AN APPLICATION OF CLUSTER ANALYSIS ON THE POLISH HOUSING MARKET

Patrycja Kowalczyk –Rólczyńska

Abstract

The real estate market is one of the most important markets in the economic development of the country. This market is strongly related to the mortgage market, the labor market and the building sector. Analysis of the real estate market is related to the assessment of the development of the many factors which describe this market. The factors describing the real estate market are: prices on the real estate market, real estate entities, investments in this market, transactions on this market and the features of the real estate. It follows that the analysis of the real estate market should be multidimensional.

The purpose of this article is application of cluster analysis to housing market in selected cities in Poland. In the article there was used one of hierarchical methods of multi-feature object clustering – the Ward's method. The analysis included housing markets in sixteen biggest Polish cities in the period 2009-2012.

Key words: housing market, cluster analysis, factors influencing housing market

JEL Code: R310, C180, C380

Introduction

The housing market plays an extremely significant role in the economy. It is under the influence of and it exerts influence on the development of the economy among others through generating gross domestic product, supplying the necessary space to satisfy housing needs, engaging a considerable number of employees, participating in taxes (local, on income generated by real estate, on sales, on the appreciation of real estate), developing financial products needed for real estate purchase and insurance. Given also the high number of factors describing the housing market and factors contributing the development of this market, it has become the object of numerous studies (see e.g. Fischer & Stamos 2013; Renaud et al. 2001; Genesove & Mayer, 2001).

An analysis of the Polish housing market is conducted in this paper with the use of one of hierarchical methods of multi-feature object clustering – the Ward's method. The analysis included housing markets in sixteen biggest Polish cities[1] in the period 2009-2012. Due to the fact that real estate markets, and housing markets in particular, are local in nature (which arises from the association of real estate with land), the cities were clustered by adopted criteria to enable the capturing of convergences and identical trends occurring on local voivodship markets.

1. Factors describing the housing market

The Polish housing sector, although it is often analysed as a whole, is a heterogeneous market that is largely diversified across the individual 16 voivodship capital cities. In order to cluster local real estate markets and conduct an analysis of the identical trends occurring on these markets, factors describing each of the local housing markets were chosen. These factors include (with reference to each analysed year and each analysed city)[2]:

- Transaction prices on the primary housing market,
- Transaction prices on the secondary housing market,
- Total number of apartments,
- Number of delivered apartments,
- Number of issued residential building permits,
- Average usable floor space in an apartment,
- Average usable floor space in a delivered apartment,
- Average number of rooms in an apartment,
- Average number of rooms in a delivered apartment,
- Number of entities operating on the real estate market (including developers, real estate agents, real estate surveyors, housing cooperatives),
- Number of transactions concluded on the housing market,
- Value of the transactions concluded on the housing market.

Transaction prices of residential real estate on the primary housing market depend on the location. It is observed that the majority of projects in cities are implemented in their suburbs, outside the city centre. This arises mainly from the availability and price of land on such areas. Where the price of land is high and the location is good (near or in the city centre),

[1] Poland is divided into 16 voivodships. The object of the analysis is housing markets in each capital city of a given voivodship.

[2] These are factors used for the purpose of analysing local housing markets carried out by the National Bank of Poland (see: NBP, 2013)

apartment prices are extremely high, which results in low interest in them. A large number of persons purchasing apartments use a mortgage loan and not many households can obtain a high loan amount (exceeding PLN 500,000)[3]. It needs to be emphasised that the secondary housing market offers cheaper apartments than the primary one. This regards mainly apartments situated in big housing estates constructed with the so-called large panels or in the post-war period, the wear of which is considerable. However, location is an asset of such real estate.

The development of the housing market is proved by the number of delivered apartments and the number of issued building permits. The increasing number of delivered apartments and the growing number of issued building permits proves the continuous housing needs on a given local market. Moreover, this means that the enterprises constructing new residential real estate achieve the expected rates of return on investment and willingly implement new projects (since they are profitable).

The average usable floor space in an apartment and the average number of rooms in an apartment are often adapted to the demand on the local housing market. Moreover, the usable floor space is limited in Poland by government programmes aimed to support the housing development industry. The programmes have been implemented for several years and are dedicated to young people (i.e. ones not older than 35 years). In both significant government programmes, i.e. "Family in its own home" ("Rodzina na swoim")[4] and "A home for the young" ("Mieszkanie dla Młodych") the maximum apartment floor space is 75 sq m and 85 sq m (if the apartment purchaser raises at least three children), respectively.

An important position on the real estate market is held by the entities that operate thereon. These include developers, real estate agents, real estate surveyors, and housing cooperatives. The increasing number of such entities contributes to competition on the real estate market and to lower costs of the services provided thereby (commission fees for agents, charges for preparing the property appraisal report) and to lower apartment prices.

The transactions concluded on the housing market (both with reference to the number and value of such transactions) imply the local nature and the development of the market. The low number of concluded transactions might be a reflection of a too low demand or a demand that fails to satisfy needs and expectations of apartment purchasers. The growing number of

[3] In 2013, only 5.63% of borrowers obtained mortgage loans of over PLN 500,000 (i.e. circa EUR 119,000)

[4] Submitting applications for funding as part of this programme has already completed while the programme was intended to apply for 8 years and therefore the state support is still being granted.

concluded transactions and the increasing number of delivered apartments might prove that the supply on this real estate segment corresponds to the demand.

2. Research methods – the Ward's method

Hierarchical methods are classification methods used for the purpose of clustering by similarities and dividing by differences of the objects subject to examination. Classification methods and related problems are quite specifically discussed in the relevant literature (see e.g. Jardine & Sibson, 1968; Ward, 1963; Rand, 1971; Ball, 1965).

The Ward's method is one of agglomerative hierarchical methods. Hierarchical methods of multi-feature object clustering, proposed by G. M. Lance and W. T. Williams (1967), are characterised by several variants differing by the manner of designating the matrix of distance between object clusters in a multidimensional classification space. The Ward's method is a specific case of a general object clustering scheme developed by G. M. Lance and W. T. Williams.

The pattern of conduct is the same for all agglomerative methods and can be called the central agglomerative procedure. It is assumed that the point of departure is matrix D of Euclidean distances d_{ij} between the classified objects $O_1, O_2, ..., O_N$.

The central agglomerative procedure allowing for the Ward's method is as follows (Nowak, 1990):

1. Each object O_i $(i = 1,2,..., N)$ is treated as a single-element cluster,
2. The minimum value $d_{pq} = \min_{i,j}\{d_{ij}\}$ is sought in the distance matrix;
3. Objects O_p and O_q are treated as single-element clusters, A_p and A_q are merged in one two-element cluster A_r:
   $A_r = A_p \cup A_q$;
4. Next the distance d_{ir} of the newly created cluster A_r from all other clusters A_i is designated in accordance with the selected method. Distance d_{ir} is put to matrix **D** in the place of the p-th line and p-th column, eliminating at the same time q line and column;
5. Steps 2–4 are repeated until all objects create one cluster.

After each iteration of the agglomerative hierarchical procedure a division of comparable objects into an increasingly smaller number of clusters is obtained. The outcome is the current matrix of distance between clusters.

The general formula for transforming the distance matrix while merging clusters A_p and A_q to form a new cluster A_r for hierarchical agglomerative methods operating in accordance with the principle of central agglomerative procedure is as follows:

$$d_{ir} = a_p d_{ip} + a_q d_{iq} + b d_{pq} + c\left|d_{ip} - d_{iq}\right| \qquad (1)$$

where:

d_{ir} – distance between clusters A_i and A_r,

d_{ip} – distance between clusters A_i and A_p,

d_{iq} – distance between clusters A_i and A_q,

d_{pq} – distance between clusters A_p and A_q

Values a_p, a_q, b and c are distance matrix transformation parameters. For the Ward's method they are designated based on the following formulas:

$$a_p = \frac{N_i + N_p}{N_i + N_r} \qquad (2)$$

$$a_q = \frac{N_i + N_q}{N_i + N_r} \qquad (3)$$

$$b = -\frac{N_i}{N_i + N_r} \qquad (4)$$

$$c = 0\,.$$

Symbols N_i, N_p, N_q and N_r denote the number of elements in clusters A_i, A_p, A_q and A_r.

A dendrogram, which is a graphic interpretation of the outcome, is obtained from the analysis.

The Ward's method intends to minimise the sum of squared deviations of any two clusters, which can be formed at each stage. It is believed that the Ward's method is a very effective method and it intends to obtain small clusters (Grabiński & Sokołowski, 1984).

In order to use the above-described method, appropriate diagnostic features need to be selected. The selection of such features can be considered an extra-statistical or a statistical task. In the former case, the diagnostic features are the values that most accurately illustrate the features of the compared objects in light of the acquired substantive knowledge about the

examined phenomenon. In the latter case, the selection of appropriate diagnostic features is determined by selected statistical methods. One of the statistical criteria for selecting features is variation of a given diagnostic feature which is determined for the purpose of checking whether a given diagnostic feature is capable of discriminating the examined objects or is characterised by sufficiently high variation. The selection of features with the use of this criterion consists in eliminating the features whose variation coefficients[5] $V_i = \frac{s_i}{\bar{x}_i}$ (where s_i is standard deviation of feature, $\bar{x}_i$ is average value of feature x_i) are lower in terms of the module than the assumed critical value V^* from a given set of potential diagnostic features since such features are believed to be characterised by too low variation.

3. Results of the research

The research included 16 local housing markets in Poland[6] in the period 2009-2012. Potential diagnostic features were created based on the factors described in section 1, which were selected substantively as features describing the housing market. Additionally, variability coefficients of potential diagnostic features were calculated in order to check whether they are capable of discriminating the examined housing markets. The values of the coefficients correspond to the 2012 data since they are the most up-to-date. It is worth mentioning that variability coefficients of diagnostic features were not calculated for each year separately as it could turn out that different diagnostic features would be characterised by too low variability each year and they should be removed from the set of potential diagnostic features. This would mean that the set of diagnostic features would be different for each examined year. In such a situation, a comparison of local housing markets over the examined period, that is four years, would not be possible. The values of variability coefficients for the largest cities on the Polish housing market are presented in Table 1.

Tab. 1: Variability coefficients of diagnostic features for 2012

Name of a diagnostic feature	Variability coefficient
Transaction prices on the primary housing market	0.155
Transaction prices on the secondary housing market	0.261

[5] Variation coefficient is a measure of relative dispersion of a data series (Makridakis et al., 1998)

[6] The analysis included real estate markets in the following Polish cities: Białystok, Bydgoszcz, Katowice, Kielce, Kraków, Lublin, Łódź, Olsztyn, Opole, Poznań, Rzeszów, Szczecin, Gdańsk, Warsaw, Wrocław, Zielona Góra.

Number of apartments per resident	0.075
Number of delivered apartments per resident	0.488
Number of issued residential building permits per resident	0.441
Average usable floor space in an apartment	0.051
Average usable floor space in a delivered apartment	0.174
Average number of rooms in an apartment	0,058
Average number of rooms in a delivered apartment	0.126
Number of developers operating on the real estate market per resident	0.968
Number of real estate surveyors operating on the real estate market per resident	0.541
Number of real estate agents operating on the real estate market per resident	0.828
Number of housing cooperatives operating on the real estate market per resident	0.885
Number of transactions concluded on the housing market per resident	0.569
Value of the transactions concluded on the housing market per resident	0.729

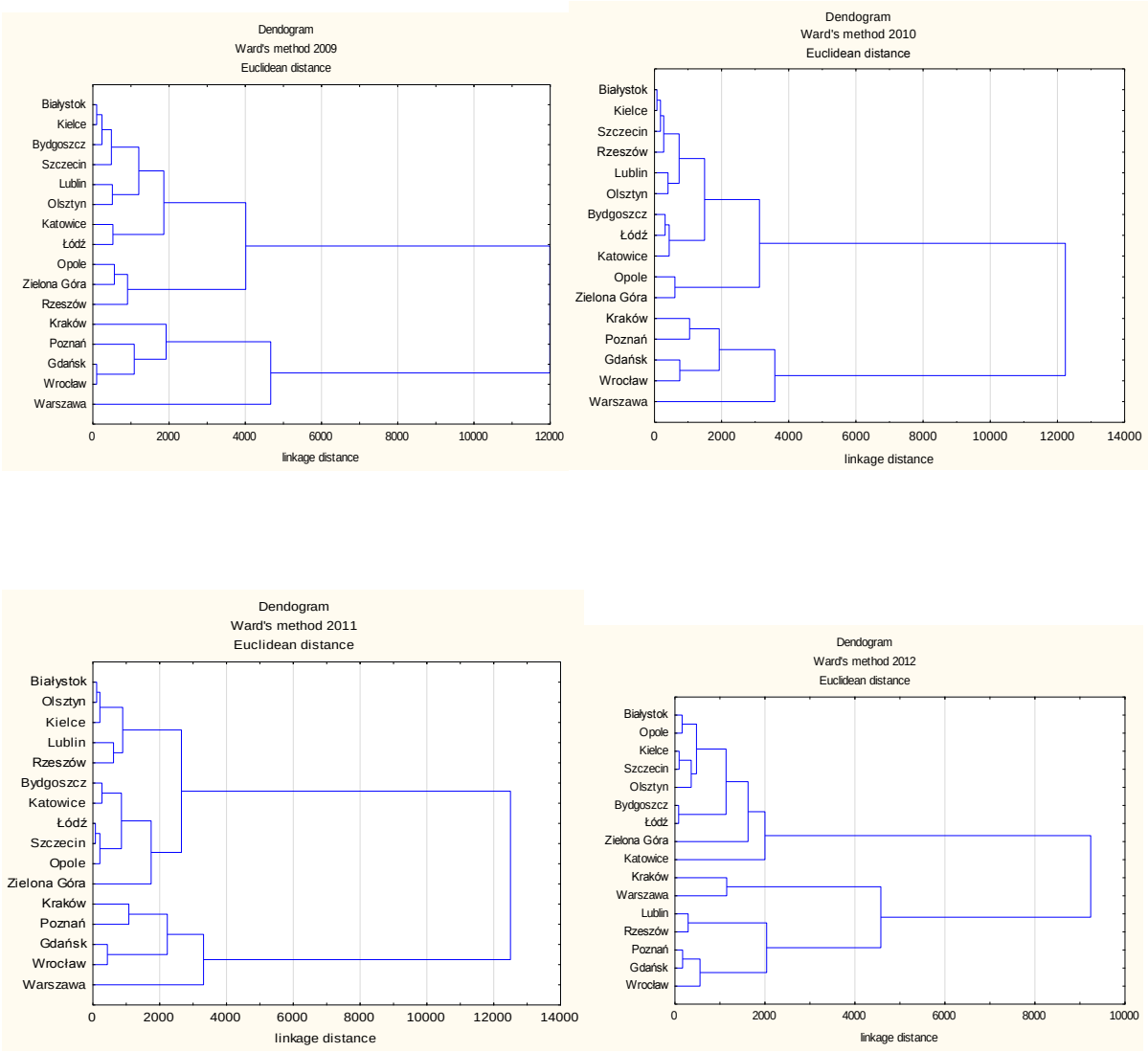
Source: own calculations based on data from NBP (2013)

The obtained values of variability coefficients prove that the features describing the number of entities operating on the real estate market and the transactions concluded on the housing market are characterised by the highest variability. The lowest variability, in turn, is inherent in the features related to the average number of rooms in an apartment, the average usable floor space in an apartment, and the number of apartments per resident.

Due to the fact that the critical value of the variability coefficient was assumed to be $|V^*| = 0{,}05$, all of the above potential diagnostic features were included in the set of diagnostic features. Then, local housing markets were clustered with the use of the Ward's method[7], and the results of the research are presented on the dendrograms below.

[7] The computer program Statistica was used for the calculations.

Fig. 1: Hierarchical ordering of local housing markets with the use of the Ward's method



Source: own work with the use of Statistica program

The conducted clustering of local housing markets allowed the separation of four clusters in the period 2009-2010 and three clusters in the period 2011-2012. The table below includes the classification results.

Tab. 2: Results of clustering local housing markets with the use of the Ward's method in the period 2009-2012

Class	2009	2010	2011	2012
I	Łódź, Katowice, Olsztyn, Lublin, Szczecin, Bydgoszcz,	Łódź, Katowice, Olsztyn, Lublin, Szczecin, Bydgoszcz,	Łódź, Katowice, Olsztyn, Lublin, Szczecin, Opole,	Łódź, Katowice, Olsztyn, Szczecin, Opole, Bydgoszcz,

	Kielce, Białystok	Kielce, Białystok, Rzeszów	Bydgoszcz, Kielce, Białystok, Rzeszów, Zielona Góra	Kielce, Białystok, Zielona Góra
II	Opole, Zielona Góra, Rzeszów	Opole, Zielona Góra	----------	----------
III	Kraków,Poznań, Gdańsk, Wrocław	Kraków,Poznań, Gdańsk, Wrocław	Kraków, Poznań, Gdańsk, Wrocław	Poznań, Gdańsk, Wrocław, Lublin, Rzeszów
IV	Warsaw	Warsaw	Warsaw	Kraków, Warsaw

Source: own work based on Fig. 1

The results of clustering unambiguously prove that the housing market in Warsaw, which is the capital city of Poland, stood out in the analysed years. It needs to be stressed that housing markets in Poznań, Gdańsk and Wrocław are similar to each other (since they are always in the same class), like in the case of Opole and Zielona Góra housing markets. Moreover, Łódź, Katowice, Olsztyn, Szczecin, Bydgoszcz, Kielce, Białystok housing markets are also similar markets.

Conclusion

The research carried out above permitted the classification of local housing markets in Poland and identification of similar markets. Diagnostic features were the ones describing the housing market that are used in the research conducted by the National Bank of Poland. The results of clustering local housing markets with the use of the Ward's method in the period 2009-2012 prove that the housing market in the capital city of Poland is a market that stands out from the rest (it was only in 2012 that the Kraków housing market was similar). It needs to be emphasised that changes occur on local housing markets in Poland, which is proved by shifts of certain markets between classes.

References

Stamos, M. Z., & Fischer, M. (2013). Optimal Life Cycle Portfolio Choice with Housing Market Cycles. *Review of Financial Studies, 26,* 2311-2352.

Renaud, B., Zhang, M., & Koeberle, S. (2001). Real estate and the asian crisis: Lessons of the Thailand experience. *A global perspective on real estate cycles* (Brown S.J.& Liu C). New York: Springer Science+ Business Media.

Mayer, C. (2001). Loss Aversion And Seller Behavior: Evidence From The Housing Market. *The Quarterly Journal of Economics, 116*, 1233-1260.

Lance, G.M., & Williams, W.T. (1967). A General Theory of Classificatory Sorting Strategies I. Hierarchical System. *The Computer Journal*, *9*, 373–380.

Lance, G., & Williams, W. (1967). A General Theory of Classificatory Sorting Strategies II. Hierarchical System. *The Computer Journal*, *10*, 271-277.

Ball, G. H. (1965). Data analysis in the social sciences: What about the details?, *AFIPS Conference Proceedings*, *Fall Joint Computer Conference*, 533-559

Jardine N. & SibsonR. (1968). The construction of hierarchic and non-hierarchic classifications, *The Computer Journal, 11*, 177-184.

Ward, J. H. (1963). Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association, 58*, 236-244.

Rand W.M. (1971). Objective Criteria for the Evaluation of Clustering Methods, *Journal of the American Statistical Association, 66,* 846-850.

Nowak, E. (1990). *Metody taksonomiczne w klasyfikacji obiektów społeczno-gospodarczych*. Warszawa: Państwowe Wydawnictwo Ekonomiczne, (in Polish).

Makridakis, S., Wheelwright, S., & Hyndman, R. (1998). *Forecasting. Methods and Applications*. New York: John Willey & Sons.

Grabiński, T., & Sokołowski, A. (1984). Research over the Effectiveness of Selected Taxonomic Procedures. *Zeszyty naukowe AE w Krakowie*, *181*, 63–79, (in Polish).

Raport o sytuacji na rynku nieruchomości mieszkaniowych i komercyjnych w Polsce w 2012 r.. (2013, April 1). *NBP*. Retrieved April 1, 2014, from http://www.nbp.pl/home.aspx?f=/publikacje/rynek_nieruchomosci/index1.html

Contact

Patrycja Kowalczyk-Rólczyńska
Wrocław University of Economics, Department of Insurance
Komandorska 118/120
53-345 Wrocław, Poland
Mail: patrycja.kowalczyk@ue.wroc.pl