

KRYTERIA STOSOWANIA MODELI STOCHASTYCZNYCH W PREDYKCJI RYNKOWEJ WARTOŚCI NIERUCHOMOŚCI

Anna Barańska

Katedra Informatyki o Terenie, Akademia Górniczo-Hutnicza w Krakowie

1 SFORMUŁOWANIE TEZ PRACY

Do najczęściej stosowanych metod określania rynkowej wartości nieruchomości należy niewątpliwie wycena realizowana przy pomocy podejścia porównawczego. Podstawową zasadą tego podejścia jest porównanie nieruchomości wycenianej z nieruchomościami podobnymi, o znanych cenach transakcyjnych oraz cechach różnicujących te nieruchomości i wpływających na ich wartość. Wszystkie informacje o nieruchomościach mają charakter probabilistyczny, gdyż ich wybór wraz z ich opisem zależy od rzeczoznawcy. Stosowne Rozporządzenie Rady Ministrów mówi o 3 metodach w podejściu porównawczym: metodzie porównywania parametrów, metodzie korygowania ceny średniej i metodzie analizy statystycznej rynku. Trzecia metoda dopuszcza wszelkie sposoby wyceny bazujące na modelach statystycznych. Uzyskana w ten sposób wartość rynkowa nieruchomości powinna odpowiadać prawdopodobnej cenie, jaką uzyskałaby wyceniana nieruchomość na wolnym rynku, jako przedmiot ustalonego prawa rzeczowego.

Na bazie obowiązujących przepisów prawnych, dotyczących zagadnienia wyceny nieruchomości – istnieje bardzo wiele koncepcji zastosowania metod statystycznych i tworzenia wielowymiarowych modeli matematycznych, możliwie najlepiej opisujących dany rynek nieruchomości. Wielowymiarowość zagadnienia wynika przede wszystkim z mnogości czynników wpływających bezpośrednio na wartość nieruchomości. Tworzenie modelu matematycznego dla wybranego rynku jest zagadnieniem niezwykle złożonym, gdyż wymaga stosownego

przygotowania bazy danych do modelowania pod względem jej kompletności i wiarygodności, z uwzględnieniem optymalnego doboru zmiennych. Sposób znalezienia tego „najlepszego” modelu związany jest z testowaniem dużej ilości informacji w bazie danych. Istnieje zatem potrzeba sformułowania jakichś kryteriów doboru optymalnego modelu spośród wielu, jak również kryteriów związanych z samą postacią bazy danych do modelowania. Dotyczy to głównie relacji pomiędzy liczbą zmiennych i liczebnością bazy danych.

Zasadniczym celem rozprawy jest opracowanie kryteriów doboru modeli matematycznych i oceny wiarygodności danych służących do modelowania. Do sformułowania tych kryteriów zostaną wykorzystane parametry niezmiennicze, które będą zdefiniowane w oparciu o macierz korelacyjną dla zmiennych występujących w modelu wyceny i macierz kowariancji dla prognozowanych rynkowych wartości nieruchomości.

Zakres pracy obejmuje:

- charakterystykę wykorzystywanej bazy wiedzy o nieruchomościach sprzedanych z różnych rynków Polski południowo-wschodniej,
- zdefiniowanie parametrów niezmienniczych, służących sformułowaniu kryteriów doboru modelu wyceny i postaci bazy,
- określenie trendu zmiany cen nieruchomości w czasie dla poszczególnych rynków lokalnych,
- wstępną analizę baz nieruchomości pod kątem wyboru atrybutów do modelowania rynkowej wartości nieruchomości,

- testowanie wielowymiarowych modeli liniowych i nieliniowych w poszczególnych bazach,
- wstępną selekcję modeli na podstawie testu *Fishera-Snedecora* o równości wariancji części wyjaśnionej i niewyjaśnionej przez model regresji,
- wyznaczenie wartości zdefiniowanych niezmienników,
- sformułowanie propozycji kryteriów doboru modelu i oceny wiarygodności bazy nieruchomości na podstawie analiz wartości parametrów niezmienniczych.

Teza rozprawy doktorskiej.

Na podstawie, zdefiniowanych przez autorkę pracy, parametrów niezmienniczych istnieje możliwość doboru optymalnej liczby nieruchomości i optymalnej liczby opisujących je cech w modelowaniu rynkowej wartości nieruchomości. Kryteria doboru optymalnego modelu wyceny mogą być sformułowane na podstawie przedziałów wartości zaproponowanych parametrów niezmienniczych.

2 OPIS WYKORZYSTYWANEJ BAZY WIEDZY

Wykorzystywane w pracy bazy danych o nieruchomościach odnoszą się do rynków nieruchomości południowo-wschodniej Polski, monitorowanych od 1998 roku do chwili obecnej.

Dane wykorzystane w niniejszej pracy obejmują informacje o nieruchomościach gruntowych, przeznaczonych pod budownictwo, będących przedmiotem obrotu, z następujących miejscowości: Bolesław, Busko Zdrój, Kraków, Nowy Sącz, Proszowice, Przeworsk, Rzeszów, Świdnik, Trzyciąż.

Rozważane rynki nieruchomości charakteryzują się dużym zróżnicowaniem cenowym i wykazują zmienną dynamikę transakcji. Może to być spowodowane zmienną koniunkturą gospodarczą kraju. Zebrane bazy zawierają od 20 do 130 nieruchomości i dają łącznie 530 nieruchomości gruntowych.

Rynek nieruchomości tworzony jest dla każdego miasta (gminy) oddzielnie. A dla dużych miast poszczególne dzielnice stanowiły oddzielne rynki nieruchomości. Jest to związane z trudnością znalezienia łatwo identyfikowalnych i mierzalnych cech (zmiennych), w oparciu

o które możliwe byłoby przetransformowanie cen do wspólnego, jednego rynku nieruchomości. W większości przypadków, nawet dla gmin o zbliżonej liczbie mieszkańców, ceny nieruchomości najczęściej zdecydowanie się różnią. Decydują o tym głównie takie czynniki jak: rynek pracy, atrakcyjność miasta, czystość środowiska, perspektywy rozwoju, położenie, krajobraz. Zachodzi też odmienne reagowanie na działanie czynników lokalnych, uwidaczniające się w charakterystykach weryfikowanych modeli.

2.1 Charakterystyka przyjętych atrybutów i ich skal.

W przypadku rynku nieruchomości trudno jest zidentyfikować cechy (zmienne), których wpływ na wartość nieruchomości jest jednoznacznie określony.

W niniejszej pracy analizowano atrybuty (wyróżnione cechy), które są łatwe do identyfikacji i rejestracji. Atrybuty opisujące nieruchomości można podzielić na dwie zasadnicze grupy: obligatoryjne i fakultatywne. Pierwsza z nich służy jednoznacznej identyfikacji nieruchomości; są to dane o charakterze administracyjno-prawnym. Należą tu wszystkie dane ewidencyjne i związane z adresem nieruchomości oraz dane opisujące jej stan prawny. Natomiast druga grupa to cechy kreujące wartość rynkową, względem których powinny być wybierane nieruchomości do porównywania. Atrybuty fakultatywne mają zasadnicze znaczenie w modelowaniu matematycznym wartości rynkowej, zatem konieczne jest ustalenie dla nich skal wartości. Dla nieruchomości gruntowych można wyszczególnić następujące atrybuty:

- przeznaczenie w planie miejscowym,
- dane geometryczne (wymiary, kształt, topografia),
- dane opisujące atrakcyjność lokalizacji (strefa, moda, otoczenie),
- dane związane z dostępem do nieruchomości (dostępność środków komunikacji, droga dojazdowa),
- występowanie elementów dodatkowych (uzbrojenie terenu, ewentualna zabudowa).

Przy ustalaniu skal liczbowych dla większości atrybutów posłużono się skalą porządkową, pozwalającą na rangowanie elementów,

z uwzględnieniem natężenia wpływu atrybutu na wartość rynkową. Najczęściej korzystano ze skali 4-stopniowej. W kilku przypadkach wykorzystano również skalę nominalną lub ilorazową.

W pracy rozpatrywano kilkanaście cech nieruchomości o charakterze uznaniowym.

3 PARAMETRY NIEZMIENNICZE

Do analizy informacji rynkowych zaproponowano trzy następujące parametry:

- 1 Współczynnik determinacji R^2 , czyli standardowa miara zaufania do estymowanego modelu wyceny, zdefiniowany wzorem:

$$R^2 = 1 - \frac{\det K}{\det K_0}, \quad (1)$$

gdzie:

$$K = \begin{Bmatrix} 1 & r_{01} & r_{02} & \dots & r_{0k} \\ r_{10} & 1 & r_{12} & \dots & r_{1k} \\ r_{20} & r_{21} & 1 & \dots & r_{2k} \\ \dots & \dots & \dots & \dots & \dots \\ r_{k0} & r_{k1} & r_{k2} & \dots & 1 \end{Bmatrix} \quad (2)$$

K – macierz korelacyjna, zawierająca współczynniki korelacji zupełnej pomiędzy wszystkimi parami zmiennych (r_{ij}),

$\det K$ – wyznacznik macierzy korelacyjnej,

$\det K_0$ – wyznacznik podmacierzy, która powstaje ze skreślenia pierwszego wiersza i pierwszej kolumny w macierzy K , czyli współczynników korelacji dotyczących zmiennej zależnej (ceny).

Wielkość $1-R^2$ stanowi współczynnik niezgodności modelu z rynkowymi wartościami nieruchomości, które są wykorzystywane w estymacji modelu wyceny.

- 2 Parametr określony na podstawie śladu macierzy kowariancji, zdefiniowany wzorem:

$$\sigma_{tr} = \frac{1}{W_{sr}} \sqrt{\frac{tr\{Cov(W)\}}{n}}, \quad (3)$$

gdzie:

$tr\{Cov(W)\}$ – ślad macierzy kowariancji dla prognozowanych wartości rynkowych nieruchomości ustalających model wyceny,

W_{sr} – średnia wartość z prognozowanych wartości rynkowych nieruchomości ustalających model wyceny,

n – liczba nieruchomości służących do estymacji modelu.

- 3 Parametr określony na podstawie wyznacznika macierzy kowariancji, zdefiniowany wzorem:

$$\sigma_{det} = \frac{1}{W_{sr}^u} \sqrt[u]{\det\{Cov(W)\}}, \quad (4)$$

gdzie:

$\det\{Cov(W)\}$ – wyznacznik macierzy kowariancji dla prognozowanych wartości rynkowych nieruchomości ustalających model wyceny,

W_{sr} – średnia wartość z prognozowanych wartości rynkowych nieruchomości ustalających model wyceny,

u – liczba zmiennych niezależnych występujących w modelu.

Ze względu na charakter pierwszego parametru, jako miary zaufania do modelu, wskazane jest, aby jego wartość spełniała kryterium: $R^2 > 0,60$. Natomiast dla dwóch pozostałych parametrów, powiązanych ściśle z wariacjami wartości modelowych cen nieruchomości tworzących model, przyjmujemy, że nie powinny one przekraczać kilkunastu procent średniej z prognozowanych wartości rynkowych nieruchomości tworzących model wyceny. Te wstępne postulaty zostały zweryfikowane w dalszych rozdziałach pracy.

Na podstawie analizy wzorów (3) i (4) można stwierdzić, że wielkości te stanowią pewnego rodzaju miary rozproszenia wokół średniej wartości modelowej, stąd mogą być obiektywnymi parametrami służącymi do sformułowania kryteriów wiarygodności baz wykorzystywanych w modelowaniu wartości rynkowej.

4 ESTYMACJA TRENDU ZMIANY CEN NIERUCHOMOŚCI W CZASIE

Rozważane w pracy bazy nieruchomości odnoszą się do transakcji realizowanych w różnym czasie, czyli przy różnym stanie rynku. Stan rynku, wyrażający poziom cen nieruchomości, wiąże się z ogólną koniunkturą gospodarki kraju

i może być również kwantyfikowany za pomocą skali uznaniowej. Obiektywną skalą rejestrującą stan rynku nieruchomości jest czas, który powinien być liczony w miesiącach.

W przedstawionych rozważaniach zakłada się, że cechy poszczególnych nieruchomości zachowują niezmienną w rozpatrywanym przedziale czasu zawieranych transakcji. Zatem stan rynku rejestrowany różnym poziomem cen nieruchomości powinien być skorygowany do jednego momentu czasowego (ustalonego miesiąca).

W pracy dla każdej z baz, w której było to niezbędne, taka korekta została wykonana. Dobór funkcji do określenia trendu zmiany cen w czasie został dokonany na podstawie wartości współczynnika korelacji krzywoliniowej. Okazało się, że najlepszymi funkcjami są tu funkcje wielomianowe różnych stopni.

5 STATYSTYCZNA ANALIZA BAZ NIERUCHOMOŚCI

W celu efektywnego wykorzystywania bazy danych w badaniach naukowych konieczne jest przeprowadzenie jej wstępnej obróbki statystycznej. Poprzez taką obróbkę doprowadza się bazy danych do spójności, co pozwala na uzyskanie bardziej wiarygodnych wyników prowadzonych badań.

W pierwszej kolejności należy przeanalizować zbiór informacji pod względem jego kompletności, tzn. w miarę możliwości każdy z przypadków powinien mieć nadaną konkretną wartość, dla każdej z rozważanych zmiennych.

Następnie, w poszczególnych bazach danych, dla zmiennej zależnej wyznaczono współczynnik dyspersji. Parametr ten jest miarą rozproszenia wartości zmiennej wokół jej wartości przeciętnej.

Bardziej szczegółowe analizy opierały się na wyznaczeniu współczynników korelacji zupełnej, cząstkowej oraz semicząstkowej, a także na wynikach analizy reszt. Współczynniki korelacji pozwalają określić wpływ poszczególnych zmiennych niezależnych na zmienną zależną oraz służą do badania nadmiarowości występowania zmiennych niezależnych. Natomiast analiza reszt pozwala wyeliminować z bazy danych przypadki odstające. Pojawienie się takich przypadków w bazie może być wynikiem bądź błędnej informacji, bądź zaistnienia warunków

szczególnych (np. sprzedaż wymuszona w obrocie nieruchomościami), albo też wpływu innych czynników – trudnych do zidentyfikowania w procesie zbierania informacji o nieruchomościach i tworzenia modelu wyceny.

Owocna dla późniejszego modelowania jest również wstępna ocena typu zależności pomiędzy zmienną zależną a poszczególnymi zmiennymi niezależnymi na podstawie wykresów rozrzutu oraz analiza częstości występowania poszczególnych wartości skal przyjętych dla zmiennych niezależnych.

Na podstawie powyższych wyników można wysnuć globalne wnioski na temat cech nieruchomości mających najczęściej największy wpływ na jej cenę. Dla wybranych dziesięciu baz nieruchomości atrybutami, które najczęściej występują w grupie tych najistotniejszych na danym rynku są: *strefa miasta, dostęp komunikacyjny, moda na lokalizację, wpływ otoczenia*, a z elementów uzbrojenia terenu: *sieć gazowa* i w drugiej kolejności *kanalizacja sieciowa*.

6 WIELOWYMIAROWE LINIOWE I NIELINIOWE MODELOWANIE WARTOŚCI NIERUCHOMOŚCI

W procesie modelowania rynku nieruchomości w każdej bazie przetestowano wielowymiarowy model liniowy w postaci liniowej regresji wielorakiej:

$$F(X_i, a) = w_i = a_0 + \sum_{k=1}^u X_{ik} * a_k \quad (5)$$

gdzie:

w_i – wartość modelowa i -tej nieruchomości w danej bazie,

X_i – wektor wartości atrybutów dla i -tej nieruchomości ($1 \times u$),

X_{ik} – wartość atrybutu k dla i -tej nieruchomości,

a_0 – wyraz wolny w modelu,

a – wektor współczynników regresji wielokrotnej ($(u+1) \times 1$),

a_k – współczynnik regresji stojący przy atrybucie k .

W następnej kolejności dla każdej bazy dobierano różne postacie modeli nieliniowych. W statystyce matematycznej nie ma analitycznych sposobów, które by umożliwiały

optymalny wybór odpowiedniej postaci funkcji. Przy konstruowaniu takiego modelu przebadano najpierw charakter zależności ceny nieruchomości od każdego atrybutu z osobna:

$$Y = g(X_k), \quad (6)$$

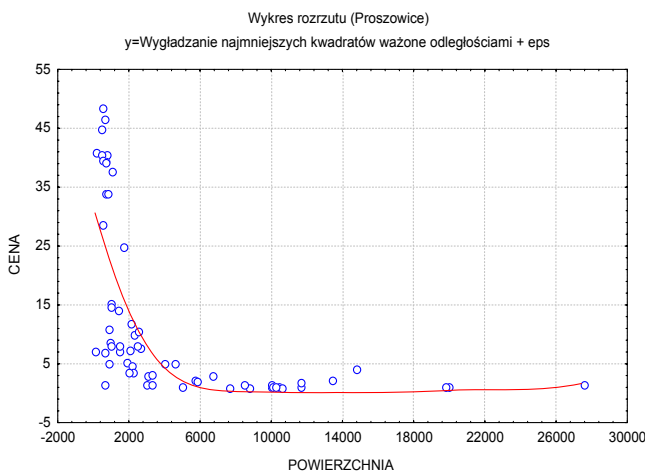
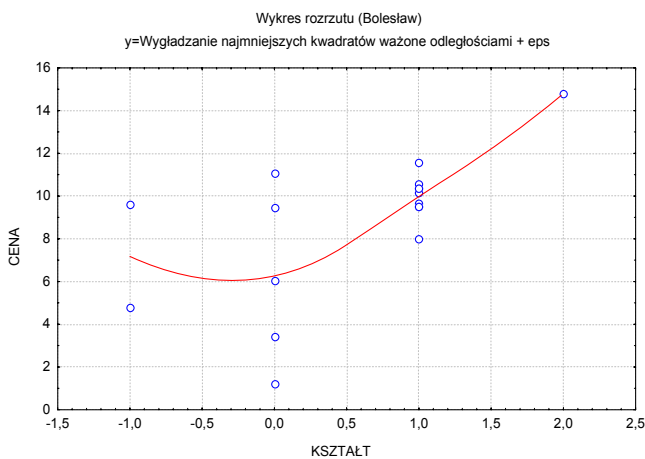
gdzie:

X_k – zmienna odpowiadająca wartościom atrybutu k ,

Y – ceny nieruchomości w bazie.

Postać funkcji g może być bardzo różna. Rozważano następujące funkcje elementarne: liniowe, wielomiany różnych stopni, funkcje logarytmiczne, wykładnicze, niewymierne.

Wykonano wykresy rozrzutu *atrybut – cena* z nałożoną linią trendu wyestymowaną metodą MNK, by w sposób optymalny dobrać typ funkcji g dla każdego atrybutu w obrębie każdej bazy nieruchomości. Poniżej zamieszczono przykładowe wykresy:



Z tak wybranych postaci funkcji g tworzone były modele wielowymiarowe, czyli globalne

funkcje nieliniowe F dla poszczególnych baz nieruchomości:

$$W = F(X, a) \quad (7)$$

gdzie:

W – zbiór prognozowanych cen nieruchomości tworzących model wyceny,

X – zmienna wielowymiarowa odpowiadająca atrybutom nieruchomości,

a – wektor parametrów modelu.

Poszczególne funkcje g dla różnych atrybutów mogą być – w obrębie funkcji F – powiązane ze sobą zależnością addytywną lub multiplikatywną. W niniejszej pracy zastosowano, w większości przypadków, addytywną postać funkcji F .

Wszystkie rozpatrywane modele regresji oszacowano metodą najmniejszych kwadratów (MNK), ze względu na jej ogólnie znane zalety. Każdorazowo estymacja parametrów modelu została wykonana wraz z analizą dokładności, tj. z podaniem błędu oszacowania parametru i jego poziomem istotności.

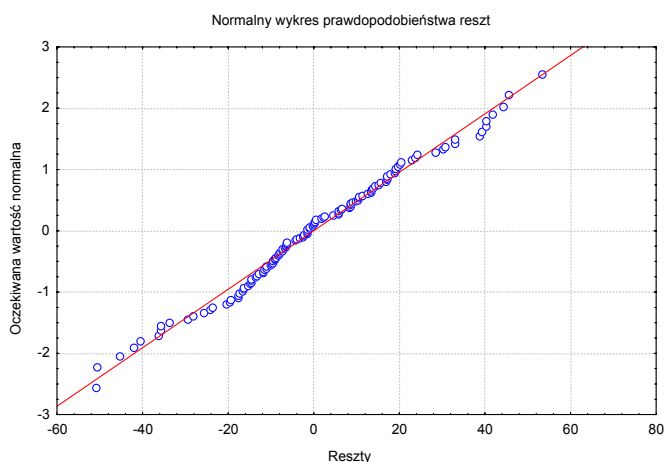
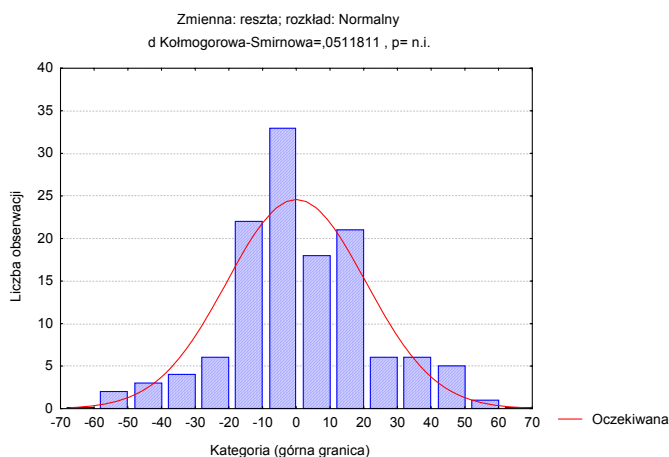
Przy okazji estymacji modelu zostały wyznaczone również inne wielkości, m.in. tzw. *wagi beta* (pozwalają one porównać relatywny wkład, jaki każda ze zmiennych niezależnych wnosi w predykcję zmiennej zależnej) oraz tzw. *tolerancja* (pozwala orzec o ewentualnej nadmiarowości wkładu danej zmiennej w równanie regresji). Jeżeli tolerancja jakiejś zmiennej w równaniu jest równa zero (lub bardzo bliska tej wartości), to nie można wyznaczyć parametrów tego równania ze względu na złe uwarunkowanie macierzy, a co za tym idzie, niemożność jej odwrócenia. W niniejszej pracy wartość progową dla tolerancji, do akceptacji bądź odrzucenia modelu, ustalono na poziomie 0,01.

Wykonano także badanie normalności rozkładu reszt przy pomocy testu zgodności *Kolmogorowa-Smirnowa* oraz analizę i wyodrębnienie reszt odstających w każdym analizowanym modelu.

Wyniki testu każdorazowo zobrazowano za pomocą wykresu rozkładu reszt w postaci histogramu, z nałożoną na niego linią przedstawiającą rozkład normalny i z wartością statystyki testu *Kolmogorowa-Smirnowa* oraz za pomocą tzw. *normalnego wykresu prawdopodobieństwa reszt*. Jeśli punkty przedstawiające reszty i ich oczekiwaną wartość „normalną” układają się, z dobrym przybliżeniem, wzdłuż linii prostej, to

dowodzi to wysokiej zgodności z rozkładem teoretycznym.

Zarówno w przypadku modeli liniowych, jak i nieliniowych pierwsza selekcja, pod kątem ich przydatności do prognozowania wartości rynkowej nieruchomości, została wykonana na podstawie wartości współczynnika korelacji wielorakiej (dla liniowych) – wzór (1) i współczynnika korelacji krzywoliniowej (dla nieliniowych) – wzór (8).



$$\rho^2 = 1 - \frac{\sum_{i=1}^n [y_i - F(X_i, a)]^2}{\sum_{i=1}^n [y_i - E(y)]^2} = \frac{\sum_{i=1}^n [F(X_i, a) - E(y)]^2}{\sum_{i=1}^n [y_i - E(y)]^2}$$

$$\Leftrightarrow \rho^2 = 1 - \frac{NSK}{CSK} = \frac{WSK}{CSK} \quad (8)$$

gdzie:

$E(y)$ – wartość oczekiwana zaobserwowanych wartości zmiennej zależnej (ceny),

$\sum_{i=1}^n [y_i - E(y)]^2$ – całkowite rozproszenie zmiennej zależnej względem jej wartości przeciętnej (*CSK*),

$\sum_{i=1}^n [y_i - F(X_i, a)]^2 = \sum_{i=1}^n \delta^2$ – suma kwadratów odchyłek pomiędzy wartościami zmiennej zależnej z próby a jej wartościami modelowymi, wyrażająca część niewyjaśnioną modelem regresji nieliniowej (*NSK*),

$\sum_{i=1}^n [F(X_i, a) - E(y)]^2$ – suma kwadratów odchyłek pomiędzy wartościami modelowymi zmiennej zależnej a wartością przeciętną z zaobserwowanych wartości tej zmiennej, wyrażająca część wyjaśnioną modelem regresji nieliniowej (*WSK*).

Oba współczynniki korelacji wskazują, w jakim zakresie wyestymowany model wyjaśnia rzeczywistą zmienność. Z tego względu przyjęto wstępne kryterium, by ich wartość przekraczała 0,60. Dalszą weryfikację wiarygodności modeli stanowił wynik testu *Fishera-Snedecora*, badający prawdziwość hipotezy o równości wariancji części wyjaśnionej i niewyjaśnionej przez model regresji, dla którego statystyka testowa ma postać:

$$F = \frac{R^2}{1 - R^2} * \frac{n - m}{m - 1} \quad (9)$$

gdzie:

R – współczynnik korelacji (liniowej lub krzywoliniowej),
 n – liczność próby (liczba nieruchomości w bazie),
 u – liczba zmiennych niezależnych w modelu,
 m – liczba estymowanych parametrów modelu wyceny.

Powyższy test uwzględnia konieczność zachowania w modelu właściwych proporcji pomiędzy liczbą przypadków i liczbą niewiadomych, a nie tylko bada bezwzględny stosunek wariancji części wyjaśnionej i części niewyjaśnionej. W ten sposób kilka modeli zostało wyeliminowanych, nawet spośród tych o bardzo wysokiej (powyżej 0,80), bezwzględnej wartości R^2 .

7 WYZNACZENIE WARTOŚCI NIEZMIENNIKÓW

Na wszystkich bazach nieruchomości przetestowano w sumie bardzo wiele różnych modeli określenia rynkowej wartości nieruchomości. Na wstępie wyeliminowano wszystkie te, dla których współczynnik korelacji wielorakiej wynosił mniej niż 0,60. Dla każdego z pozostałych 97 modeli, po wyznaczeniu macierzy korelacji dla zmiennych występujących w modelu i macierzy kowariancji dla wartości modelowych nieruchomości ustalających model wyceny, obliczono wartości trzech zdefiniowanych niezmienników. Otrzymano w ten sposób 97 zestawów tych trzech liczb. Wyniki zawarto w poniższej tabeli 1. Oprócz wartości niezmienników

zamieszczono również: nazwę miejscowości, z której pochodzi baza danych wykorzystana do modelowania, typ modelu i wartości stałych charakteryzujących bazę i model. Są to:

- liczba nieruchomości w bazie: n ,
- liczba zmiennych niezależnych występujących w modelu: u ,
- liczba estymowanych parametrów modelu wyceny: m ,
- liczba stopni swobody: $k = n - m$,
- wariancja resztowa: σ_0^2 ,
- wariancja resztowa pomnożona przez liczbę stopni swobody: $k \cdot \sigma_0^2 = Y^T Y - a^T X^T Y$.

Dodatkowo wyróżniono 9 modeli, które zostały odrzucone przez test *Fishera-Snedecora*.

Tabela 1. Wartości niezmienników.

Nr	MIEJSCOWOŚĆ	ZMIENNOŚĆ	MODEL	N	u	m	k	σ_0^2	R^2	σ_{tr}	σ_{det}
1	Bolesław	liniowa	1	18	13	14	4	2,807	,936	,166	,05308
2	Bolesław	liniowa	2	18	7	8	10	5,206	,702	,172	,00009
3	Bolesław	liniowa	3	18	8	9	9	5,364	,723	,183	,00108
4	Bolesław	liniowa	4	18	9	10	8	5,930	,728	,203	,00466
5	Bolesław	liniowa	5	18	10	11	7	4,643	,814	,189	,01180
6	Bolesław	liniowa	6	18	11	12	6	4,762	,836	,201	,02372
7	Bolesław	liniowa	7	18	12	13	5	4,750	,864	,212	,03821
8	Bolesław	liniowa	K-P-D-M	18	4	5	13	4,286	,681	,120	,00000
9	Bolesław	nieliniowa	14	18	10	11	7	2,632	,894	,148	,00983
10	Bolesław	nieliniowa	41	18	14	12	3	,325	,994	,057	,05616
11	Bolesław	nieliniowa	42	18	14	17	3	,325	,994	,057	,05586
12	Bolesław	nieliniowa	43	18	12	15	5	,355	,990	,052	,02546
13	Bolesław	nieliniowa	44	18	12	15	5	,262	,992	,047	,02530
14	Bolesław	nieliniowa	45	18	13	16	4	,266	,994	,050	,03963
15	Bolesław	nieliniowa	46	18	11	14	6	,708	,976	,076	,01648
16	Bolesław	nieliniowa	47	18	10	13	7	,834	,966	,080	,00804
17	Bolesław	nieliniowa	48	18	5	6	12	3,365	,768	,114	,00000
18	Bolesław	nieliniowa	49	18	5	8	12	3,365	,768	,114	,00000
19	Bolesław	nieliniowa	50	18	6	7	11	3,178	,800	,123	,00000
20	Bolesław	nieliniowa	51	18	8	9	9	3,064	,842	,138	,00000
21	Bolesław	nieliniowa	52	18	10	10	7	1,654	,934	,111	,00000
22	Bolesław	nieliniowa	53	18	8	11	9	,986	,949	,077	,00054
23	Bolesław	nieliniowa	54	18	9	12	8	,784	,964	,072	,00248
24	Bolesław	nieliniowa	55	18	8	11	9	,715	,963	,065	,00049
25	Busko Zdrój	liniowa	1	31	15	16	15	37,875	,810	,090	,00297
26	Busko Zdrój	liniowa	2	31	13	14	17	38,515	,781	,084	,00098
27	Busko Zdrój	nieliniowa	2a	31	17	18	13	13,227	,942	,057	,00535
28	Busko Zdrój	nieliniowa	2b	31	17	19	13	11,151	,952	,052	,00522
29	Busko Zdrój	nieliniowa	3a	31	16	18	14	10,436	,951	,049	,00373
30	Busko Zdrój	nieliniowa	4a	31	18	20	12	9,954	,960	,051	,00672
31	Busko Zdrój	nieliniowa	4b	31	18	19	12	11,193	,955	,054	,00682
32	Busko Zdrój	nieliniowa	5a	31	16	18	14	9,116	,957	,046	,00371
33	Busko Zdrój	nieliniowa	5b	31	16	17	14	10,108	,953	,048	,00377
34	Busko Zdrój	nieliniowa	6	31	15	16	15	10,212	,949	,047	,00250
35	Krowodrza	nieliniowa	1a	38	13	14	24	2324,053	,695	,342	,00033
36	Krowodrza II	liniowa	1	131	12	13	118	779,711	,919	,040	,00000
37	Krowodrza II	liniowa	2	127	12	13	114	469,446	,949	,032	,00000
38	Krowodrza II	nieliniowa	1	131	16	17	114	804,668	,920	,047	,00000

Nr	MIEJSCOWOŚĆ	ZMIENNOŚĆ	MODEL	N	u	m	k	σ_0^2	R^2	σ_{tr}	σ_{det}
39	Krowodrza II	nieliniowa	1a	127	16	17	110	473,673	,951	,037	,00000
40	Krowodrza II	nieliniowa	1b	121	16	17	104	343,562	,962	,033	,00000
41	Nowy Sącz	liniowa	1	30	13	14	16	23,934	,811	,147	,00229
42	Nowy Sącz	liniowa	2	30	12	13	17	23,438	,803	,140	,00104
43	Nowy Sącz	liniowa	3	29	13	14	15	15,782	,879	,120	,00255
44	Nowy Sącz	liniowa	4	29	12	13	16	15,086	,877	,112	,00118
45	Nowy Sącz	liniowa	5	28	12	13	15	11,458	,898	,105	,00146
46	Nowy Sącz	liniowa	6	28	11	12	16	10,983	,895	,098	,00055
47	Nowy Sącz	nieliniowa	2	30	16	17	13	21,975	,859	,150	,00949
48	Nowy Sącz	nieliniowa	3	30	16	17	13	21,942	,859	,151	,00946
49	Proszowice	liniowa	1	63	9	10	53	27,468	,894	,191	,00014
50	Proszowice	liniowa	2	60	9	10	50	18,913	,923	,197	,00247
51	Proszowice	nieliniowa	1	57	10	12	46	8,016	,964	,112	,00000
52	Proszowice	nieliniowa	2	57	10	12	46	8,016	,964	,112	,00000
53	Proszowice	nieliniowa	4	57	9	11	47	9,610	,963	,114	,00000
54	Proszowice	nieliniowa	5	57	9	10	47	9,648	,963	,115	,00000
55	Proszowice	nieliniowa	5'	54	9	10	44	5,284	,975	,116	,00000
56	Proszowice	nieliniowa	5a	57	8	9	48	9,690	,962	,115	,00000
57	Proszowice	nieliniowa	5'a	54	8	9	45	5,296	,975	,115	,00000
58	Proszowice	nieliniowa	5b	44	8	9	35	,463	,998	,046	,00000
59	Proszowice	nieliniowa	6	57	8	10	48	9,560	,963	,121	,00000
60	Przeworsk	liniowa	1	30	15	16	14	1,006	,605	,033	,00491
61	Przeworsk	nieliniowa	1	30	18	19	11	1,158	,643	,038	,01357
62	Rzeszów	liniowa	1	48	13	14	34	33,689	,824	,060	,00004
63	Rzeszów	liniowa	2	47	13	14	33	25,821	,856	,052	,00004
64	Rzeszów	liniowa	3	46	13	14	32	21,845	,875	,048	,00049
65	Rzeszów	nieliniowa	1	48	13	14	34	34,447	,820	,061	,00004
66	Rzeszów	nieliniowa	1a	47	13	14	33	26,610	,852	,053	,00004
67	Rzeszów	nieliniowa	1b	46	13	14	32	22,785	,870	,049	,00005
68	Rzeszów	nieliniowa	2	48	13	14	34	33,637	,824	,060	,00004
69	Rzeszów	nieliniowa	2a	47	13	14	33	26,123	,854	,052	,00005
70	Rzeszów	nieliniowa	2b	46	13	14	32	22,500	,871	,048	,00005
71	Rzeszów	nieliniowa	3	48	13	14	34	33,522	,825	,060	,00000
72	Rzeszów	nieliniowa	3a	47	13	14	33	25,904	,856	,052	,00004
73	Rzeszów	nieliniowa	3b	46	13	14	32	22,196	,873	,048	,00005
74	Rzeszów	nieliniowa	4	48	13	14	34	33,640	,824	,059	,00040
75	Rzeszów	nieliniowa	4a	47	13	14	33	25,669	,857	,051	,00004
76	Rzeszów	nieliniowa	4b	46	13	14	32	21,648	,876	,047	,00005
77	Świdnik	liniowa	2	42	11	12	30	30,296	,615	,094	,00001
78	Świdnik	liniowa	3	41	11	12	29	25,487	,676	,088	,00002
79	Świdnik	liniowa	4	41	7	8	33	27,622	,600	,073	,00000
80	Świdnik	liniowa	5	40	7	8	32	24,308	,652	,069	,00000
81	Świdnik	liniowa	6	40	10	11	29	22,262	,711	,079	,00000
82	Świdnik	liniowa	7	40	9	10	30	21,527	,711	,074	,00000
83	Świdnik	nieliniowa	2	41	13	14	27	25,092	,640	,093	,00023
84	Świdnik	nieliniowa	3	40	13	14	26	21,775	,736	,089	,00028
85	Świdnik	nieliniowa	4	41	13	14	27	24,974	,704	,093	,00023
86	Świdnik	nieliniowa	5	40	13	14	26	20,615	,760	,085	,00027
87	Świdnik	nieliniowa	7	40	14	15	25	16,437	,816	,080	,00060
88	Świdnik	nieliniowa	I	40	16	17	23	15,671	,839	,083	,00205
89	Świdnik	nieliniowa	II	40	15	16	24	15,515	,833	,079	,00118
90	Świdnik	nieliniowa	III	39	15	16	23	15,671	,834	,081	,00139
91	Świdnik	nieliniowa	IV	39	14	15	24	15,515	,829	,078	,00071
92	Trzyciąż	liniowa	1	50	13	14	36	1,864	,743	,063	,00005
93	Trzyciąż	nieliniowa	1	50	18	19	31	1,192	,859	,058	,00291
94	Trzyciąż	nieliniowa	2	49	18	19	30	,853	,898	,049	,00311
95	Trzyciąż	nieliniowa	3	48	18	19	29	,580	,928	,041	,00331
96	Trzyciąż	nieliniowa	4	50	14	15	35	1,289	,828	,053	,00015
97	Trzyciąż	nieliniowa	5	46	14	15	31	,523	,934	,035	,00025

Na podstawie zebranych wyników określono przedziały zmienności dla parametrów niezmienniczych i podstawowe statystyki dla nich, jak: wartość przeciętna, odchylenie standardowe, błąd standardowy wartości przeciętnej i granice przedziału ufności na poziomie ufności 0,99. Wyniki zawiera tabela 2.

By zbadać, w jakim stopniu odrzucenie modeli niezaakceptowanych przez test *Fishera-Snedecora* wpłynęło na wyniki podstawowych statystyk, w tym wartości przeciętnej i rozproszenia, przeprowadzono stosowne testy statys-

tyczne. Zostały zweryfikowane hipotezy o równości wariancji w obu zbiorach oraz o równości wartości przeciętnej.

Test dotyczący wartości przeciętnej wypadł niepomyślnie jedynie dla parametru σ_{tr} , tzn. hipotezę wyjściową o równości wartości przeciętnej należy w tym wypadku odrzucić. Taki wynik może stanowić pierwszą przesłankę do wskazania na niestabilność niezmiennika σ_{tr} lub jego związek z wynikami testu *Fishera-Snedecora*.

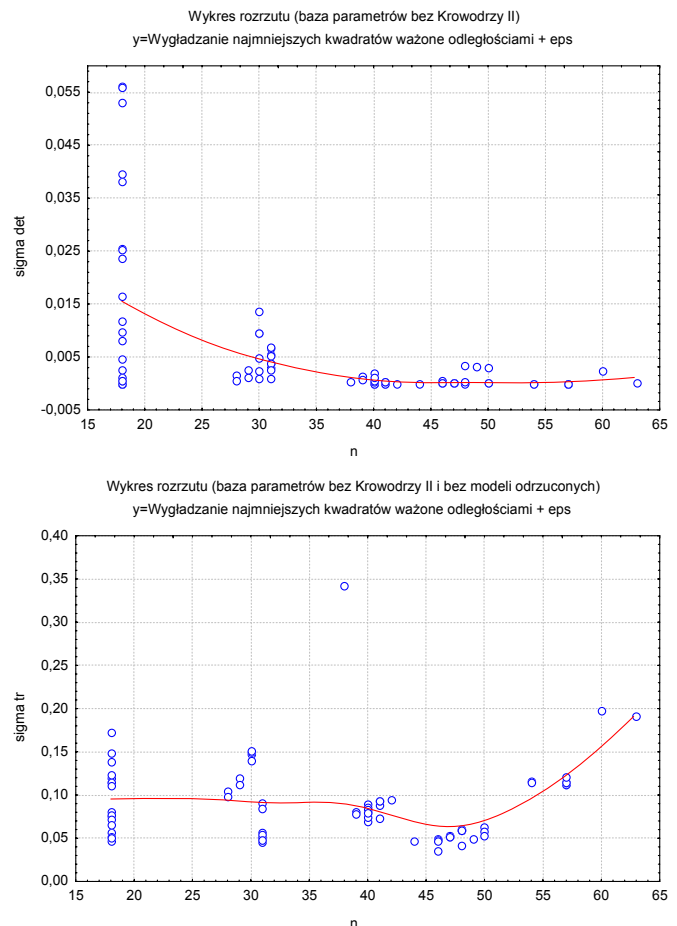
Tabela 2. Wartości podstawowych statystyk dla zbioru parametrów niezmienniczych.

	Wartość przeciętna \bar{X}	Granice przedziału ufności na poziomie 0,99		Minimum	Maximum	Odchylenie standardowe $\hat{\sigma}$	Błąd stand. wartości przeciętnej
wszystkie modele spełniające warunek $R^2 \geq 0,60$							
R^2	0,858	0,831	0,886	0,600	0,998	0,104	0,010
σ_{tr}	0,099	0,076	0,104	0,032	0,342	0,052	0,005
σ_{det}	0,00497	0,00189	0,00806	0,00000	0,05616	0,01156	0,00117
bez modeli odrzuconych przez test <i>Fishera-Snedecora</i>							
R^2	0,865	0,837	0,893	0,600	0,998	0,099	0,010
σ_{tr}	0,084	0,072	0,098	0,032	0,342	0,046	0,005
σ_{det}	0,00313	0,00079	0,00547	0,00000	0,05616	0,00833	0,00089

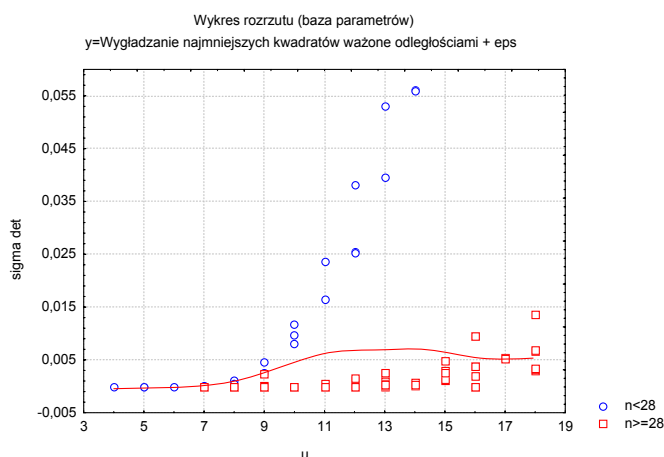
8 BADANIE ZALEŻNOŚCI NIEZMIENNIKÓW OD WIELKOŚCI CHARAKTERYZUJĄCYCH BAZĘ NIERUCHOMOŚCI I MODEL WYCENY

W celu sformułowania kryteriów oceny wiarygodności bazy nieruchomości i zastosowanego modelu wyceny sporządzono szereg wykresów rozrzutu zależności niezmienników od wielkości stałych charakteryzujących bazę nieruchomości i zastosowany model wyceny. Analizowanymi wykresami były: $R^2(n)$, $\sigma_{tr}(n)$, $\sigma_{det}(n)$, $R^2(u)$, $\sigma_{tr}(u)$, $\sigma_{det}(u)$, $R^2(\sigma_0^2)$, $\sigma_{tr}(\sigma_0^2)$, $\sigma_{det}(\sigma_0^2)$, $R^2(k \cdot \sigma_0^2)$, $\sigma_{tr}(k \cdot \sigma_0^2)$, $\sigma_{det}(k \cdot \sigma_0^2)$. Na każdy z tych wykresów nałożono, wyestymowaną metodą najmniejszych kwadratów, linię trendu. Pozwala to na stwierdzenie istnienia bądź też braku jakichkolwiek zależności pomiędzy tymi wielkościami, a co za tym idzie, na wyciągnięcie wniosków dotyczących modelu, czy też samej bazy.

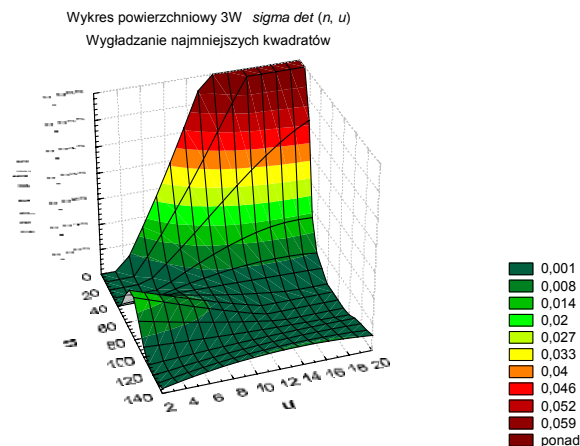
Poniżej zamieszczono dwa przykłady opisanych wykresów:



Najciekawsze wyniki dała analiza wykresów $\sigma_{det}(u)$. Wartość tego parametru rośnie wraz ze wzrostem u , ale widać wyraźnie, że tempo wzrostu jest uzależnione od wpływu innych czynników, gdyż punkty tworzące wykres rozrzutu grupują się w dwa zdecydowanie odrębne zbiory. Po dalszych analizach, polegających na próbach „rozdzielenia” zbiorów poprzez nałożenie dodatkowych warunków na poszczególne przypadki, zidentyfikowano parametry determinujące powyższy podział. Są nimi pozostałe charakterystyki baz: liczność bazy n i liczbę stopni swobody k oraz stałe, charakteryzujące wyestymowany model: wariancja resztowa σ_0^2 i wariancja resztowa pomnożona przez liczbę stopni swobody, czyli $Y^T Y - a^T X^T Y$. Najbardziej zdecydowany podział na dwie grupy uzyskano przy nałożeniu warunków na n oraz σ_0^2 . Poniżej zamieszczono wykresy $\sigma_{det}(u)$ z podziałem na podzbiory w zależności od wartości n .



Widać wyraźnie, że bardzo szybkie tempo wzrostu σ_{det} wraz ze zwiększającą się liczbą zmiennych niezależnych w modelu, otrzymujemy dla baz mało licznych, gdzie liczba nieruchomości nie przekracza 30. Natomiast, gdy jest większa od 30 sytuacja diametralnie się zmienia: parametr σ_{det} wykazuje minimalny przyrost dla zwiększającej się – nawet znacznie – liczby cech nieruchomości branych pod uwagę przy tworzeniu modelu. Obrazuje to również trójwymiarowy wykres zależności $\sigma_{det}(n, u)$.



Wynika stąd istotny wniosek, że niezbędne jest spełnienie warunku o zebraniu minimalnej liczby danych przed przystąpieniem do modelowania i stosowaniem metod statystycznych.

Na podstawie przeprowadzonego badania zależności parametrów niezmienniczych od wartości stałych, charakteryzujących bazę nieruchomości, czy też zastosowany model określenia rynkowej wartości nieruchomości – pierwszym nasuwającym się wnioskiem jest stwierdzenie, iż spośród zdefiniowanych niezmienników najlepszym wskaźnikiem jest σ_{det} . Na wykresach rozrzutu, sporządzonych z udziałem σ_{det} , punkty są najbardziej skupione wokół wyestymowanej linii trendu. Otrzymujemy najmniej odstępstw.

Najmniej jednoznaczne wyniki uzyskano za pomocą parametru σ_{tr} . Okazuje się on zatem najmniej przydatny w tego typu analizach. Punkty na wykresach rozrzutu z jego udziałem często tworzą nieregularne „chmury” i wykazują duże rozproszenie wokół wyestymowanej linii regresji.

Przy pomocy parametru R^2 uzyskano wyniki nieco gorsze od σ_{det} , ale zdecydowanie lepsze niż dla σ_{tr} . Jednak ten parametr, ze względu na swoją prostą definicję, odniesioną do współczynników korelacji zupełnej, może mieć duże praktyczne zastosowanie w wycenie nieruchomości.

9 WNIOSKI KOŃCOWE

Głównym celem pracy było zaprezentowanie możliwości oceny bazy nieruchomości i modelu zastosowanego do estymacji rynkowej wartości nieruchomości, za pomocą zdefiniowanych parametrów – niezmienników przekształcenia macierzy kowariancji dla modelowych wartości

nieruchomości. Po wyestymowaniu parametrów około stu modeli, testowanych w sumie na dziesięciu bazach nieruchomości, różniących się znacznie liczebnością (od 18 do 132 nieruchomości) i liczbą cech nieruchomości branych pod uwagę (od 4 do 18) oraz po wyznaczeniu dla nich wartości trzech parametrów niezmienniczych: R^2 , σ_{tr} , σ_{det} – na podstawie wykresów rozrzutu tych niezmienników w połączeniu z charakterystykami baz i modeli można sformułować następujące wnioski:

- ◆ Optymalna liczba nieruchomości w bazie do modelowania wartości nieruchomości powinna odpowiadać trzykrotnej liczbie wyznaczonych parametrów modelu wyceny. Przy czym należy zaznaczyć, że liczba ta powinna być co najmniej równa dwukrotnej liczbie wyznaczonych parametrów modelu, a znaczne jej zwiększanie (powyżej czterokrotnej liczby parametrów) najczęściej nie prowadzi do polepszenia modelu.
- ◆ Maksymalna liczba zmiennych niezależnych, występujących w modelach wyceny nie powinna przekraczać 14 parametrów, a wstępna analiza rynku nieruchomości powinna ustalić optymalną liczbę parametrów (atrybutów).
- ◆ Liczba stopni swobody w modelu wyceny powinna mieścić się w granicach od 28 do 42.
- ◆ Dobór bazy nieruchomości do estymacji modelu wyceny można uznać za optymalny, jeżeli wartość niezmiennika σ_{det} zawiera się w przedziale: $\sigma_{det} \in (0,0008; 0,0055)$.
- ◆ Dobór modelu do szacowania rynkowej wartości nieruchomości można uznać za zadowalający, gdy wartość współczynnika determinacji R^2 będzie spełniała nierówność: $R^2 \geq 0,837$.
- ◆ Za pomocą parametru σ_{tr} nie można sformułować kryteriów doboru bazy nieruchomości do estymacji modelu wyceny, gdyż wartość tego parametru wykazuje dużą fluktuację.
- ◆ Z porównania wniosków 4. i 6. wynika sprostowanie, że w doborze poprawnego modelu wyceny nie wystarczy uwzględnienie samych wariacji wartości modelowych, ale konieczne jest wzięcie pod uwagę całej macierzy kowariancji dla wartości modelowych $Cov(W)$.

BIBLIOGRAFIA

- 1) Adamczewski Z., Czarnecka K.; Modelowanie matematyczne ekowartości. Przegląd Geodezyjny nr 8/95, Warszawa 1995 r.
- 2) Barańska A.; Kryteria stosowania modeli stochastycznych w predykcji rynkowej wartości nieruchomości. Geodezja, tom 8, zeszyt 1, 2002 r.
- 3) Barańska A.; Zastosowanie uogólnionych modeli liniowych w wycenie nieruchomości. Geodezja, tom 5, zeszyt 2, 1999 r.
- 4) Barańska A., Mitka B.; Statystyczne przygotowanie baz danych do dalszych analiz. IX Krajowa Konferencja „Komputerowe Wspomaganie Badań Naukowych”, Polanica Zdrój, 24-26 października 2002 r.
- 5) Chow G.C.; Econometrics. New York, 1983 r.
- 6) Czaja J.; Metody szacowania wartości rynkowej i katalnej nieruchomości, Kraków 2001 r.
- 7) Czaja J.; Modele statystyczne w informacji o terenie. Książka, Wydawnictwa AGH Kraków 1997 r.
- 8) Grabowski R. J.; Ekonometria w zarysie, Wydawnictwo Wyższej Szkoły Finansów i Zarządzania w Białymstoku, Białystok 2002 r.
- 9) Hopfer A. i inni; Szacowanie nieruchomości niezurbanizowanych. Praca zbiorowa. Twigger, Warszawa 1996 r.
- 10) Krywicki W.; Statystyka matematyczna, PWN, Warszawa 1986 r.
- 11) Plucińska A., Pluciński E.; Probabilistyka, WNT, Warszawa 2000, Wyd. I.
- 12) Rao C. R.; Modele liniowe statystyki matematycznej. PWN, Warszawa, 1982 r.
- 13) Rozporządzenie Rady Ministrów z dnia 27 listopada 2002 r. w sprawie szczegółowych zasad wyceny nieruchomości oraz zasad i trybu sporządzania operatu szacunkowego. Dziennik Ustaw nr 230/2002, poz. 1924.
- 14) Standardy Zawodowe Rzeczoznawców Majątkowych – Polska Federacja Stowarzyszeń Rzeczoznawców Majątkowych. Wydanie VIII, Warszawa 2002 r.
- 15) Ustawa z dnia 21 sierpnia 1997 r. o gospodarce nieruchomościami. Dziennik Ustaw nr 115/1997, poz. 741 wraz z późniejszymi zmianami – jednolity tekst Dziennik Ustaw nr 46/2000, poz. 543.