

Konstrukcja i własności hedonicznego indeksu cen mieszkań dla Warszawy

Emilia Tomczyk*, Marta Widlak[#]

Nadesłany: 22 września 2009 r. Zaakceptowany: 21 stycznia 2010 r.

Streszczenie

Wiarygodna ocena tendencji cenowych na rynkach mieszkaniowych jest w polskiej statystyce zagadnieniem nowym i niezbadanym na gruncie empirycznym, podobnie jak problem korekty zmian jakości w indeksach cenowych oraz zastosowanie w tym celu modeli hedonicznych. Celem artykułu jest prezentacja wyników estymacji hedonicznych modeli cen oraz indeksów cen mieszkań skonstruowanych na podstawie danych na temat transakcji zawieranych na warszawskim rynku wtórnym. Stosując trzy bezpośrednie metody konstrukcji hedonicznego indeksu cen, przedstawiamy pięć różnych wskaźników hedonicznych, które w odróżnieniu od powszechnie stosowanych metod pomiaru dynamiki cen mieszkań – średniej i mediany – pozwalają uwzględnić zmiany i zróżnicowanie jakości mieszkań. Nakreślamy także zagadnienia teoretyczne związane z estymacją i interpretacją modeli hedonicznych oraz ich wykorzystaniem jako narzędzi korekty zmian jakości we wskaźnikach cen mieszkań. Przedstawione w artykule teoretyczne i praktyczne aspekty zastosowań hedonicznych indeksów cen stanowią obiecujący kierunek rozwoju polskiej statystyki cen nieruchomości.

Słowa kluczowe: indeksy cenowe, modele hedoniczne, rynek mieszkaniowy

JEL: C21, C43, R21, R31

* Szkoła Główna Handlowa w Warszawie, Instytut Ekonometrii; e-mail: emilia.tomczyk@sgh.waw.pl.

[#] Narodowy Bank Polski, Instytut Ekonomiczny; e-mail: marta.widlak@nbp.pl.

1. Wstęp

Celem artykułu jest prezentacja konstrukcji hedonicznego¹ indeksu cen mieszkań oraz praktycznych wniosków płynących z badań empirycznych nad wyznaczeniem takiego wskaźnika dla warszawskiego wtórnego rynku mieszkaniowego. Hedoniczne modele cen są narzędziem dostosowania indeksów cen do zmian jakości badanych dóbr; przedstawienie zastosowania tych modeli w przypadku indeksów cen mieszkań jest jednym z tematów poruszanych w tym artykule. Nakreśliśmy także zagadnienia teoretyczne związane z estymacją i interpretacją modeli hedonicznych. Podejmujemy ten temat ze względu na brak tego typu opracowań w Polsce, gdzie jak dotąd nie powstał wiarygodny i powszechnie dostępny wskaźnik zmian cen na rynku nieruchomości. Z jednej strony jest to spowodowane brakiem odpowiednich źródeł danych, z drugiej – trudnościami z opracowaniem odpowiedniej metodyki wyznaczania indeksu cen.

W drugiej części artykułu krótko przedstawiamy ideę tworzenia modeli hedonicznych oraz główne nurty ich zastosowań, w tym na rynkach nieruchomości mieszkaniowych. Część trzecia prezentuje bezpośrednie metody wyznaczania indeksów cen zastosowane w analizie empirycznej. Źródło danych oraz same dane, z których korzystamy w części empirycznej, opisane są w czwartej części pracy. Część piąta przedstawia ekonometryczne aspekty szacowania hedonicznych modeli cen nieruchomości. W kolejnej części prezentujemy wyniki estymacji i weryfikacji hedonicznych modeli cen mieszkań na rynku wtórnym w Warszawie oraz porównujemy skonstruowane na ich podstawie indeksy cen. Część siódma zawiera podsumowanie wyników i propozycje kierunków dalszych badań.

2. Hedoniczne indeksy cen mieszkań

Koncepcja hedonicznego modelu cen opiera się na założeniu, że heterogeniczne dobra można przedstawić jako agregat ich cech (charakterystyk). Z punktu widzenia konsumenta każdej z tych charakterystyk przypisana jest pewna użyteczność, a cena produktu jest sumą (nieobserwowalnych) częściowych użyteczności poszczególnych cech, jako że same cechy nie są przedmiotem transakcji. Bardziej formalnie, modele hedoniczne przybierają postać modeli ekonometrycznych (najczęściej jednorównaniowych i nieliniowych względem zmiennych), w których zmienną objaśnianą jest cena, a zmiennymi objaśniającymi – charakterystyki produktu, o których sądzi się, że mają znaczny wpływ na jego cenę. Cena heterogenicznego dobra jest zatem sumą wycen jego poszczególnych charakterystyk opisanych za pomocą zmiennych objaśniających oraz czynników odzwierciedlonych w składniku losowym (por. część trzecia). Główne zagadnienia ekonometryczne, w tym dobór zmiennych objaśniających i postaci funkcyjnej modelu regresji oraz własności składnika losowego, omawiamy w piątej części opracowania.

¹ W języku angielskim rozróżnia się terminy *hedonics* i *hedonism*, stosowane w różnych kontekstach znaczeniowych. W naszej pracy zdecydowaliśmy się tłumaczyć *hedonic regression* jako „regresja hedoniczna”. Określenie „hedonistyczna” (regresja, model, indeks) w języku polskim kojarzy się z pojęciem hedonizmu i rozważanym w literaturze filozoficznej pojęciem użyteczności. Jak zaznaczamy w części piątej, cytując S. Rosena, interpretacja współczynników regresji hedonicznych jako krańcowej użyteczności konsumenta nie jest uprawniona. Ponadto określenie „regresja hedoniczna” pojawia się też w innych opracowaniach, np. Dziechciarz (2005), Trojanek (2008).

Za pionierów hedonicznej analizy cen dóbr uważa się F. Waugha, który w latach 1928–1929 badał wpływ wielkości, kształtu, koloru i dojrzałości warzyw na ich ceny, oraz A. Courta, autora opublikowanej w 1939 r. pracy na temat zależności cen samochodów od takich cech, jak rodzaj silnika, waga samochodu czy powierzchnia okien (por. Berndt 1991). Rozwój formalnych metod analizy hedonicznej związany jest z pracami Z. Grilichesa i S. Rosena, publikowanymi w latach 60. i 70. XX w. W przypadku rynku nieruchomości hedoniczne modele cen stosuje się do oszacowania, jaki wpływ na wycenę nieruchomości mają jej cechy oraz dane zjawiska ekonomiczno-społeczne². Te ciekawe projekty badawcze są tworzone głównie na potrzeby polityki planowania przestrzennego i inwestycyjnej, szeroko pojętej polityki społecznej, a także w projektach tzw. masowej wyceny nieruchomości, związanej najczęściej z systemami katastralnymi. Drugim powszechnym zastosowaniem hedonicznych modeli cen (szczególnie w przypadku rynku nieruchomości) jest szacowanie dostosowań zmian jakości nowych i „znikających” (nieдоступnych w pewnych okresach, np. wycofanych z rynku) dóbr w konstrukcji wskaźników cen. Wątek ten pojawił się w literaturze dotyczącej konstruowania wskaźników cen dóbr i usług konsumpcyjnych już w latach 30. XX w., a więc równoległe z pojawieniem się koncepcji analizy hedonicznej. Aktualny stan wiedzy na temat teoretycznych własności indeksów hedonicznych przedstawiają m.in. Diewert (2003b; 2007), ILO (2004) i Triplett (2006). Dynamiczny rozwój teorii indeksów cenowych oraz ekonometrycznej teorii modeli hedonicznych odzwierciedlony jest natomiast w bogatej literaturze o charakterze empirycznym; przegląd zastosowań modeli hedonicznych oraz najnowszych kierunków ich rozwoju prezentuje Malpezzi (2002).

Cytując opinię U.S. Committee on National Statistics, Triplett (2006) podkreśla, że metody hedoniczne stanowią najbardziej obiecujący nurt badań dostosowań indeksów cen do zmian jakości – czyli zagadnienia, które powszechnie uważane jest za najbardziej skomplikowany problem związany z analizą statystyk cenowych (por. Brachinger 2002). Na szczególnie zróżnicowanym i skomplikowanym rynku nieruchomości oszacowania parametrów regresji hedonicznych nie tylko pozwalają wnioskować, jaką wartość konsumenci przypisują poszczególnym cechom mieszkania, ale również stanowią główny składnik hedonicznych indeksów cen mieszkań.

Konstrukcja wiarygodnego wskaźnika cen mieszkań jest ambitnym zadaniem głównie ze względu na charakter rynku mieszkaniowego i samego mieszkania oraz ograniczoną dostępność rozbudowanych źródeł danych na temat tego rynku. Zapoznając się z literaturą i doświadczeniem badaczy zajmujących się tą problematyką (por. Eiglsperger 2006; Nellis, Figueira-Theodorakopoulou 2005; Bover, Izquierdo 2001; Duffy 2001 i wielu innych), można wskazać dwa główne problemy związane z pomiarem tendencji cenowych na rynku mieszkań.

Pierwszym z nich jest zmiana jakości badanych dóbr, rozumiana jako zmiana użyteczności koszyka w wyniku zastąpienia pewnych dóbr z tego koszyka innymi dobrami, co prowadzi do wzrostu lub spadku użyteczności konsumpcji tego koszyka dóbr. Na przykład jeśli w okresie t kupowane były głównie mieszkania z tarasami, a w okresie $t + 1$ dominowały zakupy mieszkań z małymi balkonami, wówczas użyteczność konsumpcji koszyka zakupów z okresu t jest prawdopodobnie inna niż w okresie $t + 1$. Indeks korygujący jakość powinien mierzyć „czystą” zmianę cen, czyli zmianę cen koszyka o stałej użyteczności dla konsumenta. Heterogeniczność mieszkań powoduje,

² Na przykład Nelson (2003) szacuje wpływ hałasu z lotniska na wycenę mieszkań na rynkach mieszkaniowych w USA i Kanadzie, a Butsic i in. (2009) analizują wpływ zmian klimatycznych na wycenę domów zlokalizowanych w kurortach narciarskich. Ciekawe badanie wpływu bliskości rafinerii na ceny domów letniskowych w Szwecji przeprowadził Windstrand (2008).

że różnice między jakością zawsze występują w dwóch porównywanych okresach³. Powoduje to trudności z porównywaniem cen mieszkań i wyznaczeniem „czystej” dynamiki cen, niezaburzonej zróżnicowaniem jakości mieszkań.

Drugim problemem jest stosunkowo niewielka (w porównaniu z zasobem mieszkań) liczba transakcji, co utrudnia określenie reprezentatywnej próby danych w każdym z okresów. Mieszkania podobne lub nawet te same sprzedawane są w dużych odstępach czasu, co powoduje, że pomiędzy badanymi okresami trudno zachować reprezentatywną, taką jak w okresie bazowym, próbę transakcji mieszkaniowych. W wyniku tych dwóch problemów w każdym z badanych okresów otrzymujemy inną mieszankę mieszkań, zarówno pod względem struktury, jak i cech jakościowych. W takich warunkach badanie zmiany cen ustalonego koszyka mieszkań wydaje się być bardzo trudne lub wręcz niemożliwe.

Problem zmiany struktury może być rozwiązany przez badanie dynamiki cen populacji generalnej tam, gdzie badacz dysponuje odpowiednim zbiorem danych. Częściej jednak na rynku mieszkaniowym mamy do czynienia z ograniczonym dostępem do informacji. Wówczas problem zmiany kompozycji próby jest rozwiązany przez ustalenie stałej reprezentacji mieszkań, np. jeden uznany za reprezentatywny dla rynku typ mieszkania, średnie cech mieszkań lub sztuczny (nieistniejący w rzeczywistości) koszyk cech mieszkań o stałej strukturze.

Ustalenie miarodajnej reprezentacji służącej do badania ogólnego poziomu dynamiki cen mieszkań jest szczególnie trudne w przypadku rynku pierwotnego. Na rynku tym cechy populacji generalnej zmieniają się szybciej niż na rynku wtórnym (inwestycje pochodzą z różnych lokalizacji i od różnych producentów). W konsekwencji reprezentacja rynku dla okresu t niekoniecznie musi być miarodajna dla okresu $t + i$. Odrębnym etapem jest korekta zmiany jakości mieszkań między badanymi okresami, dokonywana najczęściej z użyciem hedonicznych modeli cen mieszkań. Budowa miarodajnego indeksu cen mieszkań wymaga przynajmniej częściowego rozwiązania problemu zmian jakości, czemu poświęcony jest nasz artykuł, jak też kompozycji reprezentacji mieszkań sprzedawanych i kupowanych w każdym z okresów.

Umiejętność wyceny wartości poszczególnych atrybutów mieszkania i znajomość zależności funkcyjnej pomiędzy nimi a całkowitą ceną mieszkania (czyli analiza hedoniczna cen) pozwalają na oszacowanie ceny dowolnej mieszanki cech mieszkaniowych (dowolnego mieszkania). Oszacowania te umożliwiają konstrukcję indeksów cenowych, które kontrolują zmiany jakościowe mieszkań. W zależności od wybranej metody hedonicznej korekta zmian jakości odbywa się na różne sposoby, za każdym razem prowadząc do pomiaru „czystej” zmiany cen, spowodowanej innymi czynnikami niż zmiany jakościowe mieszkań. Mówimy tu o „wycenie wartości poszczególnych atrybutów mieszkania” (czyli przypisaniu atrybutom mieszkania ich cen implikowanych), gdyż na cechy te nie ma odrębnych rynków (nie istnieją oddzielne rynki powierzchni użytkowej mieszkań czy typu zabudowy) i dlatego można jedynie szacować ich wartość rynkową. Przyjmujemy za ILO (2004), że wartość ta jest wypadkową użyteczności, jaką cechy mieszkania przedstawiają dla konsumenta, i ich wyceny przez producenta (por. interpretację oszacowań regresji hedonicznej w części piątej).

³ Heterogeniczność mieszkań wynika z różnic pomiędzy wieloma atrybutami, takimi jak powierzchnia użytkowa, wiek, typ, standard, lokalizacja i cechy tej lokalizacji (np. skład społeczny mieszkańców danego osiedla, dostęp do terenów zielonych, odległość od szkół, centrów komunikacji i handlu). Wydaje się, że na rynku mieszkaniowym nie istnieją dwa identyczne mieszkania. Jeśli nawet ich cechy fizyczne są takie same, to mieszkania zawsze różnią się lokalizacją (ogólną lub szczegółową) i jej cechami.

Wyceny charakterystyk determinujących cenę mieszkania mogą zmieniać się w czasie oraz przestrzeni wraz ze zmieniającymi się preferencjami nabywców i sytuacją na rynkach lokalnych. Mogą też wynikać ze zmian w sferze realnej, gdy mamy do czynienia z naturalną deprecjacją zasobu oraz jego modernizacjami. Wymagałoby to uwzględnienia w modelu hedonicznym rodzajów podmiotów nabywających mieszkanie i ich preferencji, a także budowy odrębnych modeli dla rynków lokalnych. Względnie częste zmiany wycen poszczególnych atrybutów mieszkania przez rynek powinny mieć także odzwierciedlenie w wyborze odpowiedniej metody hedonicznej. Do konstrukcji hedonicznych indeksów cen mieszkań najczęściej stosuje się hedoniczne metody bezpośrednie (por. część trzecia; więcej na ten temat: Widłak 2010).

Należy nadmienić, że inne, powszechnie stosowane w praktyce metody konstrukcji indeksów cen mieszkań częściowo zaniedbują kwestię zmian struktury i/lub zmian jakości. Do metod tych należą:

- zwykła średnia lub mediana,
- metoda śledzenia ceny reprezentatywnej nieruchomości,
- stratyfikacja i średnia ważona,
- metoda powtórnej sprzedaży (bez hedonicznej korekty zmian jakości).

3. Bezpośrednie metody wyznaczania hedonicznych indeksów cen

W bardzo obszernej anglojęzycznej literaturze dotyczącej konstrukcji indeksów cenowych z wykorzystaniem modeli hedonicznych istnieje wiele klasyfikacji wykorzystywanych metod. Brak jednej standardowej i spójnej klasyfikacji powoduje ogromne zróżnicowanie nazewnictwa, utrudniając prace empiryczne. Pewne różnice pojawiają się również w najważniejszych publikacjach (por. ILO 2004; Triplett 2006). W niniejszym artykule posłużymy się pierwszą polską klasyfikacją metod konstrukcji hedonicznych indeksów cenowych, zaproponowaną przez Widłak (2010). Autorka podkreśla, że na rynku mieszkaniowym, ze względu na jego charakter, najczęściej stosuje się metody bezpośrednie wyznaczania hedonicznych indeksów cen. Metody pośrednie, stosowane głównie w tzw. indeksach powtórnej sprzedaży (por. Łaszek, Widłak 2008), mają raczej ograniczone zastosowanie na polskim rynku nieruchomości.

Ogólnie model regresji hedonicznej można zapisać w następujący sposób:

$$P_i^t = a_0 + \sum_j a_j z_{ij}^t + \varepsilon_i^t \quad (1)$$

gdzie:

- P – cena dobra,
- a – współczynnik regresji,
- z – cecha dobra,
- ε – składnik losowy modelu;
- indeks i oznacza i -te dobro, j – j -tą cechę dobra, natomiast t jest indeksem czasu.

W zdecydowanej większości badań empirycznych modele hedoniczne są modelami log-liniowymi⁴, co zapisujemy jako:

⁴ Uzasadnienie i przykłady dotyczące rynku mieszkaniowego można znaleźć praktycznie w każdej pracy empirycznej (por. Triplett 2006).

$$\ln P_i^t = a_0 + \sum_{j=1}^J a_j z_{ij}^t + \sum_{t=2}^T b^t D^t + \varepsilon^t \quad (2)$$

Współczynniki powyższego modelu regresji (a_j) interpretuje się jako przeciętne wyceny poszczególnych cech (z_j) badanego dobra (i). Przyjmuje się, że ceny te odzwierciedlają nieobserwowalne rynkowe wyceny poszczególnych charakterystyk dobra, a więc są wynikiem działania sił popytu i podaży (por. Griliches 1961, Rosen 1974). Nieobserwowane ceny charakterystyk dobra będziemy określać mianem cen implikowanych lub domniemanych (*implicit prices*).

Założmy, że funkcja hedoniczna jest oszacowana prawidłowo, czyli poprawna jest specyfikacja modelu i zależność funkcyjna między zmienną objaśnianą a zmiennymi objaśniającymi oraz nie występują inne problemy ekonometryczne związane najczęściej ze współliniowością zmiennych czy heteroskedastycznością składnika losowego. Przy takich założeniach składnik losowy ε_i^t równania (1) obrazuje niedoszacowanie lub przeszacowanie rynkowej ceny dobra względem ceny średniej wynikającej z cech tego dobra (jego jakości mierzonej ilością poszczególnych charakterystyk, takich jak wielkość mieszkania, liczba łazienek, nasłonecznienie, typ kuchni itp.) oraz tego, jak średnio cechy te wyceniane są przez rynek.

3.1. Metoda ze zmiennymi zero-jedynkowymi czasu

Ogólną postać modelu ze zmiennymi zero-jedynkowymi (0-1) czasu można zapisać następującym równaniem:

$$\ln P_i^t = a_0 + \sum_{j=1}^J a_j z_{ij}^t + \sum_{t=2}^T b^t D^t + \varepsilon^t \quad (3)$$

gdzie D oznaczają zmienne zero-jedynkowe czasu (*time-dummy variables*).

Metodę tę stosuje się w dwóch różnych wariantach: dla dwóch sąsiadujących ze sobą okresów (*adjacent period approach*) lub dla więcej niż dwóch następujących po sobie okresów (*pooled-lub multi-pooled regression approach*). W wariantcie dla dwóch sąsiednich okresów, $t = 1$ i $t = 2$, gdzie $t = 1$ jest okresem bazowym (zgodnie ze wzorem (3) w modelu pominięta jest zmienna $D^{t=1}$), hedoniczny indeks cen przyjmuje postać:

$$Index = \exp(b^{t=2}) \quad (4)$$

Indeks ten jest skorygowanym jakościowo indeksem hedonicznym cen dobra. Dekompozycja ceny na cenotwórcze zmienne określające charakterystyki dobra (z_j) oraz zmienne czasowe, a także interpretacja współczynników regresji *ceteris paribus* sprawiają, że wyrażenie opisane równaniem (3) interpretujemy jako wskaźnik „czystej” zmiany cen po oddzieleniu efektów zmiany jakości.

Analogicznie, jeśli model (3) szacujemy na podstawie danych z kilku okresów (dwóch lub więcej, aż do wszystkich dostępnych w próbie), indeksy zmiany cen wyznaczamy każdorazowo ze wzoru (4), uogólniając go na przypadek b^t . W wariantcie modelu (3), szacowanym na danych z wielu okresów współczynniki regresji a_j odzwierciedlają przeciętne wartości wycen rynkowych poszczególnych cech dobra we wszystkich rozpatrywanych okresach. Sprowadza się to do przyjęcia zało-

żenia, że w szacowanych okresach wyceny rynkowe poszczególnych charakterystyk dobra są stałe. Założenie to jest podstawowym źródłem krytyki metody ze zmiennymi zero-jedynkowymi czasu wszystkich okresów (por. m.in. Conniffe, Duffy 1999; Diewert 2003a; Triplett 2006).

3.2. Metoda cen charakterystyk

W przeciwieństwie do metody ze zmiennymi 0-1 czasu w metodzie cen charakterystyk przyjmuje się założenie, że implikowane ceny charakterystyk dobra zmieniają się z okresu na okres. Zwolennicy tego podejścia twierdzą, że ceny całkowite dóbr zmieniają się pod wpływem zmiany cen implikowanych. W przypadku mieszkań można wyobrazić sobie sytuację, w której moda wpływa na preferencje nabywców w ten sposób, że w porównaniu z ubiegłymi okresami wolą posiadać mieszkania z aneksami kuchennymi niż zamkniętymi kuchniami. Wówczas cena zmiennej o wartości „kuchnia zamknięta” będzie maleć, co może wywołać spadek indeksu cenowego mieszkań. Inny przykład dotyczy komputerów: malejące ceny pamięci powodują obniżkę cen samych komputerów. Zgodnie z tym założeniem w metodzie cen charakterystyk wykorzystuje się wyceny cech dobra, współczynniki a_j , w ważonej, konwencjonalnej formule indeksu cenowego.

Metoda cen charakterystyk wymaga, aby dla każdego z badanych okresów szacowany był odrębny model regresji hedonicznej zapisany wzorem (1) lub (2). Interpretacja współczynników regresji stojących przy poszczególnych zmiennych objaśniających we wzorach (1) i (2) pozwala na zbudowanie indeksu cen wykorzystującego tradycyjne formuły Laspeyresa i Paaschego (w konsekwencji także Fishera), z tym że wagami tego indeksu nie są ilości dobra docelowego (mieszkania), lecz ilości charakterystyk, które tworzą to dobro. Korzystając z modelu zapisanego równaniem (2), wskaźnik zmiany cen między okresem t a $t + 1$ zapiszemy jako:

- indeks cen charakterystyk typu Laspeyresa:

$$Index = \frac{\exp \sum_j a_j^{t+1} q_j^t}{\exp \sum_j a_j^t q_j^t} \quad (5)$$

- indeks cen charakterystyk typu Paaschego:

$$Index = \frac{\exp \sum_j a_j^{t+1} q_j^{t+1}}{\exp \sum_j a_j^t q_j^{t+1}} \quad (6)$$

gdzie q oznacza ilość danej cechy.

Wagi ilościowe indeksu cenowego q mogą być sumą ilości cech dla poszczególnych jednostek produktu lub mogą wyrażać wielkość średnią danej cechy w badanej populacji (por. Nellis, Figueira-Theodorakopoulou 2005; Triplett 2006). Na przykład jeśli badamy indeks zmiany cen dwóch mieszkań o powierzchni 30 i 40 m², to q_1 może wynosić 70 lub 35 m². Dla zmiennych typu jakościowego (np. rodzaj kuchni) wagi odzwierciedlają udział danej cechy w badanej próbie.

Stosując metodę cen charakterystyk, przyjmujemy, że nabywca mieszkania płaci tak naprawdę za pewien zbiór właściwości tego mieszkania. Kupuje bowiem pewien określony zestaw cech mieszkania, tj. 60 m² powierzchni użytkowej, jasną kuchnię, balkon, centralną lokalizację budynku, dobrą komunikację z centrum, bliskość szkół itp. Zakładamy także, że zmiana ceny mieszkania wynika ze zmiany cen implikowanych. Przez zastosowanie stałych wartości średnich dla zmieniających objaśniających (wartość q we wzorach (5) i (6)) konstruujemy indeks o stałej jakości mieszkań (tzw. *constant quality index*).

3.3. Metoda imputacji

Metoda ta prezentowana jest przez ILO (2004). Korzystając z równania (2), wyznacza się model regresji hedonicznej na danych z okresu $t + 1$ (lub t , w zależności od tego, który okres, przeszły czy bieżący, ma być okresem referencyjnym). Następnie przez podstawienie do oszacowanego równania wartości charakterystyk dóbr z okresu t ($t + 1$) przypisuje się tym dobrom ceny, które odpowiadają okresowi $t + 1$ (t). Oszacowane w ten sposób ceny dóbr określamy jako ceny odbite (*shadow prices*). Nazwa wynika z interpretacji tak wyznaczonych cen: są one prawdopodobnymi (przypisanymi) cenami z okresu $t + 1$ (t) dóbr sprzedanych w okresie t ($t + 1$). Indeks wyznaczony za pomocą tej metody zapisujemy poniższym wzorem:

$$Index = \frac{\sum_{i=1}^N \frac{1}{N} \hat{p}_i^{t+1}(z_i^t)}{\sum_{i=1}^N \frac{1}{N} p_i^t(z_i^t)} \quad (7)$$

lub

$$Index = \frac{\sum_{i=1}^N \frac{1}{N} p_i^{t+1}(z_i^{t+1})}{\sum_{i=1}^N \frac{1}{N} \hat{p}_i^t(z_i^{t+1})} \quad (8)$$

gdzie $\hat{p}(z_i)$ oznacza cenę odbitą dobra i , będącą funkcją charakterystyk z_i , a $p(z_i)$ – rzeczywistą cenę dobra i o charakterystykach z_i . Na przykład $\hat{p}_i^{t+1}(z_i^t)$ oznacza cenę odbitą, obliczoną przez podstawienie do równania regresji hedonicznej oszacowanej na próbie danych z okresu $t + 1$, charakterystyk z_i dóbr sprzedanych w okresie t . Korzystając z równania (2), cenę odbitą dobra i w okresie $t + 1$ możemy zatem zapisać jako:

$$\hat{p}_i^{t+1} = \exp \left(a_0^{t+1} + \sum_j a_j^{t+1} z_{ij}^t \right) \quad (9)$$

Jak podaje ILO (2004), indeks może być oparty na średnich cenach rzeczywistych i odbitych lub może być średnią z indywidualnych indeksów cen badanych dóbr. Oznacza to, że wzory (7) i (8) możemy zapisać alternatywnie w następujący sposób:

$$Index = \frac{1}{N} \sum_{i=1}^N \frac{\hat{p}_i^{t+1}(z_i^t)}{p_i^t(z_i^t)} \quad (10)$$

lub

$$Index = \frac{1}{N} \sum_{i=1}^N \frac{p_i^{t+1}(z_i^{t+1})}{\hat{p}_i^t(z_i^{t+1})} \quad (11)$$

Metoda ta, podobnie jak dwie pierwsze z grupy metod bezpośrednich, powinna być stosowana na całej badanej populacji, choć może być również zastosowana do jej ustalonej reprezentacji (por. Triplett 2006).

W bezpośredniej metodzie imputacji kontrola zmian jakości dóbr w porównywanych okresach odbywa się przez porównanie ceny rzeczywistej z ceną odbitą wyznaczoną dla tego samego wektora charakterystyk z_t . Indeksy przedstawione wzorami (7)–(10) obrazują zatem „czystą” zmianę cen (ponieważ jakość dóbr w okresie t i $t + 1$ nie zmienia się, bo są to te same dobra). Podstawowa różnica między zastosowaniem równań (7) i (10) a (8) i (11) polega na tym, że w pierwszym podejściu model regresji hedonicznej wyznaczany jest na danych z okresu bieżącego, a w podejściu drugim – danych z okresu bazowego. Jeśli przyjmiemy jeden okres bazowy, wówczas – wykorzystując podejście drugie – wystarczy wyznaczyć jeden model regresji, a następnie stosować go dla bieżących okresów.

Bezpośrednie metody konstrukcji hedonicznych indeksów cen rozwiązują problemy dopasowania wskaźników cen do zmian jakościowych badanego produktu. Metody te powinny być stosowane niezależnie od różnych sposobów kontroli zmian struktury reprezentacji rynku mieszkań (czyli wyznaczania miarodajnej reprezentacji populacji generalnej), na której powinien opierać się indeks.

4. Dane

Dostępność danych jest największym problemem, determinującym możliwość stosowania metod hedonicznych, a w konsekwencji jakość wyników pomiaru dynamiki cen. Konstrukcja hedonicznych indeksów cen wymaga korzystania z obszernych baz danych. Niezbędne są w nich zmienne opisujące atrybuty poszczególnych nieruchomości, a także zmienne określające lokalizację tych nieruchomości. Te ostatnie często tworzone są specjalnie na potrzeby estymacji modeli hedonicznych rynku mieszkaniowego i najczęściej pochodzą z innych źródeł niż baza danych o nieruchomościach. Zróżnicowanie baz danych wynika z celów ich utworzenia, a ich wykorzystanie do modelowania hedonicznego jest najczęściej wtórne. Krótki przegląd najważniejszych źródeł danych podają Łaszek i Widłak (2007).

W Polsce ułomności oficjalnej statystyki rynku mieszkań są przyczyną rozwoju firm prywatnych lub instytucji państwowych prowadzących analizy rynku mieszkaniowego na podstawie własnych baz danych. Należą do nich mieszkaniowe portale internetowe oraz stowarzyszenia zawodowe uczestników rynku: deweloperów, pośredników, rzeczoznawców majątkowych, notariuszy, firm konsultingowych. Bazy te mają często lokalny charakter i niechętnie są udostępniane podmiotom trzecim. Często też badania tego typu są fragmentaryczne i obejmują tylko wybrane segmenty rynku. Przykładem takiej bazy jest Baza Rynku Nieruchomości (BaRN) powstająca w NBP od 2006 r., którą posługujemy się w niniejszym badaniu. Notowania cen odbywają się kwartalnie (począwszy od III kwartału 2006 r.) i dotyczą zarówno rynku pierwotnego, jak i wtórnego.

Dane są podzielone na dane ofertowe i transakcyjne z rynku kupna i sprzedaży mieszkań oraz z rynku najmu okazjonalnego. Zasięg geograficzny bazy to miasta wojewódzkie: Białystok, Katowice, Kielce, Kraków, Łódź, Opole, Poznań, Rzeszów, Szczecin, Warszawa, Zielona Góra, Bydgoszcz, Lublin, Olsztyn, Wrocław oraz Trójmiasto. Dane z rynku wtórnego dotyczą mieszkań zlokalizowanych w granicach administracyjnych miast. Dane z rynku pierwotnego zawierają także informacje o nowych mieszkaniach, powstających w sąsiedztwie tych miast. Przeważnie pełnią one funkcję zaplecza noclegowego dla miast i są naturalnym skutkiem ich rozwoju i rozrastania się.

Dane w bazie BaRN pochodzą od firm deweloperskich, biur pośrednictwa sprzedaży nieruchomości, a częściowo także z aktów notarialnych (z powiatowych Rejestrów Cen i Wartości Nieruchomości). Pozyskiwane informacje dotyczą wyłącznie transakcji zawartych na wolnym rynku; w bazie nie są gromadzone informacje dotyczące prywatyzacji zasobu komunalnego czy zakładowego, inwestycji TBS, darowizn itp. Powtarzające się (pochodzące z kilku źródeł danych) oferty i transakcje są eliminowane na etapie wprowadzania danych.

Pojedyncze obserwacje w BaRN odpowiadają poszczególnym ofertom i transakcjom. Poza podstawowymi informacjami o cenie opisywane są wybrane parametry charakteryzujące mieszkanie, a także, w ograniczonym zakresie, dane dotyczące oceny jego lokalizacji szczegółowej i ogólnej. Część zmiennych opisujących lokalizację ma charakter jakościowy. Wartości tych zmiennych są nadawane przez ekspertów zajmujących się analizą rynków lokalnych i mają przez to charakter subiektywny. Pełna lista zmiennych zawiera 19 atrybutów opisujących fizyczne cechy mieszkania oraz pięć zmiennych lokalizacyjnych. Właściwie wszystkie zmienne gromadzone w bazie są cechami cenotwórczymi mieszkań i na podstawie wiedzy eksperckiej można by je zastosować w procesie doboru zmiennych do modelu. W praktyce znaczna liczba braków danych ograniczyła zbiór zmiennych wykorzystanych w szacowanych modelach zaledwie do 12, w tym tylko dwóch zmiennych lokalizacyjnych. Listę zmiennych zastosowanych w modelach hedonicznych oraz ich krótki opis przedstawiono w punkcie 6.1.

Względy praktyczne gromadzenia i przetwarzania danych zdecydowały o niestandardowej formie uporządkowania danych. Informacje o aktualnych ofertach dotyczą pierwszych dni ostatniego miesiąca kwartału kalendarzowego (marca, czerwca, września i grudnia). Za dane transakcyjne przypisane do pierwszego kwartału uznaje się transakcje zawarte w grudniu, styczniu i lutym, transakcje z drugiego kwartału to transakcje zawarte w marcu, kwietniu i maju itd. W danych tych występuje zatem miesięczne przesunięcie wobec kwartałów kalendarzowych. Gdy rynek mieszkaniowy jest ustabilizowany, taka zmiana układu gromadzenia danych nie ma większego znaczenia dla badania tendencji cenowych. Przesunięcie danych transakcyjnych o miesiąc może obciążać analizy w sytuacji, gdy rynek zmienia się bardzo szybko.

Pomimo wielu ograniczeń zastosowania bazy BaRN oceniamy ją wysoko na tle innych baz danych gromadzących informacje o transakcjach i ofertach na rynku mieszkaniowym (bazy internetowe, AMRON, badania aktów notarialnych przeprowadzone przez GUS). Ufamy, że wraz z rozwojem baz danych i zwiększeniem zawartych w nich zasobów informacji będzie można konstruować modele o coraz lepszej jakości.

Mając do dyspozycji ceny ofertowe i transakcyjne mieszkań, decydujemy się na konstrukcję indeksu cen transakcyjnych, pomimo że liczba obserwacji w porównaniu z liczbą ofert jest stosunkowo niewielka: w okresie od III kwartału 2006 r. do II kwartału 2009 r. w bazie BaRN zgromadzono 32 905 obserwacji ofertowych oraz 1494 obserwacje dotyczące transakcji na warszawskim rynku

wtórny mieszkań. Przyjmujemy za ILO (2004) założenie, że w przeciwieństwie do ceny ofertowej cena transakcyjna jest wypadkową działania sił popytu i podaży. Dlatego traktujemy wskaźnik cen transakcyjnych, w przeciwieństwie do wskaźnika cen ofertowych, jako bardziej odpowiednią miarę rzeczywistych tendencji cenowych na rynku.

5. Zagadnienia ekonometryczne

Do podstawowych zagadnień ekonometrycznych związanych z estymacją hedonicznego modelu regresji należą: dobór postaci funkcyjnej i zbioru zmiennych objaśniających, wybór metody estymacji, analiza własności składnika losowego, ocena konsekwencji współliniowości zmiennych objaśniających.

Podsumowując wyniki dotychczasowych badań, Triplett (2006) podkreśla, że ani klasyczna teoria użyteczności, ani teoria produkcji nie dają podstaw do ustalenia jednoznacznej postaci funkcyjnej modelu regresji hedonicznej. Przesłanką selekcji postaci funkcyjnej równania regresji hedonicznej jest jakość dopasowania modelu, oceniana za pomocą kryteriów ekonometrycznych. Wśród najczęściej stosowanych postaci: liniowej, log-liniowej i podwójnie logarytmicznej (rzadziej wykładniczej lub hiperbolicznej) największą popularność, popartą wynikami badań empirycznych, zyskała postać log-liniowa, w której logarytm ceny objaśniany jest za pomocą liniowej funkcji ilościowych i (lub) jakościowych zmiennych objaśniających. W empirycznej części pracy stosujemy zatem log-liniową postać modelu regresji hedonicznej opisaną wzorem (2).

Warto podkreślić, że ocena parametru przy zmiennej objaśniającej z_i w modelu hedonicznym nie powinna być bezpośrednio interpretowana jako oszacowanie krańcowej użyteczności cechy i dla konsumenta. Rosen (1974) wykazał, że współczynniki regresji hedonicznych odzwierciedlają jednocześnie stronę popytową i podażową rynku, a zatem ich jednoznaczna interpretacja jako krańcowej użyteczności dla konsumenta nie jest uprawniona. Zagadnienie to znane jest w ekonomii jako problem identyfikacji; pojawia się w sytuacji, gdy obserwowane ceny i ilości są skutkiem interakcji czynników podażowych i popytowych, których indywidualny wpływ na obserwowane zachowania rynku nie jest obserwowalny. Oznacza to, że współczynniki funkcji hedonicznej nie można interpretować ani jako krańcowego kosztu producenta, ani jako krańcowej użyteczności konsumenta. Specyfika rynku mieszkaniowego, na którym technologia produkcyjna (czyli strona podażowa) jest podobna dla wszystkich sprzedawców, a konsumenci są zróżnicowani z punktu widzenia ich preferencji, sprawia, że można przyjąć założenie o stałości technologii dla każdego producenta. „Wtedy zamiast splątanej rodziny funkcji podaży istnieje jednoznaczna funkcja podaży z odpowiadającą jej funkcją hedoniczną opisującą ceny charakterystyk, które producent dostarcza przy danej dominującej technologii dla zaspokojenia bieżącej struktury popytu” (ILO 2004, s. 378, tłum. aut.). Oszacowania współczynników regresji hedonicznej można zatem, w pewnym uproszczeniu, interpretować jako wyceny poszczególnych cech mieszkania przy danej technologii. Głównym elementem merytorycznej weryfikacji hedonicznego modelu cen nieruchomości powinna być zatem ocena znaku i wielkości oszacowań parametrów z punktu widzenia ich ekonomicznej sensowności; wyniki tej weryfikacji przedstawiamy w części szóstej.

Zmienne objaśniające w modelu regresji hedonicznej powinny być odzwierciedleniem tych czynników, których wpływ na zmienną objaśnianą jest najsilniejszy; ideą modelu hedoniczne-

go jest bowiem przedstawienie ceny dobra jako sumy wycen jego charakterystyk. W praktyce niektóre potencjalnie najważniejsze zmienne objaśniające są subiektywne lub nieobserwowalne (np. poziom bezpieczeństwa w danej dzielnicy) lub nie są rejestrowane w bazach danych (np. odległość od parku). W badaniach empirycznych konieczne jest zatem ograniczenie zbioru zmiennych objaśniających do zmiennych obserwowalnych, rejestrowanych w bazach danych oraz dostępnych dla jak największej liczby obserwacji. Nie można przy tym wykluczyć, że nieobserwowalne lub nierejestrowane cechy są skorelowane z cechami uwzględnionymi w modelu, wprowadzając obciążenie estymatorów metody najmniejszych kwadratów (MNK). Sugerowaną w literaturze metodą rozwiązania tego problemu, znanego w ekonometrii jako „efekt pominiętych zmiennych”, jest zastosowanie metody zmiennych instrumentalnych (por. np. Wooldridge 2002). Trudności ze znalezieniem odpowiednich zmiennych – instrumentów sprawiają jednak, że w praktyce ta metoda estymacji modeli hedonicznych jest rzadko stosowana. W badaniach empirycznych wykorzystuje się zazwyczaj inne metody: poszukiwanie zmiennych zastępczych (*proxy*) lub próbę oszacowania związków między zmiennymi obserwowalnymi a nieobserwowalnymi i w rezultacie pominiętymi. W niniejszej pracy, podobnie jak w innych opracowaniach o charakterze empirycznym, zbiór potencjalnych zmiennych objaśniających jest z góry ograniczony zakresem bazy danych. Dodatkowe ograniczenie liczby zmiennych objaśniających może być konieczne z powodu współliniowości, czyli wysokiej korelacji między nimi. Przybliżona współliniowość zmiennych objaśniających nie powoduje wprowadzenia obciążenia estymatorów MNK, ale jest przyczyną ich nieefektywności, a zatem utraty precyzji oszacowań. Jeżeli dodatkowo w modelu pominięto nieobserwowalne zmienne skorelowane ze zmiennymi uwzględnionymi w modelu, to współliniowość między nimi prowadzi do obciążenia oszacowań przy zmiennych obecnych w modelu, gdyż przejmują one część efektu zmiennych nieobserwowalnych.

Współliniowość zmiennych objaśniających jest typowym problemem związanym z estymacją modeli hedonicznych, w których niektóre cechy badanych obiektów są związane z innymi (np. mieszkanie ulokowane na piątym piętrze z pewnością znajduje się w budynku wyposażonym w windę). Standardowe metody oceny siły współliniowości, w tym stosowane w empirycznej części pracy czynniki inflacji wariancji CIW (VIF – *variance inflation factors*, por. Maddala 2008), nie stanowią miar idealnych. Ponadto do ich interpretacji konieczna jest merytoryczna wiedza o powiązaniach między poszczególnymi charakterystykami obiektów, w tym intuicja i wiedza ekspercka pozwalające ocenić, jak obecność lub nieobecność jednej z cech powinna wpłynąć na znaki oszacowań pozostałych parametrów i inne statystyki modelu. Praktyka estymacji modeli hedonicznych wskazuje, że współliniowości zmiennych objaśniających można uniknąć dzięki przemyślanemu doborowi zmiennych objaśniających.

Ostatnim zagadnieniem ekonometrycznym wartym rozważenia w kontekście hedonicznych modeli cen są własności składnika losowego modelu regresji hedonicznej. Zjawiskiem typowym w hedonicznych modelach cen mieszkań (por. np. Fletcher i in. 2000) jest heteroskedastyczność (zmiennność wariancji) składnika losowego, powodująca nieefektywność estymatorów MNK, jednak bez naruszenia ich nieobciążoności. Heteroskedastyczność najczęściej wynika z heterogeniczności próby, obejmującej zarówno obiekty „małe”, jak i „duże”, cechujące się większą zmiennością. Eliminacja tej konkretnej przyczyny heteroskedastyczności – czyli podział próby na podpróby o możliwie jednolitej strukturze – na ogół nie jest możliwa z powodu ograniczonej liczebności dostępnych prób, chociaż na rozwiniętych rynkach o zinstytucjonalizowanym charakterze gromadzenia danych niekiedy podejmowane są takie analizy (por. Gouriéroux, Laferrère 2006).

Drugą własnością składnika losowego, rzadko badaną w przypadku modeli przekrojowych, ale charakterystyczną dla hedonicznych modeli cen nieruchomości, jest autokorelacja przestrzenna związana z przestrzennymi (geograficznymi) zależnościami między badanymi domami lub mieszkaniami. Jeśli obserwowane ceny są ze sobą geograficznie skorelowane, to w modelu może pojawić się autokorelacja przestrzenna składnika losowego z konsekwencjami w postaci nieefektywności estymatorów MNK (por. Brachinger 2002). W literaturze coraz większą popularność zyskują modele uwzględniające zależności przestrzenne; na przykład Bazyl (2007) wprowadza do modelu zmienną opisującą średnią cenę okolicznych mieszkań, wyznaczoną na podstawie ich współrzędnych geograficznych. Tak szczegółowe dane są jednak trudno osiągalne i nie zawierają ich powszechnie dostępne bazy danych.

W empirycznej części opracowania w celu uzyskania nieobciążonych estymatorów błędów standardowych stosujemy odporny estymator wariancji Neweya–Westa (HAC – *heteroskedasticity and autocorrelation consistent estimator*; por. np. Verbeek 2004), pozostawiając do dalszych badań uwzględnienie zależności przestrzennych w cenach nieruchomości. Ponieważ przyczyną zarówno heteroskedastyczności, jak i autokorelacji składnika losowego może być błędna specyfikacja modelu (co w przypadku modelu regresji hedonicznej oznacza najczęściej pominięcie głównych zmiennych objaśniających), szacowane w kolejnej części opracowania modele hedoniczne poddajemy standardowemu testowi specyfikacji RESET.

6. Wyniki empiryczne

W kolejnych punktach przedstawiamy procedurę przygotowania zbioru danych, wyniki estymacji modeli hedonicznych oraz porównanie skonstruowanych na ich podstawie hedonicznych indeksów cen. W procesie estymacji i weryfikacji modeli przyjęto, że:

- szacowane modele mają log-liniową postać funkcyjną,
- modele szacowane są MNK z odpornymi estymatorami błędów standardowych (HAC),
- modele muszą cechować się poprawną specyfikacją (sprawdzaną testem RESET) oraz czynnikami inflacji wariancji dla wszystkich zmiennych nieprzekraczającymi wartości 10,
- weryfikacja hipotez statystycznych przeprowadzana jest na poziomie istotności 5%.

6.1. Przygotowanie danych i wybór zmiennych

Dostępne dane obejmują 12 kwartałów: od III kwartału 2006 r. do II kwartału 2009 r. Dla tego samego zakresu czasowego wyznaczane są indeksy cen mieszkań.

Przygotowanie danych do estymacji modeli hedonicznych przebiegało dwuetapowo. Pierwszym krokiem była eliminacja błędów wynikających z pomyłek we wprowadzaniu danych oraz wybór mieszkań, które można zaliczyć do średniego segmentu rynku. Na tym etapie wykorzystano głównie wiedzę ekspercką. W celu wykluczenia obserwacji odstających, czyli obrotu mieszkaniami luksusowymi, na dane nałożono następujące warunki:

- cena za metr kwadratowy mieszkania mniejsza lub równa 18 tys. zł,
- powierzchnia użytkowa mniejsza lub równa 200 m²,

- liczba pokoi mniejsza lub równa 8,
- budynki nie wyższe niż 26 kondygnacji.

W wyniku tego czyszczenia danych liczność próby zmniejszyła się o 14 obserwacji i wynosi 1949.

Drugi etap oczyszczania danych przeprowadzany był odrębnie dla każdego z 26 szacowanych modeli i polegał na wyznaczeniu reszt studentyzowanych i usunięciu obserwacji, których reszta studentyzowana była większa bądź równa 2. Procedurę tę opisuje m.in. Maddala (2006), a przykłady jej zastosowań empirycznych w mieszkaniowych modelach hedonicznych znaleźć można m.in. w pracy Gouriéroux i Laferrière (2006), a także opracowaniu Luxa i Sunegi⁵. Dla prób przed zastosowaniem kryterium studentyzacji i po jego zastosowaniu przeprowadziłyśmy analizę rozkładów empirycznych i statystyk opisowych zmiennych. Z pewnym przybliżeniem można stwierdzić, że obserwacje eliminowane to głównie mieszkania własnościowe, o średniej wielkości, dobrym standardzie wykończenia i wybudowane głównie po 1990 r. Jednocześnie dominantę rozkładu cen metra kwadratowego mieszkań eliminowanych na tym etapie stanowią mieszkania w cenie poniżej średniej (4000–6000 zł/m²). Na podstawie wiedzy eksperckiej przypuszczamy, że cena mieszkań o powyższych cechach powinna być raczej wyższa niż niższa od średniej. Przyczyną tego niedopasowania obserwacji do estymowanych modeli może być m.in. wpływ pominiętych zmiennych (np. lokalizacja szczegółowa mieszkania lub budynku, skład społeczny mieszkańców osiedli). Na podstawie zastosowanej prostej analizy statystycznej eliminowanych obserwacji nie jesteśmy w stanie dokładnie określić zależności między obserwacjami wyeliminowanymi i pozostającymi w modelu. Określenie tych zależności może być interesujące dla dalszych prac badawczych związanych z konstrukcją hedonicznych modeli cen mieszkań.

Pierwszym etapem specyfikacji i doboru zmiennych do modelu jest wykorzystanie teorii ekonomii. Na podstawie wcześniejszych badań oraz wiedzy eksperckiej wszystkie zmienne dostępne w bazie BaRN można uznać za zmienne cenotwórcze na rynku mieszkań. Jednak ze względu na znaczny odsetek braków danych (BD) do modeli hedonicznych włączyłyśmy tylko niektóre zmienne. W celu rozstrzygnięcia trudnego dylematu pomiędzy liczbą obserwacji a liczbą zmiennych, które wykorzystamy w modelu, przyjęłyśmy zasadę, że rozpatrujemy zmienne, w których BD nie przekraczają 25% wszystkich obserwacji. Przykładami ciekawych zmiennych, które nie zostały wskutek tego zastosowane w modelach, są zmienne opisujące dostępność garażu (w której BD stanowią 76%, czyli 1120 obserwacji na 1494 przypadki), komórki lokatorskiej (gdzie BD stanowią 72%) oraz ważne zmienne lokalizacyjne, takie jak ocena subdzielniczy, ocena osiedla, ocena położenia mieszkania w budynku czy ocena lokalizacji szczegółowej (subiektywna ocena położenia bloku w odniesieniu do całego osiedla).

Eliminacja zmiennych z dużą liczbą BD jest szczególnie ważna dla konstrukcji indeksu metodą bezpośredniej imputacji. W metodzie tej wykorzystanie zmiennych ze znaczną liczbą BD wywołałaby konieczność wyznaczenia indeksu na bardzo ograniczonych próbach danych. Ponieważ badane próbki nie są próbkami losowymi (badanie jest dobrowolne, w związku z czym trudno byłoby wykorzystać operat losowania) i nie są reprezentatywne w sensie statystycznym, ważne jest, aby były jak największe.

Wybór pomiędzy liczbą zmiennych zastosowanych w modelu a liczbą obserwacji był trudny i rozstrzygany był ekspercko oraz na podstawie testów poprawności specyfikacji RESET kolejnych

⁵ Praca tych autorów nie jest publikowana. Sposób oczyszczania danych został zaprezentowany podczas spotkania warsztatowego w Pradze w 2008 r.

modeli. W Załączniku 1 przedstawiamy porównanie początkowej liczby obserwacji w każdym modelu z ich liczbą po wyeliminowaniu obserwacji odstających (metodą usunięcia obserwacji, których wartość bezwzględna reszt studentyzowanych jest większa od 2), oraz obserwacji, w których występują braki danych. Najbardziej wyraźne zmniejszenie wielkości próby można zaobserwować w przypadku metody 0-1 wszystkich okresów w wersji rozszerzonej (M1+), w której zdecydowaliśmy się dołączyć do modelu dwie zmienne niespełniające przyjętego kryterium 25% BD (model szacowany jest na danych z wielu okresów, co zwiększa liczbę obserwacji). Zmienne te, stanowiące w opinii eksperckiej ważne czynniki cenotwórcze, to ocena subdzielnic i technologia budowy.

Ostatecznie wybraliśmy 12 zmiennych; poniżej przedstawiamy ich opis wraz z odpowiadającymi im zmiennymi zero-jedynkowymi (w przypadku zmiennych jakościowych):

- ilekondygnacji: liczba kondygnacji budynku;
- ilepokoi_1_2: zmienna przyjmuje wartość 1 dla mieszkań jedno- lub dwupokojowych, 0 – w przeciwnym przypadku (dalej wpp);
- kuchnia – zmienna jakościowa o trzech kategoriach:
 - kuch_1: 1 dla mieszkań z oddzielną ciemną kuchnią, 0 – wpp,
 - kuch_2: 1 dla mieszkań z oddzielną widną kuchnią, 0 – wpp (zmienna bazowa),
 - kuch_3: 1 dla mieszkań z aneksem kuchennym połączonym z salonem, 0 – wpp;
- ocena_d – zmienna typu jakościowego o dwóch (w przypadku Warszawy) kategoriach; wartości zmiennej odzwierciedlają subiektywną ocenę ekspercką dzielnicy, w której zlokalizowane jest mieszkanie, w odniesieniu do innych dzielnic miasta, w następujących kategoriach:
 - ocena_d_2: 1 dla dzielnic o przeciętnej lokalizacji, 0 – wpp (zmienna bazowa),
 - ocena_d_3: 1 dla dzielnic o dobrej lokalizacji, 0 – wpp;
- ocena_s – zmienna typu jakościowego o trzech kategoriach; wartości zmiennej odzwierciedlają subiektywną ocenę ekspercką położenia subdzielnic w obrębie dzielnicy (np. subdzielnicy Wyględów w obrębie dzielnicy Mokotów) w następujących kategoriach:
 - ocena_s_1: 1 dla subdzielnic o złej lokalizacji, 0 – wpp,
 - ocena_s_2: 1 dla subdzielnic o przeciętnej lokalizacji, 0 – wpp (zmienna bazowa),
 - ocena_s_3: 1 dla subdzielnic o dobrej lokalizacji, 0 – wpp;
- pietrood: numer kondygnacji, na której znajduje się mieszkanie;
- rokbudowy – zmienna wyrażająca rok budowy budynku z dokładnością co do dekady (np. dla budynków wybudowanych w latach 70. zmienna ta przyjmuje wartość 1970); zgodnie z klasyfikacją przyjętą przez Narodowy Spis Powszechny (2002) zmienna została rozpisana na następujące zmienne 0-1:
 - rok_1: 1 dla budynków wybudowanych przed 1918 r., 0 – wpp,
 - rok_2: 1 dla budynków wybudowanych w latach 1918–1944, 0 – wpp,
 - rok_3: 1 dla budynków wybudowanych w latach 1945–1970, 0 – wpp,
 - rok_4: 1 dla budynków wybudowanych w latach 1971–1978, 0 – wpp,
 - rok_5: 1 dla budynków wybudowanych w latach 1979–1988, 0 – wpp,
 - rok_6: 1 dla budynków wybudowanych w latach 1989–2001, 0 – wpp (zmienna bazowa),
 - rok_7: 1 dla budynków wybudowanych w latach 2002–2005, 0 – wpp,
 - rok_8: 1 dla budynków wybudowanych po 2005 r., 0 – wpp,
 - rok_nowe – zmienna stanowi samodzielną kategorię; przyjmuje wartość 1 dla mieszkań wybudowanych najwyżej jeden rok przed rokiem notowania w celu wyodrębnienia

- mieszkań nowych lub jeszcze nie wybudowanych przez dewelopera, a już sprzedawanych na rynku wtórnym (obróć mieszkaniami i prawami do mieszkań), 0 – wpp;
- spółdzielcze: zmienna przyjmuje wartość 1 dla transakcji opartych na spółdzielczym własnościowym prawie do lokalu, 0 – dla mieszkań własnościowych;
 - *sq_pow*: kwadrat powierzchni użytkowej mieszkania wyrażonej w m²;
 - *standard_wykonczenia* – zmienna typu jakościowego o czterech kategoriach; wartości zmiennej odzwierciedlają subiektywną ocenę ekspercką (pośrednika lub analityka rynku lokalnego) o jakości wykończenia mieszkania w następujących kategoriach:
 - *stand_1*: 1 dla mieszkań o wysokim standardzie wykończenia, 0 – wpp,
 - *stand_2*: 1 dla mieszkań o przeciętnym standardzie wykończenia, 0 – wpp (zmienna bazowa),
 - *stand_3*: 1 dla mieszkań o niskim standardzie wykończenia, 0 – wpp,
 - *stand_4*: 1 dla mieszkań w tzw. stanie surowym, 0 – wpp;
 - *techn_budowy* – zmienna jakościowa o sześciu kategoriach:
 - *tech_1*: 1 dla mieszkań wybudowanych w technologii tradycyjnej, 0 – wpp (zmienna bazowa),
 - *tech_2*: 1 dla mieszkań wybudowanych w technologii tradycyjnej udoskonalonej, 0 – wpp,
 - *tech_3*: 1 dla mieszkań wybudowanych w technologii monolitycznej, 0 – wpp,
 - *tech_4*: 1 dla mieszkań wybudowanych w technologii prefabrykowanej, 0 – wpp,
 - *tech_5*: 1 dla mieszkań wybudowanych w technologii drewnianej, 0 – wpp,
 - *tech_6*: 1 dla mieszkań wybudowanych w technologii szkieletu stalowego, 0 – wpp;
 - *winda*: zmienna przyjmuje wartość 1 dla mieszkań w budynkach wyposażonych w windę, 0 – wpp.

6.2. Wyniki estymacji

W ramach estymacji hedonicznych modeli cen oszacowanych zostało 26 log-liniowych jednorównaniowych modeli ekonometrycznych. Opisana w punkcie 3.1 metoda zmiennych 0-1 została zastosowana w dwóch wersjach: na podstawie całego zbioru danych (wszystkich okresów; M1) oraz sąsiednich okresów (M2), z wykorzystaniem danych z kolejnych par kwartałów, z których bazą był zawsze kwartał wcześniejszy. Z kolei metoda zmiennych 0-1 wszystkich okresów została zastosowana do dwóch zbiorów zmiennych objaśniających: podstawowego (identycznego z zestawem zmiennych uwzględnionych w metodzie imputacji) oraz rozszerzone (M1+). Metoda zmiennych 0-1 pozwala bowiem na włączenie do modelu dodatkowych zmiennych, których uwzględnienie w metodzie imputacji nie jest możliwe z powodu brakujących danych i przyjętej przez nas zasady, że nie mogą one stanowić więcej niż 25% początkowej wielkości próby. Do modelu podstawowego M1 dołączone zostały cztery dodatkowe zmienne: ocena subdzielniczy, liczba kondygnacji, wyposażenie w windę oraz technologia budowy (która jednak okazała się współliniowa ze zmiennymi opisującymi rok budowy i została wyeliminowana w procesie estymacji). Rozszerzony model M1+ nie spełnił jednak pokładanych w nim nadziei, gdyż nie cechował się lepszymi własnościami statystycznymi niż model podstawowy, a ponadto, wskutek uwzględnienia zmiennej „ocena subdzielniczy” wielkość próby znacząco zmalała: z 1451 do 688 obserwacji (por. Załącznik 1).

Zastosowano następujące oznaczenia:

M1 – model, z którego korzysta się w metodzie 0-1 wszystkich okresów,

M1+ – model, z którego korzysta się w metodzie 0-1 wszystkich okresów z dodatkowymi zmiennymi objaśniającymi,

M2 – modele, z których korzysta się w metodzie 0-1 sąsiednich okresów (11 modeli dla kolejnych par kwartałów),

M3 – modele, z których korzysta się w metodzie cen charakterystyk (12 modeli dla kolejnych kwartałów),

M4 – model, z którego korzysta się w metodzie imputacji.

Wszystkie zaakceptowane modele cechują się poprawną konstrukcją (popartą wynikiem testu RESET) oraz umiarkowaną współliniowością – czynniki inflacji wariancji nie przekraczających wartości 10. W celu umożliwienia poprawnego wnioskowania statystycznego mimo heteroskedastyczności składnika losowego stosujemy odporne estymatory błędów standardowych (HAC). Wprawdzie zmiany poziomu cen mogą wpływać na zmiany podaży charakterystyk mieszkania odzwierciedlonych w zmiennych objaśniających (tzw. problem endogeniczności), jednak kwestia ta nie dotyczy szacowanych przez nas modeli, gdyż wpływ cen na podaż zachodzi na rynkach mieszkaniowych z około dwuletnim opóźnieniem⁶. Dla modeli ze zmiennymi 0-1 czasu weryfikujemy przyjmowane w nich założenie o stałości wycen charakterystyk mieszkania (por. punkt 3.1). Test Chowa przeprowadzony na różnych próbach modelu M1 wskazuje na brak stabilności jego parametrów w większości badanych przypadków. W modelu M2 wyniki testu stabilności dla poszczególnych par kwartałów są mieszane, a okresy niestabilności na ogół można uzasadnić sytuacją rynkową.

Skrócone wyniki estymacji wszystkich modeli prezentujemy w Załączniku 2. Z modelu oszacowanego metodą imputacji (M4) wynika na przykład, że lokalizacja mieszkania w dobrej dzielnicy Warszawy podnosi cenę metra kwadratowego o około 29% w porównaniu z mieszkaniem o przeciętnej lokalizacji, natomiast cena metra kwadratowego mieszkania jedno- lub dwupokojowego jest o około 8% wyższa niż cena metra kwadratowego większych mieszkań, *ceteris paribus*; obie te zmienne mają statystycznie istotny wpływ na zmienną objaśnianą. Wyniki estymacji sugerują również, że mieszkania bardzo duże są droższe od mieszkań małych i średnich, gdyż współczynnik stojący przy zmiennej *sq_pow* jest dodatni i istotnie różny od zera. Z analizy innych współczynników regresji wynika, że mieszkania o wysokim standardzie wyposażenia są droższe od mieszkań o przeciętnym standardzie średnio o 9%. Niski standard zaniża cenę całkowitą o około 6% w stosunku do standardu przeciętnego, a mieszkania sprzedawane bez wyposażenia (w stanie surowym) są odpowiednio droższe o blisko 6%. Pozostałe znaki oszacowań parametrów oraz ich wielkości bezwzględne są również uzasadnione z merytorycznego punktu widzenia. Stosunkowo wysoki współczynnik determinacji modelu oznacza, że zmienne uwzględnione w modelu wyjaśniają nieco ponad 59% zmienności cen metra kwadratowego warszawskich mieszkań.

Podsumowując wyniki estymacji pozostałych modeli, warto zauważyć, że wszystkie cechują się stosunkowo wysokimi współczynnikami determinacji, wynoszącymi od 42,2% do 75,5%, przy czym w połowie modeli współczynniki te są wyższe od 60%. Także stabilne znaki oszacowań parametrów przy poszczególnych zmiennych wszystkich 26 modeli potwierdzają poprawność ich specyfikacji. Z drugiej strony porównanie modeli pod względem wartości cen implikowanych (współczynników regresji) poszczególnych atrybutów ujawnia często znaczne różnice. Dzieje się tak szczególnie w przypadku modeli szacowanych na podstawie danych pochodzących z jednego

⁶ Dziękujemy Recenzentowi za zwrócenie uwagi na ten problem.

lub dwóch okresów. Dalsze badania empiryczne powinny wykazać, czy różnice te poprawnie odzwierciedlają dynamiczne zmiany preferencji i wycen atrybutów mieszkania, czy też związane są ze strukturą danych (różne próby) lub wynikają z pominięcia ważnych zmiennych objaśniających. Przykładem ważnego atrybutu o dość wysokiej (choć nie najwyższej) zmienności wartości współczynnika jest lokalizacja. Usytuowanie mieszkania w dobrej dzielnicy podnosi cenę metra kwadratowego o około 29% i waha się w zależności od modelu między 17% a 34% w porównaniu z mieszkaniem o przeciętnej lokalizacji, *ceteris paribus*. Zmienna ta wykazuje istotnie statystyczny wpływ na zmienną objaśnianą we wszystkich modelach, a przeciętna wartość tego wpływu jest zgodna z oceną ekspercką. Inna zmienna, wyróżniająca małe (1- i 2-pokojowe) mieszkania okazuje się zmienną nieistotną w większości modeli szacowanych w celu obliczenia indeksu cen charakterystyk, ale znak i wartość odpowiadającego jej współczynnika regresji są zgodne z istotnymi wynikami w pozostałych modelach (szacowanych na większej liczbie okresów). Dominanta wartości ceny implikowanej tej zmiennej wynosi około 7%, co odpowiada intuicji eksperckiej; wartość oszacowania waha się między 3 a 9%. Spostrzeżenia dotyczące zmienności cen implikowanych mają ważne konsekwencje praktyczne i powinny być brane pod uwagę przy wyborze metody wyznaczania hedonicznego indeksu cen (por. część trzecia).

Wahania wartości oszacowanych współczynników nie są zaskakujące, gdyż modele oszacowano na podstawie różnej liczby obserwacji i dla różnych okresów. Teoretycznie rzecz ujmując, wyceńna poszczególnych cech mieszkania powinna być niezależna od liczby uwzględnionych obserwacji oraz okresu analizy. W praktyce jednak zróżnicowanie rynku mieszkań, dynamiczne zmiany rynkowe zachodzące w rozpatrywanym okresie oraz błędy pomiaru danych i efekty pominiętych zmiennych nie pozwalają na uzyskanie idealnego wyniku.

6.3. Hedoniczne indeksy cen

Wykresy 1–5 ukazują przebieg średniej dynamiki cen mieszkań sprzedanych na warszawskim rynku wtórnym w porównaniu z tradycyjnymi miarami zmian cen na rynku nieruchomości: średnią i medianą. Dynamika ta jest wyrażona za pomocą pięciu różnych indeksów hedonicznych oraz wskaźnika zmiany średniej ceny mieszkania i mediany. Indeksy prezentujemy w ujęciu kwartału do kwartału (Q/Q) oraz ze stałą podstawą (III kwartał 2006 r. = 100). Przebieg wszystkich siedmiu wskaźników jest bardzo podobny; dokładne porównanie wskaźników ujawnia pewne różnice.

Metodami, które dają najbardziej podobne wyniki, są metoda ze zmiennymi zero-jedynkowymi czasu wszystkich okresów (M1) i sąsiednich okresów (M2). Indeksy wyznaczone tymi dwiema metodami nie różnią się istotnie; podobnie zachowują się indeksy tego typu prezentowane w innych pracach (por. Li i in. 2006). Założenie stałości parametrów przyjmowane jest zarówno w metodzie M1, jak i M2, choć w metodzie M1 jest ono trudniejsze do spełnienia (co potwierdzają wyniki przeprowadzonych testów). Tym bardziej należy podkreślić podobieństwo dynamiki cen wyznaczonej na podstawie obydwu metod.

Jednocześnie na tle pozostałych wskaźników zmienność indeksu wyrażonego metodą 0-1 wszystkich okresów i sąsiednich okresów jest najmniejsza. Najwyższą zmiennością charakteryzuje się metoda cen charakterystyk (M3) oraz średnia i mediana. Warto również zauważyć, że indeks oparty na metodzie cen charakterystyk cechuje się dużo większą zmiennością w niestabilnym dla

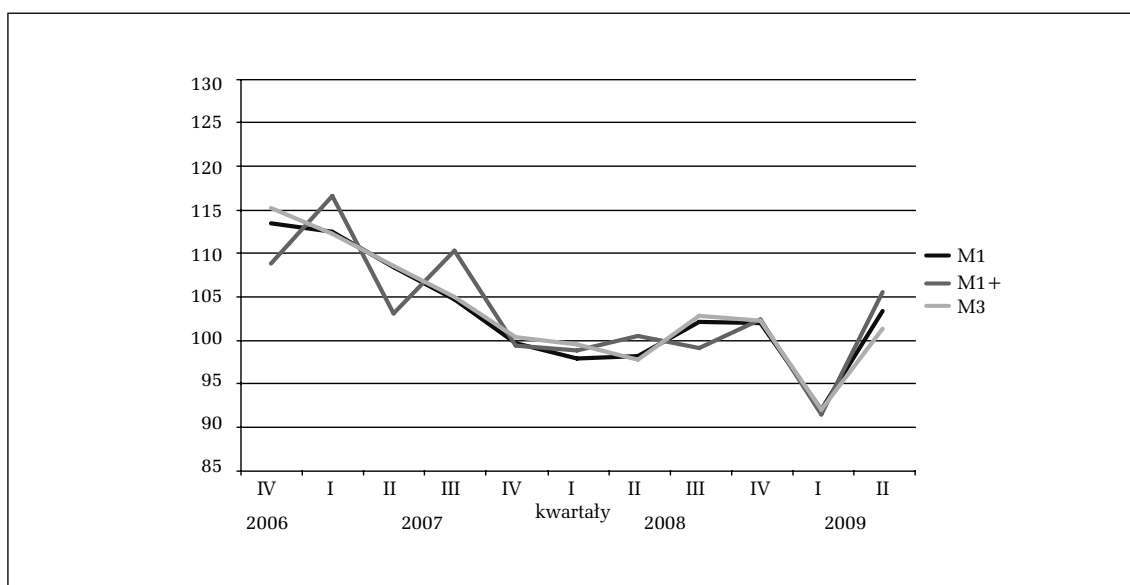
rynku mieszkaniowego okresie do III kwartału 2007 r. W okresie stosunkowo stabilnym wyniki są bardzo podobne do dwóch pozostałych indeksów hedonicznych. Zmienność indeksu wyznaczonego za pomocą metody imputacji (M4) jest stosunkowo najbardziej podobna do pierwszych dwóch indeksów, czyli dwóch wariantów metody ze zmiennymi zero-jedynkowymi czasu. Poprawa jakości estymowanych modeli (na przykład przez uwzględnienie dodatkowych zmiennych lokalizacyjnych lub efektów autokorelacji przestrzennej) prawdopodobnie wyeliminowałaby dodatkową część zmienności wskaźników.

Metoda 0-1 w wariancie rozszerzonym (z dodatkowymi zmiennymi objaśniającymi) powinna być analizowana w nieco innym wymiarze. Przypuszczamy, że powodem zróżnicowania wyników po zastosowaniu modelu rozszerzonego M1+ w porównaniu z modelem podstawowym M1 są głównie odmienne specyfikacje modeli. Indeksy z modeli M1 i M1+ kontrolują zmianę jakości innego zestawu atrybutów mieszkania, stąd indeks z modelu rozszerzonego odbiega od wartości pozostałych wskaźników. Indeks oparty na modelu M1+ wykazuje znacznie większą zmienność w okresie, w którym rynek mieszkaniowy cechował się niestabilnością (do III kwartału 2007 r.).

Analizując indeks o stałej podstawie (III kwartał 2006 r. = 100), obserwujemy konsekwentnie dodatnią dynamikę zmian cen. Najszybszy wzrost cen mieszkań notujemy w przypadku zastosowania indeksu wyznaczonego metodą 0-1 sąsiednich okresów; najwolniej rośnie wskaźnik z metody imputacji. W ostatniej metodzie jako ceny referencyjne przyjmujemy ceny implikowane wyznaczone na cały 2008 r. Testy stabilności przeprowadzone dla metody 0-1 sąsiadujących okresów wskazują na względną stabilność cen implikowanych w 2008 r. Przypuszczamy więc, że omawiane zróżnicowanie wartości indeksów wynika raczej z pominiętych zmiennych w modelach hedonicznych niż ze zmian strukturalnych na rynku.

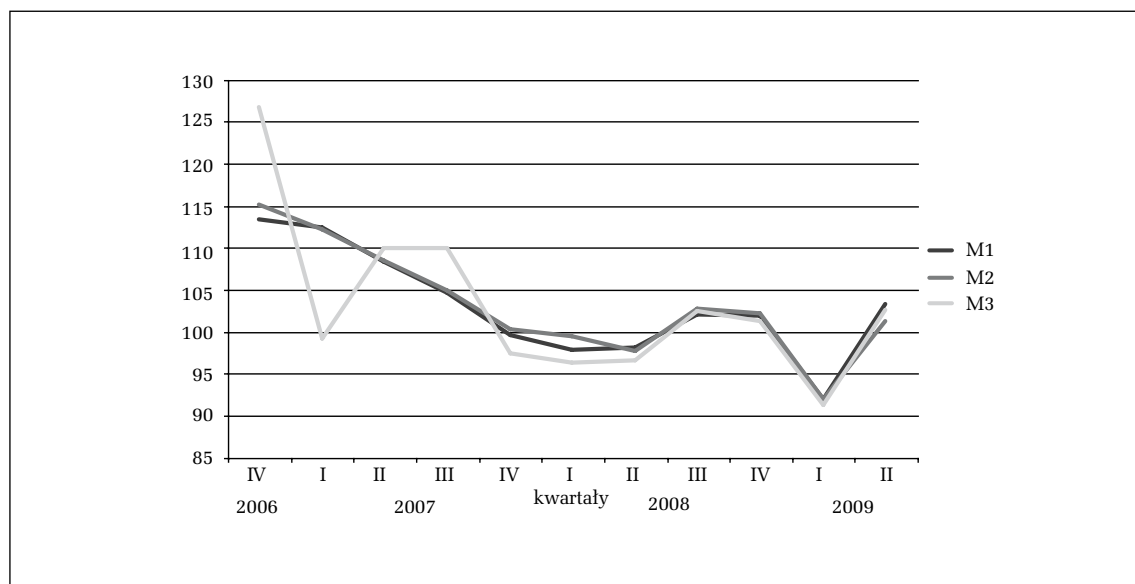
Wykres 1

Dynamika Q/Q, indeksy metod ze zmiennymi zero-jedynkowymi czasu: metoda wszystkich okresów (M1 i M1+) oraz metoda sąsiadujących okresów (M2)



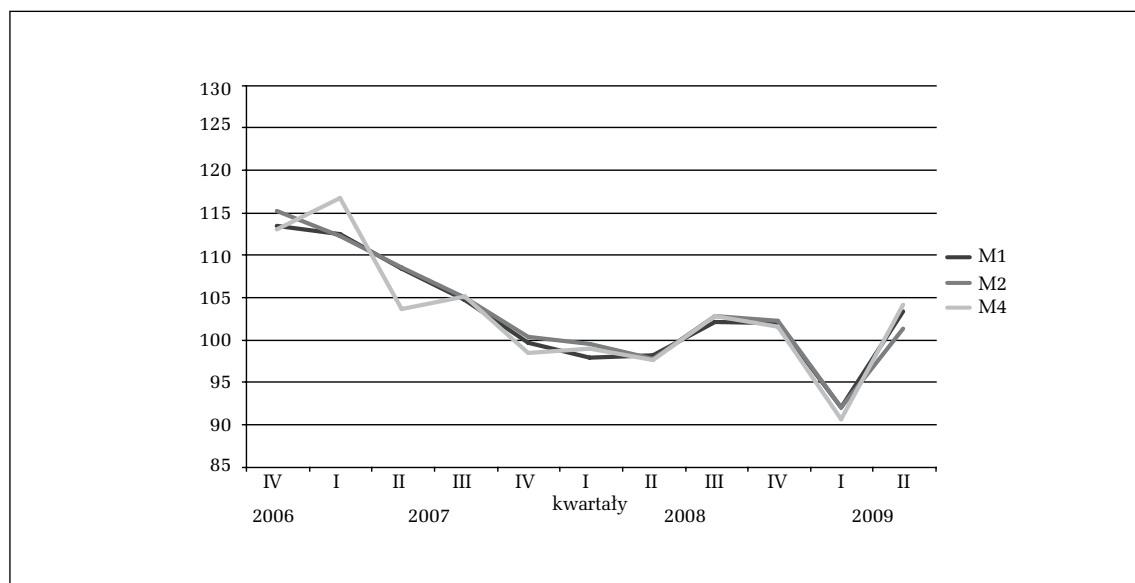
Wykres 2

Dynamika Q/Q, indeksy metod z zero-jedynkowymi zmiennymi czasu (M1 i M2) w porównaniu z metodą cen charakterystyk (M3)



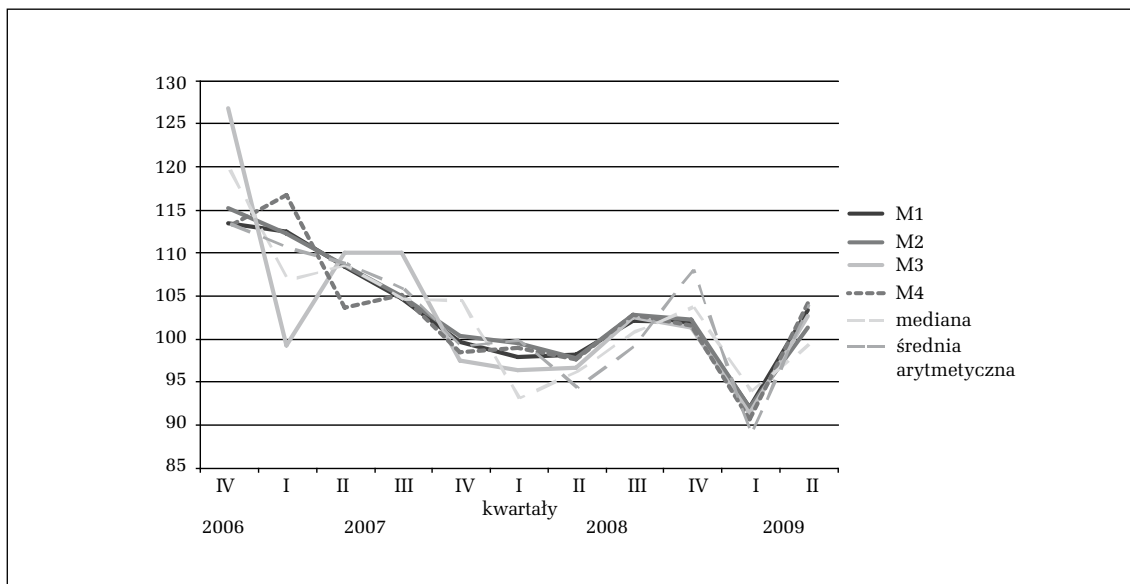
Wykres 3

Dynamika Q/Q, indeksy metod z zero-jedynkowymi zmiennymi czasu (M1 i M2) w porównaniu z metodą imputacji (M4)



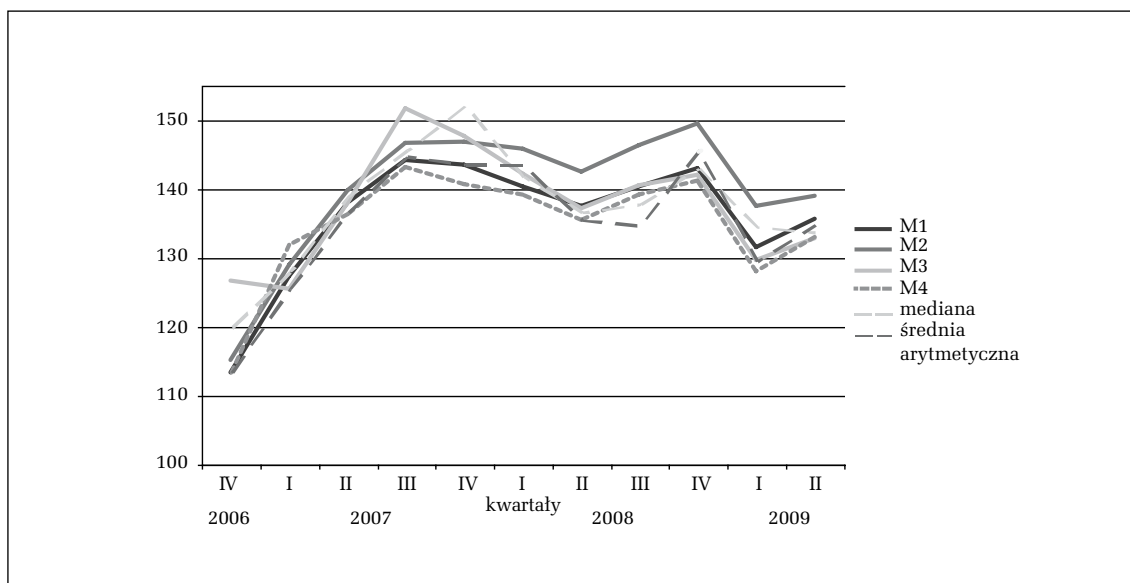
Wykres 4

Dynamika Q/Q, indeksy hedoniczne (M1, M2, M3, M4) w porównaniu z dynamiką mediany i średniej arytmetycznej



Wykres 5

Stała podstawa (baza: III kwartał 2006 r.), indeksy hedoniczne (M1, M2, M3, M4) w porównaniu z dynamiką mediany i średniej arytmetycznej



7. Wnioski

W niniejszej pracy zaprezentowano wyniki estymacji hedonicznych modeli cen nieruchomości mieszkaniowych i, po raz pierwszy w Polsce, wyniki konstrukcji hedonicznych indeksów cen mieszkań. Badanie zostało przeprowadzone dla warszawskiego rynku wtórnego mieszkań na podstawie danych transakcyjnych. W porównaniu z powszechnie stosowanymi metodami pomiaru dynamiki cen mieszkań – średnią i medianą – indeksy hedoniczne są miarą bardziej dokładną, gdyż kontrolują zmiany i różnice jakości mieszkań.

Duże zróżnicowanie oszacowań cen implikowanych w hedonicznych modelach cen sprawia, że konstrukcja indeksu metodą cen charakterystyk ma, naszym zdaniem, ograniczone zastosowanie na warszawskim rynku wtórnym mieszkań. Na podstawie porównania wyników estymacji wszystkich modeli stwierdzamy, że problem ten dotyczy w mniejszym stopniu dwóch pozostałych metod konstrukcji hedonicznego indeksu cen, czyli metody imputacji i metody ze zmiennymi zero-jedynkowymi czasu. Wyniki uzyskane dla warszawskiego wtórnego rynku mieszkań przemawiają na korzyść tych dwóch metod.

Poza prezentacją metod konstrukcji i wyników hedonicznych indeksów cen mieszkań zasygnalizowaliśmy szereg zagadnień teoretyczno-praktycznych, które mogą stanowić punkt wyjścia dla przyszłych prac badawczych. Należą do nich m.in.:

- tworzenie i ulepszanie baz danych o rynku nieruchomości w Polsce i poprawa jakości prezentowanych modeli hedonicznych,
- stworzenie analogicznych wskaźników cen dla rynku pierwotnego mieszkań oraz dla danych ofertowych,
- zastosowanie pośrednich metod konstrukcji hedonicznych indeksów cen mieszkań,
- prace nad określeniem funkcji popytu i podaży na rynku mieszkań oraz wnioski, jakie mogą stąd płynąć dla specyfikacji modeli hedonicznych,
- zastosowanie bardziej zaawansowanych metod ekonometrycznych, takich jak metoda zmiennych instrumentalnych, pozwalająca wyeliminować skutki pominięcia ważnych zmiennych objaśniających, oraz modele uwzględniające autokorelację przestrzenną.

W naszej ocenie hedoniczne modele cen stanowią użyteczne i bardzo obiecujące narzędzie modelowania cen mieszkań. Oprócz zastosowania w szeroko pojętych analizach rynków nieruchomości mogą stać się głównym elementem procesu włączania cen mieszkań do wskaźnika cen towarów i usług konsumpcyjnych. Mamy nadzieję, że ten artykuł zapoczątkuje prace i wzbogaci dyskusje na temat udoskonalenia statystyki cenowej polskiego rynku nieruchomości.

Bibliografia

- Bazyl M. (2007), *Hedonistyczny model cen mieszkań w Warszawie*, praca magisterska, Instytut Ekonometrii SGH, Warszawa.
- Berndt E. R. (1991), *The Practice of Econometrics*, Addison-Wesley, Reading.
- Bover O., Izquierdo M. (2001), *Ajustados de calidad en los precios: metodos hedonicos y consecuencias para la contabilidad nacional*, Economic studies, 70, Banco de Espana, Madrid.

- Brachinger H. W. (2002), *Statistical theory of hedonic price indices*, Seminar of Statistics, DQE Working Paper, 1, Department of Quantitative Economics, University of Fribourg.
- Butsic V., Hanak E., Valletta R. (2009), *Climate Change and Housing Prices: Hedonic Estimates for North American Ski Resorts*, Federal Reserve Bank of San Francisco Working Papers Series, San Francisco.
- Conniffe D., Duffy D. (1999), Irish house price indices – methodological issues, *Economic and Social Review*, 30, 403–423.
- Diewert E. (2003a), Hedonic Regressions A Consumer Theory Approach, w: C.R. Feenstra, M.D. Shapiro (red.), *Scanner Data and Price Indexes*, NBER Book Series in Income and Wealth, University of Chicago Press, 317–348.
- Diewert E. (2003b), *The treatment of owner occupied housing and other durables in a Consumer Price Index*, Discussion Paper, 03-08, Department of Economics, University of British Columbia, Vancouver.
- Diewert E. (2007), *Hedonic imputation versus time dummy hedonic indexes*, Discussion Paper, 007-07, Department of Economics, University of British Columbia, Vancouver.
- Duffy D. (2001), *The use of location variables in a mix-adjusted index for Dublin house prices*, Working Paper, WP138, Economic and Social Research Institute, Dublin.
- Dziechciarz J. (2005), *Regresja hedoniczna i metoda conjoint measurement jako narzędzia w kształtowaniu polityki cenowej*, Prace Naukowe Akademii Ekonomicznej, 1096, Wrocław.
- Eiglsperger M. (2006), *Residential property price statistics for the euro area and selected EU countries*, prezentacja z konferencji “Real Estate Price Indexes”, OECD – IMF, 6-7 November, Paris.
- Fletcher M., Gallimore P., Mangan J. (2000), Heteroscedasticity in hedonic house price models, *Journal of Property Research*, 17, 93–108.
- Gouriéroux C., Laferrère A. (2006), *Managing hedonic housing price indexes: the French experience*, OECD – IMF Workshop Paper, 12, Paris.
- ILO (2004), *Consumer Price Index Manual: Theory and Practice*, International Labour Organisation, Geneva.
- Li W., Prud'homme M., Yu K. (2006), Studies in Hedonic Resale Housing Price Indexes, artykuł zaprezentowany na konferencji “The Canadian Economic Association 40th Annual Meetings”, Montreal.
- Łaszek J.A., Widłak M. (2007), Indeksy rynkowych cen mieszkań jako narzędzie badania rynku, *Finansowanie Nieruchomości*, 3 (12), 4–11.
- Łaszek J.A., Widłak M. (2008), Badanie cen na rynku mieszkań prywatnych zamieszkałych przez właściciela z perspektywy banku centralnego, *Bank i Kredyt*, 8, 12–41.
- Maddala G. S. (2006), *Ekonometria*, PWN, Warszawa.
- Malpezzi S. (2002), *Hedonic pricing models: a selective and applied review*, The Center for Urban Land Economics Research, The University of Wisconsin.
- Nellis J., Figueira-Theodorakopoulou C. (2005), The importance of accurate house price measurement in the context of mortgage risk management, w: Z. Krysiak (red.), *Credit Risk of mortgage loans. Modeling and management*, ZBP, Warszawa.
- Nelson J. (2004), Meta-analysis of airport noise and hedonic property values: problems and prospects, *Journal of Transport Economics and Policy*, 38 (1), 1–24.

- Rosen S. (1974), Hedonic prices and implicit markets: product differentiation in pure competition, *Journal of Political Economy*, 1, 34–55.
- S&P (2008), *Case – Shiller Home Price Indices, index methodology*, <http://www.scribd.com/doc/14548164/Case-Shiller-Housing-Index-Method>.
- Theebe M.A.J. (2004), Planes, trains, and automobiles – the impact of traffic noise on house prices, *Journal of Real Estate Finance and Economics*, 28, 209–234.
- Triplett J.E. (2006), *Handbook on Hedonic Indexes and Quality Adjustments in Price Indexes*, OECD, Paris.
- Trojanek R. (2008), *Wahania cen na rynku mieszkaniowym*, Wydawnictwo Akademii Ekonomicznej w Poznaniu.
- Verbeek M. (2004), *A Guide to Modern Econometrics*, John Wiley & Sons Ltd, Chichester.
- Widłak M. (2010), *Hedoniczne indeksy cen – podstawowe metody wyznaczania*, wersja robocza opracowania niepublikowanego, Warszawa.
- Winstrand J. (2008), *The price of a view*, artykuł zaprezentowany na konferencji European Network for Housing Research, Dublin.
- Wooldridge J.M. (2002), *Econometric Analysis of Cross Section and Panel Data*, MIT Press, Cambridge.

Załącznik 1

Liczba obserwacji w poszczególnych modelach po eliminacji obserwacji odstających i braków danych

Metoda	Liczba obserwacji		
	początkowa	po eliminacji obserwacji nietypowych	po eliminacji braków danych
0-1 wszystkich okresów (M1)	1494	1418	1195
0-1 wszystkich okresów (M1+)	1494	1451	688
0-1 sąsiednich okresów (M2):			
baza 20063	327	311	263
baza 20064	343	316	256
baza 20071	304	280	245
baza 20072	268	253	219
baza 20073	235	220	184
baza 20074	217	202	170
baza 20081	271	257	219
baza 20082	233	219	188
baza 20083	193	177	148
baza 20084	184	169	142
baza 20091	164	154	117
cen charakterystyk (M3):			
20063	163	154	142
20064	164	158	122
20071	179	166	142
20072	125	120	109
20073	143	137	114
20074	92	85	72
20081	125	121	102
20082	146	138	119
20083	87	80	68
20084	106	101	84
20091	78	72	62
20092	86	83	56
imputacji (M4)	464	446	379

Źródło: obliczenia na podstawie bazy BaRN.

Załącznik 2

Wyniki estymacji hedonicznych modeli cen mieszkań: oszacowania parametrów

Metoda 0-1 wszystkich okresów i metoda imputacji			
	M1	M1+	M4
const	8,667	8,699	8,749
stand_1	0,078	0,068	0,087
stand_3	-0,080	-0,070	-0,064
stand_4	0,042	0,046	0,057
ocena_d_3	0,285	0,277	0,292
rok_1	0,185	0,105	--
rok_2	0,114	0,072	0,089
rok_3	0,026	-0,006	-0,079
rok_4	-0,079	-0,110	-0,152
rok_5	-0,125	-0,117	-0,136
rok_7	0,101	0,091	0,052
rok_8	0,122	0,073	0,124
rok_nowe	-0,110	-0,090	-0,156
spoldzielcze	-0,036	-0,052	-0,034
ilepokoi_1_2	0,073	0,070	0,080
sq_pow	0,000	0,000	0,000
T_20063	-0,273	-0,267	--
T_20064	-0,148	-0,183	--
T_20071	-0,031	-0,030	--
T_20072	0,048	-0,001	--
T_20073	0,093	0,096	--
T_20074	0,088	0,088	--
T_20081	0,065	0,075	--
T_20082	0,045	0,079	--
T_20083	0,065	0,069	--
T_20084	0,084	0,091	--
T_20092	0,032	0,052	--
winda_1	--	0,027	--
ilekondygnacji	--	-0,006	--
ocena_s_1	--	-0,108	--
ocena_s_3	--	0,031	--
R ²	0,632	0,619	0,593

Uwagi: wytłuszczono parametry nieróżniące się od zera na poziomie istotności 5%.

Brak oszacowania parametru przy niektórych zmiennych opisujących cechy mieszkań spowodowany jest strukturą próbek, z których nie każda zawiera obserwacje odpowiadające każdej z kategorii zmiennych 0-1.

Źródło: obliczenia na podstawie bazy BaRN.

Metoda 0-1 sąsiednich okresów (M2)											
Zmienna	2006		2007			2008			2009		
	IV	I	II	III	IV	I	II	III	IV	I	II
const	8,490	8,513	8,598	8,767	8,828	8,775	8,721	8,702	8,805	8,849	8,763
stand_1	0,123	0,130	0,122	0,050	0,003	0,090	0,111	0,078	0,056	0,024	0,005
stand_3	-0,030	-0,019	-0,054	-0,107	-0,136	-0,133	-0,061	-0,065	-0,073	-0,062	-0,093
stand_4	--	--	0,099	0,102	-0,028	0,028	0,138	0,065	-0,042	-0,118	-0,106
ocena_d_3	0,288	0,278	0,210	0,226	0,258	0,347	0,335	0,283	0,227	0,280	0,313
rok_1	0,235	0,225	0,142	0,064	0,016	--	--	--	--	--	--
rok_2	0,123	0,124	0,154	0,138	0,063	-0,002	0,119	0,169	0,052	0,030	0,087
rok_3	0,004	0,080	0,144	0,063	0,052	-0,081	-0,093	-0,032	-0,045	-0,034	-0,061
rok_4	-0,151	-0,089	0,003	0,025	-0,021	-0,164	-0,155	-0,132	-0,160	-0,174	-0,185
rok_5	-0,182	-0,175	-0,081	-0,075	-0,081	-0,110	-0,111	-0,129	-0,155	-0,148	-0,220
rok_7	0,049	0,137	0,176	0,071	0,074	0,009	0,018	0,075	0,078	0,099	0,096
rok_8	0,153	0,205	-0,021	-0,129	--	0,003	0,052	0,128	0,175	0,128	0,101
rok_nowe	-0,096	-0,170	--	--	--	--	-0,129	-0,162	-0,188	-0,119	-0,148
spoldzielcze	-0,034	0,002	-0,022	-0,092	-0,073	-0,070	-0,065	-0,014	-0,007	-0,014	0,009
ilepokoi_1_2	0,028	0,065	0,087	0,072	0,040	0,068	0,089	0,075	0,066	0,028	0,040
sq_pow	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
T_20064	0,141	--	--	--	--	--	--	--	--	--	--
T_20071	--	0,114	--	--	--	--	--	--	--	--	--
T_20072	--	--	0,081	--	--	--	--	--	--	--	--
T_20073	--	--	--	0,047	--	--	--	--	--	--	--
T_20074	--	--	--	--	0,002	--	--	--	--	--	--
T_20081	--	--	--	--	--	-0,006	--	--	--	--	--
T_20082	--	--	--	--	--	--	-0,024	--	--	--	--
T_20083	--	--	--	--	--	--	--	0,026	--	--	--
T_20084	--	--	--	--	--	--	--	--	0,022	--	--
T_20091	--	--	--	--	--	--	--	--	--	-0,084	--
T_20092	--	--	--	--	--	--	--	--	--	--	0,011
R ²	0,577	0,507	0,485	0,507	0,501	0,578	0,694	0,702	0,620	0,613	0,661

Uwagi: wyłuszczone parametry nieróżniące się od zera na poziomie istotności 5%.

Brak oszacowania parametru przy niektórych zmiennych opisujących cechy mieszkań spowodowany jest strukturą próbek, z których nie każda zawiera obserwacje odpowiadające każdej z kategorii zmiennych 0-1.

Źródło: obliczenia na podstawie bazy BaRN.

Metoda cen charakterystyk (M3)												
Zmienna	2006		2007				2008				2009	
	III	IV	I	II	III	IV	I	II	III	IV	I	II
const	8,440	8,688	8,522	8,710	8,819	8,809	8,690	8,663	8,836	8,904	8,663	8,823
stand_1	0,070	0,167	0,115	0,126	-0,010	0,059	0,162	0,114	0,014	0,024	0,034	0,042
stand_3	-0,063	-0,127	-0,006	-0,071	-0,097	-0,203	-0,070	-0,041	-0,091	-0,022	-0,117	-0,100
stand_4	--	--	--	0,137	0,145	-0,027	0,261	0,108	0,088	-0,102	-0,090	-0,066
ocena_d_3	0,293	0,276	0,249	0,212	0,235	0,323	0,300	0,311	0,170	0,228	0,299	0,248
rok_1	0,297	0,055	--	0,362	-0,008	--	--	--	--	--	--	--
rok_2	0,116	0,207	0,138	0,179	0,099	-0,049	0,075	0,198	0,145	-0,117	-0,025	0,149
rok_3	0,041	-0,064	0,174	0,099	0,042	0,037	-0,081	-0,005	-0,099	-0,079	0,052	-0,083
rok_4	-0,132	-0,134	-0,026	0,051	0,008	-0,103	-0,112	-0,090	-0,224	-0,212	-0,042	-0,193
rok_5	-0,133	-0,218	-0,082	-0,075	-0,072	-0,104	-0,034	-0,083	-0,314	-0,150	-0,091	-0,216
rok_7	0,105	-0,031	0,236	0,081	0,076	0,045	0,131	0,066	0,057	0,072	0,153	0,004
rok_8	0,077	0,665	-0,092	-0,129	--	--	0,003	0,122	0,054	0,303	0,107	0,163
rok_nowe	-0,022	-0,167	--	--	--	--	-0,049	-0,159	-0,183	-0,281	-0,150	-0,118
spoldzielcze	-0,017	-0,102	0,026	-0,103	-0,093	-0,038	-0,047	-0,044	0,105	-0,009	-0,032	0,025
ilepokoi_1_2	0,033	0,079	0,074	0,106	0,068	0,020	0,092	0,080	0,060	0,018	0,078	0,055
sq_pow	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
R ²	0,545	0,602	0,422	0,581	0,502	0,627	0,685	0,751	0,730	0,565	0,622	0,755

Uwagi: wytluszczone parametry nieróżniące się od zera na poziomie istotności 5%.

Brak oszacowania parametru przy niektórych zmiennych opisujących cechy mieszkań spowodowany jest strukturą podprób, z których nie każda zawiera obserwacje odpowiadające każdej z kategorii zmiennych 0-1.

Źródło: obliczenia na podstawie bazy BaRN.

Hedonic price index for Warsaw housing market

Abstract

Reliable evaluation of price trends in housing markets emerges as a relatively new area of Polish statistics, lacking in empirical applications. Similarly, quality change adjustments in price indices and the use of hedonic models for this purpose have not been given sufficient attention. The article presents estimates of hedonic price models and hedonic price indices constructed on the basis of transactions concluded in the Warsaw secondary market. Three variants of the direct method are used to build five hedonic price indices, which in contrast to the commonly used measures of housing price growth – the mean and the median – make it possible to account for changes in the quality of sold apartments. Theoretical aspects of estimation and interpretation of hedonic models are also discussed alongside their applications as a tool of quality change adjustment in housing price indices. Both the theoretical and empirical aspects of hedonic price index applications presented in the article offer a new and promising direction for Polish housing price statistics.

Keywords: price indices, hedonic models, housing market

