



ZASTOSOWANIE WYBRANYCH METOD STATYSTYCZNYCH DO ANALIZY RYNKU NIERUCHOMOŚCI

*Janusz Dąbrowski, Państwowa Wyższa Szkoła Zawodowa im. ks. Bronisława Markiewicza
w Jarosławiu, Instytut Inżynierii Technicznej*

*„Dane należy tak długo torturować, aż zaczną zeznawać...”
z wykładu prof. A. Małeckiego*

Wstęp

Rzeczywistość możemy opisać na wiele sposobów. W naukach ścisłych badacze starają się przede wszystkim znaleźć związek funkcyjny lub stochastyczny pomiędzy badanymi wielkościami. Jest to trudne zadanie i w wielu przypadkach wręcz niemożliwe do wykonania. W bardzo skomplikowanej i niezwykle szybko zmieniającej się rzeczywistości najczęstszym powodem braku jednoznacznych rozwiązań jest nieuwzględnienie odpowiednich danych w modelu matematycznym lub mała wiarygodność użytych danych.

Podstawowym problemem analizy rynku nieruchomości w Polsce, wielokrotnie wskazywanym w literaturze fachowej i raportach bankowych, jest brak wiarygodnych danych. W sprawozdaniach banków publikowanych za 2007 rok, można przeczytać następujące sformułowania:

- ◆ „Podmioty mające dostęp do informacji często zniekształcają je ze względu na własne interesy.”,
- ◆ „Publikowane w latach 2004-2005 średnie poziomy cen są kilkukrotnie zaniżone w stosunku do ogólnodostępnych cen ofertowych, ponadto nie obejmują umów przedwstępnych z rynku pierwotnego, które w obecnej sytuacji są właściwym miernikiem cen rynku na rynku pierwotnym.”.

Raporty analityków sektora bankowego bardzo jednoznacznie podważają wiarygodność cen transakcyjnych zawartych w aktach notarialnych. Powszechnie wiadomo, że tam, gdzie strony dążą do zminimalizowania opłat z tytułu transakcji, zazwyczaj następuje zaniżenie ceny, a dla celów kredytowych nagminnie zawiąza się ceny w umowach przedwstępnych.

W przypadku rynku nieruchomości widoczne są działania spekulacyjne połączone z manipulacją mającą na celu wywołać działania stadne. Tabela 1 ilustruje kryzysy nieruchomościowe w latach 1974-1995 w Europie.

Tabela 1. Kryzysy na rynkach nieruchomości w latach 1974-1995 powiązane z kryzysami bankowymi.

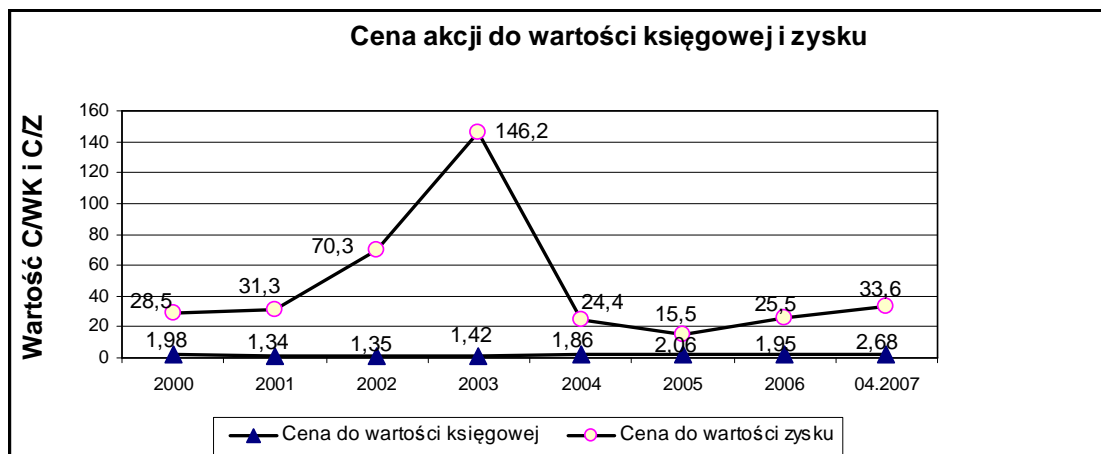
Kraje 1974-1995	1974	1975	1976	1977	1978	1979	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995
EUROPA																						
W. Brytania (1974-1976)																						
Niemcy (1976-1980)																						
Hiszpania (1977-1985)																						
Norwegia (1987-1989)																						
Szwecja (1991)																						
Finlandia (1991-1993)																						
Francja (1994-1995)																						

Źródło: opracowanie własne na podstawie: *Studia i Materiały Towarzystwa Naukowego Nieruchomości* - vol. 15.

Regularne schodkowe ułożenie przedziału czasowego kryzysu sugeruje planowy przepływ kapitału spekulacyjnego do kolejnych krajów. W latach 2007-2009 rynek nieruchomości w Polsce był zdaniem autora miejscem działań spekulacyjnych wzmocnionych niebywałymi wręcz manipulacjami medialnymi wywołującymi ściśle określone działania stadne. Od 2008 roku obserwujemy wyraźny „atak” na Bułgarię. Akademickim wręcz przykładem skutecznych działań oddziaływania na emocje społeczne jest rok 2003, kiedy to jedna z polskich firm giełdowych wykazywała 146 razy mniejszy zysk niż wartość akcji, a pomimo tego znajdowała nabywców. Rys. 1 ilustruje relacje ceny akcji do wartości księgowej i wartości zysku przykładowej spółki giełdowej.

Działania spekulacyjne w sposób absolutnie nieprzewidywalny są w stanie zaburzyć jakiegokolwiek prognozy. Zmienne opisujące działania spekulacyjne mają charakter zmiennych „nieznanych niewiadomych”, ponieważ są to działania pozaprawne i nieformalne. Jedyne, co możemy zrobić, to próbować określić parametry, jakimi kierują się spekulanci przy podejmowaniu badań. Istnienie nieznanych niewiadomych bardzo trafnie zdefiniował Donald Rumsfeld, sekretarz obrony USA w gabinecie George’a Busha, który równocześnie znakomicie wyjaśnił przyczyny zawodności prognoz, mówiąc: „Są znane wiadome. To rzeczy, o których wiemy, że je znamy. Są znane niewiadome. Czyli rzeczy, o których wiemy, że ich nie znamy. Ale są też nieznanne niewiadome. To rzeczy, których nie wiemy, że ich nie znamy”. Jeżeli nawet potrafimy zdefiniować niewiadome, to nie zawsze znamy

ich wartość. W oparciu o znane zależności pomiędzy badanymi wielkościami możemy zbudować model stochastyczny. Możemy go oceniać na wiele sposobów. Zdaniem autora najlepszą weryfikacją modelu opisującego rzeczywistość po sprawdzeniu jego poprawności statystycznej jest porównanie prognozy opartej na konkretnym modelu z danymi empirycznymi.



Rys. 1. Cena akcji do wartości księgowej i zysku
(źródło: opracowanie własne na podstawie: GPW).

Według Sobczyka [16] zależność stochastyczna występuje wtedy, gdy wraz ze zmianą jednej zmiennej zmienia się rozkład prawdopodobieństwa drugiej zmiennej. Zależność korelacyjna (statystyczna) jest wtedy, gdy określonym zmianom jednej zmiennej odpowiadają ściśle określone średnie wartości drugiej zmiennej. Brak związku stochastycznego powoduje automatyczny brak związku korelacyjnego. Niestety twierdzenie odwrotne nie jest prawdziwe, ponieważ brak związku korelacyjnego nie wyklucza istnienia związku stochastycznego. Badacze podkreślają, że badanie związku korelacyjnego ma sens jedynie wtedy, gdy między zmiennymi istnieje związek przyczynowo-skutkowy. W literaturze podaje się przykłady korelacji pozornej lub iluzorycznej. Optymalne rozwiązanie to takie, kiedy związek pomiędzy badanymi zmiennymi ma charakter jakościowy i ilościowy. Dopiero po uzasadnieniu logicznym istnienia związku pomiędzy badanymi wielkościami powinno się określić kierunki i siły zależności. Związki pomiędzy zmiennymi mające charakter przyczynowo-skutkowy mogą być jednostronne wtedy, kiedy mamy do czynienia z jednostronnym oddziaływaniem przyczyny na skutek, i związki dwustronne wtedy, gdy mamy do czynienia ze wzajemnym oddziaływaniem na siebie badanych zmiennych. W analizie rynku nieruchomości kluczowe pozostaje pytanie: „czy rozwój gospodarki wpływa na koniunkturę na rynku nieruchomości?”, czy też zasadne jest pytanie odwrotne: „czy boom w budownictwie jest przyczyną silnego wzrostu gospodarczego?”. Rozpatrując różne rynki nieruchomości na przestrzeni dziejów, możemy się spotkać z różnymi przypadkami. Co do jednego jednak większość badaczy pozostaje zgodna: gospodarka i rynek nieruchomości wzajemnie przenikają się. Dwustronne oddziaływanie na siebie zmiennych może powodować w rozważaniach umowne przyjmowanie, która zmienna jest niezależna, a która zależna. Nie wszystkie zmienne niezależne oddziałują na zmienną



zależną w danym momencie. Często mamy do czynienia z przesunięciem czasowym. Na przykład ogłoszenie decyzji o budowie metra natychmiast powoduje wzrost wartości nieruchomości odpowiednio zlokalizowanych, pomimo że w infrastrukturze nie odnotowano żadnych zmian.

W pracy autor na przykładzie badania rynku nieruchomości pokazał szerokie spektrum możliwości pakietu *STATISTICA*. Zazwyczaj programy statystyczne wykorzystujemy do weryfikacji już zbudowanych modeli. Dzięki ogromnemu przyśpieszeniu w wykonywaniu analiz statystycznych pakiet *STATISTICA* wykorzystano nie tylko do weryfikacji postawionej hipotezy, ale również do formułowania nowych hipotez. Program daje możliwość budowy i weryfikacji modeli na wiele sposobów.

Obiektem badań autora było określenie wpływu otoczenia rynku nieruchomości na wartość nieruchomości. W dotychczasowej praktyce rzeczoznawców w podejściu porównawczym skupiano się głównie na cechach lokalnych analizowanych nieruchomości. Wpływ otoczenia rynku nieruchomości uwzględniano co najwyżej poprzez współczynnik eksperta. Zmiany cen, jakie następowały w latach 2007-2009, wyraźnie wskazywały, że na wartość nieruchomości przede wszystkim wpływa stan gospodarki. Postawioną tezę udowodniono, budując ponad siedemset modeli matematycznych i weryfikując je statystycznie przy pomocy pakietu *STATISTICA*. Wykorzystując szerokie spektrum możliwości badawczych programu, zweryfikowano pozytywnie postawioną tezę różnymi niezależnymi metodami.

Zdefiniowanie zmiennych niezależnych

Największym problemem w pracy był dobór atrybutów globalnych do tworzonego modelu. Prace poprzedzono studiami literaturowymi, a ich wyniki syntetycznie przedstawiono w tabeli 2.

Autorzy wymienionych publikacji w zależności od obszaru badawczego, jaki w danym momencie analizowali, uwypuklali jedną lub kilka zmiennych. Badacze zajmujący się tylko rynkiem nieruchomości szczególną uwagę zwracali na zmienne opisujące branżę budowlaną, a w opracowaniach opisujących gospodarkę całościowo, autorzy polecali badanie wskaźników o charakterze makroekonomicznym. W literaturze nieznane jest opracowanie, które w sposób całościowy opisywałoby badanie otoczenia rynku nieruchomości. Zazwyczaj badacze analizowali wybrane elementy rynku nieruchomości i jego otoczenia.

Autor zgromadził 205 wskaźników społeczno-gospodarczych opisujących rynek nieruchomości w okresie: od grudnia 1999 do stycznia 2008 roku w interwale miesięcznym. Podstawowym problemem była redukcja 205 zmiennych i wybór siedmiu reprezentatywnych atrybutów globalnych. Wszystkie zmienne wykazywały związek przyczynowo-skutkowy ze zmienną zależną (cena transakcyjna z rynku pierwotnego dla Warszawy Woli, Milanowa i Krakowa). Zmienne niezależne zestawiono i przypisano do sześciu grup.

Tabela 2. Określenie zmiennych niezależnych na podstawie literatury.

Lp.	Autor	Obszar badawczy
1	Bajerowski [1]	Zasoby surowcowe, skażenie środowiska, demografia
2	Banarjee [2]	Badania behawioralne
3	Baum [3]	Stopy zwrotu z nieruchomości
4	Begg [4]	Efekt synergii różnych gałęzi gospodarki
5	Burns i Mitchell [5]	Cykle koniunkturalne
6	Crone, Voith [6]	Indeksy budowlane
7	Czaja, Parzych [7]	Trend zmiany cen transakcyjnych
8	Dąbrowski [8]	Równanie obiegu pieniądza, stopień koncentracji kapitału na świecie i w kraju
9	Gdakowicz [10]	Liczba transakcji kupna-sprzedaży, liczba pozwoleń na budowę, liczba pustostanów,
10	Gościński [11]	Poziom cen surowców i energii, wpływ na gospodarkę władz politycznych
11	Iwonicz-Drozdowska [12]	Powiązania rynku nieruchomości z sytuacją banków
12	Lukas [13]	Poziom realnego PKB
13	Nykiel [14]	Wielkość emigracji zarobkowej
14	Rednet [21]	Poziom popytu inwestycyjnego
15	Rekomendacja NBP [22]	Zmiany kursowe, stopy procentowe
16	Saunders [15]	Poziom wypłacalności kredytobiorców i kredytodawców
17	Trojanek [17]	Wartość udzielonych kredytów mieszk., zmiany w zatrudnieniu, koszty w budownictwie
18	Wheaton [18]	Zgodność cyklu gospodarczego z cyklem na rynku nieruchomości
19	Żelazowski [20]	Relacje: cena 1m ² /zysk z 1 m ² , wysokość dochodów gosp. dom./cena 1 m ² mieszkania

Źródło: opracowanie własne.

Pierwszą grupę stanowiły parametry gospodarcze (45) wyselekcjonowane na podstawie literatury specjalistycznej. Wskaźniki rynku nieruchomości (19) przedstawione w grupie II oddawały relacje podaży i popytu. Trzecia grupa to wskaźniki podaży i popytu pieniądza (45) silnie ze sobą skorelowane. W IV grupie zestawiono wskaźniki syntetyczne (23) obrazujące stan gospodarki, które można podzielić na dwie grupy: analitycznie obliczone według założonych formuł i ankietowe, będące odzwierciedleniem optymizmu konsumenckiego. Piąta grupa to dane demograficzne (39) zebrane głównie w oparciu o dane publikowane przez GUS. Szósta grupa zawierała wskaźniki rynku kapitałowego i surowcowego.



Narzędzia określenia zależności korelacyjnej

Najprostszym sposobem określenia związku pomiędzy badanymi zmiennymi jest zestawienie wyników badań w tablicy korelacyjnej i zobrazowanie analizowanych zmiennych na diagramie korelacyjnym. Sposób ten jest bardzo efektywny dla niewielu zmiennych niezależnych. W pracy model statystyczny budowano w oparciu o 205 zmiennych niezależnych i tylko o jedną zmienną zależną. Zmiennymi niezależnymi były wskaźniki społeczno-gospodarcze. Ceny transakcyjne były zmienną zależną. Kierując się wynikami badań przeprowadzonymi przez Riesa, i Trouta [1996] założono, że w modelu ostatecznym nie powinno być więcej niż siedem zmiennych niezależnych. Liczba kombinacji 7-elementowego podzbioru z 25-elementowego zbioru przekracza możliwości percepcji przeciętnego człowieka. Analiza przy pomocy tablic korelacyjnych, czy też diagramów jest w praktyce niemożliwa.

Do określenia miary siły korelacyjnej dwóch zmiennych można wykorzystać:

- ◆ współczynnik zbieżności Czuprowa,
- ◆ stosunki (wskaźniki) korelacyjne Pearsona,
- ◆ współczynnik korelacji liniowej Pearsona,
- ◆ współczynnik korelacji kolejnościowej (rang) Spearmana.

W pracy jako główne narzędzie badawcze wykorzystano współczynnik korelacji liniowej Pearsona i współczynnik korelacji kolejnościowej Spearmana. Było to możliwe, ponieważ zdecydowana większość zmiennych niezależnych zobrazowana na osi czasu miała charakter zbliżony do kształtu prostoliniowego. Do wyselekcjonowania spośród 205 zmiennych niezależnych grupy atrybutów, zwanych globalnymi, posłużono się kilkoma sposobami.

Najpierw wyselekcjonowano zmienne przy pomocy współczynnika korelacji Pearsona i Spearmana. Później wykorzystując założenia analizy czynnikowej, wyselekcjonowano grupę atrybutów globalnych. Dzięki regresji wielorakiej krokowej postępującej i wstecznej również otrzymano grupę atrybutów reprezentatywnych. Na bazie tak otrzymanych atrybutów globalnych zbudowano kilkaset modeli matematycznych, które to poddano analizie statystycznej i merytorycznej.

Analiza tak dużej liczby danych pochodzących z długiego interwału czasu jest trudna do przeprowadzenia (graficzne zestawienie macierzy korelacji to 42025 wykresów). Dlatego też dane poddano kilkietapowej redukcji na podstawie macierzy korelacji i analizy czynnikowej.

W pierwszym etapie obliczono dla wszystkich parametrów współczynniki korelacji Pearsona i Spearmana pomiędzy wskaźnikami gospodarczymi a cenami nieruchomości, przyjmując do dalszej analizy tylko te, których wartość zawierała się odpowiednio w przedziałach:

- ◆ $r > 0,7$ (pomiędzy ceną transakcyjną, a atrybutem) i $r < 0,3$ (pomiędzy atrybutami),
- ◆ $r > 0,8$ (pomiędzy ceną transakcyjną, a atrybutem) i $r < 0,3$ (pomiędzy atrybutami),



- ◆ $r > 0,9$ (pomiędzy ceną transakcyjną, a atrybutem) i $r < 0,3$ (pomiędzy atrybutami),
- ◆ $r > r$ czasu,

r – współczynnik korelacji liniowej Pearsona lub Spearmana pomiędzy wskaźnikiem gospodarczym i ceną 1 m² nieruchomości lokalowych,

r_{czasu} – współczynnik korelacji liniowej Pearsona pomiędzy czasem transakcji i ceną 1 m² nieruchomości lokalowych.

Podział współczynników korelacji według wartości na wyżej wymienione przedziały wynika z założenia, że współczynniki pomiędzy zmiennymi niezależnymi powinny być w możliwie jak najniższej korelacji i jednocześnie w możliwie jak największej korelacji ze zmienną zależną.

Należy zauważyć, że wartość współczynnika korelacji Pearsona bez szerszej analizy i ilustracji graficznej może być myląca, ponieważ opiera się na matematycznym wzorze „tolerującym” duże wartości brzegowe zmiennej:

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x}_i)(y_i - \bar{y}_i)}{\sqrt{\sum_{i=1}^n (x_i - \bar{x}_i)^2 \cdot \sum_{i=1}^n (y_i - \bar{y}_i)^2}} \quad (1)$$

r_{xy} – współczynnik korelacji zmiennych x i y ,

\bar{x}_i – średnia wartość x ,

\bar{y}_i – średnia wartość y ,

x_i – i -ta wartość x ,

y_i – i -ta wartość y .

Przy obliczaniu współczynnika korelacji należy pamiętać, że „jeżeli zależność między dwiema cechami jest nieliniowa, korelacja może być słaba lub nawet zerowa, a mimo to może istnieć między nimi systematyczny związek” [9]. Dlatego też ważne jest, aby zbadać, czy liniowy model regresji z wystarczającą wiarygodnością opisuje związek pomiędzy zmiennymi. W innym przypadku należy obliczyć współczynnik korelacji dla danej krzywej regresji opisującej rzeczywisty związek pomiędzy zmiennymi.

W pracy dla badanych wielkości obliczono współczynnik korelacji rangowej Spearmana, który pozwala na zbadanie związku dwóch zmiennych nawet wtedy, gdy rozkład badanego parametru odbiega od rozkładu normalnego, a związek pomiędzy zmiennymi nie jest liniowy.

Tabela 3 zawiera kilka przykładowych rekordów obliczenia współczynnika Pearsona i Spearmana dla zmiennych niezależnych. Na tej podstawie dokonano redukcji do 41 zmiennych niezależnych, czyli do 25% wartości początkowej.

Tabela 3. Wyciąg z tabeli korelacji.

Nr zmiennej (istotnej z pkt. widzenia stat.)	Współczynniki korelacji Pearsona i Spearmana obliczone dla każdego parametru makro z ceną 1m ² nieruchomości															
	k Pear 207(99)xKraków	k Pear 207(99)xKraków > 0,9803 (k czasu)	k Spear 207(99)xKraków	k Spear 207(99)xKraków > 0,9799 (k Spearmana czasu)	Zmienne o najmocniejszych korelacjach Pearsona (41)	Zmienne o najmocniejszych korelacjach Spearmana (41)	k Pear 207(99)xKraków	k Pear 207(99)xKraków > 0,9803 (k czasu)	k Spear 207(99)xKraków	k Spear 207(99)xKraków > 0,9799 (k Spearmana czasu)	Wartość bezwzględna k Pearsona	Wartość bezwzględna k Spearmana	Nr zm. o maks. wart. współ. korelacji Pearsona (36)	Nr zm. o min. wart. współ. korelacji Pearsona (5)	Nr zm. o maks. wart. współ. korelacji Spearmana (36)	Nr zm. o min. wart. współ. korelacji Spearmana (5)
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
2	0,984	0,984	0,971		2	2	0,984	0,984	0,971		0,984	0,971	2		2	
9	0,980		0,979		9	9	0,980		0,979		0,980	0,979	9		9	
11	0,984	0,984	0,980	0,980	11	11	0,984	0,984	0,980	0,980	0,984	0,980	11		11	
15	0,984	0,984	0,979		15	15	0,984	0,984	0,979		0,984	0,979	15		15	
17	0,984		0,980				0,984		0,980		0,984	0,980	17		17	

Źródło: opracowanie własne.

Po analizie przyczynowo-skutkowej wyselekcjonowano wskaźnik podaży pieniądza M1 i obliczono korelacje pomiędzy zmiennymi. Do dalszych badań przyjęto te zmienne, które miały współczynnik korelacji poniżej 0,4 z wskaźnikiem M1 i możliwie największy współczynnik korelacji z cenami nieruchomości. Wyniki zaprezentowano w tabeli 4.

Tabela 4. Wyciąg z tabeli korelacji pomiędzy M1 a zmiennymi niezależnymi.

Parametry o najsłabszej korelacji z M1 wybrane ze 186	Współ. kor. Paersona M1 z pozostałymi parametrami 185x99 i uszeregowanymi dla $r < 0,4$
106 Produkcja przemysłowa w cenach realnych	0,121
112 Sprzedaż detaliczna m/m	0,013
116 Bilans obrotów bieżących w mln EURO	-0,052
124 Inflacja r/r (Strefa Euro) cen produkcji sprzed. w %	-0,042
135 Inflacja m/m (USA) w %	-0,093
137 Produkcja przem. r/r (USA) w %	0,065
144 Zapasy firm r/r (USA) w %	0,037

Źródło: opracowanie własne.

Do dalszych badań przyjęto tylko te zmienne niezależne, które spełniały dwa warunki:

1. jak największy współczynnik korelacji Pearsona z ceną nieruchomości
2. jak najmniejszy współczynnik korelacji Pearsona z wskaźnikiem podaży pieniądza M1.



Podstawową trudnością analizy dużych baz danych dotyczących wskaźników gospodarczych jest interpretacja wyników, w przypadku silnej korelacji wielu parametrów. Macierz korelacji obliczona dla całego zbioru danych wykazuje silne zależności wzajemne pomiędzy wieloma zmiennymi. Istotą doboru parametrów jest wytypowanie kilku z nich, które reprezentowałyby cały zbiór danych i miały charakter strukturalny. Tak duża liczba danych praktycznie wyklucza możliwość intuicyjnego wyboru parametrów na podstawie macierzy korelacji. Do rozwiązania tak sformułowanego zadania najlepiej nadaje się analiza czynnikowa.

Analiza czynnikowa wg Zakrzewskiej [19] jest to wielozmiennowa metoda statystyczna wykorzystywana do analizy macierzy współczynników korelacji. Analizę czynnikową stosujemy wtedy, kiedy mamy do czynienia z dużymi macierzami współczynników korelacji i nie jesteśmy w stanie ich objąć i dokonać analizy. Rzeczą charakterystyczną jest to, że w analizie czynnikowej nie dokonujemy podziału na zmienne objaśniane i objaśniające.

Równanie analizy czynnikowej możemy zapisać w postaci:

$$C_{ij} = a_{j1} \cdot G_{j1} + a_{j2} \cdot G_{j2} + \dots + a_{jm} \cdot G_{jm} + d_j \cdot U_{ji} \quad (2)$$

C_{ij} – wynik standardowy nieruchomości i -tej w zakresie j -tej zmiennej,

a_{jm} – ładunki czynnikowe (wagi w zakresie czynników wspólnych),

G_{jm} – wynik standardowy nieruchomości j -tej w zakresie m -tego czynnika,

d_j – waga wyniku w zakresie czynnika swoistego,

U_{ji} – wynik standardowy nieruchomości i -tej w zakresie czynnika swoistego.

Analiza czynnikowa pozwala na wybór odpowiednich zmiennych pogrupowanych odpowiednio dla jednego, dwu, trzech lub czterech czynników. Dzięki takiej prezentacji danych można wprost z pogrupowanych zmiennych wyselekcjonować zmienne do modelu statystycznego, będące w silnej korelacji ze zmienną zależną i możliwie najmniejszą korelacją między sobą. Analiza czynnikowa daje również możliwość posługiwania się bezpośrednio syntetycznymi czynnikami.

Wyciąg z tabeli 5 pokazuje metodologię wyboru – typowania zmiennych na podstawie obliczonych współczynników korelacji i wartości czynników otrzymanych z analizy czynnikowej jednocześnie (kolumny 28-35).

Pakiet *STATISTICA* daje również możliwość doboru modelu na podstawie analizy statystycznej wykonanej dla całej bazy danych lub podzielonej odpowiednio na sześć grup. Z matematycznego punktu widzenia rozwiązanie układu równań, gdzie mamy więcej niewiadomych niż równań (przypadków) nie jest możliwe. Macierz jest wtedy źle uwarunkowana, ponieważ wyznacznik jest bliski zera. *STATISTICA* zupełnie dobrze daje sobie radę z tym problemem, ostrzegając przed wykonaniem operacji komunikatem, że macierz jest źle uwarunkowana. Zmienne można również wyselekcjonować, wykonując obliczenia dla każdej z sześciu grup. Wnioski, jakie autor uzyskał na podstawie analizy wyników, zawiera tabela 6.

Tabela 5. Analiza czynnikowa i redukcja zmiennych (wyciąg).

Nr zmiennej	Nr zm. o najsilniejszej korelacji z syntetycznymi zmiennymi										Wybór zmiennych							
	Anal.Czynn.1 czyn. 207x99 (Wart.wyj.)	Anal.Czynn.2 czyn. 207x99	Anal.Czynn.3 czyn. 207x99	Anal.Czynn.4 czyn. 207x99	Anal.Czynn.1 czyn. 207x99	Anal.Czynn.2 czyn. 207x99	Anal.Czynn.3 czyn. 207x99	Anal.Czynn.1 czyn. 207x99	Anal.Czynn.2 czyn. 207x99	Anal.Czynn.1 czyn. 207x99	Suma czynn. przyjętych dla zm.	Liczba występowania w analizie czynnikowej	Liczba występowania silnej korelacji Pear. lub Spearm.	Suma kol. 29+30	Nr zmiennej dla sumy (kol. 31) > = 3	Nr zmiennej dodanej na skutek weryfikacji ekonomicznej	Nr zmiennej usuniętej na skutek weryfikacji ekonomicznej	Nr nowej zmiennej do dalszych obliczeń
1*	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35
2	2				2			2			3	3	2	5	2			2
9					9						1	1	2	3	9			9
11	11				11			11			3	3	2	5	11			11
15											0	0	2	2		15		15
17										17	1	1	2	3	17			17

Źródło: opracowanie własne.

Tabela 6. Zestawienie wad i zalet trzech podejść wyznaczenia atrybutów globalnych.

Podejście	Równoczesne obliczenie dla 205 parametrów	Równoczesne obliczenie dla 186 parametrów (bez zmiennych typu „narastająco”)	Synteza analiz w sześciu grupach I-VI
Wady/zalety			
Zalety	Jednoczesna i całościowa redukcja zmiennych na podstawie całej bazy danych	Jednoczesna i całościowa redukcja zmiennych na podstawie bazy danych zawierającej tylko zmienne niezależne	Redukcja zmiennych niezależnych na podstawie macierzy dobrze uwarunkowanych
Wady	Macierz źle uwarunkowana (205 zmiennych dla 99 przypadków). Parametry z dopiskiem „narastająco” są zależne od przypadku wcześniejszego - autokorelacja	Macierz źle uwarunkowana (186 zmiennych dla 99 przypadków)	Wyłonienie parametrów etapowo. Możliwość wyłonienia tylko 2 czynników w analizowanej grupie

Źródło: opracowanie własne.

Wyniki obliczeń dla obliczeń wykonanych jednoetapowo i wieloetapowo były bardzo zbliżone.

Ostatnim sposobem wyboru zmiennych była regresja krokowa postępująca i wsteczna.



Tabela 7. Regresja krokowa dla modelu R 110-317-502-634.

Stat. podsumowujące	POSTĘPUJĄCA	WSTECZNA
R wielorakie	0,99	0,98
Wielorakie R2	0,98	0,97
Skorygowane R2	0,98	0,96
F(4,19) - postępująca F(3,20) - wsteczna	289,10	623,68
Prawdopodobieństwo - p	0,00	0,00
Błąd std. estymacji	133,6	180,3

Źródło: opracowanie własne.

Przyjęte oznaczenia:

110 – Sprzedaż nowych aut w tys. szt.,

317 – Indeks Divisia M1,

502 – Ludność ogółem w tys.,

634 – WIG w pkt.

Dane zawarte w tabeli 7 przedstawiają wyniki dla czterech parametrów globalnych.

Na podstawie wstępnych wyników zawartych w tabeli 7 można przyjąć, że regresja postępująca lepiej odwzorowuje model niż regresja wsteczna. W praktycznych obliczeniach częściej stosuje się regresję postępującą.

Dla atrybutów globalnych wyselekcjonowanych na podstawie wartości współczynników korelacji, analizy czynnikowej i regresji wielorakiej postępującej i wstecznej wykonano analizę statystyczną dla modeli regresji wielorakiej, modelu z autokorelacją, modelu nieparametrycznego GAM, wartości czynnikowych z analizy czynnikowej i dla modelu porównawczego zbudowanego w oparciu o czas.

Analiza statystyczna wykonywana była w czterech etapach poprzez analizę:

- ◆ reszt,
- ◆ współczynnika R^2 skorygowanego (umożliwiającego porównanie modeli o różnej liczbie zmiennych niezależnych),
- ◆ wartości parametrów, błędu standardowego wyznaczenia i poziomu istotności,
- ◆ graficzną: rozrzutu reszt, rozrzutu wartości bezwzględnych reszt, wykresu normalności reszt i wykresu rozkładu surowych reszt.

Zestawienie reszt dla wybranych modeli pozwala na porównanie wyników pomiędzy poszczególnymi podejściami.

Tabela 8. Zestawienie reszt i wybranych statystyk dla reprezentatywnych modeli (wyciąg).

Nr modelu	1	2	3	4	5
Czas w miesiącach	R 110-317-502-634	AUTO 202-209-134-410-217-326	GAM Czas	FA 103-321-501-628-634	CZAS
23	32	-141	-97	177	33
2-22	
1	70	20	-54	23	121
Śr. wart. mod. reszt w %	64%	33%	67%	97%	100%
Suma (reszt) ² /1000 w %	39%	18%	40%	89%	100%
Skorygowane R ²	98,1%	99,4%	98,5%	96,4%	95,9%
Odch.stand.mod. reszt	71	44	66	106	118
Odch.stand. reszt	118	61	119	178	189

Źródło: opracowanie własne.

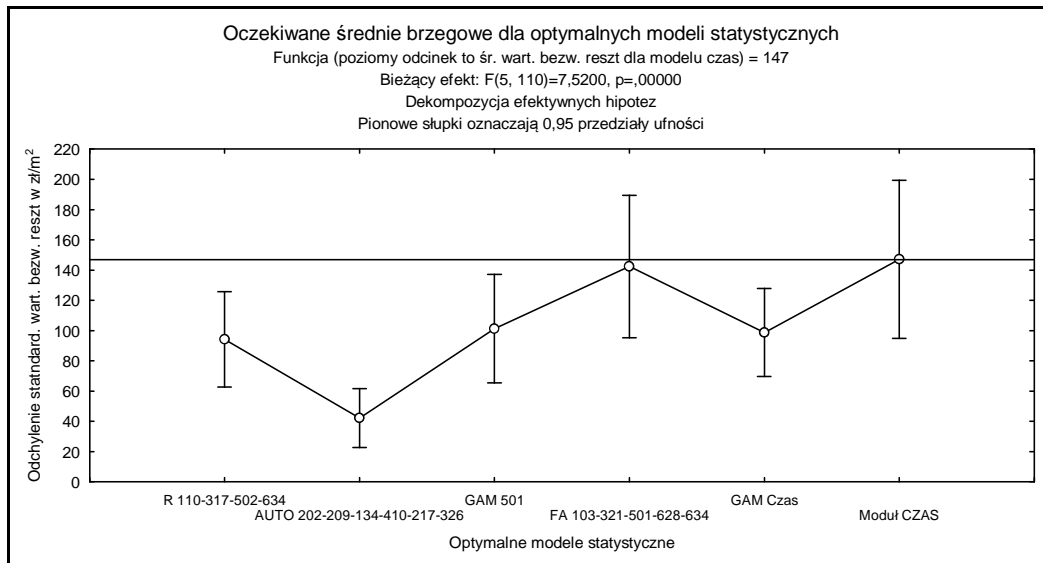
Przyjęte oznaczenia:

101 – PKB	103 – bezrobocie w %
110 – Sprzedaż nowych aut w tys. szt.	135 – Inflacja USA w %
205 – Produkcja bud.-mont. w %	317 – Indeks Divisia M1
319 – Indeks Divisia M2	321 – Indeks Divisia M3
415 – Skłonność do zakupów w pkt.	501 – Liczba emigrantów do UE w tys.
502 – Ludność ogółem w tys.	615 – OFE Aegon w zł
618 – OFE Com.U. w zł	628 – Średnia OFE w zł
634 – WIG w pkt.	

Ilustracja poniżej (rys. 2) pozwala na dokonanie syntetycznej oceny wszystkich modeli jednocześnie. Syntetyczne zestawienie wyników badań na rys. 2 potwierdza zasadność stosowania modeli zbudowanych w oparciu o atrybuty globalne.

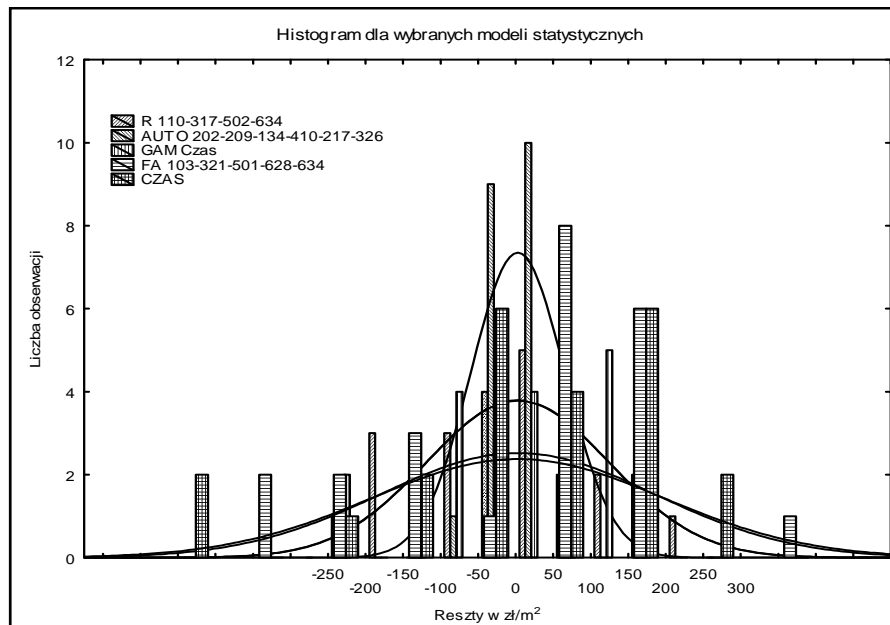
Analiza statystyczna przeprowadzona w kilku etapach potwierdziła wcześniejsze założenia i na jej podstawie można w analizie rynku nieruchomości i szacowaniu wartości rynkowej posługiwać się modelami statystycznymi zbudowanymi w oparciu o atrybuty globalne.

Dotychczas stosowany w praktyce rzeczoznawców model ma jedną zasadniczą wadę: zmienną niezależną jest czas, który de facto pełni rolę skali. Tak naprawdę można by podważać pojęcie zmiennej niezależnej. Zmienna niezależna jest zmienną przyjmującą wartości losowe. Czas przyjmuje zawsze ściśle określone wartości i jest całkowicie przewidywalny. Słabością modelu zbudowanego w oparciu o czas jest brak odporności na zmiany trendu. Wszystkie zaproponowane modele zaprezentowane na rys. 2 mają wysoką wiarygodność statystyczną. Odchylenia standardowe wartości bezwzględnych reszt dla modelu z autokorelacją i regresji liniowej są znacząco niższe niż w modelu porównawczym. Samo zastosowanie modelu GAM dla czasu poprawiło predykcję o około 30%. Zastosowanie modelu opartego na analizie czynnikowej nieznacznie poprawia wartość predykcji, ale za jego stosowaniem zdecydowanie przemawiają argumenty fundamentalne.



Rys. 2. Wykres średnich bezwzględnych wartości dla atrybutów 317,618,634 i modeli 1-5 (*pozioma linia – średni poziom wartości bezwzględnych dla modelu porównawczego z czasem);
 źródło: opracowanie własne.

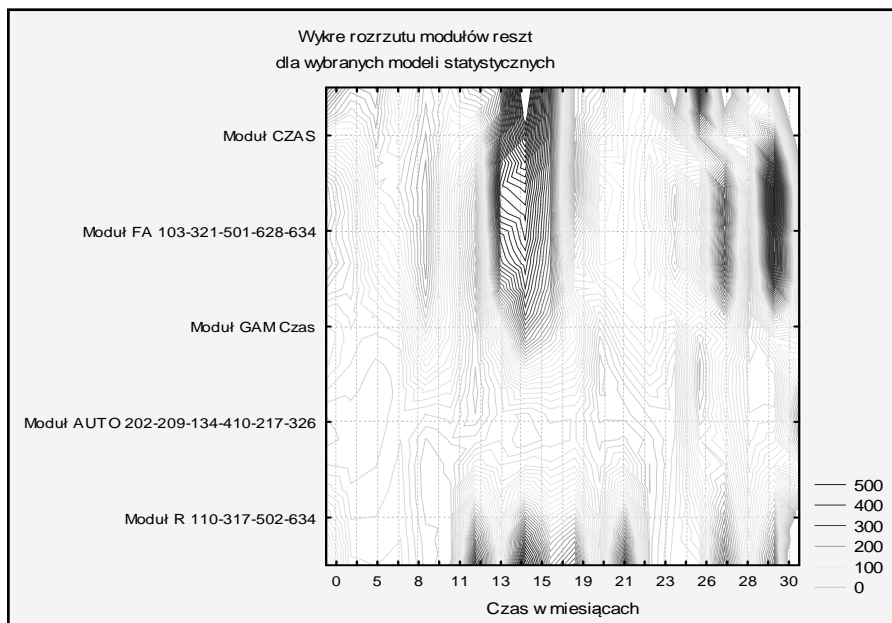
Przy wyborze atrybutów globalnych posługiwano się pomocniczo regresją krokową postępującą (gdy dodawano kolejne zmienne) i wsteczną (gdy odejmowano od modelu zdefiniowanego zmienne). Przy wystarczającej liczbie przypadków proces wyboru atrybutów globalnych można ograniczyć do zastosowania regresji krokowej.



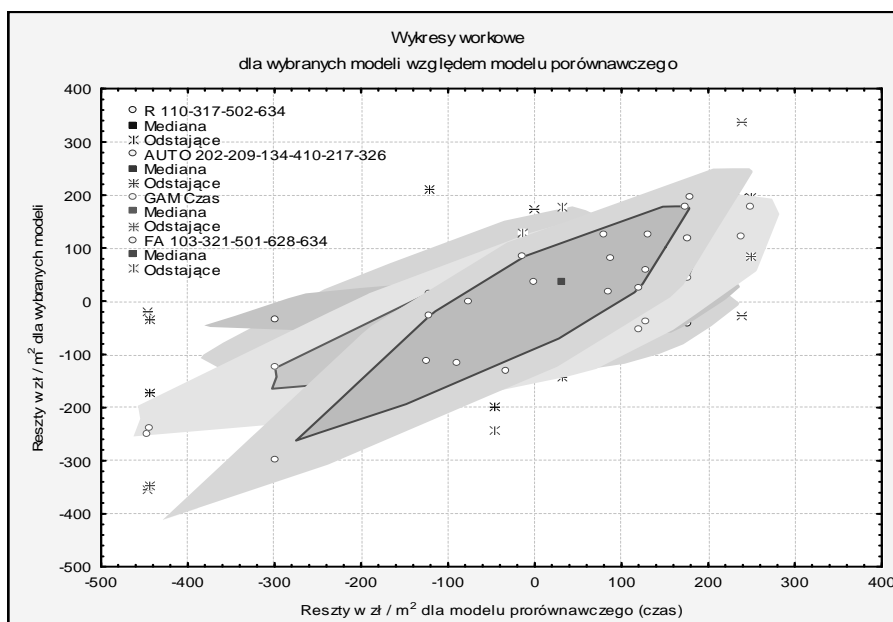
Rys. 3. Histogram dla wybranych modeli (źródło: opracowanie własne).

Rys. 3 pozwala na podstawie krzywej dzwonowej na uszeregowanie modeli od najlepszego do najgorszego. Ilustracja reszt w formie izolinii na rys. 4 pozwala na łatwe określenie mocnych i słabych odcinków (przestrzeni) zastosowanego modelu. Wykres izolinii

przestrzennych rozpatrywany w pionie pozwala na porównanie pomiędzy modelami poszczególnych stref przypisanych dla poszczególnych zmiennych zależnych.



Rys. 4. Wykres rozrzutu modułów reszt dla wybranych modeli statystycznych (źródło: opracowanie własne).



Rys. 6. Wykres workowy (źródło: opracowanie własne).

Na podstawie powyższych ilustracji model z autokorelacją i model regresji liniowej wydają się najbardziej optymalne dla celów predykcyjnych.

Zestawienie wykresów workowych na rys. 5 daje możliwość wizualnego uchwycenia korelacji pomiędzy poszczególnymi modelami. Jeżeli dla każdego modelu wyznaczymy



dłuższą oś symetrii, to kąt nachylenia pomiędzy tymi osiami będzie przedstawiał zależność pomiędzy badanymi modelami a modelem porównawczym zbudowanym w oparciu o czas. Najmniejszy związek – najmniejszy kąt pomiędzy dłuższą osią a osią X – występuje dla modelu z autokorelacją, najsilniejszy zaś dla modelu zbudowanego w oparciu o syntetyczne czynniki z analizy czynnikowej.

Podsumowując, należy stwierdzić, że jeżeli analizę rynku i szacowanie rynkowej wartości nieruchomości wykonujemy wewnątrz obszaru zawierającego zmienne zależne, to powinniśmy stosować model z autokorelacją lub modele GAM. Dla potrzeb predykcji wartości rynkowej bardziej zasadne wydaje się stosowanie modelu regresji wielorakiej opartego o atrybuty globalne lub o zmienne syntetyczne z analizy czynnikowej

Wnioski dotyczące specjalnych algorytmów do analizy rynku nieruchomości w aspekcie atrybutów globalnych

Studia literaturowe i przeprowadzone badania dotyczące estymacji modeli zbudowanych w oparciu o atrybuty globalne pozwalają na sformułowanie następujących wniosków:

1. Do fundamentalnych przyczyn wpływających na konieczność uwzględnienia większej liczby zmiennych w analizie rynku nieruchomości należy zaliczyć: wzrost swobody przepływu kapitału, wzrost popytu inwestycyjnego i zwiększenie związku pomiędzy rynkiem lokalnym i globalnym.
2. Dotychczas szeroko stosowane statyczne analizy rynku nieruchomości, bazujące głównie na danych historycznych, będą rozwijały się w kierunku analiz dynamicznych.
3. Model optymalny to taki, który zawiera atrybuty społeczno-gospodarcze mające decydujący wpływ na ceny nieruchomości na atrakcyjnych rynkach nieruchomości.
4. Potwierdzeniem słuszności zastosowania modelu zbudowanego na bazie atrybutów globalnych jest: uzyskanie mniejszych wartości reszt i porównywalnych statystyk podstawowych z modelem dotychczas stosowanym.
5. Atrybuty globalne możemy podzielić na trzy kategorie pod względem wpływu na model statystyczny: parametry makroekonomiczne i inne wskaźniki społeczno-gospodarcze, wyniki badań behawioralnych i opinie eksperckie.
6. Dane pozyskiwane do bazy powinny być w możliwie jak najmniejszym stopniu przetworzone. Zastrzeżenie to nie dotyczy wskaźników syntetycznych i wartości czynnikowych pozyskiwanych z analizy czynnikowej.
7. Optymalny sposób weryfikacji modeli statystycznych zbudowanych w oparciu o atrybuty globalne powinien być przeprowadzony w następujących etapach polegających na:
 - ◆ sprawdzeniu wartości reszt i porównaniu ich z resztami dla modelu klasycznego,
 - ◆ weryfikacji R^2 skorygowanego,
 - ◆ weryfikacji istotności statystycznej parametrów regresji,
 - ◆ graficznym badaniu przebiegu wartości modelowej.



Bibliografia

1. Bajerowski T. (red). 2003. *Podstawy teoretyczne gospodarki przestrzennej i zarządzania przestrzenią*. Wydawnictwo UWM, Olsztyn. str. 73-80.
2. Banerjee A. 1992. *A Simple model of herd behavior*. Quarterly Journal of Economics, Vol. 107, No. 3. 1992.
3. Baum A. 2001. *Evidence of Cycles In European Commercial Real Estate Markets – and Some Hypotheses*, [w:] Brown S., Liu C., *A Global Perspective on Real Estate Cycles*. Kluwer. Massachusetts.
4. Begg D., Fisher S., Dornbush R. 2007. *Makroekonomia*. Polskie Wydawnictwo Ekonomiczne, Warszawa. str. 19.
5. Burns A.F., Mitchell W.C. 1946. *Measuring Business Cycles*. Studies In Business Cycles nr 2. NBER. New York.
6. Crone T., Voith R.P. 1992. *Estimating house price appreciation: a comparison of methods*. Journal of Housing Economics. nr 2, str. 324-338.
7. Czaja J., Parzych P. 2007. *Szacowanie rynkowej wartości nieruchomości w aspekcie Międzynarodowych Standardów wyceny*, Stowarzyszenie Naukowe im. Stanisława Staszica, Kraków.
8. Dąbrowski J. 2009. *Zastosowanie metod i algorytmów statystycznych do wyznaczenia parametrów globalnych dla potrzeb badania rynku i wyceny nieruchomości*. Studia i Materiały TNN. Olsztyn. str. 31-46.
9. Ferguson G.A., Takane Y. 1997. *Analiza statystyczne w psychologii i pedagogice*. Wydawnictwo Naukowe PWN. str. 149.
10. Gdakowicz A. 2008. *Badania koniunkturalne na rynku mieszkań w Szczecinie*. Studia i Materiały TNN, vol 16 No 2, Olsztyn.
11. Gościński J.W. 1989. *Cykl życia organizacji*. PWE. Warszawa. str. 78.
12. Iwonicz-Drozdowska M. 2004. *Znaczenie rynku nieruchomości dla bezpieczeństwa banków (w:) Ryzyko kredytowe wiarygodności hipotecznych modelowanie i zarządzanie*, red. Jajuga K. 2004. ZBP. Warszawa.
13. Lukas R. 1995. *Understanding Business Cycles*, [w:] *Business Cycle Theory*, red. Kydland F.E.
14. Nykiel L. 2008. *Kierunki rozwoju i przekształceń rynku mieszkaniowego w Polsce*. Studia i Materiały TNN, vol 16 No 2, Olsztyn.
15. Saunders A. 2001. *Metody pomiaru ryzyka kredytowego*. Oficyna Ekonomiczna. Kraków.
16. Sobczyk M. 1997. *Statystyka*, Wydawnictwa Naukowe PWN. Warszawa 1997.
17. Trojanek R. 2008. *Determinanty wahań cen na rynku mieszkaniowym*. Studia i Materiały TNN, Olsztyn.
18. Wheaton W.C. 1987. *The Cyclic Behavior of the National Office Market*. Journal of the American Real Estate and Urban Economics Association, 15:4.



19. Zakrzewska M. 1993. *Słownik terminów analizy czynnikowej*.
20. Żelazowski K. 2008. *Zjawisko baniek cenowych na rynkach nieruchomości*, Studia i Materiały TNN, Olsztyn.
21. WWW.redNet.
22. WWW.NBP.