

**Iwona Foryś**

Uniwersytet Szczeciński  
e-mail: forys@wneiz.pl

---

## **OCENA PODOBIENSTWA WIELOWYMIAROWYCH OBIEKTÓW W WYCENIE NIERUCHOMOŚCI ZABUDOWANYCH DOMAMI JEDNORODZINNYMI**

---

## **ASSESMENT OF SIMILARITY OF MULTIDIMENSIONAL FACILITIES IN VALUATION OF PROPERTIES DEVELOPED WITH SINGLE-FAMILY HOUSES**

---

DOI: 10.15611/pn.2017.469.06

JEL Classification: C3, R3

**Streszczenie:** Proces szacowania wartości rynkowej nieruchomości wymaga dobierania do porównania nieruchomości podobnych, czyli różniących się jak najmniejszą liczbą cech. Mimo pewnych wskazówek zawartych w ustawie o gospodarce nieruchomościami nie rozstrzygnięto, jakie zmienne mają wyróżniać grupę nieruchomości podobnych. Przedmiotem badania jest ocena podobieństwa wielowymiarowych obiektów w wycenie nieruchomości zabudowanych domami jednorodzinnymi, opisanych zestawem zmiennych charakteryzujących ich cechy rynkowe. Wykorzystano analizę składowych głównych, która pozwoliła wskazać obiekty (nieruchomości) przynależne do jednego typu, będącego pod wpływem tych samych cech rynkowych nieruchomości podobnych. Badanie przeprowadzono na zbiorze nieruchomości sprzedanych w Szczecinie w 2015 r. Wyniki badania pozwolą rzeczoznawcom majątkowym na podejmowanie bardziej racjonalnych decyzji w procesie wyceny nieruchomości zabudowanych domami jednorodzinnymi.

**Słowa kluczowe:** nieruchomości podobne, analiza wielowymiarowa, wartość nieruchomości.

**Summary:** The process of estimation of the market value of a property requires matching similar properties, i.e. properties differing in the smallest number of features, in order to make the comparison. Despite some guidance contained in the Act on real estate management, it is not decided which variables are to distinguish similar real estates. The subject-matter of the research is the assessment of similarity of multidimensional facilities in the valuation of properties developed with single-family houses, described by the set of variables that are characterized by their market features. Principal components analysis was utilized which allowed indicating the facilities (properties) adherent to one type, under the influence of the same market prices – similar properties. The research was carried out using the set of real es-

tates sold in Szczecin in 2015. The research results will enable the valuers of the real estate market to take rational decisions concerning the right selection in the process of valuation of single-family houses.

**Keywords:** similar properties, multidimensional analysis, property value.

## 1. Wstęp

Metodyka wyceny nieruchomości w Polsce datuje się od 1990 roku ubiegłego wieku, jednak pojęcie podobieństwa nieruchomości wywołuje nadal liczne dyskusje [Cymerman, Hopfer 2012; Lis 2007; Foryś 2010]. Mimo iż podobieństwo nieruchomości w sposób ogólny definiuje ustawa o gospodarce nieruchomościami [Ustawa o gospodarce..., art. 4 pkt 16]: „Ilekcroć w ustawie jest mowa o nieruchomości podobnej – należy przez to rozumieć nieruchomość, która jest porównywalna z nieruchomością stanowiącą przedmiot wyceny, ze względu na położenie, stan prawny, przeznaczenie, sposób korzystania oraz inne cechy wpływające na jej wartość”, to nie jest ona jednoznaczna z uwagi na swobodę doboru innych cech wpływających na wartość nieruchomości. Mimo pewnych wskazówek zawartych w przepisach nie jest jednoznacznie rozstrzygnięte, jakie zmienne mają wyróżniać grupę nieruchomości podobnych, pozostawiając decyzję rzeczoznawcom majątkowym, a co w efekcie prowadzi do rozbieżności uzyskiwanych wyników szacowania. Zatem istnieje dowolność w doborze tych cech nieruchomości, która jest zdeterminowana zarówno przez rodzaj szacowanej nieruchomości, jak również cechy lokalnego rynku.

Celem przeprowadzonego badania jest ocena wpływu wybranych cech nieruchomości na ich cenę sprzedaży. Do realizacji postawionego celu zaproponowano analizę składowych głównych, której przydatność weryfikowano na szczecińskim rynku domów jednorodzinnych sprzedanych w Szczecinie w 2015 roku. Zastosowane narzędzie oraz uzyskane wyniki mogą być przydatne rzeczoznawcom majątkowym w procesie szacowania wartości rynkowej nieruchomości zabudowanych domami jednorodzinnyimi.

## 2. Analiza składowych głównych w wycenie nieruchomości

Analiza składowych głównych (*Principal Component Analysis* – PCA) służy poszukiwaniu przyczyn współzmienności i redukcji wymiaru obiektów wielowymiarowych bez utraty znaczących informacji [Hardle, Simar 2003]. Wykorzystana w badaniu analiza pozwala, poprzez transformację początkowego zbioru obserwowanych zmiennych we wzajemnie ortogonalne nieskorelowane nowe zmienne, zbudować nowy teoretyczny model opisujący strukturę zależności między badanymi cechami [Colley, Lohnes 1971]. Technikę zaproponował jako pierwszy Pearson w 1901 roku, a później w 1933

roku rozwinął w obszarze struktury zależności Hotelling [Hardle, Simar 2003; Jolliffe 2002]. Główne cechy przeprowadzanej analizy sprrowadzają się do:

- redukcji zbioru wielu zmiennych następuje przy zachowaniu dużej zmienności danych,
- każda kolejna składowa maksymalizuje zmienność, która nie została wyjaśniona przez poprzednią składową,
- analiza składowych przeprowadzana jest z wykorzystaniem macierzy korelacji z uwagi na różne rzędy wielkości zmiennych.

Natomiast kolejne etapy analizy składowych głównych można ująć w punktach:

- wybór badanych obiektów i zbioru zmiennych z uwagi na przesłanki merytoryczne i cel badania,
- odrzucenie obserwacji odstających z uwagi na analizowane cechy nieruchomości,
- analiza korelacji zmiennych opisujących badane obiekty,
- ocena istotności macierzy korelacji,
- analiza składowych głównych.

Każda nieruchomość jest wielowymiarowym obiektem opisanym zestawem cech mierzonych na różnych skalach pomiaru. Stanowią one pierwotny zbiór zmiennych, których kombinacja liniowa tworzy nowe nieobserwowalne zmienne o maksymalnej wariancji, tak aby zachować jak najwięcej zmienności danych. Każda kolejna składowa jest definiowana tak, aby maksymalizować zmienność, która nie została wyjaśniona przez poprzednie składowe. Interpretacja składowych głównych bazuje na korelacji zmiennych. Zmienne, które mają co do wartości bezwzględnej najwyższe wartości ładunków składowych, mają najwyższy wkład do zmiennych składowych. Ponieważ cechy nieruchomości mierzone są na różnych skalach oraz są różnego rzędu, stąd do obliczania składowych wykorzystuje się macierz korelacji. Kwadrat wyznaczonych ładunków składowych jest miarą udziału poszczególnych zmiennych w całkowitej zmienności danej składowej. Składowe mogą stanowić osie współrzędnych, co pozwala na interpretację graficzną obiektów wielowymiarowych w przestrzeni dwu- lub trójwymiarowej.

### 3. Charakterystyka danych empirycznych i wyniki badań

W badaniu wykorzystano dane dotyczące nieruchomości zabudowanych domami jednorodzinnymi, sprzedanych w 2015 roku w Szczecinie, opisanych w aktach notarialnych zestawem 8 zmiennych (badanie pełne):

- $X_0$  – lokalizacja (dzielnica);
- $X_1$  – rodzaj nabywanego prawa (wł, uw);
- $X_2$  – cena transakcyjna mieszkania (zł);
- $X_3$  – powierzchnia użytkowa domu ( $m^2$ );
- $X_4$  – powierzchnia działki ( $m^2$ );
- $X_5$  – wiek budynku w latach;

$X_6$  – stan zaawansowania budowy budynku (wykończony, niewykończony);

$X_7$  – rodzaj zabudowy (wolno stojący, bliźniaczy, szeregowy).

Zmienna  $X_1$ , oznaczona jako ( $wł$ ), dotyczy własności gruntu i naniesień na tym gruncie, natomiast ( $uw$ ) prawa użytkowania wieczystego gruntu i własności naniesień na tym gruncie (zmienna dychotomiczna przyjmująca wartości 1 –  $wł$  oraz 0 –  $uw$ ). Stan zaawansowania budowy został podzielony na dwie kategorie (niewykończony – oznacza najczęściej stan deweloperski lub do remontu, natomiast wykończony dotyczy przede wszystkim budynków już wcześniej zamieszkałych). Dla zmiennej  $X_6$  przyjęto dwie wartości 1 – wykończony, 0 – niewykończony. Ponieważ rynek różnicuje ceny domów wolno stojących i tych w zabudowie szeregowej lub bliźniaczej, stąd podział zmiennej  $X_7$  na trzy kategorie.

Ze względu na procedury szacowania wartości rynkowej nieruchomości i konieczność kwantyfikacji zmiennych jakościowych [Pagourtzi i in. 2003; Scarrett 2003] zaproponowano, aby zmienne jakościowe posłużyły jako zmienne: dodatkowa – stan zaawansowania budowy, dzielnica – przypadek aktywny, a jako zmienną grupującą przyjęto rodzaj zabudowy.

Badanie przebiegało w kilku etapach:

- wybór badanych obiektów i zbioru zmiennych z uwagi na przesłanki merytoryczne i cel badania;
- odrzucenie obserwacji odstających z uwagi na analizowane cechy nieruchomości [Jajuga 1993] oraz obserwacji z brakującymi danymi (tzn. takich, dla których brakuje pełnych informacji o danej zmiennej);
- analiza korelacji zmiennych opisujących badane obiekty, ocena istotności macierzy korelacji (test Bartletta, w którym testowana jest hipoteza zerowa, że macierz korelacji jest macierzą jednostkową, czyli że wszystkie współczynniki korelacji pomiędzy zmiennymi wejściowymi są równe zero [Barlett 1953]) oraz ocena adekwatności macierzy przy pomocy współczynnika Kaisera-Mayera-Olkin (KMO);
- analiza składowych głównych dla danych wg dzielnic miasta (zmienna grupująca – rodzaj zabudowy);
- na podstawie wartości własnych ( $>1$ ) lub wykresu osypiska określenie liczby składowych do interpretacji;
- sporządzenie wykresów zmiennych oraz przypadków, w których osiami układu są najbardziej istotne składowe;
- budowa modelu ceny domu z nowymi zmiennymi składowymi, którymi stały się główne składowe.

W pierwszym etapie badania dokonano analizy obiektów, odrzucając obserwacje odstające przede wszystkim z uwagi na powierzchnię działki oraz cenę transakcyjną, co wynika z metodyki wyceny nieruchomości [Cymerman, Hopfer 2012] oraz faktu, iż przypadki odstające mogą zafałszować zależności między zmiennymi [Jajuga 1993]. W efekcie uzyskano zbiór 191 obiektów, który podzielono na cztery dzielnice Szczecina: Śródmieście, Pogodno, Nad Odrą oraz Dąbie. Z uwagi na małą

liczebność obiektów w dzielnicy Śródmieście, spowodowanej przede wszystkim charakterem zabudowy (wielorodzinna), zrezygnowano w dalszej analizie z obiektów w tej lokalizacji.

Dla każdej dzielnicy wyznaczono macierze korelacji zmiennych  $X_1 - X_6$ , uzyskując istotnie statystycznie wartości współczynników korelacji Pearsona (jednak nieprzekraczające 0,6). Dodatkowo dla każdej dzielnicy zbudowano model regresji wielorakiej, w której zmienną zależną była cena nieruchomości (tab. 1).

**Tabela 1.** Podsumowanie regresji zmiennej zależnej *cena nieruchomości* dla badanych dzielnic

Zmienna	b	Bł. std.	t	p
Pogodno ( $R = 0,86332672$ , $F(6,28) = 13,658$ , $p < 0,00000$ , błąd std. estymacji: 1159E2)				
Wyraz wolny	141 125,6	195 285,9	0,722661	0,472562
Rodzaj nabywanego prawa do gruntu	27 991,9	140 254,2	0,199580	0,842452
Pow. użytkowa budynku	1 375,5	494,5	2,781742	0,007124
Powierzchnia działki	192,8	116,9	1,649361	0,104052
Stan zaawansowania	155 443,8	114 174,9	1,361453	0,178220
Wiek budynku	-1 816,8	1 934,2	-0,939308	0,351160
Nad Odrą ( $R = 0,866$ , $F(6,39) = 19,505$ , $p < 0,00000$ , błąd std. estymacji: 1305E2)				
Wyraz wolny	389 452,2	153 854,1	2,53131	0,015398
Rodzaj nabywanego prawa do gruntu	-38 594,3	131 598,5	-0,29327	0,770830
Pow. użytkowa budynku	853,5	427,0	1,99861	0,052478
Powierzchnia działki	441,7	71,9	6,14340	0,000000
Stan zaawansowania	23 483,5	57 602,1	0,40769	0,685679
Wiek budynku	-7 404,2	1 204,6	-6,14675	0,000000
Dąbie ( $R = 0,863$ , $F(6,28) = 13,658$ , $p < 0,00000$ , błąd std. estymacji: 1159E2)				
Wyraz wolny	-577 172	3 823 752	-0,15094	0,881102
Rodzaj nabywanego prawa do gruntu	1 159,9	335,0	3,46244	0,001682
Pow. użytkowa budynku	128,8	23,7	5,44406	0,000007
Powierzchnia działki	94 293,5	64 410,1	1,46395	0,153963
Stan zaawansowania	-5 952,2	1 104,9	-5,38692	0,000009
Wiek budynku	165 840,7	120 828,3	1,37253	0,180421

Źródło: opracowanie własne z wykorzystaniem pakietu Statistica 10.0.

Następnie dla każdej dzielnicy wyznaczono wartości własne, na podstawie których zdecydowano o liczbie składowych głównych (tab. 2).

**Tabela 2.** Wartości własne dla badanych dzielnic

Numer wartości	Wartość własna	% ogółu wariacji	Skumulowana wartość własna	Skumulowana %
Pogodno				
1	2,007172	33,45287	2,007172	33,4529
2	1,483756	24,72926	3,490928	58,1821
3	1,190616	19,84361	4,681545	78,0257
4	0,629452	10,49087	5,310997	88,5166
5	0,558696	9,31160	5,869693	97,8282
Nad Odrą				
1	2,247083	44,94166	2,247083	44,9417
2	1,101102	22,02205	3,348186	66,9637
3	1,007127	20,14253	4,355312	87,1062
4	0,503172	10,06343	4,858484	97,1697
5	0,141516	2,83032	5,000000	100,0000
Dąbie				
1	1,937758	38,75517	1,937758	38,7552
2	1,193028	23,86057	3,130787	62,6157
3	1,096345	21,92689	4,227131	84,5426
4	0,621004	12,42008	4,848135	96,9627
5	0,151865	3,03730	5,000000	100,0000

Źródło: opracowanie własne z wykorzystaniem pakietu Statistica 10.0.

**Tabela 3.** Ładunki czynnikowe zmiennych (na podstawie korelacji) dla badanych dzielnic

Zmienna	X1	X2	X3	X4	X5	X6
Pogodno						
Składowa 1	-0,238143	-0,663385	-0,318335	0,562933	-0,238143	0,815941
Składowa 2	0,150475	-0,538488	-0,751224	-0,623159	0,150475	-0,257046
Nad Odrą						
Składowa 1	0,035951	-0,750872	-0,838168	-0,616112	0,035951	-0,171721
Składowa 2	-0,073974	-0,446359	-0,154742	0,513114	-0,073974	0,850587
Składowa 3	0,984861	-0,083712	0,150887	-0,078207	0,984861	0,116351
Dąbie						
Składowa 1	0,082724	0,088650	-0,273199	-0,856766	0,082724	-0,856763
Składowa 2	0,677802	-0,842846	-0,290398	-0,020516	0,677802	0,091351
Składowa 3	0,575601	0,168115	0,831141	-0,093489	0,575601	-0,098568

Źródło: opracowanie własne z wykorzystaniem pakietu Statistica 10.0.

Dla dzielnicy Pogodno wyznaczono dwie składowe główne, dla pozostałych dzielnic trzy. Jako kryterium przyjęto wartości własne większe od jedności. W tabeli 3 umieszczono wartości ładunków czynnikowych składowych głównych dla poszczególnych dzielnic.

Można zauważyć, że w przypadku każdej dzielnicy inny zestaw zmiennych jest silnie skorelowany z pierwszą i kolejnymi osiami. W dzielnicy Pogodno i Nad Odrą zmienna rodzaj nabywanego prawa jest silnie skorelowana z pierwszą osią (tab. 3). W dzielnicy Pogodno pierwsza oś jest najsilniej skorelowana ze zmienną stan zaawansowania ( $X_6$ ) oraz powierzchnią działki ( $X_4$ ), w dzielnicy Nad Odrą z powierzchnią użytkową budynku, a w dzielnicy Dąbie z powierzchnią działki i stanem zaawansowania budowy domu.

W kolejnym kroku ponownie przeprowadzono analizę regresji dla zmiennej zależnej cena nieruchomości i zmiennymi niezależnymi składowymi (tab. 4).

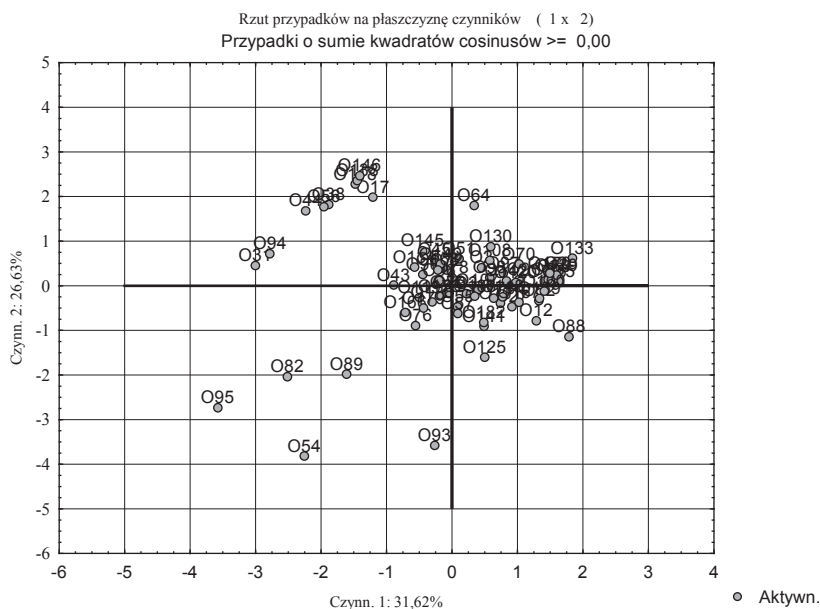
**Tabela 4.** Podsumowanie regresji zmiennej zależnej *cena nieruchomości* dla badanych dzielnic

Zmienna	b	Bł. std.	t	p
Pogodno ( $R = 0,5024$ , $F(3,65) = 7,3148$ , $p < 0,00027$ , błąd std. estymacji: 2966E2)				
Wyraz wolny	545 771	35 439,10	15,40025	0,000000
Składowa 1	-80 787	28 393,37	-2,84529	0,005904
Składowa 2	-116 366	30 939,04	-3,76115	0,000361
Nad Odrą ( $R = 0,8194$ , $F(3,42) = 28,624$ , $p < 0,0000$ , błąd std. estymacji: 1441E2)				
Wyraz wolny	583 991	21 253,46	27,47744	0,000000
Składowa 1	-106 355	16 594,85	-6,40891	0,000000
Składowa 2	-130 427	19 491,08	-6,69163	0,000000
Składowa 3	-3 079	21 283,07	-0,14465	0,885678
Dąbie ( $R = 0,6165$ , $F(3,31) = 6,3339$ , $p < 0,00178$ , błąd std. estymacji: 1718E2)				
Wyraz wolny	506 507,5	22 706,29	22,30693	0,000000
Składowa 1	44 744,7	18 460,20	2,42385	0,021385
Składowa 2	-75 983,4	20 500,09	-3,70649	0,000820
Składowa 3	124 430,0	22 283,65	5,58391	0,000004

Źródło: opracowanie własne z wykorzystaniem pakietu Statistica 10.0.

W efekcie dla badanych zmiennych uzyskano lepiej dopasowane modele niż zaprezentowane w tabeli 1, z wyjątkiem dzielnicy Pogodno. W modelu (tab. 4) dla dzielnicy Nad Odrą składowa trzecia okazała się nieistotna. Wyznaczenie wymiarów, w których można klasyfikować poszczególne obserwacje, pozwala na interpretację graficzną. Na wykresie współrzędnych czynnikowych przypadków można

zauważyć skupianie się przypadków aktywnych i dodatkowych w pewnych obszarach układu współrzędnych i ich interpretację, a zwłaszcza obserwację przypadków odstających od pozostałych. Z punktu widzenia procesu wyceny – odrzucania obserwacji odbiegających od pozostałych – jest to bardzo istotny element wykorzystanej metody. Na rysunku 1 pokazano dla przykładu rozrzut przypadków na płaszczyźnie dla dzielnicy Pogodno.



**Rys. 1.** Wykres współrzędnych czynnikowych przypadków dla dzielnicy Pogodno

Źródło: opracowanie własne z wykorzystaniem pakietu Statistica 10.0.

Na składową pierwszą największy wpływ miały zmienne  $X_1$ ,  $X_6$  (rodzaj prawa i stan zaawansowania), natomiast na składową drugą powierzchnie: użytkowa domu i działki (zmienne  $X_3$ ,  $X_4$ ). Na wykresie widoczne są dwa skupienia przypadków odstających od pozostałych. Są to obiekty: O95, O82, O89, O54, O93 w grupie pierwszej ujemnie skorelowanej z osią pierwszą i z osią drugą. Natomiast grupę drugą stanowią obiekty: O31, O94, O43, O38, O108, O17, O146 ujemnie skorelowane z osią pierwszą, a dodatnio z osią drugą. W grupie pierwszej znajdują się przede wszystkim obiekty o dużej powierzchni użytkowej domu i stosunkowo nowe. W grupie drugiej znajdują się domy o małej powierzchni działki i dużej powierzchni domu, a także dużo starsze niż w grupie pierwszej.



## 4. Zakończenie

Przeprowadzone badanie pozwala na redukcję wymiaru przestrzeni zmiennych analizowanego zjawiska, a tym samym pozwala na interpretację graficzną, co ułatwia interpretację. Umożliwia identyfikację ukrytej struktury w zbiorze zmiennych. W efekcie redukcji zmiennych i wprowadzenia nowych możliwa jest graficzna analiza skupień obiektów o podobnych cechach. W procesie szacowania wartości nieruchomości ciągle pozostaje aktualny problem doboru zmiennych, które w sposób istotny pozwalałyby wyróżnić grupy obiektów (nieruchomości) podobnych, a przede wszystkim wykorzystania w tym celu sprawdzonych metod naukowych [Colley, Lohnes 1971; Gatnar 2008; Foryś 2011]. Z tego powodu zaproponowano analizę składowych głównych, której przydatność do postawionego problemu weryfikowano na lokalnym rynku mieszkaniowym. Zbudowane funkcje regresji ze zmiennymi objaśniającymi składowymi w sposób zadawalający wyjaśniały cenę nieruchomości, a nowe zmienne składowe pozwalają na graficzną interpretację obserwowanych przypadków w adekwatnych do zjawisk układach współrzędnych.

## Literatura

- Barlett M.S., 1953, *A note on the multiplying factors for various chi square approximations*, Journal of Royal Statistical Society, Series B, vol. 16, s. 296-298.
- Colley W.W., Lohnes P.R., 1971, *Multivariate Data Analysis*, John Wiley & Sons, New York.
- Cymerman R., Hopfer A., 2012, *System, zasady i procedury wyceny nieruchomości*, PFSRM, Warszawa.
- Foryś I., 2010, *Wykorzystanie metod taksonomicznych do wyboru obiektów podobnych w procesie szacowania lokali mieszkalnych*, Studia i Materiały TNN, t. 18, nr 1.
- Foryś I., 2011, *Wielowymiarowa analiza cech mieszkań sprzedawanych na rynku warszawskim w badaniu czasu trwania oferty w systemie MLS*, Prace Naukowe Akademii Ekonomicznej we Wrocławiu, nr 176, Taksonomia 18, s. 436-446.
- Gatnar E., 2008, *Podjęcie wielomodalne w zagadnieniach dyskryminacji i regresji*, Wydawnictwo Naukowe PWN, Warszawa.
- Hardle W., Simar L., 2003, *Applied Multivariate Statistical Analysis*, Springer, Berlin.
- Jajuga K., 1993, *Statystyczna analiza wielowymiarowa*, PWN, Warszawa.
- Jolliffe I.T., 2002, *Principal Component Analysis*, Springer Series in Statistics, Springer, Berlin.
- Lis C., 2007, *Metody pomiaru podobieństwa nieruchomości w procesie wyceny w podejściach porównawczym i dochodowym*, Prace Naukowe Akademii Ekonomicznej we Wrocławiu, nr 1169, Taksonomia 14, s. 445-454.
- Pagourtzi E., Assimakopoulos V., Hatzicristos T., French N., 2003, *Real estate appraisal: A review of valuation methods*, Journal of Property Investment & Finance, vol. 21, no. 4, s. 383-401.
- Rozporządzenie Rady Ministrów z dnia 21 września 2004 r. w sprawie wyceny nieruchomości i sporządzania operatu szacunkowego, Dz.U. z 2004 r., poz. 2109.
- Scarrett D., 2003, *Property Valuation. The Five Methods*, Routledge, Taylor & Francis London – New York.
- Ustawa z 21 sierpnia 1997 r. o gospodarce nieruchomościami, Dz.U. z 2015 r., poz. 782 z późn. zm.