Babeș-Bolyai University of Cluj-Napoca

Faculty of Economics and Business Administration

THE PERFORMANCE OF STATISTICAL MODELS IN THE EVALUATION OF REAL ESTATE PROPERTIES IN THE EMERGING COUNTRIES

Ph.D. Supervisor

Full Prof. Habil. Adela DEACONU Ph. D.

Ph.D. Helga – Flavia TOTHĂZAN

2024

Thesis title:

THE PERFORMANCE OF STATISTICAL MODELS IN THE EVALUATION OF REAL ESTATE PROPERTIES IN THE EMERGING COUNTRIES

Summary

A quote says: "A man is made of body, mind and imagination". The body is imperfect, the mind is unreliable, but what has made man remarkable is his imagination". The idea of assisting users of valuation reports (sellers/buyers, banking institutions, courts, town halls) and valuation specialists in overcoming and resolving suspicions about market value is one of the aims of the thesis.

Chapter 1 is the introduction, which incorporates the definition of the terms used and an analysis of the real estate market that is considered "perfect" and for which anomalies have been observed that frequently occur in real estate transactions. These include self-interested clients or clients determined to invest, properties mortgaged (bank, judicial, with state debt), properties with long exposure to the real estate market or urgent sellers. The aforementioned practices have a direct impact on the price, the property is over or undervalued compared to the real value. In these situations, the valuer is not responsible for a wrong estimate (Tothăzan, 2022b).

Furthermore, another important framework is the field of research that defines the market value: The closer it is to the property's sale price, the more accurate it is. On the other hand, the greater the difference between the two values (the price traded on the market and the market value estimated by the valuer), the more flawed the value is considered to be.

In the area of appraisers, the estimated market value is valid on the appraisal date (Valuation Standards, ANEVAR). Thereafter, the value may fluctuate due to a constantly changing real estate market. Moreover, if we turn to the comparative analysis of the definition of the correct market value adopted by researcher, and even by the users of the reports, with that adopted in the field of valuation, we find that the definition of the term does not have the same meaning. In practice, this is an exception or a mere coincidence, since the two values do not overlap.

Eliminating the misunderstanding of what the market value expresses, through careful study, helped us bring clarification and deepening regarding the use, practice and elimination of an expectation that cannot be realized (the market value should be equal to the selling price to have a correct fair market value). We agree with Macdonald and Cannone (2003), who emphasize the need for a theoretical basis for evaluation in order to make an evaluation. The thesis addresses several analysis studies taking into account theory, research works and the practice of evaluators.

Uncertainty in appraisal's work of estimating the market value of the professional appraiser has also raised questions about the methodology approached by the specialists. The thesis approaches a comparative analysis between the traditional method (market approach, income approach and cost approach) with the use of traditional statistical methods (linear regression) and modern statistics from the group of Artificial Intelligence, AI (artificial neural networks, ANN). The thesis took shape in offering a small guide on the steps in approaching statistical models (Tothăzan and Deaconu, 2020).

Towards the end, we present the newer branch of the evaluation field, called evaluation verification (VE). We address the terminology, practice, logic and challenges in this field.

Chapter 2 of the thesis is a foray into the approach of statistical models in the field of real estate evaluation regarding market value estimation. The chapter is a study of the specialized literature that deals with the common elements and particularities of statistical models: linear regression and ANN. The theoretical presentation of these models was made with our attention directed towards the possible applicants who would apply them, the noted information outlines the most complete images regarding the methodology, structures of the models, types of models, challenges and results. In this chapter we present a ranking of the statistical models for the regression model and ANN. A more careful study was given to the statistical indicators, for which examples of works and their results were noted. The chapter ends with the regression analysis and ANN regarding the controversies, assessments or needs that have been signaled.

The theory and analysis of the literature being dissected, chapter 3 continues with the testing of statistical models, regression and ANN, for the context of emerging countries, Romania. Worzala et al. (1995) carried out an analysis to verify the efficiency of regression models with ANN, tracking the performance with which market values are estimated. Similar to the study, the testing was carried out for 900 apartments with 1, 2, 3 and 4 apartments in Cluj-Napoca city (data collected from the ARGUS platform, for July 1 - December 31, 2019). Database normalization tests were carried out, then studies for descriptive statistics, followed by ANOVA and T-Test tests. Finally, the regression and ANN approach took place. The results of the database indicated an asymmetric distribution to the right for which we ran two multiple regression models (gamma and linear model). The multiple linear regression was the one that gave us the best result and was approached for the comparison analysis with ANN, where we followed the performance of the models, the pliability for evaluation and the possibility of a future approach in practice. Some hypotheses with which we started in this study are:

1. ANN and GLM (General linear model) are models that fit in the Romanian context (Cechin et al., 2000; Şipoş and Crivii, 2008; Deaconu et al., 2022) and can estimate correct values for the field of real estate evaluation.

2. GLM is a model with a historical consistency in evaluation, therefore it can be a good tool in estimating the market value, but it is too rigid for the multivariate context of real estate properties.

3. ANN is a model that is similar in reasoning to that of the evaluator and is known to be less rigid; we believe the results will be better for ANN, not for GLM.

4. ANN and GLM are possible tools for application in the review of evaluation reports made for real estate properties.

5. The choice of the independent variables used in the model goes through the niche of several tests regarding the degree of importance, and the close relationship between the variables, and following this decantation, the choice of the variables used in the model takes place. Thus, not all available variables are used for estimation. Our interest is the process of selecting the variables and how many of them can/can't be capitalized.

The results of the study provided answers to the hypotheses noted above:

1. For hypothesis 1, the processing indicates that both models offered good results for the emerging Romanian context and the ability of the models to estimate market values is confirmed.
2. Hypothesis 2, regarding the safety and rigidity of processing the general linear model, GLM, is also confirmed by the present studies.
3. For hypothesis 3, from the comparison analysis of the results, the ANN model has better values, indicating a higher performance than the GLM. ANN has an advantage in that it can mimic the evaluator's reasoning. Another advantage of the ANN model is that it can better capture the details in the database compared to the GLM.
4. For hypothesis 4, we encourage the use of the regression model. GLM is a useful and safe tool. However, the ANN model performs better, but it is in an early stage of approach, it presents elements that are still discussed (for example, black box) and requires a longer testing time. This chapter is a contribution to the need for research regarding the testing of statistical models and the need for research for the emerging context (Deaconu et al., 2022).
5. Hypothesis 5 shows the fact that not all variables available in the database were used to estimate the market value. From the total of 33 independent variables (characteristics of the apartments), 21 were used in the model.

The thesis summarizes the advantages of using statistical models in the evaluation by increasing the evaluator's confidence in his work and by the safety it offers to the large number of users of the reports.

Chapter 4, deepens the theme of the choice and use of independent variables in the model, as well as the further testing of the ANN model. Thus, two questions arise:

1. Does the choice of independent variables have an impact on the market value?

2. If in the estimation procedure, we have made use of every valuable piece of information?

The empirical study on variable selection uses two approaches. The first is related to the arrangement of the characteristics of a variable. Thus studies use two practices. The first is represented by the group of variables, where the characters collected and noted at the beginning are kept. The second practice includes the variables we call regrouped due to the unification of two or more characteristics, caused by the small number of subjects it has and the need for a balance between the number of subjects in the groups. This fact is a benefit for the valorization of the subjects and/or the use of that variable in the model. An example would be Şipoş and Crivii (2008), who for the variable number of building floors formed a group by unifying the apartments on the ground floor with those on the top floor. In this approach, we used the comparative analysis of the graphs of the variables (ungrouped with those regrouped) and the values of the average price.

The results draw our attention to the need to deepen this topic, as it is possible to distort the market value by using several grouped independent variables in the estimation. The use of regrouping variables is a procedure often used in the estimation of market values. McCluskey and Borst (2011) suggested that a database that needs the use of several regroupings of initiators signals the need to divide the database into submarkets.

The second approach in the study of variables used running, two multi-structure ANN models. The goal was to track the performances by differentiating the variables used in the model.

The results indicate the need for a complete database, even if ANN includes the ability to self-eliminate subjects with incomplete information and does not affect the model's ability to process data. The ANN showed the positive impact of the integration of, each variable into the model, being a similarity with the evaluation practice and the recognition of the importance of the more significant variables (for example, the useful surface, the area, the type of compartmentalization).

Last but not least, (5) deals with the verification of evaluation reports. Even the profession of appraisers, Vascu (2017), has pointed out the possibility of large differences between the market value of several appraisal reports issued on the same date for the same property. A concrete example is that of, the cases encountered within judicial institutions. Isakson (1998) is among the few researchers who approach the statistical model as a tool for verifying reports. Thas, inspired by his methodology, we test the ability to verify the weights of statistical models with the tracking of, their utility by integrating them into the price equation.

To add more to my point, two cases of apartments obtained from the Court and Judiciary of Cluj-Napoca are analyzed by applying the weights resulting from the processing of Chapter 3 of the GLM and ANN models in the calculation of the price equation. Thus, there is a part analysis of the results that indicate the usefulness of the weights in checking and the performances of the models, GLM and ANN.

Our expectations were not the ideal ones, due to the fact that differences greater than 5% were obtained between the assumed result and the obtained one. Firstly, this difference we vehemently believe is due to the data differences regarding the use of the independent variables used in the evaluation (100%) and those used by us (58%). More precisely, from the total of independent variables used by the appraiser to estimate the market value, only 58% were used to estimate the market value. This fact is due to the lack of some variables from the model processed by us in chapter 3. Despite all these limitations, we consider the research valuable through its contribution to the studies of report variation.

Moreover, the image of helping the multitude of users of valuation reports, from simple men to various institutions and providing support to the profession of valuers by researching the suspicions discussed regarding the fair market value we believe, has been realized in this work.

To summ it up, in conclusion, fair market value may be a different definition than what we think. The use of statistical models does not replace the value of human reasoning, it is only an instrument that comes to the aid of the specialist. Each evaluation report is a unique art, which involves a diverse rationale, as real estate properties with specific characteristics are analyzed, this fact cannot be achieved by a statistical model. But, the regression approach, the ANN are models that are used in the evaluation of properties in the emerging Romanian context for the estimation of the market value and even for the verification of the reports.

The limits of the work are many thus offering a varied range of research possibilities in the future. The database was 900 subjects, which can be expanded and thus, allows the exploration of the functioning of the models on a larger sample, the division of the database into submarkets, the approach of several ANN architectures, and the capture of the model's behavior against variables.

Another limitation is that our database only contains apartments from one city. In this way, comparative studies can be carried out between several cities with the same significance or expansion with comparative studies between several emerging countries.

The database is composed only of apartments, studies can also be carried out for other real estate properties, for example houses, commercial spaces or studies with several types of properties.

The time for which the data was collected was 6 months, it can be extended and processing can be carried out for several years.

The topic is a vast one that expands rapidly regarding new types of models approached and statistical programs used. There are numerous testing opportunities in this field.

The present research had as its target the correct market value and the testing of statistical models for the emerging context, thus we were limited in the study of variables. There is a need for testing and deepening by approaching several models and capturing the common points, the differences, the problems that the choice, testing and use of variables involve.

Verification of reports using the method suggested by Isakson (1998) is another research topic. We encourage the continuation of testing the usefulness of the weights in verifying the reports, taking into account the limitation we encountered (by the lack of processing of the independent variables that could not be included in the estimation).

Bibliography

1. Canonne J., Macdonald R.J. (2003). Valuation without value theory: A North American "Appraisal". *Journal of Real Estate Practice and Education,* Vol. 6(1), p. 113-162. https://doi.org/10.1080/10835547.2003.12091586

2. Cechin A., Souto A., González A.M. (2000), "Real estete value at porto alege city using artificial neural networks", *Proceedings. Sixth Brazilian Symposium on Neural Networks,* Vol. 1. DOI: 10.1109/SBRN.2000.889745

3. Isakson H.R. (1998), "The review of real estate appraisals using multiple regression analysis", *Journal of real estate research,* Vol. 15(1/2), p. 177 – 190. https://doi.org/10.1080/10835547.1998.12090922

4. McCluskey J.W., Borst A.R. (2011), "Detecting and validating residential housing submarkets. A geostatistical approach for use in mass appraisal", *International Journal*

of Housing Markets and Analysis, Vol. 4(3), p. 290-318. doi.org/10.1108/17538271111153040

5. Șipoș C., Crivii A. (2008), "Modelul regresiei liniare pentru evaluarea proprietăților imobiliare"*, Revista de evaluare*, Vol. 2(5), p. 48-57.

6. Tothăzan H.F., Deaconu A., 2020. Neuronal Network Artificial Model for Real Estate Appraisal: Logic, controversies, and utility for the Romanian context. "Ovidius" University Annals, Economic Sciencis Series, Vol. XX (2), p. 1093 – 1100.

7. Tothăzan H.F. (2022a), "The contribution of statistical models in the field of real estate valuation", Timisoara Journal of Economics and Business, Vol. 15(1), p. 111-126. DOI: 10.2478/tjeb-2022-0007

8. Tothăzan H.F. (2022b), "An accurate market value: from valuation to review of real estate appraisals", *Journal: Review of Economic Studies and Research Virgil Madgearu*, Vol. 15(1), p. 161-175.

9. Vascu A. (2017), "Testul valorii de piață", *VALOAREA, oriunde ar fi ea*, nr. 15, p. 10-17, ANEVAR, București.

10. Worzala E. M., Lenk M. M., Silva A., 1995, *An explanation of neural networks and its application to real estate valuation*, Journal of Real Estate Research, Vol. 10, No. 2, p. 185-202.