BADANIA I ANALIZY

Marta WIDŁAK, Natalia NEHREBECKA

Wykorzystanie regresji kwantylowej w analizie zróżnicowania cen mieszkań

Hedoniczne modele cen mieszkań oparte są na założeniu, że całkowita cena dobra heterogenicznego jest funkcją ilości cech tego dobra i ich domniemanych (implikowanych) cen. Popyt i podaż na mieszkania są wypadkowymi popytu i podaży na ich poszczególne cechy (Lancaster, 1966; Rosen, 1974). Jednocześnie twórcy podstaw teorii hedonicznej, m.in. Rosen (1974) i Epple (1987), zwracają uwagę na trudność w empirycznym oszacowaniu funkcji popytu i podaży cech dóbr różnorodnych.

W praktyce stosunkowo najłatwiej jest oszacować hedoniczną funkcję ceny dobra heterogenicznego, którą uznaje się za wypadkową działania popytu i podaży cech mieszkaniowych. Do oszacowania tej funkcji wykorzystuje się z reguły metodę regresji wielorakiej z estymatorem MNK (metoda najmniejszych kwadratów). Można stosować także metodę najmniejszych odchyleń bezwzględnych (LAD — *Least Absolute Deviation*; własności tego estymatora szczegółowo opisano np. w Greene, 2000). MNK wyznacza parametry regresji poprzez minimalizację sumy kwadratów reszt. Prowadzi to do zbliżenia się do

średniej funkcji rozkładu warunkowego zmiennej objaśniającej. Metoda LAD minimalizuje natomiast sumę bezwzględnych błędów i zbliża się do warunkowej funkcji mediany. Chociaż średnia i mediana są dwiema ważnymi miarami tendencji centralnej, które reprezentują „średnie" zachowanie zmiennej objaśniającej, to jednak dostarczają niewielu informacji na temat zachowań danej zmiennej w „ogonach" rozkładu.

Przełomem w analizach jest regresja kwantylowa (*quantile regression*) zaproponowana przez Koenkera i Bassetta (1978). Podejście to pozwala na oszacowanie różnych funkcji kwantylowych warunkowej dystrybuanty zmiennej objaśnianej. Szczególny przypadek regresji kwantylowej dla kwantyla rzędu 0,5 (czyli mediany) jest równoważny estymatorowi LAD. Każdy kwantyl regresji charakteryzuje dany punkt (centrum lub „ogon") warunkowej dystrybuanty zmiennej objaśnianej. Wprowadzenie różnych kwantyli regresji przewiduje zatem pełniejszy opis podstawowych rozkładów warunkowych. Ta analiza jest szczególnie przydatna, gdy warunkowa dystrybuanta jest różnorodna i nie ma „standardowego" kształtu, tak jak w przypadku rozkładów asymetrycznych lub uciętych. Regresja kwantylowa zyskała dużo uwagi w literaturze stosunkowo niedawno (Koenker, 2000; Koenker, Hallock, 2001; Powell, 2002; Koenker, 2005).

W odniesieniu do rynków dóbr heterogenicznych zastosowanie estymatora MNK rodzi wiele problemów. Część z nich omawiali Tomczyk i Widłak (2010). Inne ważne kwestie dotyczą wykorzystania obserwacji nietypowych, tak by szacowany model przedstawiał w pełni złożoną rzeczywistość. Kolejną kwestią jest założenie o normalnym charakterze rozkładu zmiennej objaśnianej, czyli ceny mieszkania. Najczęściej przyjmuje się, że zmienna ta ma rozkład log-liniowy i modeluje się logarytm naturalny tej ceny. Jednak formalne testy sprawdzające czy dana zmienna ma rozkład normalny pokazują, że w praktyce założenie to nie zawsze jest spełnione także dla zmiennej zlogarytmowanej.

Niespełnienie założenia o normalności rozkładu zmiennej zależnej nie powoduje obciążenia estymatorów szacowanych za pomocą MNK. Estymatory te nadal są zgodne, jednak nie są efektywne wśród nieobciążonych estymatorów. W przypadku niespełnienia założenia o normalności rozkładu stosuje się odporne metody ekonometryczne, pozwalające na precyzyjniejsze oszacowanie nieznanych parametrów. Metodą pozwalającą na wykorzystanie obserwacji nietypowych jest m.in. regresja kwantylowa, którą tu zastosowano. Jak zauważyli Koenker i Basset (1978), w przypadku występowania heteroskedastyczności estymacja regresji kwantylowej rzędu 0,5 może okazać się bardziej efektywnym sposobem szukania wartości parametrów niż tradycyjna regresja oparta na wartości oczekiwanej zmiennej objaśnianej. Dodatkowo, jako że w metodzie regresji kwantylowej wykorzystuje się całą próbę, nie pojawia się problem obciążenia estymatorów, który zaistniałby gdyby użyto MNK do podprób wyodrębnionych na podstawie cen mieszkań. Mając na uwadze te różnice, w badaniu porównano wyniki oszacowań z wykorzystaniem standardowego podejścia MNK oraz ujęcia regresji kwantylowej.

W analizowanym modelu, w którym zmienną objaśnianą jest logarytm ceny m^2 mieszkania, metoda regresji kwantylowej umożliwia nie tylko rozwiązanie niektórych ze wspomnianych problemów metodycznych, ale także pozwala odpowiedzieć na bardzo ciekawe pytanie ekonomiczne. Czy wyceny poszczególnych cech mieszkania różnią się dla wyodrębnionych segmentów cenowych rynku mieszkaniowego? Różnice te mogą wynikać z odmiennych gustów i preferencji klientów o rozmaitych ograniczeniach budżetowych, wybierających mieszkania z różnych przedziałów cenowych.

Warto przeanalizować dwa skrajne przykłady — tzw. apartamenty oraz mieszkania o najniższym standardzie. Ceny m^2 mieszkań z obydwu grup będą zatem odmienne, co powoduje, że mieszkania te będą nabywane przez osoby różniące się gustami i ograniczeniami budżetowymi. Im mniejszy rozstęp cenowy pomiędzy grupami, tym silniejszy będzie efekt substytucji mieszkań o różnej jakości i cenie. Oznacza to, iż może zdarzyć się, że mieszkania małe o wysokiej cenie za m^2 znajdą także nabywców wśród klientów biedniejszych. Wówczas potrzeby i preferencje klientów mniej zamożnych zdecydują o substytucji wielkości mieszkania jego lokalizacją lub standardem wykończenia. Bardziej dokładna weryfikacja wpływu ograniczenia budżetowego na ceny implikowane cech mieszkania powinna zostać przeprowadzona z wykorzystaniem cen całkowitych mieszkań. Tego typu badania wpisują się w najnowszy nurt analiz związanych z modelowaniem hedonicznym cen mieszkań i wykorzystaniem bardziej zaawansowanej techniki ekonometrycznej (np. Zietz i in., 2008; Bayer i in., 2004; Liao i Wang, 2010).

W pracy dodatkowo poruszono trzecie zagadnienie, dotyczące wykorzystania szacowanych modeli hedonicznych cen mieszkań w procesie wyznaczania miernika ich dynamiki. Hedoniczne modele cen mieszkań wykorzystuje się też do korekty indeksów cenowych o zmiany jakości badanych dóbr (Widłak, 2010). Zastosowanie regresji kwantylowej umożliwi konstrukcję hedonicznych indeksów cen mieszkań odrębnie dla każdego z badanych segmentów cenowych.

Ciekawe badanie przeprowadzili McMillan i Thorsnes (2006), którzy szacowali indeksy powtórnej sprzedaży (Bailey i in., 1963) dla rynku mieszkaniowego w Chicago, wykorzystując metodę regresji kwantylowej. W ten sposób rozwiązują do pewnego stopnia problem obserwacji odstających, które w znacznej mierze odzwierciedlają, zaniedbywane w klasycznym modelu indeksu powtórnej sprzedaży, renowacje i udoskonalenia mieszkań dokonane między okresami sprzedaży.

Obserwacja warszawskiego rynku mieszkań sugeruje występowanie różnic w ocenie dynamiki cen mieszkań w poszczególnych segmentach cenowych. Tendencje te były szczególnie wyraźne wraz z początkiem światowego kryzysu gospodarczego. W 2008 r. na warszawskim rynku wtórnym obserwowaliśmy wzrost popytu na mieszkania drogie, o wyższym standardzie i lepszej lokalizacji oraz spadek zainteresowania mieszkaniami tańszymi wykonanymi w technologii wielkopłytowej. Obecnie, pomimo stabilizacji cen mieszkań, wciąż obserwuje się odmienną dynamikę w poszczególnych segmentach cenowych, np. mieszka-

nia małe, jedno- i dwupokojowe, cieszą się większym zainteresowaniem niż mieszkania duże o niższej cenie za m^2.

METODA REGRESJI KWANTYLOWEJ

Regresja kwantylowa jest metodą estymacji zależności całego rozkładu zmiennej objaśnianej od zmiennych objaśniających. W przypadku klasycznej regresji modelujemy zależność między wartością oczekiwaną zmiennej objaśnianej a zmiennymi objaśniającymi. Hiperpłaszczyzna regresji y na x w tym przypadku jest warunkową wartością oczekiwaną:

$$E(y|x) = \mu(x) \tag{1}$$

gdzie:

x — macierz zmiennych objaśniających,
y — wektor zmiennej zależnej,
$E(y|x)$ — warunkowa wartość oczekiwana zmiennej zależnej y względem zmiennych objaśniających x.

Klasyczna analiza regresji przedstawia niepełny obraz zależności pomiędzy zmiennymi, zwłaszcza w sytuacjach gdy model jest heteroskedastyczny. Regresja kwantylowa pozwala na rozszerzenie liniowej estymacji zmian wartości dystrybuanty zmiennej objaśnianej. Estymacja regresji na kwantylach ma charakter semiparametryczny, co oznacza, że nie przyjmuje się założeń o typie rozkładu dla losowego wektora reszt w modelu. Przyjmowana jest jedynie parametryczna postać modelu w deterministycznej części modelowania.

Według autorów koncepcji regresji kwantylowej, jeżeli postać dystrybuanty jest znana, to kwantyl rzędu τ można policzyć w następujący sposób:

$$\xi_\tau = F_y^{-1}(\tau) \tag{2}$$

gdzie:

ξ_τ — kwantyl rzędu $\tau \in [0,1]$ (np. $\tau = 0{,}5$ jest medianą),
F — dystrybuanta zmiennej y.

Ideą regresji kwantylowej jest badanie związku między wielkością kwantyla wybranego rzędu a zmiennymi objaśniającymi. Wówczas można zdefiniować warunkowy kwantyl postaci:

$$\xi_\tau(x) = F_{y|x}^{-1}(\tau) \tag{3}$$

Warto zwrócić uwagę, że równania regresji mogą się różnić dla poszczególnych kwantyli.

W przypadku regresji dla wartości oczekiwanej standardowo stosuje się MNK. Jako $\mu(x)$ przyjmuje się funkcję parametryczną $\mu(x,\beta)$ o nieznanym wektorze parametrów β. Wektor β można znaleźć minimalizując sumę kwadratów reszt:

$$\min_{\beta \in R^K} \sum_{i=1}^{N} [y_i - \mu(x_i, \beta)]^2 \tag{4}$$

W przypadku regresji na kwantylach estymację można także sprowadzić do rozwiązywania problemu minimalizującego. Najprostszy przypadek dotyczy estymacji parametrów regresji dla kwantyla $\tau = 0{,}5$, czyli mediany. Wówczas parametry β są wynikiem minimalizacji następującej funkcji celu:

$$\min_{\beta \in R^K} \sum_{i=1}^{N} |y_i - \xi_{0,5}(x_i, \beta)| \tag{5}$$

gdzie:

N — liczba obserwacji,
y_i — wartość zmiennej zależnej,
pozostałe oznaczenia jak poprzednio.

Estymator regresji jest tu równoważny estymatorowi LAD, który uzyskuje się za pomocą metody najmniejszych odchyleń bezwzględnych (Narula, Wellington, 1982).

W ogólnym przypadku estymacja parametrów regresji dowolnego kwantyla polega na minimalizacji ważonej sumy wartości bezwzględnych reszt, przypisując im odpowiednie wagi:

$$\min_{\beta \in R^K} \sum_{i=1}^{N} \rho_\tau(|y_i - \xi_\tau(x_i, \beta)|) \tag{6}$$

gdzie:

$$\rho_\tau(z) = \begin{cases} \tau z & \text{dla } z \geq 0 \\ (1-\tau) z & \text{dla } z < 0 \end{cases} \tag{7}$$

Estymacja każdorazowo przebiega na całej próbie, jednak dla każdego kwantyla szacowany jest inny parametr beta. Dzięki temu obserwacje nietypowe otrzymują niższe wagi, co rozwiązuje problem uwzględniania ich w modelu. W zależności od charakteru zjawiska i rozkładu danych w zastosowaniach empirycznych najczęściej estymuje się od trzech do dziewięciu różnych regresji kwantylowych (są to regresje odpowiadające kolejnym kwantylom lub decylom rozkładu). Na podstawie wszystkich otrzymanych modeli analizuje się dane zjawisko. Najczęściej estymatory błędów standardowych współczynników dla regresji kwantylowej uzyskuje się za pomocą metody bootstrap, jak zasugerował to Gould (1992; 1997). Są one mniej wrażliwe na heteroskedastyczność niż estymatory oparte na metodzie zaproponowanej przez Rogera (1993). W opisywanym badaniu replikowano reszty modelu za pomocą metody bootstrap z 500 powtórzeniami. Wartości estymatorów w tej regresji będą takie same, jak dla zwykłej regresji kwantylowej. Różnić się jednak będą wartości błędów standardowych.

W badaniu przeprowadzono również estymację regresji międzykwantylowej. Polega ona na oszacowaniu równania regresji, gdzie za zmienną objaśnianą przyjmuje się różnicę między jej kolejnymi kwantylami, a przy zmiennych objaśniających występują współczynniki zdefiniowane jako różnice wartości parametrów beta w kolejnych kwantylach.

DANE

Do analizy wykorzystano Bazę Rynku Nieruchomości (BaRN) tworzoną w NBP. W modelach wykorzystano dane dla warszawskiego rynku wtórnego mieszkań. Notowania cen prowadzone przez NBP odbywają się kwartalnie (począwszy od III kwartału 2006 r.) i dotyczą mieszkań zlokalizowanych w granicach administracyjnych miasta. Dane wykorzystane w analizie obejmują transakcje zanotowane przez biura pośrednictwa sprzedaży nieruchomości. Informacje dotyczą wyłącznie transakcji zawartych na wolnym rynku.

Poza podstawowymi informacjami o cenie w BaRN gromadzone są informacje o wybranych parametrach charakteryzujących mieszkanie, a także, w ograniczonym zakresie, dane dotyczące oceny jego lokalizacji szczegółowej i ogólnej. Część zmiennych opisujących lokalizację jest typu jakościowego. Wartość tych zmiennych jest nadawana przez ekspertów zajmujących się analizą rynków lokalnych i mają przez to charakter subiektywny. Pełna lista zmiennych zawiera dziewiętnaście atrybutów opisujących fizyczne cechy mieszkania oraz pięć zmiennych lokalizacyjnych. Specyfikację modelu oparto na pracy Tomczyk i Widłak (2010) wykorzystującej ten sam zbiór danych. Znaczna liczba braków danych ograniczyła zbiór zmiennych wykorzystanych w szacowanych modelach zaledwie do dziewięciu, w tym tylko jednej zmiennej lokalizacyjnej określającej dzielnicę. W stosunku do modelu wyjściowego inaczej ujęto kwestię lokalizacji, dołączono także zmienną *liczba kondygnacji*, wykazującą wyraźną zależność nieliniową od ceny m^2 mieszkania.

OPIS ZMIENNYCH

Zmienne	Typ	Definicja
lcena_m2	ilościowa	logarytm ceny mieszkania wyrażonej w zł za metr kwadratowy
ilekondygnacji	ilościowa	liczba kondygnacji nadziemnych budynku
sq_ilekondygnacji	ilościowa	kwadrat liczby kondygnacji nadziemnych budynku
cub_ilekondygnacji	ilościowa	sześcian liczby kondygnacji nadziemnych budynku
ilepokoi_1_2	binarna[a]	1 — mieszkania jedno- lub dwupokojowe,
kuchnia	jakościowa	kuch_1: 1 — mieszkania z oddzielną ciemną kuchnią (zmienna bazowa[b]) kuch_2: 1 — mieszkania z oddzielną widną kuchnią (zmienna bazowa) kuch_3: 1 — mieszkania z aneksem kuchennym połączonym z salonem
rokbudowy	ilościowa (przedstawiona za pomocą 8 zmiennych binarnych — zgodnie z klasyfikacją przyjętą w Narodowym Spisie Powszechnym (2002 r.)	rok_1: 1 — budynki wybudowane przed 1918 r. rok_2: 1 — budynki wybudowane w latach 1918—1944 rok_3: 1 — budynki wybudowane w latach 1945—1970 rok_4: 1 — budynki wybudowane w latach 1971—1978 rok_5: 1 — budynki wybudowane w latach 1979—1988 rok_6: 1 — budynki wybudowane w latach 1989—2001 (zmienna bazowa) rok_7: 1 — budynki wybudowane w latach 2002—2005 rok_8: 1 — budynki wybudowane po 2005
rok_nowe	binarna	1 — mieszkania wybudowane co najwyżej jeden rok przed rokiem notowania. Zmienna wskazuje na mieszkania nowe lub jeszcze niewybudowane przez dewelopera, a sprzedawane już na rynku wtórnym (obrót mieszkaniami i prawami do mieszkań)
spoldzielcze	binarna	1 — transakcje oparte na spółdzielczym własnościowym prawie do lokalu
sq_pow	ilościowa	kwadrat powierzchni użytkowej mieszkania wyrażonej w m^2
standard_wykonczenia[c]	jakościowa	stand_1: 1 — mieszkania o wysokim standardzie wykończenia stand_2: 1 — mieszkania o przeciętnym standardzie wykończenia (zmienna bazowa) stand_3: 1 — mieszkania o niskim standardzie wykończenia stand_4: 1 — mieszkania w tzw. stanie surowym
dzielnica	jakościowa (przedstawiona za pomocą 12 zmiennych binarnych)	gr1: 1 — mieszkania zlokalizowane w dzielnicach Bemowo, Bielany gr2: 1 — mieszkania zlokalizowane w dzielnicach Białołęka, Targówek gr3: 1 — mieszkania zlokalizowane w dzielnicach Praga Północ, Praga Południe gr4: 1 — mieszkania zlokalizowane w dzielnicach Rembertów, Wesoła, Wawer gr5: 1 — mieszkania zlokalizowane w dzielnicach Włochy, Ursus

a Dla zmiennych binarnych wyjaśniono, kiedy zmienna przyjmuje wartość 1; w pozostałych przypadkach zmienne binarne przyjmują wartość 0. *b* Zmienna bazowa; zmienne bazowe pominięto w estymacji i dlatego nie ma dla nich wyników w tabl. 2—4. *c* Standard_wykonczenia jest zmienną typu jakościowego i przyjmuje cztery kategorie. Wartości zmiennej odzwierciedlają subiektywną ocenę ekspercką (pośrednika lub analityka rynku lokalnego) o jakości wykończenia mieszkania według instrukcji badania NBP.

OPIS ZMIENNYCH (dok.)

Zmienne	Typ	Definicja
dzielnica (dok.)	jakościowa (przedstawiona za pomocą 12 zmiennych binarnych) (dok.)	mokotow: 1 — mieszkania zlokalizowane w dzielnicy Mokotów ochota: 1 — mieszkania zlokalizowane w dzielnicy Ochota wola: 1 — mieszkania zlokalizowane w dzielnicy Wola ursynow: 1 — mieszkania zlokalizowane w dzielnicy Ursynów zoliborz: 1 — mieszkania zlokalizowane w dzielnicy Żoliborz wilanow: 1 — mieszkania zlokalizowane w dzielnicy Wilanów srodmiescie: 1 — mieszkania zlokalizowane w dzielnicy Śródmieście (zmienna bazowa)

W okresie od III kwartału 2006 r. do II kwartału 2010 r. BaRN zgromadzono 2056 obserwacji dotyczących transakcji na warszawskim rynku wtórnym mieszkań. Metoda regresji kwantylowej umożliwia wykorzystanie pełnego zbioru informacji. Przygotowanie danych ograniczyło się do ich analizy pod kątem błędów we wprowadzaniu danych. Tabl. 1 przedstawia statystykę opisową pełnego zbioru danych oraz w podziale na trzy grupy cenowe wyznaczone przez pierwszy i trzeci kwartyl ceny m^2 mieszkania. Dodatkowo dla każdego kwartyla cen mieszkań można odczytać związaną z nim średnią cenę wraz z 95% przedziałem ufności oraz ceny minimalne i maksymalne. Ponadto można sprawdzić, ile wynosiła średnia wartość zmiennych objaśniających dla mieszkań znajdujących się w przedziałach cenowych wyznaczonych przez kwartyle. W przypadku zmiennych 0—1 wartości w tabl. 1 oznaczają udział obserwacji o wartości 1 w obserwacji ogółem.

TABL. 1. TRANSAKCJE NA WARSZAWSKIM WTÓRNYM RYNKU MIESZKANIOWYM W OKRESIE III KWARTAŁ 2006—II KWARTAŁ 2010 (2056 obserwacji)

Zmienne	Średnia	Odchylenie standardowe	Min	Max	Q0,25	Q0,5	Q0,75
cena_m2 (w zł za m. kw.)	8118	2114	2531	21537	6801	7899	9322
95% przedział ufności	—	—	—	—	6666 6894	7820 8021	9191 9480
cena całkowita w zł	460725	305947	93605	3200000	356207	424924	665981
ilekondygnacji	7,34	3,79	2	31	7,426[a]	7,586	7,156
powierzchnia	56,24	29,71	14	400	59,724	51,238	61,401
ilepokoi_1_2	0,63	—	0	1	0,516	0,663	0,615
kuch_1	0,12	—	0	1	0,096	0,126	0,106
kuch_2	0,67	—	0	1	0,764	0,694	0,593
kuch_3	0,21	—	0	1	0,138	0,180	0,301
rok_1	0,01	—	0	1	0,000	0,002	0,020
rok_2	0,07	—	0	1	0,049	0,041	0,123

a Dla zmiennych objaśniających podano wartości średnie w danej grupie cenowej mieszkań; trzy grupy cenowe wyznaczone są przez wartości pierwszego i trzeciego kwartyla ceny metra kwadratowego mieszkania.

TABL. 1. TRANSAKCJE NA WARSZAWSKIM WTÓRNYM RYNKU MIESZKANIOWYM W OKRESIE III KWARTAŁ 2006 — II KWARTAŁ 2010 (2056 obserwacji) (dok.)

Zmienne	Średnia	Odchylenie standardowe	Min	Max	Q0,25	Q0,5	Q0,75
rok_3	0,32	—	0	1	0,280	0,306	0,290
rok_4	0,11	—	0	1	0,153	0,165	0,029
rok_5	0,10	—	0	1	0,170	0,153	0,006
rok_6	0,18	—	0	1	0,202	0,173	0,159
rok_7	0,11	—	0	1	0,084	0,080	0,184
rok_8	0,11	—	0	1	0,063	0,080	0,190
rok_nowe	0,08	—	0	1	0,068	0,065	0,112
spoldzielcze	0,41	—	0	1	0,519	0,499	0,232
stand_1	0,35	—	0	1	0,263	0,305	0,499
stand_2	0,41	—	0	1	0,466	0,418	0,302
stand_3	0,18	—	0	1	0,241	0,222	0,093
stand_4	0,06	—	0	1	0,030	0,054	0,106
gr1	0,10	—	0	1	0,140	0,117	0,045
gr2	0,08	—	0	1	0,191	0,115	0,000
gr3	0,08	—	0	1	0,134	0,107	0,033
gr4	0,01	—	0	1	0,021	0,012	0,000
gr5	0,03	—	0	1	0,045	0,031	0,002
mokotow	0,20	—	0	1	0,130	0,183	0,284
ochota	0,05	—	0	1	0,043	0,062	0,053
wola	0,08	—	0	1	0,072	0,080	0,066
ursynow	0,15	—	0	1	0,142	0,148	0,154
zoliborz	0,06	—	0	1	0,037	0,060	0,068
wilanow	0,02	—	0	1	0,010	0,021	0,012
srodmiescie	0,14	—	0	1	0,035	0,064	0,284

Ź r ó d ł o: opracowanie własne.

W badaniu dokonujemy podziału mieszkań na trzy segmenty cenowe, dla których granicami są wartości pierwszego i trzeciego kwartyla ceny m^2 mieszkania. Porównując wyniki estymacji modeli hedonicznych w segmencie pierwszym (najniższa cena) i trzecim (najwyższa) możemy zakładać, że wynikają one z różnych gustów i preferencji klientów, które w znacznym stopniu zależą od ich ograniczeń budżetowych. Porównanie segmentów pierwszego i drugiego będzie natomiast odzwierciedlało w głównej mierze efekty substytucji cech mieszkaniowych, a więc przede wszystkim różnice w preferencjach nabywców.

WYNIKI ESTYMACJI

Modele hedoniczne

Hedoniczny model ceny mieszkania przyjmuje następującą postać:

$$\ln Y = \beta_0 + \sum_i \beta_i X_i + \varepsilon \tag{8}$$

gdzie:

Y — cena mieszkania wyrażona w złotych za m^2,
β_0 — stała,
β_i — nieznany współczynnik dla *i*-tej charakterystyki mieszkania,
X_i — *i*-tą zmienną objaśniającą,
ε — składnik losowy.

W ramach estymacji hedonicznych modeli cen oszacowano dwa log-liniowe jednorównaniowe modele ekonometryczne. Modele te różnią się zakresem czasowym danych, a model wykorzystujący pełen zakres czasowy próby jest wzbogacony o zmienne 0—1 (zero-jedynkowe) czasu. Taki kształt analizy jest podyktowany realizacją jednego z celów badania, jakim jest wyznaczenie hedonicznego indeksu cen. Model oparty na danych ze wszystkich okresów ze zmiennymi 0—1 czasu oznaczamy jako M1, a model oparty na danych z czterech kwartałów 2008 r. oznaczamy jako M4. Model M1 służy do wyznaczenia indeksu hedonicznego tzw. metodą ze zmiennymi 0—1 czasu, a model M4 jest wykorzystany do wyznaczenia indeksu hedonicznego w formule tzw. metody imputacji bezpośredniej. Szczegółowy opis metod wyznaczania hedonicznych indeksów cen można znaleźć w pracy Widłak (2010).

Ze względu na wielkość dostępnych prób danych, a także mając na uwadze przejrzystość wykonywanej analizy, stosując metodę regresji kwantylowej zdecydowano się na estymację w trzech różnych segmentach cenowych rynku. Segmenty te zostały wyznaczone przez wartości pierwszego, drugiego i trzeciego kwartyla rozkładu empirycznego ceny m^2 mieszkania (regresje na kwantylach rzędu: 0,25; 0,5; 0,75; w dalszej części pracy regresje te oznaczono jako: Q25, Q50 oraz Q75). Wszystkie zaakceptowane modele cechują się poprawną konstrukcją (popartą wynikiem testu RESET lub linktest) oraz brakiem współliniowości — czynniki inflacji wariancji nie przekraczają wartości 10.

W kolejnym etapie estymacji została oszacowana regresja międzykwantylowa z błędami standardowymi wyznaczonymi za pomocą metody *bootstrap*[1] dla czterech wariantów testowanych hipotez: 0,25—0,50; 0,50—0,75; 0,25—0,75 oraz 0,25—0,50—0,75.

Wyniki wskazują, że oszacowane współczynniki różnią się znacznie wśród kwantyli dla dwóch zmiennych objaśniających charakteryzujących mieszkanie — standardu wykończenia i lokalizacji. W modelu M4 w zależności od segmentu cenowego różnie wyceniany jest wysoki standard wykończenia mieszkania (stand_1). Przy zakupie mieszkań o najwyższej cenie za metr kwadra-

[1] Stosując metodę bootstrap z 20, 100, i 500 replikacjami reszt uzyskano bardzo zbliżone wartości błędów standardowych oszacowań parametrów; większa liczba replikacji wzmacnia efekt istotności statystycznej, ale nie zmienia wniosku z zastosowanej estymacji. Wyniki estymacji odnoszą się do 500 replikacji.

TABL. 2. WYNIKI ESTYMACJI MNK I REGRESJI KWANTYLOWEJ Z BŁĘDAMI STANDARDOWYMI WYLICZONYMI PRZY ZASTOSOWANIU REPLIKACJI RESZT METODĄ BOOTSTRAP 500 POWTÓRZEŃ

Zmienne	M1				M4			
	Q_25	Q_50	Q_75	MNK	Q_25	Q_50	Q_75	MNK
ilekondygnacji	0,02[a]	0,02	0,00	0,01	0,07[a]	0,06[c]	0,01	0,05[c]
sq_ilekondygnacji	–0,00[b]	–0,00[a]	–0,00	–0,00[b]	–0,01[a]	–0,01[c]	–0,00	–0,01[c]
cub_ilekondygnacji	0,00[a]	0,00[a]	0,00	0,00[c]	0,00[a]	0,00[b]	0,00	0,00[c]
stand_1	0,07[c]	0,06[c]	0,05[c]	0,07[c]	0,11[c]	0,07[c]	0,06[c]	0,09[c]
stand_3	–0,06[c]	–0,07[c]	–0,04[c]	–0,06[c]	–0,06[c]	–0,07[c]	–0,02	–0,05[c]
stand_4	–0,01	0,02	–0,03	–0,03	0,10[a]	0,05	0,01	0,04
rok_1	–0,02	0,03	0,05	0,06	0,00	0,05	0,06	0,04
rok_2	0,02	0,02	0,04	0,02	–0,07[a]	–0,12[c]	–0,12[c]	–0,10[c]
rok_3	–0,04[a]	–0,07[c]	–0,07[c]	–0,06[c]	–0,12[c]	–0,17[c]	–0,17[c]	–0,14[c]
rok_4	–0,09[c]	–0,14[c]	–0,11[c]	–0,10[c]	–0,10[c]	–0,14[c]	–0,17[c]	–0,13[c]
rok_5	–0,09[c]	–0,15[c]	–0,15[c]	–0,12[c]	0,06[a]	0,08[c]	0,05	0,06[c]
rok_7	0,10[c]	0,09[c]	0,08[c]	0,11[c]	0,10[c]	0,09[b]	0,11[c]	0,09[c]
rok_8	0,12[c]	0,08[c]	0,11[c]	0,11[c]	–0,18[c]	–0,15[c]	–0,14[c]	–0,14[c]
rok_nowe	–0,07[b]	–0,11[c]	–0,08[c]	–0,05	–0,02	–0,01	–0,01	–0,02
spoldzielcze	–0,04[c]	–0,04[c]	–0,03[c]	–0,04[c]	0,06[c]	0,05[c]	0,06[c]	0,05[c]
ilepokoi_1_2	0,07[c]	0,06[c]	0,05[c]	0,05[c]	–0,00	0,00	0,00[b]	0,00
sq_pow	–0,00	–0,00[a]	–0,00	–0,00[b]	0,04[c]	0,01	0,02	0,03[b]
kuch_3	0,02	0,03[c]	0,03[c]	0,03[c]	–0,18[c]	–0,24[c]	–0,27[c]	–0,23[c]
gr1	–0,20[c]	–0,23[c]	–0,28[c]	–0,26[c]	–0,39[c]	–0,50[c]	–0,50[c]	–0,47[c]
gr2	–0,45[c]	–0,50[c]	–0,51[c]	–0,51[c]	–0,27[c]	–0,31[c]	–0,28[c]	–0,29[c]
gr3	–0,28[c]	–0,31[c]	–0,30[c]	–0,31[c]	–0,37[c]	–0,54[c]	–0,60[c]	–0,46[c]
gr4	–0,55[c]	–0,49[c]	–0,44[c]	–0,56[c]	–0,30[c]	–0,43[c]	–0,38[c]	–0,35[c]
gr5	–0,37[c]	–0,42[c]	–0,40[c]	–0,41[c]	–0,07[b]	–0,17[c]	–0,18[c]	–0,13[c]

TABL. 2. WYNIKI ESTYMACJI MNK I REGRESJI KWANTYLOWEJ Z BŁĘDAMI STANDARDOWYMI WYLICZONYMI PRZY ZASTOSOWANIU REPLIKACJI RESZT METODĄ BOOTSTRAP 500 POWTÓRZEŃ (dok.)

Zmienne	M1				M4			
	Q_25	Q_50	Q_75	MNK	Q_25	Q_50	Q_75	MNK
mokotow	–0,10[c]	–0,14[c]	–0,18[c]	–0,14[c]	–0,11[b]	–0,16[c]	–0,18[c]	–0,14[c]
ochota	–0,12[c]	–0,14[c]	–0,20[c]	–0,15[c]	–0,14[c]	–0,18[c]	–0,12[c]	–0,15[c]
wola	–0,16[c]	–0,18[c]	–0,21[c]	–0,19[c]	–0,14[c]	–0,20[c]	–0,24[c]	–0,19[c]
ursynow	–0,15[c]	–0,18[c]	–0,23[c]	–0,21[c]	–0,12[c]	–0,17[c]	–0,21[c]	–0,16[c]
zoliborz	–0,11[c]	–0,13[c]	–0,17[c]	–0,14[c]	–0,18[b]	–0,23[a]	–0,18	–0,21[c]
wilanow	–0,21[c]	–0,26[c]	–0,30[c]	–0,28[c]	0,07[a]	0,06[c]	0,01	0,05[c]
t_20063	–0,40[c]	–0,33[c]	–0,34[c]	–0,37[c]				
t_20064	–0,12[c]	–0,14[c]	–0,18[c]	–0,14[c]				
t_20071	–0,06[b]	–0,05[a]	–0,05	–0,06[c]				
t_20072	0,07[c]	0,05[b]	–0,01	0,06[c]				
t_20073	0,12[c]	0,08[c]	0,07[b]	0,10[c]				
t_20074	0,12[c]	0,09[c]	0,07[c]	0,09[c]				
t_20081	0,10[c]	0,08[c]	0,06[b]	0,09[c]				
t_20082	0,11[c]	0,07[c]	0,02	0,07[c]				
t_20083	0,08[c]	0,06[a]	0,03	0,08[c]				
t_20084	0,08[c]	0,06[c]	0,02	0,07[c]				
t_20092	0,06[b]	0,00	–0,05	0,03				
t_20093	0,01	–0,00	–0,02	0,00				
t_20094	0,01	–0,01	–0,04	0,01				
t_20101	0,04	0,02	0,00	0,04[a]				
t_20102	0,06[a]	0,05[a]	0,01	0,04[b]				
Stała	8,97[c]	9,15[c]	9,34[c]	9,16[c]	8,91[c]	9,13[c]	9,34[c]	9,13[c]
R^2				0,65				0,65
$\bar{R}^2$				0,62				0,62
Pseudo-R^2	0,42	0,44	0,47		0,43	0,42	0,41	

U w a g a. Oznaczenia: *a* — na poziomie istotności 10%; Q_25 — regresja na kwantylu rzędu 0,25; Q_50 — regresja na kwantylu rzędu 0,50; Q_75 — regresja na kwantylu rzędu 0,75; R^2 — współczynnik determinacji liniowej (MNK); $\bar{R}^2$ — skorygowany współczynnik determinacji liniowej (MNK); *Pseudo*-R^2— nie jest porównywalny z R^2 obliczonym za pomocą MNK, ponieważ jest to lokalna, a nie globalna miara dopasowania. *b* — na poziomie istotności 5%. *c* — na poziomie istotności 1%.

Ź r ó d ł o: obliczenia własne.

towy (kwantyl rzędu 0,75) standard ten jest o 5 p.proc. niżej wyceniany przez klientów niż przy zakupie mieszkań najtańszych (kwantyl rzędu 0,25). Wynik ten potwierdza regresja międzykwantylowa, na podstawie której można zauważyć, że dla tej zmiennej występują istotne statystycznie różnice pomiędzy współczynnikami regresji kwantylowej Q25_Q75, a także Q25_Q50 na poziomie istotności 10%. Oznacza to, że klienci kupujący najtańsze mieszkania bardziej cenią sobie ich dobre wykończenie niż klienci kupujący mieszkania z dwóch pozostałych segmentów cenowych. Wniosek ten wydaje się być w pełni uzasadniony, klienci z większym ograniczeniem budżetowym prawdopodobnie wolą nie poświęcać dodatkowych funduszy na wykończenie mieszkania.

Podobnie, niski standard mieszkania jest mniej istotny dla nabywców płacących najwięcej za mieszkanie w porównaniu do pozostałych nabywców (regresja międzykwantylowa wskazuje istotne statystycznie różnice dla zmiennej stand_3 występujące pomiędzy drugą i trzecią grupą cenową).

Warto zauważyć, że zarówno w modelu M1 jak i M4 współczynnik stojący przy zmiennej reprezentującej stan surowy (stand_4) jest istotny tylko w jednym przypadku — dla mieszkań najtańszych w modelu M4. Na tej podstawie można wnioskować, że bez względu na różnice w gustach, preferencjach i ograniczeniach budżetowych nabywcy podobnie wyceniają mieszkania w stanie surowym i mieszkania o średnim standardzie wykończenia, wcześniej zamieszkane. Wniosek ten nie jest zgodny z powszechną opinią, że mieszkania nowe z rynku pierwotnego są wyżej cenione od mieszkań wcześniej zamieszkanych.

Bez względu na stosowaną metodę — MNK czy regresja kwantylowa — współczynniki przy zmiennych określających dzielnicę, w której zlokalizowane jest mieszkanie, są statystycznie istotne. Wyniki regresji międzykwantylowej dla tej zmiennej wskazują na istotne różnice w ocenie większości dzielnic dokonywanej przez nabywców wybierających mieszkania z różnych segmentów cenowych. Różnice te są najczęstsze, gdy porównujemy mieszkania z pierwszej i ostatniej grupy kwantylowej. Wyjątkiem jest tu ocena takich dzielnic, jak: Praga Północ i Praga Południe (gr3), Włochy i Ursus (gr5) oraz Wilanów. Wyceny tych dzielnic w obydwu modelach M1 i M4 nie są istotnie różne od siebie w poszczególnych segmentach cenowych rynku. Sugeruje to zgodność klientów co do oceny wartości tych dzielnic. Można wnioskować, że dzielnice te mają stałą i dość jednoznaczną ocenę. Wniosek ten może budzić zastrzeżenia niektórych ekspertów, gdyż obiegowe opinie praktyków rynku nieruchomości wskazują, że trudno o prostą i jednoznaczną ocenę dzielnic, takich jak Praga i Wilanów. Wyniki dla dzielnic: Bemowo i Bielany (gr1) oraz Białołęka i Targówek (gr2) są natomiast zbieżne z intuicją badacza rynku nieruchomości. Dzielnice te są istotnie różnie oceniane przez klientów wybierających mieszkania najtańsze i najdroższe. Na tej podstawie można przypuszczać, że klienci z większym ograniczeniem budżetowym chętniej wybierają „tańsze” lokalizacje, oceniając je wyżej niż klienci bardziej zamożni.

TABL. 3. WYNIKI ESTYMACJI REGRESJI MIĘDZYKWANTYLOWEJ Z BŁĘDAMI STANDARDOWYMI WYLICZONYMI PRZY ZASTOSOWANIU REPLIKACJI RESZT METODĄ BOOTSTRAP 500 POWTÓRZEŃ

Zmienne	M1				M4			
	Q25_Q50	Q50_Q75	Q25_Q75	Q25_Q50_Q75	Q25_Q50	Q50_Q75	Q25_Q75	Q25_Q50_Q75
	p-value							
ilekondygnacji	0,6074	0,3411	0,1960	0,4329	0,7904	0,2049	0,2751	0,4897
sq_ilekondygnacji	0,5750	0,4746	0,2614	0,5491	0,7959	0,3509	0,3809	0,6700
cub_ilekondygnacji	0,5746	0,5751	0,3431	0,6502	0,7305	0,5096	0,4536	0,7870
stand_1	0,5111	0,4795	0,2893	0,5715	0,0733	0,8380	0,0939	0,1535
stand_3	0,6739	0,0948	0,2959	0,2426	0,7546	0,0633	0,1908	0,1854
stand_4	0,4190	0,2609	0,7399	0,4989	0,3237	0,3557	0,1378	0,3380
rok_2	0,6286	0,8811	0,6261	0,7414	0,4280	0,9892	0,4948	0,7547
rok_3	0,9099	0,4310	0,6362	0,3265	0,2157	0,8799	0,2956	0,4535
rok_4	0,1587	0,6992	0,3619	0,0188	0,1884	0,9229	0,4049	0,4933
rok_5	0,0075	0,1788	0,3712	0,0087	0,2147	0,5051	0,1590	0,2840
rok_7	0,7138	0,4154	0,3954	0,6505	0,6558	0,4030	0,7750	0,6376
rok_8	0,2019	0,3639	0,9565	0,2977	0,7561	0,6250	0,8542	0,8572
rok_nowe	0,2435	0,4443	0,7388	0,4948	0,6138	0,8726	0,5911	0,8715
spoldzielcze	0,9149	0,4003	0,4618	0,6782	0,6502	0,8787	0,6214	0,8767
ilepokoi_1_2	0,8043	0,2842	0,3058	0,5240	0,7264	0,8152	0,9074	0,9286
sq_pow	0,8092	0,2132	0,5733	0,4556	0,5181	0,2528	0,1895	0,3929
kuch_3	0,2548	0,6995	0,4855	0,5196	0,1775	0,6508	0,4588	0,3827
gr1	0,0901	0,0250	0,0013	0,0067	0,1885	0,5186	0,0973	0,2318
gr2	0,0632	0,5806	0,0524	0,1239	0,0356	0,9852	0,0632	0,0947
gr3	0,2097	0,6234	0,5599	0,4725	0,3318	0,4579	0,8254	0,5514

gr4	0,4646	0,5088	0,2726	0,5637	0,0241	0,5509	0,0503	0,0903
gr5	0,2543	0,6959	0,6906	0,4632	0,0774	0,4738	0,4064	0,2592
mokotow	0,0893	0,0554	0,0020	0,0149	0,0149	0,7581	0,0170	0,0223
ochota	0,4597	0,0215	0,0246	0,0367	0,4827	0,7376	0,4210	0,6666
wola	0,3020	0,2583	0,0772	0,1775	0,4691	0,2652	0,7307	0,5186
ursynow	0,2235	0,0202	0,0024	0,0098	0,1334	0,3489	0,0423	0,1588
zoliborz	0,4507	0,0950	0,0287	0,1055	0,3601	0,3666	0,1118	0,2962
wilanow	0,2550	0,4505	0,1147	0,3182	0,5520	0,6419	0,9436	0,8007
t_20063	0,0605	0,6165	0,2069	0,1650				
t_20064	0,5894	0,1502	0,1071	0,2185				
t_20071	0,6202	0,9449	0,7530	0,8937				
t_20072	0,6325	0,0484	0,0569	0,0968				
t_20073	0,1026	0,7345	0,1385	0,2067				
t_20074	0,2909	0,5476	0,1507	0,3418				
t_20081	0,4816	0,4313	0,2247	0,4360				
t_20082	0,2244	0,0388	0,0075	0,0341				
t_20083	0,3971	0,4628	0,1908	0,4060				
t_20084	0,5939	0,1106	0,0912	0,1795				
t_20092	0,0623	0,1051	0,0078	0,0151				
t_20093	0,6097	0,5416	0,3692	0,6717				
t_20094	0,5160	0,3612	0,1964	0,4088				
t_20101	0,5712	0,5594	0,3603	0,6717				
t_20102	0,7494	0,2410	0,2324	0,4479				
Stała	0,0020	0,0003	0,0000	0,0000	0,0665	0,0795	0,0098	0,0377

U w a g a. Oznaczenia: „Q_25_Q_50” — regresja międzykwantylowa pomiędzy regresjami na kwantylach rzędu 0,25 i 0,50; pozostałe oznaczenia analogicznie.
Ź r ó d ł o: opracowanie własne.

Wyraźne zróżnicowanie cen implikowanych dzielnic Warszawy wzdłuż rozkładu ceny m^2 mieszkania może częściowo wynikać z dużej porcji informacji, jaką zawierają w sobie te zmienne. Szczegółowe położenie w obrębie dzielnicy może bowiem wiązać się z innymi ważnymi zmiennymi objaśniającymi nieuwzględnionymi w modelu.

Ceny implikowane pozostałych cech mieszkania: liczba kondygnacji budynku, rok budowy, prawo do lokalu, liczba pokoi, powierzchnia i rodzaj kuchni — jak wykazuje test oparty na regresji międzykwantylowej — nie różnią się istotnie w zależności od rozkładu cen metra kwadratowego mieszkania. Warto jednak zwrócić uwagę na interpretację niektórych zmiennych, wskazując jej znaczenie ekonomiczne.

Budynki wybudowane w latach 1945—1988 (zmienne rok_3, rok_4 i rok_5) w stosunku do budynków wybudowanych w latach 1989—2001 (rok_6)[2] są konsekwentnie niżej wyceniane, natomiast najnowsze budownictwo powstające od 2002 r. (rok_7 i rok_8) jest wyżej cenione niż mieszkania z lat 90. ub. wieku. Mieszkania zupełnie nowe, prawdopodobnie będące jeszcze w budowie, co do których odsprzedaje się właściwie prawo do zawarcia umowy ostatecznej z deweloperem (rok_nowe) wyceniane są niżej od mieszkań z lat 90. ub. wieku. Ujemna cena implikowana dla tej cechy mieszkań świadczy o właściwej ocenie ryzyka podejmowanego przy tego typu transakcjach.

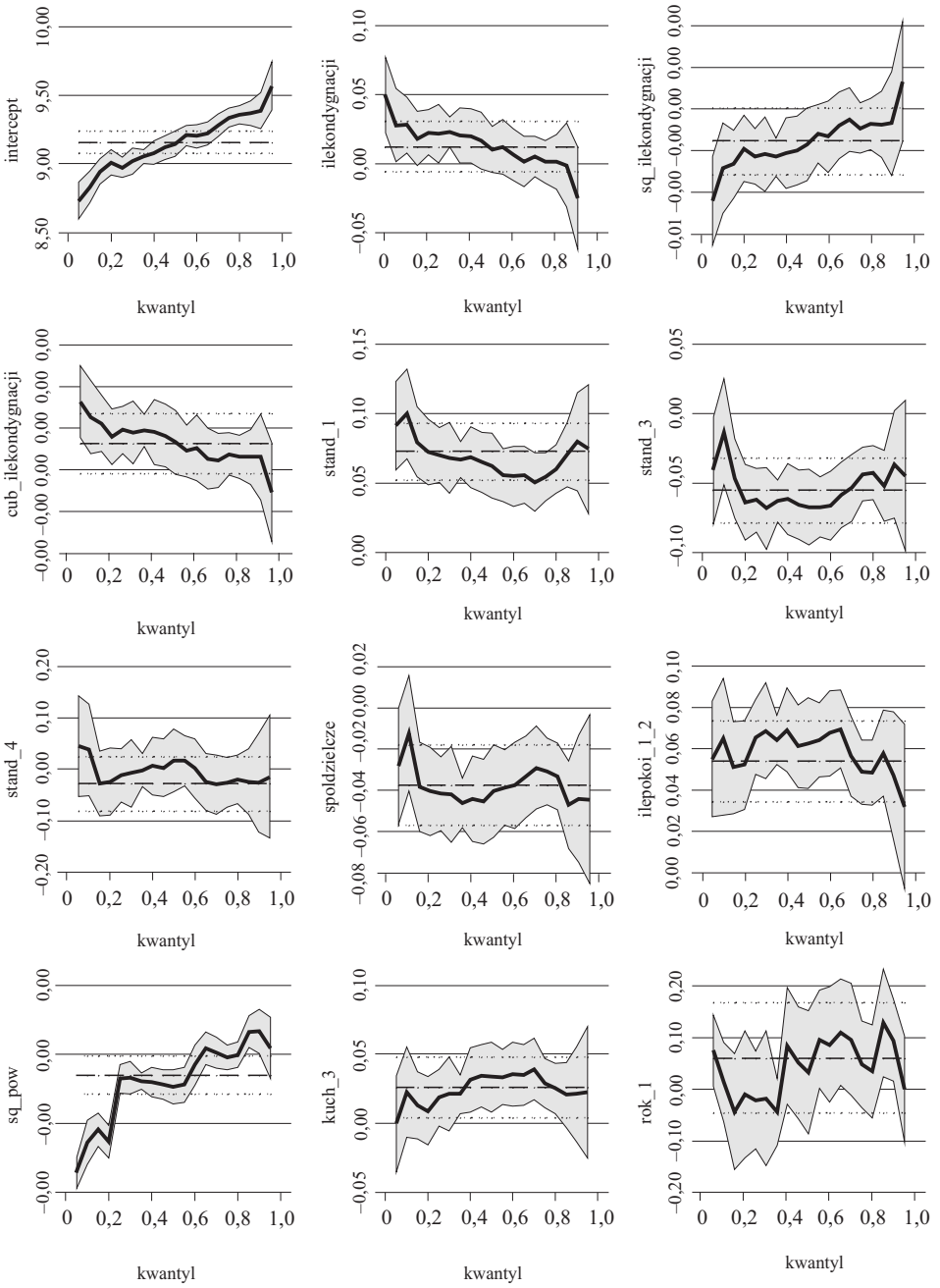
Z oszacowanych modeli wynika także, że mieszkania spółdzielcze są o ok. 4% niżej wyceniane niż mieszkania własnościowe. Cecha ta jest wyraźnie istotna w każdym segmencie cenowym modelu M1. Szacując wpływ wielkości lokalu na cenę jego metra kwadratowego można zauważyć, że istotne znaczenie bez względu na segment cenowy ma liczba pokoi, a dokładniej rozróżnienie mieszkań jedno- i dwupokojowych od pozostałych.

Na wykr. 1 dodatkowo przedstawiono rozkład wartości współczynników regresji (wraz z ich 95% przedziałem ufności oznaczonym kolorem szarym) w poszczególnych decylach rozkładu ceny m^2 mieszkania. Linie przerywane są równoważne wartości współczynników/cen implikowanych oszacowanych za pomocą estymatora MNK wraz z jej 95% przedziałem ufności.

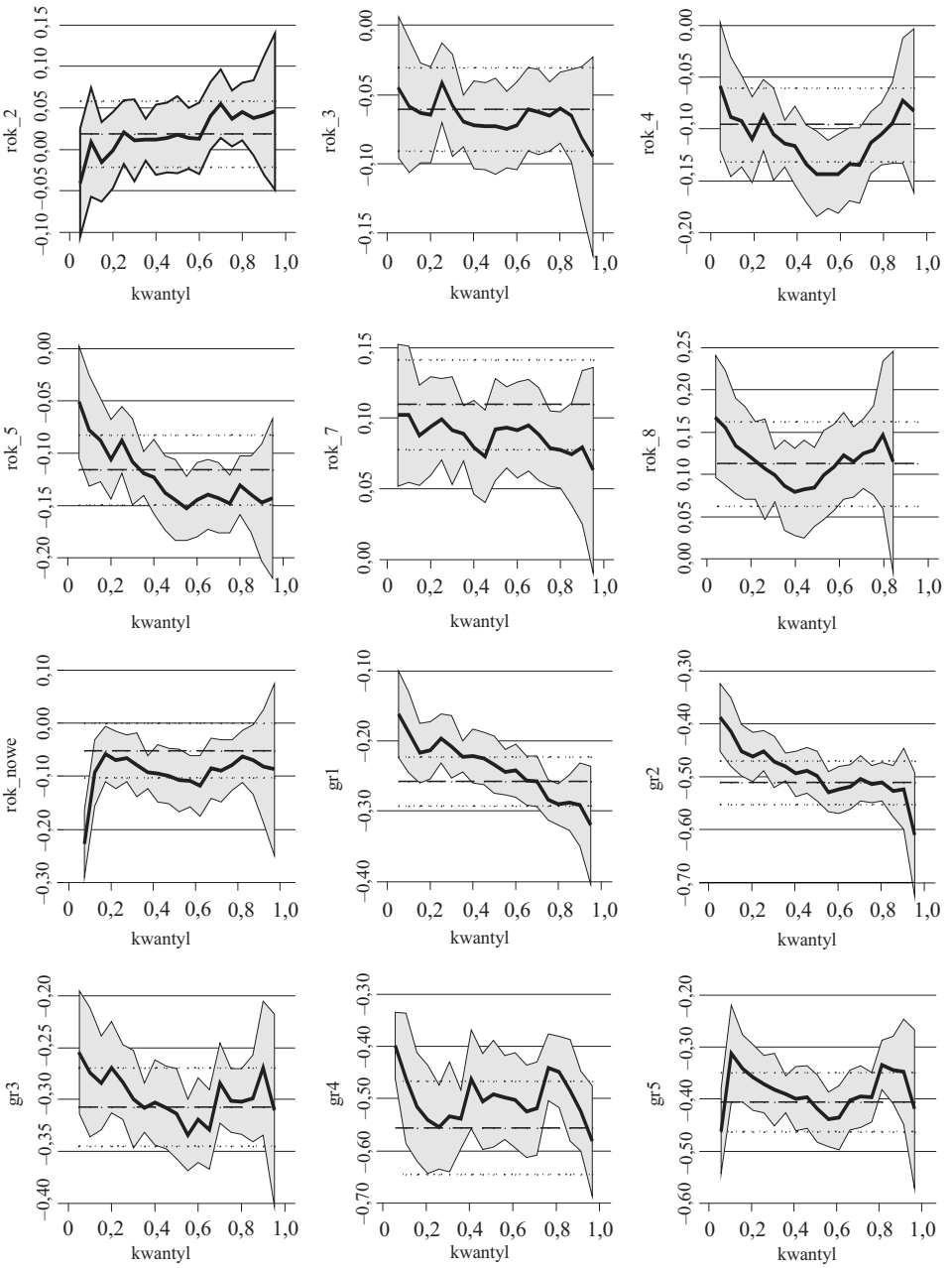
---

[2] Często w badaniach ekonomicznych mamy do czynienia ze zmiennymi jakościowymi, które przyjmują określone stany jakościowe, zwane kategoriami, i których wartości nie mogą być przedstawione za pomocą zbioru liczb rzeczywistych. Aby wprowadzić do modelu zmienną jakościową o *m* kategoriach, tworzy się zestaw zmiennych zero-jedynkowych odpowiedzialnych za wszystkie kategorie zmiennej jakościowej (tworzy się *m* zmiennych zero-jedynkowych). Do modelu dołącza się *m –1* tych zmiennych, zaś ta kategoria wyjściowej zmiennej, dla której zmienna zero-jedynkowa nie została dołączona do modelu, stanowić będzie kategorię referencyjną dla pozostałych. Jeśli jednak zostaną wprowadzone wszystkie zmienne zero-jedynkowe odpowiedzialne za wszystkie kategorie zmiennej jakościowej do modelu, to macierz zmiennych objaśniających **X** miałaby rząd niższy od rzędu kolumnowego. Uniemożliwiłoby to odwrócenie macierzy **X'X** przy wyznaczaniu estymatorów MNK. Interpretacja parametrów dla zmiennych zero-jedynkowych wprowadzonych do modelu jest interpretacją względem kategorii referencyjnej (kategorii niewprowadzonej do modelu w postaci zmiennej zero-jedynkowej).

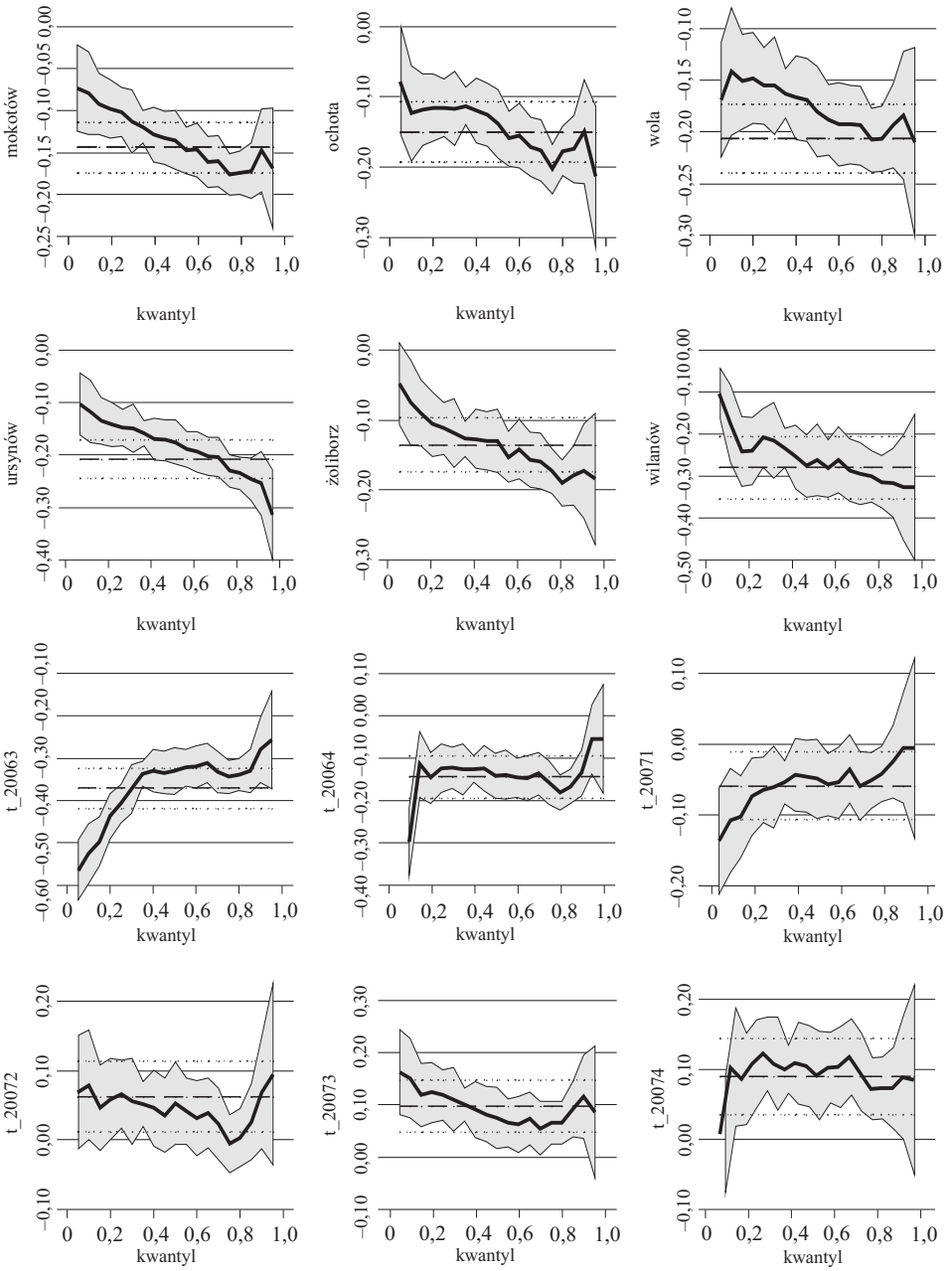
Wykr. 1. OSZACOWANE WIELKOŚCI WSPÓŁCZYNNIKÓW DLA CECH MIESZKANIOWYCH W POSZCZEGÓLNYCH DECYLACH ROZKŁADU CENY METRA KWADRATOWEGO MIESZKANIA (M4) ZA POMOCĄ REGRESJI KWANTYLOWEJ ORAZ MNK WRAZ Z 95% PRZEDZIAŁAMI ICH UFNOŚCI



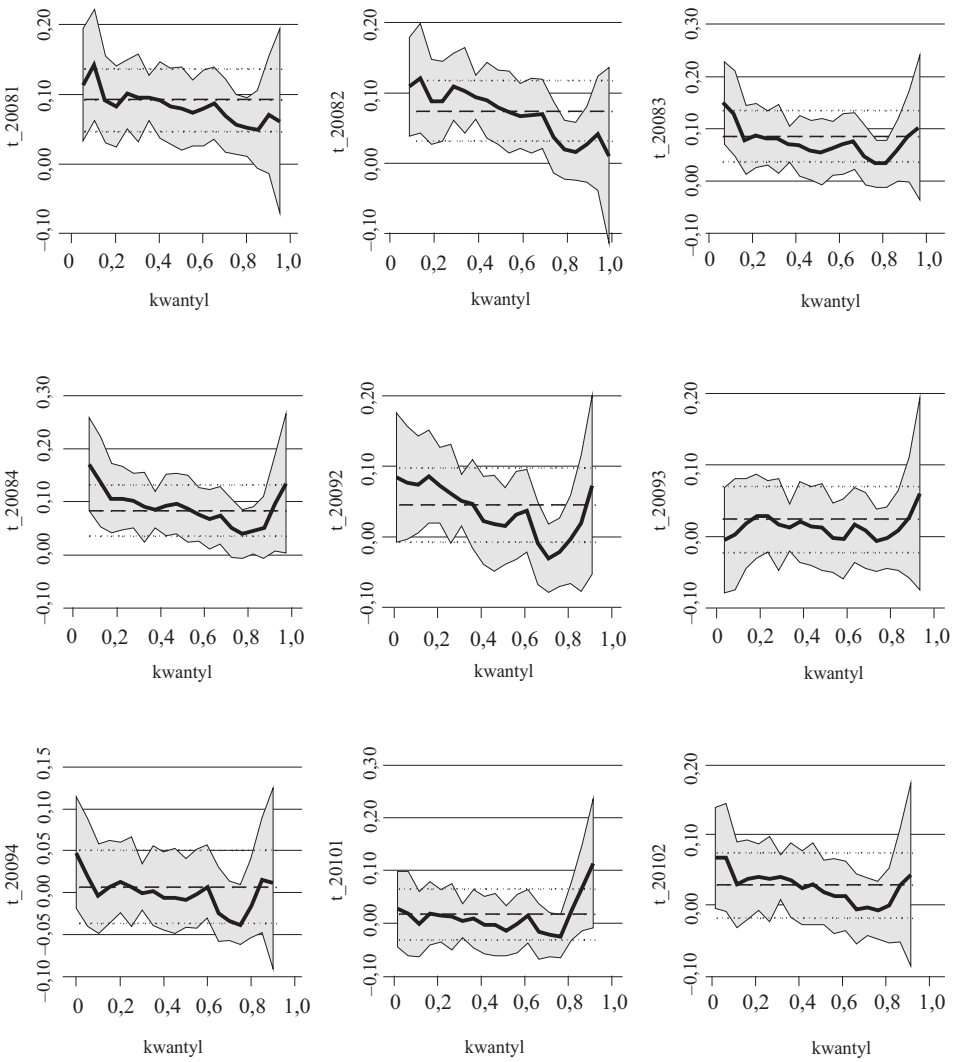
Wykr. 1. OSZACOWANE WIELKOŚCI WSPÓŁCZYNNIKÓW DLA CECH MIESZKANIOWYCH W POSZCZEGÓLNYCH DECYLACH ROZKŁADU CENY METRA KWADRATOWEGO MIESZKANIA (M4) ZA POMOCĄ REGRESJI KWANTYLOWEJ ORAZ MNK WRAZ Z 95% PRZEDZIAŁAMI ICH UFNOŚCI (cd.)



Wykr. 1. OSZACOWANE WIELKOŚCI WSPÓŁCZYNNIKÓW DLA CECH MIESZKANIOWYCH W POSZCZEGÓLNYCH DECYLACH ROZKŁADU CENY METRA KWADRATOWEGO MIESZKANIA (M4) ZA POMOCĄ REGRESJI KWANTYLOWEJ ORAZ MNK WRAZ Z 95% PRZEDZIAŁAMI ICH UFNOŚCI (cd.)



Wykr. 1. OSZACOWANE WIELKOŚCI WSPÓŁCZYNNIKÓW DLA CECH MIESZKANIOWYCH W POSZCZEGÓLNYCH DECYLACH ROZKŁADU CENY METRA KWADRATOWEGO MIESZKANIA (M4) ZA POMOCĄ REGRESJI KWANTYLOWEJ ORAZ MNK WRAZ Z 95% PRZEDZIAŁAMI ICH UFNOŚCI (dok.)



U w a g a. Pogrubiona czarna linia ciągła przedstawia oszacowanie współczynników za pomocą regresji na kwantylach (wraz z ich 95% przedziałem ufności oznaczonym kolorem szarym). Linia przerywana przedstawia oszacowanie za pomocą MNK wraz z jej 95% przedziałem ufności. Dla zmiennych: sq ilekondygnacji, cub ilekondygnacji oraz sq pow — oszacowany zakres wartości przedstawiony na wykresie mieści się w przedziale zbliżonym do zera.

Ź r ó d ł o: opracowanie własne.

W zasadzie dla większości zmiennych objaśniających szacunki oparte na metodzie regresji kwantylowej i MNK nie różnią się wzdłuż rozkładu cen m^2 mieszkania. Wyjątki stanowią: stała (można ją interpretować jako oszacowany warunkowy kwantyl rozkładu ceny mieszkania przy zmiennych objaśniających równych zero) oraz sq_pow, gr1, gr2, mokotow, ursynow i t_20063. Biorąc pod uwagę to spostrzeżenie oraz dokładną analizę wyników można uznać, że w badanym przypadku nie ma widocznej różnicy między wartością estymatora MNK i regresji kwantylowej.

Indeksy hedoniczne

Oszacowane metodą regresji kwantylowej modele hedoniczne cen mieszkań wykorzystano do wyznaczenia hedonicznych indeksów cen. Wykresy 2—13 pokazują przebieg dynamiki cen mieszkań sprzedanych na warszawskim rynku wtórnym wyrażonej w ujęciu kwartał do kwartału (Q/Q) oraz ze stałą podstawą (III kwartał 2006 r.=100). Pokazano także indeks hedoniczny wyznaczony MNK oraz wskaźnik zmiany zwykłej średniej ceny mieszkania, który jest powszechnie wykorzystywaną miarą monitorowania tendencji cenowych.

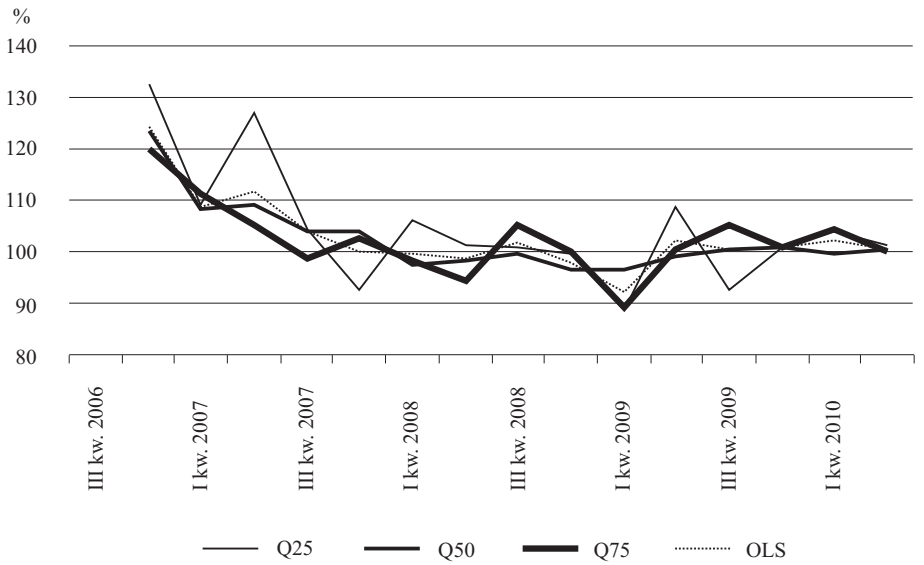
W przypadku indeksu wyznaczonego metodą imputacji (M4) analiza różnic dynamiki cen w zależności od segmentu cenowego może być dokonana wyłącznie na podstawie porównania graficznego. Dla metody ze zmiennymi 0—1 czasu wszystkich okresów (M1) przeprowadzono formalne testy wykorzystując wyniki regresji międzykwantylowej. Formalne wnioskowanie pokazuje, że tylko w 5 na 15 badanych okresów dynamika cen istotnie różniła się między pierwszą i trzecią grupą cenową (przy poziomie istotności 10%). Różnice między pozostałymi segmentami w zdecydowanej większości badanych kwartałów nie są istotne statystycznie. Niemniej jednak warto podkreślić, że wyniki z modelu można uznać za zbieżne z obserwacją rynku. Jak już wspomniano, wraz z pojawiającą się stagnacją w 2008 r. na rynku mieszkań wyraźniej zarysowały się różnice zmian cen w poszczególnych segmentach rynku. Testy formalne potwierdzają w pewnym zakresie tę obserwację, przypisując istotność różnic dynamiki między grupą pierwszą i trzecią w II i IV kw. 2008 r. oraz w II kw. 2009 r.

Graficzna analiza wartości indeksu M4 (wykr. 2 i 4) uwidacznia różnice między dynamiką cen najtańszych i najdroższych mieszkań. Są one szczególnie widoczne, gdy weźmiemy pod uwagę indeks o stałej podstawie, gdzie różnice z poszczególnych kwartałów kumulują się.

Różnice wartości indeksów hedonicznych, zarówno M1 i M4, występują przede wszystkim w pierwszej i trzeciej grupie cenowej (wykresy 11 i 13). Do dokładniejszego zatem badania tendencji cenowych i rozróżnienia zmian cen w poszczególnych segmentach cenowych rynku warto stosować modele hedoniczne, szacowane metodą regresji kwantylowej.

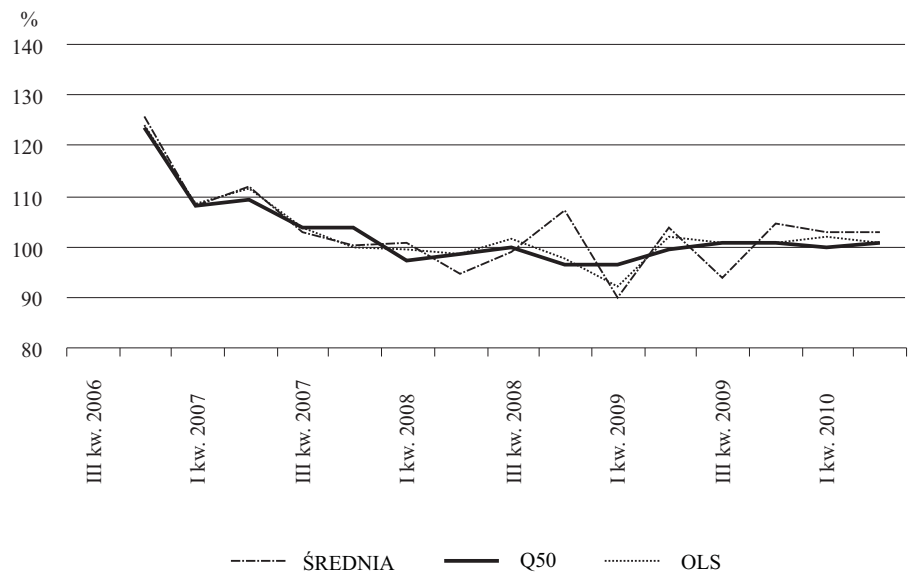
Także porównanie indeksu dla segmentu środkowego ze wskaźnikiem opartym na modelu szacowanym MNK oraz ze wskaźnikiem zwykłej średniej (wykr. 3, 5, 7, 9, 12) daje ciekawe wnioski. Po pierwsze, indeksy hedoniczne są

Wykr. 2. DYNAMIKA Q/Q. HEDONICZNY INDEKS IMPUTACJI (M4)



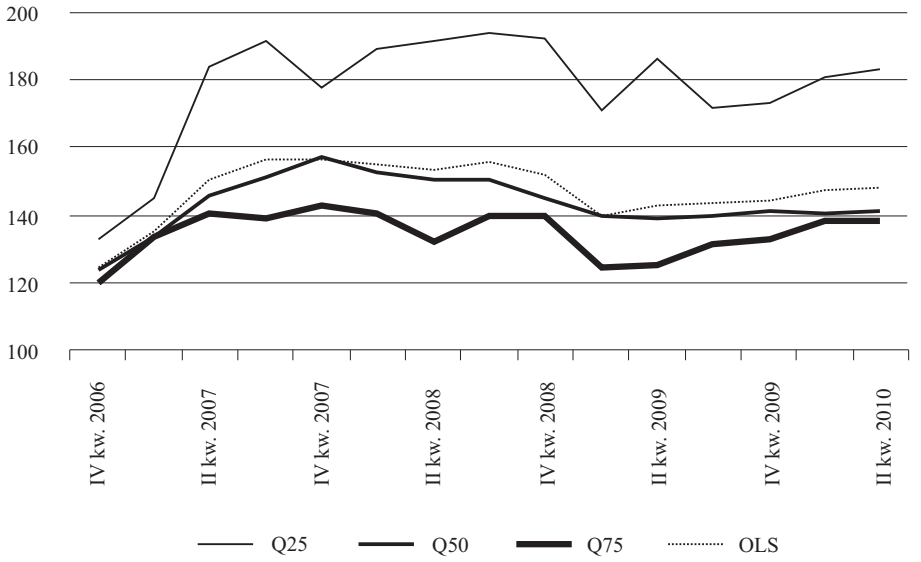
Ź r ó d ł o: opracowanie własne.

Wykr. 3. DYNAMIKA Q/Q. HEDONICZNY INDEKS IMPUTACJI (M4) (versus średnia)



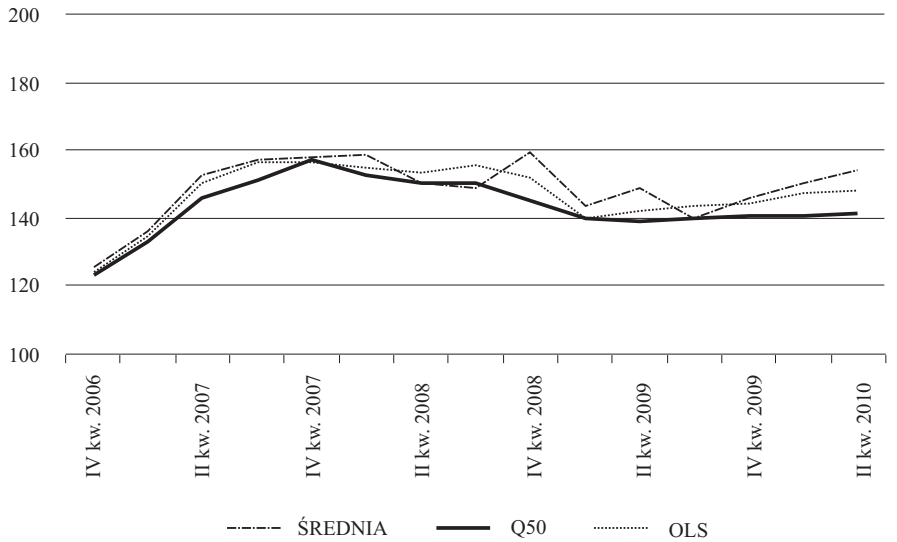
Ź r ó d ł o: opracowanie własne.

Wykr. 4. DYNAMIKA III KW. 2006=100. HEDONICZNY INDEKS IMPUTACJI (M4)



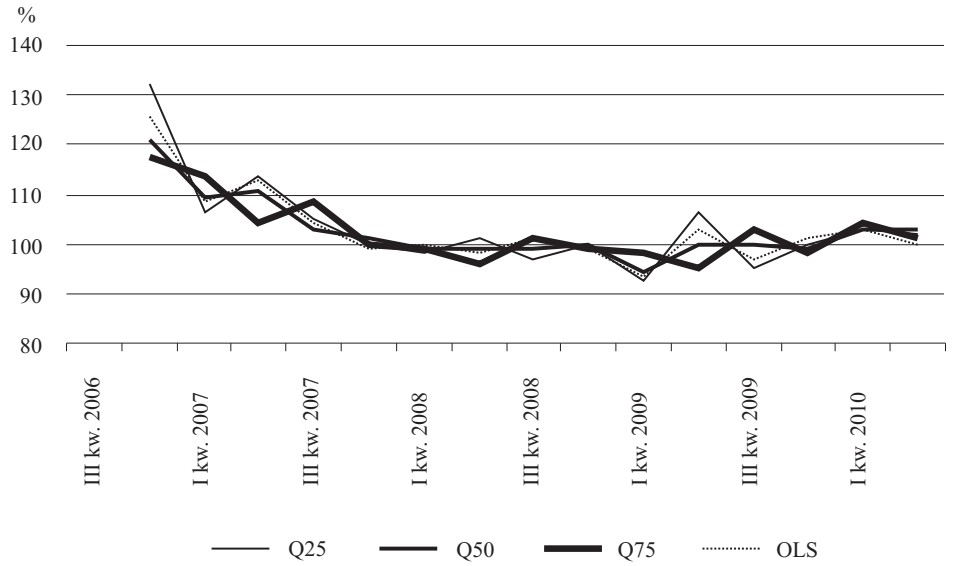
Ź r ó d ł o: opracowanie własne.

Wykr. 5. DYNAMIKA III KW. 2006=100. HEDONICZNY INDEKS IMPUTACJI (M4) (versus średnia)



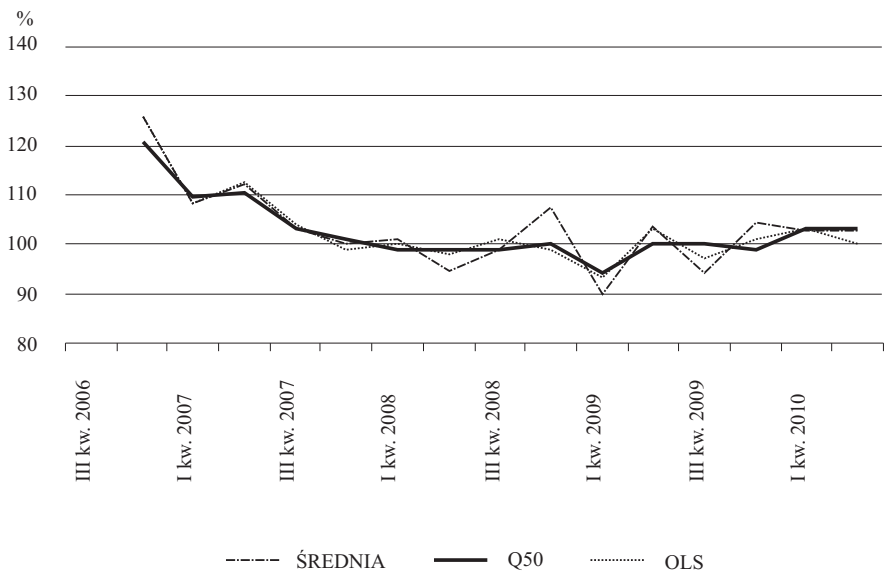
Ź r ó d ł o: opracowanie własne.

Wykr. 6. DYNAMIKA Q/Q. HEDONICZNY INDEKS 0—1 CZASU (M1)



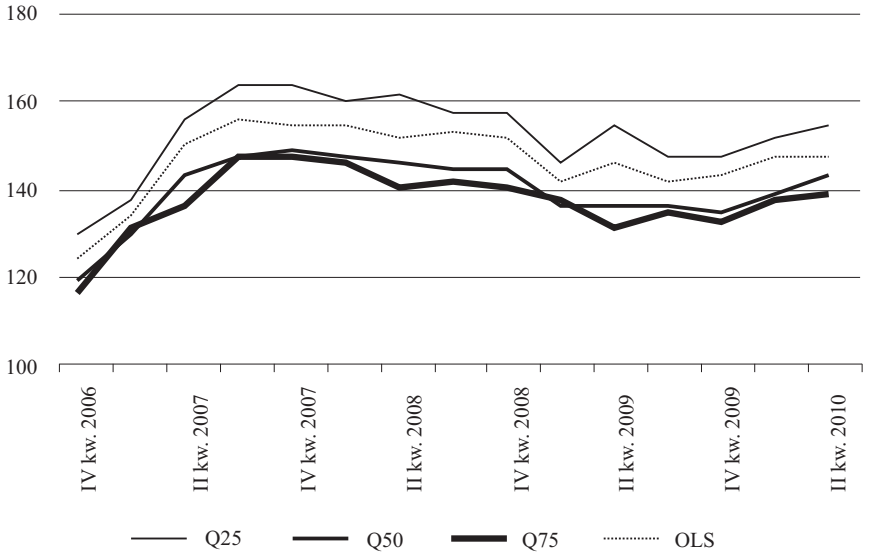
Ź r ó d ł o: opracowanie własne.

Wykr. 7. DYNAMIKA Q/Q. HEDONICZNY INDEKS 0—1 CZASU (M1) (versus średnia)



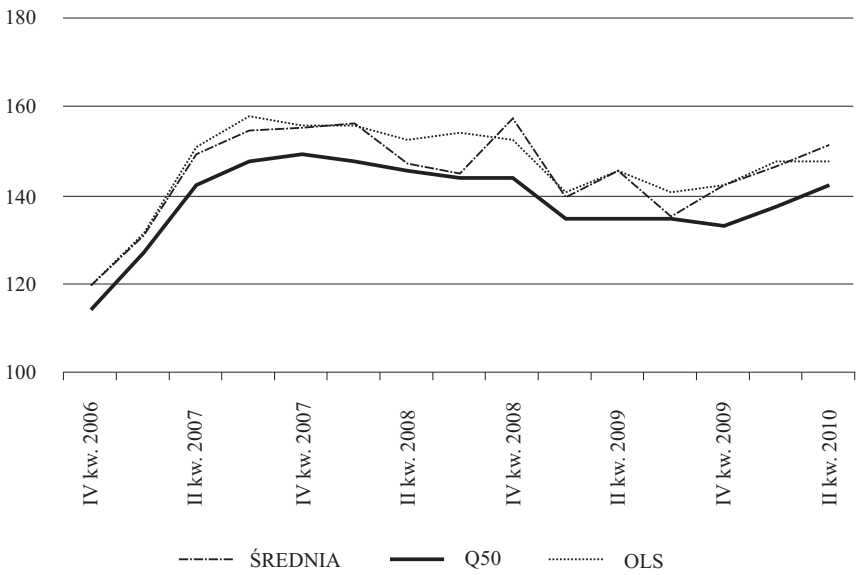
Ź r ó d ł o: opracowanie własne.

Wykr. 8. DYNAMIKA III KW. 2006=100. HEDONICZNY INDEKS 0—1 CZASU (M1)



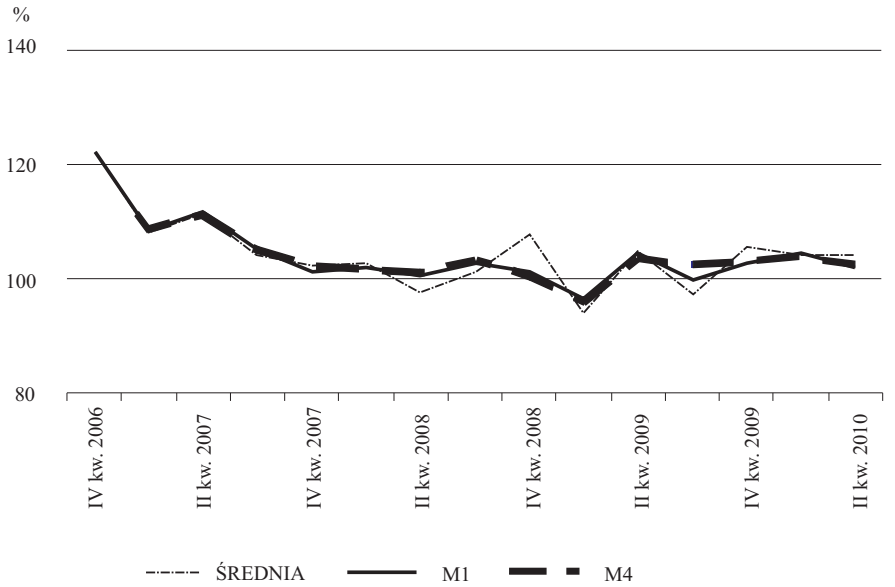
Ź r ó d ł o: opracowanie własne.

Wykr. 9. DYNAMIKA III KW. 2006=100. HEDONICZNY INDEKS 0—1 CZASU (M1) (versus średnia)



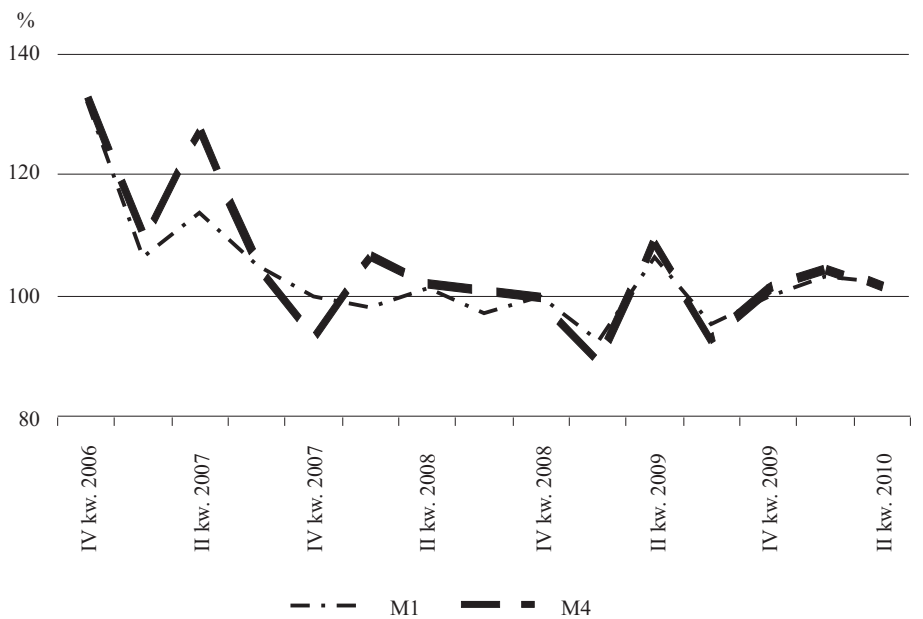
Ź r ó d ł o: opracowanie własne.

Wykr. 10. DYNAMIKA Q/Q. OLS



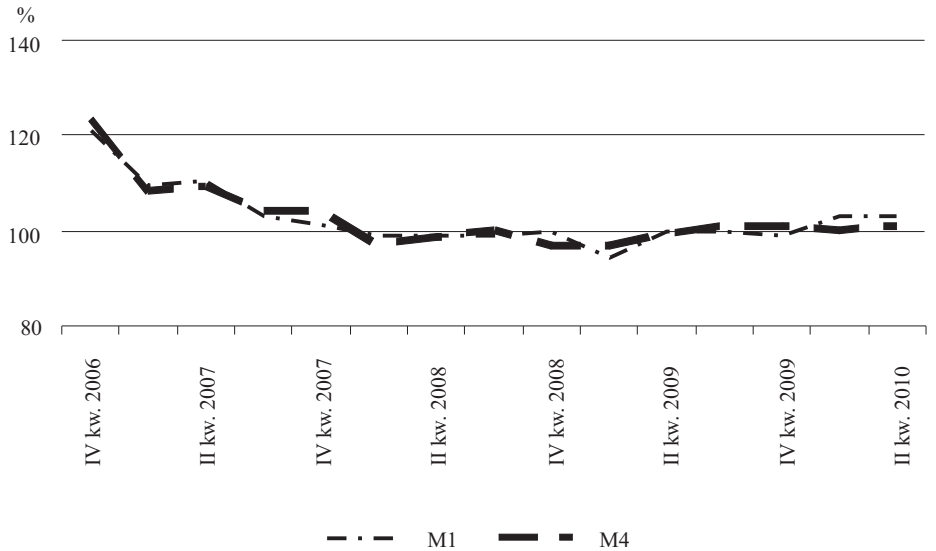
Ź r ó d ł o: opracowanie własne.

Wykr. 11. DYNAMIKA Q/Q. Q25



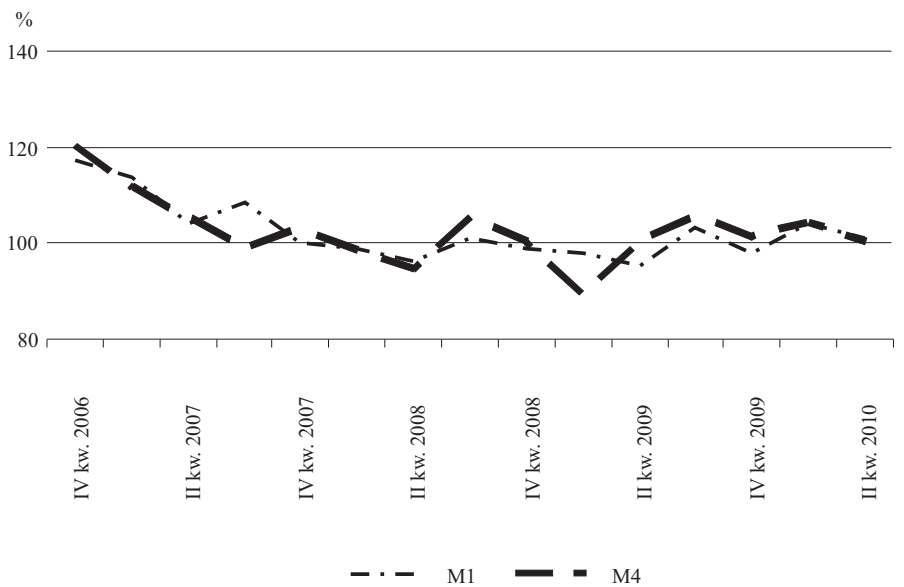
Ź r ó d ł o: opracowanie własne.

Wykr. 12. DYNAMIKA Q/Q. Q50



Ź r ó d ł o: opracowanie własne.

Wykr. 13. DYNAMIKA Q/Q. Q75



Ź r ó d ł o: opracowanie własne.

mniej zmienne niż wskaźnik dynamiki średniej ceny mieszkania. Sugeruje to, że w rzeczywistości jest to miara uwzględniająca różnice jakości badanych dóbr, dzięki czemu staje się bardziej wiarygodną miarą średniej tendencji cen na rynku. Po drugie, indeks wyznaczony metodą regresji kwantylowej dla środkowego segmentu cen jest mniej zmienny niż indeks oparty na estymatorze MNK. Obserwacja ta potwierdza wcześniej wspomniane wnioski dotyczące metody, wskazującej że w przypadku występowania heteroskedastyczności estymacja regresji kwantylowej rzędu 0,5 może okazać się bardziej efektywnym sposobem szukania wartości parametrów niż tradycyjna regresja oparta na wartości oczekiwanej zmiennej objaśnianej. Po trzecie, indeks środkowego segmentu ma właściwie zbliżoną wartość przy obydwu stosowanych metodach wyznaczania go (M1 czy M4). Te spostrzeżenia prowadzą do wniosku, że indeks środkowego segmentu cenowego spośród trzech porównywanych miar może być najdokładniejszym przybliżeniem średniej tendencji cenowej na rynku. Warto jednocześnie zaznaczyć, że wyznaczenie go z wykorzystaniem metody regresji kwantylowej jest proste i umożliwia wykorzystanie pełnego zbioru obserwacji, także tych odstających.

Podsumowanie

W artykule przedstawiono wyniki estymacji hedonicznych modeli cen nieruchomości mieszkaniowych za pomocą metody opartej na kwantylach oraz wyniki konstrukcji hedonicznych indeksów cen mieszkań. Badanie zostało przeprowadzone dla warszawskiego rynku wtórnego mieszkań na podstawie danych transakcyjnych.

Zastosowanie regresji kwantylowej do szacowania hedonicznych cen mieszkań ma tę przewagę, że pozwala uwzględniać w estymacji obserwacje odstające, ułatwia wykrycie występowania zjawiska heteroskedastyczności składnika losowego oraz nie wymaga spełnienia założenia o charakterze rozkładu zmiennej objaśnianej. Praktyczną zaletą stosowania metody regresji kwantylowej jest to, że pozwala ona oszacować model hedoniczny dla wybranego segmentu cenowego mieszkań wykorzystując jednocześnie pełen zbiór danych. Jest to szczególnie istotne w przypadku wykorzystania tych modeli do wyznaczania hedonicznego indeksu cen. Wówczas nie tracąc informacji płynącej z pełnego zbioru danych w próbie można oszacować indeks dla wybranego segmentu cenowego, np. środkowego, będącego w większości przypadków najlepszą reprezentacją mieszkań popularnych (stanowiących największą część rynku). Jak widać, indeks ten charakteryzuje się mniejszą zmiennością niż indeks z MNK, co może świadczyć o jego większej dokładności przy ocenie przeciętnej tendencji zmiany cen mieszkań reprezentatywnych dla rynku. Obserwacja ta potwierdza, że estymator LAD jest bardziej efektywną i właściwą miarą niż estymator MNK w modelu z heteroskedastyczością.

Regresja kwantylowa, w przeciwieństwie do MNK, wyznaczając oszacowania zmiennych objaśniających dla danej grupy kwantylowej nadaje mniejsze wagi obserwacjom z pozostałych grup — segmentów rynku. Zmienność tego indeksu jest mniejsza w porównaniu do zmienności indeksu wyznaczonego z wykorzystaniem estymatora MNK. Indeks segmentu środkowego liczonego metodą regresji kwantylowej jest zatem lepszą reprezentacją mieszkań przeciętnych, chociaż nie pomija zupełnie segmentów skrajnych.

Analiza pokazuje, że w badanym okresie dla warszawskiego rynku wtórnego mieszkań tylko dwie cechy — standard wykończenia oraz lokalizacja — były istotnie różnie wyceniane w zależności od badanego segmentu cenowego. Pozostałe cechy mieszkań były tak samo wyceniane w różnych segmentach cenowych. Wnioski te nie potwierdzają jednoznacznie hipotezy stawianej przez niektórych ekonomistów, że klienci rozmaicie wyceniają cechy mieszkania, a różnice te zależą od rozkładu cen mieszkań (Zietz i in., 2008; Malpezzi, 2002; Rosen, 1974).

Warto zwrócić uwagę, że transformacyjny wciąż charakter polskiego rynku nieruchomości ma istotny wpływ na wyniki tego typu badań, w szczególności na empiryczną weryfikację modeli teoretycznych, w tym wypadku hedonicznego modelu cen nieruchomości mieszkań.

---

mgr Marta Widlak — *NBP*, **dr Natalia Nehrebecka** — *NBP* i *UW*

LITERATURA

Bailey M., Muth R., Nourse H. (1963), *A Regression Method for Real Estate Price Index Construction*, „Journal of the American Statistical Association", vol. 58, No. 304

Bayer P., McMillan R., Rueben K. (2004), *An equlibrium model of sorting in an urban housing market*, NBER Working Paper, No. 10865

Epple D. (1987), *Hedonic prices and implicit markets: estimating demand and supply functions for differentiated products*, „The Journal of Political Economy", vol. 95, No. 1

Gould W. W. (1992), *Quantile Regression with Bootstrapped Standard Errors*, „Stata Technical Bulletin", No. 9

Gould W. W. (1997), *Interquantile and Simultaneous-Quantile Regression*, „Stata Technical Bulletin", No. 38

Greene W. H. (2000), *Econometric Analysis*, 4th ed., Prentice Hall, Saddle River, N. J.

Koenker R. (2000), *Galton, Edgeworth, Frisch, and prospects for quantile regression in econometrics*, „Journal of Econometrics", No. 95

Koenker R. (2005), *Quantile Regression*, Cambridge Books, Cambridge University Press, number 9780521608275

Koenker R., Bassett G. (1978), *Regression Quantiles*, „Econometrica", No. 46

Koenker R., Hallock K. F. (2001), *Quantile Regression*, „Journal of Economic Perspectives", No. 5

Lancaster K. J. (1966), *A new approach to consumer theory*, „The Journal of Political Economy", No. 74 (2)

Liao W., Wang X. (2010), *Hedonic house prices and spatial quantile regression*, Institute of Real Estate Studies Working Paper Series, Singapore

Malpezzi S. (2002), *Hedonic pricing models: a selective and applied review*, The Center for Urban Land Economics Research, The University of Wisconsin

McMillen D. P., Thorsnes P. (2006), *Housing renovations and the quantile repeat sales price index*, „Real Estate Economics", No. 34 (4)

Narula S., Wellington J. (1982), *The Minimum Sum of Absolute Errors Regression: A State of the Art Survey*, „International Statistics Review", No. 50

Rogers W. H. (1993), *Calculation of Quantile Regression Standard Errors*, „Stata Technical Bulletin", No. 13

Rosen S. (1974), *Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition*, „The Journal of Political Economy", vol. 82, No. 1

Tomczyk E., Widłak M. (2010), *Konstrukcja i własności hedonicznego indeksu cen mieszkań dla Warszawy*, Bank i Kredyt nr 1 (2010)

Widłak M. (2010), *Dostosowanie indeksów cenowych do zmian jakości. Metody wyznaczania hedonicznych indeksów cen i możliwość ich zastosowania dla rynku mieszkaniowego*, „Materiały i Studia", nr 247

Zietz E. N., Zietz J., Sirmans S. G. (2008), *Determinants of house prices: a quantile regression approach*, „Journal of Real Estate Finance and Economics", No. 37

SUMMARY

The aim of the article is to indicate theoretical and practical advantages of using quantile regression as well as a sensitivity analysis of factors influencing a dwelling price (taking into account the regression distribution). This is a new approach in comparison to previous surveys on carried in Polish market which analyzed the factors' impact on average price level. Additionally, surveys are described which make possible to estimate dynamic price differences in selected price segments using hedonic indices.

РЕЗЮМЕ

Целью статьи является представление теоретических и практических преимуществ использования метода квантильной регрессии, а благодаря ее использованию также анализ чуткости реакции факторов влияющих на цену квартиры с учетом ее распределения. Это новый подход по сравнению с предыдущими обследованиями на польском рынке, которые анализировали влияние этих факторов на средний уровень цен. Дополнительно характеризуются обследования позволяющие оценить различия динамики цен в определенных ценовых сегментах при использовании гедонистических индексов.