The Role of Model Calibration in Mass Real Estate Assessment: A Case Study of the Republic of Moldova

ALEXANDRU PALADI
Congress of Local Authorities from Moldova
Columna, 106A, Chișinău
THE REPUBLIC OF MOLDOVA
ORCID: 0000-0001-9257-6266

Abstract: The purpose of this article is to highlight the essential role of the calibration stage in mathematical modeling of real estate value for the ad valorem tax system. The analysis of a large-scale valuation project in the Republic of Moldova highlights the need to apply calibration measures to models created by specialized automated programs. Errors and deviations in results obtained from artificial intelligence bring the calibration stage to the primary level to ensure accurate and reliable valuation, reflecting the fundamental dynamics of the real estate market.

Key-Words: analytics, quality control, mathematical model, massive valuation, real estate, model calibration

Received: March 9, 2024. Revised: November 11, 2024. Accepted: December 13, 2024. Published: January 29, 2025.

1. Introduction

The ad-valorem system, used to determine the market Value as the taxable basis for real estate taxation, involves the massive valuation of real estate at a national or regional level. In this context, statistical techniques for processing voluminous market data play an essential role in the development of property valuation models. Currently, several specialized automated programs are used to calculate value factors and generate mathematical models for real estate valuation. At the same time, the concept of Artificial Intelligence and Machine Learning using in massive property evaluation has gained popularity, providing advanced and practical solutions.

However, practical experiences have highlighted certain methodological discrepancies in the results obtained by using the automated statistical techniques. A demonstrative example is presented from the practice of the massive evaluation of real estate in the Republic of Moldova. The detected discrepancies are caused mainly by the specifics of the local market, the assessment object and the transparency of the market data. To ensure a closer alignment with the realities of the real estate market, the need for a periodic analysis and recalibration of the results is required by these divergences. Despite the technological advances, it is essential that these tools must be complemented by careful human review and contextual adjustments so that the results accurately reflect the market Value of real estate.

2. Problem Formulation

Any type of real estate evaluation, whether individual or massive, can be expressed through a hedonic mathematical model, which represents the relationship between the property's Value as the dependent variable and the value factors as the independent variables that influence this Value. Currently, real estate value modelling is widely used in massive real estate valuation. The generated model determines the Value for a specific property type, in a particular area, region or country.

According to established international practice and specialized guidelines [1], the first phase of modeling involves determining the model type and identifying the key factors that influence Value, based on statistical data from real estate transactions. The basic classification of assessment models includes additive, multiplicative, and hybrid types. The primary classification of assessment types determines the additive, multiplicative and hybrid models. Each of the models includes several subtypes depending on the category of valued goods, the composition of valued objects and the variation of these characteristics. The key factors influencing property value include technical and economic characteristics of the assessed property, such as total or usable area, building volume, number of parking spaces, number of bedrooms, and similar attributes.

To assess the impact of each value factor on the property's value, an analysis of inter-factor regressions is conducted. This analysis relies on statistical calculations to explore the relationships between multiple factors (independent variables) and a dependent variable in situations where these factors are also interrelated. Several statistical indicators are used to measure the significance of a variable in the model, including Pearson's correlation coefficient, Spearman's rank correlation, and Kendall's tau statistic. At the same time, statistical indicators are used to determine multicollinearity between independent value factors: Variance Inflation Factor (VIF), Correlation Matrix, Durbin-Watson test, Covariance, etc. [2]

Provided there is a sample of representative, correct and accurate statistical data regarding the current transactions related to the category of immovable property evaluated, the generation of the model can be deduced by statistical methods using specialized electronic systems. In this context, leading specialists [3] have put forward several proposals for incorporating Artificial Intelligence (AI) and Machine Learning (ML) techniques. However, according to daily practice, the full application of statistical methods does not bring the complete accuracy of the results and requires further revision. The reasons for the complete dissatisfaction with the results are different, among them are highlighted:

1. Price dumping in real estate sales contracts, which is a significant aspect that affects the accuracy of real estate evaluations and can distort the results of statistical models used in massive evaluations. Price dumping refers to real estate transactions conducted at prices significantly lower than market value, often for economic or strategic reasons, or to obtain fiscal or competitive advantages. In the context of the real estate market of the Republic of Moldova, dumping has a massive character. It discloses the tendency of the contracting parties to evade the fiscal burdens regarding the state tax and the tax on capital growth, calculated according to the price of the transactions.
2. The problem of complexity and differentiation of market data. The real estate market includes a wide range of property types, each with unique characteristics that influence their Value. In a massive assessment model, it is difficult to fully capture all these differences, which can lead to imprecise estimates. Moreover, the cadastral data used in the Republic of Moldova for the massive evaluation are not updated regularly and are distinguished by the low level of correctness.
3. The issue of limited transaction data samples is a critical challenge in the large-scale real estate valuation process, as insufficient data can decrease the accuracy of statistical models and skew property value estimates. This problem becomes even more acute in smaller or less active real estate markets, where the number of transactions is low and does not provide a solid basis for modelling.
4. The dynamic evolution of the real estate market. The Value of real estate is influenced by economic, political and social factors that can undergo rapid changes. Statistical models based on historical data may not capture these changes in real time, resulting in inaccurate estimates. Specifically, unforeseen events such as economic crises or legislative changes can skew the results of these statistical models.
5. Collinearity between variables. In many cases, the value factors used in the models are correlated, which can affect the reliability of the results. For example, location and infrastructure can be highly correlated factors, which can create problems in interpreting the contribution of each factor in determining real estate Value.
6. Data quality and granularity. Data available for evaluation may not be sufficiently detailed or up-to-date. In addition, if not collected and processed correctly, the data may contain errors that affect the reliability of the model. The lack of granularity can lead to models that do not account for subtle differences between properties, such as building conditions or recent improvements.

7. There may be an incompatibility between the model and the characteristics of the market. Statistical models and machine learning algorithms can perform differently based on the type of market and its specific conditions. A stable real estate market will react differently to a volatile one, which requires specific adjustments depending on the kind of properties and regions analyzed.

These limitations emphasize the need for additional human intervention and manual adjustment of statistical results to reflect market reality and local specificities in a better way. In general, the role of mass assessment model calibration consists of the following critical aspects:

1. Parameter Adjustment: Calibration involves adjusting model variables and parameters to better reflect market reality. Each real estate market has unique characteristics (e.g., price trends, supply and demand), and model calibrations must take these characteristics into account.

2. Improving accuracy: The primary objective is to minimize the discrepancies between estimated values and observed market values. By calibrating the models using sales data and other real-world

information, the accuracy of the estimates can be significantly improved.

3. Ensuring fairness: In a tax assessment system, it is important that all properties must be treated fairly. Calibration can help mitigate the risk of undervaluing or overvaluing specific categories of real estate or geographic regions, thereby preventing inequalities in tax assessments.
4. Improving resilience: The real estate market is dynamic, and models must be calibrated periodically to remain relevant in the face of economic, social and legal changes that influence real estate values.

2.1 The case study

A demonstrative example that highlights the critical aspects of calibrating models generated by statistical methods can be seen in the results of the real estate massive evaluation in the Republic of Moldova. The national program for assessing and revaluating real estate for tax purposes was initiated as part of the Financing Agreement between the Republic of Moldova and the International Development Association to implement the Land Registration and Evaluation Project [4]. As outlined in the specifications, the project received support from international experts engaged as consultants to analyze the local market and develop valuation models for various categories of real estate.

To create these models, the cadastral authorities collected data on approximately 120 thousand real estate transactions from 2016-2023, covering several types of properties. In the case of urban residential houses, about 26 thousand transactions were analyzed. The traded objects were described by about 40 technical and economic parameters, recently classified and stratified. Thus, following this process, the evaluation model report for urban residential houses generated a logarithmic regression model used for evaluation.

$$V = Exp(12{,}98 + (SE \times 0{,}0066) + (SE^2 \times (-9{,}\times 10^{-6}) + Ag \times (-0{,}0038) + \sum_1^n (FT_i \times KFT_i) + KL_a); \quad (1)$$

The mathematical equation expresses the relationship between the Value of residential houses (V) and the influencing factors in a logarithmic form, having as variables: the external surface (SE), the construction age (Ag), the corrections for binary technical factors (FTi) (such as the existence the aqueduct, the sanitary block, and the technical condition of the construction), and the location coefficient (KLa), which reflects the location of the property according to the locality and the value zone. When developing this basic regression, approximately 15 thousand real estate transactions with urban houses carried out between 2021 and 2023, were analyzed. Specialized software, RStudio, was utilized to conduct the regressions. [5].

To test the proposed model, it underwent an evaluation that assessed the following: the model's ability to appraise all objects within the specified category, the vertical and horizontal equity of the assessments, the logical consistency of the regression concerning the included factors, and the incorporation of the most relevant value factors in line with local market trends. Following testing, certain shortcomings of the model were identified, which are detailed below.

1. The proposed basic regression treats the object of assessment as an inseparable element without distinguishing between the land and the dwelling house on this land. This approach does not respect the principle of separation of the object of taxation, which is essential in situations where the land and the building belong to different tax subjects. Furthermore, the absence of the 'related land area' variable in the model prevents the Value from being adjusted for land size. This structural problem highlights that the model does not reflect the reality of the real estate market, where land and buildings contribute distinctly to the total Value. Also, the synergy effect between the Value of the land and the construction is omitted since the sum of the Values of these components does not always coincide with the market Value of the entire property [6]. Thus, the structure and form of the model do not correspond to the reality of the real estate market."
2. The proposed version of the automated model includes adjustments for only seven localities out of the country's 64 cities. In addition, each locality is divided into distinct regions with different indicators, resulting in approximately 250 value zones that were not adequately represented in the model. The concentration of real estate transactions in large cities led to the distortion of the statistical results, affecting the model's validity for less developed localities. For example, in small villages, a factor such as the existence of a well-paved access road significantly increases the attractiveness of a house. In contrast in cities this factor has a negligible influence. The same applies to building infrastructure facilities, which play an essential role

in the rural environment but have minimal impact in the urban environment.

This improper aggregation of market data, coming from localities with different levels of development, reduces the relevance and correct interpretation of influencing factors. Differences in land and construction market dynamics between regions also have a significant impact. In less developed areas, land is cheap and abundantly available, while in areas with a strong economy, it is expensive and scarce. In contrast, the Value of construction depends on fluctuations in the prices of construction materials and services. Unifying the equation for calculating the location coefficient without adjusting these differences between the land and construction markets generates erroneous results. This problem can be illustrated graphically, highlighting the different shares of land Value and buildings in various cities.

Fig. 1 The share of land Value related to residential houses and buildings in different cities

The example presented clearly shows that the dynamics of land Value evolve differently compared to constructions (houses) in various localities. For instance, in central or developed regions, the increase in land size is proportionally reflected in the selling price. Conversely, in outlying cities, land area variation has little impact on land Value. At the same time, the Value of capital constructions has a noticeable tendency to increase according to the size of the house, regardless of the geographical location. This phenomenon is influenced by various social and economic factors that affect the local housing market. Among the main factors influencing this behavior are:

- The intra-urban land reserve administered by the local authorities, intended for the development of housing space, as well as the policy of the local public administration in the direction of regional development;
- The number of the population in the locality and related aspects, such as migration and demographic structure (age, gender, etc.);
- Development of building infrastructure in the region, including access to roads, water, sewage, gas and other utilities;
- The culture, traditions and occupations of the population, which influence how households are managed, the maintenance of domestic animals, and the use of products from their garden.

1. The model was tested for the extreme Values corresponding to the data set in the category of goods evaluated. The variable factors used in the test were: -Test A, which assessed the influence of the external surface (between 5 sq m and 1200 sq m) and; -Test B, which analyzed the value adjustment coefficient according to the age of the base construction. The results of these tests are shown in the figures below.

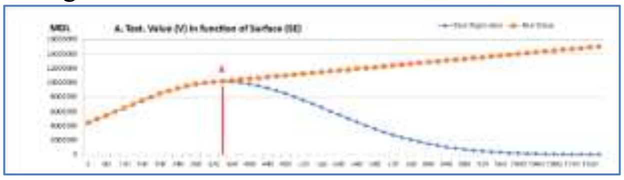

Fig. 2 Model testing results on the regression of Value and Outer Area (of the main construction)

The diagram above highlights the fact that the base regression (Base Regression curb) of the proposed model indicates a decrease in the Value of the evaluated object after reaching a critical point A (approximately 340 sqm), and the Value continues to decrease significantly after approximately 800 sqm. At the same time, the curve on the chart reflects the increase in Real Value (Real Value curb) as a function of area. This deviation is mainly determined by the low representativeness of the data on objects with large areas. The distribution of observations used in modelling, by surface category, is presented in the table below.

Tab. 1 Share of observations on townhouse transactions stratified by exterior area

Area, m^2	0÷ 80	81 ÷ 120	121 ÷ 160	161 ÷ 200	201 ÷ 240	241 ÷ 280	281 ÷ 320	> 320
Transactions	4550	5312	2459	1279	787	393	246	2
Rate, %	30,2	35,2	16,34	8,50	5,23	2,6	1,6	0,02

The small number of observations for large houses, the specificity of transactions with such properties, as well as the principle of diminishing returns led to an anomaly in the model. This situation was interpreted by the program as a dynamic decrease of the Values according to the area after a set critical point.

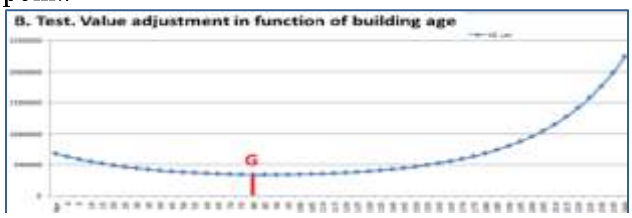

Fig. 3 The results of testing the Adjustment Coefficient for the age of the main building.

The results of model testing of the change in Value with the age of the main building show a wavy curve, with a critical point (G) at about 80 years of age. According to this model, the Value of homes older than this point increases, an assumption that flies in the face of market reality. The observed divergence, generated by the specialized software, is due to the focus of observations on the segment of houses up to 80 years old and the specificity of transactions with old houses. Most of the old townhouses are located in central areas, and these are generally purchased, due to the location and related land, for later commercial use. Thus, the attractiveness of the land is reflected in the high Values of the old houses. Still this aspect was erroneously interpreted by the software, transmitting the Value to the buildings themselves, which, in reality, are of no interest to buyers and are demolished or wholly rebuilt as a result of transactions.

These results reflect a common problem encountered in the massive estimation of residential objects. Given these findings, the models generated by the automated program required revisions and adjustments by recalibrating them. Thus, based on the evaluation model of urban houses in the Republic of Moldova studied in this work, structural changes and adjustments to the composition of the model were introduced.

3. Problem Solution

Model testing results from automated programs reflect a common problem encountered in the massive estimation of residential objects. Given these findings, the models generated by the computerised program required revisions and adjustments by recalibrating them. Thus, based on the evaluation model of urban houses in the Republic of Moldova studied in this work, structural changes and adjustments to the composition of the model were introduced. The structure and shape of the model have been revised. The component elements of the evaluated object were treated separately and later integrated into a unique equation:

$$V = (VC + VTC + VTL) x KAjC; \quad (2)$$

Where the Value of the real estate (V) is determined in the additive function of the Construction Value (VC), the Construction Land Value (VTC) and the Free Land Value (VTL). Common adjustment (AJC) refers to the common factors that affect the Value of real estate, including location.

The Value of the related land (VTC) and the free land (VTL) were edited being interchangeable calculated by separate polynomial nonlinear regression equations according to the Land surface (S) adjusted by the coefficients (K).

$$V_{TL} = (Int.TL + S_{TL}*KS_{TL}1 + S^2_{TL}*KS_{TL}2 + S^3_{TL}*KS_{TL}3 + S^4_{TL}*KS_{TL}4); \quad (3)$$

$$VTC = Ln(S_{TC})*KS_{TC}1 + (Ln(S_{TC}))^2*KS_{TC}2 + (Ln(S_{TC}))^3*KS_{TC}3 + (Ln(S_{TC}))^4; \quad (4)$$

The construct Value model was supplemented with independent variables, resulting from separate analyses. The relevant numerical factors were included as independent variables in the respective non-linear regressions: the logarithmic function for the building area (Ln(S)) and the 3rd-degree polynomial function for the age of the base building (P3(Age)). The value factors associated with individual technical and economic parameters are placed in the equation $\sum(F_i \times K_{Fi})$. The multiplicative adjustments related to the degree of construction completion (GF) and the difference in the development dynamics of the land market for each locality (KC) were attached to the created equation.

$$VC = \mathrm{Exp}(\mathrm{Ln}(S) + P_3(\mathrm{Age}) + \sum(F_i \times K_{Fi})) \times GF \times KC); \quad (5)$$

By replacing the regressions of the value factors, splitting the land value equation, and filling the equation with other distinct factors, the inconsistencies found in the primary model were reduced.

The final version of the model was submitted to the assessment authority for approval and publication [5]. Quality testing of the calibrated model and public consultations with the private sector reported a high degree of accuracy and reliability.

4. Conclusion

Following what has been studied, it can be concluded that when using specialized automated programs, a careful and extensive analysis of the market is additionally required to both specify and calibrate a model that accurately estimates Values. The results generated do not diminish the role of using automated software or technical AI/ML. They only prove that the products generated by the machine are limited by the information included in it and sometimes do not correspond to the real trends of the analyzed market. Models developed based on statistical data can be problematic if the selected variables are not well-validated or representative of the market Value. Because models predict market Value directly, they depend more on careful specification and calibration of their type and composition. Since the same model is directly applied to the valuation of all properties in the model area, it is to be built with provisions for the entire range of real estate in this area, even if transaction information on some types of objects is not in the sample.

It is also essential that the models are regularly updated to reflect the dynamic changes in the real estate market and to ensure that they are adapted to the local particularities of each assessment area. Without constant calibration and rigorous validation of variables, there is a risk that the models will generate Values that is inconsistent with the current market reality. In addition, assessment of external factors such as legislative changes, economic trends, and urban development must be integrated into the model to ensure that the results accurately reflect market Value. Thus, the involvement of the evaluator expert in the analysis remains indispensable to adjust the automatic results and to correct any discrepancies generated by the limitations of the information included in the model.

In conclusion, calibrating of massive valuation models is vital to ensure a fair, reliable valuation adapted to the realities of the real estate market. The calibration of massive real estate valuation models is an essential process to ensure the accuracy and fairness of the results obtained from these models. In the context of massive valuation, where many properties are valued simultaneously, the model must be able to reflect the characteristics of the real estate market correctly and provide valid estimates for various types of real estate.

Please, follow our instructions faithfully, otherwise you have to resubmit your full paper. This will enable us to maintain uniformity in the conference proceedings as well as in the post-conference luxurious books by Press. Thank you for your cooperation and contribution. We are looking forward to seeing you at the Conference.

Acknowledgement:

I express my deepest gratitude to the World Bank Project team and the international consultants for their invaluable contributions to the development of cadastral and real estate evaluation systems in the Republic of Moldova. A special acknowledgment goes to Paul Bidanset, whose expertise and dedication laid the foundation for the new methodology of mass real estate assessment, marking a significant milestone in the modernization of our national practices.

Additionally, I extend my heartfelt thanks to the Lincoln Institute of Land Policy for their essential support and guidance.

References:

[1] International Association of Assessing Officers, "Standard on ratio Studies," 01 April 2013. [Online]. Available: https://www.iaao.org/media/standards/Standard_on_Ratio_Studies.pdf. [Accessed 01 Mai 2024].

[2] Jamal I. Daoud, "Multicollinearity and Regression Analysis," Journal of Physics: Conference Series, 4th International Conference on Mathematical Applications in Engineering 2017 (ICMAE'17) 8–9 August 2017, International Islamic University Malaysia, vol. Volume 949, pp. 1-6, 2017.

[3] Ben p. P. Bervoets, luc d. Hermans, Ruud m. Kathmann, and Marco Kuijper, Lessons from the Netherlands. Real Estate Assessment and Property Tax Systems, Cambrige: Prepared by the Netherlands Council for Real Estate Assessment and the Lincoln Institute of Land Policy, 2024.

[4] Funding agreement, Funding agreement between the Republic of Moldova and the International Development Association for the realization of the Land Registration and Evaluation Project, Chișinău, 2018.

[5] Bidanset P., "Detached Urban Houses Model. Final Report," PIEF, Chișinău, 2023.

[6] Buzu O., "Aplicarea abordării Sinergetice În Teoria Evaluării Proprietăţii Imobiliare," *Economie Şi Sociologie / Economy And Sociology,* no. 1, pp. 73-77, 12 02 2013.

[7] Agenția Relații Funciare și Cadastru, Ordin nr. 91 din 19.12.2023, "privind aprobarea modelului de evaluare a caselor de locuit amplasate în localitățile urbane ale Republicii Moldova și localitățile rurale din municipiile Chișinău și Bălți în scopul impozitării," 2023. [Online]. Available: https://agcc.gov.md/link-type/evaluarea-bunurilor-imobile. [Accessed 01 01 2024].

Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)

The author contributed in the present research, at all stages from the formulation of the problem to the final findings and solution.

Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself

This work was suported by the Lincoln Institute of Land Policy, 113 Brattle St, Cambridge, MA 02138-3400 USA, [Lincoln Institute CLH Dissertation Fellowship - Alexandru Paladi PO-1745]

In the research, author used the date, presented by The Project of Land registration and Assessment from the Republic of Moldova (2018-2024), Implemented by the Department of Cadaster from the Public Services Agency.

AI using

The author states that no artificial intelligence generators were used in writing the article.

Conflict of Interest

The author has no conflict of interest to declare that is relevant to the content of this article.