Theoretical and practical aspects of qualitative variable descriptions of residential property valuation multiple regression models

Iwona Foryś[1], Radosław Gaca[2]

Abstract

With an increasing number of transactions and growing numbers of economic activity areas, which requires determining the value of real estate, also the numbers of attempts of practical implementation of multiple regression models increased as well (Bergh and Ketchen, 2014). However, the imperfections of the regression model, and property being real described objects in the models, lead to contradictory and unjustified conclusions. One of the reasons of the functional imperfections of regression models is a qualitative variables input method. Unfortunately, this type of variables are most usually used in decryption of property characteristics. According to the principles defining possible operations on numbers depending on the type of the scale of measurement (nominal and ordinal), the qualitative variables can be introduced into the regression models as dummy variables. In this article the authors propose an alternative implementation method of qualitative variables (describing the real estate markets) in multiple regression models, which describe the variability by means of the Osgood scale. This paper presents not only a theoretical basis of the proposed model but also the results of empirical data in relation to the classical method. Theoretical considerations are supported by the empirical study on the residential housing market in Szczecin and Bydgoszcz.

Keywords: *residential property valuation, qualitative variables, Osgood scale, semantic scale, multiple regression*

JEL Classification: C50, R30

1. Introduction

In Poland the beginnings of the theory of the residential property market valuation were closely associated with the system transformation after 1989. Along with the growing number of transactions and increasing range of economic activities requiring real property valuation, the attempts have appeared to introduce multiple regression models into practice.

However, both the imperfections of the regression model itself and the property attributes as the real objects described in the models (Green, 1991; Preachera et al., 2013) are the reason for contradictory and unjustified conclusions. One of the reasons for functional imperfection of the regression models is the mode of introducing qualitative variables, which are often the most relevant attributes describing the real property (Johnson, 2010).

[1] Corresponding author: University of Szczecin, Department of Economic and Management, Mickiewicza 64, 71-101 Szczecin, Poland, e-mail: forys@wneiz.pl.

[2] Rzeczoznawca majątkowy, Gdańska 5/3a, 81-005 Bydgoszcz, Poland, e-mail: rgaca@poczta.onet.pl.

The aim of this study is to test the method that is alternative to the one commonly used for this purpose and, basing on dummy variables, introduces qualitative variables (Bennett and Elman, 2006) to the multiple regression models describing real property markets, through using the Osgood scale for the description of their variability.

The real property attribute quantification has been the subject of numerous publications (Foryś, 2011; Hozer, 2001; Bełej and Źróbek, 2000). None of those papers, however, discussed in detail the ways of introducing qualitative variables to the models, particularly from the point of view of the acceptability of operations on numbers in terms of the measurement scale.

Stevens (Stevens, 1946) was the first to introduce the measurement scales, classifying them as: the nominal, ordinal, interval and ratio ones. According to the measurement principles, the given type of scale is associated with specific permitted groups of transformations as well as permitted arithmetic operations (Walesiak, 1996). In keeping with Stevens' quantification method, the figures expressed on the low-level scales (nominal and ordinal) do not have interpretations typical of natural numbers. For the data expressed on these scales the numbers have a character of differentiating and positioning codes. In the classical approach, the indicated numbers describe neither the interval nor the ratio between the specific variables.

According to one of the main rules of the measurement theory, the measurement results that are described on the stronger scale cannot be transformed into the numbers belonging solely to the weaker scale. The reversed transformation of data that would make them stronger is not possible. This is due to a simple fact concerning the information communicated by a given measurement (Walesiak, 1996; Wiśniewski, 1986). While it is true that there are methods to transform data measured on the ordinal scale onto the interval scale, such a transformation does not bring more information to the transformed data (Walesiak, 2014).

The proposed method to express the states of qualitative attributes by means of a scale based on the semantic differential makes the obtained results closer to the measurement made on the scale which is at least the interval one.

2. Methodology

Basing on the existing studies on modelling the qualitative variables in regression models with the view to test the hypothesis about the validity of the alternative method for the introduction of qualitative variables expressed by the semantic differential, as well as to verify the hypothesis about the interval and uniform distribution of the variability of the qualitative attributes' states, the authors compare the results of the estimation of linear regression model parameters by means of two methods of introducing the qualitative variables, i.e. the method whose qualitative variables that have been quantified with the use of the Osgood scale are introduced directly on the basis of their values and the method where the aforementioned variables have been transformed into dummy variables.

The calculations have been made with the use of the multiple linear regression model written:

$$Y = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + ... + \alpha_m X_m + \varepsilon \tag{1}$$

where: Y – dependent (explanatory) variable; X_i – independent variables where , i = 1, 2,.., m; ε – random component.

The basis for the model parameters estimation is the matrix **A** of the real property attributes for the analysed set of similar real properties.

$$\mathbf{A} = \begin{bmatrix} x_{11} & \dots & x_{1m} \\ x_{21} & \dots & x_{2m} \\ \dots & \dots & \dots \\ x_{n1} & \dots & x_{nm} \end{bmatrix}. \tag{2}$$

Information in the matrix **A** is the basis for determining the property value in a comparative approach leading to its market value estimation. The marketable attributes encompassing the qualitative variables described in the matrix **A** are quantified with the use of the Osgood scale (Foryś and Gaca, 2015).

The concept of the Osgood scale, also known as the semantic scale or the semantic differential, was proposed in 1957 (Osgood et al., 1957). The structure of the scale is based on the assessment of the phenomenon's or the object's state intensity that is evoked in the respondent's mind. The assessment is made in a semantic form. In this paper the authors assume that the measurement of the connotative meaning is limited to indicating the position the given notion or name occupies in the semantic space. Osgood's analyses revealed that the elementary dimensions of the specific semantic space were: value, strength and activity. As Osgood indicated, depending on the research purpose and needs, we can apply different sets of extreme or neutral notions that refer to the aforementioned three universal dimensions. The properties of the scales described above allow us to treat the data thus obtained as the ones of interval (semi-interval) character. The character of the scales results from their construction based on the evaluation of the "position" of the respondent's attitude or opinion in relation to the neutral state and the extreme states in a given group. In practice, the identifiers (quantifiers, codes) of individual states are either natural numbers, where the lowest rank is 1, or integers, where the mean (neutral) rank is denoted as zero, the states above are natural numbers while the states below are consecutive negative integers. However, in each of the above described coding systems it is possible to determine the interval between the specific fixed values, which lets us use the methods adequate for the interval scale for the purpose of the results analysis.

When we are using the qualitative attribute evaluation by means of the semantic differential, it is particularly important to define properly the scopes of the maximum and minimum score. Taking the above into consideration, the scale having the best properties would be the one whose extreme opinion ranges correspond to the meaning that is universal in a given community. Yet, when we want to build the real property valuation model based on a group of similar real properties, we must remember that in most cases the level of differentiation of the specific attributes' states that distinguish individual real properties from the group will be lower and will constitute just a fraction of the total score range. However, when we are evaluating the states of real property attributes by means of the Osgood scale, each of the analysed groups needs to be evaluated and scaled separately in reference to the diagnosed extreme states and the neutral state. Also, in every case the range of the scale must be adjusted to the observed differentiation of attributes.

In order to verify the assumption that the variables describing the states of real properties quantified by means of the semantic differential are of at least interval character, the estimation results of both above described regression models have been compared.

The calculations for the model with dummy variables have been made basing on the multiple linear regression model:

$$Y = \alpha_0 + \alpha_{12}D_{12} + \alpha_{13}D_{13} + \alpha_{22}D_{22} .... + \alpha_{mk}D_{mk} + \varepsilon \tag{3}$$

where: Y – dependent (explanatory) variable, D_{ij} – independent variables, where i = 1, 2, …, m; j = 1, 2, …, k; ε – random component.

The transformation of qualitative variables quantified by means of the Osgood scale into dummy variables has been done with the formulas:

- for the variables with two states:

$$d_{ij} = \begin{cases} 1 \; for\, score\, 2 \\ 0 \; for\, score\, 1 \end{cases}, \tag{4}$$

- for the variables with three states:

$$d_{ij} = \begin{cases} 1 \; for\, score \quad 2 \\ 0 \; for\, score 1 and 3 \end{cases},$$
$$d_{ij+1} = \begin{cases} 1 \; for\, score \quad 3 \\ 0 \; for\, score 1 and 2 \end{cases}. \tag{5}$$

Basing on the transformations we have the matrix **A'** of dummy variables.

$$\mathbf{A}' = \begin{bmatrix} d_{1_{12}} & d_{1_{13}} & \dots & d_{1_{mk}} \\ d_{2_{12}} & d_{2_{13}} & \dots & d_{2_{mk}} \\ \dots & \dots & \dots & \dots \\ d_{n_{12}} & d_{n_{13}} & \dots & d_{n_{mk}} \end{bmatrix}. \tag{6}$$

3. Empirical study

The assumptions were verified on the basis of two sets of similar residential properties in Bydgoszcz and Szczecin. In order to obtain the sets encompassing the differentiation solely in terms of qualitative attributes that can be described by means of the semantic differential, the authors used transactions concerning the ownership rights to a residential real properties of practically identical size and located in buildings of a similar type.

Ultimately, two sets were selected. The set O_B covering 43 transactions concerning trading cooperative member's rights to residential premises in Bydgoszcz concluded between February 2014 and October 2015. The set O_{SZ} covered 55 transactions concerning trading cooperative member's rights to residential premises in Szczecin, concluded between January 2010 and December 2012. In the set O_B the authors applied a linear correction of transaction prices due to 4% annual changes in property prices. In the case of the set O_{SZ} the correction was not necessary, as no relevant price movements were observed. For both sets the null hypothesis tests were conducted to check if the conditions for normal distribution had been met. The results of Shapiro–Wilk, Jarque–Bera, Anderson–Darling and Kolmogorov–Smirnov tests were conflicting. The Jarque–Bera and Anderson–Darling tests provided no grounds to reject the null hypothesis about the normality of the distribution, while the Shapiro–Wilk and Kolmogorov–Smirnov tests proved that hypothesis invalid.

The attributes differentiating the analysed sets in the qualitative sense included:

- location (*X1*),
- technical condition of the building (*X*2),
- the storey the flat is on (*X3*).

For the specified variables the rating scale was adopted (the Osgood scale) as in the Table below (Table 1). The plus (+) index means the intermediate state between average and good, or good and very good.

Variable	Set O_B	Set O_{SZ}
X1	1 average, 2 good, 3 v.good	1 worse, 2 better
X2	1 average(+), 2 good, 3 good(+)	1 good, 2 good(+), 3 v. good
X3	1 average, 2 good, 3 v. good	1 average, 2 good, 3 v. good

Table 1. The Osgood scale for the analysed qualitative variables.

Then, for the variables described on the Osgood scale and for dummy variables the multiple regression equation parameters were estimated by means of the least squares method. When constructing the model, in both sets the authors observed the insignificance of the attribute denoted as *X3* and describing in qualitative terms the floor on which the flat was located. In the next step, that particular variable was ignored and the estimation was limited to parameters of the models with two independent variables *X1*, *X2*. The proposed models meet the assumptions of multiple regression. In particular, a set of observations are homogeneous, the relationship between variables are linear, and the distribution of residues are normal.

For both analyzed models they were not found grounds to reject the null hypothesis of residuals normal distribution. The results of regression coefficients for the both sets and both methods for variables that were statistically significant had signs consistent with the assumptions of economics at the impact of specific factors on changes in real estate prices. (Bełej and Źróbek, 2000).

For the set O_B and variables described with the semantic differential:

$$\overline{R}^2 = 77.9\% \quad V = 0.1157\,,$$

$$\hat{y}_i = \underset{(145.29)}{2090.19} + \underset{(75.55)}{411.45}\, x_1 + \underset{(75.55)}{422.02}\; x_2\,.$$

For the set O_B and variables described with dummy variables:

$$\overline{R}^2 = 77.5\% \quad V = 0.1163,$$

$$\hat{y}_i = \underset{(112.25)}{2841.56} + \underset{(115.61)}{440.39}\, d_{12} + \underset{(153.18)}{839.31}\, d_{13} + \underset{(115.61)}{508.60}\, d_{22} + \underset{(153.18)}{860.46}\, d_{23}\,.$$

For the set O_{SZ} and variables described by means of the semantic differential:

$$\bar{R}^2 = 74.3\% \quad V = 0,0741,$$

$$\hat{y}_i = \underset{(187,03)}{2468.99} + \underset{(73,95)}{345,04}\, x_1 + \underset{(54,20)}{670,97}\; x_2 .$$

For the set O_{SZ} and variables described by means of dummy variables:

$$\bar{R}^2 = 74.6\% \quad V = 0.0871,$$

$$\hat{y}_i = \underset{(92.51)}{3424.73} + \underset{(73.60)}{350.25}\, d_{12} + \underset{(88.24)}{761.08}\, d_{22} + \underset{(109.33)}{1366.07}\, d_{23}.$$

On the basis of the models obtained with the use of dummy variables the authors determined the rate of the impact of individual attribute states described with dummy variables on the independent variable to the attributes corresponding to specific values on the Osgood scale (Table 2).

Conclusion

The results obtained for the analysed sets confirm the hypothesis that the qualitative variables quantified by means of the semantic differential can be treated as the variables that are at least the interval ones. Although the observed intervals for individual attribute states expressed with the coefficient values of dummy variables are not ideally proportionate, the observed deviations are relatively small. The findings obtained in the course of the study confirm the results of the analysis based on the correlation results comparison that was conducted for other sets of residential properties (Foryś and Gaca, 2015).

It should be noted that the proportionate representation of the impact of the qualitative variables expressed by means of the Osgood scale could be seen in the sets of residential real properties that were similar in terms of such price-relevant attributes as the size, functionality, general location, type of building, etc.). In view of the analysis results, it can be assumed that in the case of sets of residential properties that differ in terms of their qualitative attributes it is possible to include directly into the valuation models the states of qualitative attributes that have been quantified with the use of the semantic differential. The behaviour of qualitative variables that have been quantified with the use of the semantic differential for more diversified sets of real properties is the issue that requires further in-depth analyses.

	Designation of variables					
	X1			**X2**		
Model for set O_B						
Attribute states on Osgood scale	1	2	3	1	2	3
Coefficient values for dummy variables	0	440.39	839.31	0	508.6	860.46
Ratio to maximum value	0	0.5247	1	0	0.5911	1
Model for set O_{SZ}						
Attribute states on Osgood scale	1		2	1	2	3
Coefficient values for dummy variables	0		350.25	0	761.08	1366.1
Ratio to maximum value	0		1	0	0.5571	1

Table 2. Comparison of coefficient estimation results for the description of attribute states, quantification according to Osgood scale and transformation into dummy variables.

Acknowledgements

The data used in this paper come from databases collected by asset valuers based in Bydgoszcz and Szczecin.

References

Bełej, M., & Źróbek, S. (2000). *Podejście porównawcze w szacowaniu nieruchom*ości. Olsztyn: Educaterra Sp z o.o.

Bennett, A., & Elman, C. (2006). Qualitative Research: Recent Developments in Case Study Methods. *Annual Review of Political Science*, (9), 455-476.

Bergh, D. D., & Ketchen, D. J. (2014). Social Entrepreneurship and Research Methods. *Research Methodology in Strategy and Management*, (9), Emerald Group Publishing Limited.

Foryś, I. (2011). Wykorzystanie metod taksonomicznych do wyboru obiektów podobnych w procesie wyceny lokali mieszkalnych. *Studia i Materiały Towarzystwa Naukowego Nieruchomości, 18*(1). Olsztyn, 95-107.

Foryś, I., & Gaca, R. (2015). Aplikacja skal Likerta i Osgooda do kwantyfikacji stanów cech jakościowych nieruchomości. *Folia Oeconomica Stetinensia*. Wydawnictwo Uniwersytetu Szczecińskiego, w druku.

Green, S. B. (1991). How Many Subject Does It Take To Do A Regression Analysis: A Model Selection Perspective. *Multivariate Behavioral Research, 26*(3), 499-510.

Hardy, M. A. (1993). Regression with dummy variables. *SAGE University Papers series on Quantitative Applications in the Social Sciences*. USA: SAGE Publications.

Hozer, J. (2001). Regresja wieloraka a wycena nieruchomości. *Rzeczoznawca Majątkowy*, (2). Warszawa, 13-14.

Johnson, J. W. (2010). A Heuristics Method for Estimating the Relative Weight of Predictor Variables in Multiple Regression. *Multivariate Behavioral Research, 35*(1), 1-19.

Knapp, T. R (1990). Treating Ordinal Scales as Interval Scales: An Attempt To Resolve the Controversy. *Nursing Research, 39*(2), 121-123.

Preachera, K. J., Zhangb, G., Kimc, Ch., & Melsd, G. (2013). Choosing the Optimal Number of Factors in Exploratory Factor Analysis: A Model Selection Perspective. *Multivariate Behavioral Research, 48*(1), 28-56.

Osgood, C. E., Suci, G., & Tannenbaum, P. (1957). The measurement of meaning. *Urbana*. IL: University of Illinois Press.

Stevens, S. S. (1946). On the theory of scales of measurement. *Science, 103*(67).

Walesiak, M. (1996). Dopuszczalne działania na liczbach w badaniach marketingowych z punktu widzenia skal pomiarowych. *Prace Naukowe Akademii Ekonomicznej we Wrocławiu*, (718), 103-110.

Walesiak, M., & Dudek, A. (2007). Symulacyjna optymalizacja wyboru procedury klasyfikacyjnej dla danego typu danych – charakterystyka problemu. *Zeszyty Naukowe, Prace Katedry Ekonometrii i Statystyki, Prace Katedry Ekonometrii i Statystyki,* (17), 635-646.

Wiśniewski, J.W. (1986), Korelacja i regresja w badaniach zjawisk jakościowych na tle teorii pomiaru. *Przegląd Statystyczny*, (3), 239-248.