sciendo

Standardization of Regression Equation Parameters in the Case of Multiple Linear Regression for an Econometric Model Development to Determine the Price of Apartments

Silviu Gabriel SZENTESI*
Aurel Vlaicu University of Arad, Arad, România
**Corresponding author, silviu.szentesi@uav.ro*

Mioara Florina PANTEA*
Aurel Vlaicu University of Arad, Arad, România
**Corresponding author, mioara.pantea@uav.ro*

Vanina Adoriana TRIFAN
Aurel Vlaicu University of Arad, Arad, România
vanina.trifan@uav.ro

Luminița Ioana MAZURU
Aurel Vlaicu University of Arad, Arad, România
luminita.mazuru@uav.ro

Noemi Florina Gabriela SZENTESI
Aurel Vlaicu University of Arad, Arad, România
noemi.szentesi04@e-uvt.ro

Abstract. *This study examines the relationship between apartment prices in Romania and various influencing factors, for which we have constructed a model using multiple linear regression. This type of regression equation's parameters hold practical significance as they aid in establishing the regression equation. The form of the connection, specifically the parameters that define it, greatly influences our understanding of the connection between various quantified variables in the proposed model. By determining the regression function in this manner, we created a model that allows us to make predictions or draw conclusions based on the variables within the model. Microeconomic variables are crucial in influencing real estate price variations at the individual property level. These variables are more specific and relate to factors within the local market of Arad or the property itself. Some key microeconomic variables can influence real estate prices, which are important to ascertain using statistical tools.*

Keywords: Price determinants, Multicollinearity, Autocorrelation, Heteroscedasticity, Beta coefficient

Introduction

The aim of the present paper is to identify the factors that impact property prices and examine the manner in which these factors do so. Studying these factors provides valuable insights for a variety of stakeholders, including clients, investors, buyers, sellers, policymakers, and urban planners. It allows for better decision-making, risk management, and understanding of the broader economic and social context in which real estate operates. The multiple regression method and parameter standardization were utilized to accomplish this task in conjunction with a valid and solid research model from a more client perspective.

DOI: 10.2478/picbe-2024-0198

Advanced statistical approaches can be valuable for improving accuracy and gaining insights into estimating the factors that impact property prices (Li & Chu, 2017; Lu et al., 2017; Sharma & Poongodi 2023; Wang et al, 2021; Hoxha et al, 2022). It's important to note that advanced statistical methods require careful data preprocessing, model selection, and validation to ensure the robustness and generalizability of the results (Wooldridge, 2009; Manațe et al., 2023). Domain expertise and understanding of the local real estate market are also crucial for interpreting the findings accurately. Multiple linear regression is a very useful method to develop a research model to analyze the relationship between multiple predictor variables and property prices. In this study, different factors were used, such as area, age, useful surface, finishing, floor number, partitioning, and number of rooms. Our approach is from a consumer perspective, and in this case, multiple linear regression can be an instrument to quantify the impact of each factor variable and the level of influence on apartment prices and to build a practical predictive model.

In this study, the database was initially recorded in Excel and subsequently coded and processed using the SPSS application. The unstandardized model was developed using the least squares approach, and multicollinearity was assessed in the SPSS software using the tolerance (T) and VIF indicators. The beta coefficients were calculated and examined to assess the impact of each factor. The regression function was evaluated using the coefficient of determination R^2 and the F-statistic as quality measures. The t-value and beta value were used to assess the regression coefficients. The Durbin-Watson test was utilized for autocorrelation analysis, while the Goldfield-Quandt test was employed for graphical analysis of Heteroskedasticity.

Previous scholars have conducted research to examine the variables that impact the pricing of houses and properties (Bourassa et al., 2010; Choong, 2018; Ebekozien et al., 2019; Jafari & Akhavian, 2019; Kim et al., 2022; Yazdani, 2021). However, despite their substantial representation in Romania, there is a relative lack of specifically focused research on apartments. In order to achieve the study objective, a review of the specialized literature was conducted, the research methodology was described, and the model development, calculation of beta coefficients, and standardization of parameters for determining multicollinearity and autocorrelation were described. We tested both the general model and the computed unstandardized coefficients. Heteroskedasticity determination was also carried out for a better evaluation of the model's solidity. Finally, the results, limitations, and further research were presented.

Literature review

When making real estate transactions, homebuyers, sellers, and real estate professionals frequently take into account different types of criteria to make well-informed judgments. According to previous research microeconomic variables interact with macroeconomic factors to shape the real estate market because different microeconomic variables and macroeconomic variables influence real estate price variation (Ding, 2022; Cuc et al., 2023). While macroeconomic factors affect the broader market condition, microeconomic factors are more property-specific and localized.

According to Ryan (1999), the availability of infrastructure and the location of a property are directly correlated with property prices. In addition to price, according to Sundrani (2018), the quantity of rooms (surface), hallway, and kitchen have a significant impact on the purchase and, indirectly, the formation of the property price. Other significant factors include those associated with the product (age, finishing, and compartmentalization) and, of course, the area.

Recent studies also explore environmental influences, such as the presence of natural landscapes and green spaces (Jim & Chen, 2006). According to the World Green Building Council

(2014), exposure to natural environments and the presence of green areas alleviate stress and enhance the emotional well-being of individuals (Rutskiy et al., 2023). Moreover, amusement and entertainment venues, as well as shopping districts, are directly correlated with higher property prices. In this line, Rutskiy et al. (2023) explains that individuals have a requirement to fulfill their social demands by engaging in activities such as watching a movie, dining at a restaurant, or making a purchase.

According to Hong et al. (2015), access to better public transportation and health facilities are the key influential factors directly correlated with the increase in property prices because access to transportation and proximity to health facilities are essential for the well-being of occupants.

Using a cross-sectional dataset of 497 apartments, Aliyev et al. (2019) created a distinct polynomial regression model for apartments. The regression equation was found using the least squares (OLS) approach. The following variables were taken into consideration while analyzing the price: location (area), size in square meters, degree of finishing, terms of sale, and number of rooms.

It is important to note that these variables do not act in isolation, and their impact on real estate prices can vary depending on market conditions. Real estate is a complex and multifaceted asset class, and a combination of these factors, as well as market conditions and local factors, plays an important role in determining the property price. Investors, homeowners, and real estate professionals often monitor these variables to make informed decisions in the real estate market.

The purpose of the present study is to use multiple regression analysis and parameter standardization to accurately predict the price of properties and assess the impact of different variables on their value. Studying the factors that determine the price of apartments provides valuable insights for a variety of stakeholders, including investors, buyers, sellers, policymakers, and urban planners. It allows for better decision-making, risk management, and understanding of the broader economic and social context in which real estate operates. In this study the approach is from a consumer perspective, and in this case, multiple linear regression can be an important instrument to quantify the impact of each factor variable and the level of influence on apartment prices and to build a practical predictive model.

Methodology

The purpose of this study is to use multiple regression analysis and parameter standardization to accurately predict the price of properties and assess the impact of different variables on their value. The independent factors that have an impact on the dependent variable - price, were examined using data from 40 properties collected from Romanian company Blitz Imobiliare S.R.L.

The present study database was created in Excel and the data was encoded and processed in the SPSS Program. Taking into account numerous factors such as area (x_1), age (x_2), useful surface (x_3), finishing (x_4), floor number (x_5), partitioning (x_6), and number of rooms (x_7), this study quantified the impact of each predictor on property prices and built a predictive model. The method of least squares (OLS) was used to estimate the parameters of the linear model.

The linear model comprising seven regressors, as follows:

$$y=b_0+b_1x_1+b_2x_2+b_3x_3+b_4x_4+b_5x_5+b_6x_6+b_7x_7+\varepsilon$$

In this model we used the dependent variable (y); the intercept (b_0), the seven coefficients (b_1, b_2, b_3, b_4, b_5, b_6, b_7), the seven regressors ($x_1,x_2,x_3,x_4,x_5,x_6,x_7$) and error term ($\varepsilon$).

DOI: 10.2478/picbe-2024-0198, pp. 2344-2352, ISSN 2558-9652 |
Proceedings of the 18th International Conference on Business Excellence 2024

The objective is to minimize the sum of squared differences between the model-predicted values ($\hat{y}$) and the observed values (y). Mathematically, the objective is to minimize the following sum:

$$\sum_{i=1}^{n}(y_i - (b_0 + \mathrm{b}x_{1i} + b_2x_{2i} + b_3x_{3i} + b_4x_{4i} + b_5x_{5i} + b_6x_{6i} + b_7x_{7i}))^2$$

where: n - the number of observations.

Tabel 1 Dependent Variable: Price

Price-Model	**Price-Unstandardized Coefficients**		**Price-Standardized Coefficients**	**T**	**Sig.**	**Price-Collinearity Statistics**	
	B	**Std. Error**	**Beta**			**Tolerance**	**VIF**
(Constant)	12488,146	8965,722		1,393	,174		
District	1816,054	815,170	,133	2,228	,033	,770	1,299
Age	-430,721	85,219	-,295	-5,054	,000	,808	1,238
Usable surface	743,253	104,696	,711	7,099	,000	,274	3,656
Improvements level	773,735	327,998	,147	2,359	,025	,706	1,416
Floor level	548,368	276,209	,119	1,985	,056	,760	1,316
Partitioning	351,137	2384,886	,009	,147	,884	,809	1,237
Number of rooms	2903,348	2439,129	,120	1,190	,243	,272	3,676

Source: Authors' research results.

The non-standard model obtained is:

$$\hat{y} = 12488{,}146 + 1816{,}054x_1 - 430{,}721x_2 + 743{,}253x_3 + 773{,}735x_4 + 548{,}368x_5 + 351{,}137x_6 + 2903{,}348x_7$$

The standardized model by the calculation of the Beta coefficients is:

$$\hat{y} = 12487{,}3 + 0{,}133x_1 - 0{,}295x_2 + 0{,}711x_3 + 0{,}147x_4 + 0{,}119x_5 + 0{,}009x_6 + 0{,}120x_7$$

According to the findings, that the variable x_3 – usable surface, has the highest standardized regression coefficient (0.711) and thus has the strongest effect on the price, followed by x_2 - age (0.295), x_4 - degree of finishing (0.147), x_1 - area (0.133), x_7 - number of rooms (0.120) and x_5 – floor level (0.119). Partitioning (x_6) seems to have a marginal role in influencing the price of listed equipment.

The regression function must remove a variable if the examination of the regression coefficients reveals that it does not contribute to the explanation. However, we must first check the overall quality. If the model turns out to be unusable in general, it is not necessary to check the individual regression coefficients.

Tabel 2 ANOVA. Dependent Variable: Preis

Model	Sum of Squares	df	Mean Square	F	Sig.
Regression	14939370446,698	7	2134195778,100	47,620	,000[b]
Residual	1389343189,200	31	44817522,232		
Total	16328713635,897	38			

Source: Authors' research results.

Predictors: (Constant), Number of rooms, Age, Floor level, District, Partitioning, Improvements level, Usable surface.

DOI: 10.2478/picbe-2024-0198, pp. 2344-2352, ISSN 2558-9652 |
Proceedings of the 18th International Conference on Business Excellence 2024

The model, specifically the regression function, is considered correct from an econometric perspective due to the high coefficient of determination (R2) of 0.915 and the adjusted coefficient of determination of 0.896. Furthermore, the study model's factorial parameters extensively account for the price parameter. The calculated F value (47.620) is greater than the F tab. value (0.000).

Tabel. 3 R Square. Durbin-Watson test

R-model		R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics					Durbin-Watson
					R Square Change	F Change	df1	df2	Sig. F Change	
,957[a]		,915	,896	6694,58903	,915	47,620	7	31	,000	2,113

Source: Authors' research results.

We also tested autocorrelation. The formula for calculating the Durbin-Watson (DW) test for autocorrelation analysis is:

$$d = \frac{\sum_{i=2}^{N}(e_i - e_{i-1})^2}{\sum_{i=1}^{N} e_i^2}$$

where:
ei= Residual values for values observed in period i (i = 1,2,...,N)
d = Indicator value for checking autocorrelation.

If the residues (random deviations) of two successive observation values are almost the same and therefore undergo a trend, then the *d* value is also small. Low values of *d* indicate a positive autocorrelation. In contrast, strong jumps in residues lead to large values of *d* and thus indicate the existence of a negative autocorrelation. According to the specific literature, the statistical values ranging from 1.5 to the value of 2.5 in the case of this test are considered rather normal. In our model, the DW test with the value 2,113 is a normal parameter.

Heteroskedasticity analysis

Heteroskedasticity analysis can also be performed by graphical analysis of the association between the adjusted analytical parameters of the dependent variable – price, in the linear regression model, and the evolution of residual values, also known as error or random deviation denoted by $\hat{y}e_i$ or .ε_k .

Graphic analysis is quite subjective and for this reason a Breusch-Pagan test, modified Breusch-Pagan, White test or Goldfield-Qunadt test can be performed. In addition to graphical analysis, we opted for the Goldfield-Quandt test.

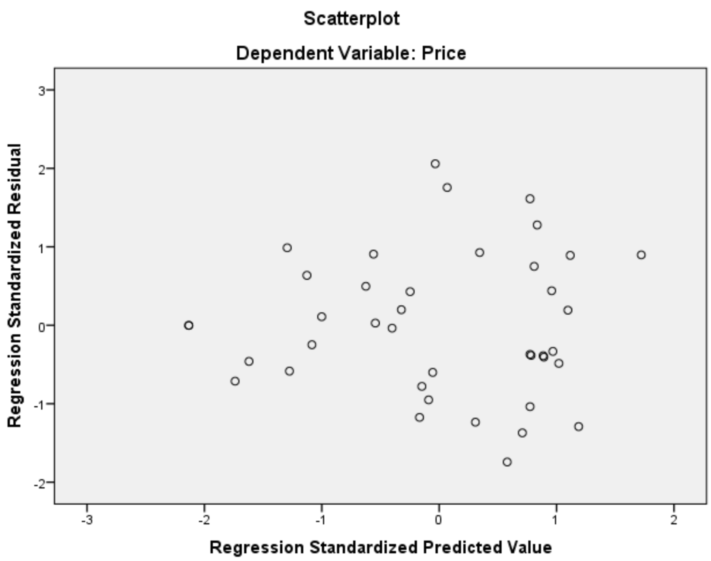

Figure 1 Standardized Residuals Plot

Source: Authors' research results.

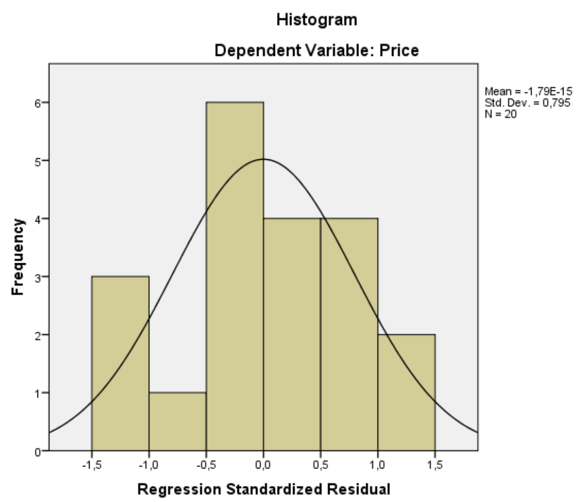

Figure 2 Standardized Residuals Fervency distribution

Source: Authors' research results

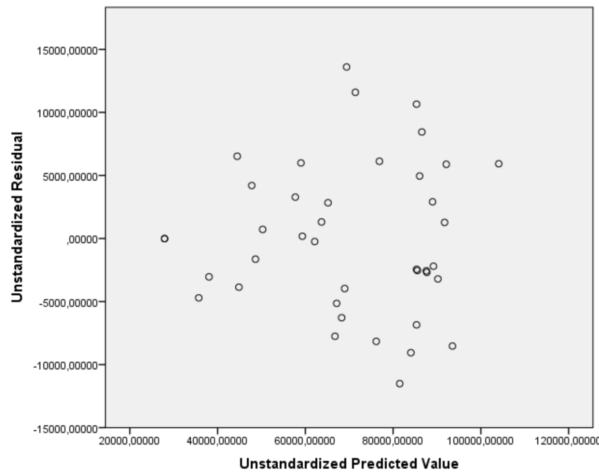

Figure 3. Unstandardized Residuals Plot

Source: Authors' research results

There is a situation of homoscedasticity, the variances are almost identical ($s_1^2 = s_2^2$). The residuals are normally distributed and the assumption of homoscedasticity applies, the ratio of variances follows an F-distribution. The F-test statistic is computed using the folowing mathematical formulas:

$$F_{emp} = \frac{s_1^2}{s_2^2} \qquad s_1^2 = \frac{\sum_{k=1}^{K_1} \varepsilon_k^2}{K_1 - J - 1} \qquad s_2^2 = \frac{\sum_{k=1}^{K2} \varepsilon_k^2}{K_2 - J - 1}$$

In this study:

$$F_{emp} = \frac{s_1^2}{s_2^2} = \frac{46294612}{34016192} = 1{,}361$$
$$F_{0.05;12;12} = 2.69$$

Since $F_{emp} = 1.361 < F_{Tab} = 2.69$ the null hypothesis is accepted and therefore $H_0 \colon \sigma_1^2 = \sigma_2^2$. Since the two distributions are identical, there is no presence of Heteroskedasticity. Therefore, the requirement of Homoskedasticity is accomplished.

Results and discussions

Standardized coefficients, Beta values (or beta weights) are the same as line regression coefficients, except that they have been standardized. This allows us to compare variables directly with each other. Standardization eliminates the units of measurement of all variables. Standardization makes it easier to compare variables with each other. If we look at the standardized coefficients, additionally following the last line in Table 1, we note that the size of the beta value is also related to the level of significance. The beta weights can range from -∞ to +∞, but their values are usually closer to a range of values between -1 and +1. In the case of higher standard deviations, variables usually strongly correlate with each other (we have multicollinearity). According to Table 1, their values in this study case are b_1=0.133 for the district; b_2=-0.295 for age; b_3=-0.711 for usable surface, b_4 =0.147 for improvements level, b_5=0,119 for floor level, b_6=0,009 for partitioning and

b_7=0,120 for the number or rooms. Their sum approaches 1, and beta values indicate the relative importance of most predictors but we notice that the value for partitioning is not relevant because b_6=0,009, and we can skip this factor from a future model.

However, there are also critical aspects regarding the use of standardized coefficients. There are some opinions that standardized coefficients can be misleading. Since standardization removes unity, a beta weight only reflects the strength of the link between the variable distributions. This has the potential to result in some trend in evolution due to sampling errors. In our study, an error can occur when because do not analyze the entire population but only a part, but then take the mean and standard deviation of this small sample to standardize variables. In addition, the distributions of standardized variables should be approximately the same so that they can be interpreted meaningfully, for example, as normal or t-type distributions Student.

For the graphical interpretation of Homoskedasticity we can notice a mild tendency to Heteroskedasticity but based on the F test is not significant. Autocorrelation is not present the value 2,113 is a normal parameter for the DW test. All the values of the VIF – Inflation Variation Factor are all situated under the value of 5 and that means the lack of multicollinearity. The variables selected are useful and good predictors of the model not influencing each other.

Conclusion

The model proposed is significant from the client's perspective. The general test entire model may be significantly determined based on ANOVA and F value, and the test of the regression coefficients is based on Student test -t. The relevance of the model is verified using an ANOVA with an F distribution. Like all significance tests, it is also susceptible to trends resulting from high degrees of freedom, for example from many predictors and/or cases.

The present study has a few limitations and for that reason, we propose a deeper and further research process. More data is needed to develop a truly solid model: cross-sectional data collected from several companies and cities in Romania from real estate market where apartments are usually traded. In this line, we propose to build a database for real estate companies that can benefit and use an actualized and develop accurate model.

References

Aliyev, K., Amiraslanova, M., Bakirova, N., & Eynizada, N. (2019). Determinants of housing prices in Baku: empirical analyses. *International Journal of Housing Markets and Analysis*, *12*(2), 281-297.

Bourassa, S., Cantoni, E., & Hoesli, M. (2010). Predicting house prices with spatial dependence: a comparison of alternative methods. *Journal of Real Estate Research*, *32*(2), 139-160.

Choong, W. C. (2018). *Statistical Analysis Of Housing Prices In Petaling District Using Linear Functional Model* (Doctoral dissertation, UTAR).

Cuc, L. D., Rad, D., Manațe, D., Szentesi, S. G., Dicu, A., Pantea, M. F., ... & Bâtcă-Dumitru, G. C. (2023). Representations of the Smart Green Concept and the Intention to Implement IoT in Romanian Real Estate Development. *Sustainability*, *15*(10), 7777.

Ding, X. (2022). Macroeconomic Factors Affecting Housing Prices: Take the United States as an Example. In *2022 7th International Conference on Financial Innovation and Economic Development (ICFIED 2022)* (pp. 2335-2339). Atlantis Press.

Ebekozien, A., Abdul-Aziz, A. R., & Jaafar, M. (2019). Housing finance inaccessibility for low-income earners in Malaysia: Factors and solutions. *Habitat International*, *87*, 27-35.

DOI: 10.2478/picbe-2024-0198, pp. 2344-2352, ISSN 2558-9652 |
Proceedings of the 18th International Conference on Business Excellence 2024

Hong, T., Koo, C., Kim, D., Lee, M., & Kim, J. (2015). An estimation methodology for the dynamic operational rating of a new residential building using the advanced case-based reasoning and stochastic approaches. *Applied Energy*, *150*, 308-322.

Hoxha, V., Hoxha, D., & Hoxha, J. (2022). Study of factors influencing apartment prices in Prishtina, Kosovo. *International Journal of Housing Markets and Analysis*, *15*(5), 1242-1258.

Jafari, A., & Akhavian, R. (2019). Driving forces for the US residential housing price: a predictive analysis. *Built Environment Project and Asset Management*, *9*(4), 515-529.

Jim, C. Y., & Chen, W. Y. (2006). Impacts of urban environmental elements on residential housing prices in Guangzhou (China). *Landscape and urban planning*, *78*(4), 422-434.

Kim, J., Lee, Y., Lee, M. H., & Hong, S. Y. (2022). A comparative study of machine learning and spatial interpolation methods for predicting house prices. *Sustainability*, *14*(15), 9056.

Li, L., & Chu, K. H. (2017). Prediction of real estate price variation based on economic parameters. In *2017 International Conference on Applied System Innovation (ICASI)* (pp. 87-90). IEEE.

Lu, S., Li, Z., Qin, Z., Yang, X., & Goh, R. S. M. (2017). A hybrid regression technique for house prices prediction. In *2017 IEEE international conference on industrial engineering and engineering management (IEEM)* (pp. 319-323). IEEE.

Manate, D., Lile, R., Rad, D., Szentesi, S. G., & Cuc, L. D. (2023). An analysis of the concept of green buildings in Romania in the context of the energy paradigm change in the EU. *Transformations in Business & Economics*, 22(1).

Rutskiy, V., García, D. S., Denisova, E., Alina, F., Okashev, N., Devederkin, I., Bystrova, N., Elisseva, E., & Tsarev, R. (2023). Modeling the Well-Being of the Population and Its Factors Using the Well-Being Index. In *Computer Science On-line Conference* (pp. 605-614). Cham: Springer International Publishing.

Ryan, S. (1999). Property values and transportation facilities: Finding the transportation-land use connection. *Journal of planning literature*, *13*(4), 412-427.

Sharma, A., & Poongodi, T. (2023). Prediction of Real-Time Estate Pricing using Train-Test Splitting Techniques. In *2023 4th International Conference on Intelligent Engineering and Management (ICIEM)* (pp. 1-6). IEEE.

Sundrani, D. M. (2018). Factors influencing home-purchase decision of buyers of different types of apartments in India. *International Journal of Housing Markets and Analysis*, *11*(4), 609-631.

Wang, P. Y., Chen, C. T., Su, J. W., Wang, T. Y., & Huang, S. H. (2021). Deep learning model for house price prediction using heterogeneous data analysis along with joint self-attention mechanism. *IEEE Access*, *9*, 55244-55259.

Wooldridge, J. M. (2009). *Introductory econometrics: A modern approach*. Mason, OH: South Western, Cengage Learning. Speziell S. 271-276.

World Green Building Council (2014). *Health, wellbeing & productivity in offices.* Retrieved from:https://worldgbc.org/wpcontent/uploads/2022/03/compressed_WorldGBC_Health_Wellbeing__Productivity_Full_Report_Dbl_Med_Res_Feb_2015-1.pdf

Yazdani, M. (2021). Machine learning, deep learning, and hedonic methods for real estate price prediction. *arXiv preprint arXiv:2110.07151*.

DOI: 10.2478/picbe-2024-0198, pp. 2344-2352, ISSN 2558-9652 |
Proceedings of the 18th International Conference on Business Excellence 2024