is

$$\frac{Y - Y_1 + Y_2}{2 \cdot Y} = \frac{1}{2} + \frac{a_2 - a_1 - b_1 \cdot sp + c_2 \cdot sb}{2 \cdot (a + b \cdot sp + c \cdot sb)}.$$

A formalized representation in the form of a game. Suppose that two people play the following game: the first player chooses the value of *sp* and gets the payoff Y_1, the second chooses the value of sb and gets the payoff Y_2; if they join in a coalition, they get the payoff $Y = V$ for two. Conditions for the existence of 'sharing' $Y_1 > 0$, $Y_2 > 0$, $Y > Y_1$, $Y > Y_2$, depend on the signs of the coefficients of the linear regression models. How to distribute the gain, if it exists, between the players (in our subject area between the LP and the improvements)?

In real estate valuation problems, a model of the form (1) is often inapplicable due to asymmetric distributions of the values *V, sp, sb*. This means asymmetric errors distribution of the linear regression model. It was shown in [13], that the price distributions formed by successive comparisons converge to a log-normal distribution. The same fact was pointed out by the authors [14, 15]. The areas of improvements often follow the same distribution (see, for example, [16]). The same tendency is also observed for the area of land plots. It should be noted that for land plots the distribution of areas in separate market sectors may not be confirmed by statistical tests due to a large number of comparison objects of the same area (e.g., suburban typical residential settlements formed in the Soviet period). Assuming that the values *V, sp, sb* are jointly distributed log-normally[4], the linear model can be considered in the form:

$$\ln(V) = a + b \cdot \ln(sp) + c \cdot \ln(sb), \quad (2)$$

and the general model of the problem in the form:

$$V = e^a \cdot sp^b \cdot sb^c. \quad (3)$$

In this case, we look for Y_1 and Y_2 in the form

$$\ln(Y_1) = a_1 + b_1 \cdot \ln(sp) \text{ and } \ln(Y_2) = a_2 + c_2 \cdot \ln(sb) \quad (4)$$

and we obtain

$$Y = e^a \cdot sp^b \cdot sb^c;\ Y_1 = e^{a_1} \cdot sp^{b_1};\ Y_2 = e^{a_2} \cdot sb^{c_2}.$$

The share on the LP is equal to

$$\frac{Y + Y_1 - Y_2}{2 \cdot Y} = \frac{1}{2} + \frac{e^{a_1} \cdot sp^{b_1} - e^{a_2} \cdot sb^{c_2}}{2 \cdot e^a \cdot sp^b \cdot sb^c},$$

the share on improvements is equal to

$$\frac{Y - Y_1 + Y_2}{2 \cdot Y} = \frac{1}{2} - \frac{e^{a_1} \cdot sp^{b_1} - e^{a_2} \cdot sb^{c_2}}{2 \cdot e^a \cdot sp^b \cdot sb^c}. \quad (5)$$

The conditions for the existence of "sharing" $Y_1 > 0$, $Y_2 > 0$ are fulfilled, for the conditions $Y > Y_1$, $Y > Y_2$ to be fulfilled, the inequalities depending on models (3) and (4) must be fulfilled

$$e^a \cdot sp^b \cdot sb^c > e^{a_1} \cdot sp^{b_1} \text{ and } e^a \cdot sp^b \cdot sb^c > e^{a_2} \cdot sb^{c_2},$$

which on the plane (*sp, sb*) give the region bounded by the conditions

$$sb > e^{\frac{a_1 - a}{c}} \cdot sp^{\frac{b_1 - b}{c}} \text{ and } sp > e^{\frac{a_2 - a}{b}} \cdot sb^{\frac{c_2 - c}{b}}. \quad (6)$$

As will be shown below, there can also be combined models. The data can be arranged in such a way that different types of models are constructed for the values Y, Y_1, Y_2. In any case, the principle of constructing the Shepley value (*Table 1*) will not change.

[4] In any case, it is often found that asymmetrically distributed prices allow one to proceed to construct a linear model in logarithms for which the conditions of the Gauss-Markov theorem are satisfied.

When estimating the market value of shares of land and improvements within the SREO, it may well turn out that either vacant land or a house on it may be valued more highly than the SREO. Such situations are well interpreted: in the first case the building deteriorates the land compared to the free (conditionally free) land; in the second case the improvements are valued so expensively that the land (as part of the SREO) is worthless or has a negative value compared to them. In cadastral valuation, such interpretation is impossible – the cadastral appraiser is, in any case, obliged to assign some positive cadastral value to both the land and the improvements. At the same time, cadastral valuation performed by mass valuation methods should be carried out as market valuation [17].

For the purposes of cadastral valuation, we modify the values , as follows

$$Y_1 = \begin{cases} f_1(sp), \text{ if } f_1(sp) < Y \\ Y = f(sp,sb), \text{if } f_1(sp) \geq Y, \end{cases}$$

$$Y_2 = \begin{cases} f_2(sb), \text{ if } f_2(sb) < Y \\ Y = f(sp,sb), \text{if } f_2(sb) \geq Y. \end{cases} \quad (7)$$

This choice of Y_1, Y_2, will result in positive cadastral values for both land and improvements.

4. Concept of information support of cadastral offices in calculating cadastral values of land and buildings included in single real estate objects

The approach proposed above is easily implemented in such environments as Python, statistical package R. Their advantage is openness and accessibility to any user. Any appraiser (researcher) interested in applying modern methods of working with large volumes of data can independently master the necessary set of skills to obtain high-quality analytical results. Cadastral offices have their own databases with which the results of apprising can be easily interfaced, as all of them can be downloaded from specialized packages in required formats (as a rule, files with .csv extension). The practical implementation of the method proposed above can be represented by the flowchart shown in *Fig. 1*.

Fig. 1. Block diagram of the programmed code for calculating the shares of values of land and buildings included in the SREO.

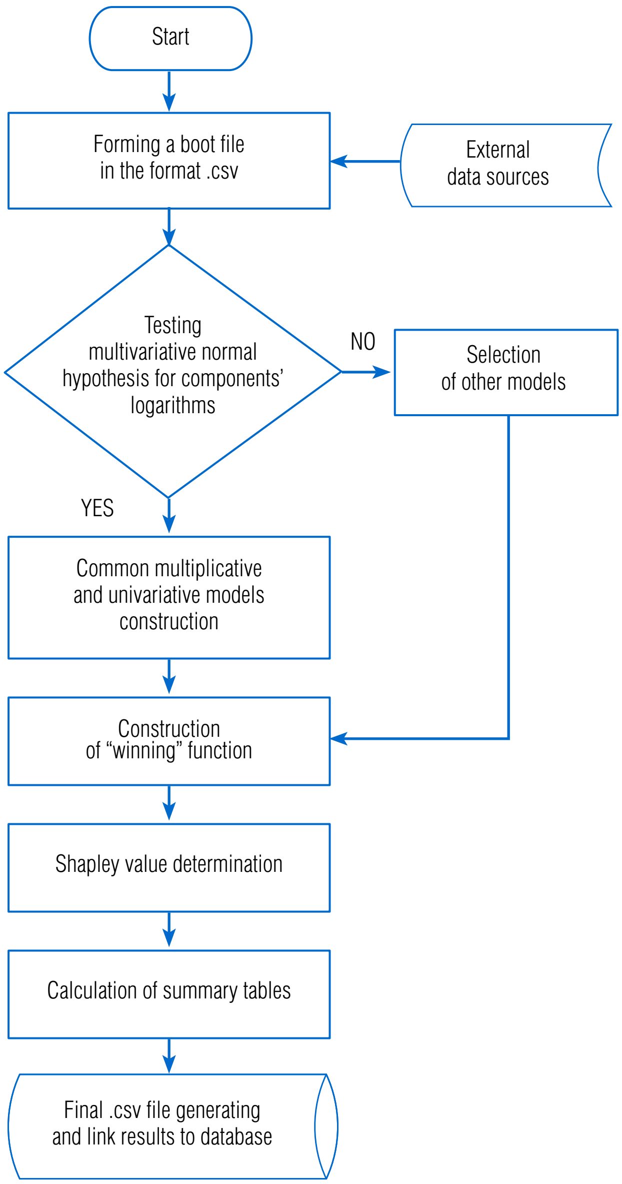

Fig. 1. Block diagram of the programmed code for calculating the shares of values of land and buildings included in the SREO.

A simple program code in an environment such as the R statistical package allows not only to calculate the cost shares for a specific object, but also to create tables (even reference tables, subject to careful selection of data by time and location) and link the results to cadastral databases.

Let us consider an example using data from the article [1] and provided by its authors (sources of primary data sites krasnodar.cian.ru, avito.ru). The total number of observations (objects of comparison) in the article [1] is 49, of which 39 objects are SREO, the remaining 10 are vacant land plots. The region is Krasnodar, and the type of permitted use is individual residential construction. The data are presented in *Table 2*. The calculations were performed in the statistical package R (readers, who are not familiar with R, can be recommended books [18, 19]).

Table 2.

Data on areas of land, buildings and prices of comparison objects

№	Land area, sq. m	Building area, sq. m	Price, million rubles	№	Land area, sq. m	Building area, sq. m	Price, million rubles
1	650.0	80.0	18.6	26	600.0	300.0	47.0
2	500.0	242.0	30.0	27	215.0	0.0	4.0
3	1120.0	720.0	78.0	28	700.0	145.0	16.0
4	580.0	300.0	65.0	29	280.0	97.1	12.5
5	120.0	62.8	6.5	30	600.0	36.0	10.0
6	580.0	49.0	3.1	31	520.0	0.0	12.0
7	800.0	370.0	125.0	32	420.0	50.0	8.8
8	500.0	0.0	15.0	33	450.0	84.0	6.3
9	215.0	123.6	7.0	34	220.0	150.0	12.0
10	640.0	260.0	16.0	35	500.0	84.0	3.9
11	383.0	43.0	9.9	36	450.0	100.0	18.5
12	600.0	50.0	7.8	37	900.0	0.0	35.0
13	616.0	149.0	18.0	38	200.0	160.0	12.0
14	707.0	0.0	8.0	39	314.0	66.3	11.0
15	400.0	90.0	10.0	40	454.0	90.0	9.5
16	300.0	90.0	6.4	41	850.0	0.0	25.0
17	600.0	160.0	19.3	42	300.0	88.0	7.8
18	2330.0	0.0	90.0	43	100.0	50.0	2.5
19	360.0	270.0	25.4	44	490.0	91.0	15.0
20	450.0	85.7	10.5	45	400.0	0.0	14.0
21	350.0	150.0	6.3	46	500.0	0.0	10.0
22	613.0	0.0	23.0	47	400.0	108.0	4.6
23	200.0	56.3	7.5	48	460.0	120.0	17.0
24	700.0	350.0	47.0	49	150.0	44.0	3.0
25	860.0	106.8	6.5				

First of all, it should be noted that vacant (conditionally vacant) land and land within the SREO are different types of real estate. For the first case, there is a market where free land is traded. For the second case, there is no market, the land within the SREO is not sold separately from the buildings and structures located on them; they can be sold only in conjunction with improvements. To build a general model $Y = f\ (sp,\ sb)$ we use data on 39 objects of SREO. To build the model $Y_1 = f_1(sp)$ we use data on vacant land; to build the model $Y_2 = f_2\ (sb)$ we use data on SREO as they have improvements.

The linear regression model on the two factors does not give a satisfactory model (*Table 3*). There is a clear asymmetry of errors and two factors of the model have unsatisfactory *t*-criterion values.

In contrast, the linear regression model built for logarithms of variables gives acceptable results (*Table 4*).

Table 3.

Results obtained using the library function *lm*() of the statistical package R for the linear regression model of the form (1)

Model: $V = a + b \cdot sp + c \cdot sb$					
Residuals:	min	1 Q	median	3 Q	max
	–26.59	–4.96	–1.15	3.91	71.34
Coefficients	Estimate	Standard error	*t*-test statistic	*p*-value *t*-test	
a	–7.191	5.677	–1.267	0.213	<0.05
b	0.015	0.014	1.075	0.289	<0.05
c	0.132	0.023	5.609	0	<0.05
RSE	14.67				
R^2	0.646	Adjusted R^2	0.626		
F-test statistic	32.81	*p*-value *F*-test	0		

Table 4.

Results obtained using the library function *lm*() of the statistical package R for the model in logarithms of the form (2)

Model: $\ln(V) = a + b \cdot \ln(sp) + c \cdot \ln(sb)$					
Residuals:	min	1 Q	median	3 Q	max
	–0.969	–0.237	0.007	0.351	0.351
Coefficients	Estimate	Standard error	*t*-test statistic	*p*-value *t*-test	
a	–4.181	0.921	–4.538	0	<0.05
b	0.425	0.170	2.500	0.017	<0.05
c	0.874	0.132	6.641	0	<0.05
RSE	0.5				
R^2	0.7	Adjusted R^2	0.682		
F-test statistic	41.83	*p*-value *F*-test	0		

For the studied real estate sector, we obtained a model (characteristic function) of the form:

$$V = e^a \cdot sp^b \cdot sb^c = e^{-4.1812} \cdot sp^{0.4246} \cdot sb^{0.8739}. \qquad (8)$$

Substituting the fixed values of *sp* and *sb* into formula (8) we obtain an estimate of the cost of the SREO.

To build the model by area of the SREO only, we select 10 free SREOs from the original data set. For this set we manage to build a satisfactory linear regression model (*Table 5*).

We obtain

$$a_1 = -7.5822,\ b_1 = -\ 0.0413,$$
$$Y_1 = -\ 7.5822 + 0.0413 \cdot sp.$$

In order to build a model by improvement area only, we use only those properties that have buildings (there are 39 of them). For this set we can build a satisfactory regression model only for logarithms (*Table 6*).

Table 5.

Results obtained using the library function *lm*() of the statistical package R for a univariate linear regression model of the form (4)

Model: $V = a_1 + b_1 \cdot sp$					
Residuals:	min	1 Q	median	3 Q	max
	–13.642	–2.393	1.343	4.49	5.38
Coefficients	Estimate	Standard error	*t*-test statistic	*p*-value *t*-test	
a_1	–7.582	3.243	–2.338	0.047	<0.05
b_1	0.041	0.003	11.867	0	<0.05
RSE	6.117				
R^2	0.947	Adjusted R^2	0.941		
F-test statistic	143.2	*p*-value *F*-test	0		

Table 6.

Results obtained using the library function *lm*() of the statistical package R for the univariate model in logarithms of the form (4)

Model: $\ln(V) = a_2 + c_2 \cdot \ln(sb)$					
Residuals:	min	1 Q	median	3 Q	max
	–0.957	–0.398	0.029	0.362	1.104
Coefficients	Estimate	Standard error	*t*-test statistic	*p*-value *t*-test	
a_2	–2.344	0.594	–3.945	0	<0.05
c_2	1.026	0.125	8.233	0	<0.05
RSE	0.532				
R^2	0.647	Adjusted R^2	0.637		
F-test statistic	67.78	*p*-value *F*-test	0		

We obtain

$$a_2 = -2.3443,\ c_2 = 1.0262,\ Y_2 = e^{-2.3443} \cdot sb^{1.0262}.$$

The share on land is $\frac{Y + Y_1 - Y_2}{2}$, the share on building is $\frac{Y - Y_1 + Y_2}{2}$.

Let $sb = 300$, $sp = 600$, then we get:

$$Y = V = e^{-4.1812} \cdot 600^{0.4246} \cdot 300^{0.8739} = \\ = 35.833 \text{ (million rubles)},$$

$$Y_1 = -7.5822 + 0.0413 \cdot 600 = 17.198 \text{ (million rubles)},$$

$$Y_2 = e^{-2.3443} \cdot 300^{1\,0262} = 33.412 \text{ (million rubles)}.$$

The share for land is equal to

$$\frac{Y + Y_1 - Y_2}{2 \cdot Y} = \frac{35.833 + 17.198 - 33.412}{2 \cdot 35.833} \approx 27\%,$$

the share for the building is equal to

$$\frac{Y - Y_1 + Y_2}{2 \cdot Y} \approx 73\%$$

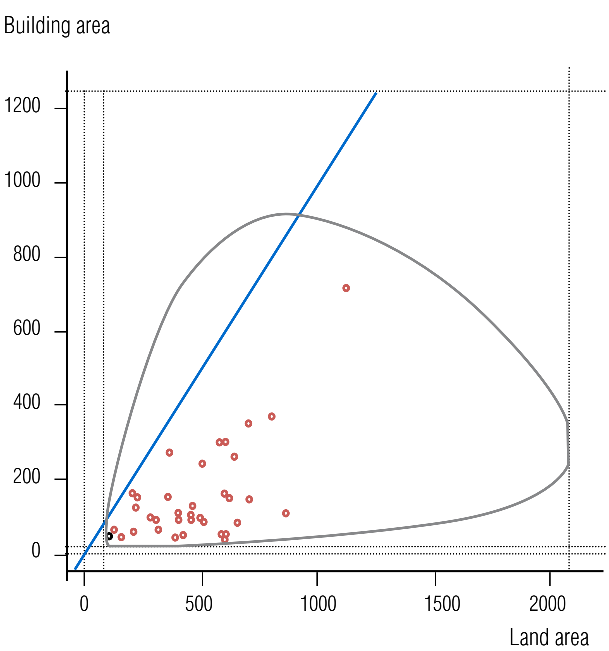

Fig. 2. Scattering diagram of the observed pairs (SP_i, SB_i), $i = \overline{1,n}$.

or in rubles 35.833 · 27% ≈ 9.675 (million rubles) for the land and 35.833 · 73% ≈ 26.158 (million rubles) for the improvements.

Thus, the total appraised value of the SREO with a land plot of 600 square meters and a house on it with an area of 300 square meters is RUR 35.833 million, of which RUR 9.675 million should be attributed to the value of the land plot and RUR 26.158 million to the value of the house. The shares to be allocated to the land and improvements for any values of *sb*, *sp* can be calculated in a similar way. *Table 2* shows the calculated shares of the value of the land within the SREO at different *sb*, *sp*. values. The improvement shares are the difference between 100% and the values in *Table 7.*

In *Table 7*, the blank fields correspond to two cases:

- the value of improvements dominates the value of LP to such an extent that the LP within the SREO has an estimated negative value (bottom left corner of the table);
- the value of the LP dominates the value of the improvements to such an extent that the improvements reduce the value of the LP within the SREO compared to the value of the free LP (upper right corner of the table).

5. Additional model justification and area of factor variation

The model of joint influence of factors on the value of the SREO of the form $V = e^a \cdot sp^b \cdot sb^c$ was constructed as a multiple linear regression model in logarithms. To construct it, it was sufficient to logarithm the factors and the target variable and fit the regression equation. Overall, the model has satisfactory statistical performance (*Table 4*), but has a standard deviation of $RSE = 0.4979$, indicating a noticeable scatter in the observations (V_i, SP_i, SB_i), $i = \overline{1,n}$. Checking the logarithms of the observations for joint normality can provide further argument in favor of this model (on joint normality see, for example, [20]). *Table 8* shows the result of the Mardia test (*MVN* library of the R statistical package).

Table 7.

Shares of the value of land within the SREO

Building area, sq. m	Land area, sq. m: 200	400	600	800	1000	1200	1400	1600
20	**15%**							
60		68%						
100		**41%**	75%					
140		**28%**	**56%**	76%	93%			
180		**20%**	**44%**	**62%**	76%	88%	99%	
220		**15%**	**36%**	**52%**	**65%**	76%	85%	94%
260		**10%**	**30%**	**45%**	**57%**	**66%**	75%	83%
300		7%	**26%**	**40%**	**50%**	**59%**	**67%**	**74%**
340		4%	**22%**	**35%**	**45%**	**54%**	**61%**	**68%**
380		2%	**19%**	**32%**	**41%**	**49%**	**56%**	**63%**
420			**17%**	**29%**	**38%**	**46%**	**52%**	**58%**
460			**15%**	**26%**	**35%**	**43%**	**49%**	**54%**
500			**13%**	**24%**	**33%**	**40%**	**46%**	**51%**
540			**11%**	**22%**	**30%**	**37%**	**43%**	**48%**
580			**10%**	**20%**	**29%**	**35%**	**41%**	**46%**
620			8%	**19%**	**27%**	**33%**	**39%**	**44%**
660			7%	**17%**	**25%**	**32%**	**37%**	**42%**
700			6%	**16%**	**24%**	**30%**	**35%**	**40%**
740			5%	**15%**	**23%**	**29%**	**34%**	**39%**
780			4%	**14%**	**21%**	**27%**	**33%**	**37%**

The *MVN* library of the R statistical package (see [21, 22] for a detailed description), in addition to the Mardia test, contains other tests of joint normality, such as the Royston, Henze-Zirkler, Dornick-Hansen and other tests, which for these data also give positive results. Thus, we have acceptable justifications for the choice of the model, albeit with noticeable standard errors (spread) of the data. Similar validations can be obtained for paired observations (SP_i, SB_i), $i = \overline{1,n}$..

Regression models are widely used in the estimation literature (see, e.g., [23–25]). The author of [26] points out the reasons why it is not recommended to extend regression models beyond the domain of observations. *Figure 2* shows the observed 39 values of the pairs: the area of LP is the area of improvements.

The curved line in *Fig. 2* is the 90% level line of the model log-normal distribution of areas of land and areas of improvements. The straight line is the bisec-

Table 8.

Mardia test result on joint normality of logarithms of initial data

Multivariative normality test Mardia				
Test:	Statistic	p-value	Result (YES – «+», NO – «–»)	
Mardia skewness	16.31	0.091	YES	
Mardia kurtosis	–0.458	0.647	YES	
Mardia MVN			YES	
Univariate normality	Components	Statistic	p-value	Result
Lilliefors (Kolmogorov-Smirnov)	Component 1	0.121	0.161	YES
Lilliefors (Kolmogorov-Smirnov)	Component 2	0.131	0.087	YES
Lilliefors (Kolmogorov-Smirnov)	Component 3	0.11	0.27	YES

Sample description										
Component numbe	Sample size	Mean	Standard error	Median	Min	Max	1 Q	3 Q	Skewness	Kurtosis
1	39	2.497	0.884	2.351	0.912	4.828	1.909	2.904	0.621	0.139
2	39	6.018	0.537	6.109	4.605	5.726	5.727	6.397	–0.752	0.129
3	39	4.718	0.692	4.575	3.584	4.288	4.288	5.043	0.569	–0.259

tor of the first coordinate angle. Above the bisector are cases where the area of improvements is larger than the area of the LP. As a rule, it is assumed that for private houses the area of improvements does not exceed the area of land, but such cases are possible and, as can be seen from *Fig. 2*, the constructed model allows it. However, in our two-dimensional case, the application of the model is better restricted to the area inside the closed curve in *Fig. 2.* A comparison of *Table 7* and *Fig. 2* shows that the shares of the value of LPs are calculated for this area. At the same time, *Table 7* leaves blank fields that have an interpretation (see above) in terms of determining the shares of land and improvements in the market value of the SREO. In cadastral valuation, for any combination of area of land (*sp*) and improvements (*sb*), the cadastral value should be reported as a positive value and, in general, the cadastral value should be estimated as market value or close to it. It is this understanding of cadastral value that minimizes possible claims against cadastral valuation. What to do? In this case, we should set Y_1 and Y_2 so that it does not exceed Y for any combination of areas of land (*sp*) and improvements (*sb*). For example,

$$Y = V = e^{a} \cdot sp^{b} \cdot sb^{c} = e^{-4.1812} \cdot sp^{0.4246} \cdot sb^{0.8739},$$

Table 9.

Share of the value of land within the EON for cadastral purposes

		Land area, sq. m							
		200	400	600	800	1000	1200	1400	1600
Building area, sq. m	20	**17%**	61%	67%	71%	74%	76%	77%	78%
	60	7%	54%	61%	66%	69%	71%	73%	74%
	100	4%	**41%**	58%	63%	66%	69%	71%	72%
	140	3%	**31%**	**56%**	61%	65%	67%	69%	71%
	180	3%	**25%**	**44%**	**59%**	63%	66%	68%	70%
	220	2%	**21%**	**36%**	**52%**	**62%**	65%	67%	69%
	260	2%	**18%**	**30%**	**45%**	**57%**	**64%**	66%	68%
	300	2%	16%	**26%**	**40%**	**50%**	**59%**	**65%**	**67%**
	340	1%	14%	**23%**	**35%**	**45%**	**54%**	**61%**	**67%**
	380	1%	13%	**21%**	**32%**	**41%**	**49%**	**56%**	**63%**
	420	1%	12%	**19%**	**29%**	**38%**	**46%**	**52%**	**58%**
	460	1%	11%	**18%**	**26%**	**35%**	**43%**	**49%**	**54%**
	500	1%	10%	**16%**	**24%**	**33%**	**40%**	**46%**	**51%**
	540	1%	9%	**15%**	**22%**	**30%**	**37%**	**43%**	**48%**
	580	1%	9%	**14%**	**20%**	**29%**	**35%**	**41%**	**46%**
	620	1%	8%	14%	**19%**	**27%**	**33%**	**39%**	**44%**
	660	1%	8%	13%	**17%**	**25%**	**32%**	**37%**	**42%**
	700	1%	7%	12%	**16%**	**24%**	**30%**	**35%**	**40%**
	740	1%	7%	12%	**15%**	**23%**	**29%**	**34%**	**39%**
	780	1%	7%	11%	**14%**	**21%**	**27%**	**33%**	**37%**

$$Y_1 = \begin{cases} -7.5822 + 0.0413 \cdot sp, \\ \text{ if } -7.5822 + 0.0413 \cdot sp < Y \\ Y, \text{if } -7.5822 + 0.0413 \cdot sp \geq Y, \end{cases}$$

$$Y_2 = \begin{cases} e^{-2.3443} \cdot sb^{1.0262}, \text{if } \ e^{-2.3443} \cdot sb^{1.0262} < Y \\ Y, \text{if } \ e^{-2.3443} \cdot sb^{1.0262} \geq Y. \end{cases}$$

Such conditions are easily realized in the script of the statistical package R and for them the shares of land and improvements in the SREO and their monetary expressions can be calculated. The table of shares for the cadastral value of land within the SREO is given in *Table 9*.

The fractions of land (similarly, the fractions of improvements as an addition up to 100%) in *Table 9* meet the objectives of cadastral valuation — the values of all land and buildings located on them will be positive, in total coinciding with the estimated model value of the

SREO. The fractions of land in *Table 9* approximately correspond to the fractions of land in *Table 7* in the main, most important part of the tables corresponding to the area of observations in *Fig. 2* (in bold), i.e., the estimation of the market value share in these fields approximately corresponds to the cadastral one. The discrepancies become apparent as one shifts to the lower left corner and to the upper right corner. These are the areas where the cadastral value is obliged to assign a positive cadastral value in any case, even if the assigned cadastral value differs from the market value share estimate. The bottom left decreases the share of the value of the land within the SREO, the top right decreases the share of the value of the improvements within the SREO.

Conclusion

Land plots within the SREO and vacant (or conditionally vacant) land plots belong to different types of real estate. Land plots within the SREO are not traded on the market without improvements located on them. Their market value can be obtained only as a result of the SREO value sweep. Free land plots are traded on the real estate market; for them objects of comparison can be selected and their market value can be estimated both by the comparative and income approaches.

The Shepley value allows to determine a fair distribution of shares of the value of land and improvements within the SREO.

When estimating the market value of the shares of land and improvements within the SREO, negative values of land and improvements may be obtained. In the first case, the value of the improvements is greater than the value of the SREO (the land within the SREO has a negative value), in the second case, the value of the land is greater than the value of the SREO (the buildings deteriorate the land plot within the SREO compared to the free land plot).

When determining the cadastral value, the land and improvements on it as part of the SREO should have a positive value. The proposed methodology, based on the Shepley value and the appropriate selection of the characteristic function, allows us to split the value of SREO into the values of land and improvements, basically corresponding to the market value. Discrepancies with the market value appear only in rare cases away from the area of observation and only due to special requirements imposed on the cadastral value (base for taxation and accounting). This is a case that illustrate the differences between cadastral and market values.

The proposed method allows us to obtain data from cadastral databases, external sources, performing calculations in a specialized environment and uploading the results of calculations in formats easily linked to cadastral databases. ■

References

1. Kuznetsov D.D., Myagkov V.N. (2023) Land – separately, building – separately. How to solve the problem of unfolding the market value of the pledged object. *Bank Lending*, no. 6(112), pp. 60–76 (in Russian).
2. Bure V.M., Staroverova K.Yu. (2019) Applying cooperative games with coalition structure for data clustering. *Automation and Remote Control*, vol. 80, no. 8, pp. 1541–1551.
3. Rozemberczki B., Watson L., Bayer P., et al. (2022) The Shapley value in machine learning. *arXiv:2202.05594v2.* https://doi.org/10.48550/arXiv.2202.05594
4. Roth J., Bajorath J. (2024) Machine learning models with distinct Shapley value explanations decouple feature attribution and interpretation for chemical compound predictions. *Cell Reports Physical Science*, vol. 5, no. 8, pp. 102–110. https://doi.org/10.1016/j.xcrp.2024.102110

5. Li M., Sun H., Huang Y., Chen H. (2024) Shapley value: from cooperative game to explainable artificial intelligence. *Autonomous Intelligent Systems*, vol. 4, article 2. https://doi.org/10.1007/s43684-023-00060-8

6. Feretzakis G., Sakagianni A., Anastasiou A., et al. (2024) Integrating Shapley values into machine learning techniques for enhanced predictions of hospital admissions. *Applied Sciences*, vol. 14(13), article 5925. https://doi.org/10.3390/app14135925

7. Kjersti A., Jullum M., Loland A. (2020) Explaining individual predictions when features are depended: More accurate approximations to Shapley values. *arXiv:1903.10464v3*. https://doi.org/10.48550/arXiv.1903.10464

8. Mikheenko A.M., Savich D.S. (2020) Application of Shepley value in regression analysis. *Vestnik of I. Kant Baltic Federal University. Ser.: Physic-mathematical and technical sciences*, no. 2, pp. 84–94 (in Russian).

9. Lundberg S.M., Allen P.G. (2017) A unified approach to interpreting model predictions. *ArXiv:1705.07874v2*. https://doi.org/10.48550/arXiv.1705.07874

10. Özdilek Ü. (2020) Land and building separation based on Shapley values. *Palgrave Communications*, vol. 6, article 68. https://doi.org/10.1057/s41599-020-0444-1

11. Petrosyan L.A., Zenkevich N.A., Semina E.A. (1998) *Game Theory*. Moscow: Higher School (in Russian).

12. Mazalov V.V. (2010) *Mathematical theory of games and applications*. St. Petersburg–Moscow–Krasnodar: Lan (in Russian).

13. Rusakov O., Laskin M., Jaksumbaeva O. (2016) Pricing in the real estate market as a stochastic limit. Log Normal approximation. *International Journal of the Mathematical models and methods in applied sciences*, vol. 10, pp. 229–236.

14. Aitchinson J., Brown J.A.C. (1963) *The Lognormal distribution with special references to its uses in economics*. Cambridge: University Press.

15. Ohnishi T., Mizuno T., Shimizu C., Watanabe T. (2011) On the evolution of the house price distribution. *Understanding Inflation Dynamics of the Japanese Economy*, Working Paper Series no. 56.

16. Laskin M.B. (2017) Adjustment of the market value by the pricing factor 'area of the object'. *Property relations in the Russian Federation*, no. 8(191), pp. 86–99 (in Russian).

17. Rosreestr Order No. N/0336 dated 08/04/2021 (as amended on 09/11/2024) "*On Approval of Methodological Guidelines on State Cadastral Valuation*" (Registered with the Ministry of Justice of Russia on 12/17/2021 N 66421). Available at: https://www.consultant.ru/document/cons_doc_LAW_403900/ (accessed 30 January 2025) (in Russian).

18. Kabakov P. (2014) *R in action. Data analysis and visualization in the R language*. Moscow: DMK-Press (in Russian).

19. Zaryadov I.S. (2010) *Introduction to statistical package R: types of variables, data structures, reading and writing information, graphics*. Moscow: PUDN Publishing House (in Russian).

20. Feller W. (1971) *An introduction to probability theory and its applications, vol. 2.* John Wiley & Sons; 2nd edition.

21. Korkmaz S., Goksuluk D., Zararsiz G. (2014) MVN: An R package for assessing multivariate normality. *The R Journal*, vol. 6(2), pp. 151–162. https://doi.org/10.32614/RJ-2014-031

22. Korkmaz S., Goksuluk D., Zararsiz G. (2019) *MVN: An R package for assessing multivariate normality*. Trakya University, Faculty of Medicine, Department of Biostatistics, Edirne, TURKEY, MVN version 5.8 (Last revision 2019-09-27). Available at: https://cran.r-project.org/web/packages/MVN/vignettes/MVN.pdf (accessed 30 January 2025).

23. Puzynya N.Y. (2020) *Modern market trends and market value assessment: collective monograph*. St. Petersburg: Izd-vo Sankt-Peterburgskogo universiteta (in Russian).

24. Barinov N.P. (2022) Application of regression analysis methods in the tasks of individual and mass valuation of real estate objects. *Valuation Issues*, no. 1(106), pp. 34–46 (in Russian).

25. Anisimova I.N., Barinov N.P., Gribovsky S.V. (2004) Accounting for different types of price-forming factors in multidimensional models of real estate valuation. *Valuation Issues*, no. 2, pp. 2–15 (in Russian).

26. Barinov N.P. (2019) Calculation of the uncertainty interval of value estimation by methods of the comparative approach. *Valuation Issues*, no. 4(98), pp. 2–10 (in Russian).

About the author

Michael B. Laskin

Doctor of Sciences (Economics), Candidate of Sciences (Physics and Mathematics), Associate Professor;

Chief Scientist, St. Petersburg Federal Research Center of the Russian Academy of Sciences, 39, 14th Line V.O., St. Petersburg 199178, Russia;

Professor, Department of Information Systems in Economics, St. Petersburg State University, 7–9 Universitetskaya Embankment, St Petersburg 199034, Russia;

E-mail: laskinmb@yahoo.com

ORCID: 0000-0002-0143-4164