Housing market heterogeneity and cluster formation: evidence from Poland

Mateusz Tomal
Department of Real Estate and Investment Economics, Cracow University of Economics, Krakow, Poland

Received 18 September 2020
Revised 3 November 2020
Accepted 11 November 2020

Abstract

Purpose – This study aims to identify clusters amongst the county housing markets in Poland, taking into account the criteria of size and quality of the housing stock, as well as price level. In addition, this work is intended to detect the socio-economic factors driving the cluster formation.

Design/methodology/approach – To group the studied housing markets into homogeneous clusters, this analysis uses a proprietary algorithm based on taxonomic and k-means++ methods. In turn, the generalised ordered logit (gologit) model was used to explore factors influencing the cluster formation.

Findings – The results obtained revealed that Polish county housing markets can be classified into three or four homogeneous clusters in terms of the size and quality of the housing stock and price level. Furthermore, the results of the estimation of the gologit models indicated that population density, number of business entities and the level of crime mainly determine the membership of a given housing market in a given cluster.

Originality/value – In contrast to previous studies, this is the first to examine the existence of homogeneous clusters amongst the county housing markets in Poland, taking into account the criteria of size and quality of the housing stock, as well as price level simultaneously. Moreover, this work is the first to identify the driving forces behind the formation of clusters amongst the surveyed housing markets.

Keywords K-means, Ripple effect, Driving forces, Cluster formation, Generalised ordered logit model, Housing market heterogeneity, Taxonomic method, Polish counties

Paper type Research paper

1. Introduction

Scientific literature investigating the existence of clusters in the real estate market, and especially in the housing market, is becoming increasingly popular. This is due to the fact that knowledge of actually existing homogeneous groups of real estate markets is important not only for researchers to properly design research methodologies but also for entrepreneurs and policymakers. In the context of the latter, the correct division of the housing market makes it possible to design individual housing policies aimed at solving problems in the identified homogeneous groups. On the other hand, real estate entrepreneurs such as developers, valuers or brokers because of their detailed knowledge of the clusters in

The publication is financed by a subsidy granted to the Cracow University of Economics.

International Journal of Housing Markets and Analysis
Vol. 14 No. 5, 2021
pp. 1166-1185
Emerald Publishing Limited
1753-8270
DOI 10.1108/IJHMA-09-2020-0114

the real estate market, can significantly increase the efficiency of their activities (better accuracy of real estate appraisals, better choice of investment location, etc).

Clustering amongst housing markets can be explained by a number of reasons. Firstly, when looking at the housing market solely in terms of one of its characteristics, i.e. the level of prices, attention should be paid to a phenomenon called the ripple effect. It involves the transmission of price shocks between even spatially distant markets as a result of the mobility of labour, capital and population. The ripple effect phenomenon in the Polish housing market has already been confirmed, both at a micro (Brzezicka *et al.*, 2019) and macro (Tomal, 2020a) level. Moreover, the allocation of housing markets to clusters may result directly from socio-economic factors such as unemployment level and wages. Such factors significantly influence the demand and supply of the housing market and shape its features, not only in terms of the price level but also the size and quality of the housing stock. On the other hand, however, it should be noted that housing is a heterogeneous good which, according to Lancaster's perspective, can be considered as a set of attributes (or characteristics) used to satisfy needs such as shelter or comfort (Maclennan and Tu, 1996). When considering the good of housing in this context, it should be stressed that buyers of potential properties may have different preferences with regard to their structural characteristics. All of this can lead to a significant diversification of residential properties in different locations.

In view of the above, the aim of this article has been formulated, which is to classify the county housing markets in Poland into groups of similar objects (clusters), taking as criteria the size and quality of the housing stock and the price level. This goal will be achieved through a simple proprietary algorithm combining taxonomic analysis and the k-means++ procedure. The second aim of the study is to explore socio-economic drivers of cluster formation using the generalised ordered logit model. Taking into account the defined research objectives, this work seeks to answer the following research questions:

RQ1. Can Polish county housing markets be grouped into homogeneous clusters, taking into account the criteria of size and quality of the housing stock, as well as the price level? If so, how many such groups of similar objects can be distinguished?

RQ2. If the clusters referred to in question *RQ1* are identified, what factors drive the membership of a given housing market in a given cluster?

This study contributes to the current literature on the grouping of housing markets. First of all, it is the first analysis taking into account the criteria of size, quality and price level in the context of cluster identification amongst Polish county housing markets. Secondly, this study is also unique in identifying the drivers of cluster formation. In the analyses to date, only voivodeship cities were studied and not all counties in Poland, i.e. 380 units. Thirdly, this article proposes a very simple author's clustering algorithm, which can be used to group housing markets in any spatial scale, i.e. at mega, macro, meso and micro levels.

The rest of the article is organised as follows. Section 2 presents a review of the existing scientific literature on empirical research on the identification of clusters in the housing market. Section 3 describes the data and methodology of the study. In turn, Section 4 contains the results of the study and its discussion. Section 5 presents the main conclusions resulting from the analysis, research limitations and directions of future studies.

2. Literature review

The division of housing markets into clusters can be done using data-based methods and contractual methods based, for example, on administrative boundaries (Usman *et al.*, 2020). When conducting a study on the delineation of the housing market into sub-markets, the spatial scale of the analysis should also be determined, i.e. clustering at macro, meso and micro levels. In the former case, the national housing market is divided into homogeneous groups of sub-markets of regional range, e.g. corresponding to the borders of counties or voivodeships. In the case of the meso scale, the analysis would aim to identify similar sub-markets of local character from one regional market. The micro-analysis, on the other hand, would cover one local market, e.g. concerning a given city and would be aimed at selecting groups of areas of homogeneous character. In this context, the spatial scale corresponding to the borders of several countries can also be considered. In this case, we should speak of a mega level, where the study would focus on creating homogeneous clusters amongst national housing markets. It should be noted, however, that a survey of the real estate market at a mega level is quite rare because it is, by its very nature, local.

Given the housing market in Poland, research into its division into homogeneous sub-market groups has been undertaken quite often by scientists. In particular, Cellmer and Jasiński (2016) classified the county residential markets in Poland into five groups taking into account their size and activity. Similarly, Brzezicka and Wisniewski (2016), but on a provincial scale, segmented the Polish housing market into four clusters taking into account its size in relation to population. Also at the voivodeship level, the classification of housing markets was performed by Bera and Śpiewak-Szyjka (2019), who divided the Polish housing market into four similar groups. In the above study, however, the authors undertook to examine only the municipal housing stock and, contrary to previous analyses, also took into account its quality. The latter element was also the subject of research in the work by Kozera and Stanisławska (2019), which identified six clusters amongst the voivodeship housing markets. It should be noted, however, that in the abovementioned study, only rural areas were analysed. A separate group of surveys are also works in which the division of the housing market into homogeneous sub-market groups was made exclusively on the basis of the average price of residential properties. One can distinguish here the analyses carried out by Tomal (2019a, 2019b, 2020a), Belej and Kulesza (2014), Brzezicka *et al.* (2019) and the study carried out by Dittmann (2018) where the subject of the research was the residential availability index (relation of average monthly salary to the average price of housing property). However, the only analysis which took into account all the above aspects of the housing market in Poland, i.e. its size, quality and value was the work done by Kowalczyk-Rólczyńska (2014). In this study, the author segmented the Polish real estate market into four groups, but the subject of the analysis was only the provincial capitals. Other studies in this aspect also include works carried out by Tomal (2020b), Kokot (2020) and Głuszak and Marona (2011), where the criterion for the division of the residential market was the degree of its smartness, the socio-economic features of the area in which it is located and the preferences of its current buyers, respectively.

When analysing the literature on the subject in the context of research on the division of the housing market in countries other than Poland, it should be noted that the vast majority of analyses concern the micro and meso scales, as indicated by the review carried out by Islam and Asami (2009). One such survey was conducted by Bourassa *et al.* (2003) for the city of Auckland in New Zealand. In particular, using the PCA method and then the k-means algorithm, the authors divided the housing market in the examined city into 14–18 homogenous clusters. In this study, the definition of housing sub-markets was based on the physical characteristics of individual properties, their location and the socio-economic characteristics of their surroundings. In methodological terms, a very similar analysis was performed by Wu and Sharma (2012), who identified 15 homogeneous areas in the housing market in Milwaukee, USA. The real estate market segmentation based on cluster analysis was also conducted by Kim and Park (2005). In this case, the study area covered the city of Seoul and neighbouring towns, and therefore it can be concluded that the analysis was performed at meso level. The authors divided the studied market into four clusters taking into account two variables, i.e. the average price of residential properties and the average rate of change of residential prices. It should be noted that in the scientific literature on the classification of the housing market at the micro and meso levels, a number of works can be distinguished based on the classical hedonic model. Amongst other things, this type of study was carried out by Watkins (2001), who investigated the city of Glasgow. In particular, the author compared the estimates of the parameters of hedonic models used for different types of residential properties and for different areas of the city. The results of the study confirmed that the analysed housing market can be divided based on spatial and structural factors. A similar study was conducted by Alkay (2008), who identified significant differences in implicit attribute prices between three predefined sub-markets based on average household income. Hedonic modelling for the division of the residential market was also used by Wilhelmsson (2004) on the example of the city of Stockholm. However, in this case, the author, contrary to the above-mentioned analyses, did not examine the parameters of the model, but the residuals to take into account the unobservable characteristics of properties. The error term was then subjected to a clustering procedure, which allowed the detection of sub-markets in Stockholm. It should be noted that the classical hedonic model is now giving way to more advanced techniques. In the context of the housing market, geographically weighted regression, which, unlike the hedonic model, takes account of spatial heterogeneity, is an increasingly common tool. The GWR model has been used, amongst others, by McCluskey and Borst (2011) to detect residential sub-markets in Catawba County, North Carolina; Sarasota County, FL; and Fairfax County, Virginia. Another increasingly popular method of grouping housing markets into homogeneous clusters is the so-called analysis of convergence clubs using the approach outlined by Phillips and Sul (2007). In this respect, the meso level study for selected US metropolitan areas was conducted by Apergis and Payne (2019, 2020), amongst others, distinguishing several convergence clubs.

With regard to the literature, in which the subject of the analysis is different from the Polish housing market, one should emphasise the small number of studies at the macro level, i.e. covering the whole selected country. In this respect, the work carried out by Kauko and Goetgeluk (2005), who grouped district housing markets into homogeneous groups using a neural network technique in The Netherlands, should be distinguished. In the above-mentioned study, the clustering was based on a number of variables describing districts in terms of socioeconomic features and property values. Another such survey was performed by Helbich *et al.* (2013), who used data on individual property transactions and their characteristics using the MGWR method and then clustering procedures to distinguish several residential sub-markets in Austria. Country-wide analyses can also be seen for the UK. In this respect, it is important to highlight the studies conducted by Montagnoli and Nagayasu (2015) and Holmes *et al.* (2019), which analysed price convergence clubs. In the context of the macro scale, mention should also be made of a study by Abraham *et al.* (1994), who grouped into clusters, using the k-means procedure, 30 metropolitan areas in the US, taking into account real annual housing returns.

On the basis of the literature review, it must be concluded that there is still a scientific gap in the issues addressed. Above all, research to date on the classification of housing markets into homogeneous groups usually ignores the aspect of the size and quality of the housing stock. Another disadvantage of the studies to date is that the number of clusters has often been assumed a priori, as for example, in the study carried out by Kowalczyk-Rólczyńska (2014), where the author assumed in advance that the Polish housing market should be divided into four groups of similar objects. In addition, the vast majority of analyses to date have omitted a detailed assessment of the characteristics of the estimated clusters in terms of their socio-economic characteristics. This applies especially to the Polish housing market, where apart from the studies at the level of voivodeship capitals performed by Matysiak and Olszewski (2019) and Tomal (2019b) there are no other analyses. To sum up, there is no study in the scientific literature to date, which would cluster all Polish district housing markets using the criteria of size and quality of the housing stock, as well as price level at the same time. Additionally, there is also a lack of an analysis, which would attempt to identify the factors driving the formation of clusters for all Polish counties.

3. Data and methodology

3.1 Study area

All Polish county housing markets are the subject of this study. It should be stressed that the choice of Poland as a place for analysis is not accidental and results from several premises. First of all, in comparison with others, especially Western European countries, the housing market in Poland is insufficiently researched. Secondly, Poland is one of the largest countries in Europe both in terms of area and population, which implies that its housing market also has an impact on the housing markets in other countries, especially the neighbouring ones. Thirdly, it should be noted that the Polish housing market is currently struggling with the problem of a lack of supply of housing. Despite the dynamic growth of new housing stock, Poland still lacks about 2–3 million flats to reach at least the EU average in terms of the number of dwellings per 1,000 citizens. This fact underlines the importance of research into the Polish housing market, which may contribute to its better functioning. Finally, in recent years, the Polish housing market has seen more and more new institutional housing projects, mainly for rent. Therefore, research on the segmentation of the residential market is extremely useful for foreign and domestic investors to choose the right location for their future investments.

3.2 Data

This study uses data from the Polish Statistical Office on the size and quality of the housing stock and the price level on the surveyed housing markets for 2018. The size criterion takes into account not only the total number of available dwellings but also the activity observed on a given market. With regard to the quality criterion, the focus was on housing deprivation indicators. In turn, within the last criterion, i.e. the price level, attention was paid not only to the average value but also to the median value, which is more resistant to outlier observations. Table 1 presents in detail the indicators describing the abovementioned aspects of the housing market and the descriptive statistics estimated for them.

On the basis of descriptive statistics presented in Table 1, it should be concluded that the surveyed county housing markets in Poland are very heterogeneous in terms of the size of the housing stock and the average price level. As far as the quality of the housing stock is concerned, the variation between the surveyed districts is quite low, with the exception of the indicator concerning the level of equipping dwellings with mains gas. The latter conclusion has already been recognised in other studies; amongst other things, reference should be made to the study carried out by Tomal (2020c), which states that poor gas supply to municipalities is one of the main obstacles to local development in Polish rural areas. It should also be noted that due to very low volatility, variables Q2-Q4 were omitted from further analysis.

The aim of the article is also to identify factors influencing the membership of a given housing market in a given cluster. In this article, due to the macro research, the focus is on factors of a socio-economic nature that have a significant impact on demand and supply in the housing market. The selection of variables was based on the existing analyses of the housing market in terms of price determinants and their detailed characteristics are presented in Table 2.

Indicators	Type	Minimum	Maximum	Mean	SD	CV
Panel A: Size						
S1: Dwellings per 1,000 inhabitants	Stimulant	248.40	548.80	356.06	50.30	0.14
S2: Dwellings completed per 1,000 inhabitants	Stimulant	0.60	16.94	3.55	2.52	0.71
S3: Number of dwellings sold under market transactions per 1,000 inhabitants	Stimulant	0.00	20.75	3.53	3.26	0.92
Panel B: Quality						
Q1: Average usable floor area of the dwelling per 1 person [m^2]	Stimulant	22.50	42.50	27.71	2.99	0.11
Q2: Percentage of dwellings equipped with water supply	Stimulant	0.80	1.00	0.96	0.04	0.04
Q3: Percentage of dwellings equipped with a flushed paragraph	Stimulant	0.70	1.00	0.92	0.06	0.07
Q4: Percentage of dwellings equipped with a bathroom	Stimulant	0.66	1.00	0.89	0.07	0.08
Q5: Percentage of dwellings equipped with central heating	Stimulant	0.56	0.99	0.79	0.09	0.11
Q6: Percentage of dwellings equipped with mains gas	Stimulant	0.00	0.97	0.43	0.28	0.66
Panel C: Price						
P1: Median dwelling price per 1 m^2 [PLN]	Stimulant	1174.00	11161.00	3193.09	1081.99	0.34
P2: Average dwelling price per 1RQ1 m^2 [PLN]	Stimulant	1029.00	11763.00	3135.08	1112.64	0.35

Notes: SD stands for standard deviation and CV for the coefficient of variation. By dwelling is meant both in multi-family buildings and also in single-family houses, semi-detached houses, etc. As of 08–09-2020, US $1 is equally about 3.80 PLN

Table 1. Descriptive statistics of indicators describing the defined dimensions of the housing market

Driver	Abbrev.	Minimum	Maximum	Mean	SD	CV	References
X1: Population density	PD	19.00	3,757.00	369.46	654.27	1.77	Hazam and Felsenstein (2007)
X2: Unemployment rate	UR	1.20	24.30	7.80	4.06	0.52	Lin *et al.* (2014) and Tomal (2019a)
X3: Average gross monthly salaries [PLN]	AS	3,183.34	8,121.08	4,142.14	561.24	0.14	Lee (2009) and Nistor and Reianu (2018)
X4: Entities included in the REGON register per 10,000 population	ER	524.00	2,440.00	965.27	280.34	0.29	Cellmer *et al.* (2020)
X5: Offences identified by the Police per 1,000 inhabitants	OP	7.38	71.78	18.41	8.44	0.46	Hazam and Felsenstein (2007) and Thaler (1978)
X6: Doctors per 10,000 population	DP	2.00	198.20	40.55	29.86	0.74	Rivas *et al.* (2019)

Table 2. Characteristics of potential factors influencing cluster membership

When analysing the Polish housing market in terms of selected socio-economic characteristics, it is important to emphasise the large variation in population density, which is not surprising. There is also a significant diversity with regard to the number of doctors per 10,000 inhabitants. It results from the fact that there is a very serious problem in Poland in terms of the number of medical staff. The data presented in Table 2 confirm this situation. In particular, there are counties where there is no more than 1 doctor per 1,000 people. On the other hand, there are units where this number exceeds 19 people. In the scope of other variables, much less heterogeneity is visible.

3.3 Methodology

An original algorithm of grouping studied housing markets into homogeneous clusters can be presented in the following stages:

- Step 1. Estimation of synthetic indexes for each county housing market describing the size and quality of the housing stock and the average price level.

This step requires, in the first stage, the standardisation of variables using the zero unitarisation formula:

$$z_{ij} = \frac{x_{ij} - \min_i[x_{ij}]}{\max_i[x_{ij}] - \min_i[x_{ij}]} \tag{1}$$

where x_{ij} (z_{ij}) is the original (standardised) value of the j-*th* indicator for the i-*th* housing market and $\max_i[x_{ij}]$ and $\min_i[x_{ij}]$ denote the maximum and minimum value of the j-*th* indicator amongst the i-*th* housing markets, respectively. It should be stressed that the standarisation of variables is necessary to bring all indicators down to comparable values. There are many standardisation formulas in the literature, but it is the zero unitarisation that meets all requirements for this type of transformation. Amongst other things, it ensures non-zero values and the stability of ranges of variability of the standardised variables (Tomal and Nalepka, 2020).

In the second stage of step 1, the following formula shall then be used to calculate a synthetic index for each defined housing market dimension:

$$CL_i^s = \frac{1}{n^s}\sum_{j=1}^{n^s} z_{ij}^s \tag{2}$$

where n^s denotes the number of indicators in criterion $s = \{size, quality, price\}$, CL_i^s is the value of the synthetic index for the i-*th* housing market in terms of criterion s. This study assumes equal weights for individual indicators describing a given dimension of the housing market. It should be noted that if the weights for the defined variables were to be adjusted on the basis of raw data, the results of the analysis could be significantly distorted. This is due to the fact that procedures (e.g. the entropy weight method) used for this purpose usually assign high weights for indicators with high variation. In the context of this study, this would mean that the quality of the housing stock in the district would be assessed largely by the variable describing the availability of a gas installation in the dwelling, which is not a desirable situation because other installations such as central heating, are equally important.

Once all the stages in the first step have been completed, each studied housing market (h_i) will be described in each dimension under consideration, i.e. in terms of quality and size of the housing stock and price level, creating a set of points in a three-dimensional space where each element can be described as $h_i\left(CL_i^{size}, CL_i^{quality}, \ CL_i^{price}\right)$.

- Step 2. Grouping the studied housing markets into homogeneous clusters and determining their exact numbers based on k-means++ method developed by Arthur and Vassilvitskii (2006). In the baseline scenario, the clustering is based on three synthetic measures calculated in step 1, describing each of the housing markets under consideration in terms of size, quality and value of the housing stock.

To identify clusters, this study uses the aspatial k-means++ method because the subject of the analysis is a new three-dimensional space, created on the basis of calculations from step 1. Therefore, the use of methods based on spatial correlation is inappropriate. The k-means++ method itself can be characterised as follows:

- Determination of the number of clusters;
- Selection of k centroids – initialisation. It should be noted that in comparison with the standard k-means method, which randomly selected the first centroids, the k-means++ procedure uses a so-called smart initialisation. It involves selecting one centroid in the first step randomly. The distance between it and all points is then calculated. The next centroid is the point with the greatest squared distance from the first centroid. To select the third centroid, the distance between each point and its nearest centroid is calculated. The point with the greatest squared distance is selected as the new centroid. The entire initialisation procedure continues until k centroids are selected;
- Matching each h_i point to the nearest centroid;
- Determination of new centroids based on the mean of all points in a given cluster; and
- Repeat two previous steps until no change in centroid position is achieved.

The key stage in the procedure described above is to determine the number of clusters. To determine the number of groups of similar objects, the elbow method will be applied, taking into account the values of the total within-cluster sum of squares after each launch of the k-means++ algorithm, assuming the existence of from 2 to 10 clusters.

The next stage of the analysis will be a detailed examination of what socio-economic factors influence cluster membership. For this purpose, the generalised ordered logit model (gologit) will be used, as when using the ordinary ordered logit model (ologit), the parallel regression assumption is violated according to the brant test. In particular, the gologit model in this study takes the following form:

$$P(Y_i > m) = \frac{\exp\left(\alpha_m + \sum_{k=1}^{6} X_{ik}\beta_{mk}\right)}{1 + \left[\exp\left(\alpha_m + \sum_{k=1}^{6} X_{ik}\beta_{mk}\right)\right]}, \quad m = \{1, 2, \ldots, M-1\} \tag{3}$$

where Y_i is the dependent variable which takes the value $m = \{1,2,\ldots,M-1\}$, M denotes the number of identified clusters, α_m is the constant of the model for the m-*th* cluster, X_{ik} represents independent variables, β_{mk} expresses model parameters. It should be noted that the advantage of the gologit model over the ologit model is that the former matches the unique parameters of the model for the $M - 1$ category of the dependent variable (Williams, 2016).

To perform the calculations under this article, the following has been used: GEODA software to identify clusters; QGIS and RGui software to visualise the study results and STATA software to estimate the gologit models.

4. Results and discussion

4.1 Cluster analysis – baseline scenario

In the first stage of the empirical study, the studied housing markets were classified into groups using the k-means++ method and assuming that 2 to 10 clusters exist in the data set. This procedure was aimed at selecting an optimal number of clusters amongst the analysed counties. The results of this study are presented in Figure 1, on the basis of which it can be concluded that the so-called elbow point is visible for 3 or 4 clusters. For clusters from 5 to 10, the decrease in the value of the total within-cluster sum of squares is much smaller. However, due to the fact that it is difficult to clearly determine whether the sample should be divided into 3 or 4 groups of similar objects, further analysis will be carried out for both variants (a detailed list of districts assigned to given clusters is available in Tables A1 and A2 in Appendix).

It was then decided to look at the characteristics of the clusters created, taking into account the average values of the size, quality and price index. In particular, on the basis of Table 3, it can be concluded that by far the highest values in all analysed dimensions of the housing market can be seen in the last clusters. In the case of clusters 1 and 2 in Panel A and 1–3 in Panel B, the differences between groups are much smaller. It can, therefore, be stated that there is quite a large heterogeneity in the Polish housing market with a group of about

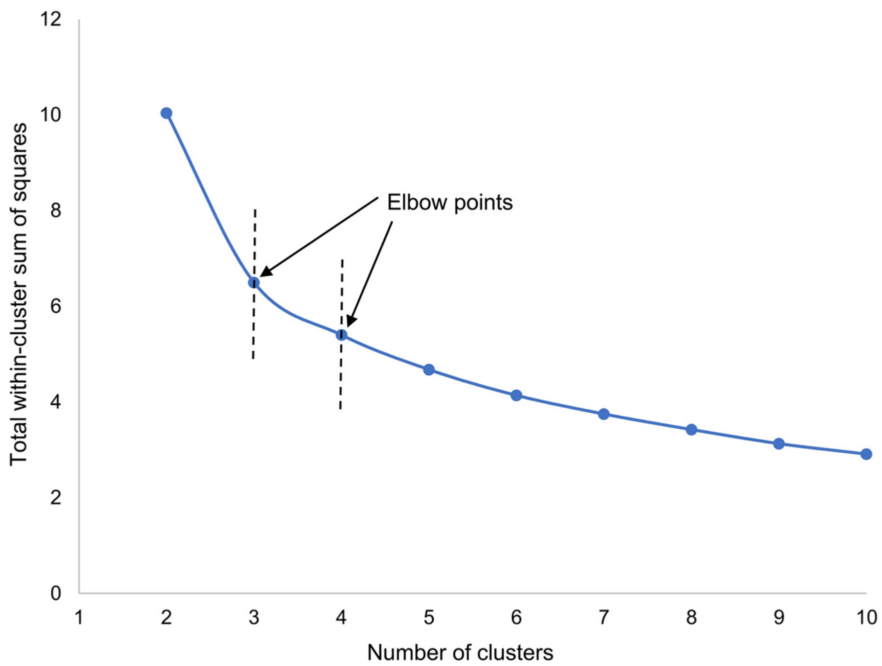

Figure 1.
Selection of the number of clusters

30–40 counties, which in terms of the size of the housing market, the quality of its stock and the prices observed differ significantly from other markets. In this context, one more characteristic feature of estimated clusters can be observed, i.e. regardless of their number (3 or 4) in each subsequent cluster the values of the studied dimensions of the housing market (size, quality, price) are increasing – this dependence is presented in detail in Figure 2(a,b). Therefore, it can be stated that in Poland the development of residential markets is taking place simultaneously in all studied aspects.

Figure 2(a,b) also provides other extremely interesting information. First of all, in the first clusters, the differences in the characteristics of the housing market in the studied counties are very small, in other words, these housing sub-markets are extremely similar in terms of size, quality and price level. Taking into account cluster 2 in Panel A and clusters 2 and 3 in Panel B, a similar relationship can also be observed, but the level of differentiation is increasing. The biggest disparities between the analysed markets can be observed in cluster 3 in panel A and cluster 4 in panel B.

It is also interesting to check what is the spatial distribution of the studied housing markets assigned to given clusters. As we can see in Figure 3(a,b) the surveyed counties within individual clusters are unevenly distributed over space. The fact that even housing

Cluster	Size index – mean	Quality index – mean	Price index – mean	Quantity
Panel A: 3-cluster division				
1	0.16	0.28	0.15	163
2	0.24	0.48	0.20	178
3	0.52	0.66	0.41	39
Panel B: 4-cluster division				
1	0.16	0.26	0.15	132
2	0.19	0.43	0.17	131
3	0.31	0.54	0.23	87
4	0.57	0.67	0.43	30

Table 3. Characteristics of estimated clusters

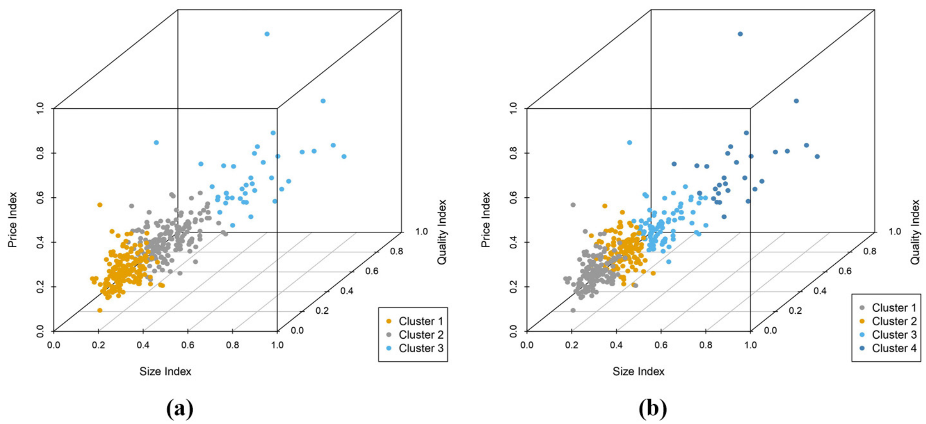

Figure 2. 3D visualisation of studied housing markets divided into (a) 3 clusters; (b) 4 clusters

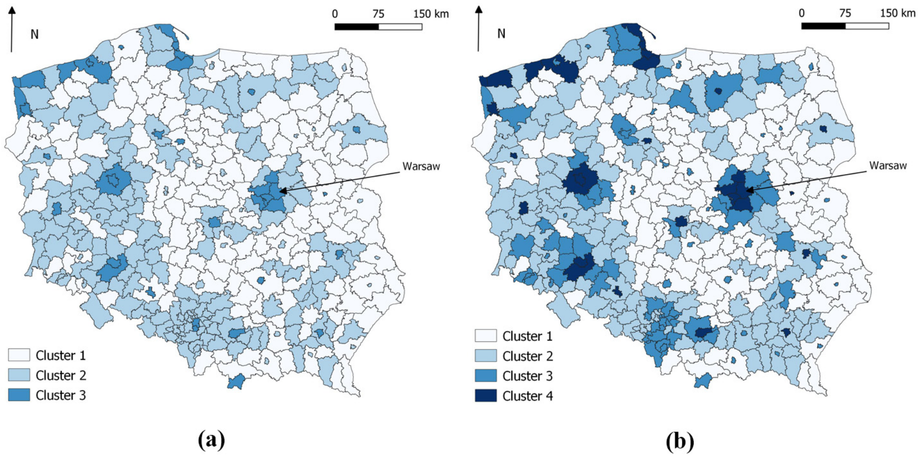

Figure 3.
Spatial distribution of studied housing markets divided into (a) 3 clusters; (b) 4 clusters

markets at the two ends of Poland belong to the same cluster can be explained by the occurrence of a phenomenon called the ripple effect and similar socio-economic characteristics of the area where these housing markets operate. The only spatial dependence that is noticeable concerns the main Polish cities and districts surrounding them, i.e. voivodeship capitals usually belong to the same cluster and the districts surrounding them form their own separate cluster. In addition, it should be noted that the districts located in the first cluster with the least developed housing markets are mostly located in the eastern part of Poland. This clearly visible division into Poland A and B, which is also noticeable in this study, is the result of the partition of Poland by Russia, Austria and Prussia, which lasted from the end of the 18th to the beginning of the 20th century.

4.2 Drivers of cluster formation – baseline scenario

In the next stage of the study, the gologit models were estimated to learn about the socio-economic factors influencing the formation of clusters amongst Polish housing markets. The estimation of the model was carried out in two ways, i.e. in the first one, independent variables were not transformed in any way, while in the second one their logarithmic form was used. Analysing in detail the results of the estimation presented in Table 4, it should be noted that out of 6 independent variables, two, i.e. UR and DP do not affect the formation of clusters in the Polish housing markets. In the context of the first of these variables, describing the unemployment rate, this conclusion is contrary to the studies carried out by Tomal (2019b). It should be noted, however, that the aforementioned studies concerned price convergence and the identified convergence clubs were analysed. On the other hand, variables describing the demographic (PD) and economic (ER) conditions of an area have an impact on the formation of clusters in almost every case. In particular, taking into account models with not transformed independent variables, an increase in population density by one unit reduces the likelihood of assigning a given housing market to cluster 1 (in panels A and B) by about 0.003 and increases the likelihood of a given market being present in clusters 2 and 3 (panel A) and 3 and 4 (panel B) by 0.00002–0.00351 depending on the specific cluster. In relation to the ER variable, on the other hand, an increase by one unit

Table 4. Estimation results from the gologit model for 3- and 4-cluster division

Variable	Average marginal effects: Not transformed independent variables				Independent variables logged			
Panel A: 3-cluster division								
	Cluster = 1	Cluster = 2	Cluster = 3		Cluster = 1	Cluster = 2	Cluster = 3	
PD	–0.00283***	0.00281***	0.00002*		–0.24079***	0.22169***	0.01910	
UR	0.00700	–0.00691	–0.00009		0.05607	–0.06059	0.00452	
AS	–0.00013***	0.00012***	0.00002		–0.54274***	0.44383***	0.09891	
ER	–0.00070***	0.00041***	0.00028***		–0.57146***	0.18796***	0.38350***	
OP	–0.00450*	0.00639**	–0.00189**		–0.10735**	0.15664***	–0.04928*	
DP	0.00066	–0.00079	0.00013		–0.00073	0.00724	–0.00651	
Panel B: 4-cluster division								
	Cluster = 1	Cluster = 2	Cluster = 3	Cluster = 4	Cluster = 1	Cluster = 2	Cluster = 3	Cluster = 4
PD	–0.00361***	0.00351***	0.00006	0.00004**	–0.29690***	0.20725***	0.06349***	0.02616*
UR	0.00166	0.00448	–0.00448	0.00172	0.00019	–0.00123	–0.01441	0.01545
AS	–0.00013***	0.00001	0.00011***	0.00002	–0.52478***	–0.03216	0.48239***	0.07456
ER	–0.00072***	–0.00006	0.00050***	0.00027**	–0.55484***	–0.14030	0.32041***	0.37474***
OP	–0.00807***	0.01019***	0.00083	–0.00295***	–0.14560***	0.19399***	0.01942	–0.06781***
DP	0.00074	–0.00291*	0.00244**	–0.00027	0.00184	0.00320	0.02599	–0.03103

Notes: *** 1% level of significance. ** 5% level of significance. * 10% level of significance. $n = 380$

results in a decrease in the likelihood of assigning a given housing market to cluster 1 by about 0.0007 for both the 3- and 4-cluster divisions. In the case of clusters 2–4, average marginal effects are positive, i.e. an increase in the ER variable by one unit results in an increase in the likelihood of the presence of a given housing market in these clusters by about 0.00027–0.0005 (excluding cluster 2 for Panel B). Very similar results can be observed for the AS variable, but in this case, average marginal effects are insignificant for more clusters.

Extremely interesting conclusions can be drawn by analysing the OP variable concerning the number of identified crimes. In particular, the probability of the presence of a given housing market in clusters 1 and 3 from panel A and 1 and 4 from panel B decreases by about 0.002–0.008 when the number of committed crimes per 1,000 inhabitants increases by one unit. It should be noted that the above clusters include in their housing markets diametrically different from each other in all studied dimensions. On the other hand, for "central" clusters, positive average marginal effects are visible, i. e. an increase in the number of crimes implies an increased chance of a given housing market belonging to these clusters. To sum up, we can see here a very strong reversed U-shaped pattern of average marginal effects across clusters (Figure 4). On the basis of these exceptionally interesting observations, it can also be stated that both the least and most developed housing markets in Poland are located in areas with less crime in comparison to housing markets with an average level of development. This phenomenon is difficult to explain on an *ad hoc* basis, and therefore it should be the subject of further studies.

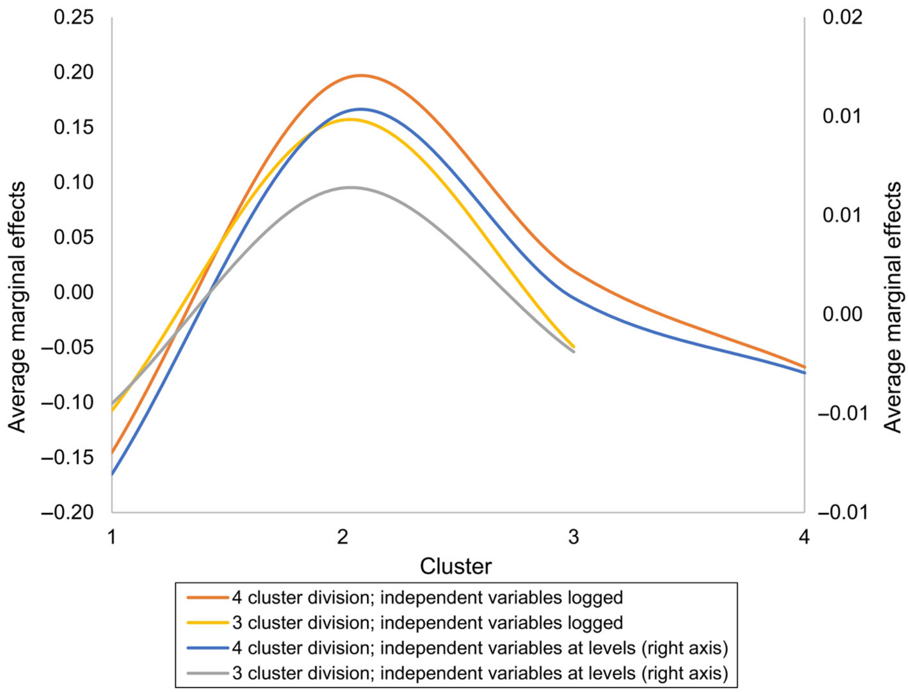

Figure 4.
Average marginal effects for the OP variable

4.3 Robustness checks

In the last stage of the empirical study, it was decided to compare the results of clustering in the baseline scenario to other alternative scenarios. In particular, the housing markets analysed were divided into groups according to the assumptions defined in seven new scenarios. In the first six of them, an analysis was done with the k-means++ method described in this article, but clustering itself was based on one or two dimensions of the housing market, which were characterised in the data and methodology section. The last alternative scenario uses all the examined criteria of the housing market, i.e. the size and quality of the housing stock, as well as price level. In this case, however, the clustering of studied markets will be performed based on the spectral technique. Unlike the k-means++ procedure, this method is more flexible and does not assume the form of clusters in advance, as in the case of the former algorithm. Detailed results concerning the compliance of the division of the studied housing markets into clusters between the baseline and alternative scenarios are presented in Table 5. The first conclusion that can be drawn from the conducted study is that in the case of the alternative scenarios taking into account the quality index, the allocation of the analysed housing markets is very similar to the baseline scenario. Therefore, if this dimension of the housing market is not taken into account, erroneous conclusions can be reached. However, with regard to the alternative scenario, which was based on the spectral clustering method and included all the criteria examined (size, quality, price), it can be concluded that it almost entirely confirms the results obtained in the baseline scenario, which confirms the previously presented research results and conclusions drawn under the baseline scenario.

5. Conclusions

5.1 Main findings

This article attempts to group Polish county housing markets into clusters based on size, quality and price. In particular, the research was aimed at answering the following research questions:

Alternative scenario definition	Clustering method	Compatibility of the match with the baseline scenario(%)
Panel A: 3-cluster division		
Size and quality indexes	k-means++	94.21
Size and price indexes	k-means++	33.42
Quality and price indexes	k-means++	87.89
Only size index	k-means++	32.37
Only quality index	k-means++	81.58
Only price index	k-means++	40.00
Size, quality and price indexes	Spectral	85.26
Panel B: 4-cluster division		
Size and quality indexes	k-means++	88.42
Size and price indexes	k-means++	34.47
Quality and price indexes	k-means++	80.79
Only size index	k-means++	32.63
Only quality index	k-means++	69.74
Only price index	k-means++	48.68
Size, quality and price indexes	Spectral	81.05

Table 5. Compliance of the alignment of the surveyed housing markets to clusters between the baseline and alternative scenarios

RQ1. Can Polish county housing markets be grouped into homogeneous clusters, taking into account the criteria of size and quality of the housing stock, as well as the price level? If so, how many such groups of similar objects can be distinguished?

RQ2. If the clusters referred to in question *RQ1* are identified, what factors drive the membership of a given housing market in a given cluster?

With regard to Question *RQ1*, this study has shown that Polish county housing markets can be classified into 3 or 4 homogeneous clusters taking into account the criteria analysed. In particular, the identified groups differ significantly in each of the studied dimensions, but the development of the housing markets takes place evenly in all analysed aspects. It should also be stressed that counties located in cluster 3 (panel A) and cluster 4 (panel B) are significantly different in terms of the analysed criteria from the other studied markets, creating a kind of base of the national housing market in Poland. This study also successfully answered question *RQ2*. Namely, the gologit models revealed that the main variables influencing the formation of clusters are PD (population density), ER (entities included in the REGON register per 10,000 population) and OP (offences identified by the Police per 1,000 inhabitants). In the case of the latter variable, an extremely interesting reversed U-shaped pattern of average marginal effects across clusters was noted, which requires further study.

5.2 Research limitations

It should be stressed that this study also has certain limitations. First of all, due to a lack of data, certain aspects of the housing market have not been taken into account, including the relationship between tenants and property owners. In addition, the lack of data has also ruled out the possibility of investigating smaller housing markets, for example, within municipalities. Finally, also due to the fact that some data on the Polish housing market are only available for the period 2015–2018, it was not possible to check whether the defined clusters change over time. The above limitations also determine the directions of future studies, which include the grouping of municipal housing markets in Poland and a study on the stability of clusters in time and space.

5.3 Research implications

It should be noted that the results of this study are very important for all real estate players, from households to policy-makers. In the context of the former, the segmentation of the real estate market into clusters, while taking into account the quality, size and value of the housing stock, allows an easy choice of alternative places to live (Gavu and Owusu-Ansah, 2019) adapted to their preferences and financial possibilities.

With regard to policy-makers, the findings of this survey make it possible to propose recommendations for housing policy. The aim of the latter, as Donner (2000) notes, is to ensure that residents have access to the adequate housing stock in terms of size and quality at a reasonable price. Therefore, Polish policy-makers should pay particular attention to the housing markets in clusters 1 from panels A and B, which are characterised by the low quality of the housing stock. It should also be noted that these markets do not have a large amount of housing stock, which increases the likelihood of effective state intervention. On the other hand, the areas in clusters 3 and 4 of Panel A and Panel B, respectively, should also be of interest to policy-makers in Poland, due to the fact that flats located in these markets are characterised by a significantly higher level of prices in comparison with other parts of the country. Therefore, policies to increase housing availability should be introduced in these housing markets.

The division of the Polish housing market by means of the criteria proposed in this article is also extremely important for real estate entrepreneurs. On the one hand, the results of this research may be useful for residential developers, both Polish and foreign, to select an appropriate investment location. On the other hand, smaller entrepreneurs and valuers, in particular, may also benefit from this study. Specifically, on the basis of the segmentation carried out, appraisers can easily identify similar housing markets to ensure the accuracy of their estimates when valuing residential properties (Royuela and Duque, 2013).

References

Abraham, J.M., Goetzmann, W.N. and Wachter, S.M. (1994), "Homogeneous groupings of metropolitan housing markets", *Journal of Housing Economics*, Vol. 3 No. 3, pp. 186-206.

Alkay, E. (2008), "Housing submarkets in Istanbul", *International Real Estate Review*, Vol. 11 No. 1, pp. 113-127.

Apergis, N. and Payne, J.E. (2019), "Convergence in condominium prices of major US metropolitan areas", *International Journal of Housing Markets and Analysis*, Vol. 12 No. 6, pp. 1113-1126.

Apergis, N. and Payne, J.E. (2020), "Florida metropolitan housing markets: examining club convergence and geographical market segmentation", *Journal of Housing Research*, Vol. 28 No. 2, pp. 145-163.

Arthur, D. and Vassilvitskii, S. (2006), "K-Means++: the advantages of careful seeding", Technical Report No. 2006–13, Stanford InfoLab, Stanford, available at: http://ilpubs.stanford.edu:8090/778/ (accessed 2 November 2020).

Belej, M. and Kulesza, S. (2014), "Similarities in time-series of housing prices on local markets in Poland", *Real Estate Management and Valuation*, Vol. 22 No. 3, pp. 45-53.

Bera, M. and Ś piewak-Szyjka, M. (2019), "Analysis of municipal real estate resources in Poland", *World of Real Estate Journal*, Vol. 2 No. 108, pp. 56-68.

Bourassa, S.C., Hoesli, M. and Peng, V.S. (2003), "Do housing submarkets really matter?", *Journal of Housing Economics*, Vol. 12 No. 1, pp. 12-28.

Brzezicka, J., Łaszek, J., Olszewski, K. and Waszczuk, J. (2019), "Analysis of the filtering process and the ripple effect on the primary and secondary housing market in Warsaw, Poland", *Land Use Policy*, Vol. 88, p. 104098.

Brzezicka, J. and Wisniewski, R. (2016), "Translocality on the real estate market", *Land Use Policy*, Vol. 55, pp. 166-181.

Cellmer, R., Cichulska, A. and Bełej, M. (2020), "Spatial analysis of housing prices and market activity with the geographically weighted regression", *ISPRS International Journal of Geo-Information*, Vol. 9 No. 6, p. 380.

Cellmer, R. and Jasiński, J. (2016), "Analysis of housing development activity in Poland from 2005-2014", *Real Estate Management and Valuation*, Vol. 24 No. 4, pp. 47-58.

Dittmann, I. (2018), "Spatial and temporal diversification of the income affordability of housing in Poland", *Real Estate Management and Valuation*, Vol. 26 No. 4, pp. 54-67.

Donner, C. (2000), *Housing Policies in the European Union*, Theory and Practice, Vienna, Austria.

Gavu, E.K. and Owusu-Ansah, A. (2019), "Empirical analysis of residential submarket conceptualisation in Ghana", *International Journal of Housing Markets and Analysis*, Vol. 12 No. 4, pp. 763-787.

Głuszak, M. and Marona, B. (2011), "Heterogeneity and clustering of housing demand: case study", *Journal of International Studies*, Vol. 4 No. 1, pp. 89-97.

Hazam, S. and Felsenstein, D. (2007), "Terror, fear and behaviour in the Jerusalem housing market", *Urban Studies*, Vol. 44 No. 13, pp. 2529-2546.

Helbich, M., Brunauer, W., Hagenauer, J. and Leitner, M. (2013), "Data-driven regionalization of housing markets", *Annals of the Association of American Geographers*, Vol. 103 No. 4, pp. 871-889.

Holmes, M.J., Otero, J. and Panagiotidis, T. (2019), "Property heterogeneity and convergence club formation among local house prices", *Journal of Housing Economics*, Vol. 43, pp. 1-13.

Islam, K.S. and Asami, Y. (2009), "Housing market segmentation: a review", *Review of Urban and Regional Development Studies*, Vol. 21 Nos 2/3, pp. 93-109.

Kauko, T. and Goetgeluk, R. (2005), "Spatial and multidimensional analysis of the Dutch housing market using the kohonen map and GIS", Louvain-la-Neuve: European Regional Science Association (ERSA), available at: www.econstor.eu/handle/10419/117425 (accessed 1 November 2020).

Kim, K. and Park, J. (2005), "Segmentation of the housing market and its determinants: Seoul and its neighbouring new towns in Korea", *Australian Geographer*, Vol. 36 No. 2, pp. 221-232.

Kokot, S. (2020), "Socio-economic factors as a criterion for the classification of housing markets in selected cities in Poland", *Real Estate Management and Valuation*, Vol. 28 No. 3, pp. 77-90.

Kowalczyk-Rólczyńska, P. (2014), "An application of cluster analysis on the polish housing market", in Loster, T. and Pavelka, T. (Eds), *Conference Proceedings, Presented at the the 8th International Days of Statistics and Economics*, Libuše Macáková, Melandrium, Prague, pp. 766-775.

Kozera, A. and Stanisławska, J. (2019), "Regional diversity of housing conditions of the population living in rural areas in Poland", *Roczniki Naukowe Stowarzyszenia Ekonomistów Rolnictwa i Agrobiznesu*, Vol. 21 No. 4, pp. 233-243.

Lee, C.L. (2009), "Housing price volatility and its determinants", *International Journal of Housing Markets and Analysis*, Vol. 2 No. 3, pp. 293-308.

Lin, W.S., Tou, J.C., Lin, S.Y. and Yeh, M.Y. (2014), "Effects of socioeconomic factors on regional housing prices in the USA", *International Journal of Housing Markets and Analysis*, Vol. 7 No. 1, pp. 30-41.

McCluskey, W.J. and Borst, R.A. (2011), "Detecting and validating residential housing submarkets: a geostatistical approach for use in mass appraisal", *International Journal of Housing Markets and Analysis*, Vol. 4 No. 3, pp. 290-318.

Maclennan, D. and Tu, Y. (1996), "Economic perspectives on the structure of local housing systems", *Housing Studies*, Vol. 11 No. 3, pp. 387-406.

Matysiak, G. and Olszewski, K. (2019), "A panel analysis of polish regional cities: residential price convergence in the primary market", *NBP Working Paper*, Vol. 316, pp. 1-39.

Montagnoli, A. and Nagayasu, J. (2015), "UK house price convergence clubs and spillovers", *Journal of Housing Economics*, Vol. 30, pp. 50-58.

Nistor, A. and Reianu, D. (2018), "Determinants of housing prices: evidence from Ontario cities, 2001-2011", *International Journal of Housing Markets and Analysis*, Vol. 11 No. 3, pp. 541-556.

Phillips, P.C.B. and Sul, D. (2007), "Transition modeling and econometric convergence tests", *Econometrica*, Vol. 75 No. 6, pp. 1771-1855.

Rivas, R., Patil, D., Hristidis, V., Barr, J.R. and Srinivasan, N. (2019), "The impact of colleges and hospitals to local real estate markets", *Journal of Big Data*, Vol. 6 No. 1, p. 7.

Royuela, V. and Duque, J.C. (2013), "HouSI: heuristic for delimitation of housing submarkets and price homogeneous areas", *Computers, Environment and Urban Systems*, Vol. 37, pp. 59-69.

Thaler, R. (1978), "A note on the value of crime control: evidence from the property market", *Journal of Urban Economics*, Vol. 5 No. 1, pp. 137-145.

Tomal, M. (2019a), "The impact of macro factors on apartment prices in polish counties: a two-stage quantile spatial regression approach", *Real Estate Management and Valuation*, Vol. 27 No. 4, pp. 1-14.

Tomal, M. (2019b), "House price convergence on the primary and secondary markets: evidence from polish provincial capitals", *Real Estate Management and Valuation*, Vol. 27 No. 4, pp. 62-73.

Tomal, M. (2020a), "Spillovers across house price convergence clubs: evidence from the Polish housing market", *Real Estate Management and Valuation*, Vol. 28 No. 2, pp. 13-20.

Tomal, M. (2020b), "Moving towards a smarter housing market: the example of Poland", *Sustainability*, Vol. 12 No. 2, p. 683.

Tomal, M. (2020c), "Analysing the coupling coordination degree of socio-economic-infrastructural development and its obstacles: the case study of polish rural municipalities", *Applied Economics Letters*, pp. 1-6.

Tomal, M. and Nalepka, A. (2020), "The impact of social participation on the efficiency of communal investment expenditure", *Economic Research-Ekonomska Istraživanja*, Vol. 33 No. 1, pp. 477-500.

Usman, H., Lizam, M. and Adekunle, M.U. (2020), "Property price modelling", *Real Estate Management and Valuation*, Vol. 28 No. 3, pp. 24-35.

Watkins, C.A. (2001), "The definition and identification of housing submarkets", *Environment and Planning A: Economy and Space*, Vol. 33 No. 12, pp. 2235-2253.

Wilhelmsson, M. (2004), "A method to derive housing sub-markets and reduce spatial dependency", *Property Management*, Vol. 22 No. 4, pp. 276-288.

Williams, R. (2016), "Understanding and interpreting generalized ordered logit models", *The Journal of Mathematical Sociology*, Vol. 40 No. 1, pp. 7-20.

Wu, C. and Sharma, R. (2012), "Housing submarket classification: the role of spatial contiguity", *Applied Geography*, Vol. 32 No. 2, pp. 746-756.

Corresponding author
Mateusz Tomal can be contacted at: tomalm@uek.krakow.pl

Appendix

Cluster	Counties
1	Augustowski; bartoszycki; bialski; białobrzeski; białogardzki; bielski; bieszczadzki; biłgorajski; braniewski; brodnicki; brzeziński; brzozowski; bytowski; chełmiński; chełmski; chojnicki; choszczeński; czarnkowsko-trzcianecki; częstochowski; człuchowski; dąbrowski; elbląski; garwoliński; golubsko-dobrzyński; gołdapski; gorlicki; gostyniński; górowski; grajewski; grudziądzki; gryfiński; hajnowski; hrubieszowski; janowski; jędrzejowski; kaliski; kamiennogórski; kazimierski; kętrzyński; kielecki; kłobucki; kolbuszowski; kolneński; kolski; konecki; koniński; kościerski; kozienicki; krasnostawski; krośnieński; kutnowski; legnicki; leski; leżajski; lidzbarski; limanowski; lipnowski; lipski; lubaczowski; lubartowski; lwówecki; łaski; łęczycki; łobeski; łomżyński; łosicki; łowicki; łukowski; makowski; miechowski; międzychodzki; międzyrzecki; milicki; mogileński; moniecki; myśliborski; nakielski; nidzicki; niżański; nowomiejski; nowosądecki; nowotarski; oleski; opatowski; opoczyński; opolski; ostrołęcki; ostrowski; pajęczański; parczewski; pińczowski; piotrkowski; piski; pleszewski; płocki; płoński; poddębicki; proszowicki; przasnyski; przemyski; przeworski; przysuski; pułtuski; pyrzycki; radomski; radomszczański; radziejowski; radzyński; rawski; ropczycko-sędziszowski; rycki; rypiński; sejneński; sępoleński; siedlecki; siemiatycki; sieradzki; sierpecki; skierniewicki; słupecki; słupski; sochaczewski; sokołowski; sokólski; starogardzki; staszowski; strzelecki; strzelecko-drezdenecki; sulęciński; suski; suwalski; szczycieński; sztumski; szydłowiecki; świdwiński; świecki; tarnowski; tomaszowski; toruński; tucholski; turecki; wałbrzyski; wąbrzeski; węgrowski; wieluński; wieruszowski; włocławski; włodawski; włoszczowski; wschowski; wysokomazowiecki; zambrowski; zamojski; ząbkowicki; zduńskowolski; złotoryjski; złotowski; zwoleński; żniński; żuromiński and żywiecki
2	Aleksandrowski; bełchatowski; będziński; białostocki; bielski; bieruńsko-lędziński; bocheński; bolesławiecki; brzeski; buski; bydgoski; chodzieski; chrzanowski; ciechanowski; cieszyński; dębicki; drawski; działdowski; dzierżoniowski; ełcki; giżycki; gliwicki; głogowski; głubczycki; gnieźnieński; goleniowski; gorzowski; gostyński; grodziski; grójecki; gryficki; iławski; inowrocławski; jarociński; jarosławski; jasielski; jaworski; jeleniogórski; kartuski; kędzierzyńsko-kozielski; kępiński; kluczborski; kłodzki; kościański; krakowski; krapkowicki; kraśnicki; krotoszyński; kwidzyński; leszczyński; lęborski; lubański; lubelski; lubiński; lubliniecki; łańcucki; łęczyński; łódzki wschodni; Biała Podlaska; Bielsko-Biała; Bytom; Chełm; Chorzów; Częstochowa; Dąbrowa Górnicza; Elbląg; Gliwice; Grudziądz; Jastrzębie-Zdrój; Jaworzno; Jelenia Góra; Kalisz; Konin; Krosno; Legnica; Leszno; Łomża; Mysłowice; Nowy Sącz; Ostrołęka; Piekary Ś ląskie; Piotrków Trybunalski; Płock; Przemyśl; Radom; Ruda Ś ląska; Rybnik; Siemianowice Ś ląskie; Skierniewice; Sosnowiec; Suwałki; Ś więtochłowice; Tarnobrzeg; Tarnów; Wałbrzych; Włocławek; Zabrze; Zamość; Żory; malborski; mielecki; mikołowski; miński; mławski; mrągowski; myszkowski; myślenicki; namysłowski; nowodworski; nowosolski; nowotomyski; nyski; obornicki; olecki; oleśnicki; olkuski; olsztyński; oławski; ostrowiecki; ostrowski; ostródzki; ostrzeszowski; oświęcimski; otwocki; pabianicki; pilski; polkowicki; prudnicki; pszczyński; puławski; raciborski; rawicki; rybnicki; rzeszowski; sandomierski; sanocki; skarżyski; sławieński; słubicki; stalowowolski; starachowicki; stargardzki; strzeliński; strzyżowski; szamotulski; szczecinecki; średzki; śremski; świdnicki; świebodziński; tarnobrzeski; tarnogórski; tczewski; tomaszowski; trzebnicki; wadowicki; wałecki; wągrowiecki; wejherowski; węgorzewski; wielicki; wodzisławski; wolsztyński; wołomiński; wołowski; wrzesiński; wyszkowski; zawierciański; zgierski; zgorzelecki; zielonogórski; żagański; żarski and żyrardowski
3	Gdański; grodziski; kamieński; kołobrzeski; koszaliński; legionowski; Warszawa; Białystok; Bydgoszcz; Gdańsk; Gdynia; Gorzów Wielkopolski; Katowice; Kielce; Koszalin; Kraków; Lublin; Łódź; Olsztyn; Opole; Poznań; Rzeszów; Siedlce; Słupsk; Sopot; Szczecin; Ś winoujście; Toruń; Tychy; Wrocław; Zielona Góra; piaseczyński; policki; poznański; pruszkowski; pucki; tatrzański; warszawski zachodni and wrocławski

Table A1.
A detailed list of counties assigned to given clusters in case of division into three groups

Cluster	Counties
1	Augustowski; bartoszycki; bialski; białobrzeski; białogardzki; bielski; bieszczadzki; biłgorajski; braniewski; brodnicki; brzeziński; bytowski; chełmiński; chełmski; czarnkowsko-trzcianecki; częstochowski; człuchowski; elbląski; golubsko-dobrzyński; gołdapski; gostyniński; górowski; grajewski; grudziądzki; gryfiński; hajnowski; hrubieszowski; janowski; jędrzejowski; kaliski; kazimierski; kielecki; kłobucki; kolneński; kolski; konecki; koniński; kościerski; krasnostawski; kutnowski; leski; lidzbarski; limanowski; lipnowski; lipski; lubaczowski; lubartowski; lwówecki; łaski; łęczycki; łobeski; łomżyński; łosicki; łowicki; łukowski; makowski; miechowski; międzyrzecki; mogileński; moniecki; nakielski; nidzicki; niżański; nowomiejski; nowosądecki; nowotarski; oleski; opatowski; opoczyński; opolski; ostrołęcki; ostrowski; pajęczański; parczewski; pińczowski; piotrkowski; piski; pleszewski; płocki; płoński; poddębicki; proszowicki; przasnyski; przemyski; przysuski; pułtuski; radomski; radomszczański; radziejowski; radzyński; rawski; rypiński; sejneński; sępoleński; siedlecki; siemiatycki; sieradzki; sierpecki; skierniewicki; słupecki; sochaczewski; sokołowski; sokólski; staszowski; strzelecko-drezdenecki; sulęciński; suski; suwalski; sztumski; szydłowiecki; świdwiński; świecki; tomaszowski; tucholski; turecki; wąbrzeski; węgrowski; wieluński; wieruszowski; włocławski; włodawski; włoszczowski; wysokomazowiecki; zambrowski; zamojski; ząbkowicki; zduńskowolski; złotowski; zwoleński; żniński; żuromiński and żywiecki
2	Aleksandrowski; bełchatowski; białostocki; bieruńsko-lędziński; bocheński; brzeski; brzozowski; buski; chodzieski; chojnicki; choszczeński; chrzanowski; ciechanowski; dąbrowski; dębicki; drawski; działdowski; dzierżoniowski; ełcki; garwoliński; gliwicki; głubczycki; gnieźnieński; goleniowski; gorlicki; gorzowski; gostyński; grodziski; gryficki; inowrocławski; jarociński; jarosławski; jasielski; jaworski; kamiennogórski; kędzierzyńsko-kozielski; kępiński; kętrzyński; kluczborski; kłodzki; kolbuszowski; kościański; kozienicki; krapkowicki; kraśnicki; krośnieński; krotoszyński; kwidzyński; legnicki; leszczyński; leżajski; lubański; lubelski; lubliniecki; łańcucki; łęczyński; łódzki wschodni; Bytom; Piekary Ś ląskie; Ruda Ś ląska; Rybnik; Ś więtochłowice; Zabrze; mielecki; międzychodzki; milicki; mławski; mrągowski; myszkowski; myślenicki; myśliborski; nowodworski; nowosolski; nowotomyski; nyski; olecki; oleśnicki; olkuski; opolski; ostrowiecki; ostrowski; ostródzki; ostrzeszowski; pilski; polkowicki; prudnicki; przeworski; pyrzycki; raciborski; rawicki; ropczycko-sędziszowski; rybnicki; rycki; rzeszowski; sandomierski; sanocki; skarżyski; sławieński; słubicki; słupski; starachowicki; starogardzki; strzelecki; strzeliński; strzyżowski; szamotulski; szczecinecki; szczycieński; świebodziński; tarnobrzeski; tarnowski; tczewski; tomaszowski; toruński; wadowicki; wałbrzyski; wałecki; wągrowiecki; węgorzewski; wodzisławski; wolsztyński; wschowski; wyszkowski; zawierciański; zgierski; zgorzelecki; zielonogórski; złotoryjski; żagański and żarski
3	Będziński; bielski; bolesławiecki; brzeski; bydgoski; cieszyński; giżycki; głogowski; grodziski; grójecki; iławski; jeleniogórski; kartuski; krakowski; lęborski; lubiński; Biała Podlaska; Bielsko-Biała; Bydgoszcz; Chełm; Chorzów; Częstochowa; Dąbrowa Górnicza; Elbląg; Gliwice; Grudziądz; Jastrzębie-Zdrój; Jaworzno; Jelenia Góra; Kalisz; Katowice; Kielce; Konin; Koszalin; Krosno; Legnica; Leszno; Łomża; Mysłowice; Nowy Sącz; Ostrołęka; Piotrków Trybunalski; Płock; Przemyśl; Radom; Siemianowice Ś ląskie; Skierniewice; Słupsk; Sosnowiec; Suwałki; Tarnobrzeg; Tarnów; Tychy; Wałbrzych; Włocławek; Zamość; Żory; malborski; mikołowski; miński; namysłowski; nowodworski; obornicki; olsztyński; oławski; oświęcimski; otwocki; pabianicki; policki; pszczyński; puławski; stalowowolski; stargardzki; średzki; śremski; świdnicki; tarnogórski; tatrzański; trzebnicki; wejherowski; wielicki; wołomiński; wołowski; wrzesiński and żyrardowski
4	Gdański; kamieński; kołobrzeski; koszaliński; legionowski; Warszawa; Białystok; Gdańsk; Gdynia; Gorzów Wielkopolski; Kraków; Lublin; Łódź; Olsztyn; Opole; Poznań; Rzeszów; Siedlce; Sopot; Szczecin; Ś winoujście; Toruń; Wrocław; Zielona Góra; piaseczyński; poznański; pruszkowski; pucki; warszawski zachodni and wrocławski

Table A2. A detailed list of counties assigned to given clusters in case of division into four groups