Peculiarities of applying methods based on decision trees in the problems of real estate valuation

Mikhail B. Laskin [a] iD
E-mail: laskinmb@yahoo.com

Lyudmila V. Gadasina [b] iD
E-mail: l.gadasina@spbu.ru

[a] St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences (SPIIRAS)
   Address: 39, 14 line, Vasilevskiy Island, St. Petersburg 199178, Russia

[b] St. Petersburg State University, Center for econometrics and business analytics (CEBA)
   Address: 7/9, Universitetskaya emb., Saint-Petersburg 199034, Russia

Abstract

The increasing flow of available market information, the development of methods of machine learning, artificial intelligence and the limited capabilities of traditional methods of real estate valuation are leading to a significant increase of researchers' interest in real estate valuation by applying methods based on decision trees. At the same time, the distribution of real estate prices is well approximated by a lognormal distribution. Therefore, traditional methods overestimate the predicted values in the region below the average of the available data set and underestimate the predicted values in the region above the average. This article shows the reasons for these features and proposes an adaptive random forest algorithm which corrects the results of the basic algorithm prediction by revising the bias of these predicted values. The results were tested on the real estate offer prices in St. Petersburg.

Keywords: decision trees, random forest, real estate market, price-forming factors, market value apprising

Introduction

The number of publications devoted to non-traditional methods of real estate valuation, focused on large samples of data, and, in particular, methods of machine learning, has significantly increased recently.

The interest of researchers in this topic is understandable: the changed information environment and the wide choice of specialized application packages allow us to consider what was not previously available from evaluation methods; see for example [1−6]. The method of hedonic pricing, linear regression models, logarithmic or partial-logarithmic dependence are considered in [7−10]. Data mining techniques, such as neural networks [11−15] and Support Vector Machines [16] are proposed, the results of such methods as decision trees, naive Bayesian classifier and ensemble algorithm Ada-Boost are compared [17]. This paper discusses the prediction biases that arise from the application of decision tree based methods in real estate valuation and proposes an algorithm to correct for these biases. Interest in this group of methods is confirmed by [18−21]. Researchers are turning to machine learning methods, in particular decision tree based methods, under conditions where there is an extensive set of input data and there are no a priori assumptions about the form of the function $F(\cdot)$, describing the dependence $V = F(x_1, x_2, ..., x_n)$ between the output or dependent variable V, which is usually the price, and predictors $x_1, x_2, ..., x_n$, which are the price-forming factors.

The advantage of decision tree based methods is that knowledge of the type of function $F(\cdot)$ is not required. However, this does not mean that the specific properties of real estate price distributions do not affect the results of such algorithms. The method of constructing a single decision tree consists in successive decomposition of the entire predictor domain into subsets of smaller size. Each element of such a subset is assigned a value of the arithmetic average of the values of the dependent variable on such a subset. This is an iterative procedure, known as recursive decomposition:

1. At each step, the datasets are divided into subsets.

2. Each of the subsets obtained on the previous step is, in turn, divided. In general, there will be a decomposition of the space of independent variables (predictors) $x_1, x_2, ..., x_n$ into some number (for example, m) of non-intersecting domains $R_1, R_2, ..., R_m$.

3. Outcome variable V values predicted as the same value for all observations in the domain R_j, $j = 1, 2, ..., m$. This value is equal to the average of all responses that fall into R_j. At each step, dividing provides a minimum of root mean square deviations of RSS (residual sum of squares):

$$RSS = \sum_{j=1}^{m} \sum_{i \in R_j} (V_i - \hat{V}_{R_j})^2,$$

where $\hat{V}_{R_j}$ is the response average value of the training observations from the set R_j.

From the computational point of view, it is not feasible to consider all combinations of decompositions to the greatest possible depth. Therefore, the basic principle of "greedy" algorithms is used: the optimal division (in terms of RSS minimum) is determined only at the current step. The depth of decomposition can be as large as the volume of data allows. Possible rules for stopping: by the number of steps, by the number of elements in subsets (tree sheets), by reaching a pre-determined improvement in the result at the next step.

The model of interest is division with resulting sets $R_1, R_2, ..., R_m$ containing a "sufficiently large number of elements" and the response values standard deviation being not too large

from its average value within the predetermined accuracy in each set.

1. Methodology

In the problems of real estate valuation, the application of the decision tree method generates additional opportunity. It allows you to split a set of objects into subsets with less variance, with more homogeneous objects within each of them. They can be studied separately. It should be noted that decision tree based algorithms can operate with both real and factor predictors.

Thus, the advantages of using decision trees are:

1. Clear model interpretation.

2. Such an algorithm reflects people's decision-making process.

3. For one decision tree there is a visual graphical representation.

4. Decision trees easily operate with factor and rank variables.

The disadvantage of such algorithms is low prediction accuracy; dispersion within each decomposition set is not small enough − on leaves). This drawback can be eliminated by applying ensemble decision tree based methods, e.g. random forest, gradient or stochastic boosting. Such algorithms do not produce interpretable results, but they allow you to analyze the importance of predictors in predicting the response and allow you to operate with factor variables. This paper considers the random forest method.

Decision tree based algorithms should be applied for real estate evaluation taking into account the peculiarities of the target variable V, since there are reasons to assume that it is a lognormally random variable. Apparently, Aichinson and Brown [22] first pointed out this fact, and later this observation was confirmed in research such as [4]. Research [23, 24] gives a theoretical substantiation of the reasons for this kind of distribution appearance.

A random variable V is called log-normally distributed with parameters μ and σ, if the random variable $\ln(V)$ is normally distributed with the same parameters. In this case mathematical expectation $E(V) = e^{\mu + \sigma^2/2}$, median $Median(V) = e^{\mu}$, and mode $Mode(V) = e^{\mu - \sigma^2}$. The fraction of empirical distribution values to the left of $E(V)$ (less than the mathematical expectation) can be estimated as

$$\frac{1}{2} + \Phi\left(\frac{\sigma}{2}\right),$$

where $\Phi(\cdot)$ is a Laplace function. The fraction of such values depends on the standard deviation and increases with its growth.

Thus, if we consider the hypothesis that the results of decision tree based algorithms prediction are also log-normally distributed, and moreover, form a joint lognormal distribution with the observed values in the sense of a two-dimensional normal distribution of logarithms, then it becomes clear that decision tree based methods are better applied not to the prices of real estate objects, but to their logarithms. Since at each step the standard deviations of RSS from the mean values in the subsets R_1, R_2, ..., R_m, are minimized, within which the principle of lognormal distribution of the dependent variable V, can be preserved, then in case of ensemble algorithms based on decision trees the prediction of results on the test set will be quite accurate only in the domain of values close to the average values of the responses. In the area of values below the average, the predictions of such algorithms will be overestimated, above the average − underestimated, increasing as we approach the boundaries of empirical distributions. An appropriate diagram showing the relationship between the true and predicted

values will, with acceptable prediction accuracy, show scattering clouds extending along some straight line, somewhat displaced relative to the bisector of the first coordinate angle by rotating around a point, with the coordinates of the average response value and the average prediction value. In this case most of the results will be overestimated, because in a lognormal distribution most of the possible values are to the left of the average (less than the average). In this case, for $\ln(V)$, which has a normal distribution, the areas of over- and under-prediction will be about the same by the number of elements. In the described approach, the prediction results remain displaced relative to the true values. This is observed, for example, in [20].

In this paper, we propose to apply an adaptive method based on the correction of the prediction results of the ensemble algorithm random forest. Adaptation consists of the following procedure. The set of initial data is divided into three parts: training, validation and test. The training procedure (selection of parameters) of the random forest algorithm is performed on the first set. Next, we analyze the dependence of the predicted values for the validation set on their true values. The prediction is corrected by rotating the scattering cloud in coordinates (response value is a prediction) by some angle, which removes the displacement from the bisector of the first coordinate angle and recalculates the predicted values. At the third step, the quality of the predictions of the random forest algorithm, taking into account the correction, is checked on the test set.

It should be noted that in any case the estimate of market value will be determined not as the most likely price, but as the average price (if the algorithm was applied to the original prices) or the geometric mean price (if the algorithm was applied to the logarithms of prices).

2. Applying the technique to real market data

Approbation of the procedure we have described was carried out on the following example. Consider the prices for secondary residential real estate in the mass-market sector in St. Petersburg in February 2017 taken from an open source (the advertising publication Real Estate Bulletin No. 1765, February 2017, not indexed in scientometric databases), the total number of records in the dataset after removing incorrect ads was 4294. The random forest method will be used for the predictive model. The dependent (target) variable is the price per square meter of secondary residential real estate in the mass-market sector in St. Petersburg in February 2017 or its logarithm. The predictors are:

♦ the number of rooms in the apartment — a quantitative variable

♦ administrative area of the location — a factor variable

♦ floor — a factor variable

♦ number of floors in the house — a factor variable

♦ living space — a quantitative variable

♦ total area — a quantitative variable

♦ subway accessibility — a binary variable

♦ house type — a factor variable

♦ number of bathrooms, their type — a factor variable.

The calculations were carried out in the open source software R. First of all, let us pay attention to the asymmetric distribution of prices (*Fig. 1*).

The verification of the null hypothesis about following the empirical distribution of prices for one m^2 logarithmically normal distribution is associated with certain difficulties. The sample size is 4294. As was rightly noted in [25], most of the common and fre-
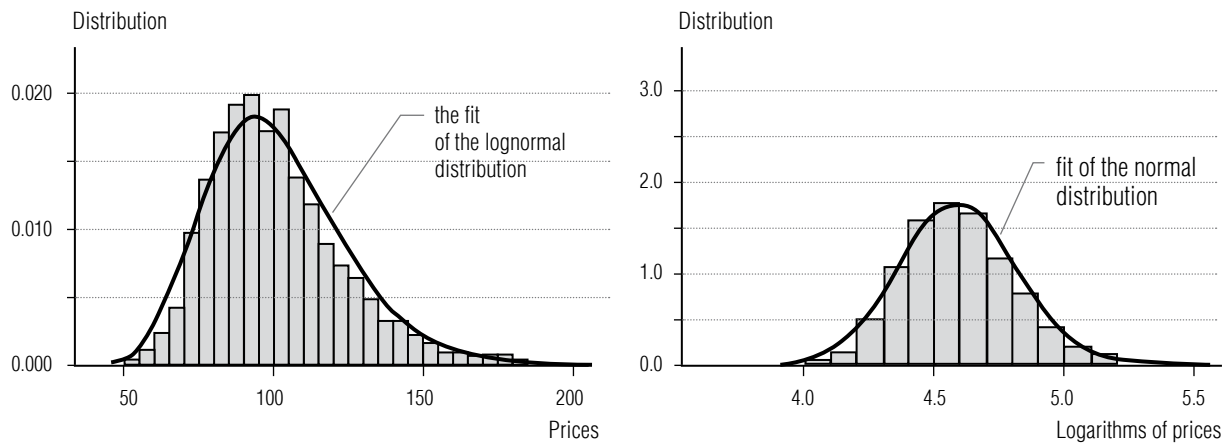
Distribution

the fit
of the lognormal
distribution

Prices

Distribution

fit of the normal
distribution

Logarithms of prices

Fig. 1. Prices distribution (left) and their logarithms (right)
of secondary residential real estate in St. Petersburg in February 2017.

quently used criteria do not work for samples of the order of even one thousand observations, since their statistics significantly depend on the sample size. Therefore, the question arises of finding, in addition to the visual correspondence shown in *Fig. 1*, additional arguments in favor of a particular type of distribution. In this context, we note the work [26], in which it is proposed to study the relations between the coefficient of asymmetry and the kurtosis of the observed sample in order to advance the null hypothesis about a particular type of distribution. In this paper, the method proposed in [27] was used to test the corresponding hypotheses. The obtained p-value values provide grounds to assume that the observed sample follows a logarithmically normal law; therefore there are grounds to solve the predictive problem in logarithms.

Let us consider sequentially what predictions are obtained by one decision tree, by the random forest algorithm, and then carry out a procedure for correcting the resulting prediction. For this purpose a training set consisting of 2000 records is formed from the initial sample (volume 4294 records) by random selection, as well as a validation set consisting of 1000 records, and the remaining 1294 records

form a test set where the model quality is evaluated.

Trained on a random sample of 2000 records, the tree model gives a price prediction diagram (decision tree cut to 11 terminal nodes), shown in the *Fig. 2*.

prices predicted by a single decision tree

bisector of the first
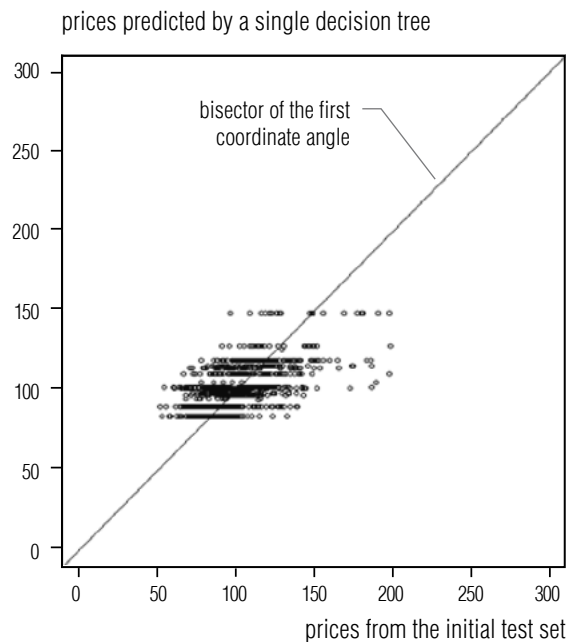coordinate angle

prices from the initial test set

Fig. 2. Prediction diagram on the test set
by a single decision tree.

Such decomposition in the evaluation tasks is not meaningless, because it allows us to form groups with a different set of pricing factors in which a group-specific average value and the standard deviation of the observed values from the group average are expected. Predictors that had a significant influence on the formation of the tree were:

♦ administrative district;

♦ building type;

♦ number of floors in the building;

♦ the number and type of bathrooms;

♦ total area;

♦ living area;

♦ metro accessibility;

♦ floor.

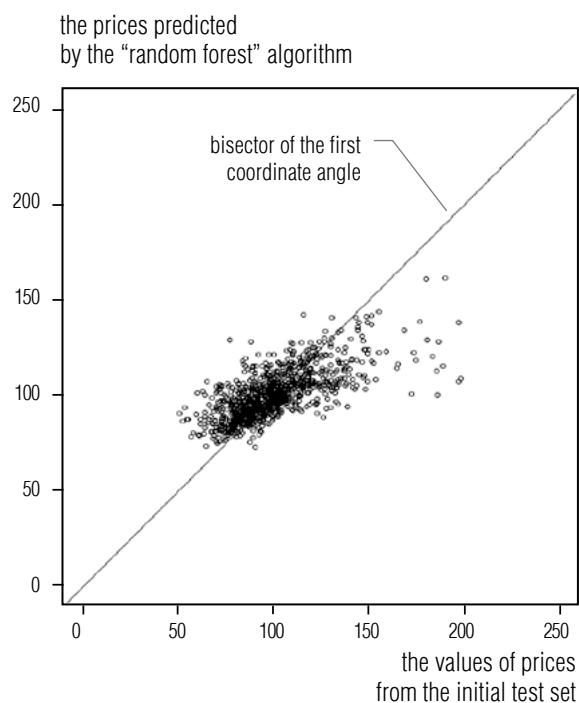The quality of the predictions shown in *Fig. 2* is unsatisfactory − too large ranges of

values are predicted to have the same price. (The ideal would be the predictions in *Fig. 2* near the bisector of the first coordinate angle). An effective way to deal with the same predictions is the random forest algorithm. This algirothm builds a large number of trees and averages the results for each object. *Figure 3* shows the result of the random forest algorithm for predicting the price per square meter of secondary residential real estate for the test set. Each tree in the algorithm was built based on 4 predictors, the total number of trees − 200.

A similar figure, with a characteristic displacement of the predictions, can be seen in the article [20] on algorithms based on decision trees when analyzing the real estate market in Ankara. In *Fig. 3* we can see the scatter of the predictions, characteristic of the joint lognormal distribution, that increases with rising price, and the fact that most of the predictions (above the bisector) appear to be overestimated, which is to be expected, given the asymmetric distribution of prices in the original set. Also note that the predictions thus obtained are predictions of average values (mathematical expectations in subsets), and not the market value as the most likely price. The estimate of market value is, in fact, somewhat lower.

Let's apply the adaptive random forest method to the logarithms of prices.

Figure 4 shows the result of the random forest algorithm for predicting the logarithm of the price per square meter of secondary residential property (the number of trees − 200, a random selection of 4 predictors on each tree, a training random sample of 2000 records).

Figure 4 shows that the areas of over- and under-prediction are approximately the same, but the axis of the scattering cloud has a characteristic trend that does not coincide with the bisector of the first coordinate angle. Note that the predictions obtained by potentiating
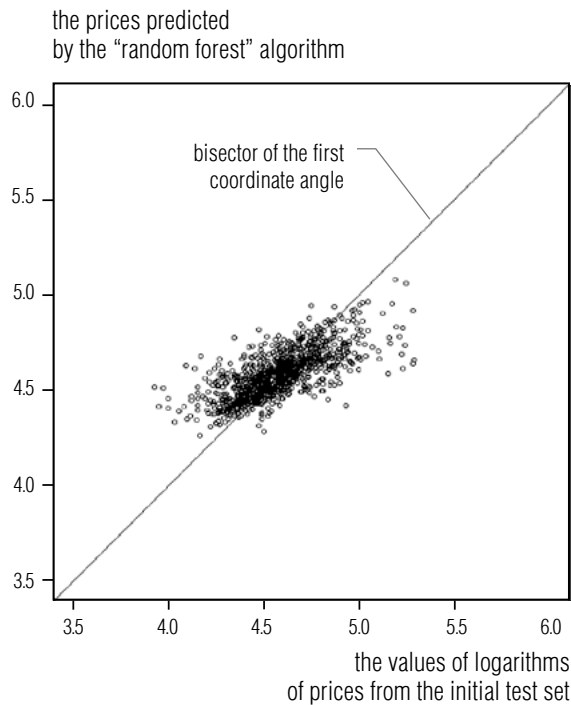

Fig. 3. Diagram of predictions
on the test set by the "random forest" algorithm.

the prices predicted
by the "random forest" algorithm

Fig. 4. Diagram of predictions
on the test set by the "random forest" algorithm.

the results are predictions of the median values (geometric averages in the subsets R_1, R_2, ..., R_{21}), and not the market value as the most likely price. The estimate of market value in this case is somewhat lower.

The results of the predictions of the random forest algorithm can be corrected by simple transformations of the results. Using the linear regression method, we determine the linear trend of the scattering cloud shown in *Fig. 4*. The result is shown in the *Fig. 5*.

In this example, the trend line equation is

$$\ln(\hat{V}) = 0.40791 \cdot \ln(V) + 2.71920, \qquad (1)$$

where $\ln(V)$ is the observed value of the logarithm of the price;

$\ln(\hat{V})$ is the predicted value of the logarithm of the price.

The statistical characteristics of the obtained trend line: *p*-value of Student's test for model

coefficients and Fisher's criterion for the created model is machine zero. The standard error is 0.086, i.e., the parameter spread with probability 0.99 is in the interval ~ +/−26%. The relatively low value of $R^2 = 0.5053$ does not spoil the situation, because there is no strictly linear relationship between $\ln(\hat{V})$ and $\ln(V)$ in this case, our expectations are related to the multivariate normal distribution of $\text{lln}(\hat{V})$ and $\ln(V)$, for which the linear trend coincides with the major axis of the ellipse of scattering (see more about multidimensional logarithmic normal distribution [27]). Equation (1) corresponds to the line shown in bold black in *Fig. 5*. For the scattering cloud shown in *Fig. 5*, the value of the standard deviation of the observed values from the predicted values is 0.168. It remains for us to correct the predictions shown in *Figs. 3*, *4*, and *5*. For this purpose, all values are centered (horizontal and vertical averages are subtracted). The scatterplot is rotated counterclockwise in the new
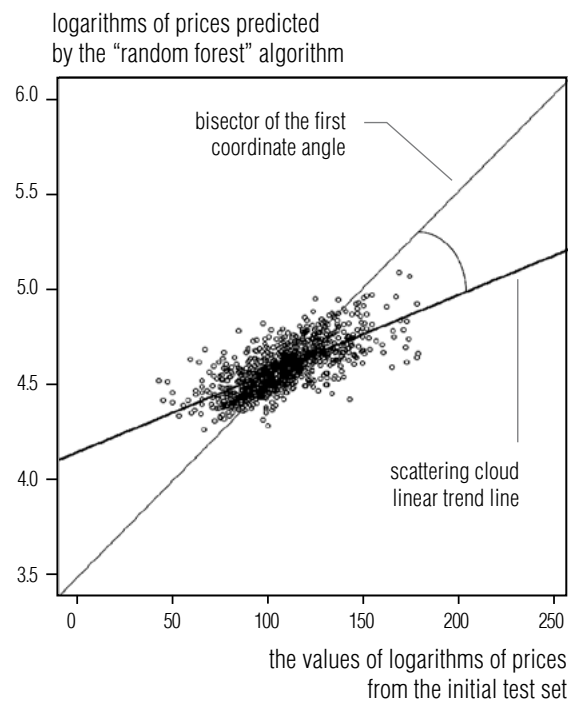
logarithms of prices predicted
by the "random forest" algorithm

Fig. 5. Diagram of predictions
on the test set by the "random forest" algorithm.

corrected predictions
of the centered logarithms of prices

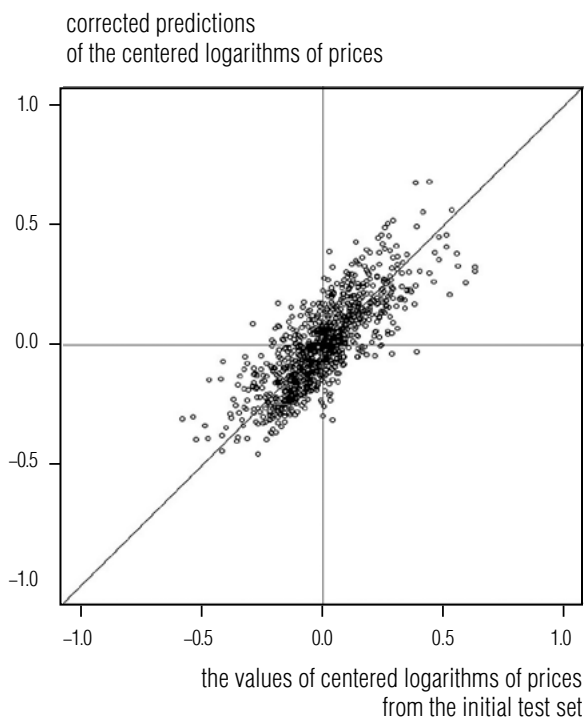the values of centered logarithms of prices
from the initial test set

Fig. 6. Corrected predictions on the test set
by the "random forest" algorithm.

coordinate system and the angle is the difference between the line given by equation (1) (*Fig. 5*, bold black) and the bisector of the first coordinate angle. Such angle φ is equal to $\frac{\pi}{4} - \text{arctg}(0.40791)$. The result of the rotation is shown in *Fig. 6*.

For the scattering cloud shown in *Fig. 6*, the value of the standard deviation of the observed values from the predicted values is 0.118. Thus, the correction performed gives a better prediction quality.

Now we have two vectors of values: a vector of predictions of centered price logarithms on the test set (we denote it by) and a vector of corrected predictions of centered price logarithms (we denote it by y^*).

In *Fig. 7*, the horizontal axis shows the values of the components of the vector y^*, the vertical shows the values of the vector components y^+.

The set shown in *Fig. 7* has a linear trend of the following form $y^+ = \alpha \cdot y^*$, which is easily determined by using the library function lm of the statistical package R. In this example we get $y^+ = 1.388$. Now we firstly apply sequentially the random forest model already obtained on the training sample and then the prediction correction obtained on the validation set to the test set (1294 records).

Figure 8 shows predictions of centered price logarithms on the verifying set (the random forest model obtained on the training set was applied).

Figure 9 shows corrected predictions of centered price logarithms on the verifying set (using the random forest model obtained on the training set and the correction obtained on the test set).

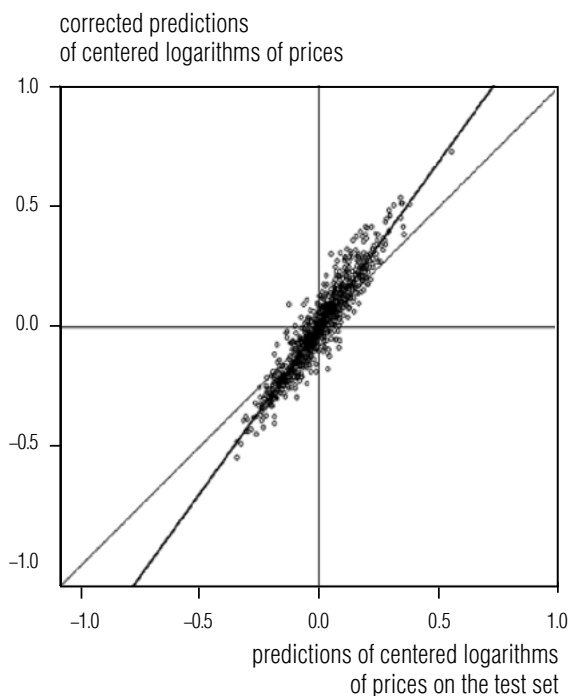Here are the necessary formulas and sequence of operations:

corrected predictions
of centered logarithms of prices

predictions of centered logarithms
of prices on the test set

Fig. 7. Correlation of predicted
and corrected values.

predicted values

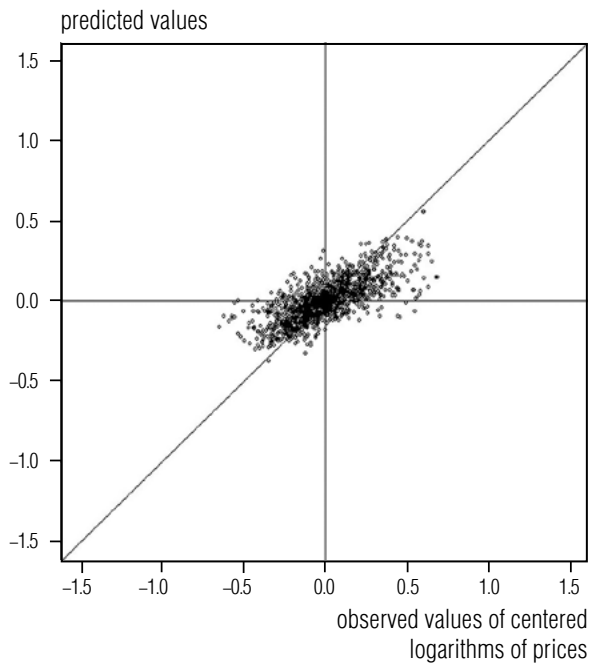*Fig. 8.* The correlation between
the observed values of centered logarithms
of prices and their predicted values
by the random forest algorithm.

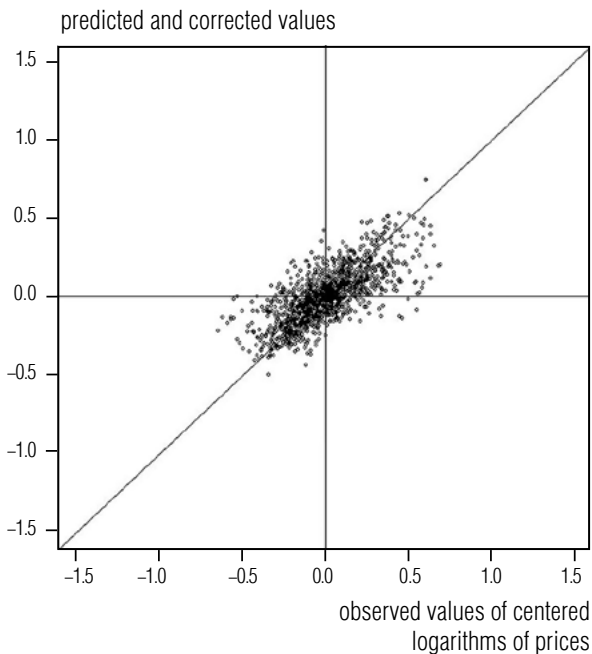predicted and corrected values

Fig. 9. The correlation of observed values
of centered logarithms of prices and their
predicted values by the random forest algorithm
and corrected by the formula
obtained on the test set.

1. Let $\overrightarrow{\ln(V)} = (\ln(V_1), \ln(V_2), \ln(V_3), ..., \ln(V_m))$ are observed values of the logarithms of prices with the average α, $\overrightarrow{\ln(\hat{V})} = (\ln(\hat{V}_1), \ln(\hat{V}_2), \ln(\hat{V}_3), ..., \ln(\hat{V}_m))$ are predicted values on the training set (test, verifying, m — can take different values) values of price logarithms with the average β. Then

$$y_i^* = \ln(\hat{V}_i) - \beta, \ x_i^* = \ln(V_i) - \alpha \ \text{and}$$

$$\begin{pmatrix} x_i^+ \\ y_i^+ \end{pmatrix} = (x_i^+, y_i^+) \cdot \begin{pmatrix} \cos\varphi & \sin\varphi \\ -\sin\varphi & \cos\varphi \end{pmatrix},$$

where rotation angle

$$\varphi = \frac{\pi}{4} - \text{arctg}(0.40791)$$

(under the arctangent sign is the tangent of the inclination angle of the linear trend of the scattering cloud). See *Figs. 4*, 5, 6.

2. We compare the predicted and corrected values of the centered logarithms on the validation sample (*Fig. 7*), determine the angular coefficient α of the trend $y^+ = \alpha \cdot y^*$.

3. The predicted and corrected value of y^+ on the test sample is compared with the observed values of centered logarithms of prices x^* (*Figs. 8, 9*). For the dataset under study, the most significant predictors (when assessing relative importance using the method of features permutation) were the city district, the type of building, the number of floors in the house and the total floor area of the premises. The least important were: the floor location of the premises, the proximity of the subway, the number of rooms.

Conclusion

1. Algorithms based on decision trees predict average values if applied for prices) and average values if applied for logarithms of prices), rather than the most likely values. The market value estimate predicted by

them is somewhat higher than the market value estimate by the most probable value.

2. Due to the lognormal distributions of prices in the original datasets, the predictions constructed using methods based on decision trees require correction. The procedure proposed in this paper allows for such a correction using double cross-validation when a validation subset of the initial data set is selected on which the adaptation of the algorithm is carried out. Then the results are evaluated on the test dataset. The approbation conducted showed the effectiveness of the proposed approach.∎

Acknowledgments

References

1. Gu S., Kelly B., Xiu D. (2019) Empirical asset pricing via machine learning. *Chicago Booth Research Paper No. 18-04, 31st Australasian Finance and Banking Conference 2018, Yale ICF Working Paper No. 2018-09.* https://doi.org/10.2139/ssrn.3159577

2. 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. https://doi.org/10.1016/j.landurbplan.2005.12.003

3. Kok N., Koponen E.-L., Martínez-Barbosa C.A. (2017) Big data in real estate? From manual appraisal to automated valuation. *The Journal of Portfolio Management,* vol. 43, no. 6, pp. 202−211. https://doi.org/10.3905/jpm.2017.43.6.202

4. 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, 296.* https://doi.org/10.7916/D8794CJJ

5. Steurer M., Hill R.J., Pfeifer N. (2021) Metrics for evaluating the performance of machine learning based automated valuation models. *Journal of Property Research,* vol. 38, no. 2, pp. 99−129. https://doi.org/10.1080/09599916.2020.1858937

6. Tay D.P., Ho D.K. (1992) Artificial intelligence and the mass appraisal of residential apartments. *Journal of Property Valuation and Investment*, vol. 10, no. 2, pp. 525−540. https://doi.org/10.1108/14635789210031181

7. 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. https://doi.org/10.1007/s00181-007-0152-3

8. Benson E.D., Hansen J.L., Schwartz Jr. A.L., Smersh G.T. (1998) Pricing residential amenities: The value of a view. *The Journal of Real Estate Finance and Economics*, vol. 16, no. 1, pp. 55−73. https://doi.org/10.1023/A:1007785315925

9. 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. https://doi.org/10.1177/0042098010371395

10. Wena H., Zhanga Y., Zhang L. (2015) Assessing amenity effects of urban landscapes on housing price in Hangzhou, China. *Urban Forestry & Urban Greening*, vol. 14, pp. 1017−1026. https://doi.org/10.1016/j.ufug.2015.09.013

11. Do A.Q., Grudnitski G. (1992) A neural network approach to residential property appraisal. *The Real Estate Appraiser*, vol. 58, no. 3, pp. 38−45.

12. Evans A., James H., Collins A. (1992) *Artificial neural networks: An application to residential valuation in the UK.* University of Portsmouth, Department of Economics.

13. McGreal S., Adair A., McBurney D., Patterson D. (1998) Neural networks: the prediction of residential values. *Journal of Property Valuation and Investment*, vol. 16, no. 1, pp. 57−70. https://doi.org/10.1108/14635789810205128

14. Peterson S., Flanagan A. (2009) Neural network hedonic pricing models in mass real estate appraisal. *Journal of Real Estate Research*, vol. 31, no. 2, pp. 147−164. https://doi.org/10.1080/10835547.2009.12091245

15. Worzala E., Lenk M., Silva A. (1995) An exploration of neural networks and its application to real estate valuation. *Journal of Real Estate Research*, vol. 10, no. 2, pp. 185−201. https://doi.org/10.1080/10835547.1995.12090782

16. Kontrimas V., Verikas A. (2011) The mass appraisal of the real estate by computational intelligence. *Applied Soft Computing*, vol. 11, no. 1, pp. 443−448. https://doi.org/10.1016/j.asoc.2009.12.003

17. Park B., Bae J.K. (2015) Using machine learning algorithms for housing price prediction: The case of Fairfax County, Virginia housing data. *Expert Systems with Applications*, vol. 42, no. 6, pp. 2928−2934. https://doi.org/10.1016/j.eswa.2014.11.040

18. Chen T., Guestrin C. (2016) XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785−794. https://doi.org/10.1145/2939672.2939785

19. Cordoba M., Carranza J.P., Piumetto M., Monzani F., Balzarini M. (2021) A spatially based quantile regression forest model for mapping rural land values. *Journal of Environmental Management*, vol. 289, no. 1, 112509. https://doi.org/10.1016/j.jenvman.2021.112509

20. Yilmazer S., Kocman S. (2020) A mass appraisal assessment study using machine leaning based on multiple regression and random forest. *Land Use Policy*, vol. 99, 104889. https://doi.org/10.1016/j.landusepol.2020.104889

21. Webb G.I. (2000) Multiboosting: A technique for combining boosting and wagging. *Machine Learning*, vol. 40, no. 2, pp. 159−196. https://doi.org/10.1023/A:1007659514849

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

23. 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.

24. Rusakov O., Laskin M., Jaksumbaeva O., Ivakina A. (2015) Pricing in real estate market as a stochastic limit. Lognormal approximation. *Second International Conference on Mathematics and Computers in Sciences and in Industry. Malta, 2015*. https://doi.org/10.1109/MCSI.2015.48

25. Lemeshko B. Yu., Lemeshko S.B., Semenova M.A. (2018) On the issue of statistical analysis of big data. *Bulletin of Tomsk University*, vol. 44, pp. 40−49 (in Russian). https://doi.org/10.17223/19988605/44/5

26. Zhukova G.N. (2016) Identification of the probability distribution by the coefficients of asymmetry and kurtosis. *Automation. Modern Technologies*, vol. 5, pp. 26−33 (in Russian).

27. Laskin M.B. (2020) Multidimensional lognormal distribution in real estate appraisals. *Business Informatics*, vol. 14, no. 2, pp. 48−63. https://doi.org/10.17323/2587-814X.2020.2.48.63

18

About the authors

Mikhail B. Laskin

Cand. Sci. (Phys.-Math.);

Associate Professor, St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences (SPIIRAS), 39, 14 line, Vasilevskiy Island, St. Petersburg 199178, Russia;

E-mail: laskinmb@yahoo.com

ORCID: 0000-0002-0143-4164

Lyudmila V. Gadasina

Cand. Sci. (Phys.-Math.);

Associate Professor, St. Petersburg State University, 7/9, Universitetskaya emb., Saint-Petersburg 199034, Russia;

E-mail: l.gadasina@spbu.ru

ORCID: 0000-0002-4758-6104