

ASSESSING THE SIMILARITY OF LOCAL REAL ESTATE MARKETS – THE USE OF TAXONOMIC METHODS (CZEKANOWSKI DIAGRAM)

Szczepańska Agnieszka^{1*}

¹ Department of Socio-Economic Geography, Institute of Spatial Management and Geography, Faculty of Ge-
oengineering, University of Warmia and Mazury in Olsztyn, Prawocheńskiego 15, 10-724 Olsztyn, Poland, e-mail:
aszczep@uwm.edu.pl, ORCID: 0000-0001-5184-0710

* Corresponding author

ARTICLE INFO	ABSTRACT
Keywords: similarity, similar markets, real estate market, numerical taxonomy, Czekanowski diagram	In the case of small or underdeveloped local real estate markets, the number of transactions is often insufficient to support the real estate valuation process. Therefore, it is necessary to expand the scope of market analysis. Since the level of socio-economic development influences the structure of the real estate market, it is assumed that this can serve as a basis for identifying similar real estate markets. In this article, similar real estate markets were grouped using a numerical taxonomy method, namely the Czekanowski diagram. The diagnostic features were quantitative in nature and represented categories such as housing and municipal services, population, housing construction, the real estate market, spatial planning, public finance, the labor market, and real estate prices. As a result of the analyses, 26 small towns located in northeastern Poland were classified into typological groups that are more homogeneous in terms of their diagnostic features and thus provide a comparable framework for the functioning of local real estate markets.
JEL Classification: C38, R31	
Citation:	
Szczepańska, A. (2026). Assessing the similarity of local real estate markets - The use of taxonomic methods (Czekanowski diagram). <i>Real Estate Management and Valuation</i> , 34(3), 104-116. https://doi.org/10.2478/remav-2026-0028	

1. Introduction

Polish regulations governing real estate valuation refer to socio-economic comparability as a factor influencing real estate prices, which is important for real estate market analysis and for identifying transaction prices for comparable properties in order to determine a property's market value.

The Regulation of the Minister of Development and Technology of 5 September 2023 on real estate valuation stipulates that the determination of a property's value must be preceded by an analysis of the real estate market -specifically with regard to transaction prices, rental rates, and transaction terms - as well as an on-site inspection of the property. When determining the value of real estate which, due to its specific characteristics and type, is not traded on the local real estate market, or in the absence of rental rates on that market, as well as when the number of transactions involving similar real estate or their rental rates is insufficient, transaction prices or rental rates

obtained for similar real estate on the regional or national real estate market may be used, where appropriate. Under the Act of 21 August 1997 on Real Estate Management, a comparable property is defined as a property that is comparable to the property being appraised in terms of location, legal status, intended use, manner of use, and other characteristics affecting its value.

In accordance with the above regulation, the relevant real estate market is defined - taking into account the subject matter, scope, purpose, valuation method, and data availability - by specifying its type, area, and the period under review. The delineation of the market area under analysis is based on the similarity of properties. It should be noted that McCluskey (1996) proposed a different approach, arguing that real estate markets should first be compared, after which the most similar properties should be identified within those markets.

The area of the real estate market under analysis should reflect the socio-economic similarities that

influence real estate prices. This area does not necessarily have to correspond to administrative boundaries. These socio-economic similarities can be used to identify comparable real estate markets.

A similar position is reflected in the case law of administrative courts, from which it can be concluded that, in certain situations where it is not possible to identify comparable properties on the local market, the scope of the search may be extended to the regional, national, or even international level (Kiedrzynek 2023). An example is the position adopted by the Provincial Administrative Court in Warsaw in its judgment of 2 December 2010 (IV SA/Wa 1221/10). That case concerned, in particular, the inclusion of real estate transactions from the local market or from areas located within a short distance of the property being appraised for comparison purposes. The court emphasized, *inter alia*, that the appraisal report should indicate the circumstances that justified the expansion of the relevant market based on the location of the property being appraised. Another example is the judgment of the Provincial Administrative Court in Łódź of 19 July 2011 (II SA/Łd 583/11). This case concerned the inclusion, for comparison purposes, of properties with different uses but located in the same area as the property being appraised. The court emphasized that a necessary condition for recognizing a property as comparable and accepting its value as reliable is the existence of a relationship based on similarity rather than identity. Accordingly, it is permissible to use transactions from different markets. In the judgment of the Supreme Administrative Court of 12 July 2022 (case no. I OSK 1267/19), it was noted that two key characteristics of comparability are the intended use and the manner of use of the real property. However, according to the court, no provision of law—in particular Article 4(16) of the Real Estate Management Act, which defines a comparable property - requires that a comparable property be a plot of land with the same area and located in the same locality as the property being appraised.

This issue is of particular significance, especially in the context of small or underdeveloped real estate markets and the need to expand the scope of market analysis. This applies to situations in which, due to their specific characteristics and type, properties are not traded on the local real estate market, or where the number of transactions involving similar properties is insufficient. In such cases, identifying a sufficiently large set of comparable properties is difficult, if not impossible (Domańska et al. 2020). A market that does

not allow for the collection of sufficient information is sometimes referred to as an underdeveloped market (Gnat 2021). Hence, the need to search for analogous markets. In Polish literature, terms such as neighboring markets (Cellmer & Kuryj, 2007), similar markets (Bełej & Cellmer, 2007), and parallel markets (Foryś & Batóg, 2017) are used.

Identifying similar markets requires determining the groups into which the real estate market as a whole can be divided. The selection of quantitative variables characterizing the markets under analysis involves identifying which comparative characteristics of real estate markets should be taken into account. In other words, this involves determining which attributes influence the assessment that the identified markets are similar. This is one of the key issues in the process of comparing real estate markets, although it is constrained by data availability. Markets can be considered similar if they are characterized by comparable locations and surroundings, common price-determining factors, similar price levels and dynamics, and comparable transaction conditions.

The aim of this manuscript is to demonstrate that taxonomic methods -used to describe and classify complex socio-economic phenomena expressed through a set of diagnostic variables -can be applied to identify similar real estate markets. The selected taxonomic method is the Czekanowski diagram. The analyses were conducted for small towns in northeastern Poland, where, due to their size, local markets are relatively limited.

1.1. Factors influencing the real estate market

A properly conducted real estate market analysis requires a broad range of information on the social, economic, and legal conditions that shape the real estate market environment and the conditions under which transactions are concluded. The literature on the subject identifies a number of factors that influence the real estate market. These are most often grouped into legal, social, and economic factors (Gołabeska, 2024; Manganelli 2014). Factors shaping real estate prices (and thus the real estate market) can be divided into exogenous (external) factors—such as economic, legal, technological, and social factors—and endogenous (internal) factors of both a static and a dynamic nature (Hermann, 2005). These factors interact with one another and with the real estate market in complex ways.

The group of legal factors comprises elements related to policy and law. These include legal

regulations concerning the broadly defined construction and real estate markets, tax regulations, housing policy, land-use planning policy, and administrative procedures. Actions taken by authorities at both the national and local levels may stimulate processes within the broadly defined real estate market or discourage investment, thereby influencing supply (Berry & McGreal, 2003). The state's housing policy, which determines the quantity and prices of the housing stock (Andrews et al., 2011; An et al., 2026), as well as land-use planning laws (Ratcliffe et al., 2021; Dudek & Nowel-Śmigaj, 2025), have a particularly significant impact on the real estate market. It is also worth noting that legal frameworks, real estate taxes, income taxes, liability protection, and rules governing foreign ownership vary significantly across countries (Gotham, 2006).

Technological and technical factors include modern solutions used in the construction industry, ranging from the design phase, through project implementation, to as-built documentation and the operation of facilities (Bac, 2014). They also encompass modern technological solutions related to real estate market services, such as artificial intelligence (AI), the Internet of Things (IoT), augmented reality (AR), virtual reality (VR), big data, cloud computing, decision automation, and machine learning. The implementation of modern technologies in the real estate market is an inevitable consequence of the fourth technological revolution (Bartkowiak et al., 2023). Many modern technologies, referred to as PropTech, are being implemented in the real estate market as part of broader technological innovations in the sector. Within this framework, the following areas of real estate research can be distinguished (after Bartkowiak et al., 2023): (1) construction technology (Baum, 2017; Annual Report..., 2017); (2) the direct implementation of the investment process and operational activities, including development activities (Maududy & Gamal, 2019b); (3) trading and management, including management, sales, and marketing (Baum, 2017; Maududy & Gamal, 2019a); and (4) indirect or financial real estate investment (Shaw, 2020).

Another group of factors consists of economic conditions. These can be analyzed at both the national and local levels and may also relate to the financial environment. It is generally accepted that the overall state of the economy and the level of economic activity have a significant impact on the generation of demand and, consequently, on prices. Key indicators in this regard include GDP and the unemployment rate (Leung

2004; Bouchouicha & Ftiti, 2012; Grum & Govekar, 2016; Kalabiska & Hlavacek 2022; Okunevičiūtė Neverauskienė et al., 2025). Strong economic growth translates into higher employment levels, which in turn lead to increased consumer spending, including in the real estate market. Additionally, GDP growth has a positive and lasting impact on mortgage lending (Filotto et al., 2018; Dogan & Topuz, 2020). A low unemployment rate influences household income and, consequently, demand. At the same time, in the long term, an increase (decrease) in real estate prices leads to a decrease (increase) in unemployment rates (Dogan & Topuz, 2020), which is associated with employment in the construction sector (Geerolf & Grjebine, 2014).

Another significant factor is interest rates, which determine the cost of mortgages and have a substantial impact on real estate prices (Bjørnland & Jacobsen, 2010; Arslan, 2014). High interest rates limit borrowing capacity and dampen demand, thereby exerting downward pressure on real estate prices. Conversely, lower interest rates stimulate demand by reducing the cost of mortgages, potentially driving up real estate prices due to increased competition among buyers. Inflation also has a significant effect on the real estate market, as real estate often serves as a hedge against inflation (Anari & Kolari, 2002). However, high inflation may adversely affect the construction sector and is closely linked to interest rates (Qabaja & Tenekeci, 2024). Demographic and demand-side factors, including population growth, net migration, divorce rates, unemployment levels, and wage growth, constitute key determinants of prices (Hlaváček & Komárek, 2011). According to other studies, on the supply side, factors influencing prices include construction activity and the housing stock, and to a lesser extent, the number of transactions in the real estate market. On the demand side, the size of the real estate market, as measured by the number of households, the age structure of the population, and regional infrastructure, has been identified as a primary determinant of real estate prices (Belke & Keil, 2018).

The final group comprises social factors such as population size and age structure, educational attainment, and migration. The demographic situation has a significant impact on the structure and development of the real estate market (Tymiński et al., 2024; Breidenbach et al., 2024; Cheung, 2024). Accordingly, demographic trends play a key role in shaping real estate needs. Rising property values affect birth rates, in particular leading to a decline in births among non-homeowners (Dettling & Kearney, 2014). At the same time, high levels of education, good health,

and high incomes increase household demand for residential real estate, even as households age and the population declines (Eichholtz & Lindenthal, 2014; Green & Lee, 2016). Demand for housing may also increase in areas experiencing significant population growth.

The selected factors discussed above shape the dynamics of supply and demand, which constitute a key indicator of the state of the real estate market. The balance between supply and demand is of fundamental importance for price stability. Consequently, identifying comparable markets requires a comprehensive analysis of these interrelated factors that shape the real estate market and influence investment opportunities.

1.2. Taxonomy in object grouping

Taxonomy based on the use of quantitative methods is referred to as numerical taxonomy, a scientific discipline concerned with the principles and procedures of ordering and classification based on quantitative data. Methods for grouping data in the study of spatial variation have a long history and are applied across many scientific fields. This is due to the fact that most spatial phenomena are not homogeneous but exhibit a certain spatial structure. As the study of spatial structure has developed independently in various branches of science and for different purposes, a wide variety of methods has emerged. However, in order to visualize the spatial structure of phenomena, it is first necessary to classify them; this is where taxonomy plays a fundamental role.

In general terms, taxonomy may be defined as the science of classification principles, dealing with procedures for organizing and grouping objects described within a multidimensional observational space (Kowerski, 2017). The purpose of classification is to identify groups of units that are similar to one another with respect to one or more features simultaneously (Gaczek et al., 1980), that is, to assign a given object to a specific class characterized by defined attributes. The task of classification methods is to divide a set of elements into statistically more homogeneous subsets that satisfy the conditions of disjointness, completeness, and non-emptiness, according to a specified criterion based on the features of the classified objects (Grabiński et al., 1989; Kaufman & Rousseeuw, 1990). The outcome of classification is the identification of typological groups that are more homogeneous in terms of the diagnostic features describing them.

By applying numerical taxonomy methods, local real estate markets can be classified through the identification of typological groups of similar markets. Quantitative diagnostic features should reflect the factors shaping the real estate market.

2. Material and methods

2.1. Czekanowski diagram

To identify similar real estate markets, the Czekanowski diagram was used, which is a method for the graphical representation of a similarity (distance) matrix between objects. The Czekanowski diagram is one of the oldest taxonomic methods, developed by the Polish anthropologist Jan Czekanowski for the analysis of the structure of series of human remains (Sołtysiak & Jaskólski, 1999; Jaskólski & Sołtysiak, 2004). Czekanowski was the author of pioneering taxonomic methods, including the concept of the square distance matrix, known for many years as the Czekanowski table (Bielicki et al., 1987).

In the analyses conducted, the basis for classifying and identifying clusters of objects was the distance matrix D , which represents the distances between pairs of objects assigned to the i -th column and the j -th row in the normalized data matrix X' (Murtagh & Contreras, 2011). According to the adopted definition of distance, object A is closer to object B than to object C (i.e., more similar to object B than to object C) if the distance between objects A and B is smaller than the distance between objects A and C (Kolenda, 2006).

The Euclidean distance was used as a measure of the distance between elements (Sołtysiak & Jaskólski, 1999; Prus & Król, 2017) and was calculated based on the known attributes (diagnostic features) of all the analyzed objects (towns). The calculations were performed using the following formula:

$$d_{ik} = \sqrt{\sum_{j=1}^m |x'_{ij} - x'_{kj}|^2} \quad (1)$$

where:

$d_{i,k}$ – the metric distance between the i -th and k -th objects,

x_{ij}, x_{kj} – the values in the i -th and k -th rows of the normalized data matrix X' (i.e., the values assigned to the i -th and k -th objects).

A Czekanowski diagram is a graphical representation of the distance matrix. The results are interpreted based on the similarity matrix, according to the principle that the smaller the values, the greater the similarity between objects. Individual rows and columns correspond to the spatial units being compared. The

size of the graphical symbol in the diagram indicates the degree of similarity between the spatial units (i.e., at the intersection of the row and column corresponding to the towns being compared). The most similar units are located closest to the main diagonal, while the degree of similarity decreases as the distance from the diagonal increases.

The Czekanowski diagram was generated using MaCzek 2.0 software, version 3.3.44 (freeware), developed by A. Sołtysiak and P. Jaskulski. This data organization program employs, among other methods, the Simply Auto algorithm and the Simple Genetic Algorithm. The primary objective of these algorithms is to rearrange the rows and columns of the distance matrix so that the most similar objects are positioned as close to one another as possible, i.e., along the main diagonal of the matrix. In practice, this involves minimizing the dispersion of similar objects within the matrix.

The application of these algorithms results in a well-structured Czekanowski diagram that reveals compact clusters of similar objects, referred to as subgroups (local clusters). Compared to the genetic algorithm, the Simply Auto algorithm operates more quickly and produces more localized and detailed groupings; however, it is less effective in terms of global optimization of the layout. In contrast, the Simple Genetic Algorithm is more resistant to local minima and ensures a higher quality of global organization. As a

result, more compact and homogeneous clusters are formed, leading to a reduction in the number of identified groups (i.e., fewer clusters than in the case of the simply auto algorithm).

2.1. Variable selection

The application of numerical taxonomy methods requires the establishment of a set of diagnostic features. The sample set of features used in the analysis comprises three groups of factors: legal, economic, and social. The selection of features was determined by the availability of statistical data provided by the Central Statistical Office. The study utilized a set of quantitative variables obtained from the Local Data Bank of the Central Statistical Office (2024) and the adresowo.pl portal, which provides median apartment prices (2024), for two provinces - Podlaskie and Warmińsko-Mazurskie. These provinces are located in northeastern Poland, are adjacent to one another, and are characterized by a similar level of socio-economic development and environmental conditions. The basic spatial units were urban municipalities and towns within urban-rural municipalities, in accordance with the Nomenclature of Territorial Units for Statistical Purposes (NTS 5(1) and NTS 5(4)). This group includes towns with similar population sizes (from 10,000 to 35,000 inhabitants). The analysis covered 26 towns, the spatial distribution of which is presented in Figure 1.



Fig. 1. The geographical location of selected towns in the Podlaskie and Warmińsko-Mazurskie voivodship. *Source: own study.*

The set of diagnostic characteristics primarily includes factors reflecting the state of the real estate market, in particular those related to the stock of municipal residential properties, land management for residential construction, the housing construction sector, and the market sale of residential units. Demographic characteristics were adopted as social factors in the analysis, as they have a significant impact on the structure and development of the real estate market. These include, in particular, total population and its age structure, natural population change,

international migration, and the number of households. The level of satisfaction of housing needs was also taken into account in quantitative terms. The group of economic factors was represented by the unemployment rate, wage levels, and GDP per capita. Data relating to spatial policy, as reflected in local zoning plans and administrative decisions concerning building permits, were adopted as legal factors. The dataset included the diagnostic features presented in Table 1 for the Warmińsko-Mazurskie Voivodeship and the Podlaskie Voivodeship.

Table 1

Initial set of diagnostic features		
No	Field	Variable name
X ₁	Housing and public utilities	Number and amounts of housing allowances paid to tenants per 1,000 people [units]
X ₂		Land held by municipalities for residential development [ha]
X ₃		Land transferred/sold to developers for residential construction [ha]
X ₄		Total housing stock [units]
X ₅		Average living area per apartment [m ²]
X ₆	Population	Average number of persons per household [persons]
X ₇		Number of apartments per 1,000 residents [units]
X ₈		Total population [persons]
X ₉		Working-age population [persons]
X ₁₀		Population density [persons/km ²]
X ₁₁	Residential construction	Net internal migration [persons]
X ₁₂		Housing units completed per 1,000 people [units]
X ₁₃		Building permits and construction notifications – residential buildings / per 1,000 inhabitants [units]
X ₁₄		New residential buildings per 1,000 inhabitants [units]
X ₁₅	Real estate market	Median price per square meter of an apartment [PLN/m ²]
X ₁₆	Land use planning	Share of land area by type of development in the total area covered by current local zoning plans - total residential development [ha]
X ₁₇		Zoning decisions regarding single-family residential development [number]
X ₁₈		Zoning decisions regarding multi-family residential development [number]
X ₁₉	Public finance	Per capita income [PLN]
X ₂₀	Labor market	Registered unemployed persons [persons]
X ₂₁		Employment in the industrial sector per 1,000 people [persons]
X ₂₂		Economic entities per 1,000 inhabitants [units]

Source: own study.

From the set of proposed variables, those that met the most important formal criteria from the perspective of the analysis were selected. The key formal and statistical criteria included relatively high spatial and spatiotemporal variability, as well as the absence of excessive correlation among the variables. As a result of the analyses conducted, the following variables were excluded from further consideration: land held by

municipalities for residential construction, housing units per 1,000 inhabitants, and the working-age population, due to their high correlation with the remaining variables. The final observation matrix had dimensions of 26 (number of towns) by 19 (number of variables characterizing the towns). The set of diagnostic features included in the analysis is presented in Table 2.

Table 2

Set of diagnostic features included in the analysis		
No	Field	Variable name
X ₁	Housing and public utilities	Number and amounts of housing allowances paid to tenants per 1,000 people [units]
X ₂		Land transferred/sold to developers for residential construction [ha]
X ₃		Total housing stock [units]

X ₄		Average living area per apartment [m ²]
X ₅		Average number of persons per household [persons]
X ₆		Number of apartments per 1,000 residents [units]
X ₇	Population	Total population [persons]
X ₈		Population density [persons/km ²]
X ₉		Net internal migration [persons]
X ₁₀	Residential construction	Building permits and construction notifications – residential buildings / per 1,000 inhabitants [units]
X ₁₁		New residential buildings per 1,000 inhabitants [units]
X ₁₂	Real estate market	Median price per square meter of an apartment [PLN/m ²]
X ₁₃	Land use planning	Share of land area by type of development in the total area covered by current local zoning plans-total residential development [ha]
X ₁₄		Zoning decisions regarding single-family residential development [number]
X ₁₅		Zoning decisions regarding multi-family residential development [number]
X ₁₆	Public finance	Per capita income [PLN]
X ₁₇	Labor market	Registered unemployed persons [persons]
X ₁₈		Employment in the industrial sector per 1,000 people [persons]
X ₁₉		Economic entities per 1,000 inhabitants [units]

Source: own study.

According to Czekanowski's method, variables describing objects must be of the same type; otherwise, they must be standardized. Standardization was performed using the standard deviation according to the formula:

$$z = \frac{x - \mu}{\delta} \quad (2)$$

where:

x – unstandardized variable,

μ – population average,

δ – population standard deviation

The input dataset prepared in this manner served as the basis for grouping similar objects in accordance with the principles of Czekanowski's method.

3. Results

Using MaCzek 2.0 software (version 3.3.44) and based on standardized diagnostic feature values, an unordered and subsequently an ordered similarity matrix were developed. The Czekanowski diagram is a graphical representation of the Euclidean distance matrix between each pair of units (the towns under study), based on the features describing them. The intensity of similarity is illustrated by the size of the dots – the larger the dot, the smaller the distance (i.e., the greater the similarity) between two units, whereas the absence of a dot indicates very low or no similarity. The names of the analyzed towns are listed in the "Name" column.

3.1. Czekanowski unordered diagram

Figure 2 presents the unordered Czekanowski diagram for the towns analyzed. The results are arranged alphabetically, which results in a lack of clear structure

within the diagram, as well as disorder and difficulty in interpreting the results, particularly in identifying clusters. In this form, the diagram has limited interpretative value and constitutes only a preliminary stage of analysis. It allows solely for a direct reading of the degree of similarity between two analyzed local real estate markets (towns), represented by the graphical symbol at the intersection of a given row and column.

The unordered Czekanowski diagram provides comprehensive information on the similarities between objects but does not clearly reveal the underlying data structure. It therefore serves as a starting point for further analysis, as it requires transformation (ordering) to enable the identification of homogeneous groups.

3.2. Czekanowski ordered diagram

In the next step, the diagram was organized using two algorithms. This process involves rearranging the rows and columns so that the elements most similar to one another are positioned along the main diagonal. This enables the identification of groups of similar objects. The identification of such groups is based on a visual inspection of the distance matrix.

Figure 3 presents a Czekanowski ordered diagram generated using the Simply Auto algorithm. The most prominent features are clusters of high-intensity points (large, dark dots) concentrated along the diagonal, which indicate groups of towns exhibiting a high degree of similarity. Within the diagram, several clusters of similar towns can be distinguished, and thus clusters of similar local real estate markets can be identified. Figure 4 presents the geographic location of the identified relevant real estate markets.

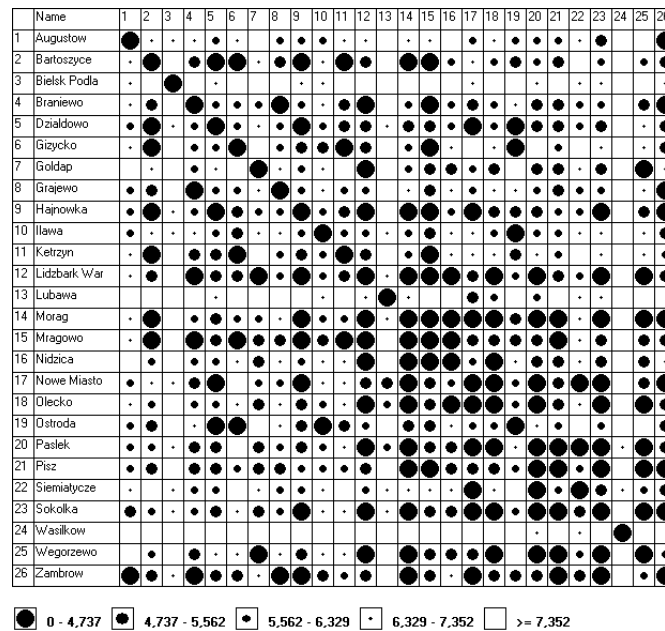


Fig. 2. Czekanowski unordered diagram. Source: own study.

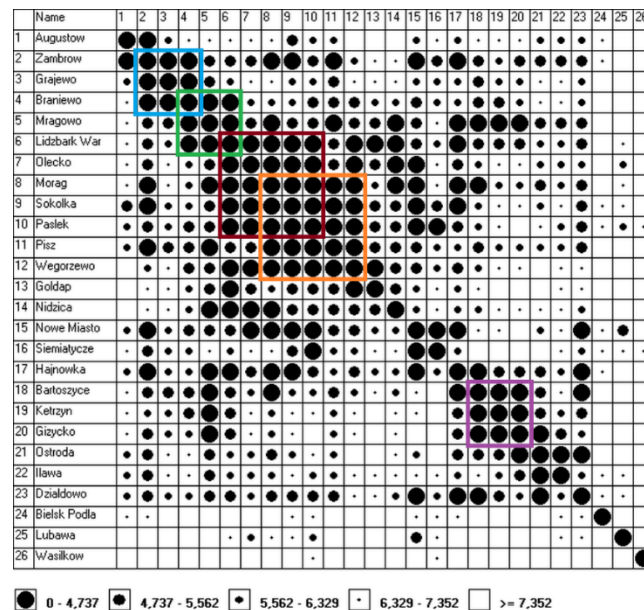


Fig. 3. Czekanowski's ordered diagram – ordering using the simply auto algorithm. Source: own study.

The two largest clusters of similar real estate markets consist of groups of five towns. They are characterized by a very high degree of internal similarity, which indicates the existence of similar local real estate markets shaped by comparable socio-economic factors.

The first group comprises Lidzbark Warmiński, Olecko, Morąg, Sokółka, and Pasłęk, towns (marked with a brown border) while the second group includes Morąg, Sokółka, Pasłęk, Pisz, Węgorzewo, and Gołdap

towns (marked with a orange border). The towns of Morąg, Sokółka, and Pasłęk are common to both groups. They exhibit connections to two different clusters, which may indicate their intermediate nature - they share characteristics with multiple groups but do not clearly belong to any single one. The towns in the first group are characterized by similar population levels, housing stock, average living space per dwelling, per capita income, and employment in the industrial sector. The observed similarities among the real estate

markets are therefore primarily influenced by demographic factors, housing conditions, and public finances. Towns assigned to the second group are characterized by similar living space per dwelling, the number of economic entities per 1,000 residents, and the median price per square meter of residential property.

The remaining clusters consist of three elements (marked with an blue, green, and purple border). These clusters are also characterized by a high degree of

similarity, although they are smaller and more compact. Three such groups of towns can be identified: the first comprises Zambrów, Grajewo, and Braniewo; the second includes Braniewo, Mrągowo, and Lidzbark Warmiński; and the third consists of Bartoszyce, Kętrzyn, and Giżycko. The first group exhibits the greatest similarity in terms of per capita income, the second in terms of population density, and the third in terms of net migration and per capita income.

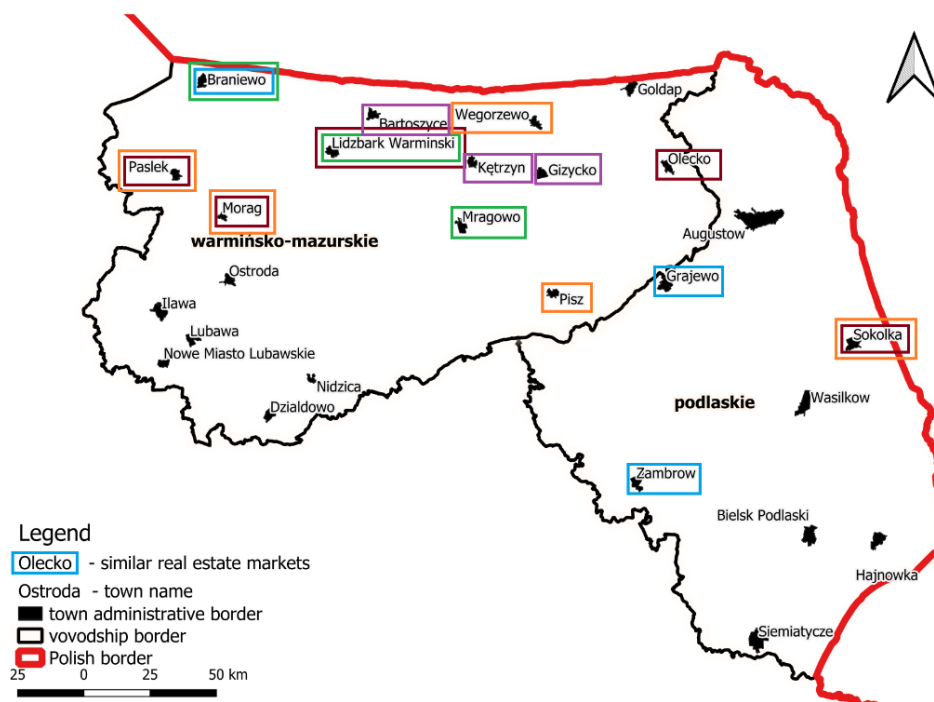


Fig. 4. The geographic location of the identified similar real estate markets – ordering using the simply auto algorithm.

Source: own study.

Figure 5 presents a Czekanowski diagram generated using a simple genetic algorithm. The largest clusters of points (large, dark dots) are concentrated in the lower-right part of the matrix, indicating the presence of a highly homogeneous group of towns. Compared to the previous ordering variant, the cluster structure is more compact and consists of a smaller number of distinct clusters; consequently, significantly fewer groups of similar elements have been identified. Figure 6 presents the geographic location of the identified relevant real estate markets.

Within the diagram, a four-element cluster (marked with a brown border) and a three-element cluster (marked with an orange border) can be distinguished. The group of four similar towns comprises Sokółka, Nowe Miasto Lubawskie, Olecko, and Paszék, which are characterized by very similar price levels. In turn, the towns of Olecko, Paszék, and Węgorzewo form a group

with very similar price levels and average living space per apartment. Most towns form a broad, continuous area of moderate similarity (numerous large and medium-sized points), suggesting a lack of strong segmentation into multiple distinct groups. This indicates that the local markets under study are largely similar, differing more in degree than in the nature of their characteristics.

It is also worth noting the towns of Bielsk Podlaski, Lubawa, and Wasilków, which exhibit a negligible degree of similarity to other localities. These towns can be regarded as outliers that differ significantly from the remaining units. Both the Simply Auto algorithm and the Simple Genetic Algorithm yield consistent results in this respect. These towns clearly stand out from the group under study in terms of the analyzed features, exhibiting little or no similarity (i.e., a negligible number of connections -very few or no points in the

corresponding rows and columns). This indicates their structural distinctiveness, both in relation to the main cluster and among themselves. Their specific characteristics may also contribute to the distinct nature of the local real estate markets they represent. Such outliers may reflect particular models of local real

estate market development, with socio-economic structures and market characteristics that differ significantly from those of the other analyzed towns. Figure 6 presents the geographic location of the identified relevant real estate markets.

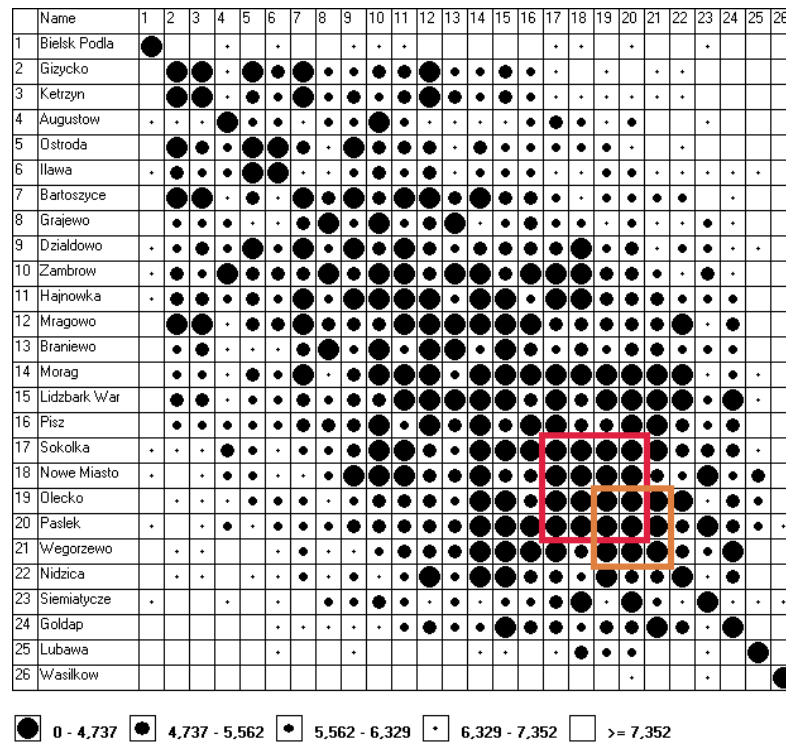


Fig. 5. Czekanowski ordered diagram – ordering using the simply genetic algorithm. Source: own study.

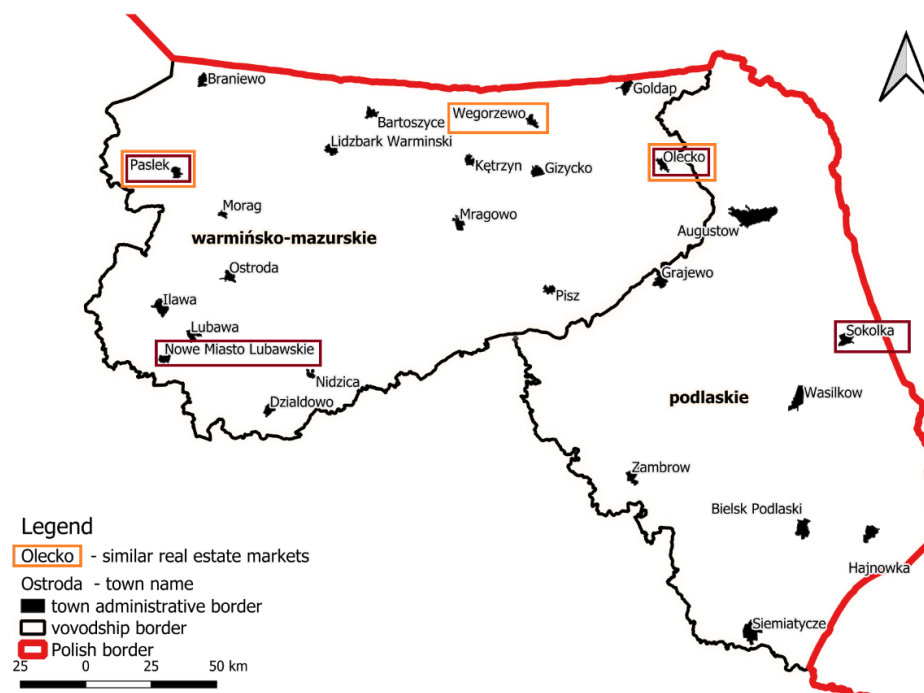


Fig. 6. The geographic location of the identified similar real estate markets – ordering using the simply auto algorithm. Source: own study.

4. Discussion

Grouping local markets makes it possible to expand the scope of the analysis to include markets that are as similar as possible. This is particularly important in limited markets with very few transactions, where it is not possible to conduct a proper valuation process in accordance with established standards and methodologies. It also hinders a reliable analysis of real estate price trends and the factors influencing them. In such cases, expanding the scope of the real estate market appears to be the only viable solution. The issues related to the identification of similar real estate markets are particularly relevant in the context of small and very small towns. Accordingly, the starting point for the analyses conducted was a set of small towns, defined by population size and characterized by similar locations (border regions in northeastern Poland), which results in similarities in terms of the level of socio-economic development.

The analysis, conducted using an ordered Czekanowski diagram based on a Euclidean distance matrix for standardized diagnostic features, enabled the identification of similarity structures among the towns under study and, consequently, the identification of groups of local real estate markets with similar characteristics. The initial, unordered diagram exhibited limited readability due to the alphabetical arrangement of the data, which made it difficult to identify clusters. Only the application of ordering procedures made it possible to clearly distinguish clusters of towns similar in terms of the analyzed socio-economic characteristics. Thus, the ordered Czekanowski diagram enabled the identification and classification of similar local real estate markets, which may serve as a basis for further comparative and decision-making analyses.

Although a comparison of the two clustering methods (i.e., the simply auto algorithm and the simple genetic algorithm) reveals differences in the structure of similarities among the analyzed units, the overall consistency of the results is preserved. The first method is more analytical in nature, offering greater detail, while the second is more aggregative, reducing the number of distinct clusters and providing a higher level of generalization. Both methods lead to convergent conclusions regarding the existence of similarities between local real estate markets and the identification of outliers that do not exhibit significant similarity either to other units or to one another. This reflects their specific socio-economic conditions and the distinct nature of the local markets they represent. The

differences between the classification methods primarily concern the level of aggregation. The first approach is more suitable for detailed analyses (e.g., market segmentation), whereas the second is better suited for strategic and synthetic overviews. Consequently, the two approaches may be regarded as complementary rather than mutually exclusive. In summary, regardless of the classification method applied, the results obtained are consistent and indicate clear differentiation among the local markets studied, while simultaneously revealing groups of towns characterized by a high degree of structural similarity.

5. Conclusions

The numerical taxonomy method applied, which is based on clustering units, has a wide range of applications, including the identification of similar real estate markets. The conducted research demonstrates that it is possible to identify similarities between local real estate markets (towns) and classify them based on a set of universal characteristics that shape these markets. The Czekanowski diagram enables the identification of groups of markets that are similar to one another based on multiple features simultaneously. The adopted set of diagnostic features is illustrative and may be modified depending on the specific requirements (type of market analyzed, size of the spatial unit, and data availability).

The results confirm that the use of taxonomic methods - in this case, the Czekanowski diagram is appropriate for achieving the intended objective. The identified groups of similar local real estate markets facilitate a more effective valuation process, particularly in the case of highly constrained markets. Although the study was conducted in small Polish towns, the methodology is universal and allows for a flexible selection of diagnostic features tailored to specific needs. Identifying even very small groups of similar localities - and thus, as intended, real estate markets (comprising two or three elements, or even a single pair of similar markets) - enables the expansion of the search area for comparable real estate transactions, which constitute the basis for determining market value. The larger the group of similar local real estate markets, the greater the likelihood of identifying an appropriate set of comparable properties (i.e., properties that have been the subject of transactions) for the property being appraised. This, in turn, allows for a more reliable analysis of the real estate market.

In light of the above, numerical taxonomy can be effectively applied to identify similar real estate markets based on specified diagnostic features representing selected factors that characterize the real estate market and influence property prices. This is consistent with the view expressed by Kahr and Thomsett (2005), who emphasize that economic, demographic, and local factors affecting real estate values should be analyzed jointly.

It should also be noted that analyzing similarities between real estate markets entails the risk of capturing random phenomena and incidental circumstances. Moreover, the characteristics that define similar markets are not necessarily stable; they may change over time or be influenced by random variation.

References

- An, S., Bae, S., Lee, G., & Ahn, K. (2026). Quantitative structure and sentiment of news articles on housing policies. *Chaos (Woodbury, N.Y.)*, 36(3), 033110. Advance online publication. <https://doi.org/10.1063/5.0310060> PMID:41773949
- Anari, A., & Kolari, J. (2002). House prices and inflation. *Real Estate Economics*, 30(1), 67–84. <https://doi.org/10.1111/1540-6229.00030>
- Andrews, D., Sánchez, A. C., & Johansson, Å. (2011). *Housing markets and structural policies in OECD countries*. OECD Economics Department Working Papers No. 836 <https://dx.doi.org/https://doi.org/10.1787/18151973>
- Annual Report Jones Lang LaSalle Incorporated. (2017). *Thinking Beyond*. Retrieved from JLL
- Arslan, Y. (2014). Interest rate fluctuations and equilibrium in the housing market. *The B.E. Journal of Macroeconomics*, 14(1), 173–204. <https://doi.org/10.1515/bejm-2013-0088>
- Bac, M. (2014). W poszukiwaniu innowacji na rynku nieruchomości [Looking for innovations at the real estate market]. *Zeszyty Naukowe Małopolskiej Wyższej Szkoły Ekonomicznej w Tarnowie*, 1(24), 11–21. <https://doi.org/10.25944/znmwse.2014.01.1121>
- Bartkowiak, P., Górka, A., Koszel, M., Mazurczak, A., Strączkowski, Ł., & Kinelski, G. (2023). *Nowe technologie na rynku nieruchomości – w poszukiwaniu zrównoważonego rozwoju [New technologies on the Real Estate Market—in search of sustainable development]*. Wydawnictwo Uniwersytetu Ekonomicznego w Poznaniu. <https://doi.org/10.18559/978-83-8211-156-9>
- Baum, A. (2017). *PropTech 3.0: The future of Real Estate*. Saïd Business School, University of Oxford Research.
- Belke, A., & Keil, J. (2018). Fundamental determinants of real estate prices: A panel study of German regions. *International Advances in Economic Research*, 24, 25–45. <https://doi.org/10.1007/s11294-018-9671-2>
- Belej, M., & Cellmer, R. (2007). Nieruchomości podobne i rynki sąsiednie - sprawozdanie z obrad Towarzystwa Naukowego Nieruchomości [Similar properties and neighboring markets – report from the proceedings of the Real Estate Scientific Society]. *Studia i materiały Towarzystwa Naukowego Nieruchomości*, 15(3–4), 23–34.
- Berry, J., & McGreal, S. (2003). European cities: the interaction of planning systems, property markets and real estate investment. In J. Berry & S. McGreal (Eds.), *European cities, planning systems and property markets* (pp. 17–31). E&FN SPON. <https://doi.org/10.4324/9780203362242>
- Bielicki, T., Krupiński, T., & Strzałko, J. (1987). Historia antropologii w Polsce [The history of physical anthropology in Poland]. *Przegląd Antropologiczny*, 53(1–2), 3–28.
- Bjørnland, H. C., & Jacobsen, D. H. (2010). The role of house prices in the monetary policy transmission mechanism in small open economies. *Journal of Financial Stability*, 6(4), 218–229. <https://doi.org/10.1016/j.jfs.2010.02.001>
- Bouchouicha, R., & Fiti, Z. (2012). Real estate markets and the macroeconomy: A dynamic coherence framework. *Economic Modelling*, 29(5), 1820–1829. <https://doi.org/10.1016/j.econmod.2012.05.034>
- Breidenbach, P., Jäger, P., & Taruttis, L. (2024). Aging and real estate prices in Germany. *Jahrbuch für Regionalwissenschaft = Review of Regional Research*, 44, 383–401. <https://doi.org/10.1007/s10037-024-00210-2>
- Cellmer, R., & Kuryj, J. (2007). Nieruchomości podobne i rynki sąsiednie-sprawozdanie z obrad Towarzystwa Naukowego Nieruchomości [Similar properties and neighboring markets – report from the proceedings of the Real Estate Scientific Society]. *Przegląd Geodezyjny*, 79, 18–20.
- Cheung, W. M. (2024). Cointegrations in house price dynamics and ageing population risks. *PLoS One*, 19(2), e0296991. Advance online publication. <https://doi.org/10.1371/journal.pone.0296991> PMID:38346072
- Dettling, L. J., & Kearney, M. S. (2014). House prices and birth rates: The impact of the real estate market on the decision to have a baby. *Journal of Public Economics*, 110, 82–100. <https://doi.org/10.1016/j.jpubeco.2013.09.009>
- Dogan, C., & Topuz, J. C. (2020). Real effects of real estate: Evidence from unemployment rates. *Studies in Economics and Finance*, 37(4), 605–623. <https://doi.org/10.1108/SEF-03-2019-0124>
- Domańska, A., Kucharska-Stasiak, E., Miklaszewska, A., Złoty, M., & Żelazowski, K. (2020). *Kierunki zmian instytucjonalnych na rynkach nieruchomości*. Wydawnictwo Uniwersytetu Łódzkiego. <https://doi.org/10.18778/8220-105-5>
- Dudek, M., & Nowel-Śmigaj, A. (2025). The Impact of ambiguous provisions in local spatial development plans on real estate valuation and investment: Case studies from Poland. *Land (Basel)*, 14(11), 2160. Advance online publication. <https://doi.org/10.3390/land14112160>
- Eichholtz, P., & Lindenthal, T. (2014). Demographics, human capital, and the demand for housing. *Journal of Housing Economics*, 26, 19–32. <https://doi.org/10.1016/j.jhe.2014.06.002>
- Filotto, U., Giannotti, C., Mattarocci, G., & Scimone, X. (2018). Residential mortgages, the real estate market, and economic growth: Evidence from Europe. *Journal of Property Investment & Finance*, 36(6), 552–577. <https://doi.org/10.1108/JPIF-09-2017-0060>
- Foryś, I., & Batóg, B. (2017). Comparison of the Structure of Apartments Sold in Two Cities in Northwest Poland. *Krakow Review of Economics and Management*, 9(957), 55–70. <https://doi.org/10.15678/ZNUEK.2016.0957.0904>
- Gaczek, W. M., Helpa, M., & Kasprzyk, A. (1980). Niehierarchiczna analiza skupień—Nowa metoda klasyfikacji zjawisk społeczno-gospodarczych [Non-hierarchical cluster analysis — a new classification method of socio-economic phenomena]. *Ruch Prawniczy, Ekonomiczny i Socjologiczny*, 2, 143–162.
- Geerolf, F., & Grjebine, T. (2014). Assessing house price effects on unemployment dynamics. *CEPII Working Paper N°2014-25*.
- Gnat, S. (2021). Property Mass Valuation on Small Markets. *Land (Basel)*, 10, 388. <https://doi.org/10.3390/land10040388>
- Gołabeska, E. (2024). *Rynek nieruchomości mieszkaniowych w Polsce w ujęciu prawnym, ekonomicznym, ekologicznym i finansowym [The residential real estate market in Poland from a legal,*

- economic, environmental, and financial perspective]. Oficyna Wydawnicza Politechniki Białostockiej.
- Gotham, K. F. (2006). The secondary circuit of capital reconsidered: Globalization and the US real estate sector. *AJS: American Journal of Sociology*, 112(1), 231–275. <https://doi.org/10.1086/502695>
- Grabiński, T., Wydymus, S., & Zeliaś, A. (1989). *Metody taksonomii numerycznej w modelowaniu zjawisk społeczno-gospodarczych [Methods of numerical taxonomy in the modeling of socio-economic phenomena]*. PWN.
- Green, R. K., & Lee, H. (2016). Age, demographics, and the demand for housing, revisited. *Regional Science and Urban Economics*, 61, 86–98. <https://doi.org/10.1016/j.regsciurbeco.2016.09.005>
- Grum, B., & Govekar, D. K. (2016). Influence of macroeconomic factors on prices of real estate in various cultural environments: Case of Slovenia, Greece, France, Poland and Norway. *Procedia Economics and Finance*, 39, 597–604. [https://doi.org/10.1016/S2212-5671\(16\)30304-5](https://doi.org/10.1016/S2212-5671(16)30304-5)
- Hermann, B. (2005). *Czynniki kształtujące wartość terenów mieszkaniowych w Poznaniu [Factors determining the value of residential land in Poznań]*. Wydawnictwo Nowak Nieruchomości.
- Hlaváček, M., & Komárek, L. (2011). Regional analysis of housing price bubbles and their determinants in the Czech Republic. *Czech Journal of Economics & Finance*, 61(1), 67–91.
- Jaskulski, P., & Sołtysiak, A. (2004). Diagram Czekanowskiego: pomysł, historia, zastosowania [Czekanowski's diagram: idea, history, implementations]. *Prace Naukowe Akademii Ekonomicznej we Wrocławiu. Taksonomia*, 11, 374–383.
- Kahr, J., & Thomsett, M. C. (2005). *Real estate market valuation and analysis*. John Wiley and Sons, Inc.
- Kalabiska, R., & Hlavacek, M. (2022). Regional Determinants of Housing Prices in the Czech Republic. *Czech Journal of Economics & Finance*, 72(1), 2–29. <https://doi.org/10.32065/CJEF.2022.01.01>
- Kaufman, L., & Rousseeuw, P. J. (1990). *Finding groups in data: An introduction to cluster analysis*. John Wiley and Sons, Inc. <https://doi.org/10.1002/9780470316801>
- Kiedrzynek, M. (2023). Cechy nieruchomości podobnej–najważniejsze uwagi na tle orzecznictwa [Features of a similar property – the most important comments in the context of jurisprudence]. *Folia Iuridica Universitatis Wratislaviensis*, 12(1), 174–184. <https://doi.org/10.34616/149322>
- Kolenda, M. (2006). Propozycja nowych mierników heterogeniczności do oceny klasyfikacji [Coefficients of classification scoring]. *Prace Naukowe Akademii Ekonomicznej we Wrocławiu*, 1100(358), 121–139.
- Kowerski, M. (2017). Zastosowanie metod taksonomicznych i ekonometrycznych do analizy poziomu rozwoju społeczno-gospodarczego regionów transgranicznych. [The application of taxonomic and econometric methods for the analysis of socio-economic development level of cross-border regions] In A. Miszczuk (Ed.), *Wyzwania rozwojowe pogranicza polsko-ukraińskiego* (pp. 13–33). Norbertinum Lublin.
- Leung, C. (2004). Macroeconomics and housing: A review of the literature. *Journal of Housing Economics*, 13(4), 249–267. <https://doi.org/10.1016/j.jhe.2004.09.002>
- Manganelli, B. (2014). *Real Estate Market. Analysis, Valuation Techniques, and Risk Management*. Springer International Publishing.
- Maududy, C. F., & Gamal, A. (2019a). Literature review: Technologies and property development. *IOP Conference Series. Earth and Environmental Science*, 396(1), 012020. Advance online publication. <https://doi.org/10.1088/1755-1315/396/1/012020>
- Maududy, C. F. i Gamal, A. (2019b). Literature review: The impact of property technology (PropTech) in property development. *Proceedings of the 33rd International Business Information Management Association Conference, IBIMA 2019: Education Excellence and Innovation Management through Vision 2020*, pp. 5370–5376.
- McCluskey, W. J. (1996). Zastosowanie sztucznej inteligencji w wycenie masowej dla potrzeb taksacji nieruchomości [Application of artificial intelligence in mass appraisal for the purposes of property taxation]. *Wycena*, 5(28), 3–7.
- Murtagh F, Legendre P (2011). *Ward's Hierarchical Clustering Method: Clustering Criterion and Agglomerative Algorithm*. Unpublished preprint.
- Okunevičiūtė Neverauskienė, L., Linkevičius, D., & Andriušaitienė, D. (2025). How macroeconomic factors impact residential real estate prices in Eastern Europe. *Business, Management and Economics Engineering*, 23(1), 30–43. <https://doi.org/10.3846/bmee.2025.22663>
- Prus, B., & Król, K. (2017). Evaluation of using selected taxonomic methods to classify socio-economic phenomena. *Acta Scientiarum Polonorum. Formatio Circumictus*, 16(2), 179–197. <https://doi.org/10.15576/ASP.FC/2017.16.2.179>
- Qabaja, M., & Tenekeci, G. (2024). Influence of inflation on the construction sector and economic growth in selected countries: A continental comparison. *Ain Shams Engineering Journal*, 15(11), 103013. <https://doi.org/10.1016/j.asej.2024.103013>
- Ratcliffe, J., Stubbs, M., & Keeping, M. (2021). *Urban planning and real estate development*. Routledge. <https://doi.org/10.4324/9780429398926>
- Rozporządzenie Ministra Rozwoju i Technologii z dnia 5 września 2023 r. w sprawie wyceny nieruchomości [Regulation of the Minister of Development and Technology of September 5, 2023, on real estate valuation] (Dz.U. 2023 poz. 1832)
- Shaw, J. (2020). Platform Real Estate: Theory and practice of new urban real estate markets. *Urban Geography*, 41(8), 1037. <https://doi.org/10.1080/02723638.2018.1524653>
- Sołtysiak, A., & Jaskulski, P. (1999). *Czekanowski's diagram: a method of multidimensional clustering (vol. 757)*. BAR International Series.
- Tymiński, A., Wróblewska, J., & Gołąbeska, E. (2024). Wprowadzenie do problematyki rynku nieruchomości. [Introduction to real estate market] In E. Gołąbeska (Ed.), *Rynek nieruchomości mieszkaniowych w polsce w ujęciu prawnym, ekonomicznym, ekologicznym i finansowym* (pp. 8–27). Oficyna Wydawnicza Politechniki Białostockiej.
- Ustawa z dnia 21 sierpnia 1997 r. o gospodarce nieruchomościami [Act of 21 August 1997 on Real Estate Management] (Tj. Dz.U. 2024 poz. 1145)