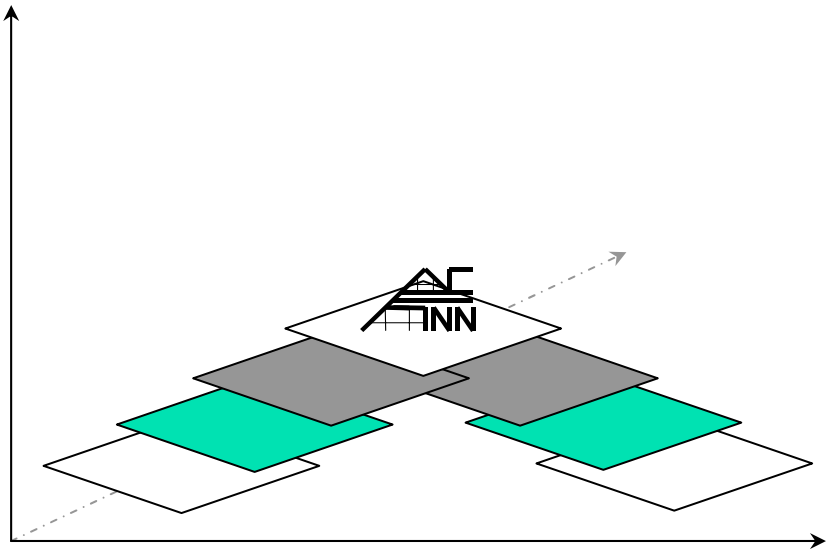
VOLUME 18 NUMBER 1

STUDIA I MATERIAŁY

TOWARZYSTWA NAUKOWEGO NIERUCHOMOŚCI

JOURNAL OF THE POLISH REAL ESTATE SCIENTIFIC SOCIETY



OLSZTYN 2010
ISSN 1733-2478

STUDIA I MATERIAŁY

TOWARZYSTWA NAUKOWEGO NIERUCHOMOŚCI

JOURNAL OF THE POLISH REAL ESTATE SCIENTIFIC SOCIETY

VOLUME 18 NUMBER 1

OLSZTYN 2010

TOWARZYSTWO NAUKOWE NIERUCHOMOŚCI

POLISH REAL ESTATE SCIENTIFIC SOCIETY

RADA WYDAWNICZA
„STUDIA I MATERIAŁY TOWARZYSTWA NAUKOWEGO NIERUCHOMOŚCI"
Editorial Council of the „Journal of the Polish Real Estate Scientific Society"

Andrzej Hopfer (przewodniczący, chairman)
Antoni Sobczak (zastępca przewodniczącego, vice-chairman)
Zdzisław Adamczewski, Józef Czaja, Józef Hozer, Ewa Kucharska-Stasiak, Maria Trojanek, Sabina Źróbek, Ryszard Źróbek, Kauko Jussi Viitanen, Radosław Wiśniewski (sekretarz, secretary)

RECENZENCI:

prof. dr hab. inż. Józef CZAJA; prof. dr hab. inż. Sabina ŹRÓBEK; prof. dr hab. inż. Ryszard ŹRÓBEK; dr inż. Radosław WIŚNIEWSKI

REDAKTOR NAUKOWY WYDAWNICTWA
Scientific Editor

Sabina Źróbek

ZASTĘPCA REDAKTORA NAUKOWEGO WYDAWNICTWA
Vice-Scientific Editor

Zdzisław Adamczewski

ADRES REDAKCJI:

TOWARZYSTWO NAUKOWE NIERUCHOMOŚCI
10-724 Olsztyn, ul. Prawocheńskiego 15

Address of the Editorial Board:

Polish Real Estate Scientific Society
10-724 Olsztyn, Prawochenskiego 15 Str.

www.tnn.org.pl
e-mail: redakcja@.tnn.org.pl

TYTUŁ DOTOWANY PRZEZ
STOWARZYSZENIE NAUKOWE IM. STANISŁAWA STASZICA

© Copyright by Towarzystwo Naukowe Nieruchomości.
© Opracowanie okładki: Małgorzata Renigier i Radosław Wiśniewski.

ISSN 1733-2478
ISBN 978-83-61564-24-9
TNN, Olsztyn 2010 r.

SPIS TREŚCI

Wprowadzenie - 5

1. **Zaawansowane metody analizy statystycznej rynku nieruchomości** - 7
   Józef Czaja, Marcin Ligas
2. **Problematyka określania wartości nieruchomości metodą analizy statystycznej rynku** - 21
   Edward Sawiłow
3. **Identification of nonstationarity type in time series of land property prices in Poland** - 33
   Radosław Wiśniewski
4. **Metody statystyczne w wycenie nieruchomości** - 49
   Marcin Ligas
5. **Modele multiplikatywne w procesie wyceny nieruchomości** - 65
   Anna Barańska
6. **Algorytm estymacji rynkowych wartości wskaźników kosztu wytworzenia i zużycia łącznego obiektów budowlanych posiadających różne okresy eksploatacji** - 83
   Tomasz Adamczyk, Janusz Dąbrowski
7. **Wykorzystanie metod taksonomicznych do wyboru obiektów podobnych w procesie wyceny lokali mieszkalnych** - 95
   Iwona Foryś
8. **Metodyka aktualizacji cen na przykładzie nieruchomości lokalowych** - 107
   Tomasz Budzyński
9. **Porównanie metod prostych oraz metody regresji hedonicznej do konstruowania indeksów cen mieszkań** - 119
   Radosław Trojanek
10. **Statystyczna analiza łódzkiego rynku nieruchomości** - 133
    Konrad Żelazowski

WPROWADZENIE

Możliwości i zastosowanie metod statystycznych do badań rynku i wyceny nieruchomości stanowią ważne pole badawcze wielu środowisk akademickich. Potrzeba i ważność takich badań wynika z faktu, iż informacje związane z rynkiem nieruchomości mają charakter probalistyczny, zatem przy określeniu wartości rynkowej nieruchomości należy stosować reguły, które rządzą wielkościami losowymi. Reguły te, ustalone zostały w dziale matematyki zwanym statystyką matematyczną, a terminologia używana w statystyce matematycznej jest zawarta w rachunku prawdopodobieństwa.

Autorzy opracowań zawartych w tym numerze kwartalnika Studia i Materiały Towarzystwa Naukowego Nieruchomości przedstawiają m.in. zaawansowane metody analizy statystycznej do analizy rynku nieruchomości gruntowych niezabudowanych, przeznaczonych pod budownictwo mieszkaniowe jednorodzinne, położonych w Krakowie.

Zwrócono także uwagę na występowanie niestacjonarności w szeregach czasowych cen nieruchomości jako na poważny problem w przeprowadzaniu analiz statystycznych.

W wycenie nieruchomości, istotnym jest dokonanie właściwego wyboru obiektów podobnych do przedmiotu wyceny. Do tego celu zaproponowano wykorzystanie metod taksonomicznych w odniesieniu do lokali mieszkalnych.

Kwartalnik zawiera także ciekawe propozycje modeli matematycznych mogących stanowić podstawę określania wartości metodą analizy statystycznej rynku czy też algorytm estymacji rynkowych wartości zużycia łącznego obiektów budowlanych.

prof. dr hab. inż. Sabina Źróbek
Redaktor Naukowy Wydawnictwa

ZAAWANSOWANE METODY ANALIZY STATYSTYCZNEJ RYNKU NIERUCHOMOŚCI[1]

Józef Czaja, Marcin Ligas
Katedra Geomatyki
Akademia Górniczo – Hutnicza w Krakowie
e-mail: czaja@agh.edu.pl

Słowa kluczowe: *modelowanie rynku nieruchomości, analiza statystyczna*

Streszczenie

Metody analizy statystycznej bazują na charakterystycznych parametrach zmiennych losowych reprezentowanych przez jednostkowe ceny nieruchomości oraz przez wyróżnione cechy nieruchomości, zwane ich atrybutami. W rozkładzie normalnym jednowymiarowej zmiennej losowej, reprezentowanej przez jednostkowe ceny nieruchomości, podstawowymi parametrami charakterystycznymi są wartość przeciętna oraz odchylenie standardowe. Natomiast dla zmiennej losowej dwuwymiarowej, reprezentowanej przez jednostkowe ceny i wybrany atrybut nieruchomości, podstawowymi parametrami charakterystycznymi, które wykorzystuje się do ustalenia rozkładu normalnego są: wartości przeciętne w rozkładach brzegowych, odchylenia standardowe w rozkładach brzegowych oraz współczynnik korelacji zupełnej (Persona).

Jeżeli w analizie rynku nieruchomości będą rozpatrywane łącznie wszystkie atrybuty i jednostkowe ceny, wtedy mamy do czynienia ze zmienną losową wielowymiarową. Charakterystycznymi parametrami wielowymiarowej zmiennej losowej są wartości przeciętne cen i atrybutów w rozkładach brzegowych, odchylenia standardowe cen i atrybutów w rozkładach brzegowych oraz macierz korelacji, która składa się ze współczynników korelacji zupełnej dla wszystkich kombinacji dwójkowych, utworzonych z jednostkowych cen i atrybutów nieruchomości z bazy.

Współczynnik determinacji R^2 zdefiniowany za pomocą elementów macierzy korelacji może stanowić miarę wyjaśniania zmienności cen nieruchomości w bazie przez rozważane atrybuty.

Na podstawie macierzy korelacji można również wyznaczyć korelacje wagowe (β_j), które określają siłę kreowania cen rynkowych przez poszczególne atrybuty nieruchomości. Wartości β_j są wyznaczane na podstawie wzajemnej współzależności jednostkowych cen i atrybutów wszystkich nieruchomości w rozważanej bazie, stąd stanowią obiektywną miarę wpływu poszczególnych

[1] Praca stanowi realizację badań statutowych w Katedrze Geomatyki AGH pod numerem 11.11.150.006

atrybutów na zmienność cen transakcyjnych na badanym rynku nieruchomości.

Wyżej opisane parametry analizy rynku nieruchomości są zilustrowane na przykładzie liczbowym dotyczącym analizy rynku nieruchomości gruntowych niezabudowanych.

1. Wprowadzenie

Wszystkie informacje związane z rynkiem nieruchomości mają charakter probabilistyczny, a to oznacza, że przede wszystkim czynniki losowe mają wpływ na to: które nieruchomości są przedmiotem umowy kupna - sprzedaży, które nieruchomości są wybrane przez rzeczoznawcę do porównywania, które atrybuty nieruchomości i w jakim udziale procentowym będą kształtowały wartość rynkową, które ceny nieruchomości są bardziej prawdopodobne, jakie wartości dochodów (czynszów) będą prognozowane do wyceny, z jakich informacji o nieruchomościach będą obliczane wartości współczynnika kapitalizacji prostej, jak będą wyznaczane wartości stopy dyskonta, jaka liczba lat projekcji dochodów będzie przyjmowana do wyceny, która metoda wyceny będzie zastosowana, a nawet, który rzeczoznawca będzie dokonywał wyceny. Powyższe uwarunkowania obligują rzeczoznawców do stosowania takich reguł, które rządzą wielkościami losowymi. Reguły te są ustalone w odrębnym dziale matematyki, zwanym statystyką matematyczną, natomiast terminologia (język) dla statystyki matematycznej jest zawarta w rachunku prawdopodobieństwa, zwanym często probabilistyką.

Każdy rzeczoznawca, w sposób świadomy lub nieświadomy, stosuje zasady statystyki matematycznej, gdyż liczenie średniej arytmetycznej z rozważanych wartości, oznacza estymację wartości przeciętnej zmiennej losowej reprezentowanej przez rozważaną wielkość. Jeżeli w tych obliczeniach rzeczoznawca uwzględnia jeszcze wagi zaufania do poszczególnych wielkości, to wtedy równocześnie stosuje rachunek prawdopodobieństwa (probabilistykę), czyli podstawowy wskaźnik statystyki matematycznej.

2. Metody analizy statystycznej rynku

Metody analizy statystycznej bazują na charakterystycznych parametrach zmiennych losowych reprezentowanych przez jednostkowe ceny nieruchomości oraz przez wyróżnione cechy nieruchomości, zwane ich atrybutami. W rozkładzie normalnym jednowymiarowej zmiennych losowych, podstawowymi parametrami charakterystycznymi są wartość przeciętna oraz odchylenie standardowe. Natomiast dla zmiennej losowej dwuwymiarowej podstawowymi parametrami charakterystycznymi, za pomocą których jest zdefiniowany rozkład normalny są: wartości przeciętne w rozkładach brzegowych, odchylenia standardowe w rozkładach brzegowych oraz współczynnik korelacji zupełnej (Persona).

Jeżeli każdy atrybut będzie rozważany z jednostkową ceną nieruchomości, to każda taka para reprezentuje zmienną losową dwuwymiarową. Graficzną ilustracją tej dwuwymiarowej zmiennej losowej jest diagram korelacyjny, który

obrazuje zależność cen i atrybutów poszczególnych nieruchomości w prostokątnym układzie współrzędnych, którego oś pozioma reprezentuje wartość rozważanego atrybutu, zaś oś pionowa reprezentuje jednostkową cenę nieruchomości. Diagram korelacyjny jest obrazem zmienności jednostkowych cen względem wartości rozważanego atrybutu. Gdy jednostkowe ceny nieruchomości zmieniają się proporcjonalnie do wartości atrybutów, wtedy wszystkie punkty diagramu korelacyjnego powinny się układać wzdłuż linii prostej. W praktycznych zastosowaniach punkty diagramu korelacyjnego tworzą linię łamaną, która obrazuje fluktuacyjną zmienność cen względem rozważanego atrybutu. Gdy linia łamana diagramu korelacyjnego zostanie zastąpiona reprezentatywną linią prostą, zwaną linią regresji, to wtedy mamy do czynienia z modelem regresji dwuwymiarowej. Zasady estymacji parametrów linii regresji są szczegółowo przeanalizowane w wielu podręcznikach statystyki matematycznej, a także w pracach dotyczących wyceny nieruchomości.

Linia regresji dla jednostkowych cen nieruchomości (c) powiązana z wyróżnionym atrybutem (a) ma następującą postać:

$$c = \hat{c} + r\frac{\sigma_c}{\sigma_a}\left(a - \hat{a}\right) \tag{1}$$

gdzie:

- $\hat{c}$ – oznacza przeciętna wartość jednostkowej ceny nieruchomości w bazie,
- $\hat{a}$ – oznacza przeciętna wartość rozważanego atrybutu w bazie,
- r – oznacza współczynnik korelacji zupełnej pomiędzy jednostkową ceną a rozważanym atrybutem,
- σ_c, σ_a – oznaczają odchylenia standardowe jednostkowej ceny i rozważanego atrybutu w rozkładach brzegowych,

Po wstawieniu do powyższego wzoru ustalonej wartości atrybutu $a = \overline{a}$ można określić prognozowaną cenę rozważanej nieruchomości. Postępując analogicznie do pozostałych atrybutów można prognozować wielokrotnie cenę rozpatrywanej nieruchomości. Wartość przeciętna z prognozowanych cen może stanowić szacowaną wartość rynkową rozważanej nieruchomości. Cały tok powyższego postępowania stanowi procedurę prognozy występującą pod nazwą modelu regresji dwuwymiarowej.

Formułę (1) można również wykorzystać do ustalenia trendu zmiany cen w czasie, czyli do prognozowania zmian jednostkowych cen transakcyjnych z tytułu przesunięcia czasu transakcji względem daty analizy.

Dla zmiennych losowych dwuwymiarowych na szczególne rozpatrzenie zasługuje współczynnik korelacji zupełnej (Persona), który - dla jednostkowej ceny oraz dla j-tego atrybutu lub czasu transakcji - jest wyznaczany według zależności:

$$r_{cj} = \frac{\frac{1}{n}\sum_{i=1}^{n}\left(c_i a_{ji}\right) - \hat{c}\hat{a}_j}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(c_i\right)^2 - \hat{c}^2} \cdot \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(a_{ji}\right)^2 - \hat{a}_j^2}} = \frac{\frac{1}{n}\sum_{i=1}^{n}\left(c_i a_{ji}\right) - \hat{c}\hat{a}_j}{\sigma_c \cdot \sigma_{aj}} \tag{2}$$

przy czym

$$\hat{c} = \frac{1}{n}\sum_{i=1}^{n} c_i \qquad \hat{a}_j = \frac{1}{n}\sum_{i=1}^{n} a_{ij} \tag{3}$$

oznaczają wartość przeciętną jednostkowych cen i ustalonego atrybutu, określoną na podstawie bazy nieruchomości wybranych do analizy rynku.

Siłę korelacji, czyli współzależności liniowej, pomiędzy jednostkową ceną a rozważanym atrybutem można określać na czterech następujących poziomach:

- dla $|r| \le 0.3$ – korelacja słaba,
- dla $0.3 < |r| \le 0.6$ - korelacja przeciętna,
- dla $0.6 \le |r| < 0.9$ - korelacja silna,
- dla $|r| \ge 0.9$ – korelacja bardzo silna.

Wartości współczynników korelacji zupełnej r_{cj} pomiędzy jednostkową ceną a j-tym atrybutem nieruchomości z bazy do porównania są dobrym wskaźnikiem do określenia udziałów wagowych k_j atrybutów. Na podstawie analizy wariancji można wykazać, że kwadrat współczynnika korelacji zupełnej stanowi miarę, jaką część zmienności jednostkowych cen wyjaśnia (CzZW) samodzielnie rozważany atrybut w odniesieniu do całkowitej zmienności jednostkowych cen (CZ), czyli:

$$r^2 = \frac{CzZW}{CZ} = \text{Część Zmienności Wyjaśnionej / Całkowita Zmienność} \tag{4}$$

Z uwagi na interakcję atrybutów, suma kwadratów współczynników korelacji r_j^2 może przybierać wartości różne od 1, stąd wyznaczone wartości r_j^2 powinny być poddane standaryzacji, czyli:

$$k_j = \frac{r_j^{\,2}}{\sum_{j=1}^{u} r_j^{\,2}} \tag{5}$$

Standaryzowane wartości udziałów wagowych k_j atrybutów można wykorzystać do wyceny nieruchomości według metod w podejściu porównawczym.

Zastosowanie zmiennej losowej dwuwymiarowej do analizy rynku nieruchomości jest procedurą uproszczoną, gdyż modele zmiany jednostkowych cen są wyznaczane przy założeniu, że atrybuty względem siebie są niezależne.

Jeżeli w analizie rynku będą rozpatrywane łącznie wszystkie atrybuty i jednostkowe ceny, wtedy mamy do czynienia ze zmienną losową wielowymiarową. Charakterystycznymi parametrami wielowymiarowej zmiennej losowej są wartości przeciętne cen i atrybutów w rozkładach brzegowych,

odchylenia standardowe cen i atrybutów w rozkładach brzegowych oraz macierz korelacji, która składa się ze współczynników korelacji zupełnej dla wszystkich kombinacji dwójkowych, utworzonych z jednostkowych cen i atrybutów nieruchomości z bazy.

Uwzględniając wcześniej przyjęte oznaczenia, macierz korelacji do wstępnej analizy rynku przyjmuje następującą postać:

$$\widetilde{\mathbf{K}} = \begin{bmatrix} 1 & r_{12} & .... & r_{1m} & r_{1c} \\ r_{21} & 1 & .... & r_{2m} & r_{2c} \\ .... & .... & .... & .... & .... \\ r_{m1} & r_{m2} & .... & 1 & r_{mc} \\ r_{c1} & r_{c2} & .... & r_{cm} & 1 \end{bmatrix} \tag{6}$$

przy czym $r_{ij} = r_{ji}$, co jest własnością macierzy symetrycznej.

W ostatnim wierszu i w ostatniej kolumnie macierzy $\widetilde{\mathbf{K}}$ znajdują się współczynniki korelacji zupełnej poszczególnych atrybutów względem jednostkowej ceny nieruchomości, zaś pozostałe elementy tej macierzy stanowią współczynniki korelacji zupełnej pomiędzy rozważanymi atrybutami, we wszystkich kombinacjach dwójkowych.

Na podstawie wartości współczynników korelacji zupełnej można dokonać wnikliwej analizy rozpatrywanego rynku nieruchomości i zweryfikować atrybuty rozważane w wybranej bazie nieruchomości.

Bezwzględna wartość współczynników korelacji dla cen w powiązaniu z poszczególnymi atrybutami powinny być, co najmniej na poziomie 0.3, czyli $\left|r_{cj}\right| \geq 0.3$. Kwadraty wartości współczynników korelacji zupełnej określają samodzielny udział każdego atrybutu w wyjaśnianiu zmienności cen nieruchomości w bazie, stąd mogą stanowić wagi istotności poszczególnych atrybutów. Atrybuty, które uzyskały bezwzględną wartość współczynników korelacji zupełnej poniżej 0.3, czyli przyjmują wartości r_j^2 mniejsze od 0.09 (9 %), w dalszych analizach mogą być pominięte. Na etapie wstępnej analizy rynku należy zwrócić szczególną uwagę na skalowanie atrybutów, które wyrażają negatywną (ujemną) korelację. W takich przypadkach należałoby powtórnie zweryfikować wartości takich atrybutów i dla nowych wartości określić elementy macierzy korelacji $\widetilde{\mathbf{K}}$.

Jeżeli bezwzględna wartość współczynników korelacji zupełnej w rozważanych parach atrybutów przekracza wartości 0.7, czyli $|r_{jk}|>0.7$, to oznacza, że między tymi atrybutami zachodzi interakcja, czyli atrybuty w rozważanych parach wyjaśniają tą samą część zmienności jednostkowych cen nieruchomości w analizowanych bazach. Dla pojedynczych parach atrybutów o współczynniku korelacji $|r_{jk}|>0.7$ powinno się eliminować ten atrybut, który posiada mniejszą bezwzględną wartość współczynnika korelacji z ceną nieruchomości.

Na podstawie dokonanej weryfikacji silnej współzależności atrybutów do wyceny należy w ostatecznym wymiarze zestawić macierz korelacji **K**, która będzie podstawą do dalszej analizy rynku nieruchomości.

Współczynnik determinacji R^2 (kwadrat współczynnika korelacji wielowymiarowej) jest zdefiniowany za pomocą elementów macierzy korelacji według następującego wzoru:

$$R^2 = 1 - \frac{\det(\mathbf{K})}{\det(\mathbf{K}_{cc})} \tag{7}$$

gdzie:

- $\det(\mathbf{K})$ – oznacza wartość wyznacznika macierzy **K**
- $\det(\mathbf{K}_{cc})$ – oznacza wartość wyznacznika z podmacierzy $\mathbf{K}_{cc}$, która powstaje ze skreślenia ostatniego wiersza i ostatniej kolumny macierzy **K**, zawierających współczynniki korelacji ceny z poszczególnymi atrybutami.

Wartość współczynnika determinacji może stanowić miarę wyjaśniania zmienności cen nieruchomości w bazie przez rozważane atrybuty.

Korelacje wagowe β_j, określające siłę kreowania cen rynkowych przez poszczególne atrybuty nieruchomości, można określić bezpośrednio z elementów macierzy korelacji **K**. Jeżeli dla każdego (j-tego) atrybutu utworzymy podmacierz $\mathbf{K}_{jc}$, która powstaje przez zastąpienie j-tego wiersza (dotyczącego j-tego atrybutu) wierszem dotyczącym współczynników korelacji ceny, to korelacje wagowe β_j można wyliczyć według następującego wzoru:

$$\beta_j = \frac{\det(\mathbf{K}_{jc})}{\det(\mathbf{K}_{cc})} \tag{8}$$

Korelacje wagowe β_j są podstawowym parametrem analizy trendów rynkowych dla szerokich rynków nieruchomości.

Na podstawie korelacji wagowych można wyprowadzić formułę na prognozowanie jednostkowej ceny nieruchomości względem ustalonych wartości wszystkich atrybutów, czyli:

$$\tilde{c} = \hat{c} + \beta_1 \frac{\sigma_c}{\sigma_{a1}}(a_1 - \hat{a}_1) + \beta_2 \frac{\sigma_c}{\sigma_{a2}}(a_2 - \hat{a}_2) + \cdots\cdots + \beta_m \frac{\sigma_c}{\sigma_{am}}(a_m - \hat{a}_m) \tag{9}$$

Z formuły tej wyraźnie wynika, że prognozowana cena $\tilde{c}$ dowolnej nieruchomości powstaje przez korektę średniej ceny $\hat{c}$ nieruchomości z bazy, za pomocą korelacji wagowych β_j mnożonych przez iloraz odchyleń standardowych σ_c / σ_{aj} oraz różnice atrybutów i ich wartości średnich $a_j - \hat{a}_j$. Ta zależność potwierdza, że obiektywnym wskaźnikiem kreatywności cen przez poszczególne atrybuty są korelacje wagowe β_j. Wartość korelacji wagowych jest uzależniona tylko od wartości elementów macierzy korelacji **K**, czyli od bazy nieruchomości do analizy rynku.

Ocenę zgodności modelu predykcji z bazą nieruchomości do analizy rynku

można wykonać według następującej procedury. Dla każdej nieruchomości występującej w podbazie realizujemy formułę (9), w wyniku tego obliczamy jej prognozowaną jednostkową cenę $\widetilde{c}_i$. Odchylenie standardowe σ_n z różnic jednostkowych cen transakcyjnych c_i i jednostkowych cen prognozowanych $\widetilde{c}_i$ każdej nieruchomości z bazy stanowi przeciętną miarę wzajemnej niezgodności jednostkowych cen transakcyjnych w rozważanej bazie. Wartość σ_n można również wyznaczyć na podstawie współczynnika determinacji R^2 oraz odchylenia standardowego cen σ_c nieruchomości z bazy, czyli według następującego wzoru:

$$\sigma_n = \sqrt{1-R^2}\cdot\sigma_c \tag{10}$$

Wskaźnik względnej niezgodności liniowej jednostkowych cen transakcyjnych w analizowanej bazie stanowi współczynnik rozproszenia, liczony jako stosunek σ_n do przeciętnej wartości z jednostkowych cen w bazie, czyli:

$$\lambda = \frac{\sigma_c}{\hat{c}} \tag{11}$$

Do oceny wskaźnika względnej zgodności jednostkowych cen w bazie nieruchomości do analizy rynku można zastosować następujące stopnie:

$(1-\lambda) > 0.95$ – zgodność bardzo wysoka
$0.95 \geq (1-\lambda) > 0.90$ – zgodność wysoka
$0.90 \geq (1-\lambda) > 0.85$ – zgodność dość wysoka
$0.85 \geq (1-\lambda) > 0.80$ – zgodność dostateczna
$0.80 \geq (1-\lambda) > 0.75$ – zgodność dopuszczalna
$0.75 \geq (1-\lambda)$ – zgodność niedopuszczalna – dobór bazy należy zweryfikować

Wyżej podane zasady analizy rynku stanowią podstawę do wyboru odpowiedniej bazy do analizy rynku oraz do procedury szacowania rynkowej wartości nieruchomości a także do jej oceny wiarygodności.

3. Przykład liczbowy

Do analizy rynku wybrano 23 nieruchomości gruntowe niezabudowane, które były przedmiotem transakcji rynkowych na przestrzeni czasu 29 miesięcy, w latach 2005 – 2007. Wybrane nieruchomości w miejscowym planie zagospodarowania mają przeznaczenie pod budownictwo jednorodzinne i są położone w dwóch sąsiednich bardzo atrakcyjnych dzielnicach miasta Krakowa.

Na podstawie dokumentacji prawnej i wizji terenowej sformułowano 6 atrybutów, których własności opisano w następującej skali pięciostopniowej:

1 – cecha bardzo niekorzystna,
2 – cecha niekorzystna,
3 – cecha przeciętna,

4 – cecha korzystna,

5 – cecha bardzo korzystna.

Czas transakcji został opisany za pomocą liczby miesięcy przesunięcia każdej transakcji względem najwcześniejszej transakcji. Jednostkowe ceny transakcyjne wraz z czasem transakcji i atrybutami rozważanych nieruchomości tworzą bazę do analizy rynku nieruchomości, której wartości zamieszczono w tabeli 1.

Tabela 1

Zestawienie bazy nieruchomości gruntowych przeznaczonych pod budownictwo jednorodzinne

L.p.	**Czas [mies.]**	**Lokali-zacja**	**Uzbro-jenie**	**Komu-nikacja**	**Otocze-nie**	**Zagospo-darowanie**	**Kształt działki**	**Cena [zł/m²]**
1	0	4	5	4	4	3	5	420
2	2	3	4	3	3	3	4	415
3	5	4	5	4	4	4	5	470
4	7	4	5	5	4	4	5	490
5	8	3	4	3	4	3	4	390
6	10	3	3	3	3	3	3	400
7	11	4	4	3	4	4	4	440
8	11	4	5	4	4	4	4	480
9	12	4	5	4	5	5	4	520
10	15	5	5	5	4	4	5	750
11	16	4	4	5	5	4	4	610
12	18	5	5	5	5	5	5	720
13	19	3	5	4	5	5	5	670
14	20	5	5	5	5	5	4	790
15	20	5	5	5	5	5	5	810
16	22	3	3	4	4	4	4	580
17	23	3	4	2	4	3	4	540
18	25	4	5	5	4	4	4	680
19	26	5	5	4	5	5	4	780
20	27	5	5	5	5	5	5	790
21	28	4	4	4	4	3	4	610
22	29	5	5	5	5	5	5	820
23	29	5	5	5	5	5	4	850
$\hat{X}$	16.652	4.087	4.565	4.174	4.348	4.130	4.348	609.783
$\hat{\sigma}$	8.855	0.793	0.662	0.887	0.647	0.815	0.573	156.078

Źródło: opracowanie własne.

Wykorzystując program STATISTICA obliczono macierz współczynników korelacji zupełnej dla zmiennej losowej 8-wymiarowej, której wartości zamieszczono w tabeli 2.

Tabela 2

Macierz $\widetilde{\mathbf{K}}$ współczynników korelacji

Atrybut	Czas	Lokali-zacja	Uzbro-jenie	Komu-nikacja	Otocze-nie	Zagospo-darowanie	Kształt działki	Cena
Czas	1.000	0.432	0.113	0.361	0.561	0.498	-0.029	**0.807**
Lokalizacja	0.432	1.000	0.681	**0.753**	0.647	0.685	0.431	**0.770**
Uzbrojenie	0.113	0.681	1.000	0.599	0.581	0.615	0.656	**0.498**
Komunikacja	0.361	**0.753**	0.599	1.000	0.602	0.659	0.502	**0.705**
Otoczenie	0.561	0.647	0.581	0.602	1.000	**0.858**	0.394	**0.750**
Zagospodarowanie	0.498	0.685	0.615	0.659	**0.858**	1.000	0.385	**0.767**
Kształt działki	-0.029	0.431	0.656	0.502	0.394	0.385	1.000	**0.336**
Cena	**0.807**	**0.770**	**0.498**	**0.705**	**0.750**	**0.767**	**0.336**	**1.000**

Źródło: opracowanie własne.

Na podstawie zamieszczonych wartości współczynników korelacji zupełnej zostanie dokonana wnikliwa analiza rozpatrywanej bazy nieruchomości. W ostatnim wierszu i w ostatniej kolumnie macierzy $\widetilde{\mathbf{K}}$ znajdują się współczynniki korelacji zupełnej czasu i poszczególnych atrybutów względem jednostkowej ceny nieruchomości, których wartości zawierają się w przedziale od 0.34 do 0.81, czyli są na poziomie przeciętnym lub wysokim. Współczynniki korelacji zupełnej pomiędzy rozważanymi atrybutami, we wszystkich kombinacjach dwójkowych, przyjmują wartości od -0.03 do 0.86. Dla dwóch par atrybutów: Komunikacja – Lokalizacja i Otoczenie – Zagospodarowanie współczynniki korelacji zupełnej przyjmują wartości większe od 0.7, a to świadczy, że atrybuty w rozważanych parach wyjaśniają tę samą część zmienności jednostkowych cen nieruchomości w bazie. W dalszej analizie rynku wyeliminowano dwa atrybuty: Komunikację i Otoczenie, stąd macierz współczynników korelacji przybiera wymiar 6 x 6, co ilustruje tabela 3.

Tabela 3

Macierz $\mathbf{K}$ współczynników korelacji po usunięciu z bazy atrybutów „Komunikacja" i „Otoczenie"

Atrybut	Czas	Lokalizacja	Uzbro-jenie	Zagospo-darowanie	Kształt działki	Cena
Czas	1.000	0.432	0.113	0.498	-0.029	**0.807**
Lokalizacja	0.432	1.000	0.681	0.685	0.431	**0.770**
Uzbrojenie	0.113	0.681	1.000	0.615	0.656	**0.498**
Zagospodarowanie	0.498	0.685	0.615	1.000	0.385	**0.767**
Kształt działki	-0.029	0.431	0.656	0.385	1.000	**0.336**
Cena	**0.807**	**0.770**	**0.498**	**0.767**	**0.336**	**1.000**

Źródło: opracowanie własne.

Na podstawie zweryfikowanej macierzy współczynników korelacji, zawartej w tabeli 3 obliczono przy użyciu programu STATISTICA korelacje wagowe dla czasu i poszczególnych atrybutów wraz z ich odchyleniami standardowymi. Wyniki tych obliczeń przedstawiono w tabeli 4.

Tabela 4

Korelacje wagowe dla poszczególnych atrybutów oraz współczynnik determinacji

Atrybuty	**Korelacja wagowa** β_j	**Odchylenie standardowe** $\sigma(\beta_j)$
Czas	0.563356	0.091848
Lokalizacja	0.326467	0.116406
Uzbrojenie	-0.009266	0.125700
Zagospodarowanie	0.216406	0.112745
Kształt działki	0.134737	0.096070
R = 0,955352 R^2 = 0,9127		
$\sigma_0 = \sqrt{1-0.9127}\cdot 156.078 = 46.12$ $\lambda = \frac{46.12}{609.783} = 0.076$		

Źródło: opracowanie własne.

Wartości zawarte w tabeli 4 pozwalają sformułować wniosek, że w rozważanej bazie nieruchomości zmienność jednostkowych cen transakcyjnych jest wyjaśniana w następujących stopniach:

- czas transakcji w części 0.56,
- lokalizacja w części 0.33,
- uzbrojenie w części 0.01,
- zagospodarowanie w części 0.22,
- kształt działki w części 0.13.

Z porównania wartości korelacji wagowych w odniesieniu do ich odchyleń standardowych wynika, że największy wpływ na zmienność jednostkowych cen w rozważanej bazie ma czas transakcji, zaś atrybut „uzbrojenie" ma zaniedbywanie mały wpływ na zmienność jednostkowych cen transakcyjnych nieruchomości w analizowanej bazie.

Model predykcji jednostkowych cen nieruchomości dla rozważanej bazy przyjmuje następującą postać:

$$\begin{aligned}\widetilde{c} &= 609.78 + 9.93(t - 16.65) + 64.28(a_{lok} - 4.09) + \\ &+ 41.45(a_{zag} - 4.13) + 36.72(a_{kszt} - 4.35)\end{aligned} \tag{12}$$

Ocena zgodności modelu predykcji (12) z bazą nieruchomości do analizy rynku prowadzi do wartości $\sigma_0 = 46.12$, a to oznacza, że przeciętna różnica pomiędzy

jednostkową ceną transakcyjną c_i i jednostkową cen prognozowaną $\tilde{c}_i$ nieruchomości z bazy przyjętej do analizy rynku jest na poziomie 46 zł/m², co w przeliczeniu na przeciętną wartość jednostkowych cen transakcyjnych $\hat{c} = 609.78$ daje $\lambda = 0.076$. Wartość wskaźnika względnej zgodności jednostkowych cen w bazie nieruchomości do analizy rynku jest na poziomie $1 - \lambda = 0.92$, a to odpowiada wysokiej wzajemnej zgodności bazy.

Dla analizy porównawczej zostały wyznaczone współczynniki zmiany jednostkowych cen transakcyjnych względem czasu transakcji według trzech następujących procedur:

- sposobem interwałowym,
- liniowej regresji,
- korelacji wagowych.

W wyniku tych obliczeń uzyskano następujące wartości:

- sposobem interwałowym B = 6.13 zł/m² x miesiąc,
- liniowej regresji B = 14.23 zł/m² x miesiąc,
- korelacji wagowych B = 9.93 zł/m² x miesiąc.

Różnice wartości otrzymanych współczynników zmiany jednostkowych cen transakcyjnych w czasie względem najdokładniejszej procedury korelacji wagowej wynoszą około ± 4 zł/m² na miesiąc, co w przeliczeniu na 29 miesięcy daje różnicę w formie poprawki korygującej 116 zł/m², a to stanowi 19 % średniej ceny transakcyjnej w rozważanej bazie. Otrzymane wartości świadczą o tym, że stosując przybliżone metody korygowania jednostkowych cen transakcyjnych ze względu na czas transakcji możemy deformować ceny nieruchomości nawet w granicach 19 % ich wartości.

Wyniki powyższej analizy potwierdzają ogólną tezę, że uproszczone metody analizy rynku nieruchomości mogą dawać bardzo rozbieżne wartości parametrów określających zmienność jednostkowych ceny transakcyjne.

4. Literatura

RAO C. R. 1982 Modele liniowe statystyki matematycznej. PWN, Warszawa.

CHOW G. C. 1983. Econometrics. New York 1983

ADAMCZEWSKI Z. 1993. Wstęp do teorii taksacji. Przegląd Geodezyjny nr 3/93. Warszawa

ADAMCZEWSKI Z., CZARNECKA K. 1995. Modelowanie matematyczne ekowartości. Przegląd Geodezyjny nr 8/95. Warszawa

ŹRÓBEK R., ŹRÓBEK S. 1996. Metoda cenowo-porównawcza w szacowaniu nieruchomości. Wycena nr 3/96. Olsztyn

ŹRÓBEK S., BEŁEJ M. 2000. Podejście porównawcze w szacowaniu nieruchomości. Educaterra 6a Olsztyn

CZAJA J. 2001. Metody szacowania wartości rynkowej i katastralnej nieruchomości. KOMP-SYSTEM. Kraków

KUCHARSKA-STASIAK E. (Redaktor naukowy) 2001. The Appraisal of Real Estate (Wycena nieruchomości) – wydanie polskie tłumaczone z wydania

amerykańskiego i kanadyjskiego. PFSRM. Warszawa
HOZER J., KOKOT S., KUŹMIŃSKI W. 2002. Metody analizy statystycznej rynku w wycenie nieruchomości. Warszawa
CZAJA J., KULCZYCKI M. 2002. Parametry oceny wiarygodności i spójności informacji rynkowych o nieruchomościach. Geodezja – Półrocznik AGH. Kraków
PARZYCH P. 2002. Zastosowania sztucznych sieci neuronowych w szacowaniu nieruchomości. Geodezja – półrocznik AGH. Kraków
PARZYCH P. 2002. Statystyczna analiza informacji o sprzedanych nieruchomościach. Geodezja – półrocznik AGH. Kraków
ADAMCZEWSKI Z. 2002. Nieliniowe i nieklasyczne algorytmy w geodezji. Warszawa
CZAJA J., PARZYCH P. 2003. Analiza rynku nieruchomości za pomocą testów zgodności. Geodezja – Półrocznik AGH. Kraków
International Valuation Standards. Seventh Edition. 2005. (Międzynarodowe Standardy Wyceny. Wydanie polskie), Polska Federacja Stowarzyszeń Rzeczoznawców Majątkowych. Warszawa
JAJUGA K., KRYSIAK Z. (praca zbiorowa) CZAJA J. (autor trzeciego rozdziału) 2005. Credit Risk of Mortage Loans Modeling and Management Związek Banków Polskich. Warszawa
ADAMCZEWSKI Z. 2006. Elementy modelowania matematycznego w wycenie nieruchomości. Podejście porównawcze. Oficyna Wydawnicza Politechniki Warszawskiej. Warszawa

ADVANCED STATISTICAL ANALYSIS FOR REAL ESTATE MARKET RESEARCH

Józef Czaja, Marcin Ligas
Department of Geomatics
University of Science and Technology, AGH, Krakow
e-mail: czaja@agh.edu.pl

Key words: *statistical analysis, property market modeling*

Abstract

Methods of statistical analysis are based on characteristic parameters of random variables being represented by the unit prices and the distinguished property's features, called attributes. In the univariate normal distribution of random variable (unit property price), the basic parameters are expected value and standard deviation. In turn, in bivariate normal distribution (unit property price and chosen attribute) the basic parameters are expected values in marginal distributions, standard deviations in marginal distributions and correlation coefficient.

If in the analysis of property market all the attributes and property prices are considered jointly we have then multivariate random variable. The characteristic parameters of multivariate random variables are expected values of property prices and attributes in marginal distributions, standard deviations in marginal distributions and correlation matrix.

Determination coefficient R^2 defined by means of elements of correlation matrix is a statistical measure which explains prices' variability according to attributes.

On the basis of correlation matrix we can also determine correlation weights (β_j), which describe the power of creating market prices by particular property characteristics.

The values of β_j are determined on the basis of mutual interdependence between property prices and all the attributes in database, thus they constitute an objective measure of the influence of particular attributes over the variability of property prices.

The aforementioned characteristics of real estate market analysis are presented in an example concerning non build-up land property.

PROBLEMATYKA OKREŚLANIA WARTOŚCI NIERUCHOMOŚCI METODĄ ANALIZY STATYSTYCZNEJ RYNKU

Edward Sawiłow
Katedra Gospodarki Przestrzennej
Uniwersytet Przyrodniczy we Wrocławiu
e-mail: e.sawilow@wp.pl

Słowa kluczowe: *nieruchomość, model, analiza statystyczna.*

Streszczenie

Wycena nieruchomości opiera się na analizie zależności zachodzących pomiędzy zbiorem cech opisujących nieruchomości w przestrzeni wielowymiarowej, a ich cenami. Zastosowanie metod analizy statystycznej rynku wydaje się być jak najbardziej właściwą metodą określania wartości nieruchomości, w przypadku dużego zbioru informacji z lokalnego rynku nieruchomości. W artykule opisano metody skalowania cech jakościowych. Zaproponowano kilka prostych, pod względem postaci analitycznej, modeli mogących stanowić podstawę do określenia wartości nieruchomości metodą analizy statystycznej rynku.

1. Wprowadzenie

Zgodnie z rozporządzeniem Rady Ministrów z dnia 21 września 2004 r. w sprawie wyceny nieruchomości i sporządzania operatu szacunkowego (Dz. U. Nr 207, poz. 2109 ze zm.) wartość nieruchomości, w metodzie analizy statystycznej rynku, określa się przy użyciu metod stosowanych do analizy statystycznej. Rozporządzenie nie precyzuje tych metod. Pod pojęciem analiza statystyczna, należy rozumieć wykrywanie prawidłowości w badanych zjawiskach i interpretowanie ich za pomocą metod statystyki matematycznej. Istnieje bardzo szeroka literatura dotycząca zastosowania metod statystycznych w modelowaniu zjawisk ekonomicznych i gospodarczych.

Metodę analizy statystycznej rynku zaliczyć należy do metod wyceny masowej. W metodach wyceny masowej wartość nieruchomości określana jest na podstawie modeli matematycznych opisujących zależności pomiędzy cechami nieruchomości, a ich cenami. Postać analityczna modeli jest mniej lub bardziej złożona. W metodzie analizy statystycznej rynku, postać analityczna powinna być prosta i łatwa w aplikacji. Bardzo istotnym elementem modelowania jest właściwy dobór cech diagnostycznych opisujących lokalny rynek nieruchomości. Szczególnego znaczenia nabiera przyjęcie optymalnej metody skalowania cech jakościowych. W dotychczasowej praktyce wyceny przyjęto skalowanie liniowe.

W przypadku modelowania matematycznego podstawą relacji model – obiekt jest uogólnienie teorii podobieństwa. Istotą jest dokonywanie wyboru, spośród elementów rozpatrywanego zbioru, połączonego wieloma różnymi relacjami, tylko tych elementów, które są obdarzone cechami istotnymi z punktu widzenia analizowanego rynku nieruchomości. Proces modelowania jest procesem przetwarzania informacji, przepływających pomiędzy dwoma wycinkami rzeczywistości, będącymi modelowanym obiektem i jego modelem. Modelowanie matematyczne wartości nieruchomości może mieć zastosowanie praktyczne przy określaniu wartości nieruchomości dla różnych potrzeb. Podstawę do określenia wartości nieruchomości metodą analizy statystycznej rynku powinny stanowić proste modele matematyczne reprezentatywne na lokalnych rynkach nieruchomości. Czynnością poprzedzającą wybór nieruchomości z lokalnego rynku powinno być określenie zakresu i rodzaju atrybutów opisujących nieruchomości, a także ustalenie sposobu ich opisu i pomiaru. Nieruchomości stanowią obiekty badawcze w przestrzeni wielowymiarowej. Bardzo istotnym czynnikiem jest ustalenie kryteriów podobieństwa nieruchomości opisywanych w przestrzeni wielowymiarowej. Bardzo przydatnym narzędziem przy analizie lokalnego rynku nieruchomości, dla potrzeb określenia wartości nieruchomości są metody wielowymiarowej analizy porównawczej (SAWIŁOW 2009).

2. Podstawy teoretyczne

Metoda analizy statystycznej rynku jest procesem, w wyniku, którego określona zostaje wartość nieruchomości w oparciu o modele matematyczne opracowane na podstawie informacji o cechach i cenach transakcyjnych nieruchomości, które były przedmiotem obrotu rynkowego. W metodzie tej zakłada się, że wartość rynkowa nieruchomości jest funkcją cen transakcyjnych i cech nieruchomości podobnych do wycenianej.

Analiza statystyczna jest dyscyplina naukowa pozwalającą na analizę obiektów i zjawisk złożonych, tj., takich, na których stan wpływa jednocześnie wiele cech. Do podstawowych pojęć w analizie statystycznej rynku zalicza się nieruchomości i cechy. Zarówno nieruchomości jak i cechy mają liczbowe charakterystyki, co umożliwia ich wszechstronne analizowanie oraz przetwarzanie za pomocą metod analizy statystycznej rynku.

Przedmiotem analizy statystycznej rynku jest zbiór nieruchomości N zawierający n nieruchomości $N=\{n_1, n_2, \dots n_n\}$ oraz zbiór atrybutów A o m elementach $A=\{\alpha_1, \alpha_2,\dots \alpha_m\}$ i cen nieruchomości $C=\{ c_1, c_2,\dots c_n\}$. Liczbowy obraz uwzględnionych w wycenie atrybutów uzyskuje się w wyniku przekształcenia X_j: $N\rightarrow R^n$, w wyniku, którego otrzymuje się macierz opisywanego fragmentu rynku nieruchomości w postaci $X=[x_{ij}]_{nxm}$ (gdzie x_{ij} – wartość atrybutu α_j dla nieruchomości n_i). Wiersze macierzy X należy traktować jako liczbowe reprezentacje nieruchomości, natomiast kolumny macierzy X należy interpretować jako liczbowe wartości atrybutów. Do określenia siły związku między nieruchomościami wykorzystywane są w analizie statystycznej rynku miary

podobieństwa.

W metodzie analizy statystycznej rynku jedną z pierwszych czynności jest opis i charakterystyka nieruchomości wycenianej. Następnie definiuje się rynek lokalny i ustala zbiór nieruchomości podobnych do wycenianej. Kolejną czynnością jest ustalenie rodzaju oraz liczby cech rynkowych wpływających na poziom cen na lokalnym rynku. Cechy nieruchomości mają właściwości jakościowe lub ilościowe. Zgodnie z obowiązującymi standardami wyceny nieruchomości, wartości tych cech muszą być podane w przyjętych skalach pomiarowych. Zakres skali ocen cech może być zróżnicowany. W praktyce najczęściej przyjmuje się skale liniową, co oznacza, że na jednostkę przyrostu skali przyrost cen jest stały. W artykule zaproponowano i poddano analizie merytorycznej skalę nieliniową dla cech jakościowych. W metodzie tej wartość współczynnika skali uzależniona jest od średniej ceny nieruchomości w wyróżnionym przedziale danego atrybutu opisującego nieruchomość. Poniżej przedstawiono procedurę postępowania skalowania nieliniowego.

1) Aktualizujemy ceny transakcyjne na lokalnym rynku nieruchomości..
2) Ustalamy zbiór atrybutów opisujących nieruchomości na lokalnym rynku i dzielimy każdy atrybut na przedziały klasowe.
3) Dla każdego atrybutu grupujemy nieruchomości w wyodrębnionych przedziałach klasowych i obliczamy w każdym z nich ceny średnie.
4) Korzystając z poniższego wzoru obliczamy dla każdego atrybutu wartości skali cech w poszczególnych przedziałach klasowych.

$$x_{ij} = 1 + \left(\max_j \left\{ S_j^i \right\} - 1 \right) \cdot \frac{Csr_j^i - \min_j \left\{ Csr_j^i \right\}}{\max_j \left\{ Csr_j^i \right\} - \min_j \left\{ Csr_j^i \right\}} \tag{1}$$

gdzie:

S_j^i - j - ta skala dla i - tej cechy,

Csr_j^i - średnia arytmetyczna cen transakcyjnych w j - tym przedziale, dla i - tej cechy.

Przy określaniu wartości nieruchomości wartość merytoryczna poszczególnych cech odgrywa rolę podstawową, niemniej jednak wpływ poszczególnych cech opisujących nieruchomości nie jest jednakowy. W celu ustalenia wpływu cech na wartość katastralną nieruchomości wprowadza się współczynniki wagowe (SAWIŁOW 2004).

W metodzie analizy statystycznej rynku, przy określaniu wartości nieruchomości, najczęściej stosowany jest liniowy model regresji wielorakiej (MLRW) (DRAPER, SMITH 1973; ŹRÓBEK S., ŹRÓBEK R., 1993), w którym ceny reprezentują zmienną zależną, zaś cechy nieruchomości stanowią zmienne niezależne. Postać analityczną modelu liniowej regresji wielorakiej można zapisać wzorem:

$$W = \alpha_o + \sum_{j=1}^{m} \alpha_j \cdot x_j \tag{2}$$

Parametry powyższego modelu wyznacza się najczęściej za pomocą metody najmniejszych kwadratów. Zastosowanie metody najmniejszej kwadratów do estymacji parametrów powyższego równania wymaga spełnienia założeń metody najmniejszych kwadratów (DRAPER, SMITH 1973). Miarą dopasowania modelu liniowej regresji wielokrotnej, do cen rynkowych nieruchomości, jest kwadrat współczynnika korelacji wielorakiej. Im bliższy jedności jest ten współczynnik, tym lepiej otrzymany model opisuje lokalny rynek nieruchomości.

Bardzo przydatną i prostą formułą możliwą do zastosowania w metodzie analizy statystycznej rynku jest metoda wielowymiarowej analizy porównawczej (MWAP) (SAWIŁOW 2009). Formułę według której możemy określić wartość nieruchomości, metodą analizy statystycznej rynku, można w tym przypadku zapisać w postaci:

$$w_i = \frac{\sum_{i=1}^{n} c_i}{n} \cdot q_i \tag{3}$$

gdzie:

$$q_i = \sum_{j=1}^{m} p_j \cdot z_{ij},$$

$$z_{ij} = \frac{x_{ij}}{\bar{x}_j}.$$

Powyższa formuła jest bardzo podobna do modelu powszechnej taksacji opublikowanego w pracach (TELEGA, BOJAR, ADAMCZEWSKI 2002; SAWIŁOW 2004b) i może być wykorzystana, zdaniem autora, jako jedna z metod analizy statystycznej rynku.

Nietrudno wykazać, że identyczną postać analityczną jak w modelu (2) mają zmodyfikowane metody porównywania parami (MPP) i korygowania ceny średniej (MKCS) (SAWIŁOW 2008). Tematyką wyceny nieruchomości w podejściu porównawczym, zajmowało się wielu autorów (ADAMCZEWSKI 2006, CZAJA 2001, SAWIŁOW 2008).

Postać analityczną modelu określenia wartości nieruchomości zmodyfikowaną metodą porównywania parami (MPP) można zapisać w postaci:
gdzie:

$$\bar{c} = \frac{\sum_{i=1}^{n} c_i}{n},$$

$$\bar{x}_j = \frac{\sum_{i=1}^{n} x_{ij}}{n},$$

$$\alpha_j = \frac{\left(\max_i\{c_i\} - \min_i\{c_i\}\right) \cdot p_j}{\max_i\{x_{ij}\} - 1},$$

$$\alpha_o = \bar{c} - \sum_{j=1}^{m} \alpha_j \cdot \bar{x}_j.$$

Zaproponowany model będący modyfikacją metody porównywania parami może być zastosowany, zdaniem autora, do wyceny nieruchomości w metodzie analizy statystycznej rynku. Na podstawie zbioru informacji o nieruchomościach na lokalnym rynku, obliczamy parametry a_o i a_j modelu. Podstawiając do otrzymanego modelu wartości cech wycenianych nieruchomości otrzymamy wartość każdej nieruchomości, dla danego lokalnego rynku nieruchomości, określoną metodą analizy statystycznej rynku. Czy otrzymane na podstawie powyższego modelu wartości nieruchomości są określone z większą dokładności od tradycyjnej metody porównywania parami? Odpowiedź nie jest jednoznaczna i zależy w dużym stopniu od przyjętej miary podobieństwa nieruchomości przy wyborze trzech porównywalnych nieruchomości. W proponowanym modelu uwzględniamy w obliczeniach cechy wszystkich nieruchomości podobnych na analizowanym rynku lokalnym, co zdaniem autora, jest pozbawione subiektywizmu wyboru bezpośrednio porównywalnych nieruchomości.

Drugą metodą wyceny nieruchomości w podejściu porównawczym, którą analizowano pod kątem przydatności do określenia wartości metodą analizy statystycznej rynku, po jej modyfikacji, jest metoda korygowania ceny średniej (MKCS). Wzór na określenie jednostkowej wartości nieruchomości, zmodyfikowaną metodą korygowania ceny średniej, może być zapisany w postaci funkcji liniowej wielu zmiennych równaniem:
gdzie:

$$\alpha_j = \frac{\left(\max_i\{c_i\} - \min_i\{c_i\}\right) \cdot p_j}{\max_i\{x_{ij}\} - 1}, \tag{5}$$

$$\alpha_o = \min_i\{c_i\} - \sum_{j=1}^{m} \alpha_j.$$

Postać analityczna powyższego równania jest identyczna z równaniem w modelu (MPP) i z równaniem funkcji liniowej regresji wielu zmiennych (MLRW). Różnice występują we wzorach na estymowane parametry a_o i a_j modelu. W metodzie tej, po ustaleniu wag cech rynkowych, nie uwzględnia się

w estymowanych parametrach modelu cech wszystkich nieruchomości, a jedynie cechy nieruchomości o skrajnych cenach. Opisany wyżej model wartości nieruchomości zmodyfikowaną metodą korygowania ceny średniej (MKCS) nie może zatem być, zdaniem autora, wykorzystany w metodzie analizy statystycznej rynku. Proponowane modyfikacje metod wyceny w podejściu porównawczym oraz metoda wielowymiarowej analizy porównawczej obok metod regresji wielorakiej mogą być wykorzystane do określenia wartości nieruchomości metodą analizy statystycznej rynku.

3. Badania własne

Proponowane w poprzednim rozdziale metody analizy statystycznej rynku, zweryfikowane zostały na danych z lokalnego rynku nieruchomości. Badania przeprowadzono w oparciu o informacje zawarte w rejestrze cen i wartości nieruchomości dotyczących transakcji sprzedaży działek budowlanych na obszarze obrębu Wojszyce we Wrocławiu. Informacje te uzupełnione zostały wynikami lustracji nieruchomości w terenie. W wyniku analizy lokalnego rynku nieruchomości i pozyskanych informacji ustalono zbiór atrybutów opisujących lokalny rynek nieruchomości. Zbiór ten obejmował następujące atrybuty:

1) dostępność (x_1),
2) wyposażenie w urządzenia infrastruktury technicznej(x_2),
3) front działki (x_3),
4) kształt działki (x_4),
5) stan zagospodarowania (x_5),
6) sąsiedztwo (x_6),
7) powierzchnia nieruchomości (x_7).

Jako zmienną zależną przyjęto zaktualizowaną cenę jednego metra kwadratowego nieruchomości (y). Aktualizacje cen przeprowadzono na podstawie parametrów modelu liniowej regresji. Wszystkie atrybuty zostały wyskalowane w skali liniowej pięciopunktowej. W tabeli 1 przedstawiono charakterystykę przyjętych do analizy nieruchomości.

Tabela 1

Charakterystyka nieruchomości

Numer nieruchomości	Atrybuty nieruchomości							Cena
	X_1	X_2	X_3	X_4	X_5	X_6	X_7	Y
1	5	4	3	3	2	3	4	213.34
2	2	3	2	3	3	4	3	236.76
3	2	3	2	4	2	2	3	198.34
4	2	3	4	3	2	2	4	219.26
5	2	4	2	5	5	2	1	265.86
6	3	5	2	4	3	3	2	259.36

7	2	5	4	3	1	5	4	255.25
8	4	3	2	3	3	4	3	236.76
9	3	5	4	4	5	2	5	302.45
10	2	3	4	3	4	3	5	251.96
11	3	2	4	2	3	1	4	199.34
12	5	2	3	4	2	2	3	229.98
13	3	3	4	5	2	3	4	258,36
14	2	3	2	2	4	3	2	267.34
15	4	3	2	4	5	2	4	276.35
16	2	3	2	3	2	3	2	165.23
17	4	5	3	4	2	3	5	259.86
18	5	4	3	3	4	3	4	276.35
19	3	4	5	3	3	4	4	223.56
20	2	3	4	5	3	4	3	275.25
21	5	5	4	4	3	2	3	265.98
$\overline{x}$	3.095	3.571	3.095	3.524	3.000	2.8571	3.476	244.62

Źródło: opracowanie własne.

Współczynniki wagowe cech obliczono w oparciu o analizę macierzy korelacji. Dla danych zamieszczonych w tabeli 1 obliczono współczynniki korelacji. Na podstawie obliczonych współczynników korelacji cech nieruchomości i cen transakcyjnych ustalono współczynniki wagowe. Współczynniki te zamieszczono w tabeli 2.

Tabela 2

Współczynniki wagowe cech

w_1	w_2	w_3	w_4	w_5	w_6	w_7
0,0731	0,1957	0,1037	0,2478	0,08791	0,1453	0,1465

Źródło: opracowanie własne.

W metodzie wielowymiarowej analizy porównawczej (MWAP) w pierwszej kolejności przeprowadzono normalizację cech nieruchomości zamieszczonych w tabeli 1, a następnie znormalizowane wartości cech nieruchomości przemnożono odpowiednio przez współczynniki wagowe, otrzymując znormalizowane ważone wartości cech i po ich podstawieniu do wzoru (3), otrzymano wartości nieruchomości. Wyniki obliczeń w metodzie (MWAP) zamieszczono w drugiej kolumnie w tabeli 3.

Określenie wartości nieruchomości metodami (MPP) oraz (MLRW) przeprowadzono również dla danych wyjściowych zamieszczonych w tabeli 1.

Wszystkie wyniki obliczeń wartości nieruchomości proponowanymi metodami analizy statystycznej rynku zawiera tabela 3.

Tabela 3

Wyniki obliczeń

Cena	Modele		
[zł/m²]	MWAP	MZMPP	MRW
213,34	251,60	247,81	233,50
236,76	221,95	232,99	235,57
198,34	207,10	228,50	209,36
219,26	216,60	232,14	208,05
265,86	248,89	247,73	280,39
259,36	248,99	247,41	252,90
255,26	273,58	257,51	235,28
236,76	233,51	238,01	242,72
302,45	298,19	270,65	299,43
251,96	253,68	248,18	259,26
199,34	186,50	217,47	200,15
229,98	219,23	232,87	215,26
258,36	269,22	256,64	245,60
267,34	189,16	217,49	227,70
276,35	250,47	247,59	276,28
165,23	192,04	219,96	202,83
259,86	286,73	265,54	254,35
276,35	265,92	253,84	270,42
223,56	276,03	257,91	261,05
275,25	272,74	257,10	266,00
265,98	274,80	259,59	260,88

Źródło: opracowanie własne.

Do oceny dokładności określenia wartości nieruchomości, zastosowano jako miarę dokładności błąd standardowy. W tabeli 4 zamieszczono błędy standardowe określonych, proponowanymi metodami analizy statystycznej rynku, wartości nieruchomości.

Tabela 4

Błędy standardowe

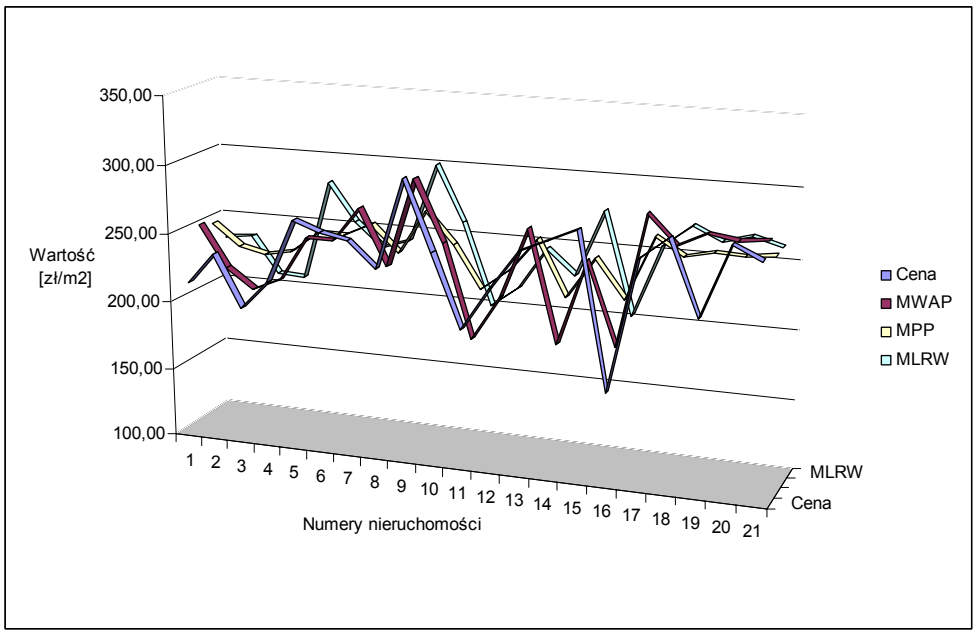
Błędy standardowe		
MWAP	MZMPP	MARW
26.55	25,02	17,82

Źródło: opracowanie własne.

Z tabeli 4 wynika, że najmniejszym błędem charakteryzuje się model liniowej

regresji wielokrotnej (MLRW). Modele (MWAP) i (MPP) są porównywalne pod względem dokładności określenia wartości nieruchomości.

W celu lepszego zilustrowania otrzymanych wyników i dokładności określenia wartości nieruchomości w poszczególnych metodach, przedstawiono graficznie relacje pomiędzy cenami transakcyjnymi, a otrzymanymi na podstawie modeli wartościami tych nieruchomości.



Rys. 1. Porównanie określonych wartości z cenami transakcyjnymi. *Źródło:* opracowanie własne.

Z powyższego rysunku widocznym jest, że określone wartości nieruchomości proponowanymi metodami analizy statystycznej rynku są bardzo zbliżone do cen transakcyjnych, co świadczy o poprawności tych modeli i ich przydatności dla potrzeb wyceny metodą analizy statystycznej rynku.

4. Podsumowanie

Z przeprowadzonych badań wynika, że proponowane w artykule modele matematyczne wielowymiarowej analizy porównawczej (MWAP), zmodyfikowanej metody porównywania parami (MPP) i metody liniowej regresji wielorakiej (MLRW) mogą znaleźć zastosowanie przy wycenie nieruchomości metodą analizy statystycznej rynku.

Zmodyfikowana metoda korygowania ceny średniej (MKCS) nie może być wykorzystana w metodzie analizy statystycznej rynku, z uwagi na wymagany opis parametrów jedynie w stosunku do nieruchomości charakteryzujących się

skrajnymi cenami.

Postać analityczna wszystkich przedstawionych w artykule modeli jest stosunkowa prosta, łatwa w interpretacji i pozwala na określania wartości nieruchomości z dokładnością satysfakcjonująca adresatów wyceny. Estymacja parametrów tych modeli nie jest trudna i może być wykonana przy wykorzystaniu prostych pakietów statystycznych.

Najdokładniejsze wyniki określenia wartości nieruchomości uzyskano w modelu (MLRW), nieco gorsze zbliżone do siebie, wyniki dokładności uzyskano w modelach (MWAP) i (MPP).

Przeprowadzone badania potwierdzają przydatność metod wielowymiarowej analizy porównawczej, metody liniowej regresji wielorakiej oraz zmodyfikowanej metody porównywania parami dla potrzeb określania wartości nieruchomości metodą analizy statystycznej rynku.

5. Literatura

ADAMCZEWSKI Z. 2006. *Elementy modelowania matematycznego w wycenie nieruchomości. Podejście porównawcze.* Oficyna Wydawnicza PW. Warszawa.

CZAJA J. 2001. *Metody szacowania wartości rynkowej i katastralnej nieruchomości*. KOMP-SYSTEM. Kraków.

DRAPER N. R., SMITH H. 1973. *Analiza regresji stosowana.* PWN. Warszawa.

SAWIŁOW E. 2004. *Analiza metod ustalania wpływy atrybutów na wartość nieruchomości*. Studia i Materiały Towarzystwa Naukowego Nieruchomości, vol. 12, nr 1. Olsztyn.

SAWIŁOW E. 2004. *Zmodyfikowana metoda powszechnej taksacji nieruchomości*. Materiały Konferencji Naukowo-Technicznej „Procedury prawne, organizacyjne i technologiczne powszechnej taksacji nieruchomości". Olsztyn.

SAWIŁOW E. 2008. *Analiza metod wyceny nieruchomości w podejściu porównawczym*. Przegląd Geodezyjny nr 2.

SAWIŁOW E. 2009. *Zastosowanie metod wielowymiarowej analizy porównawczej dla potrzeb ustalania wartości katastralnych*. Studia i Materiały Towarzystwa Naukowego Nieruchomości, vol. 17, nr 1. Olsztyn.

TELEGA T., BOJAR Z., ADAMCZEWSKI Z. 2002. *Wytyczne przeprowadzenia powszechnej taksacji nieruchomości,* Przegląd Geodezyjny nr 6.

ŹRÓBEK S., ŹRÓBEK R., 1993, *Zastosowanie regresji wielokrotnej w szacowaniu nieruchomości.* ACTA ACADEMIAE AGRICULTURAE AC TECHNICAE OLSTENENSIS, GEODAESIA ET RURIS REGULATIO 24. ART. Olsztyn.

THE PROBLEMS OF QUALIFYING THE VALUE OF REAL ESTATE WITH THE METHOD OF THE STATISTICAL ANALYSIS OF THE MARKET

Edward Sawiłow
Department of the Spatial Economy
Wroclaw University of Environmental and Life Sciences
e-mail: e.sawilow@wp.pl

Key words: *real estate, the model, the statistical analysis.*

Abstract

The pricing real estate is based on the analysis of dependences happening among the gathering the guild of describing real estate in the multidimensional space, and with their prices. The use of methods of the statistical analysis of the market seems to be as with the most proper method of qualifying of the value real estate, in the case of the large and from the local property market. In the article one described methods of graduating of qualitative features. One proposed several simple, in respect of the analytic form, models liable to underlie to the qualification of the value real estate with the method of the statistical analysis of the market.

IDENTIFICATION OF NONSTATIONARITY TYPE IN TIME SERIES OF LAND PROPERTY PRICES IN POLAND

Radosław Wiśniewski
Department of Land Management and Regional Development
University of Warmia and Mazury in Olsztyn
e-mail: danrad@uwm.edu.pl

Key words: *nonstationary time series of land property prices, nonstationarity type, unit root test*

Abstract

A stochastic process is nonstationary if its probability distribution function varies over time. Nonstationarity in time series of land property prices is considered to pose a serious problem in statistical analysis. Yet the observed nonstationarity is not homogeneous. The time series of land property prices may be marked by nonstationarity in average or nonstationarity in variance. The identified nonstationarity may be eliminated if its type is known. This paper presents the results of a study to identify the nonstationarity type in time series of land property prices in Poland.

1. Identification of nonstationarity type in time series of property prices

1.1. Data

The descriptive characteristics of the gathered observation sets are presented in Table 1. They comprise the time series of land property prices quoted on three real estate markets in Poland. The names of variables (property prices) which will be investigated in this paper are indicated in column no. 2 of Table 1. Figure 1 presents a sample price time series. Figure 2 shows the time series of price increments.

Table 1

Descriptive characteristics of particular research areas

Research Object*	Variable	N. of obs.	Mean	Median	Min	Max	Stand. Dev.	Skewness	Kurtosis
			[zł]	[zł]	[zł]	[zł]			
1	2	3	4	5	6	7	8	9	10
Gr_Ol	C_gr_Ol	1104	56.49	50.00	2.64	208.38	35.54	0.886	0.986

* Data source: WIŚNIEWSKI (2007).

Source: own calculations.

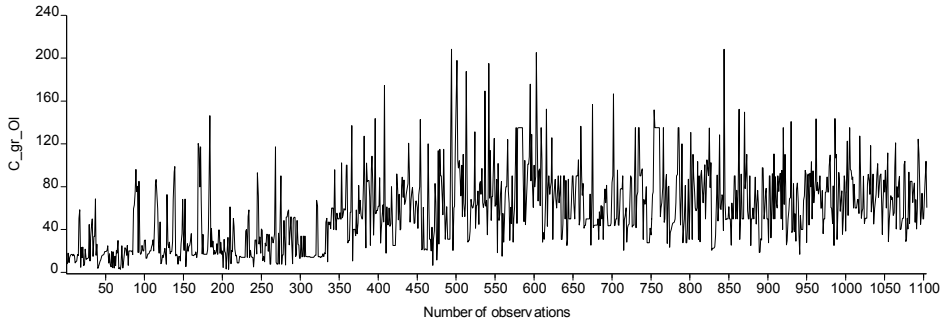


Fig. 1. Price time series for object *Gr_Ol*. *Source:* own calculations.

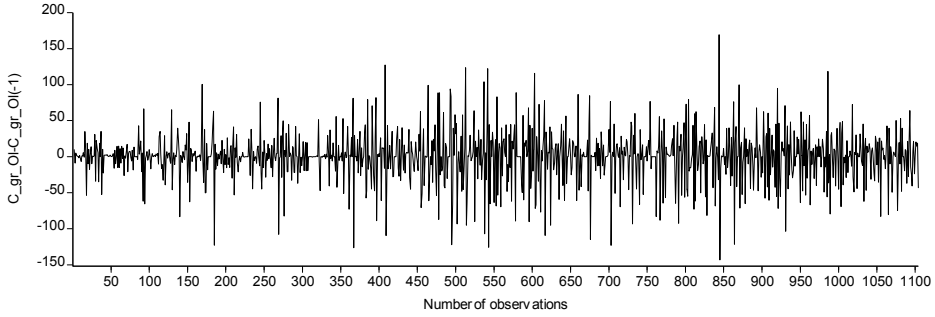


Fig. 2. Time series of price increments for object *Gr_Ol*. *Source:* own calculations.

The process of nonstationarity identification began with an analysis of Figure 1 and Figure 2 data and the development of a chart indicating the average price and standard deviation over time for object *Gr_Ol* (Fig. 3). It can be initially presumed that the time series of prices is nonstationary because both the average price and standard deviation change over time. An analysis of the above charts is indicative of a certain trend (Fig. 1) and fluctuations of various frequency. The type of nonstationarity in the time series of land property prices is difficult to ascertain beyond any doubt. Due to the above, the nonstationarity type in time series of property prices was identified with the use of the following methods:

1) comparison of the autocorrelation function (ACF) and the partial autocorrelation function (PACF),
2) specification of Ljung-Box Q-statistics values,
3) spectral density function,
4) unit root tests.

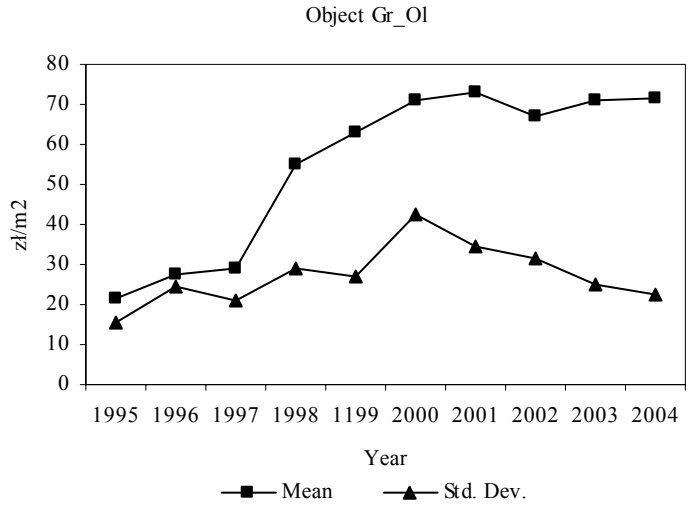


Fig. 3. Average price and standard deviation in each year for object *Gr_Ol*. *Source:* own calculations.

1.2. ACF, PACF, Ljung-Box Q-statistics

The autocorrelation function τ_k of time series of prices *C* with lag *k* is estimated as:

$$\tau_k = \frac{\sum_{t=k+1}^{T}\left(C_t - \bar{C}\right)\left(C_{t-k} - \bar{C}\right)}{\sum_{t=1}^{T}\left(C_t - \bar{C}\right)^2} \tag{1.1}$$

where $\bar{C}$ is the average of time series *C*, *T* – is number of realisations of the time series. If τ_k is equal to or approximates 0, then particular time series realisations are not autocorrelated. In practice, an absence of correlation is assumed when $|\tau_k| < 0{,}1$.

PACF ϕ_k in time series of prices *C* with lag *k* may be estimated recursively as (BOX, JENKINS 1976):

$$\phi_k = \begin{cases} \tau_1 & dla\ k = 1 \\ \tau_k - \sum_{j=1}^{k-1}\phi_{k-1,j}\tau_{k-j} & dla\ k > 1 \\ 1 - \sum_{j=1}^{k-1}\phi_{k-1,j}\tau_{k-j} & dla\ k > 1 \end{cases} \tag{1.2}$$

where: τ_k is an autocorrelation function for lag *k*, and $\phi_{k,j} = \phi_{k-1,j} - \phi_k \phi_{k-1,k-j}$. The partial autocorrelation function with lag *k* is a coefficient of regression C_{t-k} when C_t is regarded as a constant. A partial autocorrelation is a correlation of series with the same series displaced by a given lag of *k*-observations when the effect of correlations for all displacements from 1 to *k-1* has been eliminated. If ACF value is equal to 1, autoregression is less than *k*, and then the value of PACF with lag *k* will approximate 0.

Table 1

Autocorrelation function (ACF), partial autocorrelation function (PACF) and Q_{LB} values of Ljung-Box statistics of the time series of prices (C_t), first differences (ΔC_t) and residuals from linear trend $(C_t - T_L)$ for object *Gr_Ol*

k	C_t				ΔC_t				C_t-T_L			
	ACF	*PACF*	Q_{LB}	*p*	*ACF*	*PACF*	Q_{LB}	*p*	*ACF*	*PACF*	Q_{LB}	*p*
1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*
1	0.510	0.510	287.5	0.000	-0.386	-0.386	164.53	0.000	0.342	0.342	129.2	0.000
2	0.397	0.186	462.2	0.000	-0.107	-0.301	177.26	0.000	0.192	0.085	170.0	0.000
3	0.389	0.182	630.3	0.000	0.020	-0.194	177.70	0.000	0.183	0.107	207.2	0.000
4	0.362	0.112	776.1	0.000	0.003	-0.138	177.71	0.000	0.148	0.053	231.3	0.000
5	0.332	0.075	898.9	0.000	-0.045	-0.160	179.98	0.000	0.108	0.022	244.2	0.000
6	0.347	0.110	1032.6	0.000	0.030	-0.107	180.95	0.000	0.127	0.064	262.3	0.000
7	0.332	0.066	1155.4	0.000	0.050	-0.019	183.72	0.000	0.109	0.027	275.4	0.000
8	0.269	-0.018	1235.8	0.000	-0.098	-0.118	194.36	0.000	0.024	-0.057	276.0	0.000
9	0.301	0.084	1336.7	0.000	0.013	-0.108	194.55	0.000	0.067	0.046	281.0	0.000
10	0.320	0.080	1451.1	0.000	0.041	-0.065	196.40	0.000	0.094	0.047	290.9	0.000

Source: own calculations.

With the assumed lag *k*, Q-statistics is a statistical test of the null hypothesis postulating that price time series are not autocorrelated until the attainment of level *k*, estimated with the use of the below formula:

$$Q_{LB} = T(T+2)\sum_{j=1}^{k}\frac{\tau_j^2}{T-j} \tag{1.3}$$

where: τ_j is the j^{th} autocorrelation, *T* is the number of price time series realisations. Ljung-Box Q-statistics is often used to verify whether the time series is an example of white noise, and in this context, it may be deployed to validate random walk

processes and constitutes a statistical method for evaluating the efficiency of the real property market.

ACF, PACF and Ljung-Box Q-statistics tests were performed for all studied objects. The results for object *Gr_Ol* are presented in Table 1. Differences in ACF and PACF for object *Gr_Ol* are presented in Figure 4.

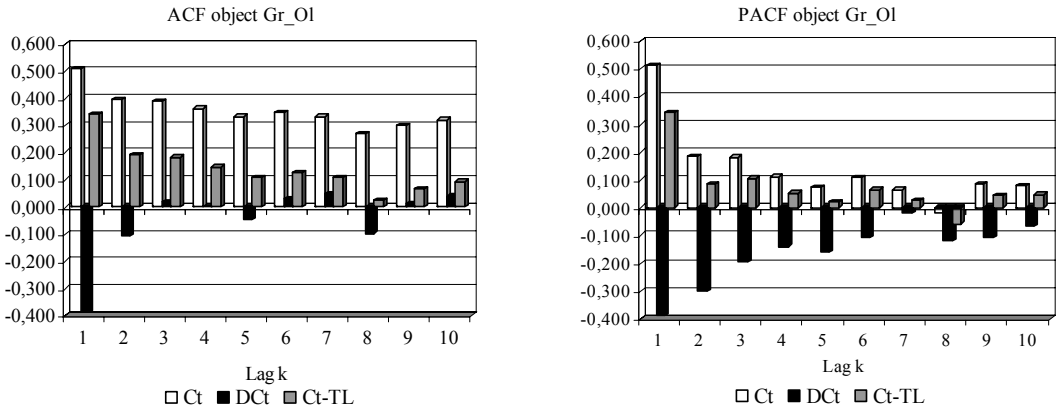


Fig. 4. Autocorrelation function (ACF), partial autocorrelation function (PACF) of the time series of prices (C_t), first differences (ΔC_t) and residuals from linear trend $(C_t - T_L)$ for object *Gr_Ol*. *Source:* own calculations.

Figure 4 indicates that the value of the autocorrelation function (ACF) of prices (C_t) for a single lag (object *Gr_Ol*) equals 0.5, while it decays exponentially for successive lags (object *Gr_Ol*) and decays rapidly for the remaining objects. The calculation of first differences (ΔC_t) in this case results in a first-order negative autocorrelation of –0.39 (cf. ACF in Fig. 4), and this value decays for successive lags. In the partial autocorrelation function (PACF) (cf. PACF in Fig. 4), this process is expressed by a clearly pronounced negative autocorrelation for several first lags. The above indicates that a real data-generating model is a model with a linear trend and a weak first-order autoregression effect, implying that we are dealing with nonstationarity in average. Nonstationarity should be eliminated if time variable *t* is introduced into the model. As regards object *Gr_Ol*, the introduction of variable *t* significantly reduces the autocorrelation effect in the price series (cf. $C_t - T_L$ in Fig. 4), but does not eliminate it entirely. It is practically not reported for lag k = 10. The above validates the assumption that a price series is nonstationary in average with an autoregression effect.

The assumption that a generating model is a model with a linear trend and first-order autoregression can be validated by analysing the curve of the spectral density function. The spectral density function of residuals from the linear trend for object *Gr_Ol*, presented in a double logarithmic scale, does not resemble

a standard spectrum curve for such processes (Fig. 5) (cf. PIŁATOWSKA 2003). The curves for the remaining objects are similar. They suggest (especially in the second part) that we are dealing with a model with a linear trend. The discrepancies between the results of an autocorrelation analysis and a spectrum analysis suggest the presence of additional solution-creating factors. Their identification is difficult because it should involve statistical analysis methods which account for higher-order correlations.

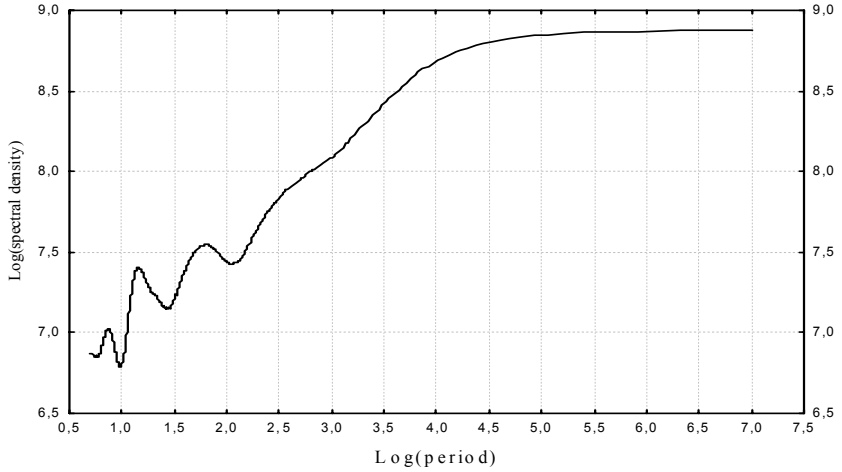


Fig. 5. Spectral density function of residuals from linear trend $\left(C_t - T_L\right)$ for object *Gr_Ol*. *Source:* own calculations.

The obtained results indicate that the analysed time series of land property prices are probably nonstationary series. Time series are marked by nonstationarity in average with an autoregressive element. The presence of unit roots was tested to validate the above dependencies.

1.3. Integration process testing – unit root tests

Nonstationary processes in average (processes with a deterministic trend) and nonstationary processes in variance (processes with a stochastic trend) have different characteristics and suggest different dynamics of economic variables. Therefore, the given process type has to be identified. Below is an overview of an analysis to determine the type of nonstationarity based on unit root tests.

Autoregressive unit root tests propose parameter ρ in the following model (BHARGAVA 1986; PIŁATOWSKA 2003):

$$C_t = d_t + u_t \tag{1.4}$$

$$u_t = \rho u_{t-1} + \varepsilon_t \tag{1.5}$$

where: d_t is a deterministic component, ε_t is white noise. $d_t = 0$ implies the absence of a deterministic component, $d_t = \alpha$ implies the presence of a constant,

$d_t = \alpha + \beta t$ implies the presence of a linear trend.

Assuming that $H_0 : \rho = 1$ i $d_t = 0$, process C_t is a random walk without drift. If $|\rho| < 0$, then the process is stationary around an average (cf. BHARGAVA 1986). If $d_t = \alpha$, where $H_0 : \rho = 1$, process C_t is a random walk without drift, when $|\rho| < 0$, the process is stationary around a non-zero average. If $d_t = \alpha + \beta t$, on the assumption that $H_0 : \rho = 1$, process C_t is a random walk with drift, i.e. $\Delta C_t = \beta + \varepsilon_t$, and if $|\rho| < 0$, process C_t has a linear trend with coefficient $\beta(1-\rho)$ (cf. PIŁATOWSKA 2003).

Unit root tests verify the hypothesis on the type of nonstationarity in time series. If the null hypothesis $H_0 : \rho = 1$ is rejected, the process belongs to a class of process with stationary deviations from the deterministic trend, and a model of such a process should be built for price levels. If the null hypothesis is not rejected, we are dealing with a process with stationary increments whose model should be built for respective increments.

The Augmented Dickey-Fuller test (ADF) is one of the basic unit root tests. Dickey-Fuller t_{DF} statistics is a quotient of estimate $(\hat{\rho} - 1)$ and standard error of estimate $s(\hat{\rho})$ (PIŁATOWSKA 2003):

$$t_{DF} = \frac{(\hat{\rho} - 1)}{s(\hat{\rho})} \tag{1.6}$$

in the following model:

$$C_t = \alpha + (\hat{\rho} - 1) C_{t-1} + \sum_{i=1}^{p-1} \hat{\rho} \Delta C_{t-i} + \varepsilon_t \tag{1.7}$$

where: C_t is a process in the form of $C_t = \alpha + \sum_{i=1}^{p} \rho_i C_{t-i} + \varepsilon_t$, and p is the autoregression order.

PHILLIPS and PERRON (1988) modified the Dickey-Fuller test by accounting for a residual autocorrelation and adjusting standard DF statistics. To account for the residual autocorrelation effect, they proposed a procedure of modifying standard statistics after the estimation of model parameters, without the need to estimate additional parameters. The PHILLIPS-PERRON (PP) test (1988) is based on the below statistics:

$$t_{PP} = t_{DF} \left(\frac{s_0}{f_0} \right)^{1/2} - \frac{T(f_0 - s_0)(s(\hat{\rho}))}{2 f_0^{1/2} s} \tag{0.1}$$

where: s_0 is a consistent estimator of residual variance (calculated as $(T-k)s^2/T$, where k is the number of regressors (lags)), f_0 is the estimator of the spectral density function at zero frequency, T is the number of observations, $s(\hat{\rho})$ is the standard error of estimate, and s is the standard error of a regression test.

Table 2

Results of verification of the hypothesis $H_0 : C_t \sim I(1)$ with the first unit root for object *Gr_Ol*. Augmented Dickey-Fuller test (ADF) and Phillips-Perron test (PP)

	C_gr_Ol			ln(C_gr_Ol)		
k	Model with constant	Model with constant and trend	Model without constant or trend	Model with constant	Model with constant and trend	Model without constant or trend
1	2	3	4	5	6	7
			ADF			
0	-6.41***	-8.48***	-2.38**	-5.56***	-8.05***	-0.61
1	-10.66***	-6.39***	-18.47***	-13.28***	-8.20***	-20.38***
2	-8.75***	-5.46***	-13.71***	-9.16***	-5.66***	-13.11***
3	-6.14***	-3.65**	-9.45***	-5.68***	-3.10	-8.29***
4	-4.71***	-2.74	-7.13***	-3.65***	-1.61	-5.58***
5	-4.33***	-2.81	-6.08***	-3.17**	-1.54	-4.62***
6	-2.25	-1.27	-3.34***	-3.06**	-1.94	-4.01***
$d_t = \alpha$	5.94	5.17	-	5.53	7.84	-
$d_t = \alpha + \beta t$	-	5.45	-	-	5.74	-
$adjR^2$	0.32	0.33	0.30	0.30	0.32	0.28
			PP			
0	-18.93***	-22.35***	-9.04***	-16.08***	-20.72***	-2.78***
$d_t = \alpha$	16.05	9.12	-	15.78	19.22	-
$d_t = \alpha + \beta t$	-	10.33	-	-	11.77	-
$adjR^2$	0.24	0.31	0.07	0.19	0.28	0.01
	Critical test values for a given significance level					
1%	-3.44	-3.97	-2.57	-3.44	-3.97	-2.57
5%	-2.86	-3.41	-1.94	-2.86	-3.41	-1.94
10%	-2.57	-3.13	-1.62	-2.57	-3.13	-1.62

***, **, * - rejection of H_0 at a significance level of 1%, 5% and 10%, respectively.

Cases when H_0 was not rejected are indicated in bold, $adjR^2$ - corrected model fitting coefficient. *Source:* own calculations.

PIŁATOWSKA (2003) also notes that an adequate number of lags in the ADF test should be selected as it has been observed that the power and the number of rejections of the true null hypothesis $H_0 : \rho = 1$ is determined by the number of first increment lags (cf. AGIAKOGLOU, NEWBOLD 1992). NG and PERRON (1995) postulated that the number of lags in the ADF test should be determined by: identifying the constant number of lags *k* in an arbitrary manner subject to the number of observations, identifying the constant number of lags *k* as a function of the number of observations, e.g. $k = 12(T/100)^{1/4}$ (SCHWERT 1989) or $k = T^{1/4}$

(DIEBOLD, NERLOVE 1990), relying on Akaike information criteria (AIC) and Schwarz Bayesian information criteria (BIC), and selecting k in line with the top-down approach or the bottom-up approach.

Table 2 presents the results of first unit root verification with the use of ADF and PP tests in object *Gr_Ol*, with different assumptions concerning price series transformations (finding the logarithm – columns 5, 6 and 7), the presence of a constant in the model (columns 2 and 5), the presence of a constant and a trend in the model (columns 3 and 6) and a model without a constant or a trend (columns 4 and 7). Table 3 presents the results of second unit root verification for the same set of assumptions which apply to Table 2. The number of lags in the ADF test was selected based on dependency $k = T^{1/4}$ (DIEBOLD, NERLOVE 1990).

The results presented in Table 2 point to a lack of uniqueness as regards stationarity in average. The ADF test performed on original (non-logarithmed) data for the model with a constant and the model without a trend or a constant indicates that unit root hypothesis H_0 is rejected in most cases, thus assuming the occurrence of nonstationarity in average. The above is validated by the PP test. An analysis of logarithmed property price series leads to similar conclusions. As regards the model with a constant and a linear trend, nonstationarity in variance to the second lag is rejected by the ADF test. Test values for the constant in the model with a constant (columns 2 and 5), and for the constant and the trend (columns 3 and 6) are significant. The above is validated by the PP test.

The indicated lack of uniqueness prompted the need to calculate the second unit root. Calculation results are presented in Table 3.

Tests to determine the degree of integration, i.e. calculation of the second unit root, brought unexpected results. The values presented in Table 3 for both tests are positive and high in most cases. It should be noted that the value of $adjR^2$ statistics in Table 3 is much higher (nearly two-fold) than in Table 2. Test values indicated in Table 3 point to overdifferencing, i.e. the use of the subtraction operator too many times (cf. CHAREMZA, DEADMAN 1997).

The KPSS stationarity test (KWIATKOWSKI et al. 1992) was performed to validate the process of overdifferencing.

Kwiatkowski-Phillips-Schmidt-Shin (1992) developed a test with the following output model:

$$Y_t = d_t + \xi_t + \varepsilon_t \tag{1.9}$$

$$\xi_t = \xi_{t-1} + \eta_t,\ \eta_t \sim iid(0, \sigma_\eta^2) \tag{1.10}$$

where: ε_t is a stationary process, ξ_t is a random walk, and $d_t = \beta_0$ lub $d_t = \beta_0 + \beta_1 t$. The null hypothesis postulating process stationarity is formulated as follows: $H_0 : \sigma_\eta^2 = 0$ or ξ_t is a constant. The null hypothesis is verified with the use of the following test (NABEYA, TANAKA 1988):

Table 3

Results of verification of the hypothesis $H_0 : C_t \sim I(2)$ with the second unit root for object *Gr_Ol*. Augmented Dickey-Fuller test (ADF) and Phillips-Perron test (PP) [1]

	ΔC_gr_Ol			ln(ΔC_gr_Ol)		
K	Model with constant	Model with constant and trend	Model without constant or trend	Model with constant	Model with constant and trend	Model without constant or trend
1	2	3	4	5	6	7
			ADF			
0	-18.33***	-18.32***	-18.33***	-18.84***	-18.83***	-18.85***
1	9.83	9.83	9.84	9.85	9.85	9.85
2	7.56	7.56	7.56	7.69	7.69	7.69
3	5.84	5.84	5.84	6.31	6.31	6.31
4	4.31	4.31	4.32	5.44	5.44	5.44
5	2.39	2.39	2.39	4.50	4.50	4.50
6	0.63	0.63	0.63	2.80	2.80	2.80
$d_t = \alpha$	0.19	0.22	-	0.26	0.29	-
$d_t = \alpha + \beta t$	-	-0.14	-	-	-0.18	-
$adjR^2$	0.74	0.74	0.74	0.75	0.75	0.75
			PP			
0	-49.81***	-49.78***	-49.83***	-51.36***	-51.34***	-51.39***
$d_t = \alpha$	0.07	0.09	-	0.11	0.13	-
$d_t = \alpha + \beta t$	-	-0.06	-	-	-0.08	-
$adjR^2$	0.69	0.69	0.69	0.71	0.71	0.71

[1] – critical test value as in Table 2.

***, **, * - rejection of H_0 at a significance level of 1%, 5% and 10%, respectively. Cases when H_0 was not rejected are indicated in bold, $adjR^2$ - corrected model fitting coefficient. *Source:* own calculations.

$$LM = \frac{\sum_{t=1}^{T} S_t}{\hat{\sigma}_e^2} \tag{1.11}$$

where: e_t - residuals from model y_t in reference to the time constant and variable t, $\hat{\sigma}_e^2$ - residual variance from the model, S_t - partial sum of residual e_t

$$S_t = \sum_{j=1}^{t} e_j \quad t = 1, 2, \ldots, T . \tag{1.12}$$

For white noise residuals, estimator $\hat{\sigma}_e^2$ in LM test approaches σ^2, where

$$\sigma^2 = \lim_{T \to \infty} T^{-1} E\left(S_T^2\right). \tag{1.13}$$

If residuals are free of white noise, the consistent estimator σ^2 is $\tilde{s}_{Tl}^2$:

$$\tilde{s}_{Tl}^2 = \frac{1}{T}\sum_{t=1}^{T} e_t^2 + \frac{2}{T}\sum_{j=1}^{l} \omega_{jl} \sum_{t=j+1}^{T} e_t e_{t-j} \ , \qquad (1.14)$$

where $\omega_{jl} = 1 - j/(l+1)$ is an optional weighting function that corresponds to the choice of a special window, e.g. Bartlett window.

Table 4

Results of verification of the hypothesis $H_0 : C_t$ is stationary for object *Gr_Ol*. KPSS test

	C_gr_Ol		ln(C_gr_Ol)	
	Model with constant	Model with constant and trend	Model with constant	Model with constant and trend
1	*2*	*3*	*5*	*6*
LM-Statistic*				
	7.55	1.09	9.05	1.36
Critical LM test values for a given significance level*				
1%	0.74	0.22	0.74	0.22
5%	0.46	0.15	0.46	0.15
10%	0.35	0.12	0.35	0.12
t-Statistic**				
$d_t = \alpha$	52.82 *(0.00)*	14.75 *(0.00)*	165.07 *(0.00)*	80.64 *(0.00)*
$d_t = \alpha + \beta t$	-	17.42 *(0.00)*	-	23.21 *(0.00)*

* - Kwiatkowski-Phillips-Schmidt-Shin (1992, Table 1)
** p-value for Student's t-statistics at a significance level of 0.05 is given in brackets. Cases when H_0 was rejected are indicated in bold. *Source:* own calculations.

Test results (Tables 4 and 5) validate the process of overdifferencing. Table 4 presents calculations indicating that the null hypothesis which postulates that process C_t is stationary should be rejected. This implies that the price time series is nonstationary which, following the rejection of the first unit root hypothesis (Table 2), points to nonstationarity in average. An analysis of data in Table 5 (KPSS test to determine increments in land property prices) shows that the time series of first price increments is stationarity. Yet statistical values are nearly equal to zero. The above clearly indicates that the use of the subtraction operator leads to overdifferencing.

2. Conclusions

The results of this study indicate that the analysed time series of land property prices are roughly nonstationary in average, and increment calculation results in overfitting. ADF, PP and KPSS tests have shown that the time series of land property prices is not integrated, implying that it may not be transformed into

a stationary time series through increment calculation. It should, however, be noted that the inability to determine integration in the time series of property prices could be due to the low power of the applied tests (cf. PHILLIPS, PERRON 1988; DIEBOLD, NERLOVE 1990; CHAREMZA, DEADMAN 1997; PIŁATOWSKA 1997).

Table 5

Results of verification of the hypothesis $H_0 : \Delta C_t$ is stationary for object *Gr_Ol*.
KPSS test

	ΔC_gr_Ol		ln(ΔC_gr_Ol)	
	Model with constant	Model with constant and trend	Model with constant	Model with constant and trend
1	*2*	*3*	*5*	*6*
	LM-Statistic*			
	0.01**	0.00**	0.01**	0.00**
	Critical LM test values for a given significance level*			
1%	0.74	0.22	0.74	0.22
5%	0.46	0.15	0.46	0.15
10%	0.35	0.12	0.35	0.12
	t-Statistic***			
$d_t = \alpha$	0.05 *(0.96)*	0.08 *(0.94)*	0.09 *(0.93)*	0.14 *(0.89)*
$d_t = \alpha + \beta t$	-	-0.06 *(0.95)*	-	-0.11 *(0.91)*

* - Kwiatkowski-Phillips-Schmidt-Shin (1992, Table 1)

** - Cases when H_0 was accepted are marked.

*** p-value for Student's t-statistics at a significance level of 0.05 is given in brackets. *Source:* own calculations.

The identified nonstationarity in average should be eliminated by subtracting trend P_t, i.e. by performing operation $X_t - P_t = \eta_t$, where P_t is a trend component isolated with the use of a given operator. The isolation of component P_t P_t is also significant for another reason – the isolated component can be introduced to the model. If components P_t are isolated for various processes, the dependency between those components, i.e. components describing changes in average values, can be investigated. Yet to date, a universal and fully objective method for selecting the trend form has not been developed. The following methods can be proposed: the growth curve method, multinomial function of time variable, investigation of absolute and relative increments, trigonometric polynomial, various filters (moving average filters, recursive filters, exponential smoothing filters). In general, the simplest functions are recommended, provided that they are effective.

Data which are characterised by trends pose a serious problem in empirical econometrics. They lead to apparent regressions, uninterpretable values of Student's t-statistics and other statistics, overestimated measures of goodness of fit,

thus distorting the evaluation of the produced results. The use of modern estimation methods, such as artificial neural networks and modular neural networks, should not eliminate the attempts to investigate the relationships governing such data. The identification of hidden relations in data contributes valuable information on the structure of the built forecasting model. But above all, it enables researchers to arrive at correct conclusions as regards the forecast value of variable in the future.

An analysis of the internal structure of time series of property prices is one of the main stages of an analysis of the gathered data. It aims to determine the type of relationships found in the analysed data. As regards classical statistical models of property prices (e.g. regressive) based on nonstationary series, the type of nonstationarity has to be correctly identified at the first stage for the appropriate parameter estimation procedures to be applied at the successive stage. A diagnosis of time series of property prices which are used to, for example, forecast property value in the context of a given trend, and an analysis of variable integration processes support the forecasting process. The above particularly applies to situations in which the forecasting strategy is based on the process of simulating the property market model, on the assumption that variables will take up the values of a given probability distribution function. The use of various prognostic methods based on stationary series is always easier, and the obtained results are marked by lower uncertainty due to the manner in which data are processed by a prognostic model.

The application of classical methods for forecasting property value under such circumstances produces apparent regressions which could adversely affect the quality of the forecast. Statistical analyses for forecasting property value should be conducted with the use of tools that account for nonstationarity in the course of the estimation process.

The identification of nonstationarity type is of paramount importance for the selection of the right method of forecasting property value. For the needs of the forecasting process, short- and long-term relationships have to be determined. Those analyses are significant because, on the one hand, they identify autocorrelation dependencies in data and validate property value forecasts on the side of data-generating processes, while on the other hand, the identification of the type of the dependency, and the power and stability of its effect, supports the correct selection of prognostic process parameters (e.g. the prognostic horizon).

3. References

AGIAKOGLOU C., NEWBOLD P. 1992. *Empirical Evidence on Dickey-Fuller Type Tests*. Journal of Time Series Analysis 13 s.471-483.

BOX G. E. P., JENKINS G. M. 1983. *Analiza szeregów czasowych. Prognozowanie i sterowanie*. PWN. Warszawa.

CHAREMZA W. W. DEADMAN D. F. 1997. *Nowa ekonometria*. Polskie Wydawnictwo Ekonomiczne, Warszawa.

DIEBOLD F. X., NERLOVE M. 1990. *Unit Roots in Economic Time Series: a Selective Survey*, w: G. F. Rhodes, T. B. Fomby (red.). Advances in Econometrics, vol. 8, Greenwich, CT: JAT Press, s. 3-69.

GRANGER C. W. J., NEWBOLD P. 1986. *Forecasting economic time series*. Academic Press, INC., New York.

HELLWIG Z. 1978. *Elementy rachunku prawdopodobieństwa i statystyki matematycznej*. PWN. Warszawa.

KILDISZEW G., FRENKEL A. 1976. *Analiza szeregów czasowych*. PWE. Warszawa.

KWIATKOWSKI D., PHILLIPS P. C. B., SCHMIDT P. SHIN Y. 1992. *Testing the Null Hypothesis of Stationary against the Alternative of a Unit Root*. Journal of Econometrics 54, 159-178

MANTEGENA R. S., STANLEY H. E. 2001. *Ekonofizyka. Wprowadzenie*. Wydawnictwo Naukowe PWN, Warszawa.

MILO W. 1983. *Analiza szeregów czasowych*. Uniwersytet Łódzki. Łódź.

NABEYA S., TANAKA K. 1988. *Asymptotic Theory of a Test for the Constancy of Regression*. Journal of Econometrics, 46, p.247-271.

NG S., PERRON P. 1995. *Unit Root Tests in ARMA Models with Data-Dependent Methods for the Selection of the Truncation Lag*. Journal of the American Statistical Association 90 s. 268-281.

PIŁATOWSKA M. 2003. Modelling non-stationary economic processes. Methodological study. Wydawnictwo Uniwersytetu Mikołaja Kopernika, Toruń, Poland.

PHILLIPS P. C. B. 1986. *Understanding Spurious Regressions in Econometrics*. Journal of Econometrics 33, s. 311-340.

PHILLIPS P. C. B., PERRON P. 1988. *Testing for Unit Root in Time Series Regression*. Biometrica 75, s. 335-346.

SCHWERT G. W. 1989. *Tests for Unit Roots: A Monte Carlo Investigation*. Journal of Business and Economic Statistics 7, s. 147-159.

WIŚNIEWSKI R. 2007. *Wielowymiarowe prognozowanie wartości nieruchomości*. Seria: Rozprawy i monografie. Wydawnictwo UWM w Olsztynie.

ZELIAŚ A. 1997. *Teoria prognozy*. PWE, Warszawa.

ZIELIŃSKI Z.1979. *Metody analizy dynamiki i rytmiczności zjawisk gospodarczych*. PWN. Warszawa.

IDENTYFIKACJA TYPU NIESTACJONARNOŚCI W SZEREGACH CZASOWYCH CEN NIERUCHOMOŚCI GRUNTOWYCH W POLSCE

Radosław Wiśniewski
Katedra Gospodarki Nieruchomościami i Rozwoju Regionalnego
Uniwersytet Warmińsko-Mazurski w Olsztynie
e-mail: danrad@uwm.edu.p

Słowa kluczowe: *testy niestacjonarności, typy niestacjonarności, testy pierwiastków jednostkowych*

Streszczenie

Proces stochastyczny jest niestacjonarny, jeśli jego funkcja rozkładu prawdopodobieństwa jest zmienna w czasie. Występowanie niestacjonarności w szeregach czasowych cen nieruchomości gruntowych uważa się za poważny problem w analizach statystycznych. Obserwowana niestacjonarność nie jest jednak jednorodna. W szeregach czasowych cen nieruchomości gruntowych występować mogą: niestacjonarność w średniej, niestacjonarność w wariacji. Zidentyfikowaną niestacjonarność można usunąć trzeba jednak znać jej typ. W pracy przedstawiono badania mające na celu identyfikację typu niestacjonarności szeregów czasowych cen nieruchomości gruntowych na polskim rynku nieruchomości.

METODY STATYSTYCZNE W WYCENIE NIERUCHOMOŚCI

Marcin Ligas
Katedra Geomatyki
Akademia Górniczo – Hutnicza im. Stanisława Staszica w Krakowie
e-mail: marcin.ligas@agh.edu.pl

Słowa kluczowe: *bootstrap, kroswalidacja, estymatory odporne, transformacje zmiennych*

Streszczenie

Artykuł ma na celu odświeżenie zapomnianych a może nawet często nieznanych metod statystycznych, które powinny znaleźć swoje miejsce w warsztacie rzeczoznawcy - statystyka. Dotyczą one zarówno małych prób, z jakimi mamy do czynienia w wycenie pojedynczych nieruchomości jak również masowej wyceny. Przedstawiona została metoda symulacyjna bootstrap i jej przewaga nad stosowaniem klasycznych metod opartych o mocne założenia, „unormalniające" transformacje zmiennych jak transformacje Box'a – Cox'a, NST (Normal Score Transform) i ich zastosowanie w metodach regresyjnych, metody „sprawdzania" modeli – metoda kroswalidacji, estymacja odporna. Niektóre z opisanych metod zostały opatrzone przykładem opartym na danych pochodzących z rynku nieruchomości.

1. Wprowadzenie

Niniejszy artykuł poświęcony jest metodom statystycznym, które mogą znaleźć trwałe miejsce w warsztacie rzeczoznawcy - statystyka. Niewielka objętość artykułu pozwala jedynie na ukazanie „mechaniki" tych metod i ich krótkiego opisu. Treść artykułu dotyczy bezpośrednio metody analizy statystycznej wyceny nieruchomości jednakże przez uniwersalność opisywanych metod mogą one z powodzeniem znaleźć miejsce we wszystkich metodach podejścia porównawczego jak i innych metodach, w których metody statystyczne przejawiają się w większym bądź mniejszym stopniu.

Przedstawione zostały „unormalniające" (oraz stabilizujące wariancję) transformacje zmiennych, mogące znaleźć zastosowanie wszędzie tam gdzie założenia stosowalności metod statystycznych wymagają normalności (bądź niewielkich odstępstw od normalności) a oryginalne dane ich nie spełniają. Opisano transformację Box'a i Cox'a oraz potężne narzędzie, jakim jest anamorfoza gaussowska wykorzystywana bardzo często w geostatystyce nieliniowej.

Przedstawiona również została pokrótce metoda Bootstrap (w literaturze polskiej czasem określana nazwą metody „sznurowadłowe") należąca do metod określanych mianem „computer intensive methods" choć przy obecnych mocach komputerów nie ma to już tak wielkiego znaczenia jak to było kiedyś. Twórca

Bootstrapu Efron (EFRON, TIBSHIRANI 1993 str. 11) definiuje go jako: opartą o komputery metodę przypisywania miar dokładności do szacownych parametrów (*The bootstrap is a computer – based method for assigning measures of accuracy to statistical estimates*). Minimum założeń statystycznych czyni z bootstrapu metodę godną uwagi a wykorzystywać ją można w metodach regresyjnych, budowaniu przedziałów ufności, testowaniu hipotez statystycznych.

Przedstawiono i porównano również metodę estymacji odpornej (LMS – Least Median of Squares) z klasyczną metodą najmniejszych kwadratów (OLS – Ordinary Least Squares) wrażliwą na przypadki odstające w danych oraz metodę kroswalidacji, czyli metodę szacowania rzeczywistych błędów predykcji.

2. Transformacje zmiennych

2.1. Transformacja Box'a Cox'a

Mechanizm transformacji potęgowej Box'a i Cox'a zostanie opisany pokrótce ze względu na fakt, iż można ją znaleźć w większości dobrych pozycji traktujących o ekonometrii oraz statystyce. Oryginalne zmienne x_i podlegają transformacji zgodnie z następującą formułą (RAWLINGS i inni 1998, BILODEAU 1999):

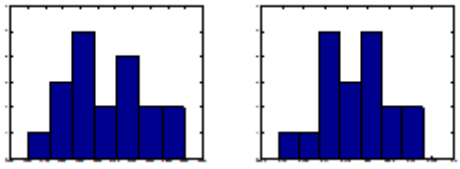
$$x_i(\lambda) = \begin{cases} \dfrac{x_i^{\lambda} - 1}{\lambda}, \lambda \neq 0 \\ \ln(x_i), \lambda = 0 \end{cases} \tag{1}$$

Najczęściej stosowaną metodą szacowania wartości parametru λ jest metoda oparta o maksymalizację logarytmicznej funkcji wiarygodności postaci:

$$f(x,\lambda) = -\frac{n}{2}\ln\left[\sum_{i=1}^{n}\frac{(x_i(\lambda)-\hat{x}(\lambda))^2}{n}\right] + (\lambda-1)\sum_{i=1}^{n}\ln(x_i) \tag{2}$$

gdzie:

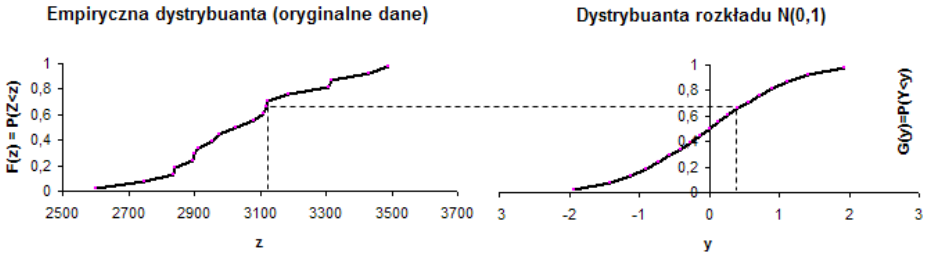
$$\hat{x}(\lambda) = \frac{1}{n}\sum_{i=1}^{n} x_i(\lambda)$$



Rys. 1. Transformacja Box'a i Cox'a, histogram oryginalnych zmiennych oraz histogram zmiennych po transformacji dla danych z Tabeli 1. *Źródło*: opracowanie własne.

2.2. Anamorfoza gaussowska

Anamorfoza gaussowska (gaussian anamorphosis, normal score transform) jest „potężnym" narzędziem „unormalniania" zmiennych. Narzędzie często używane w geostatystyce nieliniowej (RIVOIRARD 1990, WACKERNAGEL 2003, ORTIZ i inni). W wyniku transformacji otrzymujemy zmienne pochodzące ze standaryzowanego rozkładu normalnego $N(0,1)$. Idea transformacji „w przód", czyli otrzymanie zmiennych gaussowskich, jest bardzo prosta, budujemy empiryczną dystrybuantę z oryginalnych danych otrzymując w ten sposób wartości funkcji $F(z) = P(Z<z)$. Znając wartości funkcji $F(z)$ przechodzimy na wartości funkcji $G(y) = P(Y<y)$ oczywiście zachodzi równość $F(z) = G(y)$ i na tej podstawie dostajemy zmienną losową o rozkładzie normalnym $y = G^{-1}(F(z))$ (kwantyl rozkładu $N(0,1)$). Przejście z oryginalnych zmiennych na zmienne gaussowskie przedstawia rysunek 2.



Rys. 2. Schemat anamorfozy gaussowskiej (transformacja „w przód"). *Źródło*: opracowanie własne.

Trochę kłopotów nastręcza transformacja powrotna, trzeba znaleźć funkcję przekształcającą zmienne losowe gaussowskie na zmienne losowe z oryginalnego rozkładu, czyli: $z = F^{-1}(G(y))=\phi(z)$. Funkcja ϕ nazywa się funkcją anamorfozy. Empiryczna funkcja anamorfozy nie nadaje się do wykorzystania – nie jest bijekcją (odwzorowaniem wzajemnie jednoznacznym), dlatego też wykorzystuje się tutaj rozwinięcie w szereg wielomianów ortogonalnych Hermite'a, który przybliża kształt funkcji anamorfozy ϕ (oczywiście dopasowanie funkcji anamorfozy zależy od stopnia rozwinięcia).

Wielomiany ortogonalne Hermite'a definiowane są na podstawie formuły Rodrigues'a, czyli (np. WACKERNAGEL 2003, RIVOIRARD 1990):

$$H_n(y) = \frac{1}{\sqrt{n!}\cdot g(y)} \cdot \frac{d^n g(y)}{dy^n} \quad \forall n \geq 0 \tag{3}$$

gdzie: n jest stopniem wielomianu, $n!$ jest współczynnikiem normalizującym, y zmienną losowa pochodzącą ze standaryzowanego rozkładu normalnego a $g(y)$ funkcją gęstości standaryzowanego rozkładu normalnego, czyli:

$$g(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} \tag{4}$$

Tabela 1

Zmienne oryginalne „*z*", zmienne po transformacji „*y*" oraz 10 wielomianów Hermite'a

z=C	y	G(y)	0	1	2	3	4	5	6	7	8	9	10
2600,92	-1,938	0,061	1,000	1,938	1,949	0,598	-1,108	-1,495	-0,171	1,259	1,023	-0,526	-1,293
2748,37	-1,412	0,147	1,000	1,412	0,703	-0,580	-1,018	-0,124	0,858	0,573	-0,517	-0,783	0,140
2835,79	-1,119	0,213	1,000	1,119	0,178	-0,798	-0,601	0,413	0,738	-0,071	-0,718	-0,201	0,610
2840,2	-0,899	0,266	1,000	0,899	-0,136	-0,804	-0,244	0,621	0,451	-0,422	-0,556	0,231	0,593
2896,38	-0,716	0,309	1,000	0,716	-0,345	-0,727	0,038	0,663	0,159	-0,570	-0,293	0,468	0,384
2897,93	-0,555	0,342	1,000	0,555	-0,489	-0,610	0,254	0,609	-0,094	-0,583	-0,026	0,545	0,121
2912,97	-0,407	0,367	1,000	0,407	-0,590	-0,471	0,415	0,497	-0,296	-0,506	0,204	0,504	-0,129
2953,72	-0,267	0,385	1,000	0,267	-0,657	-0,319	0,526	0,348	-0,442	-0,367	0,379	0,380	-0,328
2977,49	-0,132	0,395	1,000	0,132	-0,695	-0,161	0,591	0,179	-0,530	-0,192	0,487	0,202	-0,453
3024,91	0	0,399	1,000	0,000	-0,707	0,000	0,612	0,000	-0,559	0,000	0,523	0,000	-0,496
3080,38	0,132	0,395	1,000	-0,132	-0,695	0,161	0,591	-0,179	-0,530	0,192	0,487	-0,202	-0,453
3110,27	0,267	0,385	1,000	-0,267	-0,657	0,319	0,526	-0,348	-0,442	0,367	0,379	-0,380	-0,328
3120,41	0,407	0,367	1,000	-0,407	-0,590	0,471	0,415	-0,497	-0,296	0,506	0,204	-0,504	-0,129
3122,02	0,555	0,342	1,000	-0,555	-0,489	0,610	0,254	-0,609	-0,094	0,583	-0,026	-0,545	0,121
3187,5	0,716	0,309	1,000	-0,716	-0,345	0,727	0,038	-0,663	0,159	0,570	-0,293	-0,468	0,384
3310,34	0,899	0,266	1,000	-0,899	-0,136	0,804	-0,244	-0,621	0,451	0,422	-0,556	-0,231	0,593
3315,37	1,119	0,213	1,000	-1,119	0,178	0,798	-0,601	-0,413	0,738	0,071	-0,718	0,201	0,610
3431,29	1,412	0,147	1,000	-1,412	0,703	0,580	-1,018	0,124	0,858	-0,573	-0,517	0,783	0,140
3491,06	1,938	0,061	1,000	-1,938	1,949	-0,598	-1,108	1,495	-0,171	-1,259	1,023	0,526	-1,293

Źródło: opracowanie własne.

Pierwsze 3 wielomiany Hermite'a przedstawiają się jak poniżej:

$$H_o(y) = 1, \quad H_1(y) = -y, \quad H_2 = \frac{1}{\sqrt{2}} \cdot \left(y^2 - 1\right) \tag{5}$$

Znajomość dwóch pierwszych wielomianów wystarcza, aby obliczyć wartość wielomianu dowolnego stopnia, korzysta się przy tym z rekurencyjnej formuły zgodnie z następującym wzorem:

$$H_{n+1}(y) = -\frac{1}{\sqrt{n+1}} \cdot y \cdot H_n(y) - \sqrt{\frac{n}{n+1}} \cdot H_{n-1}(y) \tag{6}$$

Zgodnie z tym, co napisano powyżej przejście ze zmiennych normalnych

standaryzowanych na oryginalne następuje za pomocą funkcji anamorfozy ϕ, czyli:

$$z = F^{-1}(G(y)) = \phi(y) \approx \sum_{p=0}^{P} \phi_p H_p(y) \tag{7}$$

gdzie współczynniki ϕ_p wyrażają się następującymi wzorami (żmudne i długie wyprowadzenie pominięto) (na podstawie Ortiz, Oz, Deutsch):

$$\phi_o = E(z) \tag{8}$$

$$\phi_p = \int \phi(y) \cdot H_p(y) \cdot g(y) dy = \sum_{i=2}^{n} (z_{i-1} - z_i) \cdot \frac{1}{\sqrt{p}} H_{p-1}(y) \cdot g(y) \tag{9}$$

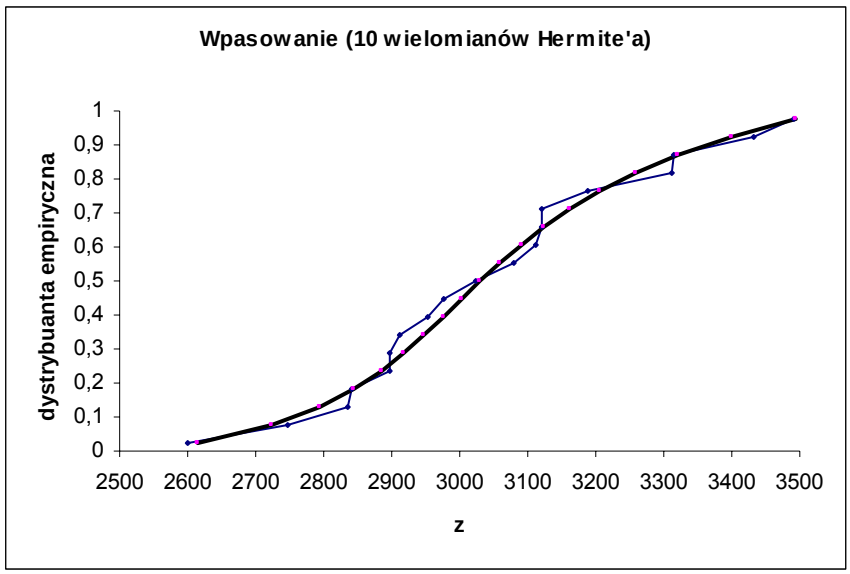
Przykład:

Tabela 2

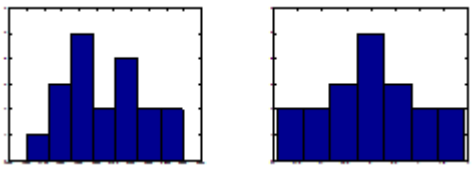
Obliczenie współczynników funkcji anamorfozy (wiersz ostatni)

z=C	y	G(y)	1	2	3	4	5	6	7	8	9	10
2600,92	-1,938	0,061	–	–	–	–	–	–	–	–	–	–
2748,37	-1,412	0,147	-21,683	-21,649	-8,796	6,288	9,872	1,098	-7,030	-4,390	3,733	5,369
2835,79	-1,119	0,213	-18,616	-14,730	-1,916	7,432	5,005	-3,141	-5,190	0,465	4,455	1,184
2840,2	-0,899	0,266	-1,170	-0,744	0,092	0,471	0,128	-0,297	-0,199	0,175	0,217	-0,086
2896,38	-0,716	0,309	-17,366	-8,792	3,455	6,313	-0,296	-4,697	-1,043	3,502	1,696	-2,569
2897,93	-0,555	0,342	-0,513	-0,201	0,145	0,156	-0,058	-0,127	0,018	0,106	0,004	-0,088
2912,97	-0,407	0,367	-5,542	-1,595	1,888	1,305	-1,029	-1,124	0,621	0,990	-0,378	-0,884
2953,72	-0,267	0,385	-15,669	-2,958	5,941	2,501	-3,687	-2,228	2,619	2,034	-1,980	-1,882
2977,49	-0,132	0,395	-9,401	-0,877	3,771	0,755	-2,485	-0,686	1,883	0,638	-1,525	-0,601
3024,91	0	0,399	-18,913	0,000	7,721	0,000	-5,179	0,000	3,996	0,000	-3,297	0,000
3080,38	0,132	0,395	-21,923	2,046	8,794	-1,762	-5,795	1,599	4,391	-1,487	-3,557	1,402
3110,27	0,267	0,385	-11,512	2,173	4,365	-1,837	-2,708	1,637	1,924	-1,494	-1,455	1,383
3120,41	0,407	0,367	-3,707	1,067	1,263	-0,873	-0,688	0,752	0,415	-0,662	-0,253	0,591
3122,02	0,555	0,342	-0,547	0,215	0,155	-0,167	-0,062	0,136	0,020	-0,113	0,005	0,094
3187,5	0,716	0,309	-20,240	10,247	4,027	-7,358	-0,345	5,474	-1,215	-4,082	1,977	2,994
3310,34	0,899	0,266	-32,665	20,765	2,558	-13,138	3,566	8,286	-5,567	-4,874	6,053	2,389
3315,37	1,119	0,213	-1,086	0,860	-0,112	-0,434	0,292	0,183	-0,303	-0,027	0,260	-0,069
3431,29	1,412	0,147	-17,037	17,011	-6,912	-4,941	7,757	-0,863	-5,524	3,450	2,934	-4,219
3491,06	1,938	0,061	-3,648	4,999	-4,104	1,091	1,808	-2,227	0,236	1,623	-1,244	-0,607
ϕ_p			**-221,236**	**7,835**	**22,332**	**-4,197**	**6,094**	**3,775**	**-9,947**	**-4,146**	**7,646**	**4,402**

Źródło: opracowanie własne.



Rys. 3. Wpasowany rozkład oryginalnych zmiennych za pomocą 10 wielomianów ortogonalnych Hermite'a. *Źródło*: opracowanie własne.



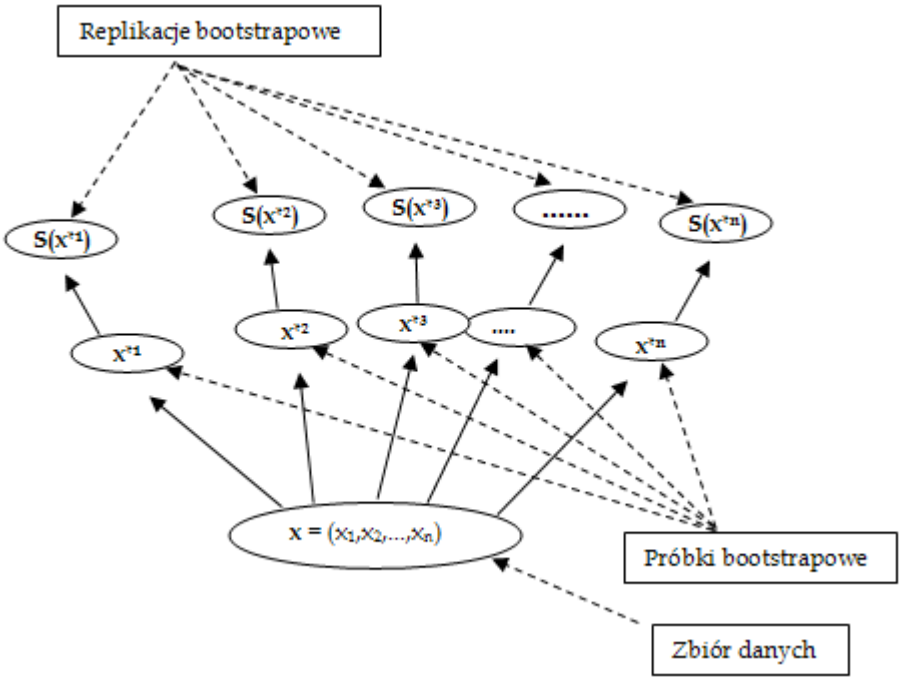
Rys. 4. Anamorfoza gaussowska, histogram oryginalnych zmiennych oraz histogram zmiennych po transformacji. *Źródło*: opracowanie własne.

3. Bootstrap

Idea bootstrapu polega na wielokrotnym losowaniu ze zwracaniem próbek o liczności n z pierwotnego n – elementowego zbioru danych, zatem ten sam element może zostać wylosowany więcej niż jeden raz. Przy pobieraniu próbek bootstrap (ze zwracaniem) każdy z n elementów może zostać wybrany n sposobami, zatem liczba możliwych próbek jest równa n^n.

Na podstawie dużej ilości losowań prób bootstrapowych (przy szacowaniu błędów standardowych przyjmuje się od 25 – 200) jesteśmy w stanie oszacować błąd dowolnej statystyki będącej funkcją próby. Oszacowania metodą bootstrap wolne są od „mocnych" założeń, co do rozkładu statystyki w populacji generalnej, nie wymagają skomplikowanych wyprowadzeń mocno opartych o teorię statystyki i rachunku prawdopodobieństwa aż w końcu mają te zaletę, iż są w stanie „opanować" każdą statystykę niezależnie jak skomplikowana jest jej postać

analityczna. Rysunek 5 przedstawia ideę metody bootstrap.



Rys. 5. Schemat procesu estymacji błędów standardowych za pomocą metody Bootstrap (na podstawie EFRON, TIBSHIRANI 1993). *Źródło*: opracowanie własne.

Procedurę szacowania błędu standardowego metodą bootstrap można przedstawić w następujących krokach (EFORN 1993):

1) Wylosowanie B n – elementowych prób bootstrapowych z n elementowego pierwotnego zbioru danych.
2) Oszacowanie wartości parametru dla każdej z B prób bootstrapowych,
3) dostajemy B oszacowań statystyki s^{*b}, b=1,2,...,B.
4) Oszacowanie błędu standardowego estymatora za pomocą formuły:

$$se_B^* = \sqrt{\frac{\sum_{b=1}^{B}\left(s^{*b} - \hat{s}^*\right)^2}{B-1}} \tag{10}$$

gdzie: $\hat{s}^* = \frac{\sum_{b=1}^{B} s^{*b}}{B}$

Procedura szacowania błędu standardowego dowolnej statystyki przedstawiona zostanie na przykładzie metody korygowania ceny średniej (PKZW NI1, Standard III.7*). Jak wiadomo metoda ta nie posiada zwartej formuły na

określenie błędu szacunku jednostkowej wartości nieruchomości. Tabela 3 zawiera zbiór 19 nieruchomości opisanych 6 atrybutami oraz jednostkową ceną transakcyjną, jak również wszystkie konieczne wielkości oraz przebieg obliczeń w tej metodzie. Liczba możliwych prób bootstrapowych 19^{19}. Tabela 4 przedstawia średnią arytmetyczną oraz błąd szacunku wartości dla kolejno 25, 50, 100, 250, 500, 750 oraz 1000 prób bootstrapowych. Tabela ta również zawiera wyniki dla tego samego zbioru danych, gdy średnią arytmetyczną w metodzie korygowania ceny średniej zastąpiono medianą. Korzystając zarówno ze średniej arytmetycznej jak i z mediany otrzymano zbieżne wyniki: błąd szacunku w granicach 60 zł, oraz średnia arytmetyczna z „wartości" z B prób bootstrapowych na poziomie 3120 zł. Można by nawet wysunąć stwierdzenie, iż zastąpienie pierwotnego oszacowania nieruchomości na poziomie 3153 zł na rzecz 3120 zł jest w pełni uzasadnione. Dodatkowo należy stwierdzić, iż metoda bootstrap świetnie nadaje się dla małych prób, ponieważ rośnie z n – tą potęgą n – elementowego zbioru danych, dla n=3 mamy 9, dla n=5 mamy 25, dla n = 7 mamy już 49 różnych próbek bootstrapowych.

Tabela 3

Dane i obliczenia do metody korygowania ceny średniej

Lp.	Moda	Położenie	Otoczenie	Kondygnacja	Powierzchnia	Standard	Cena
1	1	1	1	-1	37,12	1	2912,97
2	1	2	1	2	24,29	1	3315,37
3	1	0	1	2	32,96	1	2953,72
4	1	2	1	2	30	2	3080,38
5	2	1	2	2	35	1	3120,41
6	1	1	1	-1	22,92	0	2600,92
7	2	2	2	2	37,51	2	3024,91
8	2	2	2	0	37	2	3431,29
9	1	1	1	2	24	1	2896,38
10	2	2	1	-1	23,27	1	3110,27
11	1	2	2	1	38,7	2	3122,02
12	1	1	1	0	25,34	1	2835,79
13	2	1	1	2	21,7	1	2840,2
14	1	1	2	0	24,58	1	3187,5
15	1	1	1	2	27,22	2	2977,49
16	2	2	1	0	36,89	0	2748,37
17	2	2	0	-1	36,01	1	2897,93
18	2	1	2	2	29	1	3491,06
19	2	2	2	2	33,3	1	3310,34
Wyceniana	1	2	1	2	24	1	?

Lp.	Moda	Położenie	Otoczenie	Kondygnacja	Powierzchnia	Standard	Cena
Wagi	5%	25%	20%	15%	15%	20%	C_{max}=3491,06 C_{min}=2600,92 C_{sr}=3045,12 u_{max}=1,15 u_{min}=0,85
u_{imin}	0,04	0,21	0,17	0,13	0,13	0,17	
u_{imax}	0,06	0,29	0,23	0,17	0,17	0,23	
a_{imin}	1	0	0	-1	21,7	0	
a_{imax}	2	2	2	2	38,7	2	
a_{imax} - a_{imin}	1	2	2	3	17	2	
u_{imax}-u_{imin}	0,01	0,07	0,06	0,04	0,04	0,06	
$\Delta_{i/j}$	0,015	0,037	0,029	0,015	0,003	0,029	
a_{iw}-a_{imin}	0	2	1	3	2,3	1	
u_i	0,04	0,29	0,20	0,17	0,13	0,20	Σu_i=1,04
							3153,07

Źródło: opracowanie własne.

Tabela 4. Oszacowanie błędu średniego wartości nieruchomości w metodzie korygowania ceny średniej oraz „poprawiona" wartość, dla 25, 50, 100, 250, 500, 750, 1000 replikacji próby

Na podstawie średniej arytmetycznej							
	25	**50**	**100**	**250**	**500**	**750**	**1000**
Średnia	3133,25	3124,30	3119,46	3117,71	3117,12	3117,75	3117,42
Odchylenie standardowe	62,17	56,31	55,35	60,95	58,46	59,06	57,56
Na podstawie mediany							
	25	**50**	**100**	**250**	**500**	**750**	**1000**
Średnia	3121,56	3119,19	3109,63	3115,28	3114,28	3114,72	3114,24
Odchylenie standardowe	69,18	58,31	58,89	61,87	59,75	60,52	58,52

Źródło: opracowanie własne.

4. Estymacja odporna

Metoda najmniejszych kwadratów jest chyba najczęściej stosowaną metodą estymacji parametrów, ma to uzasadnienie po części w tradycji (posługujemy się nią od czasów LEGENDRE'A (1805) i GAUSSA (1809)) oraz jej zwartym i prostym formułom (zwłaszcza w postaci macierzowej). Jednakże już sami „wynalazcy" metody byli świadomi, że jest ona mało odporna na przypadki odstające w danych. Legendre pisał o eliminacji danych z „dużymi" resztami i ponownej estymacji opartej o pozostałe dane. Jednak jak się okazuje metoda najmniejszych kwadratów potrafi „ukrywać" odstające przypadki, co prowadzi do obciążania wyników. Metoda ta charakteryzuje się niskim (tak naprawdę zerowym) punktem przełamania „breakdown point", wystarczy jedna jedyna obserwacja odstająca, aby całkowicie zaburzyć wyniki estymacji.

Druga połowa XX wieku to lata narodzin „statystyki odpornej" kojarzonej z takimi nazwiskami jak: Tukey, Hampel oraz Huber. Poszukiwali oni odpornych na przypadki odstające metod estymacji parametrów (jak również statystyk

diagnostycznych), nabierających znaczenia ze względu na powiększające się ilości danych pochodzących z obserwacji różnych zjawisk i obiektów. W przypadku małej ilości danych oraz w zagadnieniach „nisko" wymiarowych (2D, 3D) nie stanowi to wielkiego problemu, wystarczą z powodzeniem ludzkie zmysły, jednakże w wielkich zbiorach danych i zagadnieniach wielowymiarowych poszukiwanie przypadków odstających staje się bardzo trudne a wręcz niemożliwe.

Rousseeuw w 1984 wprowadził metodę LMS (Least median of squares – najmniejsza mediana kwadratów), polega ona na takim doborze estymatorów parametrów modelu, aby zminimalizować medianę kwadratów reszt, czyli:

$$\min_{\hat{\theta}} \; med\left(r_i^2\right) \tag{11}$$

Metoda ta charakteryzuje się maksymalnym możliwym 50% „punktem przełamania", oznacza to, że radzi sobie nawet z n/2 przypadkami odstającymi, gdzie n – liczność próby, dodatkowo jest niezależna od liczby estymowanych parametrów. Geometrycznie jest przedstawiana jako najwęższy pas pokrywający połowę obserwacji. Niestety posiada również wadę - brak zwartej formuły podobnej do tej w metodzie najmniejszych kwadratów, aby wyznaczyć parametry modelu niezbędne jest odpowiednie oprogramowanie.

Poniżej przedstawiono 3 przykłady wpasowania liniowego trendu zmiany cen w czasie stosując metodę najmniejszych kwadratów (OLS) oraz metodę najmniejszej mediany kwadratów (LMS). Dla tych samych wartości zmiennej „czas transakcji" t, przedstawiono 3 zestawy cen: C_1, C_2 oraz C_3. W pierwszym przypadku bez obserwacji odstających (Rysunek 6), w drugim przypadku z umiarkowanymi „outlierami" (Rysunek 7) oraz przypadek trzeci – katastrofalne zachowanie metody najmniejszych kwadratów (Rysunek 8).

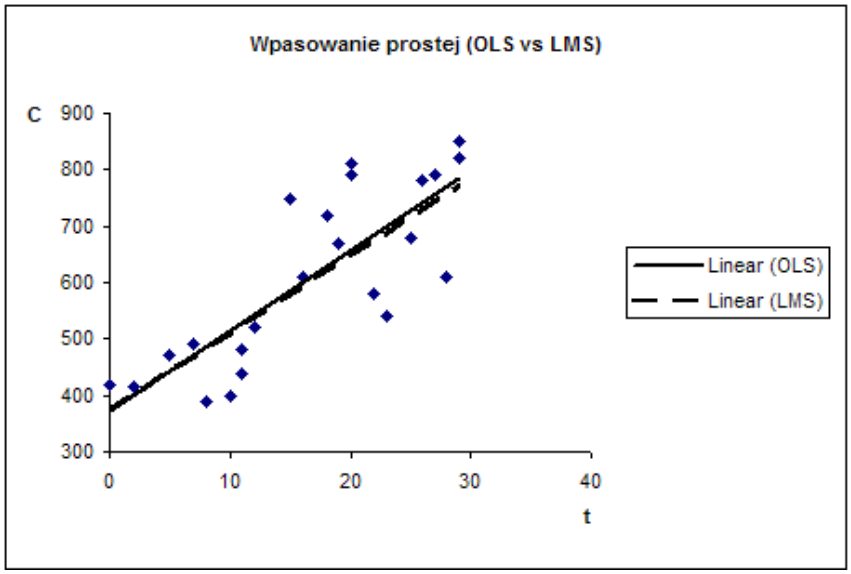
Tabela 5

Dane do oszacowania liniowego trendu zmiany cen w czasie oraz porównania estymacji metodą OLS oraz LMS

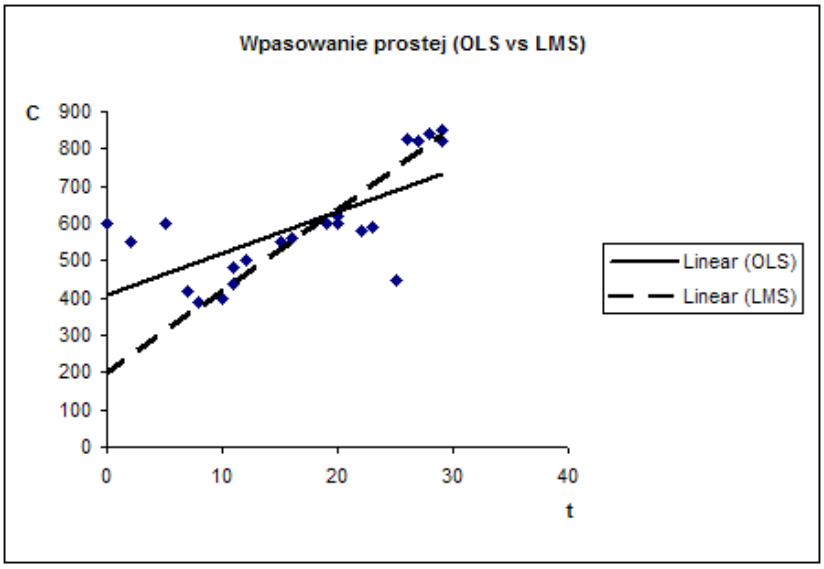
Lp.	t	C_1	C_2	C_3
1	0	420	600	550
2	2	415	550	550
3	5	470	600	500
4	7	490	420	420
5	8	390	390	430
6	10	400	400	400
7	11	440	440	440
8	11	480	480	480
9	12	520	500	500

Lp.	t	C_1	C_2	C_3
10	15	750	550	550
11	16	610	560	560
12	18	720	600	600
13	19	670	600	600
14	20	790	620	620
15	20	810	600	600
16	22	580	580	580
17	23	540	590	590
18	25	680	450	550
19	26	780	825	250
20	27	790	820	300
21	28	610	840	280
22	29	820	820	310
23	29	850	850	290

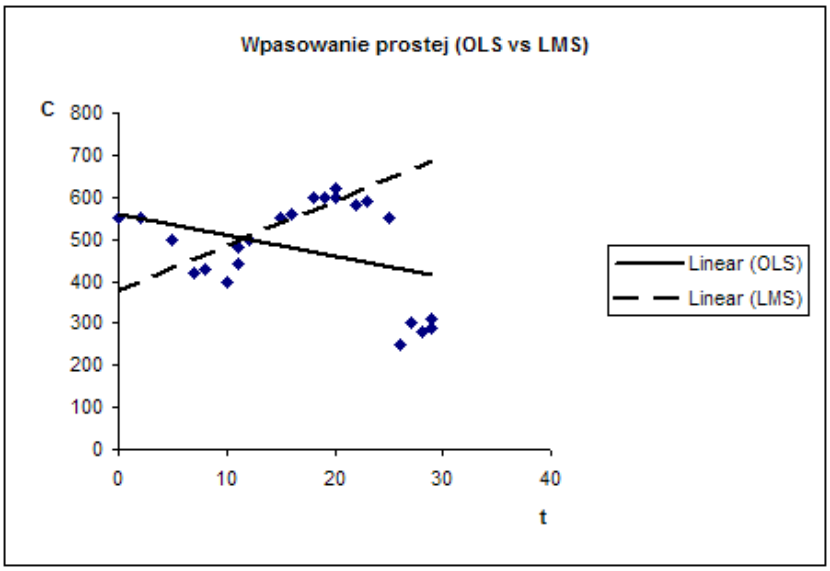
Źródło: opracowanie własne.



Rys. 6. Porównanie OLS z LMS – przypadek pierwszy.
Źródło: opracowanie własne.



Rys. 7. Porównanie OLS z LMS – przypadek drugi. *Źródło*: opracowanie własne.



Rys. 8. Porównanie OLS z LMS – przypadek trzeci. *Źródło*: opracowanie własne.

5. Kroswalidacja

Idea kroswalidacji jest bardzo prosta, polega na podziale zbioru danych na zbiór „treningowy" oraz zbiór „walidacyjny". Na podstawie zbioru „treningowego" budujemy model i na jego podstawie dokonujemy prognozy

danych ze zbioru „walidacyjnego". W zależności od ilości danych zbiór możemy podzielić albo na tylko dwa rozłączne podzbiory (duże zbiory danych) bądź więcej (K – fold cross – validation) bądź też każdorazowo (*N* razy) usuwać tylko jedną obserwację (Leave One Out cross – validation) ze zbioru danych i dokonywać prognozy jej wartości na podstawie pozostałych *N*-1 obserwacji.

Dla każdej krotki w tabeli danych (cena transakcyjna oraz zestaw wyróżnionych atrybutów) możemy wyznaczyć różnicę ceny rzeczywiście obserwowanej oraz tej przewidzianej przez model określony bez jej udziału:

$$W_i - \hat{W}_i^{(n-1)} \tag{12}$$

Jeśli uśrednimy powyższe różnice ze wszystkich obserwacji w zbiorze danych dostaniemy średnią z błędów pochodzących z kroswalidacji. Jeśli wartość ta jest bliska zera to można powiedzieć, że dokonano predykcji bez obciążenia natomiast, jeśli widać znaczące odstępstwo od wartości zero na „plus" lub „minus" oznacza to systematyczne przeszacowywanie lub niedoszacowywanie wartości (WACKERNAGEL 2003).

$$\frac{\sum_{i=1}^{n}\left(W_i - \hat{W}_i^{(n-1)}\right)}{n} \approx 0 \tag{13}$$

Jeśli chcemy porównać rzeczywiste zdolności predykcyjne różnych modeli możemy skonstruować średnio kwadratowy błąd kroswalidacyjny dla każdego z modeli zgodnie z formułą poniżej. Model charakteryzujący się najmniejszą wartością średnio – kwadratowego błędu kroswalidacyjnego będzie pretendował do miana modelu najlepszego.

$$\frac{\sum_{i=1}^{n}\left(W_i - \hat{W}_i^{(n-1)}\right)^2}{n} = \sigma_{CV}^2 \tag{14}$$

Jeżeli metoda predykcji pozwala na oszacowanie błędu prognozy, oszacowanie to może zostać skonfrontowane ze wspomnianą wyżej różnicą wartości rzeczywiście obserwowanej oraz tej pochodzącej z kroswalidacji. Konstruując ilorazy jak w ponizszym wzorze możemy obserwować na ile poszczególne oszacowania różnią się one od jedynki (idealna wartość).

$$\frac{W_i - \hat{W}_i^{(n-1)}}{\sigma\left(\hat{W}_i^{(n-1)}\right)} \tag{15}$$

Jako sumaryczną miarę porównawczą rzeczywistych i tych pochodzących z modelu błędów prognozy możemy zastosować formułę postaci (na podstawie EFRON 1993 oraz WACKERNAGEL 2003):

$$\frac{1}{n}\cdot\sum_{i=1}^{n}\frac{\left(W_i-\hat{W}_i^{(n-1)}\right)^2}{\sigma^2\left(\hat{W}_i^{(n-1)}\right)}\approx 1 \qquad (16)$$

6. Podsumowanie i wnioski

Spośród opisanych metod na szczególną uwagę zasługują bootstrap oraz kroswalidacja. Pierwsza z nich zastosowana do oszacowania błędu standardowego wartości nieruchomości w metodzie korygowania ceny średniej przyniosła dobre wyniki również w oszacowaniu samej wartości nieruchomości (zastąpienie jednorazowego oszacowania zgodnie z formułami na wartość nieruchomości w metodzie korygowania ceny średniej średnią z oszacowań bootstrapowych). Należy mieć jednak świadomość, iż nie był to czysty bootstrap a raczej połączenie bootstrapu z odtwarzaniem brakujących danych bo za taką brakującą daną można potraktować wartość/cenę nieruchomości wycenianej przy znanych atrybutach. Takie podejście do wyceny nieruchomości poszerza obszar poszukiwań metod wyceny nieruchomości o metody odtwarzające brakujace dane (missing data methods). Kroswalidacja jest natomiast świetną metodą testowania spójności analizowanego modelu, odrzucając kolejno jedną bądź wiele obserwacji i starając się je przewidzieć na podstawie pozostałych daje nam wgląd w zachowanie się modelu dla nieznanych danych ale oczywiście przy założeniu, iż pochodzą z tej samej populacji.

W treści artykułu przedstawiono kilka metod statystycznych spośród wielu możliwych do stosowania w analizie rynku nieruchomości i samej wycenie. Należy jednak pamiętać, iż prawidłowości najlepiej zauważalne są w zjawiskach masowych (masowa wycena) i właśnie takim zjawiskom służy statystyka. Wyjątkami mogą tu być: posiadana wiedza a priori o zachowaniu się zjawiska (np. rozkład prawdopodobieństwa badanej zmiennej), stosowanie metod właściwych dla małych prób (metody symulacyjne, np. opisany w treści bootstrap). Stosowanie natomiast wyszukanych metod statystycznych dla małych prób może wcale nie przynieść oczekiwanych skutków a wręcz przeciwnie, gorsze wyniki przy większym nakładzie pracy.

7. Literatura

BILODEAU M., BRENNER D. 1999 Theory of multivariate statistics, Springer

EFRON B., TIBSHIRANI R. J. 1993 An introduction to the Bootstrap, Chapman&Hall/CRC

GOOD P. 2005 Permutation, Parametric and Bootstrap Tests of Hypotheses, Springer

GOOVAERTS P. 1997 Geostatistics for natural resources evaluation, Oxford University Press

HUBER P. J. 1981 Robust statistics, John Wiley & Sons

MARONNA R. A., DOUGLAS MARTIN R., YOHAI V. J. 2006 Robust Statistics, John Wiley & Sons

OLEA R. A. 1999 Geostatistics for engineers and earth scientists, Kluwer Academic Publishers
ORTIZ J. M., OZ B., DEUTSCH C. V. A step by step Guide to Bi – Gaussian Disjunctive Kriging, dostępny na stronie:
http://www.uofaweb.ualberta.ca/ccg//pdfs/2003%2007-DK.pdf
RIVOIRARD J. 1990 Introduction to disjunctive kriging and nonlinear geostatistics, Centre de Geostatistique, Ecole Nationale Superieure des Mines de Paris
ROUSSEEUW P. J., LEROY A. M. 1987 Robust regression and outlier detection, John Wiley & Sons
RAWLINGS J. O., PANTULA S. G., DICKEY A. D. 1998 Applied regression analysis: a research tool, Springer
WACKERNAGEL H. 2003 Multivariate Geostatistics, Springer

STATISTICAL METHODS IN REAL ESTATE VALUATION

Marcin Ligas
Department of Geomatics
University of Science and Technology, AGH, Krakow
e-mail: marcin.ligas@agh.edu.pl

Key words: bootstrap, crossvalidation, robust statistics, data transforms

Abstract

The purpose of the article is to refresh forgotten or perhaps often unknown statistical methods, which should find their place in the workshop of appraiser – statistician. They concern small samples (for individual valuations) and mass appraisal as well. The paper presents bootstrap method and its advantage over classical methods based on strong assumptions, normalizing transformations as: Box and Cox transformation and Normal Score Transform, crossvalidation as a method for testing models and a single example of robust estimation technique. Some of the methods presented in the paper were shown on a simple example.

MODELE MULTIPLIKATYWNE W PROCESIE WYCENY NIERUCHOMOŚCI[2]

Anna Barańska
Katedra Geomatyki
Akademia Górniczo-Hutnicza w Krakowie
e-mail: abaran@agh.edu.pl

Słowa kluczowe: *wielowymiarowe modelowanie rynku nieruchomości, model multiplikatywny, model wykładniczy, model potęgowy*

Streszczenie

W pracy zostanie zaprezentowany algorytm estymacji wartości rynkowej nieruchomości z wykorzystaniem wielowymiarowych modeli funkcyjnych w postaci multiplikatywnej. Rozważania teoretyczne i przykłady praktyczne zostaną zaprezentowane przy zastosowaniu dwóch modeli multiplikatywnych:

- model wykładniczy

$$c = a_0 \cdot a_1^{x_1} \cdot a_2^{x_2} \cdot \ldots \cdot a_m^{x_m} \tag{1.1}$$

- model potęgowy

$$c = a_0 \cdot x_1^{a_1} \cdot x_2^{a_2} \cdot \ldots \cdot x_m^{a_m} \tag{1.2}$$

Powyższe modele są nieliniowe zarówno względem zmiennych jak i swoich parametrów, dlatego proces estymacji parametrów rozpoczyna się od linearyzacji modelu. Linearyzacji możemy dokonać na dwa sposoby:

1) logarytmowanie równań stronami - otrzymujemy odpowiednio:
   - $\ln c = \ln a_0 + x_1 \cdot \ln a_1 + x_2 \cdot \ln a_2 + \cdots + x_m \cdot \ln a_m$
   - $\ln c = \ln a_0 + a_1 \cdot \ln x_1 + a_2 \cdot \ln x_2 + \cdots + a_m \cdot \ln x_m$
2) rozwinięcie wyjściowych równań w szereg Taylora.

W pracy zostały zamieszczone przykłady estymacji parametrów obu powyższych postaci modeli multiplikatywnych, wraz z wynikami parametrycznych testów istotności różnic pomiędzy odpowiadającymi sobie parametrami, uzyskanymi z obu wariantów.

1. Wprowadzenie

Większość wielowymiarowych modeli wyceny nieruchomości opiera się na addytywnych postaciach funkcji modelujących. Są one znacznie prostsze do - estymowania, zwłaszcza, gdy zachowują liniowość względem swoich parametrów. Proces estymacji parametrów znacznie się komplikuje, gdy model nie jest liniowy,

[2] Temat realizowany w ramach badań statutowych w Katedrze Geomatyki, WGGiIŚ, AGH, Kraków

a jego ogólna postać nie ma charakteru addytywnego.

W niniejszej pracy będą rozpatrywane dwa modele w formie multiplikatywnej. Porównane zostaną wyniki estymacji ich parametrów po zastosowaniu dwóch różnych procedur wstępnej linearyzacji modelu. Uwzględniono modele (1.1) i (1.2)

2. Linearyzacja

Chcąc dokonać estymacji parametrów powyższych modeli metodą najmniejszych kwadratów, należy w pierwszej kolejności przeprowadzić proces linearyzacji każdego z nich. Najprostszym sposobem sprowadzenia modelu wykładniczego lub potęgowego do postaci liniowej jest obłożenie stronami równań (1.1) i (1.2) funkcją różnowartościową, którą w tym wypadku jest funkcja logarytmiczna. Drugim, uniwersalnym, aczkolwiek przybliżonym sposobem linearyzacji, jest rozwinięcie w szereg Taylora. Oba te sposoby zastosowano do każdego z rozpatrywanych modeli multiplikatywnych.

2.1. Przekształcenie różnowartościowe (logarytmowanie)

Po zlogarytmowaniu stronami równań (1.1) i (1.2), przy założeniu, że ich liczba będzie większa niż liczba estymowanych parametrów (nadliczbowy układ równań), otrzymujemy odpowiednio:

$$\ln c = \ln a_0 + x_1 \cdot \ln a_1 + x_2 \cdot \ln a_2 + \cdots + x_m \cdot \ln a_m + \ln \delta \tag{2.1}$$

$$\ln c = \ln a_0 + a_1 \cdot \ln x_1 + a_2 \cdot \ln x_2 + \cdots + a_m \cdot \ln x_m + \ln \delta \tag{2.2}$$

Równania te zachowują liniowość względem parametrów przekształconych w sposób różnowartościowy, monotoniczny. Zatem w obu przypadkach można bez przeszkód zastosować metodę najmniejszych kwadratów do ich estymacji, przy uwzględnieniu zapisu macierzowego dla *n* równań, powstałych z *n* rekordów bazy danych, będących podstawą estymacji:

$$[Y] = [X] \cdot [a] + [\delta] \tag{2.3}$$

gdzie:

$[Y] = \begin{bmatrix} \ln c_1 \\ \ln c_2 \\ \vdots \\ \ln c_n \end{bmatrix}$ – wektor logarytmów cen nieruchomości z bazy danych (n×1),

$[X] = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1m} \\ 1 & x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix}$ – macierz zmiennych niezależnych modelu wykładniczego (n×(m+1)),

$[X] = \begin{bmatrix} 1 & \ln x_{11} & \ln x_{12} & \cdots & \ln x_{1m} \\ 1 & \ln x_{21} & \ln x_{22} & \cdots & \ln x_{2m} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \ln x_{n1} & \ln x_{n2} & \cdots & \ln x_{nm} \end{bmatrix}$ - macierz zmiennych niezależnych modelu

potęgowego (n×(m+1)),

$$[a] = \begin{bmatrix} \ln a_0 \\ \ln a_1 \\ \vdots \\ \ln a_m \end{bmatrix}$$ - wektor parametrów zlinearyzowanego modelu wykładniczego ((m+1)×1),

$$[\alpha] = \begin{bmatrix} \ln \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_m \end{bmatrix}$$ - wektor parametrów zlinearyzowanego modelu potęgowego ((m+1)×1),

$$[\delta] = \begin{bmatrix} \ln \delta_1 \\ \ln \delta_2 \\ \vdots \\ \ln \delta_n \end{bmatrix}$$ - wektor odchyłek losowych modeli zlinearyzowanych (n×1),

n – liczba nieruchomości w bazie danych,
m – liczba rozpatrywanych cech nieruchomości.

2.2. Zastosowanie szeregu Taylora

Rozwinięcie funkcji w szereg Taylora jest uniwersalnym sposobem na jej linearyzację. Im wyższa pochodna, do której zastosujemy takie rozwinięcie, tym wyższa dokładność uzyskanych wyników estymacji parametrów modelu. W zagadnieniach geodezyjnych rozwinięcie do pierwszej pochodnej jest zazwyczaj wystarczające, gdyż znamy przybliżone wartości niewiadomych.

Modele (1.1) i (1.2) zapisane w postaci szeregu Taylora do pierwszej pochodnej, przyjmują postać:

$$c = \dot{c} + \sum_{i=1}^{m} \frac{\partial c}{\partial a_i} \cdot da_i \qquad (2.4)$$

gdzie:

$\dot{c}$ – przybliżona wartość funkcji (1.1) lub (1.2), określona dla przybliżonych wartości parametrów,

$\frac{\partial c}{\partial a_i}$ – pochodna cząstkowa funkcji (1.1) lub (1.2) po parametrze a_i,

da_i – poprawka do przybliżonej wartości parametru a_i.

Odpowiednik układu równań (2.3) dla tego sposobu linearyzacji ma postać:

$$[L] = [X] \cdot [da] + [\delta] \qquad (2.5)$$

gdzie:

$[L] = [C] - [C_0] = \begin{bmatrix} c_1 - \dot{c}_1 \\ c_2 - \dot{c}_2 \\ \vdots \\ c_n - \dot{c}_n \end{bmatrix}$ – wektor wyrazów wolnych, różnic cen nieruchomości z bazy danych i ich przybliżonych wartości (n×1),

$[X] = \left[\frac{\partial c_i}{\partial a_j}\right]_{\substack{i=1,2,\ldots,n \\ j=0,1,2,\ldots,m}}$ – macierz pochodnych cząstkowych równań (1.1) lub (1.2) po nieznanych parametrach (n×(m+1)),

$[da]$ - wektor przyrostów (poprawek) do przybliżonych wartości niewiadomych ((m+1)×1),

$[\delta] = \begin{bmatrix} \delta_1 \\ \delta_2 \\ \vdots \\ \delta_n \end{bmatrix}$ – wektor odchyłek losowych modelu (1.1) lub (1.2).

3. Estymacja parametrów modeli zlinearyzowanych

Do estymacji parametrów modeli zlinearyzowanych w obu przypadkach stosujemy metodę najmniejszych kwadratów, która pozwala wyznaczyć estymatory nieznanych parametrów tak, by spełniony był warunek minimalizacji sumy kwadratów odchyłek:

$$[\delta]^T \cdot [\delta] = min \tag{3.1}$$

Jednakże w przypadku układu równań (2.3), po zastosowaniu pierwszego sposobu linearyzacji, polega ona *de facto* na minimalizacji sumy kwadratów logarytmów odchyłek wyjściowego modelu. Chodzi zatem o spełnienie warunku:

$$[\ln \delta]^T \cdot [\ln \delta] = min \tag{3.2}$$

Należy zauważyć, że na składnik losowy uzyskany w wyniku estymacji parametrów modelu zlinearyzowanego przy pomocy rozwinięcia w szereg, składa się nie tylko losowość danych rynkowych będących podstawą estymacji, czy nadliczbowość układu równań (2.5), ale również nie uwzględnione w rozwinięciu funkcji pochodne wyższych rzędów i tzw. reszta Lagrange'a.

Szczegółowo proces estymacji został opisany m.in. w pracach (BARAŃSKA 2005) oraz (BARAŃSKA 2006).

3.1. Estymacja po linearyzacji logarytmicznej

Rozwiązując układ równań normalnych otrzymany z równania (2.3), uzyskujemy nieobciążony estymator wektora [a] w postaci:

$$[\hat{a}] = [X^T \cdot X]^{-1} \cdot [X^T \cdot Y] \tag{3.3}$$

którego charakterystykę niedokładności zawiera jego macierz kowariancji:

$$Cov(\hat{a}) = \hat{\sigma}^2 \cdot [X^T \cdot X]^{-1} \tag{3.4}$$

przy czym estymator wariancji resztowej $\hat{\sigma}^2$ wyznaczamy według wzoru:

$$\hat{\sigma}^2 = \frac{Y^T \cdot Y - \hat{a}^T \cdot X^T \cdot Y}{n-(m+1)} \tag{3.5}$$

W ten sposób możemy oszacować każdy z parametrów modelu punktowo i przedziałowo na zadanym poziomie ufności.

3.2. Estymacja po rozwinięciu w szereg Taylora

Rozwiązując układ równań (2.5), uzyskujemy wektor poprawek do przybliżonych parametrów modelu wykładniczego lub potęgowego:

$$[d\hat{a}] = [X^T \cdot X]^{-1} \cdot [X^T \cdot L] \tag{3.6}$$

z którego wyznaczamy wyrównane wartości nieznanych parametrów:

$$[\hat{a}] = [a] + [d\hat{a}] \tag{3.7}$$

Macierz kowariancji estymatorów wyznaczanych parametrów jest analogiczna do postaci (3.4):

$$Cov(d\hat{a}) = Cov(\hat{a}) = \hat{\sigma}^2 \cdot [X^T \cdot X]^{-1} \tag{3.8}$$

przy czym wariancję resztową wyznaczamy z wektora [L]:

$$\hat{\sigma}^2 = \frac{L^T \cdot L - [d\hat{a}]^T \cdot X^T \cdot L}{n-(m+1)} \tag{3.9}$$

4. Porównanie wyników estymacji

W celu porównania wyników estymacji parametrów modeli multiplikatywnych, uzyskanych dla dwóch różnych sposobów ich wstępnej linearyzacji, przy pomocy stosownych testów statystycznych badamy statystyczną istotność różnic między nimi. Wykorzystujemy tu funkcję testową postaci:

$$Z = \frac{\hat{a}_{i_I} - \hat{a}_{i_{II}}}{\sqrt{\sigma^2(\hat{a}_{i_I}) + \sigma^2(\hat{a}_{i_{II}})}} \tag{4.1}$$

gdzie:

$\hat{a}_{i_I}$, $\hat{a}_{i_{II}}$ – estymatory parametru i modelu, uzyskane z dwóch rożnych sposobów linearyzacji,

$\sigma(\hat{a}_{i_I})$, $\sigma(\hat{a}_{i_{II}})$ – odchylenia standardowe powyższych estymatorów.

Weryfikujemy w ten sposób hipotezę zerową:

$$H_0: a_{i_I} = a_{i_{II}} \tag{4.2}$$

przeciw hipotezie alternatywnej:

$$H_1: a_{i_I} \neq a_{i_{II}} \tag{4.3}$$

Przy pomocy parametrycznego testu istotności możemy również porównać odchylenia standardowe uzyskanych estymatorów parametrów, badając w ten sposób wpływ sposobu linearyzacji funkcji modelującej na dokładności oszacowanych parametrów. W tym teście korzystamy z następującej funkcji testowej:

$$F = \frac{\sigma^2(\hat{a}_{i_I})}{\sigma^2(\hat{a}_{i_{II}})} \tag{4.4}$$

Weryfikowana hipoteza zerowa ma postać:

$$H_0: \sigma^2(\hat{a}_{i_I}) = \sigma^2(\hat{a}_{i_{II}}) \tag{4.5}$$

przeciw hipotezie alternatywnej:

$$H_1: \sigma^2(\hat{a}_{i_I}) > \sigma^2(\hat{a}_{i_{II}}) \tag{4.6}$$

W obu przypadkach wnioskowanie będzie prowadzone na standardowym poziomie istotności α=0,05.

5. Prognoza wartości nieruchomości

Każdy z wyestymowanych modeli został poddany weryfikacji pod kątem jakości dopasowania do danych rynkowych i wiarygodności uzyskanych parametrów. Pozytywnie zakończona weryfikacja modelu pozwala wykorzystać go do prognozy wartości modelowej wybranej nieruchomości, której cechy są znane i mieszczą się w rozpiętości cech nieruchomości będących podstawą estymacji modelu. Dla obu sposobów linearyzacji estymacja wartości modelowej przebiega nieco inaczej.

5.1. Prognoza po linearyzacji logarytmicznej

W wyniku estymacji parametrów modeli zlinearyzowanych za pomocą przekształcenia logarytmicznego, uzyskujemy estymatory parametrów „obarczone" logarytmem. Na ich podstawie wyznaczamy bezpośrednio wartości funkcji (2.1) i (2.2), czyli logarytmy modelowych cen wybranych nieruchomości:

$$\ln w_M = [X_w] \cdot [\hat{a}] \tag{5.1}$$

gdzie:

$[X_w] = [1 \quad x_{w1} \quad x_{w2} \quad \ldots \quad x_{wm}]$ – wektor wartości atrybutów wycenianej nieruchomości (1×(*m*+1)); zmienne niezależne w modelu wykładniczym zlinearyzowanym,

$[X_w] = [1 \quad \ln x_{w1} \quad \ln x_{w2} \quad \ldots \quad \ln x_{wm}]$ – wektor logarytmów atrybutów wycenianej nieruchomości (1×(*m*+1)); zmienne niezależne w modelu potęgowym zlinearyzowanym,

$[\hat{a}] = \begin{bmatrix} \ln \hat{a}_0 \\ \ln \hat{a}_1 \\ \vdots \\ \ln \hat{a}_m \end{bmatrix}$ – wektor estymatorów parametrów zlinearyzowanego modelu wykładniczego ($(m+1)\times 1$),

$[\hat{a}] = \begin{bmatrix} \ln \hat{a}_0 \\ \hat{a}_1 \\ \vdots \\ \hat{a}_m \end{bmatrix}$ – wektor estymatorów parametrów zlinearyzowanego modelu potęgowego ($(m+1)\times 1$).

Stąd modelowe wartości nieruchomości wycenianej wyznaczamy według wzoru:

$$w_N = \exp(\ln w_N) \tag{5.2}$$

Analizę dokładności uzyskanej wartości modelowej wykonujemy na podstawie macierzy kowariancji $Cov(\hat{a})$, wyznaczonej według wzoru (3.4), wyliczając najpierw odchylenie standardowe $\sigma(\ln w_N)$:

$$V(\ln w_N) = \sigma^2(\ln w_N) = [X_w] \cdot Cov(\hat{a}) \cdot [X_w]^T \tag{5.3}$$

a następnie, z prawa przenoszenia wariancji, odchylenie standardowe wartości modelowych wyjściowych modeli multiplikatywnych (1.1) lub (1.2), czyli wartości modelowych wycenianej nieruchomości $\sigma(w_N)$:

$$V(w_N) = \sigma^2(w_N) = w_N^2 \cdot \sigma^2(\ln w_N) \tag{5.4}$$

5.2. Prognoza po rozwinięciu funkcji w szereg Taylora

W przypadku estymacji parametrów funkcji zlinearyzowanej za pomocą szeregu Taylora, uzyskujemy poprawki $[d\hat{a}]$ (3.6) do przybliżonych wartości parametrów wyjściowych modeli multiplikatywnych (1.1) lub (1.2) i ich macierze kowariancji. Prognozę wartości modelowej nieruchomości wykonujemy według wzoru:

$$w_N = c_w + \left[\frac{\partial c_w}{\partial a}\right] \cdot [d\hat{a}] \tag{5.5}$$

gdzie:

c_w – przybliżona wartość modelowa wycenianej nieruchomości, wyznaczona z przybliżonych wartości parametrów,

$\left[\frac{\partial c_w}{\partial a}\right] = \left[\frac{\partial c_w}{\partial a_0} \quad \frac{\partial c_w}{\partial a_1} \quad \frac{\partial c_w}{\partial a_2} \quad \dots \quad \frac{\partial c_w}{\partial a_m}\right]$ – wektor pochodnych cząstkowych równania (1.1) lub (1.2) po nieznanych parametrach ($1\times(m+1)$), wyznaczony na

podstawie wartości atrybutów wycenianej nieruchomości.

Dokładność powyższej prognozy szacujemy na podstawie macierzy kowariancji (3.8):

$$V(w_N) = \sigma^2(w_N) = \left[\frac{\partial c_W}{\partial a}\right] \cdot Cov(\hat{a}\hat{a}) \cdot \left[\frac{\partial c_W}{\partial a}\right]^T \tag{5.6}$$

Celem porównania prognoz wartości uzyskanych z rożnych modeli, czy z rożnych sposobów linearyzacji, możemy zastosować parametryczne testy istotności, analogiczne do opisanych w punkcie 4.

6. Przykład praktyczny

Przedmiotem analiz była baza danych złożona ze 142 nieruchomości lokalowych z terenu miasta Chrzanowa, opisanych za pomocą 15 atrybutów. W ramach wstępnej analizy zgromadzonych danych, usunięto przypadki odstające i nieistotne cechy nieruchomości.

Przypadki odstające eliminowano na podstawie kryterium dotyczącego wartości standaryzowanej reszty, czyli różnicy pomiędzy między wartością rzeczywistą i modelową zmiennej objaśnianej. Standaryzowaną resztę wyznacza się dzieląc resztę przez pierwiastek ze średniego kwadratu reszt według wzoru:

$$\frac{c_i - w_i}{\sqrt{\frac{1}{n} \cdot \sum_{i-1}^{n} (c_i - w_i)^2}} \tag{6.1}$$

gdzie:
c_i – wartość obserwowana zmiennej objaśnianej dla przypadku i,
w_i – wartość modelowa zmiennej objaśnianej dla przypadku i,
n – liczba przypadków.

Za odstające uznano przypadki, dla których wartość powyższego wyrażenia przekraczała 2. Ciekawym może być porównanie podstawowych charakterystyk opisowych zmiennych występujących w modelu, przed i po usunięciu przypadków odstających. Tabele z wartościami charakterystyk zawiera załącznik.

W iteracyjnym procesie estymacji parametrów modeli, z wykorzystaniem algorytmu Gaussa-Newtona, poddawano weryfikacji ich statystyczną istotność, pozostawiając tylko te spośród nich, które okazały się decydujące w kształtowaniu ceny nieruchomości w danym modelu. Zastosowano tutaj klasyczny test istotności statystycznej na poziomie istotności 0,05, w którym statystyka ma postać ilorazu wartości estymatora parametru stojącego przy danej zmiennej losowej, do wartości jego odchylenia standardowego. Za zmienne istotne uznano te, dla których parametry przy nich stojące, wykazały statystyczną istotność.

Zabieg redukcji odstających przypadków i nieistotnych zmiennych ma na celu poprawienie statystycznej jakości modelu. Standardowym miernikiem oceny statystycznej modelu jest współczynnik determinacji, czyli kwadrat współczynnika korelacji wielorakiej R^2, określający udział wariancji części wyjaśnionej

w całkowitym rozrzucie zmiennej zależnej wokół jej wartości przeciętnej. Jego wartość powinna przekraczać 0,5. Dodatkowo wykonujemy test porównujący wariancje części wyjaśnionej i niewyjaśnionej modelem regresji, weryfikujący na ile wiarygodna jest „duża" wartość R^2. Pozostałe mierniki oceny statystycznej modelu omówiono szczegółowo w pracy (BARANSKA, CZAJA 2007a).

6.1. Model wykładniczy

Ostateczne zadowalające wyniki estymacji parametrów modelu wykładniczego uzyskano na bazie 134 nieruchomości oraz 8 atrybutów. Zatem postać wyjściowa modelu była następująca:

$$c = a_0 \cdot a_1^{x_1} \cdot a_2^{x_2} \cdot \ldots \cdot a_8^{x_8} \tag{6.2}$$

Tabela 1

Wyniki estymacji parametrów modelu wykładniczego dla bazy lokali

logarytmowanie		szereg Taylora		wyniki testów			
$\hat{a}_i$	$\sigma(\hat{a}_i)$	$\hat{a}_i$	$\sigma(\hat{a}_i)$	Z_{obl}	Z_{tab}	F_{obl}	F_{tab}
1061,817	140,804	1066,973	261,312	0,02		**3,44**	
1,057	0,012	1,074	0,017	0,82		**2,11**	
1,052	0,016	1,075	0,023	0,82		**2,06**	
0,966	0,014	0,942	0,023	-0,90		**2,50**	
1,058	0,015	1,085	0,022	1,01	1,96	**2,08**	1,35
1,149	0,046	1,160	0,080	0,12		**3,11**	
1,056	0,018	1,066	0,027	0,30		**2,21**	
0,993	0,001	0,991	0,002	-0,71		**2,39**	
1,101	0,029	1,077	0,045	-0,45		**2,32**	

Źródło: opracowanie własne.

Tabela 1 zawiera estymatory parametrów $a_0, a_1, a_2, \ldots, a_8$ i ich odchylenia standardowe wyznaczone po zastosowaniu dwóch sposobów linearyzacji funkcji (6.2) oraz wyniki parametrycznych testów istotności, porównujących odpowiadające sobie estymatory i ich dokładności. Pogrubioną czcionką zapisano te wartości funkcji testowych, które wskazują na statystyczną istotność analizowanych różnic.

Okazuje się, że sposób linearyzacji modelu nie wpływa znacząco na wartości jego parametrów, jednakże zauważalny jest wpływ na dokładności wyznaczonych estymatorów. Istotnie dokładniejsze estymatory parametrów uzyskujemy po wykonaniu linearyzacji przy pomocy różnowartościowego przekształcenia funkcyjnego. Warto zauważyć, że wszystkie odchylenia standardowe parametrów uzyskanych po zastosowaniu rozwinięcia w szereg Taylora okazały się istotnie większe od odpowiadających im odchyleń standardowych estymatorów parametrów wyznaczonych z przekształcenia logarytmicznego.

6.2. Model potęgowy

W przypadku modelu potęgowego, najlepsze wyniki dopasowania modelu do danych rynkowych otrzymano po usunięciu z bazy danych 24 przypadków odstających oraz 10 nieistotnych zmiennych objaśniających (cech nieruchomości). Zatem proces estymacji parametrów został przeprowadzony na podstawie grupy 118 nieruchomości i 5 atrybutów cenotwórczych. Postać estymowanego modelu potęgowego jest następująca:

$$c = a_0 \cdot x_1^{a_1} \cdot x_2^{a_2} \cdot x_3^{a_3} \cdot x_4^{a_4} \cdot x_5^{a_5} \tag{6.3}$$

Po przeprowadzeniu procesu estymacji w dwóch omówionych powyżej wariantach, uzyskano wartości oszacowań parametrów modelu wraz z ich odchyleniami standardowymi, zamieszczone w tabeli 2. Również i w tym przypadku, przeprowadzone parametryczne testy istotności nie wykazały istotności różnic pomiędzy odpowiadającymi sobie estymatorami, uzyskanymi w wyniku różnie przeprowadzonej linearyzacji modelu. Zidentyfikowano natomiast istotne różnice pomiędzy dokładnościami oszacowań niektórych parametrów. Podobnie jak dla modelu wykładniczego, oszacowania uzyskane po rozwinięciu funkcji w szereg, okazały się w połowie przypadków istotnie mniej dokładne.

Tabela 2

Wyniki estymacji parametrów modelu potęgowego dla bazy lokali

logarytmowanie		szereg Taylora		wyniki testów			
$\hat{a}_i$	$\sigma(\hat{a}_i)$	$\hat{a}_i$	$\sigma(\hat{a}_i)$	Z_{obl}	Z_{tab}	F_{obl}	F_{tab}
1252,256	132,407	1187,366	160,444	-0,31	1,96	**1,47**	1,37
0,185	0,030	0,200	0,039	0,31		**1,68**	
0,065	0,032	0,076	0,037	0,24		1,36	
0,175	0,032	0,153	0,038	-0,43		1,35	
0,036	0,018	0,036	0,020	-0,02		1,25	
0,124	0,068	0,169	0,081	0,43		**1,44**	

Źródło: opracowanie własne.

6.3. Obliczenie wartości modelowej

Kolejnym etapem obliczeń było wyznaczenie wartości modelowej wybranej nieruchomości lokalowej z rozpatrywanego rynku lokali, wraz z jej odchyleniem standardowym. Obliczenia wykonano dla obu modeli multiplikatywnych i na podstawie parametrów uzyskanych z obu sposobów linearyzacji funkcji modelowej. W celu porównania uzyskanych wyników, przeprowadzono stosowne testy statystyczne na poziomie istotności 0,05, zgodnie ze schematem opisanym w punkcie 4. Porównywano zarówno efekty zastosowania różnych postaci modeli multiplikatywnych, jak i rożnych sposobów linearyzacji funkcji modelującej. Wszystkie wyniki zamieszczono w tabeli 3. Pogrubione wartości funkcji testowych

wskazują na istotność analizowanej różnicy.

Wyniki zamieszczone w tabeli 3 pokazują, iż dobrze dopasowane do danych rynkowych modele multiplikatywne pozwalają zaprognozować modelową wartość nieruchomości na tym samym poziomie (statystyka Z_{obl} w żadnym z testów nie wykazuje statystycznej istotności). Na wynik prognozy nie wpływa ani postać modelu, ani zastosowana metoda jego linearyzacji. Natomiast dokładność prognozy w każdym przypadku zależy od postaci modelu (model wykładniczy okazał się bardziej dokładny) i może też zależeć od sposobu wstępnej linearyzacji funkcji modelującej. Model wykładniczy okazał się bardziej wrażliwy na sposób linearyzacji.

Tabela 3

Wyniki estymacji wartości modelowej lokalu

	logarytmowanie		szereg Taylora		porównanie sposobów linearyzacji			
Model	w_M	$\sigma(w_M)$	w_M	$\sigma(w_M)$	Z_{obl}	Z_{tab}	F_{obl}	F_{tab}
wykładniczy	2153,98	54,43	2181,14	68,79	0,31	1,96	**1,60**	1,35
potęgowy	2113,77	79,58	2140,95	86,42	0,23		1,18	1,37
porównanie modeli multiplikatywnych								
	0,42	**2,14**	0,36	**1,58**				
	1,96	1,35	1,96	1,35				
	Z_{obl}	F_{obl}	Z_{tab}	F_{obl}				

Źródło: opracowanie własne.

7. Podsumowanie

W pracy zaprezentowano przebieg procesu estymacji parametrów modeli multiplikatywnych w postaci wykładniczej i potęgowej, modelujących wartość nieruchomości, z zastosowaniem metody najmniejszych kwadratów, ale odniesionej do funkcji zlinearyzowanych w różny sposób.

Na podstawie wyników zawartych w tabelach 1 i 2 widać, iż oszacowanie parametrów modeli multiplikatywnych nie zależy w sposób istotny od sposobu linearyzacji pierwotnej postaci funkcji modelującej. Jednakże dokładność takiego oszacowania może się okazać istotnie lepsza, gdy linearyzujemy model za pomocą różnowartościowej funkcji monotonicznej. Wartości odchyleń standardowych estymatorów parametrów są wtedy zawsze bezwzględnie mniejsze od odchyleń standardowych odpowiadających im estymatorów wyliczonych po rozwinięciu w szereg wielowymiarowej nieliniowej funkcji modelującej. Dla modelu wykładniczego 100% parametrów oszacowanych z rozwinięcia funkcji w szereg okazało się istotnie mniej dokładne niż odpowiadające im parametry oszacowane po linearyzacji logarytmicznej. W przypadku modelu potęgowego połowa parametrów uzyskanych z szeregu Taylora jest istotnie mniej dokładna. Widać

zatem, iż tam, gdzie jest możliwe zastosowanie różnowartościowego przekształcenia funkcyjnego w celu linearyzacji modelu, tam zawsze warto to przeprowadzić, by uzyskać istotnie dokładniejsze oszacowania parametrów funkcji modelującej.

Wyniki zawarte w tabeli 3 wskazują na fakt, iż wartość prognozy uzyskanej z rozpatrywanych modeli multiplikatywnych nie zależy od ich postaci, ani od sposobu linearyzacji. Natomiast dokładność takiej prognozy jest zawsze lepsza, gdy zastosowano linearyzację za pomocą przekształcenia różnowartościowego. Poprzez „prognozowanie" lub „przewidywanie" rozumiemy tutaj wyznaczenie wartości nieruchomości wynikającej z aktualnie przyjętego równania regresji (modelu).

Warto zaznaczyć, że porównywano modele jednakowo dobre pod względem dopasowania do danych rynkowych. Osobno wykonane testy statystyczne weryfikujące jakość dopasowania oraz porównujące wariancje części wyjaśnionych różnymi modelami (wykładniczym i potęgowym) lub tym samym modelem, ale wyestymowanym po zastosowaniu różnych metod linearyzacji – nie wykazują statystycznej istotności. Zatem wyeliminowano ewentualny wpływ różnej jakości uzyskanych funkcji modelujących lokalny rynek nieruchomości lokalowych, na wnioski wyciągnięte powyżej.

Wprawdzie przeprowadzone parametryczne testy istotności, porównujące wyniki estymacji punktowej modelowej wartości wycenianej nieruchomości (tabela 3), dowodzą nieistotności różnic pomiędzy wartościami modelowymi i jednocześnie uznają za istotne różnice pomiędzy odchyleniami standardowymi tych wartości, to jednak uzasadnianie tego wpływem eliminacji przypadków lub zmiennych na ten fakt, wydaje się pochopne. Wpływ ten trudno ocenić, gdyż w każdym modelu inne przypadki okazały się odstające i inne zmienne okazały się nieistotne. Proces usuwania przypadków i zmiennych miał na celu poprawienie stabilności wyznaczanych estymatorów parametrów oraz stopnia dopasowania do danych rynkowych estymowanego modelu, co za tym idzie, zwiększenie wiarygodności wyników uzyskanych na jego podstawie. Był to zatem proces odnoszący się do każdego z modeli indywidualnie.

Prezentowane modele niewątpliwie wymagają stosownej (znacznej) liczby danych rynkowych. Nie mogą zatem być stosowane do rynków nieruchomości nietypowych, czy rzadko występujących w obrocie. Zgromadzenie wystarczająco licznej bazy danych, będącej podstawą estymacji parametrów modeli, jest warunkiem koniecznym zwłaszcza w przypadku modeli nieliniowych.

8. Załącznik

Poniższe tabele zawierają podstawowe charakterystyki opisowe zmiennych występujących w obu modelach (wykładniczym i potęgowym), przed o po usunięciu przypadków odstających.

Tabela 4

Model wykładniczy – charakterystyki opisowe zmiennych w pełnej bazie danych

n=142	Średnia	Mediana	Moda	Liczność mody	Min	Max	Dolny kwartyl	Górny kwartyl	Odchyl. standard	Współ. zmien.
Kom.	3,32	4,00	4,00	60	1,00	5,00	3,00	4,00	1,21	36,51
Sąs.	3,61	3,00	3,00	62	1,00	5,00	3,00	5,00	1,12	31,15
Tech.	2,80	3,00	3,00	43	1,00	5,00	2,00	4,00	1,18	42,07
SBud.	3,20	3,00	3,00	42	1,00	5,00	2,00	4,00	1,18	36,92
SLok.	2,94	3,00	3,00	137	1,00	3,00	3,00	3,00	0,34	11,62
Funk.	2,85	3,00	3,00	61	1,00	5,00	2,00	3,00	0,90	31,51
Pow.	47,31	47,47	60,59	5	17,57	103,00	35,50	57,02	15,31	32,36
Pokoje	2,00	2,00	2,00	56	1,00	4,00	1,00	3,00	0,81	40,39

Źródło: opracowanie własne.

Tabela 5

Model wykładniczy – charakterystyki opisowe po usunięciu 8 przypadków

n=134	Średnia	Mediana	Moda	Liczność mody	Min	Max	Dolny kwartyl	Górny kwartyl	Odchyl. standard	Współ. zmien.
Kom.	3,30	4,00	4,00	54	1,00	5,00	3,00	4,00	1,23	37,37
Sąs.	3,60	3,00	3,00	58	1,00	5,00	3,00	5,00	1,14	31,64
Tech.	2,82	3,00	3,00	41	1,00	5,00	2,00	4,00	1,17	41,43
SBud.	3,19	3,00	3,00	40	1,00	5,00	2,00	4,00	1,17	36,74
SLok.	2,95	3,00	3,00	130	1,00	3,00	3,00	3,00	0,31	10,46
Funk.	2,84	3,00	3,00	59	1,00	5,00	2,00	3,00	0,87	30,76
Pow.	46,75	45,88	60,59	5	17,57	100,93	35,47	56,64	14,61	31,24
Pokoje	1,98	2,00	2,00	54	1,00	4,00	1,00	3,00	0,79	39,93

Źródło: opracowanie własne.

Tabela 6

Model potęgowy – charakterystyki opisowe zmiennych w pełnej bazie danych

n=142	Średnia	Mediana	Moda	Liczność mody	Min	Max	Dolny kwartyl	Górny kwartyl	Odchyl. standard	Współ. zmien.
Lok	3,82	4,00	4	57	1,00	5,00	3,00	5,00	1,15	30,08
Otocz	3,03	3,00	3	45	1,00	5,00	2,00	4,00	1,06	35,17
SBud	3,20	3,00	3	42	1,00	5,00	2,00	4,00	1,18	36,92
Kond	3,37	4,00	5	64	1,00	5,00	2,00	5,00	1,69	50,34
Prawo	4,26	4,00	4	89	2,00	5,00	4,00	5,00	0,63	14,71

Źródło: opracowanie własne.

Tabela 7

Model potęgowy – charakterystyki opisowe po usunięciu 24 przypadków

n=118	Średnia	Mediana	Moda	Liczność mody	Min	Max	Dolny kwartyl	Górny kwartyl	Odchyl. standard	Współ. zmien.
Lok	3,69	4,00	4	48	1,00	5,00	3,00	5,00		
Otocz	3,03	3,00	3	37	1,00	5,00	2,00	4,00		
SBud	3,17	3,00	3	34	1,00	5,00	2,00	4,00		
Kond	3,32	4,00	5	51	1,00	5,00	2,00	5,00		
Prawo	4,25	4,00	4	74	2,00	5,00	4,00	5,00		

Źródło: opracowanie własne.

Tabela 8

Charakterystyki opisowe zmiennej zależnej w obu modelach

	Model wykładniczy		Model potęgowy	
Charakterystyka	n=142	n=134	n=142	n=118
Średnia	2591,86	2608,50	2591,86	2554,88
Mediana	2659,91	2661,30	2659,91	2652,79
Min	1240,40	1240,40	1240,40	1240,40
Max	3878,90	3878,90	3878,90	3373,30
Dolny kwartyl	2218,71	2218,71	2218,71	2284,80
Górny kwartyl	2929,13	2929,13	2929,13	2865,69

Rozstęp	2638,50	2638,50	2638,50	2132,91
Odchyl. standardowe	553,30	532,99	553,30	460,28
Współcz. zmienności	21,35	20,43	21,35	18,02

Źródło: opracowanie własne.

W powyższej tabeli wyróżniono charakterystyki zmiennej zależnej, które nie uległy zmianie po redukcji liczby przypadków o 6% lub o 17%. Wynika z nich, że odstające przypadki nie musiały posiadać jednocześnie ekstremalnych cen, pomimo tego, że kryterium odstawania jest ściśle związane z wartością reszty, czyli różnicą ceny rzeczywistej lokalu i jej modelowej wartości.

Na podstawie tabel 4-7, można stwierdzić, że charakterystyki zmiennych objaśniających zmieniły się nieznacznie dla obu modeli, po odrzuceniu przypadków odstających. Jest to związane ze spójnością danych rynkowych zebranych wstępnie do modelowania. Nie ma wśród nich przypadków o cechach odstających.

9. Literatura

ADAMCZEWSKI Z. 2006. Elementy modelowania matematycznego w wycenie nieruchomości Podejście porównawcze. Oficyna Wydawnicza Politechniki Warszawskiej. Warszawa.

BARANSKA A. 2004. Criteria of database quality appraisement and choice stochastic models in prediction of real estate market value. FIG Working Week 2004 and 27th General Assembly in Athens, Greece, 22 – 27 May 2004.

BARANSKA A. 2005. Estymacja parametrów nieliniowych modeli funkcyjnych dla potrzeb predykcji rynkowej wartości nieruchomości. Kraków, UWND AGH, Geodezja, t.2.

BARANSKA A. 2006. Estimation of parameters of multiplicative exponential function model for real estate market value prediction. XXIII FIG Congress "Shaping the Change", Munich, Germany, October 8-13, 2006.

BARANSKA A., CZAJA J. 2007a. Statistical verification of real estate estimation models. FIG Working Week 2007 "Strategic Integration of Surveying Services", Hong Kong SAR, China, 13-17 May 2007.

BARANSKA A. 2007b. Comparing the results of function model estimation for the prediction of real estate market values in additive and multiplicative form. Kraków, UWND AGH, Geomatics and Environmental Engineering, v. 1, no 3.

BARANSKA A. 2008. Jakościowo – ilościowe metody rynkowej wyceny nieruchomości. Studia i materiały Towarzystwa Naukowego Nieruchomości, v. 16, nr 3, Olsztyn.

CZAJA J., PARZYCH P. 2007. Szacowanie rynkowej wartości nieruchomości w aspekcie międzynarodowych standardów wyceny. Kraków, Wydawnictwo Stowarzyszenia Naukowego im. ST. Staszica.

CZAJA J., PREWEDA E. 2000. Analiza statystyczna zmiennej losowej wielowymiarowej w aspekcie korelacji i predykcji. Kraków, UWND AGH,

Geodezja, t.2.
CZAJA J. 2001. Metody szacowania wartości rynkowej i katastralnej nieruchomości. Kraków.
KRYSICKI W. i in. 1986. Rachunek prawdopodobieństwa i statystyka matematyczna w zadaniach. PWN, Warszawa.
RADHAKRISHNA RAO C. 1982. Modele liniowe statystyki matematycznej. PWN, Warszawa.

MULTIPLICATIVE MODELS IN REAL ESTATE ASSESSMENT PROCESS

Anna Barańska
Department of Geomatics
AGH University of Science and Technology
e-mail: abaran@agh.edu.pl

Key words: *multi-dimensional modelling of the real estate market, multiplicative model, exponential model, power model*

Abstract

The algorithm of real estate value estimation using multidimensional function models in multiplicative form will be presented in the work. Theoretical analysis and practical examples will be presented applying two multiplicative models:

- exponential model

$$c = a_0 \cdot a_1^{x_1} \cdot a_2^{x_2} \cdot \ldots \cdot a_m^{x_m}$$

- power model

$$c = a_0 \cdot x_1^{a_1} \cdot x_2^{a_2} \cdot \ldots \cdot x_m^{a_m}$$

The models above are nonlinear both in relation to the variables and to their parameters, therefore the process of parameters assessment begins with the linearization of the model. The linearization can be done in two ways:

1) finding the logarithms of the equations by sides; we get correspondingly:
   - $\ln c = \ln a_0 + x_1 \cdot \ln a_1 + x_2 \cdot \ln a_2 + \cdots + x_m \cdot \ln a_m$
   - $\ln c = \ln a_0 + a_1 \cdot \ln x_1 + a_2 \cdot \ln x_2 + \cdots + a_m \cdot \ln x_m$
2) expansion of initial equations in a Taylor series.

The paper includes some examples of estimation of the parameters for both presented forms of multiplicative models with the results of parametric tests of significance of differences between corresponding parameters obtained from both variants.

ALGORYTM ESTYMACJI RYNKOWYCH WARTOŚCI WSKAŹNIKÓW KOSZTU WYTWORZENIA I ZUŻYCIA ŁĄCZNEGO OBIEKTÓW BUDOWLANYCH POSIADAJĄCYCH RÓŻNE OKRESY EKSPLOATACJI

Tomasz Adamczyk
Katedra Geomatyki
Akademia Górniczo - Hutnicza w Krakowie
e-mail: tomasz.adamczyk@agh.edu.pl

Janusz Dąbrowski
Instytut Inżynierii Technicznej
Państwowa Wyższa Szkoła Zawodowa w Jarosławiu
e-mail: geo-staszic@wp.pl

Słowa kluczowe: *modelowanie rynku nieruchomości*

Streszczenie

Międzynarodowe standardy wyceny wskazują na możliwość wyznaczenia rynkowej wartości nieruchomości zabudowanych według podejścia kosztowego. Założenie to pozwoliło autorowi odnieść ceny transakcyjne obiektów budowlanych, po odpowiednim rozdzieleniu z cen transakcyjnych nieruchomości zabudowanych, do kosztów wytworzenia tych obiektów.

W ramach przedstawionej pracy zaprezentowano algorytm, który pozwala na szacowanie rynkowej wartości wskaźnika kosztu wytworzenia obiektów budowlanych i rynkowej wartości wskaźnika zużycia łącznego tych obiektów określonego na jednostkę czasu eksploatacji. Zużycie łączne zawarte w algorytmie ujmuje zużycie fizyczne, funkcjonalne i zużycie zewnętrzne obiektu budowlanego.

Algorytm bazuje na parametrycznym modelu Gausa-Markowa, w którym liczba analizowanych obiektów przewyższa liczbę estymowanych parametrów.

Dla tego algorytmu została wykonana pełna analiza wariancji, która prowadzi do oceny niedokładności estymowanych parametrów.

Ilustarcja liczbowa poszczególnych etapów analizowanego algorytmu przedstawiona jest na przykładowej bazie budynków mieszkalnych jednorodzinnych.

1. Wprowadzenie

Nieruchomości zabudowane są szczególnym przypadkiem nieruchomości głównie ze względu na ich złożoność, gdyż obejmują grunt oraz części składowe gruntu, które niejednokrotnie są zabudowaniami o różnym charakterze. Składniki mienia tego typu opisywane są przede wszystkim za pomocą pola powierzchni

gruntu, powierzchni użytkowych poszczególnych budynków, powierzchni lub kubatur obiektów produkcyjnych i magazynowych, a także za pomocą atrybutów mających bezpośredni wpływ na ich wartość rynkową. Ceny transakcyjne nieruchomości zabudowanych określone w aktach notarialnych ujmują łącznie cenę gruntu oraz poszczególnych składników. Rozdzielenie cen nieruchomości zabudowanych na ceny cząstkowe gruntu i elementów zabudowy możliwe jest między innymi na podstawie analizy statystycznej rynku nieruchomości tego typu (CZAJA, PARZYCH, 2008).

Opierając się na Międzynarodowych Standardach Wyceny, które wskazują na możliwość szacowania wartości rynkowej przy wykorzystaniu podejścia kosztowego wyceny, możliwe jest określenie zależności pomiędzy cenami transakcyjnymi obiektów budowlanych możliwymi do uzyskania przy zastosowaniu przytoczonych powyżej algorytmów a kosztami wytworzenia tych obiektów. Koszty wytworzenia części składowych nieruchomości są z kolei możliwe do określenia przy znajomości parametrów obiektów budowlanych, ich elementów konstrukcyjnych i wykończeniowych. Wyznaczenie kosztów wytworzenia umożliwiają obecne na rynku periodyczne katalogi cenowe publikowane przez specjalistyczne wydawnictwa.

Budując zależności funkcyjne pomiędzy ceną transakcyjną obiektu budowlanego a jego kosztem wytworzenia, należy również uwzględnić w przypadku starszych składników mienia ich zużycie łączne obejmujące zużycie fizyczne, funkcjonalne i zewnętrzne lub każdy z rodzajów zużycia z osobna. Zużycie może być szacowane łącznie lub poprzez określenie i zsumowanie poszczególnych rodzajów zużycia. Łączne oszacowanie zużycia wymaga znajomości efektywnie uzasadnionej trwałości obiektu oraz jego wieku efektywnego, który uwzględnia aktualny poziom utrzymania obiektu oraz przeprowadzone prace remontowe wpływające na jego stan fizyczny (ŹRÓBEK, 2009).

Oprócz wymienionych parametrów powinno się także uwzględnić charakterystyczne atrybuty takie jak lokalizacja, otoczenie, komfort, itp., które mogą mieć istotny wpływ na wartość rynkową w odniesieniu do obiektu budowlanego. Dla obiektów budowlanych mających różny czas eksploatacji, a co się z tym wiąże wykazujących znacznie różne zużycia, uzasadnione wydaje się zastosowanie dodatkowego parametru jakim jest rzeczywisty czas eksploatacji takiego obiektu. Założenia dotyczące rzeczywistego czasu eksploatacji wydają się być zasadne pod warunkiem, że do modelu przyjmuje się obiekty budowlane podobne ze względu na czynniki warunkujące głównie zużycie fizyczne, czyli przede wszystkim prowadzoną gospodarkę remontową.

Ostatecznie warunek funkcyjny, w przypadku gdy wartość zużycia jest określona w funkcji kosztu wytworzenia, przyjmuje postać:

$$C_{Tj} = K_j \times u_k - K_j \times s_{zj} \times \Delta t_j \times u_{z/\Delta t} + a_1 \times u_1 + a_2 \times u_2 + \ldots + a_s \times u_s \qquad (1)$$

gdzie:

C_{Tj} - cena transakcyjna j-tego obiektu budowlanego,

K_j - koszt wytworzenia j-tego obiektu budowlanego,

s_{zj} - stopień zużycia łącznego j-tego obiektu budowlanego,

Δt_j - czas eksploatacji j-tego obiektu budowlanego,

u_k - wskaźnik rynkowy kosztu wytworzenia obiektu budowlanego,

$u_{z/\Delta t}$ - wskaźnik rynkowy zużycia łącznego na jednostkę czasu eksploatacji obiektu budowlanego,

$a_1, a_2, \ldots, a_s$ - wartości poszczególnych atrybutów w odniesieniu do obiektu budowlanego,

$u_1, u_2, \ldots, u_s$ - wskaźniki rynkowe dla poszczególnych atrybutów.

Estymowanymi parametrami w równaniu (1) są wskaźnik rynkowy kosztu wytworzenia obiektu budowlanego (u_k), wskaźnik rynkowy zużycia łącznego na jednostkę czasu eksploatacji obiektu budowlanego ($u_{z/\Delta t}$) oraz wskaźniki rynkowe dla poszczególnych atrybutów ($u_1, u_2, \ldots, u_s$).

2. Estymacja parametrów modelu

Rozpisując dla kilku obiektów budowlanych układ równań (1) w formie macierzowej otrzymuje się:

$$\begin{bmatrix} C_{T1} \\ C_{T2} \\ \vdots \\ C_{Tj} \end{bmatrix} = \left\{ \begin{bmatrix} K_1 \\ K_2 \\ \vdots \\ K_j \end{bmatrix} \begin{bmatrix} -K_1 \times s_{z1} \times \Delta t_1 \\ -K_2 \times s_{z2} \times \Delta t_2 \\ \vdots \\ -K_j \times s_{zj} \times \Delta t_j \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1s} \\ a_{21} & a_{22} & \cdots & a_{2s} \\ \vdots & \vdots & \cdots & \vdots \\ a_{j1} & a_{j2} & \cdots & a_{js} \end{bmatrix} \right\} \times \begin{bmatrix} u_k \\ u_{z/\Delta t} \\ u_{a1} \\ u_{a2} \\ \vdots \\ u_{as} \end{bmatrix} + \begin{bmatrix} \delta_{T1} \\ \delta_{T2} \\ \vdots \\ \delta_{Tj} \end{bmatrix} \quad (2)$$

gdzie:

- indeks 1÷j oznacza kolejny obiekt budowlany ze znaną ceną transakcyjną,
- indeks 1÷s oznacza kolejny atrybut w odniesieniu do obiektu budowlanego,
- w nawiasie klamrowym zawarta jest macierz blokowa.

Parametryczny model Gaussa-Markowa można zastosować do estymacji rynkowych wskaźników dla poszczególnych parametrów przy założeniu, że liczba estymowanych wskaźników jest mniejsza od liczby obiektów budowlanych o znanych cenach transakcyjnych. Liczba estymowanych wskaźników zależeć będzie w głównej mierze od liczby atrybutów cenotwórczych przyjętych do modelu, gdyż pozostałymi parametrami są wskaźniki dla kosztu wytworzenia i zużycia łącznego na jednostkę czasu eksploatacji.

Zapisany w postaci macierzy układ równań (2) przyjmuje postać:

$$[C_T] = \{[K] \quad [Z] \quad [a]\} \times [u] + [\delta_T] \tag{3}$$

gdzie:

$[C_T]$ oznacza wektor cen transakcyjnych obiektów budowlanych,

$[K]$ oznacza wektor utworzony z kosztów wytworzenia poszczególnych obiektów budowlanych,

$[Z]$ oznacza wektor utworzony z iloczynów kosztu wytworzenia, stopnia zużycia łącznego kolejnego obiektu budowlanego oraz czasu eksploatacji takiego obiektu ($Z_j = -K_j \times s_{zj} \times \Delta t_j$),

$[a]$ oznacza macierz prostokątną pionową utworzoną z wartości atrybutów rozważanych obiektów budowlanych,

$[u]$ oznacza wektor wskaźników rynkowych: kosztu wytworzenia obiektu budowlanego, zużycia łącznego na jednostkę czasu eksploatacji oraz atrybutów w odniesieniu do obiektu budowlanego,

$[\delta_T]$ oznacza wektor odchyłek losowych z modelu do cen transakcyjnych obiektów budowlanych.

Warunki wieloparametrowego modelu Gaussa-Markowa dla układu równań (3) są określone następująco:

$$E[C_T] = \{[K] \quad [Z] \quad [a]\} \times [u] \tag{4}$$

$$E[\delta_T] = \{0\} \tag{5}$$

$$E\left[[\delta_T]^T \times [\delta]\right] = V[C_T] = \sigma^2 \{I\} \tag{6}$$

Estymacja wskaźników rynkowych z układu równań (3) opiera się na minimum dla formy kwadratowej:

$$F = [\delta_T]^T \cdot [\delta_T] = \left([C_T] - \{[K] \quad [Z] \quad [a]\} \times [u]\right)^T \times \left([C_T] - \{[K] \quad [Z] \quad [a]\} \times [u]\right) = \min. \tag{7}$$

Na podstawie warunku koniecznego, czyli po wyliczeniu pochodnych cząstkowych funkcji F i przyrównaniu ich do zera, otrzymuje się układ równań normalnych postaci:

$$\left(\{[K] \quad [Z] \quad [a]\}^T \cdot \{[K] \quad [Z] \quad [a]\}\right) \times [u] = \{[K] \quad [Z] \quad [a]\}^T [C_T] \tag{8}$$

Drugie pochodne funkcji F dają macierz:

$$\left(\{[K] \quad [Z] \quad [a]\}^T \cdot \{[K] \quad [Z] \quad [a]\}\right) \tag{9}$$

Macierz ta jest dodatnio określona, przez co układ równań normalnych (8) spełnia warunki wieloparametrowego modelu Gaussa-Markowa dla układu równań (3).

Estymatory wskaźników rynkowych poszczególnych parametrów wynikają z układu równań normalnych, można je zatem zapisać wzorem:

$$[\hat{u}] = \left(\{[K] \quad [Z] \quad [a]\}^T \cdot \{[K] \quad [Z] \quad [a]\} \right)^{-1} \times \{[K] \quad [Z] \quad [a]\}^T [C_T] \tag{10}$$

oraz po rozpisaniu:

$$[\hat{u}] = \begin{Bmatrix} [K^T K] & [K^T Z] & [K^T a] \\ [Z^T K] & [Z^T Z] & [Z^T a] \\ [a^T K] & [a^T Z] & [a^T a] \end{Bmatrix}^{-1} \times \begin{Bmatrix} [K]^T \\ [Z]^T \\ [a]^T \end{Bmatrix} [C_T] \tag{11}$$

Formuła na określenie wektora odchyłek losowych modelu $\left[\hat{\delta}_T\right]$ po uwzględnieniu estymowanych wskaźników rynkowych zapisanych w postaci wektora $[\hat{u}]$ jest wyrażona wzorem:

$$\left[\hat{\delta}_T\right] = [C_T] + \{[K] \quad [Z] \quad [a]\} \times [\hat{u}] \tag{12}$$

3. Analiza wariancji

Dla zaprezentowanego modelu wariancja resztowa $\hat{\sigma}_0^2$ przyjmuje postać:

$$\hat{\sigma}_0^2 = \frac{\left[\hat{\delta}_T\right]^T \left[\hat{\delta}_T\right]}{j - R\{[K] \quad [Z] \quad [a]\}} \tag{13}$$

przy czym:

$\left[\hat{\delta}_T\right]$ — oznacza estymator wektora odchyłek losowych,

j — oznacza liczbę cen transakcyjnych nieruchomości zurbanizowanych,

$R\{[K] \quad [Z] \quad [a]\}$ — oznacza rząd bloku macierzy $\{[K] \quad [Z] \quad [a]\}$, który zależał będzie głównie od liczby przyjętych do modelu atrybutów.

Macierz kowariancji estymowanych wskaźników rynkowych wyrażona jest zależnością:

$$V[\hat{u}] = Cov[\hat{u}] = \hat{\sigma}_0^2 \cdot \left(\{[K] \quad [Z] \quad [a]\}^T \times \{[K] \quad [Z] \quad [a]\} \right)^{-1} \tag{14}$$

Określona w analizie macierz kowariancji wektora odchyłek losowych jest postaci:

$$V\left[\hat{\delta}_T\right] = Cov\left[\hat{\delta}_T\right] = \hat{\sigma}_0^2 \cdot ([I] - \{[K] \quad [Z] \quad [a]\} \times \left(\{[K] \quad [Z] \quad [a]\}^T \times \{[K] \quad [Z] \quad [a]\}\right)^{-1} \times \times \{[K] \quad [Z] \quad [a]\}^T \tag{15}$$

4. Szacowanie wartości rynkowej w oparciu o algorytm

Do szacowania wartości rynkowej obiektu budowlanego w oparciu o prezentowany algorytm, konieczne jest zapisanie w jednowierszowej macierzy kosztu wytworzenia szacowanego obiektu budowlanego, iloczyn kosztu wytworzenia, stopnia zużycia łącznego i okresu eksploatacji tego obiektu oraz wartości poszczególnych atrybutów:

$$\{[K] \quad [Z] \quad [a]\} = \left\{\overline{K}_1 \quad -\overline{K}_1 \times \overline{s}_{z1} \times \Delta\overline{t}_1 \quad \overline{a}_1 \quad \overline{a}_2 \quad \ldots \quad \overline{a}_k\right\} \tag{16}$$

przy czym:

$\overline{K}_1$ — oznacza koszt wytworzenia szacowanego obiektu budowlanego,

$\overline{K}_1 \times \overline{s}_{z1} \times \Delta\overline{t}_1$ — oznacza iloczyn kosztu wytworzenia, stopnia zużycia łącznego szacowanego obiektu budowlanego oraz jego okresu eksploatacji,

$\overline{a}_1 \quad \overline{a}_2 \quad \ldots \quad \overline{a}_k$ — oznaczają wartości atrybutów, przy założeniu że w modelu wykorzystanych jest k atrybutów.

Wartość rynkowa obiektu budowlanego przy zastosowaniu modelu wyraża się zależnością:

$$WR = \left\{\overline{K}_1 \quad -\overline{K}_1 \times \overline{s}_{z1} \times \Delta\overline{t} \quad \overline{a}_1 \quad \overline{a}_2 \quad \ldots \quad \overline{a}_k\right\} \times \begin{Bmatrix} \hat{u}_k \\ \hat{u}_{z/\Delta t} \\ \hat{u}_{a1} \\ \hat{u}_{a2} \\ \vdots \\ \hat{u}_{ak} \end{Bmatrix} \tag{17}$$

Wykorzystując macierz kowariancji dla estymowanych wartości wskaźników można szacować wariancję rynkowej wartości, a następnie określić jej przedział ufności.

Wariancja dla modelu jest wyrażona wzorem:

$$\sigma^2(WR) = \left\{ \overline{K}_i \quad -\overline{K}_i \times \overline{s}_{zi} \times \Delta\overline{t}_i \quad \overline{a}_1 \quad \overline{a}_2 \quad \ldots \quad \overline{a}_k \right\} \times Cov[\hat{u}] \times \begin{Bmatrix} \overline{K}_i \\ -\overline{K}_i \times \overline{s}_{zi} \times \Delta\overline{t}_i \\ \overline{a}_1 \\ \overline{a}_2 \\ \vdots \\ \overline{a}_k \end{Bmatrix} \tag{18}$$

Dla szacowanej wartości rynkowej części składowej gruntu można wyznaczyć symetryczny przedział ufności. Przedział ten bazuje na odchyleniu standardowym szacowanej wartości oraz kwantylu rozkładu Studenta $(t(p,\nu))$, gdzie p uwzględnia poziom ufności, natomiast ν oznacza liczbę stopni swobody:

$$\begin{Bmatrix} WR_{\max} \\ WR_{\min} \end{Bmatrix} = WR \pm \sigma(WR) \times t(p,\nu) \tag{19}$$

5. Przykład estymacji wskaźników rynkowych

Zastosowanie modelu parametrycznego do estymacji wskaźników rynkowych zilustrowano na podstawie budynków mieszkalnych jednorodzinnych. Do tego celu pozyskano informacje zawarte w aktach notarialnych. Stworzona w ten sposób baza zawierała 34 nieruchomości gruntowe zabudowane budynkami mieszkalnymi jednorodzinnymi, które były przedmiotem obrotu rynkowego. Dane pochodziły z podkrakowskiego lokalnego rynku nieruchomości – gminy Michałowice, Zielonki, Wielka Wieś i Zabierzów. Nieruchomości zawarte w bazie były przedmiotem obrotu w okresie od listopada 2005 r. do listopada 2007 r.

Na takiej bazie dokonano pierwotnie korekty cen ze względu na czas, a następnie wyestymowano jednostkowe ceny dla gruntu i budynku uwzględniając przy tym wpływ najważniejszych atrybutów cenotwórczych.

Następnie z wcześniejszej bazy wybrano 15 budynków, które wykonane były w podobnej technologii (technologia tradycyjna) oraz mających różne okresy eksploatacji i zarazem różne stopnie zużycia. Dla tych budynków obliczono ceny transakcyjne, określono stopień zużycia na podstawie bezpośrednich oględzin, obliczono koszt wytworzenia bazując na Scalonych Normatywach wyd. Wacetob, a także ustalono wartości atrybutów. Wszystkie dane zestawiono w tabeli 1.

W modelu uwzględniono dwa atrybuty, które mają największy wpływ na wartość w przypadku budynków mieszkalnych jednorodzinnych:

- Lokalizacja/Otoczenie: najbardziej korzystne (30000), bardzo korzystne (20000), korzystne (10000), przeciętne (0), niekorzystne (-10000),
- Komfort budynku: bardzo wysoki (20000), wysoki (10000), przeciętny (0), niski (-10000).

Wartości te są tak wysokie ze względu na standaryzację, aby uzyskać

porównywalne wartości w macierzy kowariancji dla wszystkich wskaźników.

Okres eksploatacji uwzględniony w przykładzie jest czasem rzeczywistym, gdyż obiekty budowlane na bazie, których dokonano estymacji parametrów modelu, charakteryzowały się podobnymi czynnikami warunkującymi przebieg zużycia, m.in. prowadzoną i planowaną gospodarką remontową.

Stopień zużycia łącznego ujmuje zużycie fizyczne i funkcjonalne wyznaczane oddzielnie, gdyż obiekty zebrane do weryfikacji algorytmu nie wykazywały zużycia zewnętrznego lub było ono zaniedbywalnie małe.

Tabela 1

Dane dla budynków do estymacji wskaźników modelu

Lp	Cena transakcyjna C_i [zł]	Koszt wytworzenia K_i [zł]	Stopień zużycia łącznego s_{zi} [%]	Okres eksploatacji wyrażony w latach	Wartość atrybutu lokalizacja/ otoczenie	Wartość atrybutu komfort
1	827 135,22	758 251,62	2,184	3	10000	20000
2	587 610,93	522 389,76	8,074	10	30000	20000
3	340 074,84	331 691,28	3,561	8	0	10000
4	383 377,47	431 869,68	27,129	25	-10000	0
5	325 381,35	233 499,32	1,710	4	20000	10000
6	324 299,14	252 128,68	2,559	6	20000	10000
7	439 847,01	335 577,44	0,666	1	30000	10000
8	250 299,95	217 812,55	0,108	0,5	-10000	20000
9	95 135,78	121 974,16	49,448	57	30000	-10000
10	190 124,36	198 184,71	24,799	21	30000	-10000
11	503 185,58	408 639,71	1,309	2	30000	10000
12	279 548,03	352 037,55	34,008	37	20000	0
13	186 426,78	277 493,42	34,547	34	10000	-10000
14	368 806,57	295 305,67	3,795	7	20000	20000
15	300 074,43	333 883,41	31,134	27	20000	-10000

Źródło: Opracowanie własne.

W tabeli 2 zebrane zostały wartości rynkowe poszczególnych wskaźników otrzymane w wyniku estymacji.

Wariancja resztowa $(\hat{\sigma}_0^2)$ estymacji modelu określona na podstawie odchyłek losowych modelu przyjmuje wartość:

$$(\hat{\sigma}_0^2) = 476922687 \Rightarrow \hat{\sigma}_0 = 21839\,zł$$

Natomiast macierz kowariancji estymowanych wskaźników przyjmuje postać:

$$Cov[\hat{u}] = \frac{1}{100000} \cdot \begin{bmatrix} 143 & 15 & -622 & -2682 \\ 15 & 3 & 3 & -347 \\ -622 & 3 & 13602 & 4412 \\ -2682 & -347 & 4412 & 75242 \end{bmatrix}$$

Tabela 2

Zestawienie wskaźników rynkowych z modelu

Parametr modelu	Wartość z estymacji
Wskaźnik rynkowy kosztu wytworzenia	1,0165
Wskaźnik rynkowy zużycia łącznego budynku na jednostkę czasu eksploatacji	0,0225
Wskaźnik rynkowy dla atrybutu „Lokalizacja/Otoczenie"	1,8080
Wskaźnik rynkowy dla atrybutu „Komfort"	2,1361

Źródło: Opracowanie własne.

Na podstawie macierzy kowariancji estymowanych wskaźników można określić odchylenia standardowe dla wszystkich wskaźników:

Tabela 3

Zestawienie wskaźników modelu i ich odchyleń standardowych

Parametr modelu	Wartość z estymacji	Odchylenie standardowe
Wskaźnik rynkowy kosztu wytworzenia	1,0165	0,0378
Wskaźnik rynkowy zużycia łącznego budynku na jednostkę czasu eksploatacji	0,0225	0,0052
Wskaźnik rynkowy dla atrybutu „Lokalizacja/Otoczenie"	1,8080	0,3688
Wskaźnik rynkowy dla atrybutu „Komfort"	2,1361	0,8674

ŹRÓDŁO: Opracowanie własne.

Wykorzystując kwantyl rozkładu Studenta dla 11 stopni swobody i poziomu istotności $(1-\alpha) = 0.95$, czyli $t(0.95; 11) = 2,20$, symetryczne przedziały ufności dla szacowanych wartości modelowych przyjmują następujące zakresy:

$$\tilde{u}_k = \hat{u}_k \pm 2,20 \times \sigma(\hat{u}_k) = 1,012 \pm 0,083$$
$$\tilde{u}_{z/\Delta t} = \hat{u}_{z/\Delta t} \pm 2,20 \times \sigma(\hat{u}_{z/\Delta t}) = 0,023 \pm 0,011$$
$$\tilde{u}_{l/o} = \hat{u}_{l/o} \pm 2,20 \times \sigma(\hat{u}_{l/o}) = 1,808 \pm 0,811$$
$$\tilde{u}_{kf} = \hat{u}_{kf} \pm 2,20 \times \sigma(\hat{u}_{kf}) = 2,136 \pm 1,908$$

Z określonych zakresów wynika, że z najniższą wiarygodnością wyznaczony

jest wskaźnik rynkowy atrybutu „komfort".

6. Wnioski

Przedstawiony algorytm ukazuje, że możliwe jest probabilistyczne modelowanie danych rynkowych i nierynkowych prowadzące do określenia najbardziej prawdopodobnych rynkowych wskaźników kosztu wytworzenia obiektu budowlanego oraz zużycia łącznego tych obiektów. Świadczyć o tym mogą między innymi uzyskane wartości wariancji resztowej modelu, odchylenia standardowe wyestymowanych wskaźników rynkowych oraz określone przedziały ufności dla poszczególnych parametrów modelu.

Wartości uzyskane w wyniku estymacji są charakterystyczne dla rynku, z którego pochodzą dane wykorzystane w procesie estymacji i zarazem charakterystyczne dla obiektów budowlanych, na bazie których dokonano modelowania. Pokazano przy tym, że powinno się przeprowadzać również pełną analizę wariancji, która obrazuje stopień zaufania do otrzymanych wyników.

Przy wykorzystaniu otrzymanych z estymacji rynkowych wskaźników możliwe jest szacowanie wartości rynkowych obiektów podobnego typu. Korzystając z modeli na podstawie których wykonany został rozdział cen transakcyjnych możliwe jest przejście na wartość rynkową nieruchomości zabudowanej poprzez dodanie części dotyczącej wartości samego gruntu. Możliwe jest przy tym określenie niedokładności otrzymanych wyników zarówno w przypadku gruntu jak i budynku, a następnie przejście na niedokładność całej szacowanej nieruchomości.

7. Literatura

Adamczewski Z., *Elementy modelowania matematycznego w wycenie nieruchomości. Podejście porównawcze.* Oficyna Wydawnicza Politechniki Warszawskiej. Warszawa 2006.

Czaja J., Parzych P., *Szacowanie rynkowej wartości nieruchomości w aspekcie międzynarodowych standardów wyceny*, Stowarzyszenie Naukowe im. St. Staszica, Kraków 2007.

Czaja J., Parzych P., *Zaawansowane modele statystyczne wyceny nieruchomości zurbanizowanych*, Studia i materiały Towarzystwa Naukowego Nieruchomości, Olsztyn 2008.

Międzynarodowe Standardy Wyceny 2007 Międzynarodowego Komitetu Standardów Wyceny (IVSC).

WACETOB, *Zasady ustalania zużycia obiektów budowlanych*, Warszawa 1997.

Źróbek R., *Zasady określenia zużycia łącznego w szacowaniu wartości budynków oraz zarządzaniu nieruchomościami*, Przegląd Geodezyjny Nr 5, 2003.

Źróbek R., *Teoretyczne i metodyczne aspekty określania wartości zużycia łącznego nieruchomości*, Studia i materiały Towarzystwa Naukowego Nieruchomości, Olsztyn 2009.

ESTIMATE ALGORITHM OF DEVELOPMENT COST AND COMBINE WEAR COST INDICATORS ASSOCIATED WITH THE MARKET VALUE OF REAL PROPERTIES WITH DIFFERENT EXPLOITATION TIME PERIODS

Tomasz Adamczyk
Department of Geomatics
AGH University of Science and Technology
e-mail: tomasz.adamczyk@agh.edu.pl

Janusz Dąbrowski
The Bronisław Markiewicz State School of Higher
Vocational Education in Jarosław
e-mail: geo-staszic@wp.pl

Key words: *valuation models*

Abstract

The international valuation standards clearly indicate existence of vital possibility that allows estimating real estate properties market value by using the cost approach. Such assumption lets the author to refer commercial objects transaction prices (after previous precise separation of all built-up properties from the transaction prices) to the cost associated with creating such objects.

The algorithm presented and used solely for the purpose of this study helps in estimating market value indicators associated with commercial facility construction cost as well as market value of combine wear such facility that is based on strictly defined time unit (period) exploitation. Combine wear used as a factor in the algorithm contains all physical, functional and external wear of a construction.

The algorithm is based on the Gauss-Markov parametric model, where the number of objects under investigation is always greater than the number of the parameters being estimated.

Complete and comprehensive variance analysis was conducted solely for the purpose of such algorithm. The results and findings of such analysis help to assess inaccuracy of all the parameters that we were trying to estimate.

The number approach to consecutive stages of the analyzed algorithm is explicitly illustrated by using one-family structure base examples.

WYKORZYSTANIE METOD TAKSONOMICZNYCH DO WYBORU OBIEKTÓW PODOBNYCH W PROCESIE WYCENY LOKALI MIESZKALNYCH

Iwona Foryś
Katedra Ekonometrii i Statystyki
Uniwersytet Szczeciński
e-mail: forys@uoo.univ.szczecin.pl

Słowa kluczowe: *analiza wielowymiarowa, obiekty podobne, wycena lokali mieszkalnych*

Streszczenie

Wśród metod wycen nieruchomości wykorzystywanych w pracy rzeczoznawcy majątkowego znalazły uznanie również metody oparte o wielokryterialne porównywanie cech nieruchomości. W praktyce oznacza to najczęściej przyjęcie kilku cech, a pozycjonowanie nieruchomości z uwagi na jedną spośród nich, przy założeniu stałości pozostałych (metoda porównywania parami). Zaproponowane w artykule analizy wielowymiarowe pozwalają na wybór nieruchomości podobnych z uwagi na wiele cech, a w grupie nieruchomości najbardziej podobnych zastosowanie uproszczonych rachunków wyznaczania ceny przeciętnej.

Rozważania oparto na dwóch procedurach porządkowania wielowymiarowego: taksonomicznej mierze atrakcyjności inwestycji (TMAI) oraz syntetycznym mierniku rozwoju z odległością miejską, których zalety w procesie wyceny pokazano na przykładzie spółdzielczych lokali mieszkalnych. Dokonano również oceny podobieństwa klasyfikacji wybranymi metodami.

1. Wstęp

Szacowanie wartości rynkowej nieruchomości wymaga od rzeczoznawcy majątkowego przeprowadzenia wszechstronnej analizy rynkowej mającej na celu między innymi zidentyfikowanie czynników wpływających na pozycję rynkowa szacowanej nieruchomości. Oprócz katalogu cech, które w procesie wyceny, zgodnie z przepisami prawa, rzeczoznawca majątkowy zobligowany jest uwzględnić, są również cechy specyficzne dla każdego rodzaju nieruchomości, a nawet dla lokalnych rynków nieruchomości. O ich przyjęciu lub nie decyduje arbitralnie rzeczoznawca na podstawie przeprowadzonej analizy lokalnego rynku, znajomości ogólnych mechanizmów rynkowych, a przede wszystkim zobiektywizowaniu indywidualnych preferencji uczestników obrotu na danym rynku rodzajowym. Zgodnie z wymogami prawa wybór nieruchomości podobnych determinuje szacowane prawo do lokalu, okoliczności zawieranej umowy sprzedaży oraz znajomość cen transakcyjnych. Wybór cech

porównawczych jest elementem sztuki wyceny i znajomości rynku rodzajowego.

W przypadku lokali mieszkalnych wybór nieruchomości porównawczych jest zdeterminowany „lokalnością" tego rynku, którą należy utożsamiać z najbliższym sąsiedztwem, osiedlem czy dzielnicą miasta. Częstość zawieranych transakcji na rynku mieszkaniowym nie stwarza bariery braku nieruchomości do porównania. Trudności w pozyskaniu informacji mogą wynikać jedynie z rodzaju praw, które są przedmiotem obrotu. W przypadku własności czy współwłasności nieruchomości lokalowej o funkcji mieszkalnej umowa przeniesienia praw znajdzie się w zasobie ewidencyjnym właściwego miejscowo wydziału katastru nieruchomości, natomiast umowy dotyczące przeniesienia praw spółdzielczych najczęściej znajdują się jedynie w aktach członkowskich spółdzielni mieszkaniowych. W tym przypadku szczególnie istotne jest w pierwszym kroku określenie jednolitego rodzajowo obszaru, który będzie przedmiotem szczegółowej analizy. Należy w ty miejscu podkreślić, że wybór na tym etapie nie dotyczy tak naprawdę nieruchomości lokalowych, które będą przyjęte do porównania ale nieruchomości zabudowanych budynkami wielolokalowymi o funkcjach mieszkalnych. Czyli nieruchomości budynkowe, a przede wszystkim technologia wznoszenia, wiek budynków i ich stan techniczny zbliżony do budynku, w którym znajduje się szacowany lokal mieszkalny decydują w pierwszym etapie o wyznaczonym przestrzennie lokalnym rynku. Charakter zabudowy polskich miast, będący wielowiekowym dorobkiem urbanistycznym, a także zakłócenia w ładzie przestrzennym jakie wprowadziła gospodarka nakazowa w realizacji inwestycji mieszkaniowych, w wielu przypadkach ułatwiają określenie zasięgu przestrzennego prowadzonej analizy. Są to bardzo często jednolite w zabudowie osiedla, czy zespoły budynków, wzniesione w podobnej technologii, z mieszkaniami w tym samym standardzie, niejednokrotnie znormalizowane w efekcie restrykcyjnej polityki mieszkaniowej państwa. W przypadkach zabudowy mieszanej, z różnych okresów a tym samym wznoszonych w różnych technologiach o doborze zadecyduje „podobieństwo" lokalizacji obiektów w stosunku do założonych priorytetów (odległość od centrum, sąsiedztwo terenów rekreacyjnych lub innych o specyficznej funkcji, dogodny dojazd, lokalizacja obiektów handlowych, oświatowych, czy wreszcie moda na daną dzielnicę).

Podobieństwo obiektów na potrzeby prowadzonych w artykule rozważań należy utożsamiać z brakiem różnic w zestawie cech obiektów przyjętych do analizy albo istnieniem jednolitej skali przekształcenia obiektów w podobne. Potocznie są to obiekty najmniej się różniące, czyli różniące się najmniejszą liczbą cech spośród zestawu wybranych cech porównawczych decydujących o ich wartości rynkowej (minimalizacja liczby cech różnicujących dane obiekty). Obiekty porównawcze w skrajnej sytuacji z uwagi na wybrany zestaw cech mogą być „takie same" lub „identyczne" ale nigdy nie „te same" czyli tożsame. W procesie wyceny obiekty „identyczne" powinny mieć oszacowaną tą samą wartość rynkową. Jednak w rzeczywistości nieruchomości nie są identyczne, chociaż

porównywane z uwagi na określony zestaw cech mogą stwarzać takie wrażenie. Często jednak cechy je różnicujące obiekty są na tyle subtelne lub są odbiciem tak indywidualnych preferencji nabywców, których rzeczoznawca majątkowy nie jest w stanie wyspecyfikować, że pozostaje zwrócenie jedynie uwagi na „niemierzalne czynniki" wpływające na rozbieżność cen transakcyjnych. Matematycznie podobieństwo jest przekształceniem geometrycznym (funkcją podobieństwa), które w określonej przestrzeni metrycznej zachowuje stosunek odległości punktów. Cechą każdej przestrzeni metrycznej jest zbiór punktów i sposób pomiaru odległości pomiędzy nimi. W praktyce wyznacza się „funkcję niepodobieństwa" obiektów szukając odległości pomiędzy obiektami. Im większa odległość tym większe niepodobieństwo obiektów. W procesie szacowania wartości nieruchomości porównując nieruchomości operujemy jednocześnie kilkoma skalami pomiaru (od nominalnej, po ilorazową), nie dokonując ani normalizacji ani ujednolicenia zmiennych. Obiekty są podobne jeżeli istnieje podobieństwo (przekształcenie) przeprowadzające jeden obiekt w drugi obiekt, czyli istnieje skala zachowująca stosunek „odległości" cech tych obiektów. Takie podejście oznaczałoby zastosowanie jednej skali podobieństwa dla wszystkich cech (jednej skali pomiaru lub jej normalizacji), którymi opisujemy obiekty podobne. W teorii ekonomiki nieruchomości podkreśla się fakt „niepodobieństwa" nieruchomości oraz brak nieruchomości identycznych (KUCHARSKA-STASIAK 1999). Zatem rzeczoznawca majątkowy wybiera obiekty porównywalne z uwagi na wybrane cechy, szukając obiektów najmniej „niepodobnych". Problemy z wyborem obiektów porównywalnych pozwalają rozwiązać wielowymiarowe analizy porządkowania liniowego obiektów, w których przeprowadzona jest zarówno normalizacja jak i ujednolicanie zmiennych, a w efekcie również porządkowanie obiektów w stosunku do wybranego wzorca.

Odrębnej dyskusji wymaga zestaw cech „indywidualnych mieszkania", które niejednokrotnie przesądzają o decyzjach kupujących, a są nieidentyfikowalne dla rzeczoznawcy majątkowego. Przyjmując nieruchomości lokalowe do porównania rzeczoznawca majątkowy nie ma możliwości „wejścia" i oceny lokali, które zakwalifikował jako obiekty podobne. Jego wiedza na temat wspomnianych nieruchomości ogranicza się do stanu prawnego lokalu, informacji pozyskanych od zarządcy nieruchomości z dokumentacji technicznej obiektu (w tym rzutu lokali przyjętych do porównania, ich usytuowania w budynku, wyposażenia technicznego) oraz własnej oceny nieruchomości ale tylko budynkowej. Nie jest w stanie ocenić walorów poszczególnych lokali w aspekcie ich utrzymania, ulepszeń, aranżacji powierzchni, które w większości przypadków decydują o wyborach kupujących. Z tego powodu można zaproponować kilka grup cech będących podstawą oceny podobieństwa lokali mieszkalnych na potrzeby szacowania wartości tych nieruchomości:

- lokalizacja (położenie) w odniesieniu do lokalnych centrów miejskich, obiektów użyteczności publicznej, oświatowych, centrów usługowo-handlowych, dostępności komunikacji miejskiej;

- stan techniczny budynku (wiek budynku, zużycie techniczne i funkcjonalne, technologia wznoszenia, kubatura, udziały powierzchni funkcjonalnych w budynku, wyposażenie w instalacje techniczne, modernizacje);
- sąsiedztwo nieruchomości budynkowej (usytuowanie budynku wewnątrz osiedla, zagospodarowanie terenu w sąsiedztwie, dojazd do posesji, dostępność miejsc parkingowych, bliskość obiektów o niekorzystnym oddziaływaniu na sąsiedztwo albo o walorach atrakcyjnych dla funkcji mieszkalnej);
- lokal mieszkalny (powierzchnia, liczba i struktura pomieszczeń, układ funkcjonalny-rozmieszczenie pomieszczeń na planie mieszkania, pomieszczenia przechodnie, ograniczenia konstrukcyjne- skosy, filary, brak światła dziennego, układ lokalu w stosunku do stron świata, usytuowanie na kondygnacji, przynależny taras, loggia, balkon lub inne pomieszczenia o funkcjach wspomagających, dodatkowe walory użytkowe nieruchomości lub ograniczenia jak liczba mieszkańców w budynku, wysokość opłat eksploatacyjnych jako wskaźnik kosztów utrzymania gotowości użytkowej lokalu);

Wybór konkretnych cech zależy od subiektywnej oceny nieruchomości dokonanej przez rzeczoznawcę majątkowego, w oparciu o wiedzę na temat preferencji nabywców na lokalnym rynku mieszkaniowym.

2. Taksonomiczna miara atrakcyjności inwestycji (TMAI) oraz odległość miejska w metodach porządkowania liniowego

Wybór metody analizy wielowymiarowej zależy od charakteru prowadzonych badań, dostępności metody w formie gotowych pakietów obliczeniowych oraz użyteczności metody w konstruowaniu zestawu nieruchomości porównawczych. Zaproponowane taksonomiczne (syntetyczne) mierniki rozwoju, wykorzystywane do porządkowania liniowego obiektów, są jednymi z najbardziej popularnych metod klasyfikacji obiektów. Punktem wyjścia w metodach taksonomicznych jest konstrukcja macierzy obserwacji $\mathbf{X} = [x_{ij}]$,w której i=1,2,...,n badanych obiektów (nieruchomości) oraz j=1,2,...m zmiennych czyli cech diagnostycznych (atrybutów nieruchomości, które mogą być symulantami, destymulantami lub nominatami dla badanego zjawiska). Typowe formuły zamiany destymulant na stymulanty można zapisać wzorami:

- ilorazowa $x_{ij} = b \cdot D_{ij}^{-1}$, gdzie $b>0$ (1)
- różnicowa $x_{ij} = a - c \cdot D_{ij}^{-1}$, gdzie $c>0$ (2)

oraz D_{ij} jest wartością j-tej destymulanty zaobserwowaną w i –tym obiekcie, a – stała (w szczególnych przypadkach $a=0$ lub $a = \max_i \{D_{ij}\}$, b –stała (w szczególnych przypadkach $b = \min_i \{D_{ij}\}$ lub b=1), c –stała (w szczególnych przypadkach $b=1$). Zamiany nominat na stymulanty (na skali ilorazowej) można dokonać

wykorzystując wzory według formuły (WALESIAK 1996):

- ilorazowej
$$x_{ij} = \begin{cases} \frac{N_{ij}}{n_j} & (N_{ij} \le n_j) \\ \frac{n_{ij}}{N_{ij}} & (N_{ij} > n_j) \end{cases} \quad (3)$$

- różnicowej
$$x_{ij} = \begin{cases} N_{ij} - n_j & (N_{ij} \le n_j) \\ n_{ij} - N_j & (N_{ij} > n_j) \end{cases} \quad (4)$$

gdzie N_{ij} to wartość j –tej nominanty zaobserwowana w *i* –tym obiekcie, a n_j jest nominalnym poziomem *j* –tej zmiennej.

Syntetyczny miernik rozwoju (NOWAK 1990) jest funkcją przekształcającą wektor **X** w wektor **Z**=f(x_1, x_2, ..., x_m)=[z_i], gdzie i=1,2,...,n są syntetycznymi miernikami rozwoju dla badanych obiektów. Metody wyznaczania syntetycznych mierników różnią się sposobem wyznaczania odległości między badanymi obiektami a obiektem wzorcowym oraz metodami normalizacji zmiennych.

Obiekt wzorcowy to obiekt posiadający najlepsze współrzędne spośród wartości znormalizowanych macierzy Z, najczęściej są to wartości maksymalne z_{01}, z_{02}, ..., z_{0m} spośród {z_{ij}}. W kolejnym etapie wyznaczana jest odległość każdego badanego obiektu od wzorca, które w artykule zostanie wykonane z wykorzystaniem formuły (WALESIAK 1996):

- odległość euklidesowa
$$d_i = \left[\frac{1}{m} \sum_{j=1}^{m} \alpha_j \left(z_{ij} - z_{oj} \right)^2 \right]^{\frac{1}{2}} \quad (5)$$

- odległość miejska
$$d_i = \frac{1}{m} \sum_{j=1}^{m} \alpha_j \left| z_{ij} - z_{oj} \right| \quad (6)$$

gdzie z_{0j} – to *j*-ta współrzędna obiektu wzorca, z_{ij} – znormalizowana wartość *j*-tej zmiennej w i- tym obiekcie (przyjęto jednakowe wagi zmiennych, czyli w_j=1*)*. TMAI (taksonomiczna miara akcyjności inwestycji) wykorzystuje odległość euklidesową oraz poniższy proces normalizacji i standaryzacji zmiennych (TARCZYŃSKI 1994, FORYŚ 2009). Dla porównania w podobnej procedurze wykorzystano metrykę miejską do wyznaczenia SMR (syntetycznego miernika rozwoju). Pozbawienie zmiennych mian oraz ujednolicenie rzędów wielkości wyników pomiaru (dla skali przedziałowej i ilorazowej) należy wykonać wykorzystując normalizację zmiennych. W tym celu można zastosować wzór na normalizację zmiennych poprzez standaryzację:

$$z_{ij} = \frac{x_{ij} - \overline{x_j}}{S_j} \quad (7)$$

gdzie S_j –odchylenie standardowe, $\overline{x_j}$ -wartość średnia *j*-tej zmiennej.

Wyznaczone wzorem (5) lub (6) odległości jako miary syntetyczne nie są

unormowane. Normalizacja pozwala na zmianę preferencji zmiennej –wartości większe świadczą o wyższym poziomie badanego zjawiska. W tym celu należy wyznaczyć:

$$z_i = 1 - \frac{d_i}{d_0} \tag{8}$$

gdzie $d_0 = \overline{d} + a \cdot S_d$ oraz $a = \frac{d_{i\max} - \overline{d}}{S_d}$, natomiast $\overline{d}$ -średnia arytmetyczna, d_{imax} –maksymalna wartość a S_d odchylenie standardowe zmiennej d_i. Im bliższą jedności osiąga wartość syntetyczny miernik określony wzorem (8) tym obiekt jest bliższy wzorcowi.

Ostatni etap analizy polega na grupowaniu obiektów z uwagi na wyznaczony miernik rozwoju. Liczbę klas można przyjąć arbitralnie lub wykorzystując inne dostępne (opracowane w pakietach obliczeniowych) metody grupowania. W literaturze przedmiotu proponuje się podział na cztery klasy wykorzystując średnią arytmetyczną lub średnią arytmetyczną i odchylenie standardowe (w uzasadnionych przypadkach można wykorzystać również zmienne pozycyjne). Dla wzorca zmiennego można przyjąć podział na cztery grupy według wzoru (ZELIAŚ 2000):

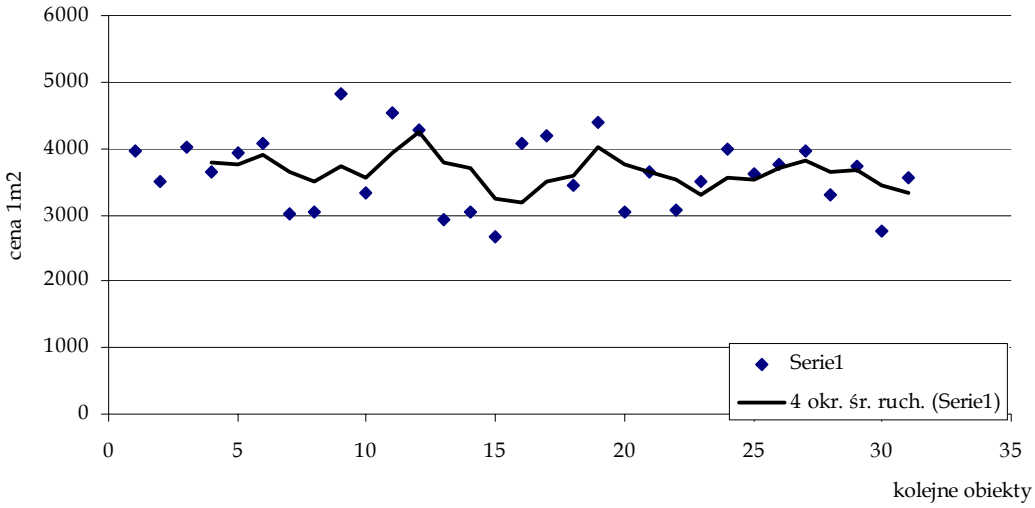
$$\begin{aligned} &G1: z_i \in \left\langle \bar{z} + S_z ; \max_i \{z_i\} \right\rangle, G2: z_i \in \left\langle \bar{z}; \bar{z} + S_z \right\rangle \\ &G3: z_i \in \left\langle \bar{z} - S_z ; \bar{z} \right\rangle, G4: z_i \in \left\langle \min_i \{z_i\}; \bar{z} - S_z \right\rangle \end{aligned} \tag{9}$$

gdzie z_i -zmienna syntetyczna, $\bar{z}$ -średnia arytmetyczna oraz S_z –odchylenie standardowe zmiennej syntetycznej.

3. Ocena podobieństwa wyników porządkowania liniowego zbioru obiektów z wybranymi formułami wyznaczania odległości obiektów od wzorca

Przedstawioną procedurę budowy syntetycznej miary odległości zastosowano na potrzeby wyboru obiektów podobnych w procesie szacowania wartości spółdzielczego prawa do lokalu mieszkalnego. Ponieważ przedmiotowy lokal znajduje się na osiedle Stare Miasto w Stargardzie Szczecińskim, do analizy przyjęto 31 obiektów (30 obiektów, którymi nastąpił obrót w 2009 roku w formie transakcji rynkowych oraz 1 obiekt szacowany na listopad 2009 roku), które oznaczono symbolem O_i, gdzie i= *1,2,...31*. Każdy obiekt opisano zestawem dwunastu zmiennych oznaczonych symbolem x_j, gdzie j= *1, 2,..., 12*. Zmienne reprezentują cechy nieruchomości lokalowych, które można przedstawić na skali ilorazowej (najmocniejszej) i do niej wybrać odpowiednie procedury taksonomiczne. Są to odpowiednio:

x_1 odległość od szkoły podstawowej rejonowej (km),
x_2 odległość od lokalnego centrum handlowo-usługowego (km),
x_3 udział liczby mieszkańców budynku w liczbie mieszkańców osiedla (%),
x_4 miesięczne opłaty eksploatacyjne za użytkowanie lokalu (zł/m^2),
x_5 zużycie techniczne budynku (%),
x_6 zużycie moralne ocieplenia budynku (%),
x_7 liczba miesięcy od daty sprzedaży do daty szacowania (miesiące),
x_8 powierzchnia użytkowa lokalu mieszkalnego (m^2),
x_9 cena transakcyjna 1m^2 powierzchni użytkowej lokalu mieszkalnego (zł/m^2),
x_{10} kondygnacja, na której znajduje się lokal mieszkalny,
x_{11} liczba pokoi w lokalu mieszkalnym.



Rys 1. Dynamika zmian cen transakcyjnych 1 m^2 powierzchni użytkowej spółdzielczego prawa do lokalu mieszkalnego na osiedlu Stare Miasto w Stargardzie Szczecińskim w 2009 roku. *Źródło*: opracowanie własne.

Zużycie techniczne budynku wyznaczono w oparciu o informacje uzyskane od zarządcy obiektów (SM w Stargardzie Szczecińskim), oszacowane na podstawie przeglądów budowlanych. Zużycie moralne ocieplenia budynku przyjęto jako relację „wieku ocieplenia" w latach do stałej wielkości 20 lat, po których upływie (od 1989 roku) zmieniły się w sposób istotny wymagane parametry cieplne budynków oraz technologia i materiały docieplenіowe. Udział liczby mieszkańców budynku w liczbie mieszkańców osiedla odzwierciedla komfort zamieszkiwania w budynku, w tym między innymi oddaje zapotrzebowanie na miejsca parkingowe. Spośród zidentyfikowanych atrybutów lokali na tym etapie analizy wyeliminowano zmienną x_7 oraz x_9. Nie zauważono wyraźnego trendu zmiany

ceny 1m² w analizowanym okresie, a przynajmniej takiego który dałby się opisać dobrze dopasowaną funkcją o prostej postaci analitycznej, co widoczne jest na rysunku 1. W 2009 roku najwyższą cenę uzyskano w marcu za lokal o powierzchni 49,85 m² (4814 zł/m²) w budynku niskim, a najniższą w kwietniu za lokal na 7 kondygnacji o powierzchni 46,76 m², w budynku wysokim, (2673 m²). Oba lokale miały dwa pokoje, znajdowały się w budynkach o przeciętnym zużyciu technicznym i podobnej technologii wznoszenia ale istotnie różniły się z uwagi na zużycie moralne ocieplenia budynku i odległość od lokalnego centrum handlowo-usługowego.

Natomiast zmienna x_9 będzie podstawą ostatecznych wyliczeń, a na etapie analizy taksonomicznej byłaby zmienną z brakującymi dnami (dla obiektu szacowanego). Kolejna weryfikacja zestawu zmiennych oparta jest na wartości współczynnika zmienności, który powinien gwarantować dużą zmienność klasyfikowanych obiektów. W tabeli 1 zaprezentowano podstawowe statystyki opisowe analizowanych zmiennych.

Tabela 1

Podstawowe statystyki opisowe analizowanych zmiennych

Statystyki	X_1	X_2	X_3	X_4	X_5	X_6	X_8	X_{10}	X_{11}
Maximum	0,500	1,150	0,031	0,610	47,78	100	98,91	9,0	4,0
Minimum	0,200	0,200	0,002	0,120	12,22	25	25,09	1,0	1,0
Rozstęp	0,300	0,950	0,029	0,490	35,56	75	73,82	8,0	3,0
Średnia	0,308	0,735	0,013	0,283	29,27	62,9	42,58	3,42	2,0
Odchylenie standardowe	0,130	0,315	0,007	0,098	8,581	23,56	15,006	2,046	0,817
Wspólczynnik zmienności	42,34	42,90	54,52	34,46	29,32	37,44	35,24	59,83	40,82
Współczynnik asymetrii	0,714	-0,514	0,578	0,983	0,418	-0,281	2,006	1,010	0,786

Źródło: opracowanie własne.

Dla obiektu szacowanego (O31) poszczególne zmienne przyjmują wartości:

X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}
0,3	0,68	0,01	0,13	32,3	55	**0**	46,73	**3,552**	5	3

Ostateczna weryfikacja zmiennych polega na wyeliminowaniu ze zbioru zmiennych wysoko skorelowanych (zmienna x_{11}). W efekcie pozostał zbiór cech, które podzielono na: stymulanty $\{x_1, x_2, x_8\}$ oraz destymulanty $\{x_3, x_4, x_5, x_6,\}$ w zależności od wpływu zmiennej na cenę transakcyjną lokalu. Dokonano zamiany destymulant na stymulanty korzystając z formuły różnicowej dla zmiennych mierzonych na skali ilorazowej (wzór 2, gdzie $a = \max_i \{D_{ij}\}$). Tak uzyskany zbiór zmiennych dla badanych obiektów podlegał następnie standaryzacji (wzór 7). W kolejnym etapie wyznaczono odległości obiektów od wzorca (obiekt o wartościach maksymalnych dla każdej analizowanej zmiennej)

według dwóch zaproponowanych miar (wzory 5 oraz 6) a następnie unormowano uzyskane zmienne syntetyczne (wzór 8). Uzyskane wartości syntetycznych mierników wraz z grupowaniem obiektów (wzór 9) zaprezentowano w tabeli 2.

Tabela 2

Syntetyczne miary rozwoju z odległością euklidesową (TMAI) oraz miejską wraz z grupowaniem obiektów

Obiekt	TMAI (metryka euklidesowa)	TMAI grupowanie	SMR (metryka miejska)	SMR grupowanie
O1	0,904	O11	0,850	O11
O2	0,711	O16	0,741	O16
O3	0,822	O20	0,895	O29
O4	0,748	O9	0,731	O19
O5	0,831	O29	0,806	O9
O6	0,677	O6	0,693	O6
O7	0,863	**O31**	0,907	O20
O8	0,914	O2	0,950	O4
O9	0,670	O24	0,684	O2
O10	0,844	O19	0,856	**O31**
O11	0,425	O4	0,351	O18
O12	0,926	O18	0,964	O30
O13	0,842	O25	0,874	O5
O14	0,854	O22	0,914	O24
O15	0,806	O23	0,854	O22
O16	0,648	O15	0,614	O21
O17	1,000	O3	1,000	O1
O18	0,760	O5	0,769	O25
O19	0,747	O13	0,663	O15
O20	0,649	O10	0,724	O10
O21	0,855	O14	0,839	O13
O22	0,788	O26	0,819	O26
O23	0,801	O21	0,888	O23
O24	0,734	O7	0,816	O3
O25	0,771	O28	0,853	O28
O26	0,855	O27	0,884	O7
O27	0,894	O1	0,909	O27
O28	0,872	O8	0,900	O14
O29	0,671	O12	0,658	O8
O30	0,948	O30	0,779	O12
O31	0,688	O11	0,745	O17

ŹRÓDŁO: opracowanie własne.

Ostatecznie szacowany obiekt O31 znajduje się w grupie obiektów podobnych odpowiednio:

- dla metryki euklidesowej {O2, O4, O6, O18, O19, O22, O24, O25},
- dla metryki miejskiej {O2, O4, O6, O9, O18, O20, O30}.

W tak wyznaczonych klasach obiektów podobnych, przy braku trendu zmian cen transakcyjnych, wystarczy wyznaczyć średnią cenę 1 m^2 powierzchni użytkowej lokalu, która wynosi odpowiednio:

- dla metryki euklidesowej $x_{śr}$=3714 zł/m^2 (przeszacowana o 4,3% rzeczywistej ceny transakcyjnej)
- dla metryki miejskiej $x_{śr}$=3579 zł/m^2 (przeszacowana o 0,7% rzeczywistej ceny transakcyjnej),

gdzie uzyskana cena transakcyjna w listopadzie 2009 roku wyniosła 3552 zł/m^2. Średnia wartość transakcyjna z całej próby wyniosła 3639 zł/m^2, stąd posłużenie się całą 30-elementową próbą oznaczałoby błąd oszacowania na poziomie 2,4%.

Powtarzając grupowanie metodą Warda (dostępną w pakiecie Statistica 7.0) dla każdej ze stosowanych metryk uzyskano większy błąd oszacowania wartości nieruchomości niż stosując metody porządkowania liniowego (w metodzie Warda dla metryki miejskiej błąd wyniósł 6,7%, a dla metryki euklidesowej 8,4%).

4. Podsumowanie

Przedstawiona w artykule analiza wskazuje na przydatność metod taksonomicznych, a zwłaszcza porządkowania liniowego do wyboru nieruchomości podobnych w procesie szacowania wartości spółdzielczego prawa do lokalu mieszkalnego. Wybór zmiennych wynikał z dostępności, ich obiektywnego pomiaru oraz wymogów formalnych dotyczących wysokiej zmienności i niskiej korelacji. Zaproponowana metoda syntetycznego miernika rozwoju dla dwóch wariantów wyznaczania odległości wskazuje na lepsze wyniki oszacowania przy wykorzystaniu metryki miejskiej niż metryki euklidesowej w metodzie TMAI. Metody porządkowania liniowego pozwalają na obiektywną ocenę podobieństwa obiektów porównawczych z uwagi na wiele cech, a w efekcie dobór tych z nich, które zapewnią najmniejszy błąd szacowania, mimo wykorzystania w obrębie wyznaczonej klasy tylko średniej. Metoda zyskuje w praktyce, gdy zbiór obiektów jest większy, podobnie jak liczba cech opisujących każdy obiekt. Pozwala również na ominięcie subiektywnej oceny wpływu poszczególnych cech na wartość szacowanej nieruchomości.

5. Literatura

FORYŚ I. 2009. *Społeczno-gospodarcze uwarunkowania rozwoju rynku mieszkaniowego. Ujęcie ilościowe.* (maszynopis w druku), Wydawnictwo Uniwersytetu Szczecińskiego, Szczecin.

KUCHARSKA-STASIAK E. 1999. *Nieruchomość a rynek*, PWN Warszawa.

NOWAK E. 1990. *Metody taksonomiczne klasyfikacji obiektów społeczno-gospodarczych,* PWE, Warszawa.
TARCZYŃSKI W. 1994. *Taksonomiczna miara atrakcyjności inwestycji w papiery wartościowe,* Przegląd Statystyczny Nr 3, Warszawa.
WALESIAK M. 1996. *Metody analizy danych marketingowych*, PWN Warszawa.
WALESIAK M. 2002. *Uogólniona miara odległości w statystycznej analizie wielowymiarowej,* Wydawnictwo Akademii Ekonomicznej we Wrocławiu, Wrocław.
ZELIAŚ A. 2000. *Taksonomiczna analiza przestrzennego zróżnicowania poziomu życia w Polsce w ujęciu dynamicznym*, Kraków.

*****

THE MULTIVARIATE ANALYSIS USING TO THE CHOICE THE SIMILAR OBJECT IN THE HOUSING VALUATION PROCESS

Iwona Foryś
Department of Econometric and Statistic
University of Szczecin
e-mail: forys@uoo.univ.szczecin.pl

Key words: *the multivariate analysis, similar object, the housing valuation*

Abstract

Among many methods of property valuation the estate valuers also recommend the use of multivariate methods to compare property attributes. In practice it means taking into account a few attributes and positioning the property with one of them in mind assuming that the remaining attributes are constant. The proposed multivariate analyses make it possible to choose the properties whose many attributes are similar and, in a group of most similar properties, to apply the simplified average price valuation. The considerations were based on two ranking procedures: a) the classification according to the variable (TMAI) norm and b) the municipal distance. Their advantage in the valuation process is shown on the example of cooperative housing. The authors also evaluated the similarity of classification by means of the chosen methods.

METODYKA AKTUALIZACJI CEN NA PRZYKŁADZIE NIERUCHOMOŚCI LOKALOWYCH

Tomasz Budzyński
Zakład Katastru i Gospodarki Nieruchomościami
Politechnika Warszawska
e-mail: t.budzynski@gik.pw.edu.pl

Słowa kluczowe: aktualizacja cen, modele regresji, sztuczne sieci neuronowe

Streszczenie

W procesie wyceny nieruchomości rzeczoznawca majątkowy zobowiązany jest do aktualizacji cen transakcyjnych na datę wyceny. Prawidłowo wykonana aktualizacja jest szczególnie ważna w przypadku istotnie zauważalnych zmian poziomu cen na rynku nieruchomości w badanym okresie czasu.

W pracy przedstawiono metody określania wpływu daty transakcji na ceny nieruchomości oraz ich aktualizacji. Rzeczoznawcy majątkowi, obliczając współczynnik zmiany cen, porównują ceny nieruchomości identycznych lub bardzo podobnych sprzedanych w różnych okresach czasu.

W przypadku licznych zbiorów cen mogą zastosować metody statystyczne m.in. wyznaczając linię trendu o postaci liniowej lub nieliniowej. W celu aktualizacji cen na datę wyceny można zastosować również bardzo zaawansowane metody np. sztuczne sieci neuronowe.

W pracy umieszczono przykłady zastosowania ww. opisanych metod dla potrzeb aktualizacji cen. Obiektami badań były 2 rynki lokalne nieruchomości lokalowych obejmujące: dzielnicę miasta Płocka oraz dzielnicę m. st. Warszawy. Wyniki uzyskane przy zastosowaniu wspomnianych metod zostały omówione jak również dokonano ich porównania.

1. Wstęp

Rzeczoznawca majątkowy, określając wartość nieruchomości, zobowiązany jest do aktualizacji cen transakcyjnych na datę wyceny. Jej prawidłowe przeprowadzenie jest szczególnie ważne w przypadku istotnie zauważalnych zmian poziomu cen na rynku nieruchomości w badanym okresie czasu.

Aktualizacja cen obok określenia cech rynkowych i ich wpływu na zróżnicowanie poziomu cen jest jednym z elementów procesu analizy rynku nieruchomości. Proces ten wymaga zgromadzenia danych o cenach transakcyjnych nieruchomości podobnych a także ich charakterystyk pod kątem cech rynkowych. Przyjęty sposób aktualizacji cen transakcyjnych powinien zależeć od liczby zebranych cen (PRYSTUPA 2001).

2. Metody określania wpływu czasu na poziom cen

W celu określenia wpływu czasu na poziom cen oraz ich aktualizacji stosowanych jest kilka metod omówionych poniżej. W przypadku bazy liczącej co najwyżej kilkanaście cen transakcyjnych często stosowaną przez rzeczoznawców majątkowych metodą określenia wpływu czasu na poziom cen jest metoda porównania cen nieruchomości podobnych w parach. Polega ona na porównaniu cen transakcyjnych nieruchomości ocenionych pod kątem cech rynkowych identycznie lub bardzo podobnie a będących przedmiotem sprzedaży w różnych okresach czasu.

Zagadnienie aktualizacji cen z wykorzystaniem tej metody można opisać z wykorzystaniem następujących wzorów:

$$W_{z,i} = \frac{C_{p,i} - C_{w,i}}{C_{w,i}} \bullet \frac{100\%}{t} \tag{1}$$

$$W_{z,c} = \frac{\sum_{1}^{n} W_{z,i}}{n} \tag{2}$$

$$C_a = C_{nom} \bullet (1 + \frac{W_{z,c}}{100\%} \bullet m) \tag{3}$$

gdzie:

$W_{z,i}$ - jednostkowy współczynnik zmiany cen obliczony dla jednej *i*-tej pary nieruchomości podobnych,

$C_{w,i}$, $C_{p,i}$ - cena nieruchomości sprzedanej wcześniej oraz później w *i*-tej parze,

t - liczba jednostek czasu pomiędzy datami sprzedaży nieruchomości w *i*-tej parze,

n - liczba par porównawczych – co najmniej kilka,

$W_{z,c}$ - współczynnik zmian cen na danym rynku,

C_{nom}, C_a - cena nominalna (transakcyjna) i zaktualizowana na datę wyceny,

m - liczba jednostek czasu od daty transakcji do daty wyceny.

Jeśli lokalny rynek nieruchomości jest dobrze rozwinięty to możliwym jest utworzenie bazy zawierającej co najmniej kilkadziesiąt transakcji. Wówczas wpływ czasu na poziom cen możemy z dużym powodzeniem oszacować z wykorzystaniem modeli statystycznych. Budując je zakładamy, że inne czynniki wpływające na poziom cen nie są istotne dla ogólnej tendencji.

Jeśli zmiany cen w czasie są w przybliżeniu liniowe stosuje się model liniowej regresji:

$$c = a + b \bullet t \tag{4}$$

gdzie:
c - jednostkowa ceny nieruchomości
a, b - parametry modelu
t - data transakcji (kolejna jednostka czasu wyrażona liczbą)

Estymacja parametrów modelu opisanego równaniem (4) (oszacowanie współczynnika kierunkowego i stałej) odbywa się najczęściej metodą najmniejszych kwadratów. Współczynnik kierunkowy wyraża zmianę cen przypadającą na jednostkę czasu np. miesiąc (ŹRÓBEK 2007).

CZAJA (2001) oprócz opisywanego modelu proponuje również model liniowej regresji ważonej. Model ten uwzględnia stopień podobieństwa poszczególnych nieruchomości. Przyjęte wagi są równe odwrotności wyrażenia zawierającego 1 plus liczbę różniących się atrybutów w nieruchomości wycenianej i porównywanej uwzględnianych w oszacowaniu modelu.

Modele liniowej regresji można próbować stosować również dla zbiorów niezbyt licznych, zawierających kilkanaście transakcji.

W przypadku, gdy zmiany cen w czasie nie są wyraźnie liniowe, wówczas należy zastosować model regresji nieliniowej. Najczęściej wykorzystywane funkcje nieliniowe to: potęgowa, wykładnicza, logarytmiczna oraz wielomian *n*-stopnia. Estymacja parametrów modelu nieliniowego może odbywać się nieliniową metodą najmniejszych kwadratów za pomocą algorytmów np. uogólniony proces Seidela czy Centrum Quadricae (ADAMCZEWSKI 2002).

Właściwy dobór funkcji nie jest prosty. Pomocnym może być analiza sporządzonego wykresu – zależności cen od daty transakcji, a także obliczenie współczynnika determinacji R^2.

W przypadku funkcji nieliniowych aktualizację cen na datę wyceny dokonuje się współczynnikami dla poszczególnych jednostek czasu, w których zawarto transakcję, obliczonymi na podstawie funkcji nieliniowej.

Gdy dynamika cen jest zmienna dla potrzeb aktualizacji cen można stosować zaawansowane metody np. sztuczne sieci neuronowe, które umożliwiają wykrycie złożonych – nieliniowych zależności pomiędzy ceną jednostkową a datą transakcji. Wśród modeli sztucznych sieci neuronowych stosuje się m.in. perceptrony wielowarstwowe czy sieci o radialnych funkcjach bazowych. Omówienie modeli oraz zastosowania sztucznych sieci neuronowych w wycenie nieruchomości zawierają m.in. prace (MCCLUSKEY1996), (WIŚNIEWSKI 1998), (WILKOWSKI, BUDZYŃSKI 2006).

3. Zastosowanie metod aktualizacji cen na przykładzie nieruchomości lokalowych

Przedmiotem badań był lokalny rynek nieruchomości lokalowych obejmujący

dzielnicę miasta Płocka – osiedle Dobrzyńska w okresie 2003-2007. Dane o cenach transakcyjnych nieruchomości i ich charakterystyce uzyskano z pracy (DOMŻALSKA, ŻURAWSKA 2008). Po weryfikacji cen transakcyjnych pod kątem ich przydatności do wyznaczenia trendu czasowego przyjęto do obliczeń zbiór 237 cen jednostkowych. Wartość średnia w tym zbiorze wynosi 1669,98, wartość minimalna 1135,01, wartość maksymalna 3206,09 a odchylenie standardowe 450,32.

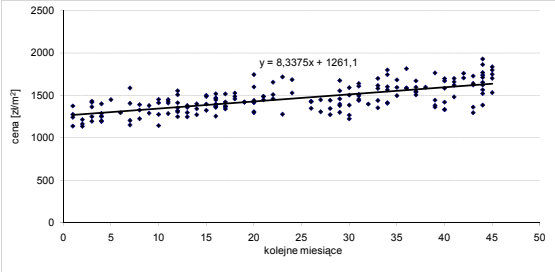
Określono wpływ czasu na poziom cen przy zastosowaniu:

1) metody porównania cen nieruchomości podobnych w parach – określono współczynnik zmiany cen,
2) modeli statystycznych – wyznaczono trendy czasowe typ liniowy i wielomianowy,
3) modelu sztucznych sieci neuronowych – wyznaczono trend czasowy.

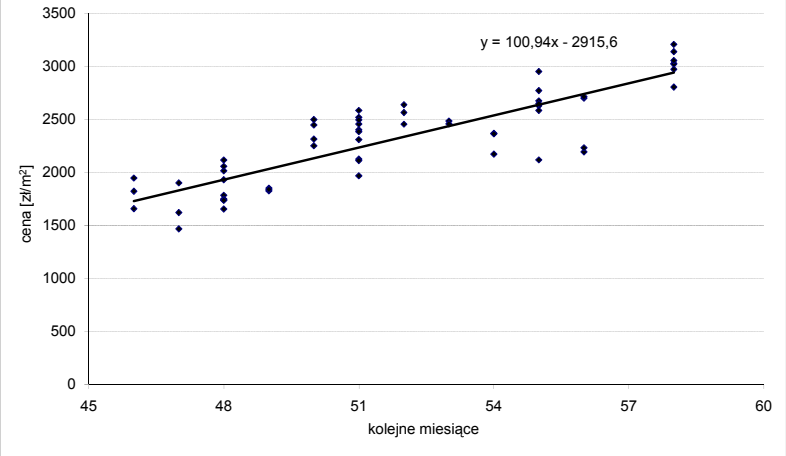
Obliczenia rozpoczęto od zastosowania metody porównania cen nieruchomości w parach. W oparciu o 10 jednostkowych współczynników zmian cen obliczono jako średnią arytmetyczną współczynnik zmian cen dla okresu 2003 -2007. Wyniósł on 0,0206 w skali miesiąca tj. 23,7% w skali roku. Nie jest on adekwatny do realiów rynku ze względu na zmienną dynamikę cen w badanym okresie - patrz rysunek 3. W związku z tym obliczono współczynniki zmiany cen w skali miesiąca oddzielnie dla okresu 01.2003-09.2006 – w oparciu o 10 par oraz 10.2006-10.2007 w oparciu o 4 pary. Wyniosły one odpowiednio: 0,0066 tj. 7,9% w skali roku oraz 0,0540 tj. 64,8 % w skali roku.

Następnie dla okresu - lata 2003-2007 wyznaczono trend czasowy liniowy wyrażony funkcją y = 20,823x + 1030,4. Wartość współczynnika determinacji R^2 wyniosła 0,6073. Trend ten nie opisuje we właściwy sposób zmian cen w czasie, gdyż w badanym okresie można wyróżnić dwa okresy charakteryzujące się zmienną dynamiką wzrostu cen. W pierwszym okresie 01.2003 – 09.2006 (kolejne miesiące oznaczone 1 – 45) można zauważyć powolny wzrost cen, który począwszy od 10.2006 (określono na podstawie rysunku 3), przechodzi we wzrost bardzo szybki. W związku z tym wykonano analizę wzrostu cen dla ww. dwóch okresów, wyznaczając dla nich trend czasowy typu liniowego.

W okresie 01.2003-09.2006 zmienność cen opisuje funkcja y = 8,3375x + 1261,1 a w okresie 10.2006-10.2007 funkcja y = 100,94x - 2915,6. W pierwszym okresie oznacza to wzrost cen w skali roku 7,9% czyli identyczny jak wyznaczony pierwszą metodą, a w okresie drugim 74,4 % wyższy o prawie 10%.Współczynnik R^2 dla trendu liniowego określonego odrębnie dla 2 okresów czasu był bardzo wysoki i wyniósł 0,8843.

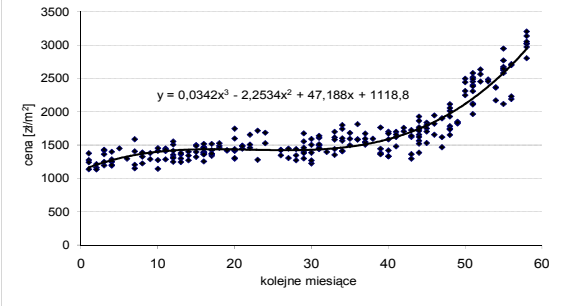


Rys. 1. Ceny nieruchomości lokalowych w okresie 01.2003-09.2006. Linia trendu – typ liniowy. *Źródło*: opracowanie własne.



Rys. 2. Ceny nieruchomości lokalowych w okresie 10.2006-10.2007. Linia trendu – typ liniowy. *Źródło*: opracowanie własne.

Następnie opracowano trend nieliniowy. Zastosowano funkcje: wielomianu, logarytmiczną, potęgową i wykładniczą. Spośród ww. funkcji kierując się najwyższą wartością R^2 wybrano wielomian 3 stopnia o równaniu $y = 0{,}0342x^3 - 2{,}2534x^2 + 47{,}188x + 1118{,}8$ – rysunek 3. Współczynnik R^2 wyniósł 0,8726 – podobnie jak dla ww. trendu liniowego.

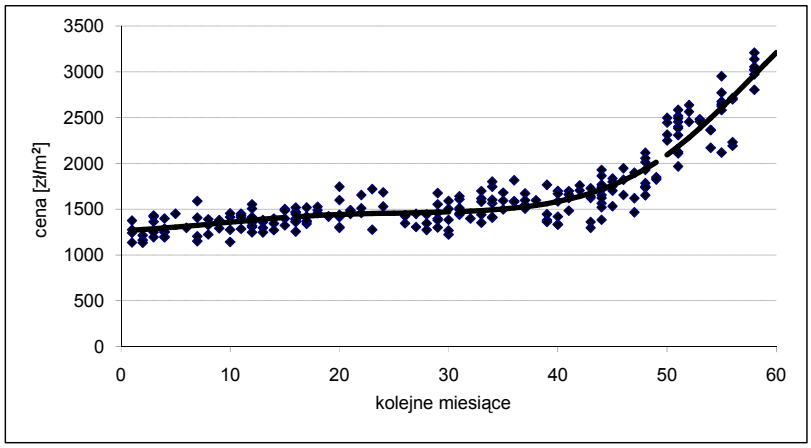


Rys. 3. Ceny nieruchomości lokalowych. Linia trendu – wielomian 3 stopnia.
Źródło: opracowanie własne.

Wykorzystując program komputerowy Statistica po podzieleniu zbioru cen na 2 podzbiory (uczący – 190 przypadków i testowy 47 przypadków), zbudowano modele sztucznych sieci neuronowych typu perceptron wielowarstwowy, poszukując wśród nich modeli najlepiej opisujących zjawisko zmiany cen w czasie – o najmniejszej wartości sumy kwadratów reszt w podzbiorze testowym. Utworzono modele posiadające od 2 do 5 neuronów w warstwie ukrytej. Perceptrony wielowarstwowe były uczone 3 algorytmami uczącymi: najszybszego spadku, gradientów sprzężonych oraz BFGS. Jako funkcję aktywacji warstwy ukrytej i wyjściowej stosowano jedną z następujących funkcji: liniową, logistyczną, tanh i wykładniczą. Jako najlepszą sieć wybrano: perceptron wielowarstwowy o 5 neuronach ukrytych, funkcji aktywacji warstwy ukrytej – tanh, warstwy wyjściowej – wykładniczej, uczony za pomocą algorytmu BFGS (Broyden-Fletcher-Goldfarb-Shanno – wykorzystującego przybliżenie Hessianu za pomocą pierwszych pochodnych).

Z wykorzystaniem ww. perceptronu wielowarstwowego dokonano predykcji zmiennej – ceny transakcyjnej dla poszczególnych miesięcy. Przez uzyskane wartości przeprowadzono linię (rysunek 4), stanowiącą trend czasowy dla którego współczynnik R^2 był bardzo wysoki i wyniósł 0,8796.

Zbudowano również bardziej skomplikowane modele np. sieci o radialnych funkcjach bazowych z 36 neuronami ukrytymi. Te modele jednak zbytnio dopasowywały się do danych nie pokazując ogólnego trendu zmiany cen, co stanowi ich ewidentną wadę.



Rys. 4. Ceny nieruchomości lokalowych. Linia trendu opracowana z wykorzystaniem perceptronu wielowarstwowego. *Źródło*: opracowanie własne.

W celu porównania dopasowania poszczególnych trendów czasowych do cen transakcyjnych zestawiono wartości R^2 w tabeli 1.

Tabela 1

Wartości współczynników R^2 dla poszczególnych typów trendu

Lp.	Typ trendu	R^2
1	Liniowy	0,6073
2	Liniowy – 2 okresy	0,8843
3	Wielomianowy	0,8726
4	Perceptron wielowarstwowy	0,8796

Źródło: opracowanie własne.

Na podstawie obserwacji wykresów pokazujących zmienność cen w czasie oraz obliczonych wartości współczynnika R^2 dla poszczególnych trendów czasowych można stwierdzić, iż zastosowanie trzech rodzajów trendów: liniowego – obliczonego odrębnie dla 2 okresów, wielomianowego oraz trendu opracowanego z wykorzystaniem perceptronu wielowarstwowego, daje w tym przypadku bardzo zbliżone wyniki. Oddają one bardzo dobrze rzeczywistość rynku nieruchomości w zakresie zmian cen w przeciwieństwie do trendu liniowego obliczonego dla całego badanego okresu dla którego wartość współczynnika była wyraźnie niższa.

Podstawą zatem jest znajomość dynamiki zmian cen do której można następnie dobrać odpowiednią funkcję począwszy od najprostszej – liniowej – dzieląc okres analizy na krótsze przedziały czasu, do najbardziej złożonych wykorzystujących

sieci neuronowe.

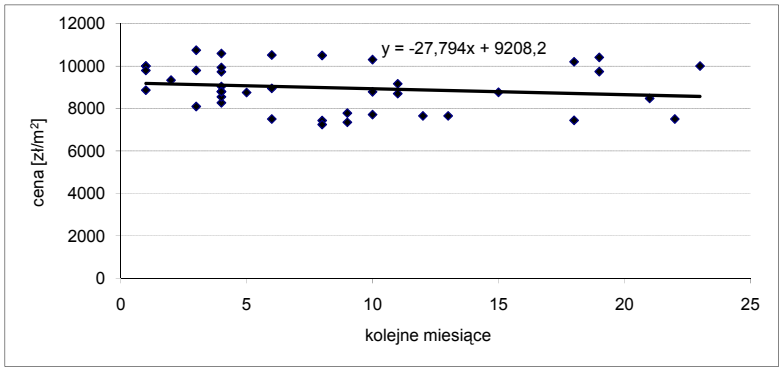
Jak widać na powyższym przykładzie nie zawsze funkcja bardziej skomplikowana opisuje lepiej zjawiska występujące na rynku nieruchomości.

Przydatna przy określaniu zmian cen może być też metoda porównania cen nieruchomości w parach. Należy wcześniej jednak upewnić się korzystając m.in. z dostępnych analiz rynku, jaki charakter mają zmiany cen.

Kolejnym przedmiotem badań był lokalny rynek nieruchomości lokalowych obejmujący dzielnicę Miasta Stołecznego Warszawy – Ochotę w okresie 07.2007-06.2009. Dane o cenach transakcyjnych nieruchomości i ich charakterystyce uzyskano z portalu www.oferty.net. Po weryfikacji cen transakcyjnych pod kątem ich przydatności do wyznaczenia trendu czasowego przyjęto do obliczeń zbiór 40 cen jednostkowych. Wartość średnia w tym zbiorze wynosi 8969,15, wartość minimalna 7241,00 wartość maksymalna 10746,00 a odchylenie standardowe 1084,08.

Określono wpływ czasu na poziom cen przy użyciu metod zastosowanych dla dzielnicy miasta Płocka. Na początku posłużono się metodą porównywania cen nieruchomości w parach. Uzyskane współczynniki jednostkowe miały zbliżone wartości zarówno dodatnie jak i ujemne, co nie pozwoliło przesądzić ostatecznie o spadku czy wzroście cen. Można co najwyżej przyjąć, że na rynku nieruchomości występowała w badanym okresie stabilizacja cen.

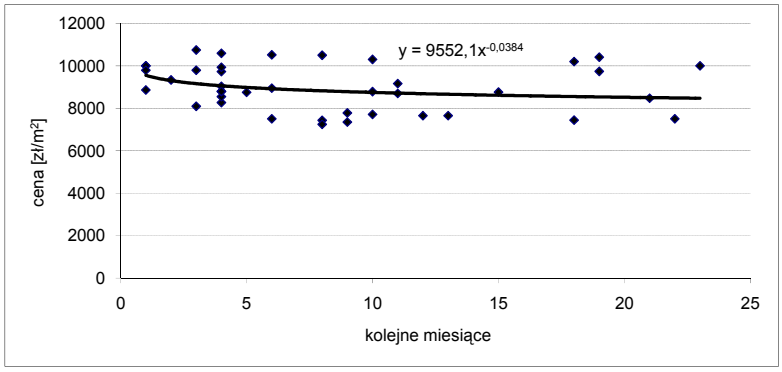
Następnie wyznaczono trend czasowy liniowy opisany funkcją y = -27,794x + 9208,2. Wskazuje on na niewielki – 3,6% spadek cen w skali roku. Obliczona wartość współczynnika R^2 dla tego trendu wyniosła zaledwie 0,0269.



Rys. 5. Ceny nieruchomości lokalowych. Trend czasowy – typ liniowy. *Źródło*: opracowanie własne.

Dla badanego okresu czasu wyznaczano również trendy wielomianowe, potęgowy, logarytmiczny i wykładniczy. Najwyższy współczynnik R^2 = 0,0782 uzyskał oraz zapewniał najlepszą zgodność z obserwacjami rynku trend typu

potęgowego y = 9552,1x$^{-0,0384}$. Zgodnie z nim w badanym okresie czasu ceny malały, przy czym największa dynamika spadku była w początkowym okresie.



Rys. 6. Ceny nieruchomości lokalowych. Trend czasowy – typ potęgowy. *Źródło*: opracowanie własne.

Wykorzystując program komputerowy Statistica po podzieleniu zbioru cen na 2 podzbiory (uczący – 33 przypadki i testowy – 7 przypadków) zbudowano modele sztucznych sieci neuronowych typu perceptron wielowarstwowy o tej samej architekturze i pozostałych parametrach jak dla dzielnicy miasta Płocka. Jako najlepszą sieć – o najmniejszej wartości sumy kwadratów reszt w zbiorze testowym, wybrano: perceptron wielowarstwowy o 4 neuronach ukrytych, funkcji aktywacji warstwy ukrytej – liniowej, warstwy wyjściowej – wykładniczej, uczony za pomocą algorytmu BFGS.

Z wykorzystaniem ww. perceptronu wielowarstwowego dokonano predykcji zmiennej – ceny transakcyjnej dla poszczególnych miesięcy. Przez uzyskane wartości przeprowadzono linię, stanowiącą trend czasowy. Zbliżony jest on do trendu czasowego typu potęgowego. Przedstawiono go na rysunku 7 na tle cen transakcyjnych. Wartość współczynnika R^2 wyniosła dla niego 0,1143.

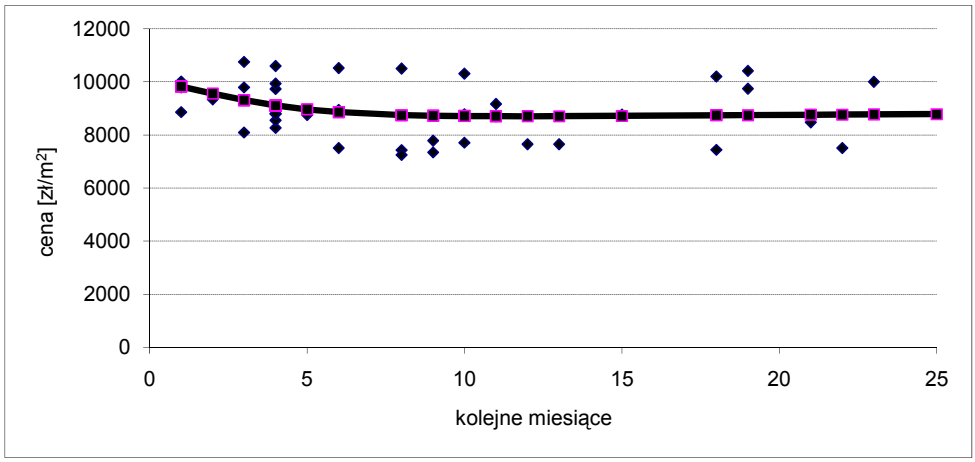
W celu porównania dopasowania poszczególnych trendów czasowych do cen transakcyjnych zestawiono wartości R^2 w tabeli 2.

Tabela 2

Wartości współczynników R^2 dla poszczególnych typów trendu.

Lp.	Typ trendu	R^2
1	Liniowy	0,0269
2	Potęgowy	0,0782
3	Perceptron wielowarstwowy	0,1143

Źródło: opracowanie własne.



Rys. 7. Ceny nieruchomości lokalowych. Linia trendu opracowana z wykorzystaniem perceptronu wielowarstwowego. *Źródło*: opracowanie własne.

Wszystkie trendy czasowe pokazują niewielki spadek cen w badanym okresie. Najlepiej dopasowana jest linia trendu opracowana z wykorzystaniem perceptronu wielowarstwowego – rysunek 7 dla której współczynnik R^2 był najwyższy. Jednocześnie należy zauważyć, iż obliczone wartości współczynników determinacji były bardzo niskie. Wynika to z faktu stosunkowo dużego zróżnicowania cen w zbiorze w porównaniu do zróżnicowania spowodowanego upływem czasu. Istotny wpływ na poziom cen na tym lokalnym rynku nieruchomości w badanym okresie miały jedynie cechy rynkowe nieruchomości a nie upływ czasu – gdyż ceny nieruchomości podobnych sprzedanych w różnych okresach czasu były do siebie zbliżone.

4. Podsumowanie

Aktualizacja cen jest procesem dosyć złożonym obarczonym niepewnością co do prawidłowości uzyskanych wyników. Kluczową kwestią przed rozpoczęciem procesu obliczeniowego jest określenie prawdopodobnej dynamiki zmian cen w czasie. Wiedza w tym zakresie jest pomocna przy wyborze odpowiedniej metody aktualizacji cen. Dotyczy to w szczególności przypadku niezbyt licznej bazy cen transakcyjnych, których zróżnicowanie jest stosunkowo duże. Taka sytuacja występowała w przypadku dzielnicy Ochota. Zastosowanie nawet bardzo zaawansowanych metod jakimi są sztuczne sieci neuronowe nie wyjaśniło w dużym stopniu zmienności cen w bazie. Trend czasowy opracowany z wykorzystaniem perceptronu wielowarstwowego był co prawda najlepiej dopasowany do zbioru cen spośród obliczonych trendów, ale i tak współczynnik R^2 wynosił zaledwie 0,1143. Wynikało to z faktu, że istotny wpływ na poziom cen na tym lokalnym rynku nieruchomości w badanym okresie miały jedynie cechy rynkowe nieruchomości – a nie upływ czasu, gdyż ceny nieruchomości podobnych

sprzedanych w różnych okresach czasu były do siebie zbliżone. W przypadku posiadania licznego zbioru cen, odpowiednie zastosowanie opisywanych metod – nawet tych najprostszych, pozwala na uzyskanie zadawalającego wyniku, co widać na przykładzie dzielnicy miasta Płocka Wartość współczynnika R^2 dla trendu liniowego określonego odrębnie dla 2 okresów czasu, trendu wielomianowego oraz trendu opracowanego z wykorzystaniem perceptronu wielowarstwowego były bardzo zbliżone i wynosiły odpowiednio 0,8843, 0,8726 oraz 0,8796.

5. Literatura

ADAMCZEWSKI Z. 2002. *Nieliniowe i nieklasyczne algorytmy w geodezji*. Oficyna Wydawnicza PW, Warszawa

CZAJA J. 2001. *Metody szacowania wartości rynkowej i katastralnej nieruchomości.* KOMP-SYSTEM. Kraków

DOMŻALSKA D. ŻURAWSKA I. 2008. *Analiza rynku nieruchomości lokalowych na potrzeby wyceny.* Praca dyplomowa magisterska Wydział Geodezji i Kartografii 2008.

MCCLUSKEY W.J. 1996. *Zastosowanie sztucznej inteligencji w wycenie masowej dla potrzeb taksacji nieruchomości.* Wycena nr 5(28), Olsztyn.

PRYSTUPA M. 2001. *Wycena nieruchomości przy zastosowaniu podejścia porównawczego.* PFSRzM. Warszawa.

WILKOWSKI W., BUDZYŃSKI T. 2006. *Application of Artificial Neural Networks for Real Estate Valuation*. Shaping the Change XXIII FIG Congress Munich, Germany.

WIŚNIEWSKI R. 1998. *Zastosowanie sztucznych sieci neuronowych do wyceny masowej*. Wycena 1(36), Olsztyn

ŹRÓBEK S. 2007. *Metodyka określania wartości rynkowej nieruchomości*. Educaterra.

METHODOLOGY OF THE UPDATE OF PRICES ON THE EXAMPLE OF HOUSE REAL ESTATES

Tomasz Budzyński
Department of Cadastre and Land Management
Warsaw University of Technology
e-mail: t.budzynski@gik.pw.edu.pl

Key words: update of prices, regression models, artificial neural networks

Abstract

In the process of valuation a valuer is obliged to update of real estate prices for the evaluation date. A correct update is particularly important in case of noticeable indeed changes of the price level on the real estate market in the examined period of time.

The paper presents methods of determing the influence of the date of transaction on real estate prices and methods of their update. Valuers compare prices of identical or very similar real estates sold in different period of times, calculating the rate of the changes in prices.

In case of numerous sets of prices valuers can apply statistical methods among others, determing the line of the trend about linear or non – linear form. It is possible to apply also others advanced metods e.g. artificial neural networks for the purpose of the update of prices for the evaluation date.

The paper contains examples of applying mentioned above methods for the purposes of update of prices. Two local house real estate markets were objects of research: the district of the city Płock and the district of Warsaw. Results obtained using mentioned above methods were discussed and compared as well.

PORÓWNANIE METOD PROSTYCH ORAZ METODY REGRESJI HEDONICZNEJ DO KONSTRUOWANIA INDEKSÓW CEN MIESZKAŃ

Radosław Trojanek
Katedra Inwestycji i Nieruchomości
Uniwersytet Ekonomiczny w Poznaniu
e-mail: r.trojanek@ue.poznan.pl

Słowa kluczowe: indeksy cen nieruchomości mieszkaniowych, rynek mieszkaniowy

Streszczenie

W opracowaniu podjęto próbę oceny średniej oraz średniej ważonej jako metod konstruowania indeksów cen nieruchomości mieszkaniowych. Problematyka ta została omówiona na przykładzie wtórnego rynku mieszkaniowego w Poznaniu w latach 2008-2009.

1. Istota indeksów cen nieruchomości mieszkaniowych

Idealny indeks cen mierzy generalną zmianę poziomu ceny grupy towarów w danym okresie. Określenie „generalną" odnosi się do typowej zmiany ceny w wybranej grupie towarów, przy założeniu, że taka typowa zmiana istnieje (CHAU i in. 2005). Konstrukcja indeksu cen nieruchomości mieszkaniowych stwarza problemy na poziomie już samej koncepcji jak również w momencie praktycznego zastosowania. Niemożliwym wydaje się być zastosowanie metod szeroko stosowanych do obliczania indeksów cen innych dóbr czy też usług, głównie z trzech przyczyn (WOOD 2005). Po pierwsze, nieruchomości mieszkaniowe są heterogeniczne – nie istnieją dwa identyczne domy, mieszkania, zawsze występuje różnica choćby w jednym elemencie, np. lokalizacji. Ponadto, uchwycenie różnic w cechach fizycznych jak i jakościowych nieruchomości mieszkaniowych w danej grupie nieruchomości czy też w danym okresie sprawia wiele problemów, głównie ze względu na jakość baz danych. Należy podkreślić ponadto, że cechy ilościowe i jakościowe mogą się zmieniać w czasie, poprzez modernizację starszych obiektów czy też wzrost liczby sprzedanych obiektów wybudowanych w tradycyjnej technologii, zlokalizowanych na osiedlach strzeżonych. Po drugie, obserwacja zmiany cen danej nieruchomości w czasie jest niezwykle utrudniona i ujawnia się w momencie dokonania sprzedaży. Należy podkreślić, że od momentu zawarcia transakcji do chwili, kiedy informacja ta jest dostępna i może zostać wykorzystana upływa nawet kilka miesięcy. Po trzecie, transakcje na rynku nieruchomości, w relacji do innych, dóbr zawierane są stosunkowo rzadko.

Prawidłowe wskaźniki, czyli takie, które będą pokazywały zmiany czystej ceny

i nie będą obciążone błędami wynikającymi ze zmian chociażby cech jakościowych, wymagają takiego systemu kryteriów, który pozwoli uwzględnić zróżnicowanie w grupie obiektów. Innymi słowy, pierwotne dane muszą być odpowiednio przygotowane (CASE, WACHTER 2005). Pomijając samą jakość danej próby, wybór metody do oceny jednorodności w tej próbie, jest głównym elementem wpływającym na ocenę przydatności określonego systemu mierników.

Zważywszy na fakt, że rynek nieruchomości jest rynkiem lokalnym, wszelkie indeksy cen nieruchomości odnoszą się do danego obszaru geograficznego. W większości przypadków zasięg geograficzny indeksów nie przekracza granic miasta czy metropolii. Budowa indeksów dla większych obszarów wymaga dysponowania bogatymi bazami danych, dlatego takie indeksy najczęściej są liczone głównie w krajach o rozwiniętych rynkach nieruchomości. Należy podkreślić, że indeksy skonstruowane dla danego typu nieruchomości, np. dla domów jednorodzinnych, opisują zmiany cen tylko w obrębie tej grupy nieruchomości, a nie np. dla mieszkań zlokalizowanych w budynkach wielorodzinnych.

2. Metody proste oraz regresji hedonicznej jako metody konstruowania indeksów cen nieruchomości mieszkaniowych

Metody konstruowania wskaźników cen[3] nieruchomości można podzielić, biorąc za kryterium możliwość uwzględnienia zmian cech jakościowych i ilościowych nieruchomości, na dwie grupy: metody proste (takie, które tych zmian nie uwzględniają) oraz metody złożone (takie, które te zmiany uwzględniają). Do metod prostych zalicza się metody oparte na średniej oraz medianie. W grupie metod złożonych wyróżnia się: metody regresji hedonicznej, powtórnej sprzedaży, średniej ważonej oraz hybrydowe. Przedmiotem szerszych rozważań w niniejszym opracowaniu będą metoda proste oraz regresji hedonicznej.

2.1. Metoda średniej i mediany

Najprostszą metodą określenia głównej tendencji kształtowania się cen nieruchomości mieszkaniowych jest wyznaczenie średniej lub mediany ceny w danym okresie. Zważywszy na fakt, że ceny nieruchomości mieszkaniowych przeważnie charakteryzuje dodatnia asymetria (spowodowane głównie jest to przez heterogeniczność nieruchomości) prosta średnia używana jest stosunkowo rzadko (MARK, GOLDBERG 1984). Średnia arytmetyczna jest stosunkiem wartości globalnej badanej cechy do liczebności zbiorowości. Średnią arytmetyczna w szeregu szczegółowym wyznaczyć można na podstawie wzoru (1):

[3] Omówienie poszczególnych metod można znaleźć np. w: R. Trojanek, *Wahania cen na rynku mieszkaniowym*, Wydawnictwo Akademii Ekonomicznej w Poznaniu, Poznań 2008r.

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}, \quad (1)$$

gdzie:
x_i – wartość i-tej obserwacji,
n – liczba obserwacji.

Mediana jest to wartość cechy dzieląca uporządkowaną zbiorowość statystyczną na dwie jednakowe pod względem liczebności części w ten sposób, że połowa jednostek zbiorowości ma wartości mniejsze lub równe medianie, a połowa jednostek zbiorowości przyjmuje wartości większe lub równe wartości mediany. W szeregu szczegółowym medianę wyznaczyć można z poniższego wzoru (2):

$$Me = \frac{x_{n+1}}{2}, gdy\ n\ jest\ nieparzyste\ \text{lub}\ \frac{1}{2}\left(x_{n/2} + x_{n/2+1}\right), gdy\ n\ jest\ parzyste, \quad (2)$$

gdzie:
x_i – wartość i-tej obserwacji,
n – liczba obserwacji.

W celu skonstruowania indeksu cen nieruchomości opartego na średniej, należy najpierw wyznaczyć te wartości dla każdego z okresów, a następnie z uzyskanych rezultatów zbudować szereg czasowy.

W odniesieniu do metody średniej, największą jej wadą jest nieuwzględnianie zmian jakościowych i ilościowych nieruchomości mieszkaniowych w czasie (ENGLUND I IN. 1999). Odnosi się to zarówno dla jednego okresu jak również dla kilku, powodując, że zmiana w strukturze nieruchomości może obciążyć wskaźniki oparte na średniej.

Zważywszy na powyższe fakty, indeks cen skonstruowany z wykorzystaniem średniej, może dostarczyć wiarygodnych informacji o czystej zmianie ceny nieruchomości wyłącznie, gdy spełnione zostaną następujące warunki:

- istnieje mała zmiana w strukturze analizowanych nieruchomości (np. wszystkie mieszkania zlokalizowane są w budynkach wykonanych w technologii wielkopłytowej)
- zmiany jakościowe analizowanych nieruchomości są ograniczone (np. wszystkie mieszkania posiadają podobny standard wykończenia).

2.2. Metoda regresji hedonicznej

Idea modelu hedonicznego sprowadza się do założenia, że cena nieruchomości może zostać odpowiednio oszacowana na podstawie cech jakościowych i ilościowych danej nieruchomości. Regresja hedoniczna to sposób określania wpływu poszczególnych cech nieruchomości na jej wartość. Zamiast dostarczać jedynie zwykłego podsumowania dotyczącego tempa wzrostu cen lub samych cen, metoda hedoniczna pozwala uzyskać matematycznie poprawną postać funkcji

ceny nieruchomości. Regresja hedoniczna jest przykładem metody parametrycznej. Modele parametryczne są takimi modelami gdzie postać modelu jest przyjęta a priori. Termin „parametryczny" oznacza, że liczba i natura parametrów jest ustalona z góry (LIM, PAVLOU 2007).

Indeks cen nieruchomości w oparciu o metodę hedoniczną może zostać zbudowany na dwa sposoby(BOURASSA, HOESLI 2006). W pierwszym podejściu budowane są modele regresji dla każdego okresu i następnie na podstawie oszacowania parametrów modelu konstruowany jest indeks. Drugi sposób polega na zbudowaniu jednego równania regresji, zawierającego zmienną binarną czasu. Zasadnicza różnica między tymi dwoma podejściami polega na tym, że w pierwszym przypadku zarówno średnia jak i odchylenie standardowe składnika losowego różni się w badanych okresach, natomiast w drugim są one stałe.

Poniżej scharakteryzowana została metoda, która jest opisem standardowej regresji hedonicznej, opisanej szczegółowo przez Fleming'a i Nellis'a (FLEMING, NELLIS 1994). W danym momencie poszczególne nieruchomości są różnie wyceniane ze względu na ich cechy jakościowe (np. typ zabudowy, lokalizacja) i ilościowe (liczba pokoi, liczba łazienek, wiek budynku). Cena każdej nieruchomości może zostać przedstawiona jako funkcja jej atrybutów mierzalnych X_i oraz niemierzalnych, które są specyficzne dla każdej nieruchomości, ale dla których dane nie są dostępne, e_i. Zależność ta może zostać wyrażona za pomocą równania (3):

$$\ln(P_i) = b_0 + b_1 X_{1,i} + b_2 X_{2,i} + b_3 X_{3,i} + \ldots + b_j X_{j,i} + e_i , \qquad (3)$$

gdzie b_1, b_2 ,...b_j są współczynnikami regresji odpowiadającymi zmiennym jakościowym i ilościowym, X_j.

Ograniczenia, wypływające z danych i metodologii, powodują, że jakościowe cechy nieruchomości muszą być reprezentowane przez zmienną zero-jedynkową (przyjmując wartość 0, jeśli dana nieruchomość nie posiada danej cechy i 1 jeśli ją posiada). Wskaźniki b_1 ,b_2 ,...b_j powiązane z kolejnymi zmiennymi objaśniającymi X_j są szacowane za pomocą klasycznej metody najmniejszych kwadratów (klasyczna mnk). Rolą współczynników jest wskazanie względnej ważności zmiennych w wyjaśnianiu różnic (rozbieżności) w cenach nieruchomości w danym okresie.

Następnym krokiem w analizie regresji jest standaryzacja. Uzyskuje się to poprzez zastosowanie systemu wag odpowiadających atrybutom w wybranym okresie (zazwyczaj wybiera się okres początkowy). Obliczony numer wskaźnika reprezentuje średni ruch cen dla nieruchomości posiadających te same atrybuty jak nieruchomości w okresie początkowym. Cena, ustalona za pomocą metody ceny skorygowanej, jest wyrażona przy zastosowaniu wag W_j, stałych w czasie równaniem (5):

$$\ln(P_i) = b_{0,i} + W_1 b_{1,i} + W_2 b_{2,i} + W_3 b_{3,i} + \ldots + W_j b_{j,i} + e_i , \qquad (4)$$

Wagi te są proporcjami liczby nieruchomości z daną cechą w okresie początkowym. Ostatni etap sprowadza się do:

- obliczenia wag, W_j: proporcja zmiennych jakościowych i średnich ilościowych reprezentowanych w wybranym początkowym okresie,
- użycia klasycznej mnk, dla oszacowania wskaźników regresji $b_1,b_2,...b_j$ dla j zmiennych objaśniających, zarówno dla okresu początkowego jak i dla okresów następnych,
- obliczenia początkowego wskaźnika Laspeyres'a ważonego okresem obliczeniowym dla bieżącego okresu (I_t) jako:

$$I_t = \frac{anti \ln \sum b_{jt} W_j}{anti \ln \sum b_{jt=0} W_j} x100; \quad (5)$$

gdzie:
Wj – wagi wyznaczone dla każdej zmiennej objaśniającej dla okresu początkowego, np. dla zmiennej lokalizacja będzie to udział sprzedanych mieszkań w danej dzielnicy w ogóle zawartych transakcji
b_{jt} - współczynniki regresji odpowiadające zmiennym jakościowym i ilościowym.
Suma dotyczy wszystkich zmiennych w każdej funkcji regresji.

Wykorzystanie każdej z wyżej wymienionych metod pozyskiwania informacji o cenach/wartościach nieruchomości mieszkaniowych niesie za sobą pewne korzyści jak i obciążenia. W tabeli 1 przedstawiono główne zalety jak i ograniczenia poszczególnych metod.

Tabela 1

Korzyści i obciążenia metod konstrukcji indeksów cen nieruchomości

Metoda	Korzyści	Obciążenia
Prosta średnia lub mediana	-łatwa do liczenia	-nie uwzględnia zmian cech jakościowych jak i strukturalnych nieruchomości
Metody hedoniczne	-kontrola nad zmianami cech jakościowych -nie odnosi się wyłącznie do wybranej grupy nieruchomości	-wymagania co do danych -potencjalne obciążenie wynikające ze specyfikacji modelu

Źródło: Opracowanie własne.

Nie ma idealnej metody konstrukcji indeksów cen nieruchomości. Zastosowanie metod złożonych często jest niemożliwe ze względu na brak odpowiednich baz danych. W sytuacji występowania znikomej liczby cech opisujących nieruchomości różnice wynikające z wykorzystania metod złożonych i prostych są nieduże. Potwierdzają fakt ten liczne badania przeprowadzone w krajach, w których problematyka konstrukcji indeksów cen na rynku

mieszkaniowym jest bardzo rozwinięta[4].

3. Metodyka badania oraz źródła danych

W celu określenia znaczenia wyboru metody konstruowania indeksów cen nieruchomości mieszkaniowych zebrano informacje o cenach ofertowych dla miasta Poznania w okresie I kw. 2008- IV kw. 2009 r. Pierwotnie dane obejmowały ponad 50000 ofert sprzedaży mieszkań w latach 2008-2009. Usunięto puste rekordy, rekordy powtarzające się, czy też takie, w których określenie ceny ofertowej 1 m² było niemożliwe. Powtórzenia danych było spowodowane ogłaszaniem jednej oferty przez kilka biur pośrednictwa nieruchomości, a więc wielokrotnym umieszczaniem w bazie danych tej samej oferty.

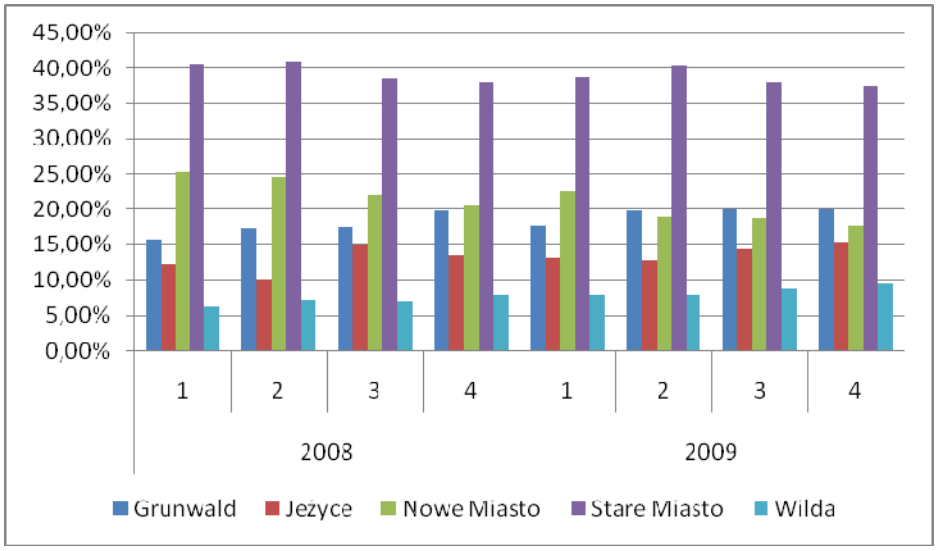
Kolejny etap analizy dotyczył otrzymanych danych pod kątem ich wiarygodności. Etap ten miał na celu wyeliminowanie tych ofert, które, bez jasno określonej przyczyny, znacznie odbiegały od średniej. Do klasyfikacji danych wykorzystano informacje o średnich cenach lokali mieszkalnych w poszczególnych dzielnicach w danym miesiącu. Ponadto przyjęto, że analizie poddane zostaną mieszkania o powierzchni do 120 m² oraz o liczbie pokoi nie większej niż cztery. Przedmiotem zainteresowania były prawo własności jak i spółdzielcze własnościowe prawo do lokalu. W wyniku powyższych zabiegów liczebność bazy danych zmniejszyła się do ponad 30 tysięcy informacji o ofertach sprzedaży mieszkań. Liczba zebranych ofert spełnia warunek reprezentatywności próby.

Kolejnym krokiem była analiza struktury oferowanych mieszkań w Poznaniu w latach 2008-2009. Na wykresach 1, 2, 3 i 4 przedstawiono strukturę oferowanych na sprzedaż mieszkań ze względu na położenie (dzielnica), liczbę pokoi, okres budowy oraz materiał z którego wykonany był budynek.

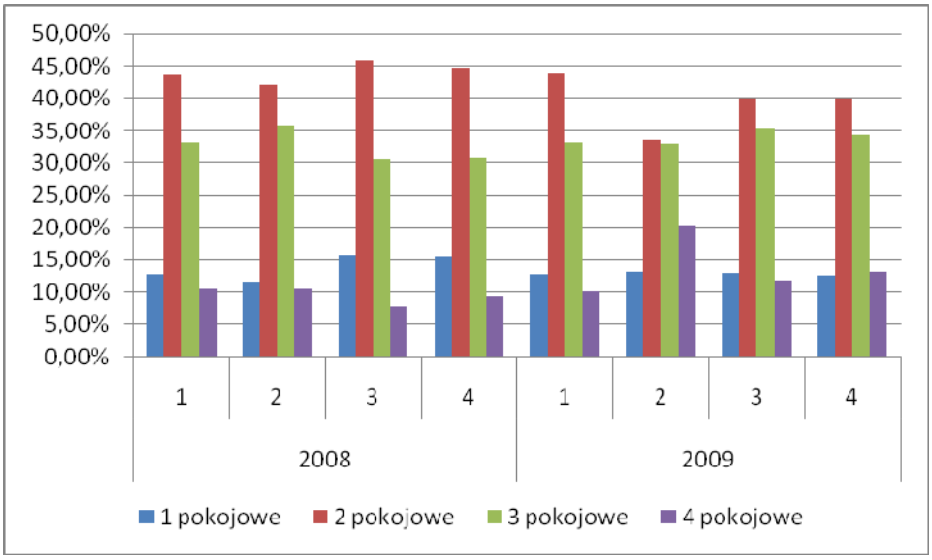
Struktura oferowanych do sprzedaży mieszkań ze względu na przyjęte kryterium w analizowanym okresie podlegała zmianom. Należy jednak zaznaczyć, że zmiany te nie były gwałtowne – w porównaniu między kwartałami zazwyczaj sięgały kilku punktów procentowych. Fakt ten jednak powoduje, że indeksy cen mieszkań zbudowane na metodach prostych będą obciążone - nie będą odwzorowywać w sposób prawidłowy zachodzących zmian na rynku mieszkaniowym, co może prowadzić do błędnych wniosków.

W celu określenia znaczenia wyboru metody konstruowania indeksu na osiągnięte wyniki wyznaczono średnią i medianę oraz cenę heodniczną mieszkań w poszczególnych kwartałach a następnie na ich podstawie określono indeksy cen oraz procentowe zmian rok do roku. W badaniu wykorzystano metodę hedoniczną zaproponowaną przez FLEMING'A I NELLIS'A. W danym momencie poszczególne nieruchomości są różnie wyceniane ze względu na ich cechy jakościowe (np. typ zabudowy, lokalizacja) i ilościowe (liczba pokoi, liczba łazienek, wiek budynku).

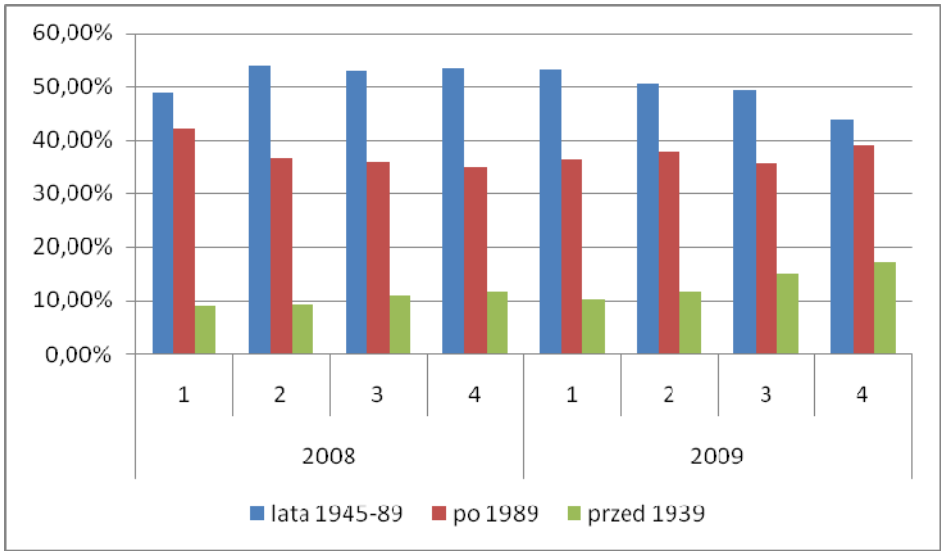
[4] Por. J. Hansen, *Australian House Prices: A Comparison of Hedonic and Repeat-sales Measures*, Reserve Bank of Australia 2006, s.10.; R. Meese, N. Wallace, *The Construction of Residential Housing Price Indices: a Comparison of Repeat Sales, Hedonic Regression, and Hybrid Approaches*, Journal of Real Estate Finance and Economics 1997, vol. 14, nr 1/2, ss. 51-74.



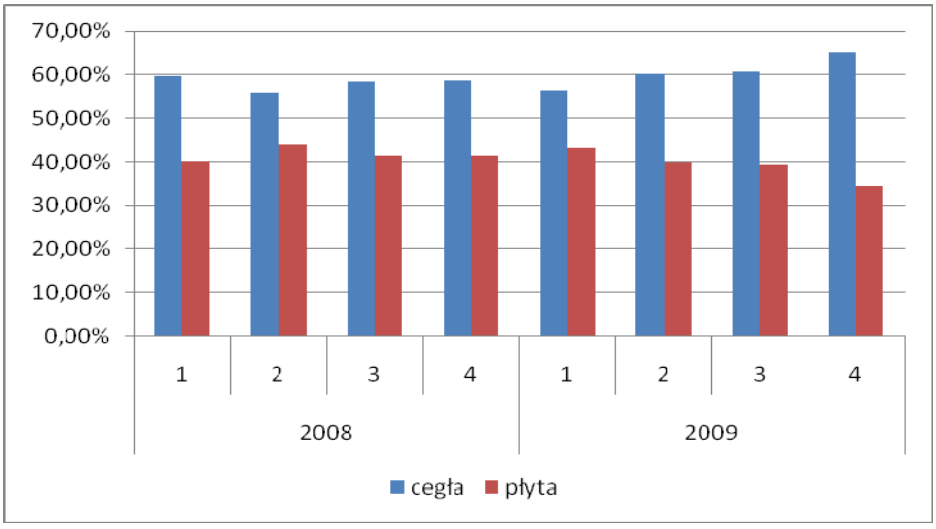
Wykres 1. Struktura mieszkań oferowanych na sprzedaż mieszkań ze względu na położenie (dzielnica) w Poznaniu w latach 2008-2009 (w %). *Źródło*: Opracowanie na podstawie danych CARN.



Wykres 2. Struktura mieszkań oferowanych na sprzedaż mieszkań ze względu na liczbę pokoi w Poznaniu w latach 2008-2009 (w %). *Źródło*: Opracowanie na podstawie danych CARN.



Wykres 3. Struktura mieszkań oferowanych na sprzedaż mieszkań ze względu na okres budowy mieszkania w latach 2008-2009 (w %). *Źródło*: Opracowanie na podstawie danych CARN.



Wykres 4. Struktura mieszkań oferowanych na sprzedaż mieszkań ze względu na materiał, z którego wykonany był budynek w latach 2008-2009 (w %). *Źródło*: Opracowanie na podstawie danych CARN.

Cena każdej nieruchomości może zostać przedstawiona jako funkcja jej atrybutów mierzalnych X_i oraz niemierzalnych, które są specyficzne dla każdej nieruchomości, ale dla których dane nie są dostępne, e_i. Zależność ta może zostać wyrażona za pomocą równania (6):

$$\ln(P_i) = b_0 + b_1 X_{1,i} + b_2 X_{2,i} + b_3 X_{3,i} + ... + b_j X_{j,i} + e_i , \qquad (6)$$

gdzie b_1, b_2 ,...b_j są współczynnikami regresji odpowiadającymi zmiennym jakościowym i ilościowym, X_j.

Wybór zmiennych jakościowych i ilościowych ograniczony był przez informacje dostępne w bazie danych. W tabeli 2 przedstawiono wykorzystane w badaniu zmienne.

Tabela 2

Zmienne jakościowe i ilościowe wykorzystane w modelu

Zmienna	Symbol	Opis
Lokalizacja	*L1* – Grunwald *L2*- Jeżyce *L3*- Nowe Miasto *L4*- Stare Miasto *L5*- Wilda	5 zmiennych binarnych. W przypadku, gdy mieszkanie znajduje się w danej dzielnicy wówczas przyjmuje się 1, w innym przypadku 0.
Materiał	*M1*-cegła *M2*-płyta	2 zmienne binarne. W przypadku, gdy mieszkanie znajduje się w budynku wykonanym z danego materiału wówczas 1, w innym przypadku 0.
Okres budowy	*R1* – 1945-1989 *R2* – po 1989 *R3* – przed 1939	3 zmienne binarne. W przypadku, gdy mieszkanie znajduje się w budynku wykonanym w danym okresie wówczas 1, w innym przypadku 0.
Powierzchnia	*pow.*	Powierzchnia danego mieszkania wyrażona w metrach kwadratowych.
Liczba pokoi	*l_pok.*	Liczba pokoi

Źródło: Opracowanie własne.

Następnie, przy wykorzystaniu programu GRETL, oszacowano równania ekonometryczne o postaci równania (3), dla każdego kwartału w latach 2008-2009, w których zmienną objaśnianą była cena mieszkania natomiast zmiennym objaśniającymi były lokalizacja, materiał z którego wykonany był dany budynek, okres budowy, powierzchnia mieszkania oraz liczba pokoi. W tabeli 3 przedstawiono wyniki funkcji regresji dla równania w I kw. 2008r.

Na podstawie otrzymanych rezultatów można stwierdzić, że użyte w równaniu zmienne objaśniające w 78% wyjaśniają kształtowanie się cen mieszkań

w Poznaniu w I kw. 2008 r. Ponadto wszystkie zmienne użyte w modelu okazały się statystycznie istotne.

Tabela 3

Wyniki funkcji regresji cen mieszkań w Poznaniu w I kw. 2008r.

	Współczynnik	*Błąd stand.*	*t-Student*	*wartość p*	
Const	11,6806	0,0303665	384,6553	<0,00001	***
L1	0,0581721	0,0194086	2,9972	0,00277	***
L2	0,0783065	0,0201582	3,8846	0,00011	***
L3	0,128985	0,0195016	6,6141	<0,00001	***
L4	0,125597	0,0185133	6,7841	<0,00001	***
M1	0,0696551	0,0169289	4,1146	0,00004	***
R1	-0,0614445	0,0201142	-3,0548	0,00229	***
R2	0,21414	0,0161616	13,2499	<0,00001	***
pow.	0,01238	0,000436166	28,3837	<0,00001	***
l_pok	0,0551372	0,00926963	5,9482	<0,00001	***

Źródło: Opracowanie własne.

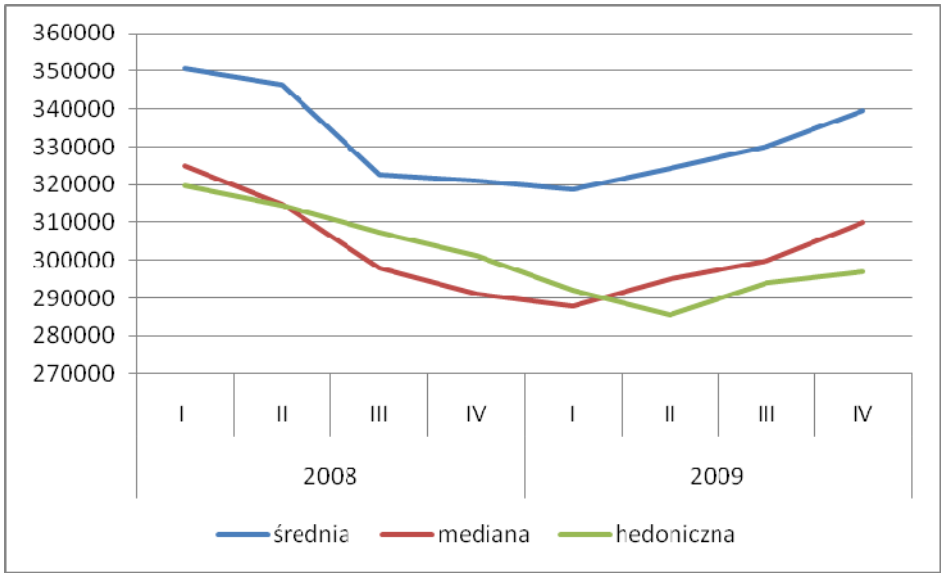
Średnia arytmetyczna zmiennej zależnej = 12,705
Odchylenie standardowe zmiennej zależnej = 0,344745
Suma kwadratów reszt = 37,7477
Błąd standardowy reszt = 0,15853
Wsp. determinacji R^2 = 0,78980
Skorygowany R^2 = 0,78854

Następnie obliczono wag, W_j: proporcja zmiennych jakościowych i średnich ilościowych reprezentowanych w I kw. 2008 r. Kolejny etap sprowadzał się do obliczenia początkowego wskaźnika Laspeyres'a ważonego okresem obliczeniowym dla bieżącego okresu (I_t) jako (7):

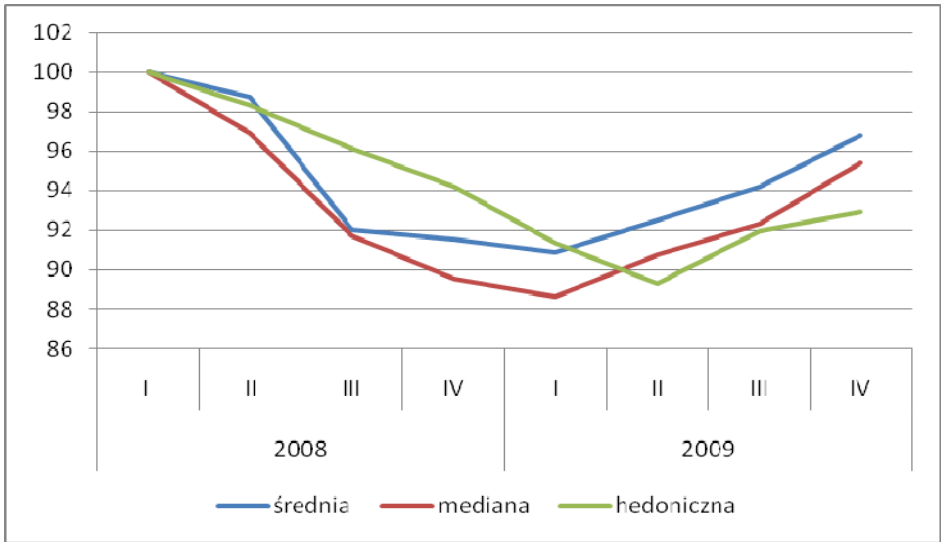
$$I_t = \frac{anti \ln \sum b_{jt} W_j}{anti \ln \sum b_{jt=0} W_j} x100; \qquad (7)$$

Na wykresach 5, 6 i 7 przedstawiono kształtowanie się średnie, mediany oraz ceny hedonicznej mieszkań na rynku w wtórnym w Poznaniu w latach 2008-2009, indeksy cen mieszkań oraz procentowe zmiany rok do roku wyznaczone w oparciu o te miary.

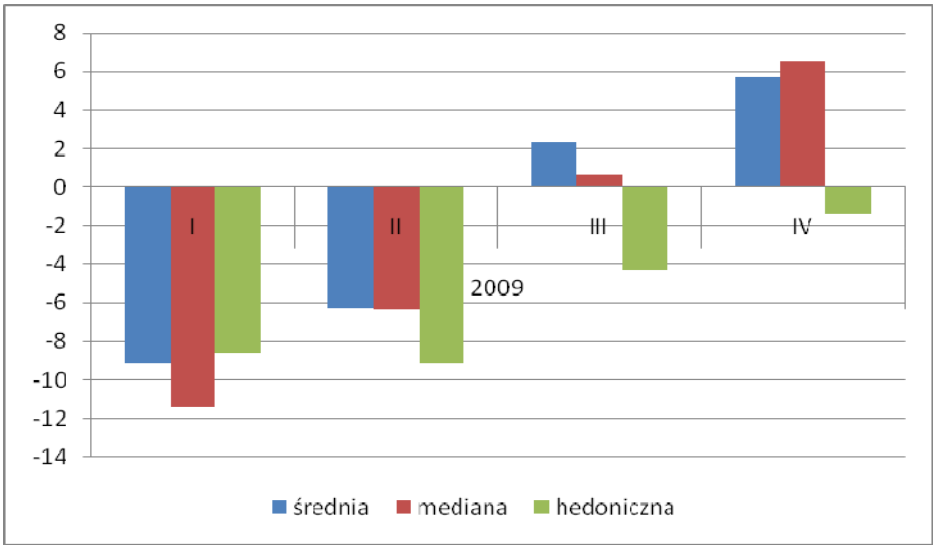
Analiza wykresów 5,6 i 7, pozwala zauważyć, że mimo tego, że wyznaczone miary mają podobny przebieg to, co do wartości różnią się. Różnice między przebiegiem średniej a mediany cen mieszkań są nieznaczne w porównaniu do przebiegu ceny hedonicznej. Miara średniej przyjmuje wartości wyższe niż mediana, co potwierdza ogólnie znaną prawdę, że ceny mieszkań nie mają rozkładu normalnego i charakteryzuje je asymetria prawostronna.



Wykres 5. Średnia, mediana oraz cena hedoniczna mieszkań w Poznaniu w latach 2008-2009 (w zł). *Źródło*: Opracowanie na podstawie danych CARN.



Wykres 6. Indeksy cen mieszkań wyznaczone w oparciu na średniej, medianie oraz cenie hedonicznej mieszkania w Poznaniu w latach 2008-2009. *Źródło*: Opracowanie na podstawie danych CARN.



Wykres 6. Procentowe zmiany średniej oraz średniej ważonej cen nieruchomości mieszkaniowych rok do roku w Poznaniu w latach 2008-2009 (w %). *Źródło*: Opracowanie na podstawie danych CARN.

Indeksy cen zbudowane na tych miarach pokazują różne zależności. W okresie między I kwartałem 2008 r a IV kwartałem 2009 r. najmniejszy spadek odnotował indeks oparty na średniej – 3,2%, następnie medianie – 4,6% i największy oparty na cenie hedonicznej -7,1%. Różnica między indeksem oparty na średniej a cenie hedonicznej wyniosła ponad 100%. Wynika to, z faktu, że indeks oparty na cenie hedonicznej pokazuje kształtowanie się ceny mieszkania reprezentatywnego w I kw. 2008 r. Na wykresach 1,2,3 i 4 pokazano strukturę mieszkań ze względu na przyjęte kryteria. Struktury te ulegały zmianom, co powodowało, że proste miary nie mogły w pełni odwzorować zachodzących zmian. Zbudowane indeksy różni również przebieg – indeksy oparte na metodach prostych zaczynają wzrastać od II kw. 2009 r., natomiast na regresji hedonicznej od III kw. 2009 r. Analiza procentowych zmian rok do roku (eliminuje wpływ wahań sezonowych) w przypadku metod prostych oraz metody hedonicznej dostarczy różnych informacji. Procentowe zmiany cen mieszkań rok do roku w przypadku średniej oraz mediany w III kw. i IV kw. 2009 r osiągają wartości dodatnie, natomiast w przypadku ceny hedonicznej przez cały 2009 r. przyjmują wartości ujemne.

4. Podsumowanie

Zważywszy na duży wpływ rynku nieruchomości mieszkaniowych na rozwój systemów gospodarczych, w wielu krajach konstruowane są indeksy cen nieruchomości mieszkaniowych mające na celu uchwycenie aktualnych zmian w ich poziomie. Wynika to z faktu, iż po pierwsze, poziom cen mieszkań wpływa na decyzje przedsiębiorstw budowlanych o rozpoczęciu nowych projektów

inwestycyjnych. Boom na rynku budowlanym powoduje wzrost zatrudnienia również w sektorach powiązanych z nim. Wzrost cen mieszkań prowadzi do wzrostu opłacalności takich inwestycji. Po drugie, ceny mieszkań mogą wywierać wpływ na popyt gospodarstw domowych – wyższe ceny oznaczają dla właścicieli nieruchomości wzrost bogactwa, co może się przełożyć na wyższą konsumpcję (GIROUARD, SVEINBOJRN 2001).

W opracowaniu podjęto próbę porównania metody prostych oraz regresji hedonicznej jako miar wartości nieruchomości w czasie. Przeprowadzona analiza wykazała, że metody proste w bardzo podobny sposób pokazują zmiany na rynku mieszkaniowym w Poznaniu w latach 2008-2009. Wykorzystana w badaniu metoda regresji hedonicznej, która uwzględnia część zmian zachodzących na rynku, dostarcza znacznie różniących się informacji na temat zachowania się cen mieszkań w okresie tylko 2 lat. Wydaje się niezmiernie istotnym wybór metody konstruowania indeksów cen mieszkań, gdyż osiągnięte rezultaty przy zastosowaniu metod prostych prowadzić mogą do błędnych interpretacji.

5. Literatura

BOURASSA S.C., HOESLI M., SUN J. 2006. *A Simple Alternative House Price Index Method*, Journal of Housing Economics 2006, vol. 15.

CASE B., WACHTER S. 2005. *Residential Real Estate Price Indices as Financial Soundness Indicators: Methodological Issues*. BIS Paper 2005, nr 21.

CHAU K. W., WONG S. K., YIU C. Y., LEUNG H. R. 2005. *Real Estate Price Indices in Hong Kong*. Journal of Real Estate Literature 2005, vol. 13, nr 5.

ENGLUND P., QUIGLEY J. M., REDFEARN C. L. 1999. The *Choice of Methodology for Computing Housing Price Indexes: Comparisons of Temporal Aggregation and Sample Definition*. Journal of Real Estate Finance and Economics 1999, vol. 19.

FLEMING M.C, NELLIS J.G. 1994, *The Measurement of UK House Prices: a Review and Appraisal of the Principal Sources*, Journal of Housing Finance 1994, vol. 24.

GIROUARD N., SVEINBOJRN B., 2001. *House Prices and Economic Activity*. OECD Economics Department Working Papers 2001.

HANSEN J. 2006. *Australian House Prices: A Comparison of Hedonic and Repeat-sales Measures*. Reserve Bank of Australia 2006.

MARK J. H., GOLDBERG M. A. 1984. *Alternative House Price Indices: An Evaluation*. AREUA Journal 1984, vol. 12, nr 1.

MEESE R., WALLACE N. 1997. *The Construction of Residential Housing Price Indices: a Comparison of Repeat Sales, Hedonic Regression, and Hybrid Approaches*. Journal of Real Estate Finance and Economics 1997, vol. 14, nr 1/2.

TROJANEK R. 2008. *Wahania cen na rynku mieszkaniowym*. Wyd. Akademii Ekonomicznej w Poznaniu, Poznań.

TROJANEK R. 2009. *Porównanie metody średniej oraz średniej ważonej konstruowania indeksów cen nieruchomości mieszkaniowych*. Studia i Materiały Towarzystwa Naukowego Nieruchomości, vol. 17, nr 2.

WOOD R. 2005. *A Comparison of UK Residential House Price Indices*. BIS Paper 2005, nr 21.

THE CONSTRUCTION OF RESIDENTIAL HOUSE PRICE INDEXES - A COMPARISON OF SIMPLE MEAN, MEDIAN AND HEDONIC REGRESSION APPROACHES

Radosław Trojanek
Departament of Investment and Real Estate
Poznań University of Economics
e-mail: r.trojanek@ue.poznan.pl

Key words: residential market, house price indexes

Abstract

The main aim of the paper is the to compare simple mean, median and hedonic regression approaches as methods used in computing house price indexes on the secondary housing market in the years 2008-2009 in Poznań. The subject scope results from the aim of the paper and includes price's on the secondary housing market, involving both property rights and cooperative property rights for private accommodation. In this research only dwellings located in multifamily buildings are analyzed.

STATYSTYCZNA ANALIZA ŁÓDZKIEGO RYNKU NIERUCHOMOŚCI

Konrad Żelazowski
Katedra Inwestycji i Nieruchomości
Uniwersytet Łódzki
e-mail: kzelazowski@uni.lodz.pl

Słowa kluczowe: *statystyczna analiza rynku nieruchomości, łódzki rynek nieruchomości*

Streszczenie

Metody statystyczne oferują szerokie spektrum ilościowej analizy rynku nieruchomości. Ze względu na swoje niezaprzeczalne zalety stanowić powinny obowiązkowe narzędzie analityczne rzeczoznawców majątkowych, inwestorów, doradców inwestycyjnych oraz pozostałych uczestników rynku nieruchomości.

Celem opracowania jest zaprezentowanie dynamiki zmian cen dla wybranych segmentów łódzkiego rynku nieruchomości oraz oszacowanie wpływu wybranych atrybutów nieruchomości na ich cenę.

1. Wstęp

Nieodzownym elementem procesu szacowania wartości nieruchomości jest szczegółowa analiza lokalnego rynku nieruchomości. Podstawowym celem takiej analizy jest dostarczenie rzeczoznawcy majątkowemu istotnych informacji dotyczących uwarunkowań ekonomicznych, społecznych i przestrzennych funkcjonowania rynku lokalnego, określenie przeciętnego poziomu cen nieruchomości, wyznaczenie trendu czasowego ich zmian, zidentyfikowanie głównych atrybutów nieruchomości wraz z określeniem ich wpływu na wartość [GAWRON, 2009].

Zaprezentowana w artykule statystyczna analiza łódzkiego rynku nieruchomości dotyczy:

- rynku nieruchomości lokalowych mieszkalnych,
- rynku nieruchomości gruntowych pod zabudowę.

Przedmiotowa analiza zrealizowana została w oparciu o udostępnioną przez środowisko łódzkich rzeczoznawców majątkowych bazę danych zawierającą transakcje lokalami mieszkalnymi oraz nieruchomościami gruntowymi w latach 2005-poł. 2009 [ŻELAZOWSKI, 2009]. Niniejsza baza obejmuje **wszystkie** transakcje zrealizowane na terenie Łodzi, jest zatem kompletnym źródłem informacji o rynku lokalnym.

2. Rynek lokali mieszkalnych

W badanym okresie, głównym czynnikiem kształtującym aktywność rynku lokali mieszkalnych był zmieniający się potencjał finansowy gospodarstw

domowych, który należy traktować szeroko uwzględniając zarówno środki własne kupujących (poziom dochodów do dyspozycji powiększony o zakumulowane oszczędności) oraz kwotę kapitałów obcych, głównie kredytów hipotecznych, możliwych do pozyskania z sektora bankowego. Malejące stopy procentowe oraz rosnąca dostępność kredytów hipotecznych stymulowały wzrost liczby transakcji lokalami mieszkalnymi na rynku wtórnym (rys. 1). Okresem przełomowym był rok 2006, charakteryzujący się największą liczbą zawartych transakcji (1496 transakcji). Wraz z rosnącym poziomem cen oraz pierwszymi sygnałami globalnego kryzysu finansowego aktywność badanego rynku ewidentnie zmalała. Stan rynku wtórnego miał istotny wpływ na sytuację rynku pierwotnego. Wraz z rosnącymi cenami lokali mieszkalnych na rynku wtórnym rosła opłacalność przedsięwzięć deweloperskich oraz liczba transakcji nowymi lokalami (rys. 2).

Zmiany aktywności rynku lokali znalazły swoje odzwierciedlenie w poziomach cen. Dla wtórnego rynku lokali mieszkalnych funkcją trendu, która najpełniej odzwierciedla dynamikę zmian cen w kolejnych kwartałach 2005-2009 r. jest wielomian stopnia 3 postaci:

$$y_t = -3{,}7647 \cdot t^3 + 95{,}728 \cdot t^2 - 456{,}34 \cdot t + 1894 + \xi_t \qquad (1)$$

y_t – średnia cena 1 m² pow. użytkowej lokalu,

t – numer kolejnych kwartałów: 1,2,3,…,17,

ξ_t – składnik losowy.

Funkcja trendu wskazuje na pewne tendencje zmian cen: stabilizację cen lokali mieszkalnych w okresach 1Q2005-2Q2006 oraz 3Q2007-4Q2008, a także na okres gwałtownego wzrostu cen w przedziale czasowym 3Q2006-2Q2007 (rys. 3).

Dla pierwotnego rynku lokali mieszkalnych funkcją trendu, która najpełniej odzwierciedla dynamikę zmian cen lokali w kolejnych kwartałach 2005-2008 r. jest również wielomian stopnia 3 postaci:

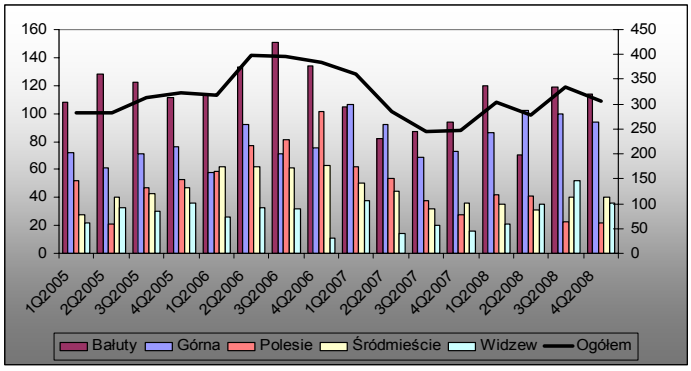
$$y_t = -1{,}9631 \cdot t^3 + 64{,}112 \cdot t^2 - 422{,}53 \cdot t + 2851 + \xi_t \qquad (2)$$

y_t – średnia cena 1 m² pow. użytkowej lokalu,

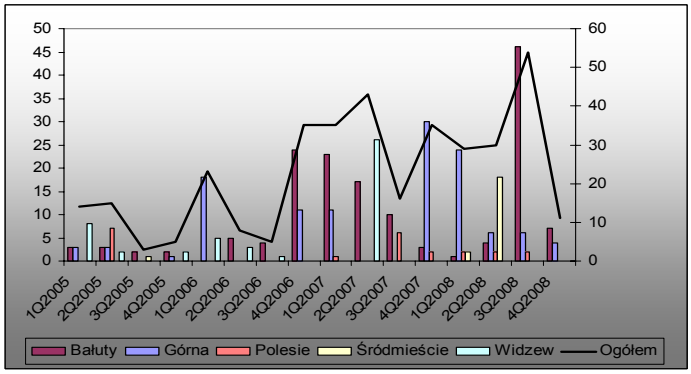
t – numer kolejnych kwartałów: 1,2,3,…,16,

– składnik losowy.

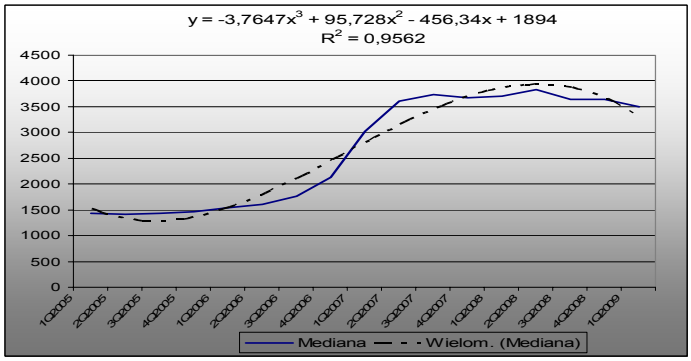
Dynamiczny wzrost cen lokali mieszkalnych na rynku pierwotnym stanowił konsekwencję gwałtownego wzrostu popytu na lokalne mieszkalne po 2005 r. Znaczna część tego popytu skierowana została na rynek wtórny, jednak wraz ze wzrostem cen nieruchomości rosła opłacalność inwestycji deweloperskich, liczba nowych lokali sprzedawanych przy wyższych poziomach cen (rys. 4).



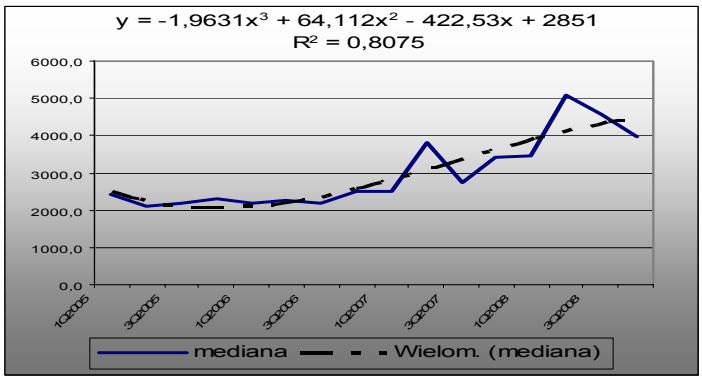
Rys. 1. Liczba transakcji lokalami mieszkalnymi – rynek wtórny.
Źródło: opracowanie własne.



Rys. 2. Liczba transakcji lokalami mieszkalnymi – rynek pierwotny.
Źródło: opracowanie własne.



Rys. 3. Trend czasowy zmian cen 1 m² p.u. lokali mieszkalnych-rynek wtórny.
Źródło: opracowanie własne.

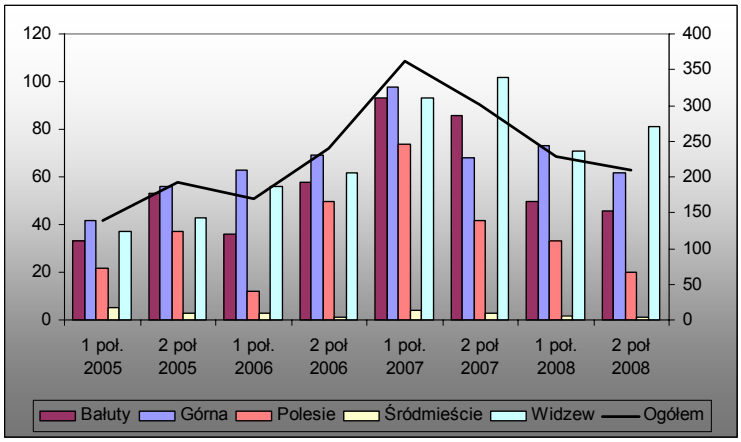


Rys. 4. Trend czasowy zmian cen 1 m² p.u. lokali mieszkalnych -rynek pierwotny.
Źródło: opracowanie własne.

3. Nieruchomości gruntowe pod zabudowę

Drugim segmentem łódzkiego rynku nieruchomości stanowiącym przedmiot niniejszej analizy był rynek gruntów pod zabudowę. Segment ten obejmował transakcje gruntami z wyłączeniem gruntów przeznaczonych w planie zagospodarowania przestrzennego na cele rolne oraz gruntów przeznaczonych pod komunikację i zieleń.

Badany rynek pozostawał w ścisłym związku z rynkiem mieszkaniowym. Wraz ze wzrostem liczby nowych inwestycji mieszkaniowych realizowanych na terenie miasta rosło zapotrzebowanie na nieruchomości gruntowe pod zabudowę (rys. 5). Faza szczytowa aktywności tego rynku przypadła na 2007r. (362 transakcje), a więc rok po maksimum aktywności rynku mieszkaniowego.



Rys. 5. Liczba transakcji gruntami pod zabudowę**.**
Źródło: opracowanie własne.

Dla rynku gruntów pod zabudowę funkcją trendu, która dobrze odzwierciedla dynamikę zmian cen tych nieruchomości w kolejnych kwartałach 2005-2009r. jest wielomian stopnia 3 postaci:

$$y_t = -0{,}0786 \cdot t^3 + 2{,}2229 \cdot t^2 - 13{,}969 \cdot t + 70{,}116 + \xi_t \qquad (3)$$

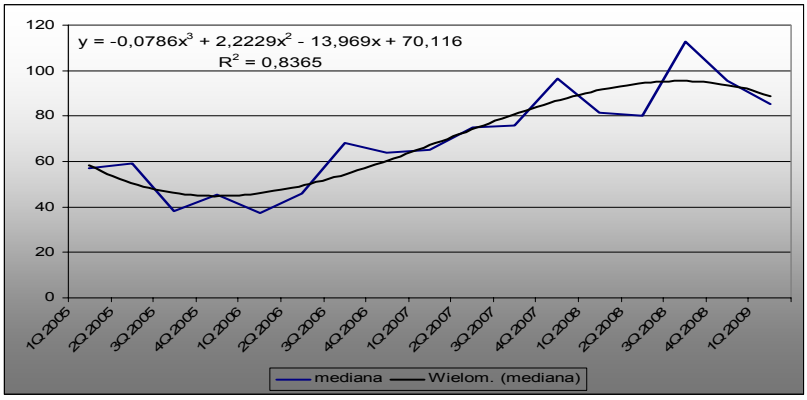
gdzie:

y_t – średnia cena 1 m^2 gruntu pod zabudowę,

t – numer kolejnych kwartałów: 1,2,3…17,

ξ_t – składnik losowy.

Funkcja trendu wskazuje na silny wzrost cen gruntów pod zabudowę w okresie od 2Q 2006 do 4Q 2007r (rys. 6).



Rys. 6. Trend czasowy zmian średniej ceny 1 m^2 gruntu pod zabudowę.
Źródło: opracowanie własne.

4. Analiza siły i kierunku powiązań pomiędzy cenami a wybranymi atrybutami nieruchomości

Oprócz badania aktywności wybranych segmentów łódzkiego rynku nieruchomości oraz identyfikacji trendów czasowych zmian cen przedmiotem analizy była próba określenia związku pomiędzy wybranymi atrybutami nieruchomości a ich cenami. Wybór atrybutów podyktowany był dostępnością danych statystycznych jak również ich obiektywnym charakterem[5]. W analizie uwzględniono następujące atrybuty:

1. Rynek lokali mieszkalnych:

[5] Zasadniczym celem opracowania było zweryfikowanie powiązania wybranych atrybutów nieruchomości i ich cen, nie zaś budowa modelu wyceny nieruchomości, dlatego też do analizy przyjęte zostały tylko te atrybuty, których sposób pomiaru nie budził wątpliwości. Pominięte zostały atrybuty o charakterze ocennym np. lokalizacja, stan techniczny.

a) Wiek budynku.
b) Numer kondygnacji.
c) Powierzchnia użytkowa.

2. Rynek nieruchomości gruntowych pod zabudowę:
   a) Powierzchnia działek

Przy doborze postaci funkcyjnej szacowanych modeli uwzględniono dwa podstawowe kryteria: zgodność modelu z teorią ekonomii i wyceny nieruchomości oraz stopień dopasowania modelu do danych empirycznych [GAJDA, 2004].

4.1 A. Analiza zależności pomiędzy ceną 1 m² p.u. lokalu mieszkalnego a wiekiem budynku

W celu zweryfikowania zależności pomiędzy ceną 1 m² lokalu mieszkalnego a wiekiem budynku oszacowano funkcję regresji postaci:

Tabela 1

Wyniki estymacji funkcji regresji

	Współczynniki	*Błąd standardowy*	*t Stat*	*Wartość-p*
Wyraz wolny	**1249,481**	23,01917	54,28001	0
40te	**-126,003**	165,4607	-0,76153	0,446452
50te	**168,5519**	29,10287	5,79159	8,34E-09
60te	**199,2627**	25,08864	7,942345	3,65E-15
70te	**196,9991**	32,44816	6,071195	1,57E-09
80te	**323,7389**	32,60788	9,92824	1,34E-22
90te	**595,2036**	55,92866	10,64219	1,27E-25
>2000	**703,7232**	37,3621	18,83521	8,57E-72
R kwadrat		0,216866		
Dopasowany R kwadrat		**0,213521**		
Błąd standardowy		283,7994		

Źródło: opracowanie własne

$$y = \alpha_0 + \alpha_1 \cdot Z1 + \alpha_2 \cdot Z2 + \alpha_3 \cdot Z3 + \alpha_4 \cdot Z4 + \alpha_5 \cdot Z5 + \alpha_6 \cdot Z6 + \alpha_7 \cdot Z7 + \xi \quad (4)$$

y -cena 1 m² p.u. lokalu mieszkalnego,

α_i -parametry modelu,

Zi – zmienne zero-jedynkowe reprezentujące lokale mieszkalne umiejscowione w:
Z1 – budynkach wybudowanych w latach 40tych,
Z2 – budynkach wybudowanych w latach 50tych,
Z3 – budynkach wybudowanych w latach 60tych,
Z4 – budynkach wybudowanych w latach 70tych,
Z5 – budynkach wybudowanych w latach 80tych,
Z6 – budynkach wybudowanych w latach 90tych,
Z7 – budynkach wybudowanych w latach 2000-2009,

ξ - składnik losowy.

Interpretacja wyników:

W oparciu o wyestymowaną funkcję regresji należy stwierdzić, iż zróżnicowanie w wieku (roku budowy) budynków, w których usytuowane były przedmiotowe lokale mieszkalne wyjaśnia w około 20% zróżnicowanie cen lokali mieszkalnych. Ponadto uwzględniając wartości współczynników przy kolejnych zmiennych można określić ilościowe relacje pomiędzy cenami lokali w budynkach o różnym wieku. Zgodnie z przedstawioną funkcją regresji w budynkach przedwojennych należy oczekiwać średniej ceny 1249,48 zł za m^2. Lokale w budynkach z lat 40-tych powinny być nieznacznie tańsze od lokali przedwojennych o około 126 zł za 1 m^2 (10%, warto jednak zwrócić uwagę na statystyczną nieistotność zmiennej Z1). Lokale w budynkach z lat 50-tych powinny być średnio droższe o 168, 55 zł (13%) od lokali w budynkach przedwojennych, lokale z lat 60-tych o 199,3 zł (16%), w budynkach z lat 70-tych o 197 zł (16%), w budynkach z lat 80-tych o 324 zł (26%), w budynkach z lat 90-tych o 595 zł (48%), w budynkach wybudowanych po 2000 r. o 703 zł (56%) droższe za 1 m^2 od lokali w budynkach przedwojennych[6].

4.1 B. Analiza zależności pomiędzy ceną 1 m^2 p.u. lokalu mieszkalnego a kondygnacją, na której lokal jest usytuowany

W celu zweryfikowania zależności pomiędzy ceną 1 m^2 p.u. lokalu mieszkalnego a kondygnacją, na której usytuowany jest lokal oszacowano model postaci:

$$y = \alpha_1 \cdot x^2 + \alpha_2 \cdot x + \alpha_3 + \xi \qquad (5)$$

gdzie:

y - cena 1 m^2 p.u. lokalu mieszkalnego

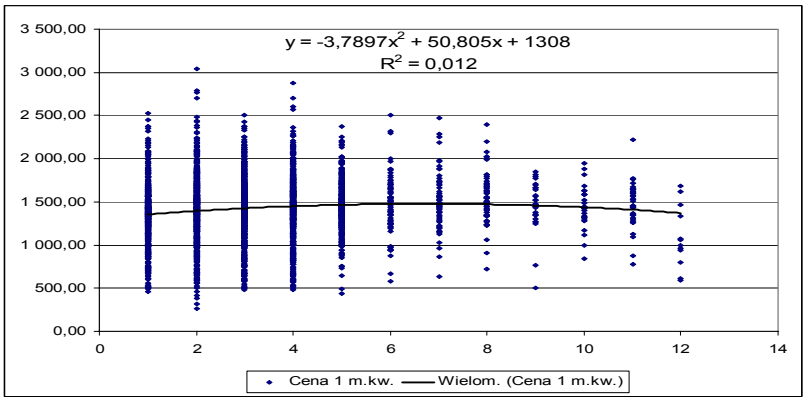
x - numer kondygnacji na której znajduje się lokal

α_i - parametry modelu

ξ - składnik losowy

Zróżnicowanie w usytuowaniu lokali mieszkalnych na kolejnych kondygnacjach budynku w niewielkim stopniu wyjaśnia zróżnicowanie cen lokali mieszkalnych (1,2% - rys. 7). Numer kondygnacji, na której usytuowany jest lokal mieszkalny nie jest zatem zasadniczym czynnikiem wyjaśniającym zróżnicowanie cen nieruchomości lokalowych. Warto jednak podkreślić, iż model potwierdza powszechne przekonanie o preferowaniu przez uczestników rynku lokali na „środkowych" kondygnacjach.

[6] Należy zwrócić uwagę, iż prezentowane różnice cenowe zostały określone na 1Q 2005r. Ceny wszystkich transakcji lokalami mieszkalnymi po 1 kwartale 2005 zostały skorygowane o trend czasowy zmian cen nieruchomości. Celem korekty było wyeliminowanie wpływu czasu na zróżnicowanie cen lokali mieszkalnych w latach 2005-2009.



Rys. 7. Zależność pomiędzy ceną 1 m² p.u. lokalu mieszkalnego a kondygnacją, na której lokal jest usytuowany[7]
Źródło: opracowanie własne

Interpretacja wyników

4.1 C. Analiza zależności pomiędzy ceną 1 m² p.u. lokalu mieszkalnego a powierzchnią użytkową lokalu

W celu zweryfikowania zależności pomiędzy ceną 1 m² p.u. lokalu mieszkalnego a powierzchnią użytkową lokalu, oszacowano model postaci:

$$y = \alpha_1 \cdot x^2 + \alpha_2 \cdot x + \alpha_3 + \xi \qquad (6)$$

gdzie:

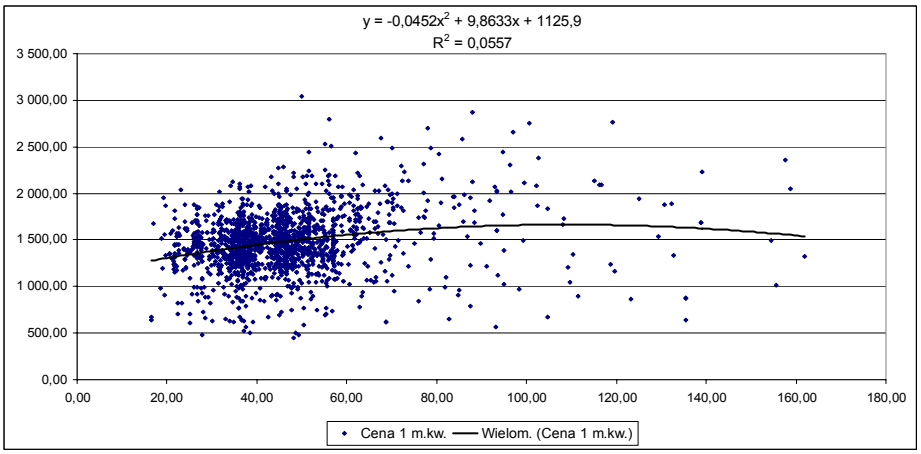
y - cena 1 m² lokalu mieszkalnego

x - powierzchnia użytkowa lokalu mieszkalnego

α_i - parametry modelu

ξ - składnik losowy

[7] Ceny lokali zostały skorygowane o trend zmian cen w badanym okresie 2005-2009 (ceny na stan 1Q 2005r.)



Rys. 8. Zależność pomiędzy ceną 1 m² p.u. lokalu mieszkalnego a powierzchnią użytkową lokalu[8]. *Źródło*: opracowanie własne.

Interpretacja wyników

Różnice w powierzchni użytkowej lokali mieszkalnych zgodnie z wynikami oszacowanego modelu wyjaśniają w 5% zróżnicowanie cen tychże lokali (rys. 8). Powierzchnia użytkowa nie jest zatem zasadniczym atrybutem lokali mieszkalnych wpływającym na zróżnicowanie cen 1 m² p.u. lokali mieszkalnych.

5.2. Zależność pomiędzy ceną 1 m² gruntu pod zabudowę a powierzchnią działek

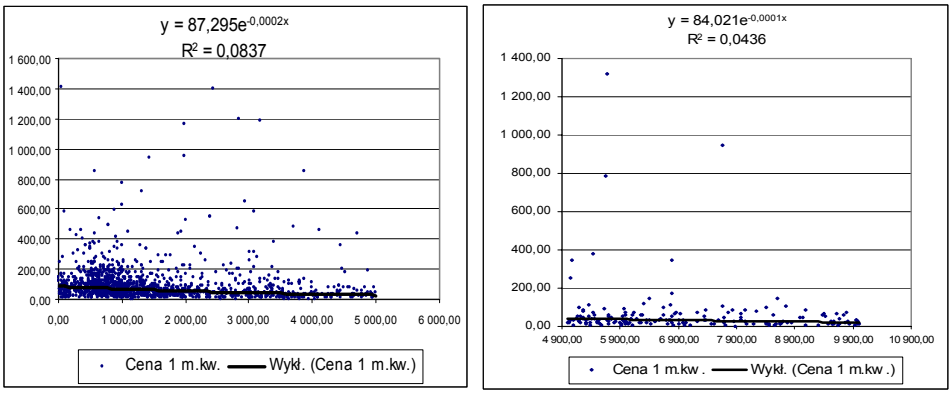
Przedmiotem analizy była zależność pomiędzy ceną 1 m² gruntu pod zabudowę a powierzchnią działek[9]. Ze względu na duże zróżnicowanie powierzchni działek dla poszczególnych transakcji analiza została przeprowadzona dla wybranych przedziałów powierzchni:

- transakcje nieruchomościami o pow. działek do 5000 m² (rys. 9);
- transakcje nieruchomościami o pow. działek w przedziale 5000 – 10000 m² (rys. 10);
- transakcje nieruchomościami o pow. działek w przedziale 10000 – 20000 m² (rys. 11);
- transakcje nieruchomościami o pow. działek powyżej 20000 m² (rys. 12).

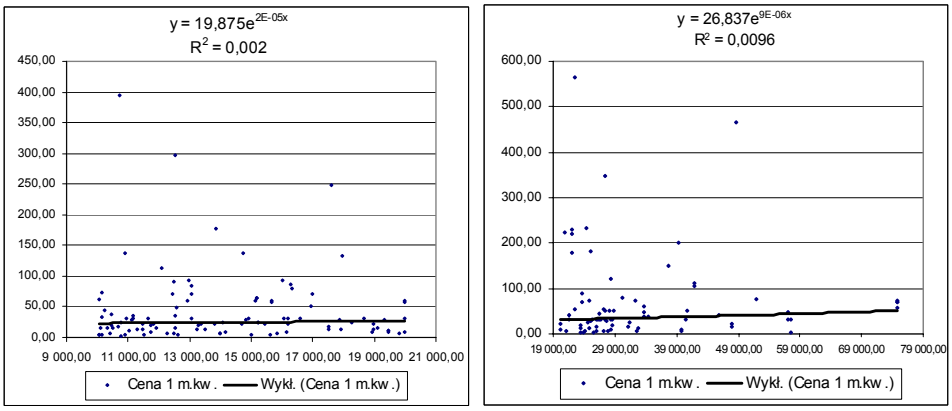
Dla rynku gruntów pod zabudowę o powierzchni do 5000 m² różnice w powierzchni działek wyjaśniają w 8% różnice w poziomie cen 1 m². Zgodnie z wynikami oszacowanego modelu wraz ze wzrostem powierzchni działek o 1 m² ceny średnio spadają o 0,02%.

[8] Ceny lokali zostały skorygowane o trend zmian cen w badanym okresie 2005-2009 (ceny na stan 1Q 2005r.)

[9] Ceny gruntów zostały skorygowane o trend zmian cen w badanym okresie 2005-2008 (ceny na stan 1Q 2005r.)



Rys. 9. i 10. Zależność pomiędzy ceną 1 m² gruntu pod zabudowę a powierzchnią działek



Rys. 11. i 12. Zależność pomiędzy ceną 1 m² gruntu pod zabudowę a powierzchnią działek
Źródło: opracowanie własne

Interpretacja wyników

Dla rynku gruntów pod zabudowę o powierzchni 5000 -10000 m² różnice w powierzchni działek wyjaśniają w 4% różnice w poziomie cen 1 m². Zgodnie z wynikami oszacowanego modelu wraz ze wzrostem powierzchni działek o 1 m² ceny średnio spadają o 0,01%.

Dla rynku gruntów pod zabudowę o powierzchni 10000 - 20000 m² różnice w powierzchni działek wyjaśniają tylko w 0,2% różnice w poziomie cen 1 m². Wartość współczynnika R^2 bliski zeru wskazuje na brak istotnego związku pomiędzy ceną 1 m² gruntu a powierzchnią działek. Potwierdzają to również wartości oszacowanych parametrów modelu. Zgodnie z wynikami oszacowanego modelu wraz ze wzrostem powierzchni działek o 1 m² ceny średnio rosną

o 0,002%[10].

Dla rynku gruntów pod zabudowę o powierzchni powyżej 20000 m² różnice w powierzchni działek wyjaśniają tylko w 0,9% różnice w poziomie cen 1 m². Wartość współczynnika R^2 bliski zeru wskazuje na brak istotnego związku pomiędzy ceną 1 m² gruntu a powierzchnią działek. Potwierdzają to również wartości oszacowanych parametrów modelu. Zgodnie z wynikami oszacowanego modelu wraz ze wzrostem powierzchni działek o 1 m² ceny średnio rosną o 0,0009%.

W oparciu o wyniki oszacowanych modeli dla wyszczególnionych przedziałów transakcji gruntami pod zabudowę należy stwierdzić, że wraz ze wzrostem powierzchni działek spada jej znaczenie w kształtowaniu ceny 1 m². Najsilniejsza zależność pomiędzy ceną 1 m² gruntu a jego powierzchnią zachodzi dla gruntów o powierzchni do 5000 m², dla gruntów o powierzchni powyżej 10000 m² znaczenie powierzchni jest marginalne.

6. Podsumowanie

Łódzki rynek lokali mieszkalnych w latach 2005-2008 doświadczał gwałtownych zmian cen, (z historycznie wysokim 42% wzrostem cen lokali rynku wtórnego w 1Q 2007r.). Powyższy proces był zbieżny z tendencją ogólnokrajową wywołaną boomem kredytowym. Rynek gruntów pod zabudowę pozostawał w badanym okresie w ścisłym związku z rynkiem mieszkaniowym, z fazą szczytową aktywności i wzrostu cen w 2007r.

Statystyczna analiza rynku łódzkiego umożliwiła również zweryfikowanie istotności wybranych atrybutów nieruchomości w kształtowaniu ich cen, dzieląc je na cechy istotne (wiek budynku) oraz cechy drugorzędne (kondygnacja, powierzchnia użytkowa lokalu mieszkalnego, powierzchnia działki).

7. Literatura:

GAJDA J.B. 2004. *Ekonomeria*, Wyd. C.H. Beck. Warszawa.

GAWRON H. 2009. *Analiza rynku nieruchomości. Materiały dydaktyczne 224*, Wyd. Akademii Ekonomicznej w Poznaniu.

ŻELAZOWSKI K. 2009. *Wybrane elementy analizy statystycznej łódzkiego rynku nieruchomości*, maszynopis, Łódź.

[10] Wartość parametru przy zmiennej reprezentującej powierzchnię działek jest bliski zeru co oznacza brak istotnej zależności pomiędzy powierzchnią działek i ceną 1 m². Bez znaczenia pozostaje tu zmiana kierunku wpływu powierzchni na cenę z dotychczasowego negatywnego na pozytywny.

STATISTICAL ANALYSIS OF REAL ESTATE MARKET IN LODZ

Konrad Żelazowski
Department of Investments and Real Estate
University of Lodz
e-mail: kzelazowski@uni.lodz.pl

Key words: statistical analysis of real estate market, real estate market in Lodz

Abstract

Statistical methods offer a wide range of quantitative analyses of real estate market. Considering their undeniable advantages they should be an obligatory technique used by real estate valuers, investors, investment advisors and other market participants.

The aim of the article is to present dynamics of real estate prices in Lodz and to estimate the influence of selected real estate attributes on price level.

ISSN 1733-2478
ISBN 978-83-61564-24-9