STATISTICAL MODELING APPLIED IN REAL ESTATE VALUATION

Silvia ZEICU (CHIOREAN), PhD Student, Faculty of Geodesy, Technical University of Civil Engineering, silvia.zeicu@icloud.com
Dumitru ONOSE, Prof.dr.eng., Faculty of Geodesy, Technical University of Civil Engineering, Bucharest
Mircea ORTELECAN, Prof.dr.eng., Faculty of Horticulture, University of Agricultural Sciences and Veterinary Medicine of Cluj-Napoca
Maricel PALAMARIU, Prof.dr.eng., Faculty of Horticulture, University of Agricultural Sciences and Veterinary Medicine of Cluj-Napoca

Abstract: *The use of statistical models and special applications of statistical processes provides extra support for developing an opinion on the value or analysis. This article presents the method of obtaining statistical models that help stakeholders to better know the specific real market estate. Currently, specialist appraisers use various statistical methods for appraising real estate, a multitude of solutions provided by the mathematical statistics to be applied in this field: MRA - multiple regression analysis, time series, fuzzy logic techniques, ANN – artificial neural networks, NFS - neuro-fuzzy systems and others. The immediate and very palpable advantages of applying these valuation solutions are the automatization of the appraisal process and the existence of control methods for the results achieved. The main disadvantage is the complexity of applying the suggested solutions, requiring a team of specialists from different fields - programmers, statisticians, mathematicians, appraisers, market analysts - to develop and operate systems of this nature and scale.*

Keywords*: real estate appraisal, statistical models, automated valuation model;*

1. Introduction

In real estate appraisal, the quality and quantity of data available for analysis are just as important as the methods and techniques used to process them and to complete the valuation mission. Therefore, the ability to discern between the multitude of information, to look for sources of reliable data and organise the volume of information is of essential importance for the valuation practice.

Data management about the market, real estate and transactions, required in the appraisal process, involves three steps: data collection; organising data collected, and analysing them. For appraisers, understanding statistics has become important with the merger of the valuation methodology of each real estate separately with the overall appraisal techniques. The traditional real estate appraiser plays an important role in this changing environment, bridging pure mathematical models used in the valuation, local market conditions and the analysed physical real estate. The use of statistical models and special applications of statistical processes provides extra support for developing an opinion on the value or analysis.

A great variety of mathematical techniques is used in the field of real estate appraisal, starting from simple arithmetic and algebraic formulas up to techniques for multiple regression statistical analysis. Given the continuous upgrading of the computational tools, i.e. computers, the use of sophisticated techniques is increasing in the valuation practice.

2. Additions and Results

Statistics in Real Property Appraisal

Statistics was defined as information science which “involves collecting, classifying, synthesizing, organising, analysing and interpreting numerical information”. According to this definition, the traditional methods of appraisal have always been part of the statistics. Although the study of information was always positioned at the core of the valuation process, the tools required for rigorous analysis and interpretation of numerical data gave rise to expectations in terms of obtaining a more “statistical” overall to draw conclusions on the valuation. Therefore, the profession has evolved into the direction in which the appraisers are expected to have solid knowledge of statistics, statistical models and automated valuation model (AVM).

In general, statistical applications are classified into two categories: descriptive statistics and inferential statistics. Descriptive statistics deals with the use of charts, graphs and scales of synthesising data to describe a data sample or population. Inferential statistics involves using the sample data as the basis in formulating opinions (inferences) about a population represented by that sample. Statistical inferences can include, among others, the estimation of the core tendency and dispersion of current but unknown population, the result prediction and structure detection underlying the cause-effect relationships.

The development of personal computers, spreadsheet programs and statistical processing programs allows appraisers to incorporate, easily and precisely, the statistics in their analyses and valuation reports. In the early years of the advent of personal computers, statistical analysis was generally limited to generating descriptive statistics and appropriate diagrams, tables and graphs. As operating systems begun to predominate graphical interfaces, statistical programs such as SPSS, Minitab and SAS have become more friendly, especially as it was no longer necessary for the user to write programming codes.

Moreover, as computer users have become more sophisticated, new statistical tools were added to spreadsheet programs to cope with the needs. Currently, Microsoft Excel includes a set of statistical tools that generate the results of correlation matrices, F variant tests, t-tests for means and linear regression models. However, Excel offers very little in terms of diagnosis indicators that accompany these statistical tools. The statistical power of Excel still lies in its ability to execute graphics.

Although modern statistical analysis programs are easy to handle, their use can contribute to achieving implausible results when the necessary steps are omitted to ensure the creation of a credible model.

Overall valuation - statistical valuation models

At present, specialist appraisers use various statistical methods for appraising real estate, a multitude of solutions provided by the mathematical statistics to be applied in this field: MRA - multiple regression analysis, time series, fuzzy logic techniques, ANN – artificial neural networks, NFS - neuro-fuzzy systems and others.

The immediate and very palpable advantages of applying these valuation solutions are the automatization of the appraisal process and the existence of the control methods for the results achieved. The main disadvantage is the complexity of applying the suggested solutions, requiring a team of specialists from different fields - programmers, statisticians, mathematicians, appraisers, market analysts - to develop and operate systems of this nature

and scale.

AVM automated valuation models - mathematical software that produce an estimate of the market value based on the market location analysis, market conditions and real estate features of market information previously collected.

Creating an AVM - Automated Valuation Model can be a solution for appraising real estate collateral portfolios held by banks. AVMs are software that produce estimators of market values for the subject real estate, based on the analysis of market conditions and characteristics of comparable real estates, market information collected in advance. AVM are particularly applicable in the real estate appraisal through direct market comparison.

AVM credibility and accuracy of its results depend on the quantity and quality of the data used in the valuation, experience and training of the staff that designs and develops the model. The amount of data refers to the volume of sample data used in the valuation. As regards the quality of data in the process of developing an AVM, a special stage for the management and analysis of the data quality is designated.

Given the impossibility of using the entire data population (entire portfolio), it is necessary to extract a representative sample of the whole population. A sample is representative of the entire population if its structure is identical or very similar with the structure of the population where it has been extracted so that it is possible to extrapolate the findings to the entire population. The sample will be representative when using an objective, random sampling method, with anticipated probability and if the sample has a sufficiently high volume of data.

This process is presented in the Standard automated valuation models developed by the International Association of Assessing Officers, a complex process that requires close cooperation between the appraisers, real estate market analysts, statisticians and software developers.

Starting from the recommendations of this standard, we propose the following sequence of steps needed to build the automated valuation model:

1. identifying the subject real estate (portfolio of bank guarantees to be valuated);
2. establishing the extraordinary assumptions, hypothetical and limiting conditions;
3. data management and data quality analysis;
4. sample stratification (if it was not built by stratified sampling);
5. establishing the model specifications;
6. model calibration;
7. model testing and quality assurance;
8. model validation;
9. application of the model;
10. regular verification of the model accuracy.

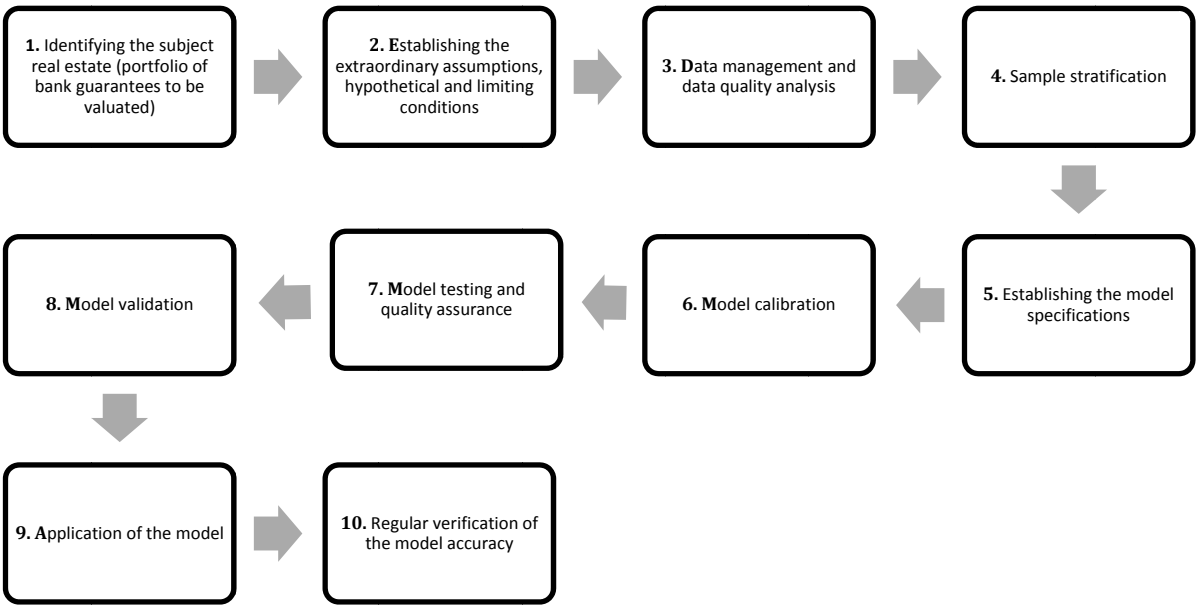

Fig. 1. Standard for automated valuation models developed by the International Association of Assessing Officers (IAAO)

Identification of the subject real estate (bank guarantee portfolio to be valuated)

The database contains a total of 46 transactions performed in Cluj-Napoca City, Marasti District, Cluj County. Cluj-Napoca is one of the most important cultural, academic, medical and economic centres in Romania.

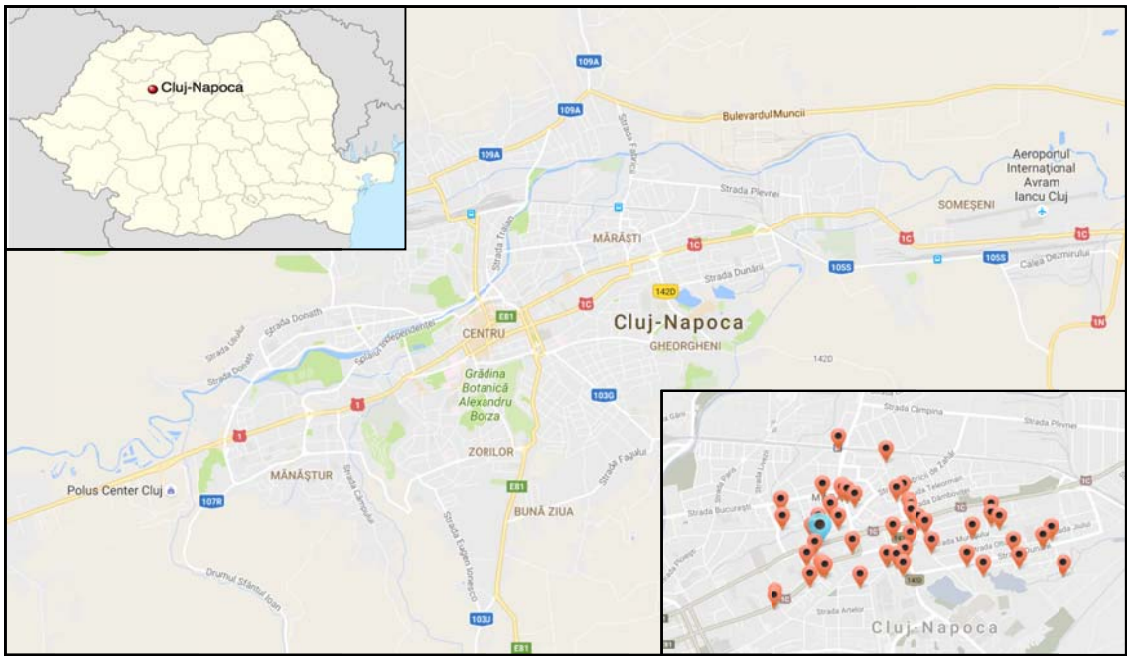

Fig. 2. Location of subject real estate

Establishing the valuation assumptions

The database consists of real estate transactions conducted in August-September of the current year, and includes 2-bedroom apartments, with areas ranging between 40 and 65 sq m. For creating the database, it was taken into account that the dependent variable is the trading price of the apartments in EUR and the independent variables are the usable area of the real estate, the number of rooms, floor and their finishing according to the table below:

Table 1. Description of data base variables

Type of real estate	Apartment
Area of real estate	Marasti District, Cluj-Napoca City, Cluj County
Number of rooms	2
Useful floor area [sq m]	40-65 sq m
Floor	Ground / intermediate / last floor
Finishing	low/average/high
Transaction price	Stated in EUR
Unit transaction price	Stated in EUR / sq m

Data management and data quality analysis

In support of this stage, statistical analysis software provides graphical techniques and quantitative techniques to explore the data series. These techniques can be used to establish the representativeness of the sample to identify the frequency distribution function and its characteristics or to detect outliers in the series analysed.

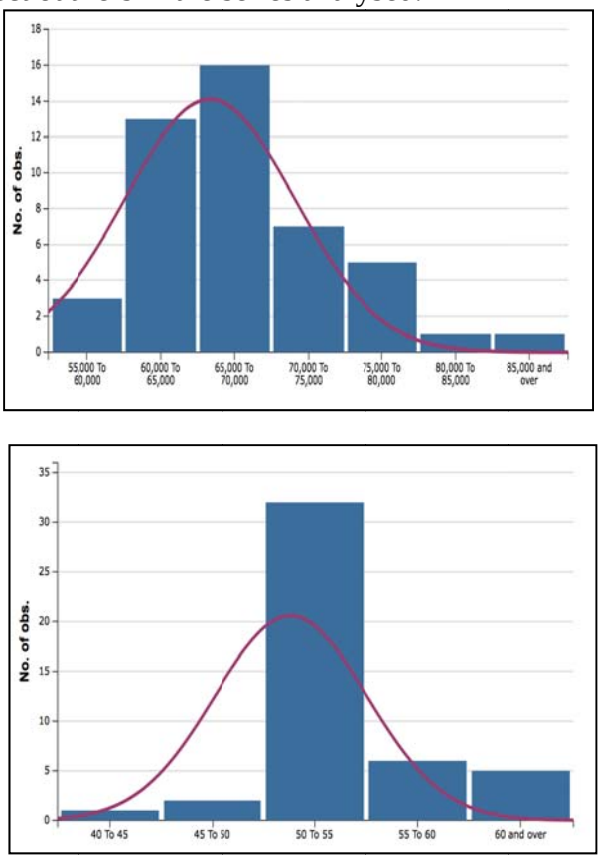

Fig. 3. Frequency of distribution of real estate depending on the sale price and their surface

Count	46	Mean Deviation	3.09546
Mean	53.58696	Second Moment	19.41635
Mean LCL	52.26396	Third Moment	34.67334
Mean UCL	54.90995	Fourth Moment	1,830.36039
Variance	19.84783		
Standard Deviation	4.45509	Sum	2,465
Mean Standard Error	0.65687	Sum Standard Error	30.21589
Coefficient of Variation	0.08314	Total Sum Squares	132,985
		Adjusted Sum Squares	893.15217
Minimum	40		
Maximum	65	Geometric Mean	53.40705
Range	25	Harmonic Mean	53.22664
		Mode	52
Median	53		
Median Error	0.12138	Skewness	0.40527
Percentile 25% (Q1)	52	Skewness Standard Error	0.34240
Percentile 75% (Q3)	54	Kurtosis	4.85514
IQR	2	Kurtosis Standard Error	0.64247
MAD (Median Absolute Deviation)	6	Skewness (Fisher's)	0.41906
Coefficient of Dispersion (COD)	0.05455	Kurtosis (Fisher's)	2.21650

Count	46	Mean Deviation	5,044.06427
Mean	67,458.47826	Second Moment	41,316,104.20605
Mean LCL	65,528.57658	Third Moment	2.1561E+11
Mean UCL	69,388.37995	Fourth Moment	6.2E+15
Variance	42,234,239.85507		
Standard Deviation	6,498.78757	Sum	3,103,090
Mean Standard Error	958.19395	Sum Standard Error	44,076.92178
Coefficient of Variation	0.09634	Total Sum Squares	2.1E+11
		Adjusted Sum Squares	1,900,540,793.47826
Minimum	57,000		
Maximum	87,000	Geometric Mean	67,164.20574
Range	30,000	Harmonic Mean	66,880.93296
		Mode	63,000
Median	66,495		
Median Error	177.06570	Skewness	0.81187
Percentile 25% (Q1)	63,000	Skewness Standard Error	0.34240
Percentile 75% (Q3)	70,000	Kurtosis	3.64642
IQR	7,000	Kurtosis Standard Error	0.64247
MAD (Median Absolute Deviation)	8,000	Skewness (Fisher's)	0.83950
Coefficient of Dispersion (COD)	0.07555	Kurtosis (Fisher's)	0.86531

Fig. 4. Statistical analysis of data quality

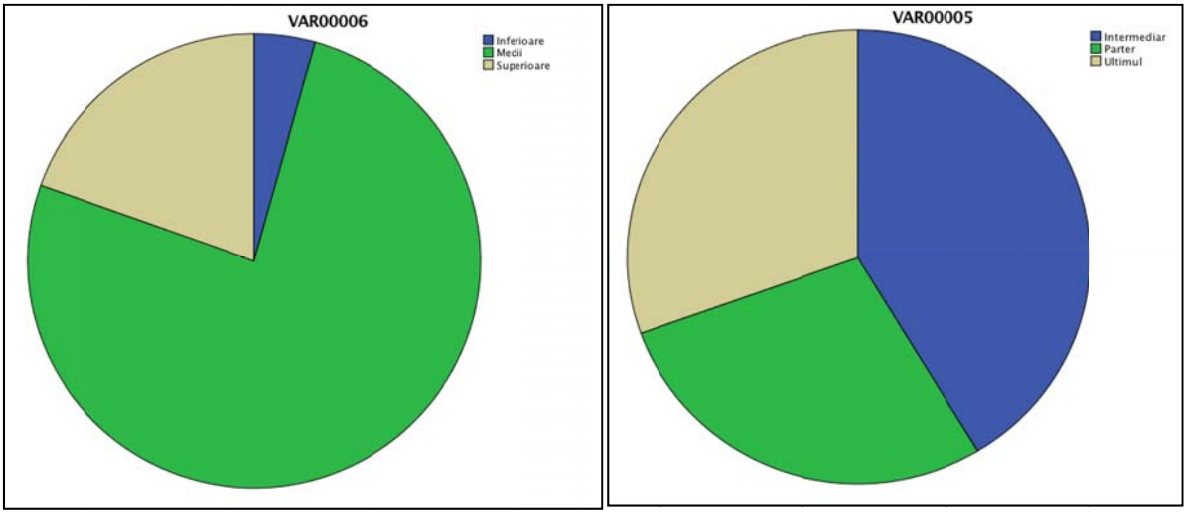

Fig. 5. Statistical analysis of real estate by finishing and by floor

Stratification

It is the process by which the real estate forming a heterogeneous sample are grouped by type of comparable real estate according to their physical and economic characteristics in order to minimise the differences within a layer and maximise the differences between layers.

Establishment of model specification

The appraiser may select from the multitude of variables that characterise the real estate analysed (potential explanatory variables), those which could have an economic impact on the real estate value (dependent variable). The higher the number of explanatory variables selected, the higher the risk that the information contained therein becomes redundant. On the other hand, the inclusion of a too low number of explanatory variables could cause the regression model not to explain enough the true evolution of the dependent variable, which would lead to a shift of estimators (regression coefficients).

The choice of explanatory variables must be made by the appraiser, not only in terms of size of the correlation coefficients with the dependent variable, but also by taking into consideration the economic significance that could characterise the relationship between the explanatory variable and the dependent variable.

Moreover, it examines whether there is a strong correlation between the selected explanatory variables. Where such a correlation is identified, the analysis will include only one of the explanatory variables in the group of the interrelated. One can opt for the strongest correlated with the dependent variable. The more the correlation coefficient value matches 1 or -1, the stronger the variables are correlated.

Correlations

		Pretvanzare	Suprafata
Pretvanzare	Pearson Correlation	1	.615**
	Sig. (2-tailed)		.000
	N	46	46
Suprafata	Pearson Correlation	.615**	1
	Sig. (2-tailed)	.000	
	N	46	46

**. Correlation is significant at the 0.01 level (2-tailed).

Fig. 6. Calculation of correlation coefficient

The correlation coefficient was calculated in order to determine the correlation between the sales price dependent variable and the usable floor area variable. If its value is close to zero – the two variables show a very weak linear dependence, if it is zero - the variables are independent, if it is close to 1 - variables have a direct linear dependence (positive) and if it is close to -1 - inverse linear dependence (negative).

Model calibration

Among the methods of calibration, the most used are those based on statistical methods such as multiple linear regression and nonlinear regression.

The term regression is used to describe any process in which one or more known variables (characteristics of the analysed real estate) are used to determine an unknown variable (in this case the market value of the real estate appraised) or to predict its value.

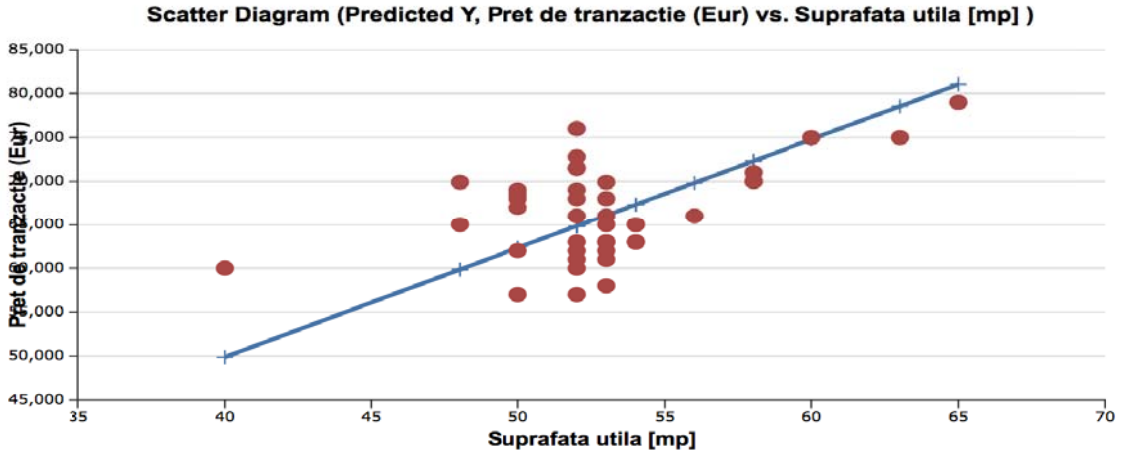

Fig. 7. Scatter Diagram - price/surface

Regression Statistics	
R	0.99728
R-square	0.99457
Adjusted R-square	0.99457
S	4,985.34531
N	44

Pret de tranzactie (Eur) = 1,246,69698 * Suprafata utila [mp]

Linear regression coefficient r shows that there is a strong dependence between the two variables, i.e. between the transaction price and the useful floor area of the apartments. R square determination coefficient is 99%. The standard error indicates that the market value predicted by the model established may be different by EUR 4,985.34 from the average apartments sold.

Model testing and quality assurance

Model testing is intended to determine whether it ensures the accuracy and integrity necessary for the estimates so achieved to be extrapolated by statistical inference to the entire population of a bank guarantee. For this purpose, a set of data on the real estate is used, for which the sale prices or market values are known, based on which the testing shall be performed. This set of data, used to test the model, was the basis for the execution and calibration of the model and represents 80% of the sample volume (n).

As stated above, there are a variety of statistical tests that can be applied to obtain the necessary information, some of them parametric, applicable to the sets of data with normal frequency distribution, such as Student (t), Fisher (F) Testing and other non-parametric for series that do not have normal frequency distribution, such as Chi-square, Shapiro-Wilk, Kolmogorov-Smirnov, Man-Withney, Durbin-Watson tests.

In practice, the model specification and the calibration stage is performed iteratively:

model specification - testing model specification by calibration - model specification correction - testing the specifications corrected by repeating these operations until obtaining the confidence level designed.

Model validation

Once the model was calibrated and tested, its validation is performed based on the set of data representing the remaining 20% of the sample, using statistical tests.

Applying the model

The automated valuation model takes into account, as values of the explanatory variables, the valuescorresponding to the characteristics of the subject real estate and their market value is estimated as such.

Periodic verification of the model accuracy

It is recommended to check the model accuracy by comparing the estimated market values to those estimated using the valuation methods and techniques recommended by IVS. The large differences existing between the two estimated values are a signal of the need to update both the analysis sample and the automated valuation model.

3. Conclusions

The advantages of the statistical modelling in real estate appraisal include:

- it allows the use of unlimited data series;
- it allows the use of variables in any high number if they are necessary and relevant;
- it allows to determine the estimation accuracy of the value that can be exceptionally good when using sufficient market data;
- it ensures uniform appraisals;
- the resulting model can be used for both targeted valuations (singular) and overall valuations;
- the automatization of the valuation process, therefore little time for processing and analysing data and determining the conclusion of the appraisal;
- objectivity and efficiency;

These advantages can be achieved only with the basic condition that the model be substantiated on a very serious and coherent statistical and mathematical reasoning, the database must be constantly updated, the model must be tested before it is actually used in the appraisal.

Statistical modelling disadvantages are:

- it cannot be applied to restricted markets, with few transactions or to atypical markets;
- since the results and conclusions arising from the application of the statistical modelling are valid only in the population under study, the appraisal of real estate atypical for the population is not credible;
- the need of highly developed stock of knowledge in mathematical statistics and specialised software.

4. References

1. *T. Andrei, S. Stancu, D. Traian Pele: Statistica. Teorie şi aplicaţii, Editura Economică, ediţia a doua, 2002*
2. *Bruce L. Bowerman, Richard T. O'Connell, McGraw Hill: Business Statistics in Practice, fourth edition, 2007*
3. *Evaluarea proprietatii imobiliare, Editia a 13-a, Bucuresti 2011*
4. *Standarde Internationale de Evaluare a bunurilor ANEVAR 2016*
5. *Evaluarea proprietatii imobiliare, Editia a 12-a, Bucuresti 2011*
6. *Mass Appraisal Methods – An international perspective for property valuers, Wiley-Blackwell, 2008;*
7. *Practical Applications in Appraisal Valuation Modeling, Statistical Methods for Real Estate Practitioners, M.S. Kane, M.R. Linné, J.A. Johnson, Appraisal Institute, 2004;*
8. *Statistică aplicată pentru modelarea evaluării imobiliare, Iuliu Paşca, Editura Politehnica, 2008 pg. 16*
9. *Standard on Automated Valuation Models (AVMs), IAAO, 2003*
10. *Lecture notes - Real estate appraisal and urban planning, I. Bene*
11. *Various articles and Internet*