DOI: 10.17323/2587-814X.2020.2.48.63

Multidimensional log-normal distribution in real estate appraisals

Michael B. Laskin
E-mail: laskinmb@yahoo.com

St. Petersburg Institute for Informatics and Automation, Russian Academy of Sciences
Address: 39, 14 Line, St. Petersburg 199178, Russia

Abstract

The purpose of the research was to develop a market value appraisal methodology based on a set of a joint logarithmically normal distribution of price-forming factors. Joint logarithmically normal distribution means random vector component logarithms are distributed together jointly normally. This article suggests a method for appraising the real estate market value based on the statistical hypothesis of a joint logarithmically normal distribution and conditional distribution of prices with fixed values of pricing factors. The article suggests a method of offer price analysis from the point of view of its relevance to pricing factor values. We consider the features of the coefficient of development depending on the area of the land plot. Additional arguments are given in favor of estimating market value as a mode of conditional laws of price distribution. An example of a multidimensional log-normal distribution of prices and pricing factors such as the area of the improvements (improvements mean buildings and constructions) area and the land area in real data, i.e. for the case of a three-dimensional random vector. We present a formula for determining the absolute maximum density point of a multidimensional logarithmically normal random vector. The proof is given in the Appendix. The results obtained can be used to create information systems to support decision-making in valuation activities for real estate properties.

Key words: market value; logarithmically normal law of price distribution; multidimensional logarithmically normal distribution, valuation of real estate.

Citation: Laskin M.B. (2020) Multidimensional log-normal distribution in real estate appraisals. *Business Informatics*, vol. 14, no 2, pp. 48–63. DOI: 10.17323/2587-814X.2020.2.48.63

Introduction

One of the most common methods in market value real estate appraisals is the linear regression model of prices with some price-forming factors as regressors. The factors can be qualitative (type of home, encumbrance, floors, window view, the condition of the apartment/room, etc.) and quantitative (area of the object or the land plot, distance to the city center, to the metro, to other infrastructure objects, etc.).

There are various views on division the pricing factors into classes. In the context of this article, we mean splitting into qualitative and quantitative factors in terms of the possibility of representing the values of the factor as a real number (if such a possibility exists, then the factor is quantitative). Combining quantitative and qualitative factors in a single regression model presents a certain difficulty for analysts. However, this problem goes beyond the scope of the present article: here we will limit ourselves only to quantitative (real) factors. Very often such factors, considered as random variables on a set of objects of comparison, can usually be approximated by a logarithmically normal distribution. There are reasons to assume that the prices of properties formed by sequential comparisons follow the logarithmically normal distribution law.

The theoretical reason for the formation of a lognormal general population for prices formed by successive comparisons was given in [1]. The fact of subordination of rental rates in real estate to the logarithmically normal distribution was pointed out by Aitchinson and Brown in 1963 [2]. More recent researchers have also pointed to the logarithmic distribution of prices in real estate [3]. This approach is not yet traditional from the point of view of the existing practice of real estate valuation, since it requires the use of special applied statistical packages that are not used by practicing appraisers, who use a small number of objects for comparison. At the same time, the changing information environment encourages researchers to look for new, non-traditional approaches to real estate valuation. As an example, we can cite the works [4–9] devoted to the method of hedonistic pricing, i.e. the identification of a statistical relationship between the average or median cost of housing, internal and external price-forming factors.

Statistical dependence is usually estimated using models of linear, logarithmic, or partially logarithmic dependence. In general, this same ideology is the basis for the report on cadastral value [10] made by the St Petersburg government department "Cadastral assessment" in 2018. A number of works use non-regression models for estimating residential real estate objects: for example, in [11, 12] neural networks are used to predict the value of residential property, in [13, 14] machine learning methods (random forest, support vector method) are used, and in [15] the results of using such methods as decision trees, naive Bayesian classifier, and AdaBoost are compared. These methods require the use of large data samples. Another approach is to use price indices. For example, the Case-Shiller housing price index is considered in [16]. Articles [17–19] study the re-sale index, which predicts changes in the value of a resold property based on the difference in time and changes in its attributes between the initial sale and subsequent resale. The authors of [20–23] consider a hybrid method that combines a hedonistic approach and a method of re-selling.

The main approach to the study of price bubbles is to use variations of auto regression methods applied to average prices, for example, in [24–28]. Thus, the use of multidimensional logarithmically normal distributions is also in line with current trends in the search for non-traditional methods in real estate valuation.

1. Estimation of market value based on conditional distributions with fixed values of price-forming quantitative factors

Let V – be the price of the offer (or transaction), X_1, ..., X_n – are the quantitative (real) price – forming factors. Let $W = \ln(V)$, $Y_i = \ln(X_i)$, $i = \overline{1, n}$ (then $v = e^W$, $X_i = e^{Y_i}$).

Consider a multidimensional normal random vector $(W, Y_1, ..., Y_n)$ with a mean vector $(\mu_W, \mu_{Y_1}, ..., \mu_{Y_n})$. Let's write the covariance matrix in block form:

$$CV = \begin{pmatrix} \sigma_W^2 & cov\left(W, \vec{Y}\right) \\ cov\left(W, \vec{Y}\right)^T & COV \end{pmatrix},$$

where COV – covariance matrix of a random vector $\vec{Y} = (Y_1, ..., Y_n)$;

$cov\left(W, \vec{Y}\right)$ – vector $\left(\rho_{WY_1}\sigma_w\sigma_{Y_1}, \ldots, \rho_{WY_n}\sigma_w\sigma_{Y_n}\right)$;

$\sigma_W^2, \sigma_{Y_1}^2 \ldots, \sigma_{Y_n}^2$ – variances of random variables W, Y_1, ..., Y_n;

$\rho_{WY_1}, \ldots, \rho_{WY_n}$ – corresponding correlation coefficients.

Then, the conditional expectation of W, if $Y_1 = y_1$, ..., $Y_n = y_n$ equals to

$$E\left(W | Y_1 = y_1, \ldots, Y_n = y_n\right) = \\ = \mu_W + \left(COV^{-1} \times cov\left(W, \vec{Y}\right)^T, \left(\vec{Y} - \overline{\mu_Y}\right)\right),$$

where $\overline{\mu_Y} = \left(\mu_{Y_1}, \ldots, \mu_{Y_n}\right)$. Conditional variance of W, if $Y_1 = y_1$, ..., $Y_n = y_n$ is equal to

$$D\left(W | Y_1 = y_1, \ldots, Y_n = y_n\right) = \\ = \sigma_W^2 - \left(COV^{-1} \times cov\left(W, \vec{Y}\right)^T, cov\left(W, \vec{Y}\right)\right)$$

For fixed values of price-forming factors $X_1 = x_1$, ..., $X_n = x_n$ the most probable value of the offer price (or transaction, depending on what prices were in the source data) V is calculated using the conditional mode formula:

$$Mode(V | X_1 = x_1 = e^{y_1}, \ldots, X_n = x_n = e^{y_n}) = \\ \exp\left(\mu_W + (COV^{-1} \times cov(W, \vec{Y}))^T, \left(\vec{Y} - \overline{\mu_Y}\right) - \right. \\ \left. -\sigma_W^2 + \left(COV^{-1} \times cov(W, \vec{Y})^T, cov(W, \vec{Y})\right)\right). \quad (1)$$

Under the terms of Federal Law No 135 [29], the market value is the most probable price at which the evaluation object can be alienated on the open market in conditions of perfect competition. In practice, appraisers tend to use an average or median estimation. Such estimations can be based on conditional expectation and a conditional median:

$$E(V | X_1 = x_1 = \exp(y_1), \ldots, X_n = x_n = \exp(y_n)) = \\ = \exp(\mu_W + (COV^{-1} \times cov\left(W, \vec{Y}\right)^T, \left(\vec{Y} - \overline{\mu_Y}\right)) + \\ + \frac{1}{2}\sigma_W^2 - \frac{1}{2}\left(COV^{-1} \times cov\left(W, \vec{Y}\right)^T, cov\left(W, \vec{Y}\right)\right)). \quad (2)$$

$$Median\left(V | X_1 = x_1 = \right. \\ \left. \exp(y_1), \ldots, X_n = x_n = \exp(y_n)\right) = \\ = \exp(\mu_W + COV^{-1} \times cov(W, \vec{Y})^T, (\vec{Y} - \overline{\mu_Y}). \quad (3)$$

Thus, if with respect to some ensemble of quantitative pricing factors and the prices of objects of comparison can be adopted as a working hypothesis on the joint log-normal distribution (joint normal distribution of the logarithms) component of a random vector, then the valuation can be accepted in the evaluation by the formula (1). Estimates according to the formulas (2) and (3) can also be taken; but it should be noted that they do not follow the definition of market value in accordance with Federal Law No 135.

Let's consider an example that uses real data collected by well-known Russian appraisers and was published on the resource [30]. The data set includes 40 real estate objects of industrial and warehouse use with a location in the same region (St Petersburg), on offer for sale in the same time period. Since the authors of the example justified the rejection of a number of adjustments, in our example we will also consider the data comparable and comparable without additional adjustments. Industrial and warehouse purpose real estate is considered as a unit complex consisting of a land plot and improvements (buildings). The data set is presented in *Table 1.*

The items compared are considered as existing industry and warehouse properties that are offered for sale in the current use. We will build a general method for estimating the market value (without auction discount), if the area

Table 1.

Source data

Building area (sq. m)	Land area (sq. m)	Offer prices (rubles)	Price to improvements square ratio (rubles per 1 sq. m of improvements)	Building area (sq. m)	Land area (sq. m)	Offer prices (rubles)	Price to improvements square ratio (rubles per 1 sq. m of improvements)
400	2 500	20 500 000	51 250	5 292	11 143	56 000 000	10 582
750	5 000	18 000 000	24 000	5 300	16 000	220 000 000	41 509
1 081	3 378	26 000 000	24 052	6 011	11 319	135 000 000	22 459
1 130	6 638	27 500 000	24 336	6 013	20 781	90 000 000	14 968
1 320	4 167	31 500 000	23 864	6 060	21 790	179 000 000	29 538
1 440	10 000	160 000 000	111 111	6 123	2 390	152 490 000	24 904
1 790	3 462	93 000 000	51 955	6 479	7 337	119 000 000	18 367
1 900	13 000	85 000 000	44 737	6 756	4 220	90 000 000	13 321
2 125	5 623	85 000 000	40 000	10 000	12 000	420 000 000	42 000
2 642	5 183	75 000 000	28 388	10 300	17 000	312 000 000	30 291
2 700	6 800	59 000 000	21 852	10 672	12 194	350 000 000	32 796
1 820	2 737	32 000 000	17 582	10 990	30 000	480 000 000	43 676
2 250	9 252	84 000 000	37 333	12 000	30 000	300 000 000	25 000
2 973	5 388	90 000 000	30 272	13 000	55 000	200 000 000	15 385
3 513	10 000	80 000 000	22 773	14 428	33 000	385 000 000	26 684
3 600	5 000	95 000 000	26 389	15 000	37 000	840 000 000	56 000
4 000	13 558	140 000 000	35 000	18 924	20 600	800 000 000	42 274
4 124	12 866	91 000 000	22 066	22 312	40 162	338 541 000	15 173
4 167	5 000	125 000 000	29 998	34 082	478 000	2 500 000 000	73 353
4 257	6 861	128 500 000	30 186	35 000	160 000	2 400 000 000	68 571

of improvements and land are fixed (of course, at the same time period, same real estate class, and the same region).

In this case, there are random variables V – the offer price per 1 sq. m of improvements, SB – the area of improvements, SP – the area of the land plot. They form a three-dimensional random vector (V, SB, SP). Let $W = \ln(V)$, $Y = \ln(SB)$, $Z = \ln(SP)$ (then $v = e^{W}$, $SB = e^{Y}$, $SP = e^{Z}$). For a three-dimensional normal random vector (W, Y, Z) the mean vector is equal to (μ_W , μ_Y, ..., μ_Z). The covariance matrix looks like:

$$CV = \begin{pmatrix} \sigma_W^2 & \rho_{WY}\sigma_W\sigma_Y & \rho_{WZ}\sigma_W\sigma_Z \\ \rho_{YW}\sigma_W\sigma_Y & \sigma_Y^2 & \rho_{YZ}\sigma_Y\sigma_Z \\ \rho_{ZW}\sigma_W\sigma_Z & \rho_{ZY}\sigma_Y\sigma_Z & \sigma_Z^2 \end{pmatrix},$$

or:

$$CV = \begin{pmatrix} \sigma_W^2 & cov\left(W,\vec{Y}\right) \\ cov\left(W,\vec{Y}\right)^T & COV \end{pmatrix},$$

where $COV = \begin{pmatrix} \sigma_Y^2 & \rho_{YZ}\sigma_Y\sigma_Z \\ \rho_{ZY}\sigma_Y\sigma_Z & \sigma_Z^2 \end{pmatrix}$; $\vec{Y} = (Y,Z)$

$$cov\left(W,\vec{Y}\right) = \left(\rho_{WY}\sigma_W\sigma_Y,\ \rho_{WZ}\sigma_W\sigma_Z\right);$$

$\sigma_W^2, \sigma_Y^2, \sigma_Z^2$ – variances of random variables W, Y, Z;

$\rho_{WY} = \rho_{YW}$, $\rho_{WZ} = \rho_{ZW}$, $\rho_{YZ} = \rho_{ZY}$ – corresponding correlation coefficients.

Conditional expectation of W, if Y, $Z = z$:

$$E\left(W|Y = y, Z = z\right) = \\ = \mu_W + \left(COV^{-1} \times cov\left(X,\vec{Y}\right)^T,\ \left(y-\mu_Y, z-\mu_Z\right)\right).$$

Conditional variance of W, if Y, $Z = z$:

$$D\left(W|Y = y, Z = z\right) = \\ = \sigma_W^2 - \left(COV^{-1} \times cov\left(X,\vec{Y}\right)^T,\ cov\left(X,\vec{Y}\right)\right).$$

Let's set the values of the area of improvements $SB = sb$ and the area of the land plot $SP = sp$. In accordance with the above notation $Y = \ln(SB)$, $Z = \ln(SP)$, $y = \ln(sb)$, $z = \ln(sp)$. The most probable value of the offer price V for known values of the area of improvements and land area is calculated using the formula [31]:

$$Mode(V|SB = sb, SP = sp) = \\ = \exp(\mu_W + (COV^{-1} \times cov\left(X,\vec{Y}\right)^T, (y-\mu_Y, z-\mu_Z) - \\ -\sigma_W^2 + \left(COV^{-1} \times cov\left(W,\vec{Y}\right)^T,\ cov\left(W,\vec{Y}\right)\right))). \quad (4)$$

Conditional expectation:

$$E(V|SB = sb,\ SP = E) = \\ \mu_W + (COV^{-1} \times cov\left(X,\vec{Y}\right)^T, (y-\mu_Y, z-\mu_Z)) + \\ +\frac{1}{2}\sigma_W^2 - \frac{1}{2}\left(COV^{-1} \times cov\left(W,\vec{Y}\right)^T,\ cov\left(W,\vec{Y}\right)\right)). \quad (5)$$

Conditional median:

$$Median(V|SB = sb,\ SP = sp) = \exp(\mu_W + \\ + \left(COV^{-1} \times cov\left(W,\vec{Y}\right)^T,\ \left(y-\mu_Y, z-\mu_Z\right)\right)). \quad (6)$$

Before applying formulas (4)–(6) to the data in *Table 1*, let's check whether there are grounds to assume the lognormality of distributions of components of a random vector (V, SB, SP) (joint normality of logarithms of their components). The Kolmogorov–Smirnov parametric test was used to check marginal distributions. The following p-value figures are received:

V – price of 1 sq. m of improvements: with parameters meanlog = 10.3 and sdlog = 0.43 p-value is equal 0.7016;

SB – improvements area: with parameters meanlog = 8.45 and sdlog = 1.02 p-value is equal to 0.9761;

SP– plot of land area: with parameters meanlog = 9.3 and sdlog = 1.01 p-value is equal to 0.8963.

Let's check the three studied random variables for the joint normality of logarithms. To do this, we use a well-known condition of joint normality: in order for a multidimensional random vector to have a multidimensional normal distribution, it is necessary and sufficient that any linear combination of its components is distributed normally. The following procedure was implemented in the statistical package R environment:

✦ we take the logarithm from variables

✦ the resulting logarithmic values of the variables are centered and normalized, each with its own standard deviation;

✦ using the standard function R unif(3,0,1), three weight coefficients are generated, and the coefficients are normalized by their sum;

✦ a linear combination of centered and normalized logarithms is formed with random positive coefficients equal to one;

✦ the resulting linear combination is tested using the Kolmogorov-Smirnov normality test, and the test result is written as a *p*-value in the array;

✦ the procedure with a random linear combination is repeated a specified number of times, each time the *p*-value is written. Then the total *p*-value array is compared to the critical level (0.05).

Figure 1 shows a histogram in which *p*-values were obtained when the test was repeated 100 000 times.

The minimum of *p*-value is equal to 0.2867691; it is more than 0.05. The mentioned procedure of 100 000 time test repeating of random linear combinations of components of random vector (W, Y, Z) seems like a reason for keeping the joint logarithmically component distribution hypothesis as the working one.

For the logarithms of the variables "Ratio of price to area of improvements," "Area of buildings," "Area of land" specified in *Table 1*, the following values of the mean vector and covariance matrix are obtained (*Table 2*).

The means are the following: $\mu_W = 10{,}2993$; $\mu_Y = 8{,}4469$; $\mu_Z = 9{,}3506$, $\sigma_W^2 = 0{,}2381$, $\rho_{WY}\sigma_W\sigma_Y = \rho_{YW}\sigma_W\sigma_Y = 0{,}0108$; $\rho_{WZ}\sigma_W\sigma_Z = \rho_{ZW}\sigma_W\sigma_Z = 0{,}1467$; $\sigma_Y^2 = 1{,}0635$, $\rho_{YZ}\sigma_Y\sigma_Z = \rho_{ZY}\sigma_Y\sigma_Z = 0{,}8978$; $\sigma_Z^2 = 1{,}2140$.

In the statistical package R, a program code was implemented that allows us to calculate the market value estimation based on the specified values of the parameters "Building (improvements) area" and "Land area" (formula (4)).

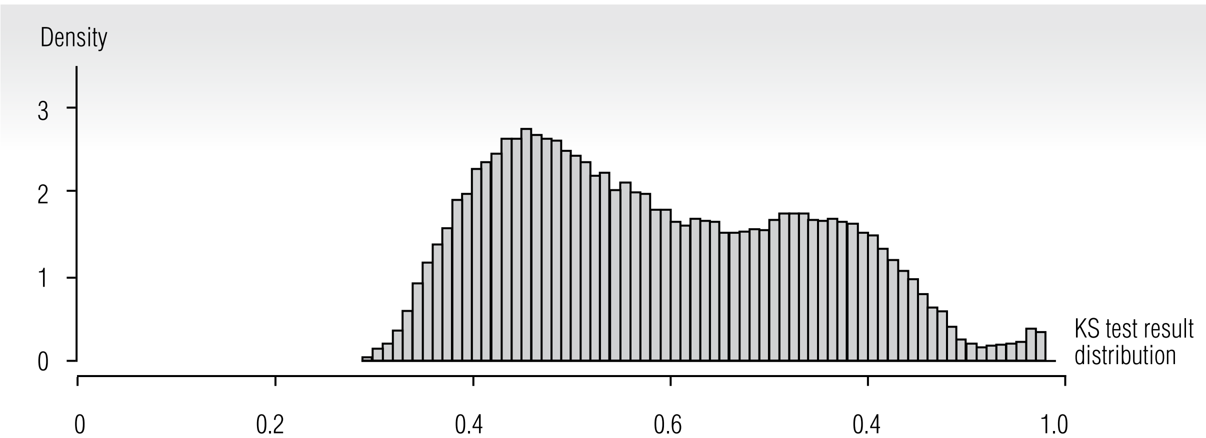

Fig. 1. Results of testing random linear combinations of centered and normalized vector components (W, Y, Z) = (ln(V), ln(SB), ln(SP)) on joint normality by the Kolmogorov–Smirnov test

Table 2.

Means of the logarithms

Ratio of price to area of improvements (V)	Area of buildings (SB)	Area of land (SP)
Means of logarithm		
10.2993	8.4469	9.3506
Covariance matrix		
0.2381	0.0108	0.1467
0.0108	1.0635	0.8978
0.1467	0.8978	1.2140

Similar calculations can be performed for estimates based on median values or on mathematical expectations (formulas (5) and (6)). The results are shown in *Table 3.*

This table shows the following:

✦ the mode estimate is always lower than the median estimate, and the median estimate is always lower than the mathematical expectation estimate (author's opinion: the market value should be defined as a mode estimate, in accordance with the terms of Federal Law No 135, taking into account the asymmetric distribution of prices, areas and distances observed in the market);

✦ if the area of the land plot is constant, the market value (1 sq. m of improvements) decreases as the area of improvements increases;

✦ if the area of improvements is constant, the market value (1 sq. m of improvements) increases as the land area increases;

✦ the formula (4) can be used to calculate the market value of a property of the same class as the comparison items for any values of improvement areas and land (on the same date, for the same location). Since there is no general consensus in the evaluation community regarding the numerical characteristics used for estimating market value (mode, median, mathematical expectation), formulas (5) and (6) can be applied, but, strictly speaking, these do not follow the definition of market value in accordance with Federal Law No 135.

2. The ratio of the area of improvements to land square if the price offers are fixed

In [30] the authors considered the question of pricing trends in the property market for industrial and storage purposes, and the dependence of the market value on the "density factor" (development coefficient) of the land, which is defined as the ratio of the area of improvements to the area of land. The model of joint logarithmically normal distribution of components of a random vector (V, SB, SP) considered in this article also allows us to look at the problem of forming price trends. The difference is that all the components of the random vector (V, SB, SP) are distributed on a positive half-axis; for each given value $V = v$, we can specify the most probable values of the components SB (improvement area) and SP (land area) corresponding to the offer price. In contrast to the previous case (estimation of market value based on the specified values SB and SP), the area of possible deviations from the most probable (median, average) values is not on the numeric axis, but on the plane and consists (as will be shown below) of nested sets obtained from the scattering ellipses of logarithmic values SB and SP in the inverse exponential transformation of the plane.

Let the offer price $V = v$ be known. It is necessary to estimate the ratio of the area of improvements and land for a class of objects with such an initial offer price, i.e. to select objects with lower, middle and upper price trends [30]. Denote the former: V – bid price, SB – area of improvements, SP – area of land, $W = \ln(V)$, $Y = \ln(SB)$, $Z = \ln(SP)$ (then $V = e^W$, $SB = e^Y$, $SP = e^Z$).

Table 3.

Estimates of market value per 1 sq. m of improvements for various values of improvement areas, land plots

Moda estimation		Plot of land in sq. m.									
		2 000	**7 000**	**12 000**	**17 000**	**22 000**	**27 000**	**32 000**	**37 000**	**42 000**	**47 000**
Improvements area in sq. m.	**400**	26 247	38 298	45 058	50 049	54 096	57 543	60 568	63 279	65 745	68 014
	2 400	16 938	24 714	29 076	32 297	34 909	37 133	39 085	40 835	42 426	43 890
	4 400	14 605	21 310	25 072	27 849	30 101	32 019	33 702	35 211	36 583	37 845
	6 400	13 327	19 445	22 877	25 411	27 466	29 216	30 752	32 129	33 381	34 533
	8 400	12 469	18 194	21 406	23 777	25 700	27 337	28 775	30 062	31 234	32 312
	10 400	11 835	17 269	20 317	22 567	24 392	25 947	27 311	28 533	29 645	30 668
	12 400	11 337	16 542	19 462	21 618	23 366	24 855	26 161	27 332	28 397	29 377
	14 400	10 930	15 948	18 763	20 842	22 527	23 962	25 222	26 351	27 378	28 323
	16 400	10 588	15 449	18 176	20 189	21 822	23 212	24 433	25 527	26 521	27 436
	18 400	10 294	15 021	17 672	19 629	21 217	22 569	23 755	24 818	25 786	26 675
Median estimation		**Plot of land in sq. m.**									
		2 000	**7 000**	**12 000**	**17 000**	**22 000**	**27 000**	**32 000**	**37 000**	**42 000**	**47 000**
Improvements area in sq.m.	**400**	31 947	46 615	54 843	60 918	65 844	70 039	73 722	77 021	80 023	82 784
	2 400	20 616	30 081	35 391	39 311	42 490	45 197	47 573	49 703	51 640	53 421
	4 400	17 777	25 938	30 517	33 897	36 638	38 972	41 021	42 858	44 528	46 064
	6 400	16 221	23 668	27 846	30 930	33 431	35 561	37 431	39 106	40 630	42 032
	8 400	15 177	22 146	26 055	28 941	31 281	33 274	35 023	36 591	38 017	39 329
	10 400	14 405	21 019	24 729	27 468	29 690	31 581	33 242	34 730	36 083	37 328
	12 400	13 799	20 134	23 688	26 312	28 440	30 252	31 843	33 268	34 564	35 757
	14 400	13 304	19 412	22 838	25 368	27 419	29 166	30 700	32 074	33 324	34 473
	16 400	12 887	18 804	22 123	24 574	26 561	28 253	29 739	31 070	32 281	33 395
	18 400	12 530	18 283	21 510	23 892	25 824	27 470	28 914	30 208	31 385	32 468
Expectation estimation		**Plot of land in sq. m.**									
		2 000	**7 000**	**12 000**	**17 000**	**22 000**	**27 000**	**32 000**	**37 000**	**42 000**	**47 000**
Improvements area in sq. m.	**400**	35 246	51 428	60 506	67 208	72 643	77 271	81 334	84 974	88 285	91 332
	2 400	22 744	33 187	39 045	43 370	46 877	49 864	52 485	54 835	56 971	58 937
	4 400	19 612	28 616	33 668	37 397	40 421	42 996	45 257	47 283	49 125	50 820
	6 400	17 895	26 112	30 721	34 124	36 883	39 233	41 296	43 144	44 825	46 372
	8 400	16 744	24 432	28 745	31 929	34 511	36 710	38 640	40 369	41 942	43 389
	10 400	15 893	23 189	27 283	30 305	32 755	34 842	36 674	38 315	39 809	41 182
	12 400	15 224	22 213	26 134	29 029	31 377	33 376	35 130	36 703	38 133	39 449
	14 400	14 677	21 416	25 196	27 987	30 250	32 178	33 869	35 385	36 764	38 033
	16 400	14 218	20 746	24 408	27 111	29 304	31 171	32 810	34 278	35 614	36 843
	18 400	13 824	20 170	23 731	26 359	28 491	30 306	31 899	33 327	34 626	35 821

As before, we consider a three-dimensional normal random vector (W, Y, Z) with a mean vector (μ_W, μ_Y, μ_Z) and covariance matrix

$$CV = \begin{pmatrix} \sigma_W^2 & \rho_{WY}\sigma_W\sigma_Y & \rho_{WZ}\sigma_W\sigma_Z \\ \rho_{YW}\sigma_W\sigma_E & \sigma_Y^2 & \rho_{YZ}\sigma_Y\sigma_Z \\ \rho_{ZW}\sigma_W\sigma_Z & \rho_{ZY}\sigma_Y\sigma_Z & \sigma_Z^2 \end{pmatrix},$$

or:

$$CV = \begin{pmatrix} \sigma_W^2 & cov\left(W,\vec{Y}\right) \\ cov\left(W,\vec{Y}\right)^T & COV \end{pmatrix},$$

where $COV = \begin{pmatrix} \sigma_Y^2 & \rho_{YZ}\sigma_Y\sigma_Z \\ \rho_{ZY}\sigma_Y\sigma_Z & \sigma_Z^2 \end{pmatrix}$;

$\vec{Y} = (Y, Z)$

$cov\left(W,\vec{Y}\right) = \left(\rho_{WY}\sigma_W\sigma_Y,\ \rho_{WZ}\sigma_W\sigma_Z\right)$;

$\sigma_W^2, \sigma_Y^2, \sigma_Z^2$ – variances of random variables W, Y, Z;

$\rho_{WY} = \rho_{YW}, \rho_{WZ} = \rho_{ZW}, \rho_{YZ} = \rho_{ZY}$ – corresponding correlation coefficients.

Conditional expectation of vector $\vec{Y} = (Y, Z)$ if $W = w$:

$$E\left(\vec{Y}|W = w\right) = \bar{\mu} + \frac{cov\left(W,\vec{Y}\right)^T}{\sigma_W^2}\left(w - \mu_E\right) =$$

$$\begin{pmatrix} \mu_Y \\ \mu_Z \end{pmatrix} + \begin{pmatrix} \rho_{YW}\dfrac{\sigma_Y}{\sigma_W}\left(w - \mu_W\right) \\ \rho_{ZW}\dfrac{\sigma_Z}{\sigma_W}\left(w - \mu_W\right) \end{pmatrix} = \tag{7}$$

$$\begin{pmatrix} \mu_Y + \rho_{YW}\dfrac{\sigma_E}{\sigma_W}\left(w - \mu_W\right) \\ \mu_Z + \rho_{ZW}\dfrac{\sigma_Z}{\sigma_W}\left(w - \mu_W\right) \end{pmatrix}.$$

Conditional covariance matrix if $W = w$:

$$COV\left(\vec{Y}|W = w\right) =$$

$$= COV - \frac{cov\left(W\vec{Y}\right)^T \times cov\left(W,\vec{Y}\right)}{\sigma_W^2} =$$

$$= \begin{pmatrix} \sigma_Y^2 & \rho_{YZ}\sigma_Y\sigma_Z \\ \rho_{ZY}\sigma_Y\sigma_Z & \sigma_Z^2 \end{pmatrix} -$$

$$- \begin{pmatrix} \rho_{YW}^2\sigma_Y^2 & \rho_{YW}\rho_{ZW}\sigma_Y\sigma_Z \\ \rho_{YW}\rho_{ZW}\sigma_Y\sigma_Z & \rho_{ZW}^2\sigma_Z^2 \end{pmatrix} = \tag{8}$$

$$= \begin{pmatrix} \sigma_Y^2\left(1 - \rho_{YW}^2\right) & \sigma_Y\sigma_Z\left(\rho_{YZ} - \rho_{YW}\rho_{ZW}\right) \\ \sigma_Y\sigma_Z\left(\rho_{YZ} - \rho_{YW}\rho_{ZW}\right) & \sigma_Z^2\left(1 - \rho_{ZW}^2\right) \end{pmatrix}.$$

Let $V = v$. In accordance with the notation introduced above $W = \ln(V)$, $w = \ln(v)$. The most probable combination of *SB* and *SP* in the condition if $V = v$:

$$Mode(\vec{Y}|V = v) =$$

$$= \exp\left(\bar{\mu} + \frac{cov\left(W,\vec{Y}\right)^T}{\sigma_W^2}\left(w - \mu_W\right) - \right.$$

$$\left. - COV + \frac{cov\left(W\vec{Y}\right)^T \times cov\left(W,\vec{Y}\right)}{\sigma_W^2}\right).$$

Table 4 shows the results: most probable combination of SB and SP in a few cases of bid prices.

It should be noted that for each value of the offer price, the most probable pair of *SB, SP* values is the only one (the building density coefficient in this case corresponds to the most probable pair *SB, SP*). Trying to present as the most convenient another pair of components of *SB* and *SP* means choosing a point when there are many other equally probable points, with a density less than the maximum, and for which the building density coefficients will obviously be different. *Figure 2* shows images of two-dimensional distributions of *SB* and *SP* for the offer price of 7.000 rubles, 28.000 rubles (from left to right), 100.000 rubles per 1 sq. m of improvements.

It is possible to see that any other points in plane *SP, SB* have any set of equal-probability points. The sets of such points shown on *Figure 3* (for $V = 28\,000$ rubles per 1 sq. m of improvements).

Table 4.

Most probable combination of *SB* and *SP*, density factor in a few cases of bid prices

Bid price on 1 sq.m. of improvements	7 000	12 000	21 000	28 000	40 000	60 000	80 000	100 000
Most probable pair:								
Area of improvements	619	634	650	659	669	682	691	698
Plot of land in sq.m.	630	878	1 239	1 479	1 843	2 365	2 824	3 240
Dencity factor	0.98	0.72	0.52	0.45	0.36	0.29	0.24	0.22

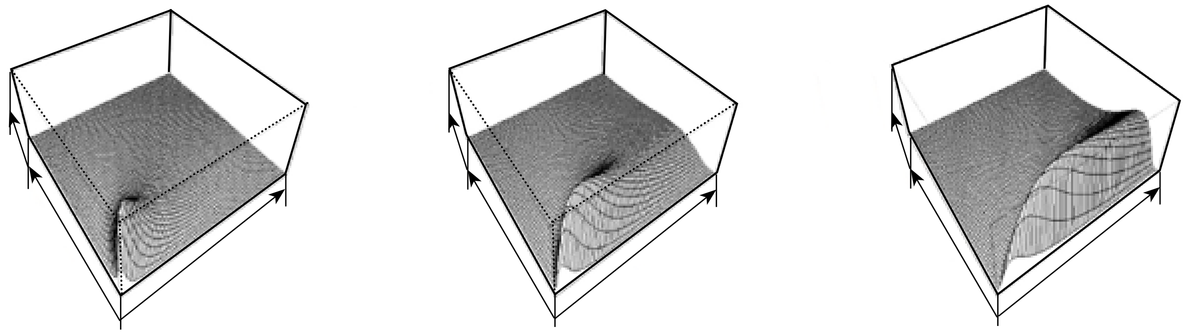

Fig. 2. Two–dimensional distributions of *SP* (improvements area)

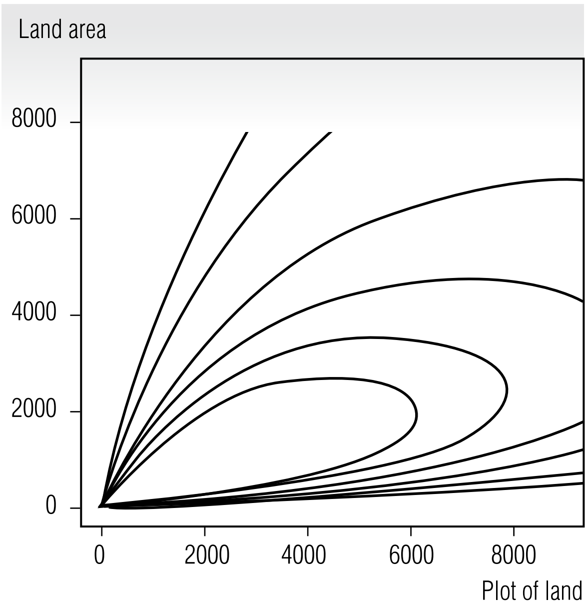

Fig. 3. Equal–probability level lines for *SP* and *SB*

3. The density factor (ratio of the improvements area to land area)

Let's assume that the price of the offer (or transaction) is known. Our goal is to estimate what the coefficient of building density should be at a given price and the given area of the land plot. Let's use the formulas (7) and (8). For a fixed offer price ($V = v$), formulas (7) and (8) give the calculated values of the conditional mathematical expectations of the improvement area logarithms ($SB| V = v$), the land area ($SB| V = v$), and the conditional covariance matrix. Additionally let's assume that the area of the land plot is also known. We introduce new notation for conditional logarithms of the improvement area ($SB| V = v$) and the land area ($SB| V = v$):

$$\mu_{condSB} = \mu_Y + \rho_{YW} \frac{\sigma_Y}{\sigma_W}\left(w - \mu_W\right),$$

$$\mu_{condSP} = \mu_Z + \rho_{ZW} \frac{\sigma_Z}{\sigma_W}\left(w - \mu_W\right),$$

$$\sigma^2_{condSB} = \sigma^2_Y\left(1 - \rho^2_{YW}\right),$$

$$\sigma^2_{condSP} = \sigma^2_Z\left(1 - \rho^2_{ZW}\right),$$

$$\rho = \sigma_Y \sigma_Z\left(\rho_{YZ} - \rho_{YW}\rho_{ZW}\right)$$

(”cond” subscripts mean “conditional”). Consider a two-dimensional random vector ($SB\,|\,V = v$, $SP\,|V = v$) with the specified parameters. For a given value of the land plot area and a given value of the price (in the example of the offer price) ($SP = sp$, $V = v$), the conditional mode of SB (improvement area) is equal to (by analogy with the proof given in [32]):

$$Mode\left(SP|\,\mathrm{SP} = \mathrm{sp}, \mathrm{V} = \mathrm{v}\right) =$$

$$= \exp(\mu_{condSP} + \rho \times \frac{\sigma^{2}_{condSP}}{\sigma^{2}_{\alpha lSB}}\left(\ln\left(sb - \mu_{condSB}\right)\right) - \quad (9)$$

$$- \sigma^2_{condSP}\left(1 - \rho^2\right)).$$

Conditional median of SP is equal to:

$$Median(SP\,|\,\mathrm{SP} = \mathrm{sp}, \mathrm{V} = \mathrm{v}) =$$

$$= \exp(\mu_{condSP} + \rho \times \frac{\sigma^2_{condSP}}{\sigma^2_{condSB}}\left(\ln\left(sb - \mu_{condSB}\right)\right)).$$

Conditional expectation of SP is equal to:

$$E(SP\,|\,\mathrm{SP} = \mathrm{sp}, \mathrm{V} = \mathrm{v}) = \exp(\mu_{condSP} +$$

$$+ \rho \times \frac{\sigma^2_{condSP}}{\sigma^2_{condSB}}\left(\ln\left(sb - \mu_{condSB}\right)\right) + \frac{1}{2}\sigma^2_{condSP}\left(1 - \rho^2\right)).$$

Let’s assume the need to estimate the density factor in a group of items in the lower, middle, or upper price category. Such estimates can be constructed depending on the area of the land plot by modal, median or average values. However, the appearance of the surfaces shown in *Figure 2* suggests that the most conservative estimates will be based on modal values. Estimates for the median or average values seems overestimated (for $V = 28.000$ rubles/sq. m approximately 1.4 and 1.7 times, *Figure 4*). Let’s assume that we are interested in the following question: if the offer price is 28.000 rubles per sq. m and if the area of the land plot is equal to 30,000 square meters, then what area of improvements (and, accordingly, what coefficient of building density) should be considered adequate for such a price and land area. Under the development coefficient (density factor), we will understand the ratio of the estimated value of the area of improvements to the area of the land plot, i.e.

$$\frac{Mode\left(SB|SP = sp\right)}{sp}$$

(alternatively, $\frac{Median\left(SB|SP = sp\right)}{sp}$ or

$$\frac{E\left(SB|SP = sp\right)}{sp}).$$

Figure 4 shows that estimates for modal, median, and average values can differ significantly. Applying formula (9) to the results of calculating conditional parameters at the price of 28.000 rubles/sq. m and the value of the land area equal to 30.000 sq. m gives the result of 7.165 sq. m of improvements, and then the building density coefficient (density factor) is 7 165 / 30 000 = 0.24. Thus, based on the example data (table 1), the other coefficient of the building area at the price of 28,000 rubles/sq. m, the land area of 30.000 sq. m may be understood as not appropriate to the price set. The same result could be obtained by applying a formula similar to formula (1). In this section, sequential accounting of conditions (first prices $V = v$, then land area $SP = sp$) is used to show that the coefficient of building density is not a constant within one price group or even for one specific price, and has a power-law dependence on the land area. Left part of *Figure 4* shows lines of modal, median and average values of the area of improvements, depending on the area of the land plot for the case when

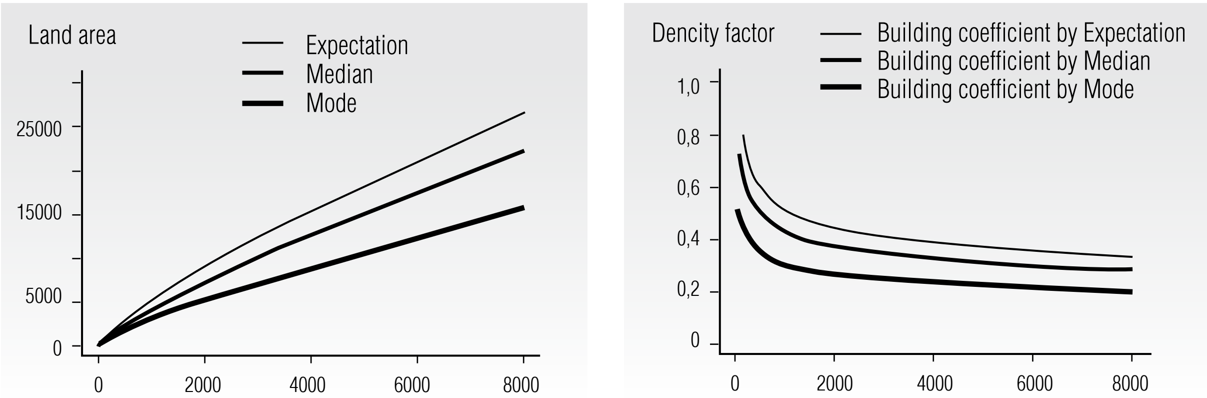

Fig. 4. Values of the area of improvements and building coefficients, depending on the area of the land plot

the offer price is equal to 28.000 rubles/sq. m of the area of existing improvements. The right figure shows lines of building coefficients for corresponding estimates of the area of improvements. *Figure 4* shows that at a given price (price group), the coefficient of development with acceptable accuracy for evaluation purposes can be estimated as a constant only if the land area is large enough. For plots with a small area, the development coefficient cannot be estimated as a constant and must be studied individually taking into account the area of the plot.

4. A note regarding the form of the joint logarithmically normal distribution of the vector (*V*, *SB*, *SP*)

Multidimensional distribution of vector components $(W, Y, Z) = (\ln(V), \ln(SB), \ln(SP))$ it is normal and has symmetry. The scattering clouds of empirical observations will take the form of three-dimensional ellipsoids. The density maximum point has coordinates equal to the mean values of the components W, Y, Z. The distribution of the components of the vector (V, SB, SP) is asymmetric, the density maximum point is not the center of symmetry and can be calculated (see Appendix) using the following formulas:

$$V_{max} = \exp(\mu_W - \sigma_W^2 - \rho_{WY}\sigma_W\sigma_Y - \rho_{WZ}\sigma_W\sigma_Z),$$
$$SB_{max} = \exp(\mu_Y - \sigma_Y^2 - \rho_{YW}\sigma_W\sigma_Y - \rho_{YZ}\sigma_Y\sigma_Z),$$
$$SP_{max} = \exp(\mu_Z - \sigma_Z^2 - \rho_{ZY}\sigma_Z\sigma_Y - \rho_{ZW}\sigma_W\sigma_Z).$$

Figure 5 shows the following: the scattering of source data and the scattering of logarithms of source data, the point of maximum density in space (V, SB, SP) with coordinates $V_{max} = 20\,004$ rubles per 1 sq. m, $SB_{max} = 649$ rubles per 1 sq. m, $SP_{max} = 1\,202$ rubles per 1 sq. m and the point of maximum density in logarithmic space $(W, Y, Z) = (\ln(V), \ln(SB), \ln(SP))$ with coordinates $\mu W = 10.30$; $\mu Y = 8.45$; $\mu Z = 9.35$. Black marks the points of maximum density: on the left – in the space (V, SB, SP), on the right – in the space $(W, Y, Z) = (\ln(V), \ln(SB), \ln(SP))$.

Figure 6 shows the result of 1000 generations three-dimensional random vectors with the same parameters.

It is obvious that (see Appendix) the maximum density point of a multidimensional vector (mode) whose logarithms are normally distributed together is unique. All other density values correspond to the sets described in the logarithmic dimension by hollow three-dimensional ellipsoids, and in the original coordinates, the sets corresponding to a single density value represent the result of distortion (stretch-

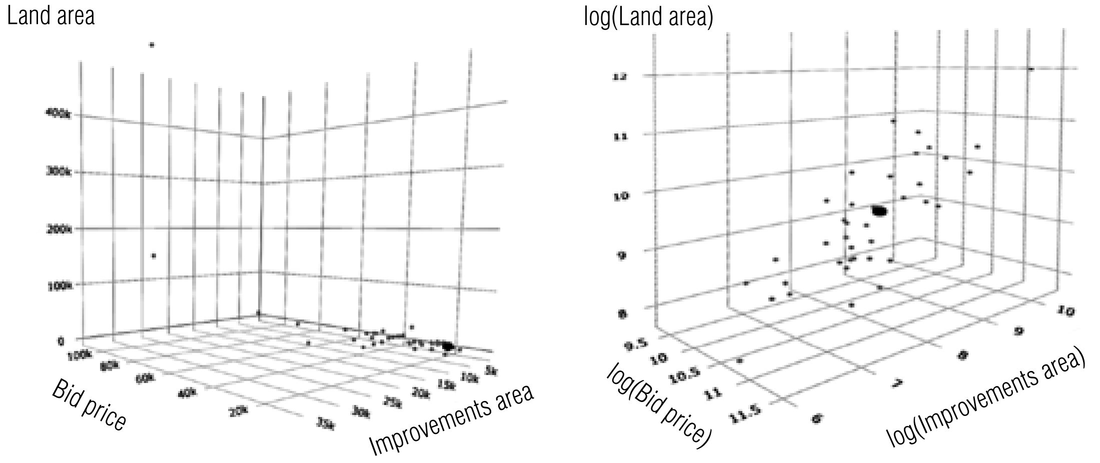

Fig. 5. The scattering of original data (left), the scattering of the logarithms of the original data (right)

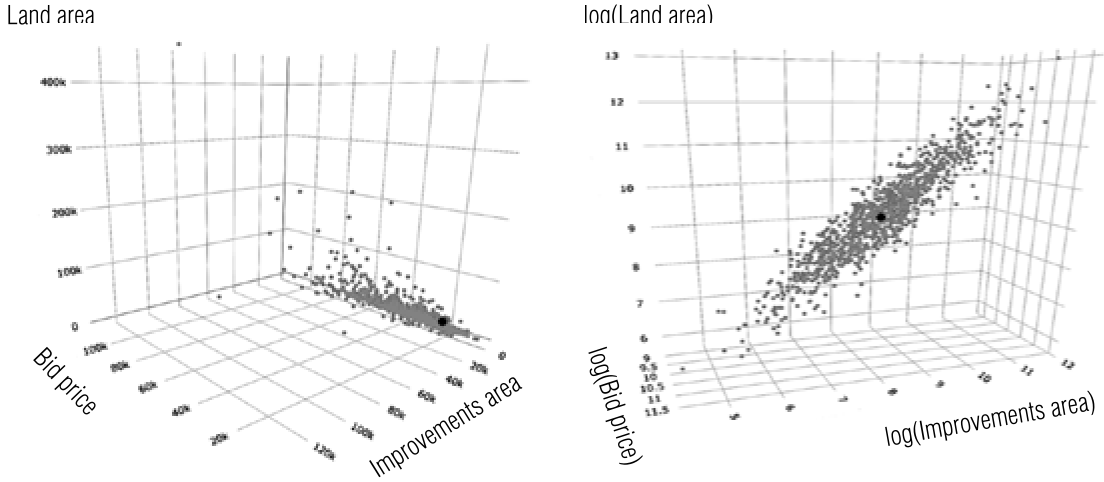

Fig. 6. Result of 1000 generations three–dimensional random vectors

ing) of the hollow ellipsoids during the inverse exponential transformation of space. Thus, it is the modal assessment of the market value that should lead to a correct result that does not create conflict situations. All other (non-modal) market value estimates are potentially a source of constant disputes about the market value of the object of valuation.

Conclusion

Considering the prices of objects of comparison and the values of price-forming factors as multidimensional random variables opens up new opportunities in the assessment of real estate. It often turns out that empirical observations of prices and their corresponding values of price-forming factors are well approximated by the logarithmically normal distribution law, including the multidimensional one, which allows us to derive calculation formulas for various estimation problems. The bulkiness of these formulas is compensated by the capabilities of modern applied statistical packages (in particular, R). In addition, the ability to reduce calculations to a well-studied multidimensional normal law by logarithm of components makes this choice of model distribution preferable.

Conditional price distributions with known values of price-forming factors make it possible to estimate the market value in full accordance with its definition fixed in Russian legislation and foreign standards, as the maximum point of the density of the conditional price distribution.

Conditional distributions of price-forming factors at a given offer price allow us to assess the adequacy of the offer price in terms of a set of price-forming factors.

It is hardly to be expected that practicing appraisers are prepared to apply the formulas given in this article in their daily practice of valuation and business analysis. This is not required. Once written and debugged, the script (in the statistical package R or in other specialized packages) will allow is to easily solve such problems practically in real time. It should be recognized that in the period of digital transformation of the economy and business analysis, it is time for the valuation business to move to advanced statistical packages and automatic data processing. ■

Appendix

Statement. The absolute maximum (mode) density of a random logarithmically normal vector $\bar{\boldsymbol{x}}$ is reached at the point with coordinates $\exp(\bar{\mu} - \Sigma \times \mathbf{1})$, where $\bar{\mu}$ is the vector of mathematical expectations of the logarithms of the component, Σ is the covariance matrix of the logarithms of the component, and $\mathbf{1}$ is a vector consisting of units.

Proof. Consider the density of a multidimensional normal distribution of a centered random vector $\bar{\boldsymbol{y}}$:

$$f(\bar{\boldsymbol{y}}) = \frac{1}{(2\pi)^{\frac{n}{2}} \sqrt[2]{\det \Sigma}} \exp\left(-\frac{1}{2}\left(\Sigma^{-1}\bar{\boldsymbol{y}}, \bar{\boldsymbol{y}}\right)\right).$$

When replacing variables $\bar{\boldsymbol{y}} = \ln(\bar{\boldsymbol{x}})$, the density of the lognormal distribution of the random vector $\bar{\boldsymbol{x}}$:

$$f(\bar{\boldsymbol{x}}) = \frac{1}{(2\pi)^{\frac{n}{2}} \sqrt[2]{\det \Sigma}} \times \frac{1}{\prod_{i=1}^{n} x_i} \times$$
$$\times \exp\left(-\frac{1}{2}\left(\Sigma^{-1}\overline{\ln(\boldsymbol{x})}, \overline{\ln(\boldsymbol{x})}\right)\right),$$

where $\frac{1}{\prod_{i=1}^{n} x_i}$ – coordinate transformation Jacobian,

Σ – covariance matrix, $\overline{\ln(\boldsymbol{x})}$ – centered random vector. At the point of absolute maximum density of the joint logarithmically normal distribution, the derivative in any direction must be zero, which means that all partial derivatives are equal to zero.

$$\frac{\partial f(\bar{\boldsymbol{x}})}{\partial \boldsymbol{x}_j} = \frac{1}{(2\pi)^{\frac{n}{2}} \sqrt[2]{\det \Sigma}} \times \frac{1}{\prod_{i=1}^{n} x_i} \left(-\frac{1}{x_j}\right) \times$$
$$\times \exp\left(-\frac{1}{2}\left(\Sigma^{-1}\overline{\ln(\boldsymbol{x})}, \overline{\ln(\boldsymbol{x})}\right)\right) + \frac{1}{2\pi)^{\frac{n}{2}} \sqrt[2]{\det \Sigma}} \times$$
$$\times \frac{1}{\prod_{i=1}^{n} x_i} \exp\left(-\frac{1}{2}\left(\Sigma^{-1}\overline{\ln(\boldsymbol{x})}\ \ \overline{\ln(\boldsymbol{x})}\right)\right) \times$$
$$\times \left(-\frac{2}{2}\Sigma^{-1}\overline{\ln(\boldsymbol{x})}\right) \times \frac{1}{x_j} = 0$$

After removing the common multipliers from brackets, the condition remains:

$$-\mathbf{1} + \left(-\Sigma^{-1}\overline{\ln(\boldsymbol{x})}\right) = 0 \text{ or } \left(-\Sigma^{-1}\overline{\ln(\boldsymbol{x})}\right) = \mathbf{1},$$

where $\mathbf{1}$, $\mathbf{-1}$ – vectors with dimension n, consisting from units/negative units.

Let multiply the last equality on the left by Σ:

$$\Sigma\Sigma^{-1}\overline{\ln(\boldsymbol{x})} = -\Sigma \times \mathbf{1},$$
$$\boldsymbol{E} \times \overline{\ln(\boldsymbol{x})} = -\Sigma \times \mathbf{1}.$$

Here $\boldsymbol{E}$ is a unit matrix (on the main diagonal – units, the other elements are zero), $\mathbf{1}$ – a vector consisting of units. I.e., the values of the vector $\overline{\ln(\boldsymbol{x})}$ in which all partial derivatives are zero, are equal to the line-by-line sums of the covariance matrix, taken with the reverse sign.

It remains to remember that $\bar{\boldsymbol{y}} = \ln(\bar{\boldsymbol{x}})$ is a centered random vector. If the expectation vector $\bar{\mu}$ contains non-zero values, then the final solution is:

$$\overline{\ln(\boldsymbol{x})} = \bar{\mu} - \Sigma \times \mathbf{1} \text{ or } \bar{\boldsymbol{x}} = \exp(\bar{\mu} - \Sigma \times \mathbf{1}).$$

Taking into account negative definiteness of the quadratic form composed of second partial derivatives in point $\bar{\boldsymbol{x}} = \exp(\bar{\mu} - \Sigma \times \mathbf{1})$ (the author omits this bulky record since the result is obvious), the point $\bar{\boldsymbol{x}} = \exp(\bar{\mu} - \Sigma \times \mathbf{1})$ is a point of maximum density of lognormal random vector $\bar{\boldsymbol{x}}$.

The statement is proven.

References

1. Rusakov O.V., Laskin M.B., Jaksumbaeva O.I. (2016) Pricing in the real estate market as a stochastic limit. Log Normal approximation. *International Journal of Mathematical Models and Methods in Applied Sciences*, no 10, pp. 229–236.
2. Aitchinson J., Brown J.A.C. (1963) *The Lognormal distribution with special references to its uses in economics*. Cambridge: University Press.
3. Ohnishi T., Mizuno T., Shimizu C., Watanabe T. (2011) *On the evolution of the house price distribution*. Columbia Business School. Center of Japanese Economy and Business. Working Paper Series, no 296.
4. Anselin L., Lozano-Gracia N. (2008) Errors in variables and spatial effects in hedonic house price models of ambient air quality. *Empirical Economics*, vol. 34, no 1, pp. 5–34. DOI: 10.1007/s00181-007-0152-3.
5. Benson E.D., Hansen J.L., Schwartz Jr. A.L., Smersh G.T. (1998) Pricing residential amenities: The value of a view. *Journal of Real Estate Finance and Economics*, vol. 16, no 1, pp. 55–73. DOI: 10.1023/A:1007785315925.
6. Debrezion G., Pels E., Rietveld P. (2011) The impact of rail transport on real estate prices: an empirical analysis of the Dutch housing market. *Urban Studies*, vol. 48, no 5, pp. 997–1015. DOI: 10.1177/0042098010371395.
7. Jim C.Y., Chen W.Y. (2006) Impacts of urban environmental elements on residential housing prices in Guangzhou (China). *Landscape and Urban Planning*, vol. 78, no 4, pp. 422–434. DOI: 10.1016/j.landurbplan.2005.12.003.
8. Rivas R., Patil D., Hristidis V., Barr J.R., Srinivasan N. (2019) The impact of colleges and hospitals to local real estate markets. *Journal of Big Data*, vol. 6, no 1, article no 7 (2019). DOI: 10.1186/s40537-019-0174-7.
9. Wena H., Zhanga Y., Zhang L. (2015) Assessing amenity effects of urban landscapes on housing price in Hangzhou, China. *Urban Forestry & Urban Greening*, no 14, pp. 1017–1026. DOI: 10.1016/j.ufug.2015.09.013.
10. Saint Petersburg State Budget Department "Cadastral Valuation City Department" (2018) *Report on determining the cadastral value of real estate objects on the territory of Saint Petersburg*, no 1. Available at: http://www.ko.spb.ru/interim-reports/ (accessed 05 June 2019).
11. Peterson S., Flanagan A.B. (2009) Neural network hedonic pricing models in mass real estate appraisal. *Journal of Real Estate Research*, vol. 31, no 2, pp. 147–164.
12. Rafiei M.H., Adeli H. (2018) Novel machine-learning model for estimating construction costs considering economic variables and indexes. *Journal of Construction Engineering and Management*, vol. 144, no 12, article no 04018106. DOI: 10.1061/(asce)co.1943-7862.0001570.
13. Antipov E.A., Pokryshevskaya E.B. (2012) Mass appraisal of residential apartments: An application of Random forest for valuation and a CART-based approach for model diagnostics. *Expert Systems with Applications*, no 39, pp. 1772–1778. DOI: 10.1016/j.eswa.2011.08.077.
14. Kontrimas V., Verikas A. (2011) The mass appraisal of the real estate by computational intelligence. *Applied Soft Computing*, no 11, pp. 443–448. DOI: 10.1016/j.asoc.2009.12.003.
15. Park B., Baem J.K. (2015) Using machine learning algorithms for housing price prediction: The case of Fairfax County, Virginia housing data. *Expert Systems with Applications*, no 42, pp. 2928–2934. DOI: 10.1016/j.eswa.2014.11.040.
16. Case K.E., Shiller R.J. (1987) Prices of single-family Homes since 1970: New indexes for four cities. *New England Economic Review*, September, pp. 45–56. DOI: 10.3386/w2393.

17. Englund P., Quigley J.M., Redfearn C.L. (1999) The choice of methodology for computing housing price indexes: comparison of temporal aggregation and sample definition. *Journal of Real Estate Finance and Economics*, vol. 19, no 2, pp. 91–112. DOI: 10.1023/A:1007846404582.

18. Epley D. (2016) Assumptions and restrictions on the use of repeat sales to estimate residential price appreciation. *Journal of Real Estate Literature*, vol. 24, no 2, pp. 275–286. DOI: 10.5555/0927-7544.24.2.275.

19. Malpezzi S. (2002) Hedonic pricing models: A selective and applied review. *Housing economics and public policy: Essays in honor of Duncan Maclennan* (T. O'Sullivan, K. Gibb, eds.). Oxford, UK: Blackwell Science, pp. 67–89. DOI: 10.1002/9780470690680.ch5.

20. Case B., Quigley J.M. (1991) The dynamics of real estate prices. *Review of Economics and Statistics*, vol. 73, no 1, pp. 50–58.

21. Englund P., Quigley J.M., Redfearn C.L. (1998) Improved price indexes for real estate: Measuring the course of Swedish housing prices. *Journal of Urban Economics*, vol. 44, no 2, pp. 171–196.

22. Jones C. (2010) House price measurement: The hybrid hedonic repeat-sales method. *Economic Record*, vol. 86, no 272, pp. 95–97. DOI: 10.1111/j.1475-4932.2009.00596.x.

23. Wang F., Zheng X. (2018) The comparison of the hedonic, repeat sales, and hybrid models: Evidence from the Chinese paintings. *Cogent Economics & Finance*, no 6, pp. 1–19. DOI: 10.1080/23322039.2018. 1443372.

24. Brunnermeier M.K. (2009) Bubbles. *The new Palgrave dictionary of Economics* (L.E. Blume, S.N. Durlauf, eds.). New York: Palgrave Macmillan.

25. Fabozzi F.J., Xiao K. (2019) The timeline estimation of bubbles: The case of real estate. *Real Estate Economics*, vol. 47, no 2, pp. 564–594. DOI: 10.1111/1540-6229.12246.

26. Fernandez-Kranz D., Hon M.T. (2006) A cross-section analysis of the income elasticity of housing demand in Spain: Is there a real estate bubble? *Journal of Real Estate Finance and Economics*, vol. 32, no 4, pp. 449–470. DOI: 10.1007/s11146-006-6962-9.

27. Phillips P.C.B., Shi S.-P., Yu J. (2015) Testing for multiple bubbles: Historical episodes of exuberance. *International Economic Review*, vol. 56, no 4, pp. 1043–1078. DOI: 10.1111/iere.12132.

28. Phillips P.C.B., Shi S.-P., Yu J. (2015) Testing for multiple bubbles: Limit theory of real time detectors. *International Economic Review*, vol. 56, no 4, pp. 1079–1134. DOI: 10.1111/iere.12131.

29. The Federal Law of 29.07.1998 No 135-FZ (edition of 29.07.2017) "*About assessment activity in the Russian Federation*". Available at: http://www.consultant.ru/document/cons_doc_LAW_19586/ (accessed 14 March 2020).

30. Slytsky A.A., Slytskaya I.A. (2020) *The modified extraction method and the generalized modified method of allocation. Use for analyzing the market segment that the item is being evaluated belongs to*. Available at: http://tmpo.su/sluckij-a-a-sluckaya-i-a-mmv-i-ommv-primenenie-dlya-analiza-rynka-3/ (accessed 14 March 2020).

31. Laskin M.B. (2014) Logarithmically normal distribution of prices and market value in the real estate market. *Saint Petersburg State Technological Institute Review*, no 25 (51), pp.102–106.

32. Laskin M.B. (2017) Market value adjustment for the pricing factor "square". *Property Relations in the Russian Federation*, no 8 (191), pp. 86–99.

About the author

Michael B. Laskin

Cand. Sci. (Phys.-Math.), Associate Professor;

Senior Researcher, St. Petersburg Institute for Informatics and Automation, Russian Academy of Sciences (SPIIRAS),
39, 14 Line, St. Petersburg 199178, Russia

E-mail: laskinmb@yahoo.com

ORCID: 0000-0002-0143-4164