www.degruyter.com/view/j/remav

sciendo

THE ASSESSMENT OF A BUILDING'S INSURABLE VALUE USING MULTIVARIATE STATISTICS: THE CASE OF THE CZECH REPUBLIC

Martin Cupal
Institute of Forensic Engineering
Brno University of Technology
e-mail: martin.cupal@usi.vutbr.cz

Marek Sedlačík
Faculty of Military Leadership
University of Defence, Brno
e-mail: marek.sedlacik@unob.cz

Jaroslav Michálek
Faculty of Military Leadership
University of Defence, Brno
e-mail: jaroslav.michalek@unob.cz

Abstract

When concluding a property insurance agreement, adjustment of the insured amount poses a certain risk. From the policyholder's point of view, the risk measure translates into the chosen target amount, which should correspond to the insurable value.

The aim of the research is to determine a statistical model for prediction of the insurable value with using current models in the Czech Republic. The model for insurable value prediction proposed in this paper accepts the risk of decision making under uncertainty suitably. The model's foundation is a synthesis of four core models discussing the addressed issue. The methodology is based on a classification tree created by the CART method, and multivariate linear regression. After the classification tree is created, the input variables which contributed to the classification are used in the regression model.

The database consists of 125 family houses which went through a detailed examination (they were documented, measured, and their technical state and legal status were determined), and described in experts' reports.

The obtained results showed a high degree of statistical association of selected predictors with the estimated insurable value of property, as well as with the acceptable risk, and subsequently, a relatively low percentage of misclassified objects. The proposed multiple regression model proved to be statistically significant and can be used for objective estimations of insurable values free of insurance companies' strategy. The designed methodology may be applied in other areas as well, for example, in decision-making processes at the population level in crisis situations.

Key words: *insurable value, building, CART algorithm, regression, classification tree.*

JEL Classification: *D46, C40, R30.*

Citation: Cupal M., Sedlačík M., Michálek J., 2019, *The Assessment of Building's Insurable Value Using Multivariate Statistics: the Case of Czech Republic*, Real Estate Management and Valuation, vol. 27, no. 3, pp. 81-96.

DOI: 10.2478/remav-2019-0027

1. Introduction

When concluding a property insurance contract, one of the most important decisions lies in setting the insured amount. It is in the policyholder's interest to set an amount corresponding to the insurable value. In case these two variables (the insurable value and the insured amount) are unequal, the policyholder is going to suffer a loss. The first loss situation can occur accidentally after submission of an insurance claim, when the insurable value is lower that the insured amount (i.e. in the case of underinsurance). The second scenario takes places when the insurance premium is being inadequately overpaid; i.e. the insured amount is excessively higher than the insurable value. This phenomenon is related to information asymmetries on the insurance market (SANDRONI, SQUINTANI 2013).

On that account, it would be beneficial to the policyholder if they could assess the insurable value themselves, using a simpler and more understandable mechanism than the usual method utilized by appraisers and valuators—a mechanism which would make it possible to set an objective property value for the purpose of insurance. In cases when the insurer's valuation models are applied, the assessment of insurable value is not objective or impartial. For that reason, the new model represents an effective solution, which would bring similar results as those of appraisers' and valuations' valuation process. However, it could also be used by non-experts, who represent the target group of policyholders. Basically, this would mean a significant reduction of risks related to incorrectly set insurable value of property.

The modelling process as such is essentially similar to basic approaches used with real property where the explained variable (price) is estimated using various explanatory variables which influence it (ADAIR et al. 1996). Quite often, regression methods (for example, Ordinary Least Squares estimator (OLS) with logarithmic transformation of variables, robust MM estimator, structural time series estimator, and also robust local regression (HANNONEN 2008)) are applied in modeling real estate prices. For instance, for data containing prices of land in Espoo, Finland, the structural time series estimator proved to be the most suitable of these methods. A hierarchical trend model (HTM) also presents an effective tool for modelling the selling prices of houses. It solves the spatial and temporal dependence of selling prices and the dependency of price index changes on housing quality, see Francke (FRANCKE, VOS 2004). The HTM is used for property valuation and for determining local price indices.

The fundamental difference when predicting a building's insurable value compared to its market price lies in the property's (real estate's) housing characteristics. Although its heterogeneity is the same as with classic market valuation, the commonly dominant influence of location on the property's market price is almost non-existent, so division into local markets or other location-related characteristics are not going to be discussed here. In market valuation, these location characteristics are dominant in the form of explanatory variables in a model, as was shown, for example, in a study on the prediction of house prices in Istanbul (KESKIN 2008). Other dominant characteristics are ones that are similar to the characteristics of the insurable value of a building, and are applied in market value estimations using hedonic regression models (e.g. type of building)—its size, structure types, etc. (SELIM 2008). A certain approach related to information asymmetry by real estate transaction presents decomposition of real estate into two components: land and the building structure (WONG et al. 2012).

In real estate price market price modeling, fuzzy logic started to be used as well (KUSAN et al. 2010). ANN (Artificial Neural Network) proved to be another alternate method to multiple regression models (SELIM, H. 2009); in data from Turkey, it showed more predictive power than hedonic regression. It is also interesting to compare the predictive ability using conventional parametric and semi-parametric models containing data on 2,595 houses in North Carolina (BIN 2004), where a semi-parametric regression also showed to be substantially useful for price prediction. Another approach to determining the relationship of a particular housing characteristic with selling price is using a quantile regression to identify the coefficients of diverse variables across different quantiles (ZIETZ et al. 2008; EBRU, EBAN 2011), or the Box-Cox quantile regression method, which was applied on the determinants of housing prices in Hong Kong in a recent article (KIM et al. 2015). Sirmans (SIRMANS et al. 2006) explained, by Meta regression analysis, significant variation of the estimated coefficient by geographical location with nine housing characteristics that appear most often in hedonic pricing models for single-family housing: square footage, lot size, age, bedrooms, bathrooms, garage, swimming pool, fireplace, and air conditioning. His research, as well as relatively often applied geographically weighted regression (e.g. CICHULSKA, CELLMER 2019), showed the uniqueness of geographical location compared to other price-setting factors, and therefore the insurable value can be considered as relatively constant and more predictable.

Application of the decision tree approach for price prediction and its comparison to the regression approach is described, for instance, in Breiman (BREIMAN et al. 1984). Also Fan (FAN et al. 2006) showed an application of this approach in a Singapore housing market case study. The purpose of this methodology is to find significant determinants (housing characteristics) affecting house prices, and use them to predict real estate prices.

In order to model a relationship between housing characteristics and house prices, this article uses a combination of two multivariate statistical methods—the CART (Classification and Regression Trees) method, and multiple linear regression models. The CART method was also used in article dealing with mortgage default issue (FELDMAN 2005), where many advantages of this method were mentioned and discussed in relation to traditional methods such as linear logistic regression and others. Based on studied data sets, it was possible to create a statistical model to predict the insurable value unaffected by factors of local markets or other factors resulting from the fact that buildings cannot be moved. In the field of real estate appraisal, this approach (price value) can be classified as a part of the basic "cost approach" to valuation (TEGOVA 2012). This means that the resulting value represents only costs that were or could be expended on construction of that particular building. On that account, the insurable value is set primarily using the cost method. Variables affecting it might be identified and added to the list of explanatory variables for the prediction of insurable value as a dependent variable. The impact of natural risks on a building is excluded here, but some research papers and articles present their influence on property value (see SPEYRER, RAGAS 1991).

The insurable value is considered only for one real estate type — single-family houses, which appear to be the most representative type of property. A family house is not a piece of real estate (such as a garage) too simple or too complicated and sectioned (such as a production facility), nor does it not include aspects of historic buildings and monuments, which require a special perspective and inclusion of other parameters (see STENDEBAKKEN et al. 2015); however, the determination of insurable value is often relevant to such buildings.

The aim of the research is to determine a statistical model for predicting insurable value using models currently used in the Czech Republic. The model for predicting the insured value described in this article is based on four core models. The first model (further referred to as “M_1”) is applied by experts and appraisers, and can be considered objective with the desired level of valuation accuracy. This model is based on methods set by the legislation of the Ministry of Finance of the Czech Republic (MFCR 2013). The second model (“M_2”) is provided by the Czech Insurance Association for the general public, whereas the remaining two models were created by commercial insurers (Allianz—“M_3”, Česká pojišťovna—“M_4”) for the public, and are used for determining premiums. Several observed variables present in individual models were identical, some were slightly different, and other variables were included only in certain models and completely absent from others. All models contain variables, so-called quantifiers (enclosed construction space, built-up floor space, number of floors, etc.), as well as the type of construction, type of roof, attic usability, or the factor of construction and equipment quality.

Subsequently, all these variables were used to classify the examined objects with respect to their insurable values. A classification tree was created based on methodology described in Breiman (BREIMAN et al. 1984). It allowed for the prediction of insurable value using established value categories. Afterwards, thanks to this created classification tree, it was possible to determine variables from individual models which play the most important role in the classification. These selected variables were later used in the creation of a regression model (SEBER, LEE 2003) which can help estimate a property’s insurable value based on values of selected classification variables. The selected classification variables are readily available, and even laymen (potential policyholders) can apply it in order to estimate an independent insurable value of a building. The created model is very similar to the first one (which serves as a reference for our purposes, and is marked as M_1 further in the text), utilized by experts, and objective with a reasonable level of accuracy.

2. Theoretical background - Assessment of insurable value

Insurable value represents a value which has to be considered in damage coverage resulting from maximum damage to the building. Subsequently, this value is the basis of the insured amount, which is indicated in the insurance contract and relevant to setting indemnification in the case of an insured event occurring. The insured amount usually serves to set the premium.

A correctly set insured amount should equal the building's insurable value. Otherwise, the policyholder may face risky situations, regardless of higher or lower insured amount. In the case that the insured amount is lower, underinsurance may take place (insufficient coverage of loss). On the contrary, if the insured amount exceeds the insurable value, the policyholder will have to pay proportionally higher premiums (interim loss).

In the current market turbulence, operational risk management is also a basic duty of every financial institution, including insurance companies. In a situation when principles based on prudent business activities and company solvency are malfunctioning, and systematic risk is an increasing threat, a new system should be applied. Chorafas (CHORAFAS 2004) provides basic information on the quantitative method, offering a practical classification and ranking of risks (risk scenarios).

The policyholder bears the risk of a correctly set insured amount, or rather the correctly set insurable value, since they are to the policyholder's benefit. On the contrary, the insurance company can profit from excessively high insured amounts. It is up to policyholders to set insured amounts; however, as they are usually non-experts, they cannot estimate these values easily, but need to have expert appraisals determining insurable values carried out. Of course, these service fees are not low, and their results may be of uneven quality, although experts and appraisers should work in compliance with ethical and professional codes. Nevertheless, in many cases, experts' and appraisers' credibility is problematic, since their activities are often not monitored by regulatory bodies (AMIDU, ALUKO 2010), even though setting up such a mechanism is essential.

Insurance companies offer calculation of insured amounts, though it is non-transparent, as it often includes overestimation or gross inaccuracies (CUPAL 2014). The insurer may be motivated by an insurable value exceeding the objective value, which would lead to the payment of higher premiums. On that account, it would be helpful to propose an objective model for insurable value estimation, so that even non-experts could use it to assess insurable values on their own.

When defining insurable value, the EVS (European Valuation Standards) definition should be taken into consideration (TEGOVA 2012 – EVA 4 Section). It states that, among others, during the process of insurable value estimation, two categories may be applied—new replacement cost (NRC), and depreciated replacement cost (DRC). The dynamics between house prices, rent and structure age (the age depreciation of house price, defined as the decline of house price with respect to house age) are mentioned by Xu (XU et al. 2017) and Cannaday (CANNADAY, SUNDERMAN 1986), who performed methods for the estimation of depreciation within the cost approach to the appraisal of a single-family residential property. This article deals only with the main category of new replacement cost (NRC). Regarding DRC, a rather important discussion is held on what kind of depreciation should be considered (more detailed, e.g. COPIELLO, BONIFACI 2018). The EVS recommends factoring only physical deterioration, not functional or economic defects. Moreover, defect evaluation methods based on current processes are often inflexible, and do not render the market situation well (MANSFIELD, PINDER 2008). This discussion brings up the question of adequacy, whether in the case of compensation for damage covering only accounted technical defects or an implicit enrichment of the policyholder may occur, and so the probability of insurance fraud increases. Those risks are related to big differences between market and insurable values of the same property.

3. Current Assessment Models Used in the Czech Republic

3.1. An Overview of Current Models

At present, several models are applied in the Czech Republic which can be used to predict insurable values of most common building types (usually a single-family house, an apartment, a cottage, and a garage). Such models are offered by the most important insurance companies on the market. This article focuses on models offered by Allianz (the M_3 model) and Česká pojišťovna (the M_4 model). An essentially similar model is applied by the Czech Insurance Association; its model was also included in the research, and called the M_2 model.

At the same time, it is possible to indicate insurable value based on general valuation procedures which return unbiased values with the required level of accuracy. In this research, the most sophisticated, yet least useful model from the policyholder's point of view is the M_1 model, resulting from the Czech statutory regulation on appraising (MFCR 2013). In comparison to other models (provided by commercial insurance companies or associations), this model may be considered the most objective and accurate one; however, it differs from them regarding its calculation principle and the specific choice of variables. The M_1 model is commonly used in real estate appraisals, and is based on the principle of highly aggregated cost prices.

3.2. An Overview of All Variables

Each of these four models contains a set of explanatory variables – predictors, and each of them has one explained variable – the insurable value. The selected models and variables were also applied in previous studies addressing the accuracy of obtained estimates (CUPAL 2014). A comprehensive overview of all variables, including their markings, types and descriptions can be found in Table 1.

Table 1

List of All Variables with Description

Denotation	Name	Affiliation	Type	Description
		DEPENDENT VARIABLES		
Denotation	**Name**	**Affiliation**	**Type**	**Description**
Y_{M1}	*Reproduction costs (Insurable value)*	M_1	ratio	It corresponds to the price of a new building [CZK].
Y_{M2}	*CAP insurable value*	M_2	ratio	It should correspond to the price of a new building [CZK]. It is suitable for only common types of property. Note: The insurable value of high-class or luxury property can be several times higher.
Y_{M3}	*Insurable value*	M_3	ratio	It should correspond to the price of a new building [CZK].
Y_{M4}	*Recommended insurable value*	M_4	ratio	It should correspond to the price of a new building [CZK].
		PREDICTORS		
Denotation	**Name**	**Affiliation**	**Type**	**Description**
ES	*Enclosed space*	M_1	ratio	Enclosed space of the whole house [m³].
S	*Type of structure*	M_1	nominal	Brick, ferroconcrete—monolith, prefabricated reinforced concrete, and wooden structure.
R	*Type of roof pitch*	M_1,M_2,M_4	dichotom.	Sloping roof or flat roof.
B	*Basement area*	M_1,M_2,M_4	dichotom.	Basement area over 50% *BA 1.AGF*; the second option is basement covering up max. one half of *BA 1.AGF*).
NF	*Number of floors*	M_1,M_2	ratio	The number of floors (excluding attic and basement).
BP	*Basic price*	M_1	ratio	Basic price per unit [CZK/m³].
USA	*Usable attic*	M_1	ordinal	Unusable attic; options of the Unusable Attic variable—usable up to ⅓ of *BA 1.AGF*, or usable up to ⅔ of *BA 1.AGF*, more than ⅔ of *BA 1.AGF* usable.

K_A	*Usable attic coefficient*	M_1	ratio	It has a range of <1.00; 1.12>. Value 1.00 represents no usable attic; 1.12 means fully usable; divided into thirds of usable attic space (cf. USA).
K_4	*Equipment and construction coefficient*	M_1	ratio	All standard (quality and performance) equipment and constructions provide a value of 1.00. Substandard features less than 1.00, above standard - vice versa.
K_i	*Price change coefficient*	M_1	ratio	Gives annual building price changes.
ABP	*Adjusted basic price*	M_1	ratio	Basic unit price adjusted using coefficients and based on the particular object's category.
UA	*Usable attic*	M_2	dichotom.	Usable or unusable attic.
SBA	*Sum of built up area*	M_2	ratio	Sum of built-up area of all above-ground floors including attic space (living space in the attic or uninhabited attic) and basement [m^2]. Thus *SBA* = Σ*BA* for each floor of the house.
PC	*Property condition*	M_3	nominal	First category for property younger than 30 years; second one for that older than 30 years after an total renovation, and the last one for property older than 30 years without an overall reconstruction.
RT	*Roof type*	M_3	nominal	3 types: sloping roof, flat roof, and sloping roof with usable attic.
BA 1.AGF	*Built-up area of the first above-ground floor*	M_3, M_4	ratio	Built-up area of the first above-ground floor [m^2].
BBA	*Basement*	M_3	dichotom.	Existing basement or no basement.
BUA	*Basement usable area*	M_3	ratio	Basement usable area [m^2].
NIF	*Number of inhabited floors*	M_3	ratio	It represents a number of inhabited floors, including attic.
QW	*Quality of workmanship*	M_3	ordinal	3 levels of quality: basic, standard and luxury (at least 3 elements above standard).
TF	*Type of floors*	M_4	dichotom.	Houses with first above-ground floor plus at least one additional floor which is not an attic. The other option is a single-storey house.

Source: own study.

The marking of variables listed in the first column in Table 1 will be used further on for the purpose of their brief identification, and also in calculation formulas, graphs and results.

3.3. The M_1 Model

The insurable value was first assessed by the Property Valuation Act (Czech legislation), which is used mostly by experts (qualified appraisers) to obtain the valuation, based on the cost approach (TEGoVA 2012). The principle of this procedure lies in the highly aggregated degree of budget costs of buildings. The procedure is contained in the Property Valuation Act No. 151 of 1997 with its implementing regulation No. 441 of 2013 (MFCR 2013). The regulation of the Ministry of Finance closely defines particular procedures of valuation for various types of real estate. The cost approach is included in Part 2, Head 1 (Building cost valuation method). With a focus on single-family houses (the database), the following procedures and comments will only concern these. This leads to Section 5 (Single-family houses, holiday cottages). The methodology is designed as a selection of appropriate categories in the first step, according to the corresponding appendix of the regulation (MFCR 2013); the basic price is then obtained (as unit price). This price must be adjusted by particular coefficients selected correctly from the appendix. We thus get the adjusted price, which is a unit price (price for 1 cubic meter of enclosed space) that must be multiplied by the corresponding number of cubic meters of enclosed space detected by the procedure in the appendix of the regulation for individual family houses. The total price corresponds with the expert's assessment of insurable value, which is identical to the category of reproduction costs. The calculation scheme is as follows.

The whole procedure is based on basic price (BP) adjustment, formally expressed in two steps:

$$1.\ \text{Step: } ABP = BP\,(S, R, B, NF) * K_A * K_4 * K_i, \tag{1}$$

$$2.\ \text{Step: } Y_{M1} = ABP * ES. \tag{2}$$

In the first step, BP denotes the unit price variable (CZK/m^3), where BP (S, R, B, NF) is the basic price BP as a function of variables S, R, B, NF (see Table 1), and ABP denotes adjusted unit price. Variable S contains 4 types of structure (brick, ferroconcrete—monolith, prefabricated reinforced concrete, and wooden structure), variable R defines the type of roof pitch (sloping roof or flat roof), variable B is categorized according to basement area percentage (total basement or only mid basement), and NF represents the number of floors (excluding the attic). The appropriate combination of these variables gives the correct BP (basic price per unit). These variables (categories) are similar to building classes in risk assessment (see ALGERMISSEN, STEINBRUGGE 1984). The following adjustment of BP (still part of the first step) is performed by multiplying coefficients. The regulation of the property valuation act contains two others—K_5 and K_p. However, these two are unsuitable for insurable value assessment Y_{M1}, because both of them express the dependence of Y_{M1} on the house location. The K_A coefficient takes into account the usable attic based on its percentage in the built-up area. Its values are categorized into thirds, and they get combined with NF, e.g. (0;1/3>:=1.05) means a usable attic up to a maximum built-up area of 1/2 of total first floor built-up area and for a single story house.

Another coefficient—K_4 adjusting the unit price, is called "the equipment coefficient". Its normal range of values is 0.8-1.2, but it can be higher or lower if appropriately justified by experts. This coefficient is very important for insurable value assessment, as equipment should always be considered as permanently fixed to the real estate. Calculation of the coefficient is represented as a weighted mean with price proportions of structures and equipment which represent weights (of a weighted mean) in K_4 calculations. The particular weight is then multiplied by a selected coefficient for each item; the selection is performed by an expert with a list of appropriate items for each type of equipment (structure). This list sets the standard equipment (in such cases, K_4 = 1.00); in other cases, K_4 increases or decreases by 0.54 (above or below standard) for one item. Consequently, the K_4 coefficient is calculated as a weighted mean of all items. The constant value (plus or minus 0.54) is one of the weaknesses of the current procedure for assessing insurable value.

The last coefficient - K_i reflects the price changes of buildings in annual periods. This is also very important for insurable value, because it is necessary to update the initial price for potential claims. Finally, all coefficients multiply the initial BP to get the ABP adjusted unit price (CZK/m^3).

In the second step, ES denotes the enclosed space of a house measured in cubic meters; the procedure to set the correct ES for a particular house is contained in the regulation appendix (it is not considered a purely geometric body). It describes which parts of the house (for instance, recessed construction elements, subconstruction elements, various roof geometries, and balconies) constitute the enclosed space. Finally, after receiving *ES* and multiplying it by *ABP*, then Y_{M1}, the insurable value (in CZK), can be obtained. The Y_{M1} relates to the new price (the NRC principle, see above).

This procedure of the M_1 model can continue with the next step, which enables us to get another output based on a different DRC principle (see above). This basically means the reduction of the Y_{M1} output using a corresponding percentage of physical depreciation of the property. In some cases, such an output might be identified as insurable value in accordance with the terms and conditions of insurance. However, this is not a prevalent method of insurable value determination, and therefore it is not considered in the article (MFCR 2013).

3.4. The M_2 Model

The M_2 model and other models considered in the paper are similar to or to some extent represent a simplification of the first model. However, they do not require data sources, and are frequently used in practice by insurance companies and policyholders. Its main advantage is also automatic calculation. Nevertheless, unlike in the previous model, the *BP* of other models is unknown; therefore, it takes part in the calculation only implicitly, in combination with values of input variables.

The insurable value estimation Y_{M2} is based on a simplified cost approach provided by insurance specialists. The common procedure by the Czech Insurance Association (CAP) leading to the determination of insurable value should be a method providing sufficient outputs for insurance parties. This is, like in the previous method, based on adjusting basic cost unit price, where a unit corresponds to 1 square meter of sum of built-up area *SBA* of all above-ground floors, including attic space (living space in attic or uninhabited attic).

It can be formally displayed by the following notation:

$$Y_{M2} = BP(R, B, NF, UA) * f(SBA), \tag{3}$$

where *BP*(*R*, *B*, *NF*, *UA*) is a function of *R*, *B*, *NF*, *UA* and denotes a unit price variable (CZK/m^2), but, as a unit, one square meter of sum of built-up area of each floor is considered. *BP* is determined based on *R*, *B*, *NF*, *UA* variables. *R*, *B* and *NF* variables have the same meaning as in the first model. Just like the K_A coefficient, the *UA* variable represents a category with a usable attic, but in this case only as a dichotomous variable (the attic is either usable or unusable). However, the *BP* variable is unknown in this model.

The variable *SBA* means a sum of built-up area for each floor of the house. When multiplied by f(*SBA*) (specified function of *SBA*), the insurable value Y_{M2} can be assessed (CUPAL 2014).

This model has certain limitations that might affect valuation, especially that of luxurious equipment.

3.5. The M_3 Model

Similarly, models M_3 and M_4 can be written down formally. The M_3 model has been published by the insurance company Allianz.

$$Y_{M3} = BP(PC, RT, BBA) * g_1(BA\ 1.AGF) * g_2(BUA) * g_3(NIF) * g_4(QW), \tag{4}$$

BP denotes a cost unit price variable (CZK/m^2) again, but a square meter of built-up area of the first above-ground floor including attic (*BA 1.AGF*) is considered as a unit. The variable *BA 1.AGF* enters here as a functional argument of function g1. *BP*(*PC*, *RT*, *BBA*) represents a function of the *PC*, *RT*, *BBA* variables. Variable *PC* gives 3 categories; the first is for properties younger than 30 years, the second for those more than 30 years old after an total renovation, and the last one - for property more than 30 years old without an overall reconstruction. *RT* denotes the roof type in combination with a usable attic; it also provides for 3 possibilities: sloping roof, flat roof and sloping roof with a usable attic. *BBA* expresses the existence of any basement as a dichotomous variable. The *BP* variable is unknown in this model too.

Other variables entering the M_3 model are *BUA* (denoting basement usable area), *NIF* (representing a number of inhabited floors, including the attic) and *QW* (representing an ordinal scale of the quality of workmanship at three levels: basic, standard and luxury). They have an influence on the insurable value Y_{M3}, which is described by special functions g_2, g_3 and g_4.

3.6. The M_4 Model

The last model, M_4, belongs to Česká pojišťovna, a leading insurance company in the Czech Republic.

$$Y_{M4} = BP(R, B, TF) * \mathrm{h}(BA\ 1.AGF), \tag{5}$$

Function *BP* of *R*, *B*, *TF* again denotes a cost unit price variable (CZK/m²), a square meter of built-up area of the first above-ground floor (*BA 1.AGF*) is considered as a unit. The *B* variable is divided into two categories based on basement area percentage (basement area over ½ of *BA 1.AGF*, or basement area up to ½ of *BA 1.AGF*), same as in the M_1 model. The *R* variable defines the type of roof pitch (sloping roof or flat roof), which is the same as in the M_2 model. *TF* divides the type of floors into 2 categories; the first one involves houses with an above-ground floor plus at least one additional floor (not attic), with the others falling into the second category. When *BP* is multiplied by h(*BA 1.AGF*) (specified function of *BA 1.AGF*), the insurable value Y_{M4} can be assessed. The *BP* variable is unknown in this model too (CUPAL 2014).

4. Data

The starting point for creating an input database for model designs were comprehensive experts' reports containing the valuation of family houses. These reports were obtained from individual appraisers or expert institutes. Altogether, 125 appraisals were successfully collected.

Hence the database consists of 125 family houses which went through a detailed examination (they were documented, measured, and their technical state and legal status were determined), and were described in experts' reports. Documentation on each piece of real estate is available, and so is its accurate up-to-date survey and description of materials.

This set of family houses represents an average sample of family houses in the Czech Republic, which is mainly shown by the fact that, based on material characteristics, brick houses and pitched roofs are prevalent in the database. The reasons are, above all, rooted in history and geography. Less common are houses made of concrete (monolithic constructions or prefabricated elements). The sample, however, does not include wooden houses or structures built of wooden materials. This type of buildings has started to develop in the Czech Republic only recently, and is very slowly winning customers on the market. In the course of evaluation, structures of different ages need to be appraised; this database complies with this requirement, since it contains new buildings as well as houses that are a hundred years old.

The resulting database compiled using experts' reports of all 125 houses contains all variables of individual models which are needed for given models to calculate insurable values. The price level (construction costs) for the insurable value of all models corresponds to the price level of 2014.

The data set later used for object classification was created as follows. For all 125 objects, values of input variables used for estimated insurable value calculation in each of four accounted models were determined (18 variables altogether, see regressors in Table 1), and for each model, input variables were used to set a model value of each individual object based on functional instructions from Chapter 3. The number of regressors further applied in classification was reduced from 21 (all included in Table 1) to 18. Three variables (*ABP*, K_A a K_i) which were not used, are basically further rendered by other variables, and they represent only an intermediate step in the M_1 model calculations (as for *ABP* and K_A). K_i variable represents only a constant for adjusting basic cost price to current price level.

To each object, an average insurable value of property was assigned (*AY234*, a new dependent variable), which had been calculated as an average of values estimated using the M_2, M_3 and M_4 models. Such values were obtained by calculations performed with these models; see for Chapters 3.4 to 3.6, formulas (3), (4), and (5). The precise calculation process of these models in unknown (the given models contain unspecified functions, which belong to insurance companies and the Czech Insurance Association's know-how). However, along with the values of input variables, values of a dependent variable for individual objects corresponding to insurable values were obtained from insurers and the Association. The resulting average insurable value *AY234* was further divided into four categories (categorized *CAT_AY234*):

Low (A) – objects with a maximum value of 2,000,000 CZK
Medium (B) – objects with a value range between 2,000,000 and 4,000,000 CZK
High (C) – objects with a value range between 4,000,000 and 6,000,000 CZK
Extreme (D) – objects with a value higher than 6,000,000 CZK.

The value category for each given object obtained this way was a part of the data set.

Furthermore, independent variables were assigned to every object; these were individual variables from M_1 to M_4 models, which were used in individual models to calculate the insurable values of property.

This way, a data set was created, with two dependent *AY234* and categorized *CAT_AY234* variables, and 18 independent variables—predictors, which were later utilized to estimate categorized property insurable values and model property insurable values.

5. Methodology

5.1. Construction of a Classification Tree

The purpose of performing a classification was to divide 125 individual objects based on determined values of 18 independent variables, which describe these buildings into four groups so that the insurable values of objects in a given group correspond to a maximum degree to the categorized value of the object *CAT_AY234* contained in the data set. After a thorough statistical analysis of the data set, the classification tree method described by Breiman (BREIMAN et al. 1984) was chosen, and the calculation itself was carried out using the STATISTICA software product. This methodology is advantageous, since classification can be performed using cardinal and ordinal, as well as nominal variables. The CART classification algorithm proved to be the best for the calculation. A priori probabilities of individual categories were determined using a teaching set with five-fold cross-validation. As a splitting rule, Gini splitting criterion was applied, and the minimum number of observations in one nod was 1. Given the number of observations, an optimally pruned tree was selected applying multifold cross-validation using the 1S rule, which is based on an estimate of the total price resulting from misclassification. After the classification tree was constructed, the classification ability of all 18 independent variables was assessed using their significance. Details on the applied algorithm can be found in Breiman (BREIMAN et al. 1984), and basic information on parameters of the performed classification is given in Fan (FAN et al. 2006).

5.2. Regression Model

The classification performed using a classification tree allowed for the classification of individual objects into four price categories A, B, C and D, based on determined values of 18 independent variables. It also enabled the selection of variables with the best classification ability. Subsequently, these variables were used to construct a linear regression model for property price prediction (see SEBER, LEE 2003). In the regression model, the dependent variable was the average insurable value, i.e. the *AY234* variable. The model was based on the assumption, that the *AY234* variable is linearly dependent on accompanying variables chosen from all 18 independent variables which were used to construct the resulting classification tree. The authors of this article gave preference to this approach over the instant application of a classification regression tree, because it was possible to compare results obtained using a classification tree to results of a regression analysis immediately. The level of consistency between both sets of results then allowed for an assessment of the suitability of both methods.

6. Results and Findings

6.1. Results Obtained Using a Classification Tree

The classification tree obtained for the data set using the CART method is represented in Figure 1. The number above each classification tree node gives the number of objects—pieces of a building in the given node; the upper left number represents the node's marking, and the bar graph inside each node provides representation of individual property groups (based on the observed value levels) which were classified for this node, i.e. the histogram of the *CAT_Y234* variable for that node's objects. Under each internal node, the respective splitting rule can be found. The symbol indicated in the upper right-hand corner of every node of the classification tree represents the resulting value group assigned to the node. From the first classification tree split, which is shown in Figure 1, it is apparent that, based on the accompanying *SBA* variable, the original 125 buildings may be divided into two unequally sized groups of objects. Pieces of real estate whose *SBA* values are lower than 156.5 were included in the end note marked as node No. 2., which is depicted in the left branch of the classification tree. Buildings in this node are therefore assigned into a group of properties with a low average insurable value (category A). End node No. 2. includes pieces of real estate with the lowest average insurable value, which have the lowest *SBA* variable values.

The right branch of the classification tree represents the classification of the remaining 102 buildings into 9 end nodes indicated in red. In addition to the previously mentioned *SBA* variable, a gradual split of all internal classification tree nodes using *BA 1.AGF*, K_4 and *NIF* descriptors was performed, so that each piece of a building is assigned one of the insurable value classes: low (A), medium (B), high (C), or extreme (D). End node No. 2, made of 23 objects, and node No. 10, made of 4 objects, represent group (A) with low mean insurable values. Those are pieces of property with the lowest values of the *SBA* variable, or rather of *BA 1. AGF* and K_4 descriptors. The remaining end clusters, i.e. No. 11, 12, 14, 16 and 18, contain objects with medium insurable values (class B); similarly, objects with high values are sorted into end nodes No. 15 and 19. In the case of their *SBA* variable value exceeding 449.5, buildings with extreme insurable values (D) belong to node No. 9.

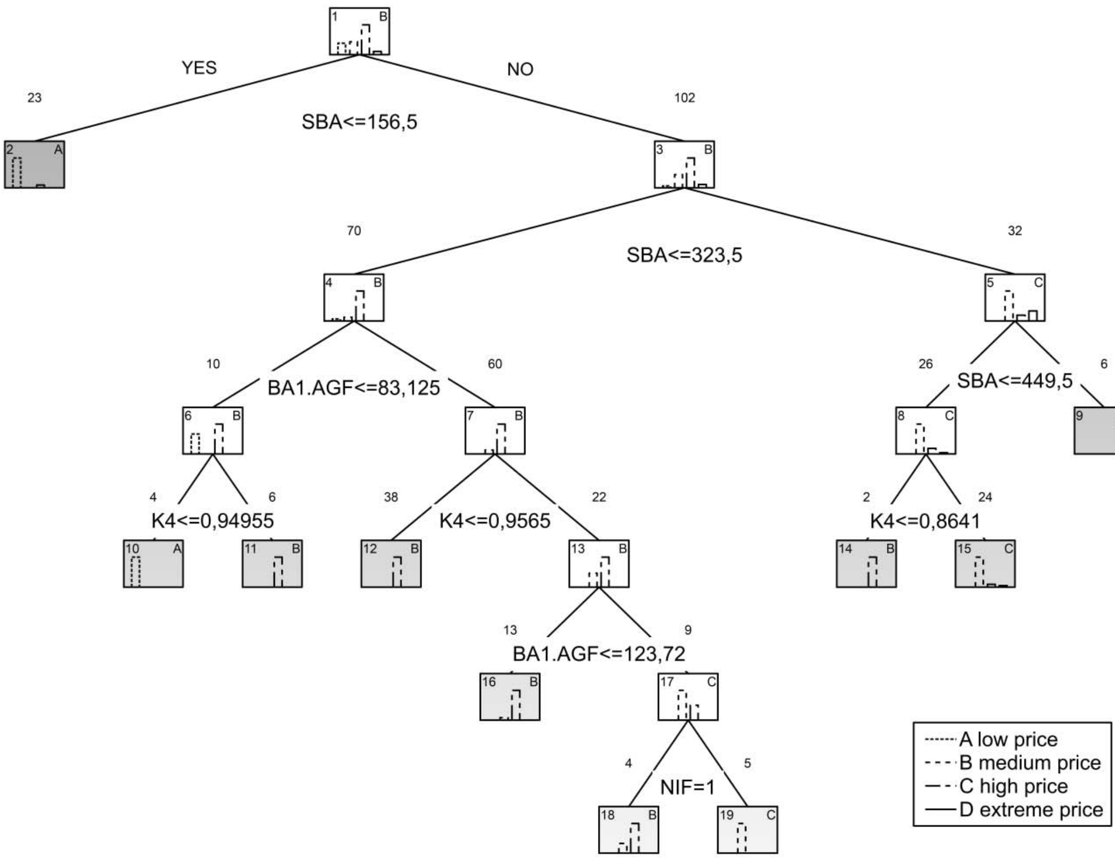

Fig. 1. Classification Tree for the *CAT_Y234* Variable. *Source*: own study.

The classification tree shows that, out of a total of 18 available predictors, only 4 accompanying variables were considered during the classification: *SBA*, *BA 1. AGF*, K_4 and *NIF*. This means an important advantage when the classification tree is used by individual non-expert appraisers. In order to compare classification abilities of all accompanying variables, their significance to the resulting classification was calculated using the CART method. The resulting classification is based on a measure called "Predictor Variable Significance Rankings" (PVSR) (see BREIMAN et al. 1984). The determined PVSR for a 0 to 100 scale is shown graphically in Figure 2. A classification value of 100 indicates its high classification significance; on the contrary, a value of 0 means a low classification ability of such a variable (see SEDLACIK et al. 2015). This figure implies that the most important variable for the performed classification is the *SBA* variable, and also the *ES* variable, which can work as a substitution variable. Other significant values exceeding 60 are *BA 1. AGF*, K_4 and *BUA*.

Results obtained using the CART method lead to estimation of the categorized value, which is written in the upper right-hand corner of each node (see Fig. 1). Later on, these estimates were compared to the object's original real mean insurable value calculated (the categorized value) using M_1 to M_4 models. A comparison is carried out in Table 2.

This table contains frequencies of correct classifications and misclassifications; more precisely, a frequency of objects with the given categorized insurable value sorted into a selected category with a categorized value set using the classification tree. So, for example, value 2 in line C and column B indicates that exactly 2 objects with a mean insurable value in category C were assigned to value category B using the classification tree.

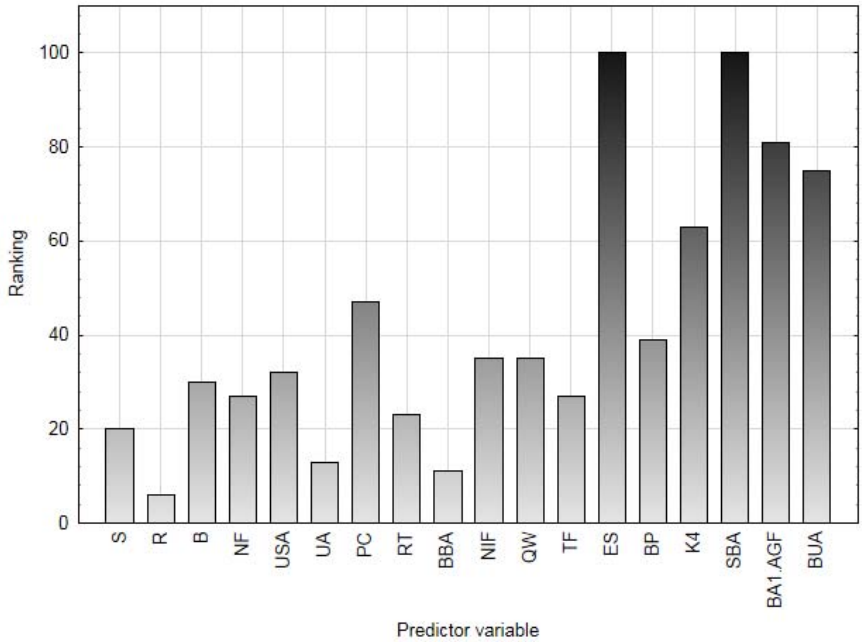

Fig. 2. Predictor Variable Significance Rankings (PVSR). *Source*: own study.

Table 2

Comparison of CART Estimates and Real Mean Categ. Values of M_1 to M_4 Models

Predicted Insurable Value Category	Real Mean Categorized Insurable Values of M_1 to M_4				Total
	A	B	C	D	
A	**25**	2	0	0	27
B	0	**61**	2	0	63
C	0	2	**26**	1	29
D	0	0	0	**6**	6
Total	25	65	28	7	**125**

Source: own study.

Out of a total number of 125 objects, 7 were misclassified, which is 5.6%. An explanation of the characteristics of these buildings is summarized in Chapter 7.

A similar approach was applied for the comparison of the classification of 125 observed buildings based on CART classification tree with independent classification in accordance with the previously described M_1 model. The comparison once again does not show significant differences in this case, as documented in Table 3. Of a total number of 125 objects, 23 were misclassified, which makes for 18.4%.

6.2. Prediction of Insurable Value Using Regression Analysis

In the previous chapter, results obtained using the CART method show that the categorized insurable value of an object can be predicted using four accompanying variables (predictors), i.e.: *SBA*, *NIF*, *BA 1. AGF* a K_4 (see Table 4) with a rather low misspecification error. Based on their significance, these four variables were chosen from 18 available predictors; therefore, from a practical point of view, they make the estimation of a categorized value of property easier. In order to independently test how good the categorized insurable value estimates based only on the four selected variables are, and to calculate an estimation of the mean (non-categorized) property insurable value, a linear regression model was used (SEBER, LEE 2003).

Table 3

Comparison of CART Estimates and Real Categ. Values Based on the M_1 Model

Predicted Insurable Value Category	Real Categorized Insurable Value Y_1 of M_1 Model				Total
	A	B	C	D	
A	**25**	2	0	0	27
B	5	**53**	5	0	63
C	0	7	**19**	3	29
D	0	0	1	**5**	6
Total	30	62	25	8	**125**

Source: own study.

Table 4

List of Predictors Selected by CART Method

Denotation	Name	Affiliation
K_4	*Equipment and construction coefficient*	M_1
SBA	*Sum of built-up area*	M_2
NIF	*Number of inhabited floors*	M_3
BA 1.AGF	*Built-up area of the first above-ground floor*	M_3, M_4

Source: own study.

The created regression model can be formally described by the following equation:

$$AY234 = b_0 + b_1{*}NIF + b_2{*}K_4 + b_3{*}SBA + b_4{*}BA\ 1.AGF + e, \tag{6}$$

where the non-categorized insurable value *AY234* is a dependent variable, *SBA*, *NIF*, *BA 1. AGF* and K_4 are independent variables that help to predict the object's insurable value, *e* is the random error of the model, and b_0, b_1, b_2, b_3 and b_4 are model parameters that have to be estimated. These estimates were determined using the least square method (OLS under common assumptions; see SEBER, LEE 2003), and are listed in Table 5.

All model parameters were statistically significantly different from zero at the 5 percent level of significance. Overall, the model proved to be statistically highly significant (multiple correlation coefficient R=0.95745, the relevant F statistics is F=330.21 for 4 and 120 degrees of freedom with the corresponding p-value p=0.000000). Results of regression analysis are systematically summarized in Table 5.

Table 5

Linear Regression Results

Variable	b	Std. Err. of b	t(120)	p-value
	-2 385 090	374 166.3	-6.374	0.000000
K_4	1 584 349	447 633.6	3.539	0.000572
SBA	8 651	639.3	13.532	0.000000

NIF	450 961	101 537.3	4.441	0.000020
BA 1.AGF	10 863	1 276.1	8.512	0.000000

Source: own study.

Given the overall significance of model parameters, it can be claimed that an object's insurable value can be predicted using equation (6), with coefficients listed in Table 5. The predicted value $\hat{A}_{Y234}$ in thousands of Czech Crowns (1,000 CZK) can be estimated using the following equation:

$$\widehat{AY234} = -2\,385.09 + 450.961{*}NIF + 1\,584.349{*}K_4 + 8.651{*}SBA + 10.863{*}BA\ 1.AGF. \qquad (7)$$

A comparison of results of classifications performed using the classification tree and of regression analysis can be found in Table 6. This table lists frequencies of objects assigned in the classification tree into a category states in the first column and assigned by the regression model into a category corresponding to the first line. Thus, the respective table's diagonal contains frequencies of objects classified equally, in both the regression tree and in the regression analysis; outside the diagonal, frequencies of objects classified differently in both methods are listed. Table 6 implies a high degree of concordance, while 14 objects out of 125 building were classified differently only in neighboring categories of the monitored mean insurable value of property. There are only 11.2% misclassified observations. Despite these incongruent results, in comparison with original classification using the reference M_1 model, a significant concordance apparently exists between both mentioned classifications.

Table 6

Comparison of Predictions of Insurable Value Categories Using CART and Using Regression Analysis

Prediction Using CART Method	Prediction Using Regression Analysis				Total
	A	B	C	D	
A	**24**	3	0	0	27
B	2	**57**	4	0	63
C	0	2	**24**	3	29
D	0	0	0	**6**	6
Total	26	62	28	9	**125**

Source: own study.

In conclusion, a comparison of the classification of 125 observed buildings was performed based on regression analysis with the previously described M_1 model. In this case too, the comparison does not show significant differences, which is documented in Table 7. Out of the total number of 125 objects, 17 were misclassified, which is 13.6%. As a matter of fact, this comparison represents the main connection of the model proposed in this paper to the objective M_1 source model.

Table 7

Comparison of Predictions of Insurable Value Categories Using Regression Analysis and Real Categorized Insurable Values Based on the M_1 Model

Prediction Category Based on Regression Analysis	Real Categorized Insurable Value Y_1 of M_1 Model				Total
	A	B	C	D	
A	**26**	0	0	0	26
B	4	**56**	2	0	62
C	0	6	**20**	2	28
D	0	0	3	**6**	9
Total	30	62	25	8	**125**

Source: own study.

7. Discussion and conclusions

Results obtained using the CART method and regression model suggest that this method can be used in practice to estimate insurable value. The key is the proposed model's link to the M_1 reference model applied by experts and appraisers. A more detailed analysis showed that, in relation to the M_1 reference model, certain inaccuracies occurred mostly due to the following reasons:

– In the case of non-standard building heights, variables concerning floor space and enclosed space differed (high ceilings, areas with very low clear height, etc.). Moreover, generally speaking, these predictors strongly influenced the property insurable value.
– The M_3 model showed to be very sensitive to the *QW* variable (Quality of Workmanship), which could change the insurable value abruptly. On the whole, the quality of construction and equipment is reflected only in M_1 (K_4) and M_3 (*QW*) models. On that account, in case of above-standard constructions, estimates of both M_1 and M_3 models are often significantly higher than those of M_2 and M_4 models for the same building.
– A combination of high surface area values (the *BA 1.AGF* variable) and cellars (the *B* variable) are very significant (nearly inadequately) in the M_4 model. The insurable values are overestimated (see also CUPAL 2014).
– In the M_4 model, attic never qualifies as a stand-alone storey, which leads to distortion, mainly in comparison to the M_1 model. The M_4 model interprets a house with an extensive enclosed space (*ES*) with a habitable attic as a single storey house with no basement, which is very misleading, and causes big differences between insurable value estimates.

The estimated regression model can be applied in common assessments of insurable values of family house constructions, because its strong statistical link with the objective M_1 model was proven. It was shown that the best predictors are variables concerning certain surface areas of buildings (sum of built-up area – *SBA*, and built-up area of the first aboveground floor – *BA 1.AGF*), equipment quality, and the number of habitable floors.

Predictors of a quantitative nature, which were clearly dominant here (*SBA*, *BA 1.AGF*, and *NIF*) allow for very good objective assessment. Moreover, common users can find out, or rather measure values of surface area variables more easily than spatial variables (*ES* in the M_1 model), which gives them a practical advantage.

The last selected predictor, i.e. K_4, is a qualitative one, and relates directly to the quality of property constructions and equipment. When applying this variable, it is necessary to consider individual elements and their presence in the given building, and to assess their value deviation from the defined standard. The entire procedure of obtaining K_4 is described in a statutory regulation (MFCR 2013) and should be respected. Since specific items are also listed in the regulation procedure, selection and calculation can be done simply. This approach ensures a high degree of objectivity of a variable, which is otherwise difficult to assess by a common user.

8. References

ADAIR A. S., BERRY J. N., MCGREAL W. S., 1996, *Hedonic modelling, housing submarkets and residential valuation*, Journal of Property Research, vol. 3, no. 1, pp. 67-83.

ALGERMISSEN S. T., STEINBRUGGE K. V., 1984, *Seismic Hazard and Risk assessment: Some Case Studies*, The Geneva Papers on Risk and Insurance, vol. 9, no. 1, pp. 8-26.

AMIDU A., ALUKO B. T., 2010, *Client influence on valuation: Perceptual analysis of the driving factors*, International Journal of Strategic Property Management, vol. 11, no. 2, pp. 77–89.

BIN O., 2004, *A prediction comparison of housing sales prices by parametric versus semi-parametric regressions*, Journal of Housing Economics, vol. 13, no. 1, pp. 68-84.

BREIMAN L., FRIEDMAN J., STONE C. J., OLSHEN R. A., 1984, *Classification and Regression Trees (Wadsworth Statistics/Probablility)*, Chapman and Hall/CRC, UK.

CANNADAY R. E., SUNDERMAN M. A., 1986, *Estimation of Depreciation for Single-Family Appraisals*, Real Estate Economics, vol. 14, no. 2, pp. 255–273.

CHORAFAS D. N., 2004, *Operational Risk Control Business Opportunity and Challenges for the Insurance Industry*, The Geneva Papers on Risk and Insurance: Issues and Practice, vol. 29, no. 1, pp. 87-101.

CICHULSKA A., CELLMER R., 2019, *Analysis of Prices in the Housing Market Using Mixed Models*, Real Estate Management and Valuation, vol. 26, no. 4, pp. 102-111.

COPIELLO S., BONIFACI P., 2018, *Depreciated Replacement Cost: Improving the Method Through a Variant Based on three Cornerstones*, Real Estate Management and Valuation, vol. 26, no. 2, pp. 33-47.
CUPAL M., 2014, *Models for insurable value assessment, their accuracy and optimal value of buildings*, Economic Challenges in Enlarged Europe conference proceedings. Tallinn: Tallinn University of Technology.
EBRU C., EBAN A., 2011, *Determinants of house prices in Istanbul: a quantile regression approach*, Quality & Quantity, vol. 45, no. 2, pp. 305–317.
FAN G. Z., ONG S. E., KOH H. C., 2006, *Determinants of house price: A decision tree approach*, Urban Studies, vol. 43, no. 12, pp. 2301-2316.
FELDMAN D., GROSS S., 2005, *Mortgage Default: Classification Trees Analysis*, The Journal of Real Estate Finance and Economics, vol. 30, no. 4, pp. 369–396.
FRANCKE M. K., VOS G. A., 2004, *The hierarchical trend model for property valuation and local price indices*, The Journal of Real Estate Finance and Economics, vol. 28, no. 2, pp. 179–208.
HANNONEN M., 2008, *Predicting Urban Land prices: A comparison of four approaches*, International Journal of Strategic Property Management, vol. 12, no. 4, pp. 217–236.
HARRISON D., RUBINFELD D. L., 1978, *Hedonic housing prices and the demand for clean air*, Journal of Environmental Economics and Management, vol. 5, no. 1, pp. 81-102.
KESKIN B., 2008, *Hedonic analysis of price in the Istanbul housing market*, International Journal of Strategic Property Management, vol. 12, no. 2, pp. 125–138.
KIM H. G., HUNG K. C., PARK S. Y., 2015, *Determinants of Housing Prices in Hong Kong: A Box-Cox Quantile Regression Approach*, The Journal of Real Estate Finance and Economics, vol. 50, no. 2, pp. 270–287.
KUSAN H., AYTEKUN O., OZDEMIR I., 2010, *The use of fuzzy logic in predicting house selling price*, Expert Systems with Applications, vol. 37, no. 3, pp. 1808-1813.
MANSFIELD J. R., PINDER J. A., 2008, *'Economic' and 'functional' obsolescence: their charakteristics and impacts on valuation practice*, Property Management, vol. 26, no. 3, pp. 191-206.
MFCR - MINISTRY OF FINANCE OF THE CZECH REPUBLIC, 2013, *Zákon č. 151/1997 Sb. v platném znění pro rok 2014 a jeho prováděcí vyhláška č. 441/2013 Sb.* (Property Valuation Act Nr. 151 of 1997 and its implementing regulation Nr. 441 of 2013 incl. Appendix), Prag.
SANDRONI A., SQUINTANI F., 2013, *Overconfidence and asymmetric information: The case of insurance*, Journal of economic behavior & organization, vol. 93, pp. 149-165.
SEBER G. A. F., LEE A. J., 2003, *Linear regression analysis*, John Wiley and Sons, Hoboken, NJ.
SEDLACIK M., ODEHNAL J., FOLTIN, P., 2015, *Classification of Terrorism Risk by Multidimensional Statistical Methods,* AIP Conference Proceedings 1648 720011.
SELIM H., 2009, *Determinants of house prices in Turkey: Hedonic regression versus artificial neural network*, Expert systems with Applications, vol. 36, no. 2, pp. 2843-2852.
SELIM S., 2008, *Determinants of house prices in Turkey: Hedonic regression model*, Doğuş Üniversitesi Dergisi, vol. 9, no. 1, pp. 65-76.
SIRMANS G. S., MACDONALD L., MACPHERSON D. A. ET AL., 2006, *The Value of Housing Characteristics: A Meta Analysis*, The Journal of Real Estate Finance and Economics, vol. 33, no. 3, pp. 215-240.
SPEYRER J. F., RAGAS W. R., 1991, *Housing prices and flood risk: An examination using spline regression*, The Journal of Real Estate Finance and Economics, vol. 4, no. 4, pp. 395-407.
STENDEBAKKEN M. O. G., GRYTLI E. R., OLSSON, N. O. E., 2015, *Proposed aspects for evaluation of the value of spaces in historic buildings*, Procedia Economics and Finance, vol. 21, pp. 23-31.
THE EUROPEAN GROUP OF VALUERS ASSOCIATIONS – TEGOVA, 2012, *European Valuation Standards*, Gillis, Belgium.
WONG S. K., YIU C. Y., CHAU K. W., 2012, *Liquidity and Information Asymmetry in the Real Estate Market*, The Journal of Real Estate Finance and Economics, vol. 45, no. 1, pp. 49-62.
XU Y., ZHANG Q., ZHENG S. ET AL., 2017, *House Age, Price and Rent: Implications from Land-Structure Decomposition*, The Journal of Real Estate Finance and Economics, vol. 55, no. 1, pp. 1-22.
ZIETZ J., ZIETZ E. N., SIRMANS G. S., 2008, *Determinants of House Prices: A Quantile Regression Approach*, The Journal of Real Estate Finance and Economics, vol. 37, no. 4, pp. 317–333.