

Uniwersytet Warszawski
Wydział Nauk Ekonomicznych

Mikołaj Rymajdo
Nr albumu: 432511

**Charakterystyka nieruchomości mieszkaniowych, a
ich ceny ofertowe - analiza rynku warszawskiego na
podstawie ofert z serwisu Otodom**

Praca zaliczeniowa
Programowanie narzędzi analitycznych

Praca wykonana pod kierunkiem
dr Janusza Gajdy
z Katedry Statystyki i Ekonometrii
WNE UW

Warszawa, styczeń 2024

WSTĘP

Nieruchomość to dobro, które wyróżnia się na tle innych ze względu na kluczową rolę jaką odgrywa w życiu ludzkim. Zapewnia ono podstawowe potrzeby ludzkie takie jak schronienie, poczucie bezpieczeństwa i prywatność. Prawie wszystkie mieszkania, bez względu na jego cechy, są w stanie spełnić podstawowe potrzeby ludzkie, a więc konsumenci przy zakupie nieruchomości mieszkaniowych kierują się głównie preferencjami i ograniczeniem budżetowym. Preferencje co do mieszkania dotyczą głównie jego cech fizycznych, w tym lokalizacji. Na kształtowanie się ceny nieruchomości wpływa wiele czynników ekonomicznych, jednak można przekrojowo dokonać wyceny nieruchomości wykorzystując ich cechy charakterystyczne. Dobra heterogeniczne, takie jak nieruchomości, są agregatem cech, które mają swoją hierarchię atrakcyjności i poprzez preferencje wpływają na cenę dobra. Przy pomocy modelu ekonometrycznego można wycenić przybliżoną wartość poszczególnych cech nieruchomości dla konsumentów, dokonując w ten sposób rozwarstwienia ceny dobra na wartość czynników, z których jest zbudowane. Celem pracy jest zbadanie determinantów cen nieruchomości w Warszawie na rynku wtórnym przy pomocy modelu ekonometrycznego, wykorzystując do tego dane ofertowe pozyskane metodą web-scrapingu z serwisu <https://www.otodom.pl/>, ujawniając w ten sposób cechy nieruchomości mieszkaniowych, które w największym stopniu wpływają na cenę jednego metra kwadratowego nieruchomości. Temat pracy jest ważny ze względu na ciągle zmieniający się rynek mieszkaniowy. W 2022 roku wystąpił potężny napływ nowych konsumentów ze wschodnich krajów do Polski, który wpłynął na zmiany preferencji i stworzył zapotrzebowanie na przeprowadzenie aktualnych badań dotyczących atrakcyjności poszczególnych charakterystyk mieszkań na rynku wtórnym w Warszawie. Dodatkowo w ciągu ostatnich lat pandemia, zmienna sytuacja gospodarcza w kraju oraz Bezpieczny Kredyt 2% miał istotny wpływ na kształtowanie się popytu i podaży nieruchomości, a więc liczba oddawanych mieszkań do użytkowania i dostępność kredytów uległa zmianie.

1. Zbiór danych i zmienne wykorzystane w badaniu

Aby dokonać przekrojowej analizy determinantów cen nieruchomości mieszkaniowych, wykorzystano dane ze strony <https://otodom.pl>, która jest jednym z największych serwisów ogłoszeniowych nieruchomości w Polsce. Ogłoszenia na Otodom.pl zawierają ceny ofertowe oraz wiele informacji dotyczących charakterystyk nieruchomości. Na potrzeby badania wykonanego w niniejszej pracy dane zostały pozyskane z wykorzystaniem metody web-scrapingu w dniu 15.12.2023 r. Uzyskane dane po ograniczeniach zawierają 3213 pozycji. Na początku dane zostały ograniczone do mieszkań na rynku wtórnym. Następnie, w przypadku zmiennej dotyczącej statusu prawnego własności, usunięto wszystkie obserwacje dotyczące mieszkań współdzielonych ze względu na niewielką liczbę obserwacji, tym samym pozostawiając jedynie następujące formy własności: „spółdzielcze własnościowe” i „pełna własność”. Brakujące dane dotyczące własności zostały uzupełnione jako „pełna własność”, jeżeli budynek został wybudowany po 2007 roku. Wynika to z tego, że od 2007 r. przepisy zabraniają ustanawiania własnościowego prawa spółdzielczego do lokalu.¹ Pozostałe wybrakowane dane zostały również usunięte. W modelu bazowym wykorzystano cenę ofertową jako zmienną zależną i 13 zmiennych niezależnych, co ilustruje zamieszczona dalej Tabela 1. Zmienne określające odległości od centrum i od metra zostały obliczone z wykorzystaniem współrzędnych geograficznych odzwierciedlając najkrótszą drogę biegnącą po powierzchni sferycznego modelu ziemi. Promień kuli założony w obliczeniach wynosi

¹ Ustawa z dnia 14 czerwca 2007 r. - o zmianie ustawy o spółdzielniach mieszkaniowych oraz o zmianie niektórych innych ustaw (Dz. U. 2007 nr 125 poz. 873).

6371.009 km, a więc tyle co średni promień planety według standardu WGS-84². Współrzędne mieszkań zostały zebrane z map zawartych w ogłoszeniach. Za punkt określający centrum ustanowiono Pałac Kultury i Nauki, którego współrzędne, wraz ze współrzędnymi stacji metra, pochodzą z Google Maps.

Tabela 1. Opis zmiennych wykorzystanych w modelu

Nazwa zmiennej	Objaśnienie	Miara	Typ zmiennej:
<i>cena_m2</i>	Cena ofertowa mieszkania za m ²	PLN	Ciągła
<i>wiek_budynku</i>	Wiek budynku wyrażony w latach	Lata	Dyskretna
<i>powierzchnia</i>	Powierzchnia użytkowa mieszkania	Metry kwadratowe	Ciągła
<i>pietro</i>	Piętro, na którym znajduje się mieszkanie	Liczba	Dyskretna
<i>pietro_max</i>	Wysokość budynku liczona w piętrach	Liczba	Dyskretna
<i>odleglosc_metro</i>	Odległość mieszkania od najbliższej stacji metra	Kilometry	Ciągła
<i>odleglosc_centrum</i>	Odległość mieszkania od Pałacu Kultury i Nauki	Kilometry	Ciągła
<i>agencja</i>	Zmienna określająca czy ogłoszenie zostało wystawione przez agencję nieruchomości	0 – osoba prywatna 1 – agencja nieruchomości	Binarna
<i>balkon</i>	Zmienna określająca czy mieszkanie zawiera balkon	0 – nie 1 – tak	Binarna
<i>winda</i>	Zmienna określająca czy budynek jest wyposażony w windę	0 – nie 1 – tak	Binarna
<i>spoldzielcze</i>	Zmienna określająca czy jest to mieszkanie spółdzielcze własnościowe	0 – pełna własność 1 – spółdzielcze własnościowe	Binarna
<i>remont</i>	Zmienna określająca czy mieszkanie wymaga remontu	0 – nie 1 – tak	Binarna
<i>wykonczenie</i>	Zmienna określająca czy mieszkanie wymaga wykończenia	0 – nie 1 – tak	Binarna
<i>taras</i>	Zmienna określająca czy mieszkanie zawiera taras	0 – nie 1 – tak	Binarna

Źródło: Opracowanie własne na podstawie danych z <https://otodom.pl>.

Dokładne statystyki opisowe zmiennych przedstawiono w Tabeli 2. Nieruchomości w stolicy są wyjątkowo drogie, średnia cena ofertowa za m² w badanej próbie wyniosła aż

² National Geospatial-Intelligence Agency (NGA) Standardization Document, 2014, *Department of Defense World Geodetic System 1984: Its Definition and Relationships with Local Geodetic Systems*. Dostęp 12.2023 r. <https://nsgreg.nga.mil/doc/view?i=4085>

18053,28 zł. Dodatkowo mieszkania w Warszawie nie są zbyt duże, ponieważ średnia powierzchnia użytkowa wyniosła 66.19 m², a najmniejsze mieszkanie ma tylko 11,28 m². Obecnie zgodnie z polskim prawem mieszkanie powinno mieć powierzchnię użytkową nie mniejszą niż 25 m², aczkolwiek przed 2018 r. zapis ten nie był tak restrykcyjny, a także można dalej go obejść sprzedając mieszkanie jako lokal użytkowy.³ Linie metra wydają się być dobrze zoptymalizowane, ponieważ średnia odległość nieruchomości od najbliższej stacji metra wynosi zaledwie 2,34 km. Większość mieszkań wyposażona jest w balkon (65.70%) oraz windę (69.65%), aczkolwiek taras przylega tylko do 15.66% z nich. Niewiele nieruchomości jest w drugorzędnym stanie, ponieważ tylko 8.71% z nich wymaga remontu, a wykończenia potrzebuje 9.55%. Zaskakująco dużo mieszkań sprzedawanych jest przez agencje nieruchomości, gdyż wskaźnik ten wyniósł w badanej próbie 76.53%.

Tabela 2. Statystyki opisowe zmiennych wykorzystanych w modelu

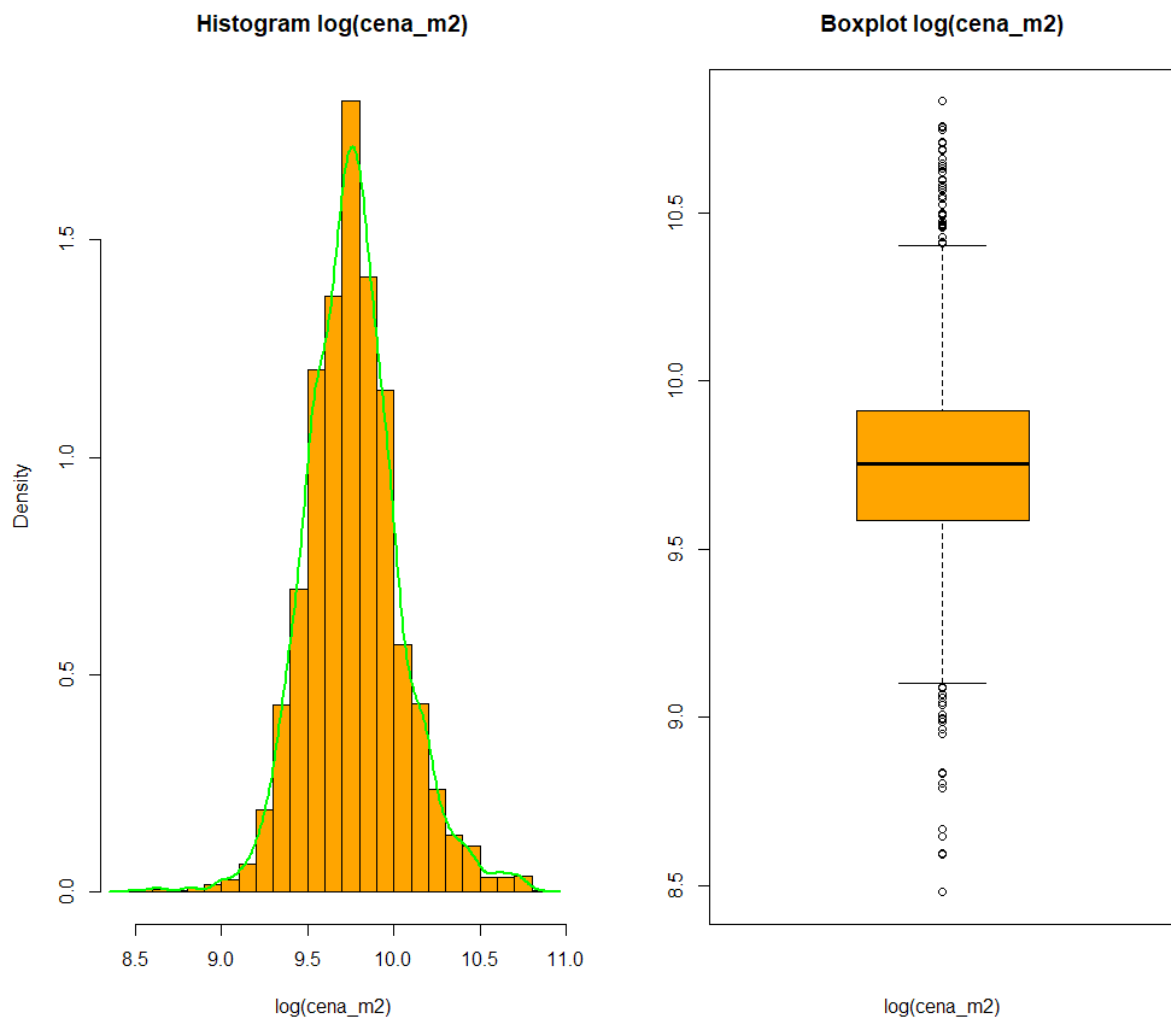
Nazwa zmiennej	Średnia	Odchylenie standardowe	Minimum	Maksimum
cena_m2	18053,28167	5321,384233	4821	50595
powierzchnia	66,187893	38,829953	11,28	487,75
wiek_budynku	1998,792406	22,985718	1944	2023
pietro	2,838469	2,341565	0	10
pietro_max	5,833126	3,313799	1	30
winda	0,696545	0,459821	0	1
balkon	0,657018	0,474779	0	1
taras	0,156552	0,363434	0	1
wspolrzedne_szerokosc	52,224488	0,048854	52,1046	52,362261
wspolrzedne_wysokosc	21,007044	0,067322	20,856345	21,25031
odleglosc_metro	2,344504	2,137635	0,023278	14,459002
odleglosc_centrum	6,382423	3,23092	0,189656	16,683793
agencja	0,765328	0,423859	0	1
spoldzielcze	0,101463	0,301988	0	1
remont	0,087146	0,282093	0	1
wykonczenie	0,095549	0,294018	0	1

Źródło: Opracowanie własne na podstawie danych z: <https://www.otodom.pl/>

Podjęto decyzję o zastosowaniu transformacji logarytmicznej zmiennej objaśnianej. W przypadku występowania prawostronnej skośności rozkładu danych, takie działanie redukuje wpływ wartości odstających na wyniki estymacji, a rozkład zmiennej staje się bliższy rozkładowi normalnemu. Co więcej, taki krok może niwelować heteroskedastyczność składnika losowego, poprzez zbliżenie wysokich wartości cenowych, które z reguły charakteryzują się wyższą wariancją, do środka rozkładu. Histogram zmiennej zależnej i wykres pudełkowy przedstawiono na rysunku 1. Boxplot wskazuje na występowanie licznych obserwacji odstających, a histogram pokazuje wysokie skoncentrowanie obserwacji w okolicach średniej. Wykonany test Jarque-Bera pozwolił na odrzucenie H₀ o normalności rozkładu, gdyż p-value wyniosło 0.00.

³ Rozporządzenie Ministra Infrastruktury z dnia 12 kwietnia 2002 r. - w sprawie warunków technicznych, jakim powinny odpowiadać budynki i ich usytuowanie (Dz.U. 2022 poz. 1225).

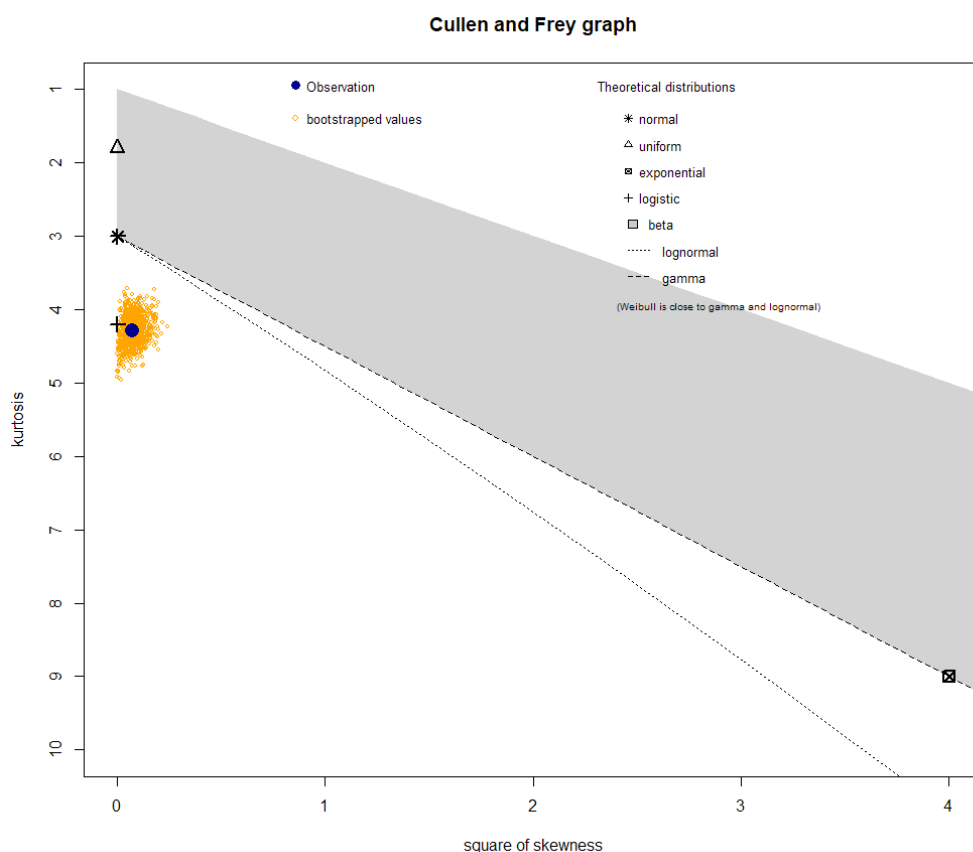
Rys. 1. Histogram i wykres pudełkowy zmiennej zależnej



Źródło: Opracowanie własne w R

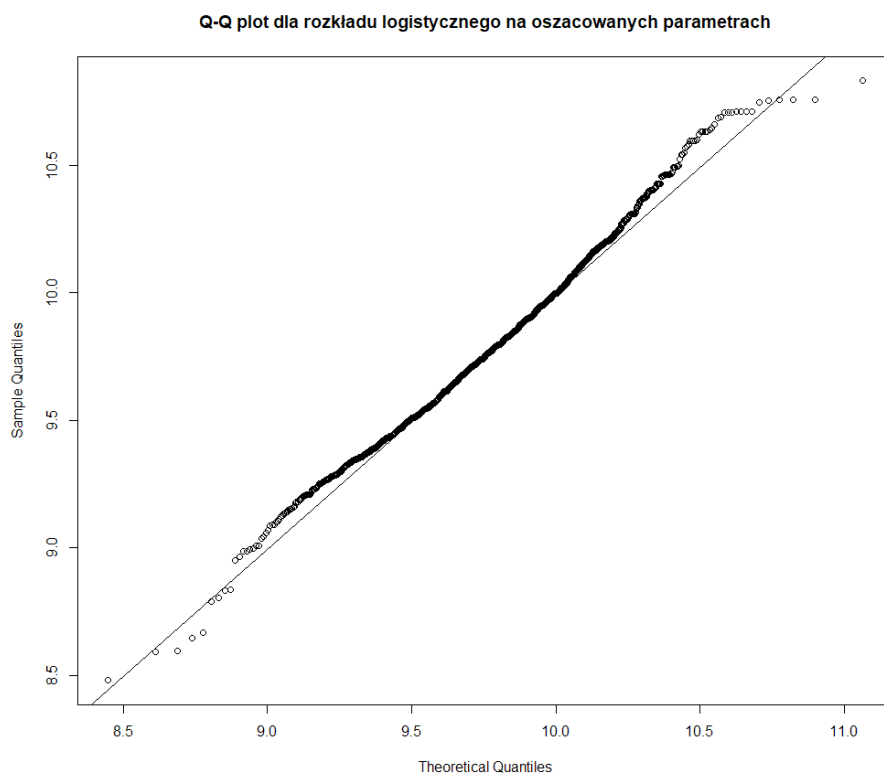
Wykonano wykres Cullena i Freya, aby odnaleźć właściwy rozkład (Rys. 2). Kurtoza i kwadrat skośności dla naszej zmiennej (niebieska kropka) osiąga wartości zbliżone do wartości teoretycznych rozkładu logistycznego. Wprowadzenie niepewności w parametrach opisujących rozkład danych empirycznych (żółte kropki) nie przyniosło istotnych odchyleń dla tysiąca symulacji. Oznacza to, że niewielkie zmiany wartości pojedynczych obserwacji, a tym samym oszacowanej średniej i wariancji z danych nie wpływają na celność dopasowania do rozkładu logistycznego. Wyestymowano właściwe parametry rozkładu logistycznego dla danych empirycznych poprzez maksymalizację funkcji wiarygodności. Oszacowane parametry są równe: $location = 9.7559653$ oraz $scale = 0.1491862$. Wykonano Q-Q plot dla teoretycznego rozkładu logistycznego na powyższych parametrach (Rys. 3). Dopasowanie nie jest idealne, lecz nie ma bardzo dużych odchyleń i całkiem dobrze wydaje się odzwierciedlać rozkład zmiennej.

Rys. 2. Wykres Cullena i Freya dla zmiennej zależnej



Źródło: Opracowanie własne w R

Rys. 3. Q-Q plot dla teoretycznego rozkładu logistycznego na oszacowanych parametrach



Źródło: Opracowanie własne w R

2. Badanie

Wykonano estymację modelu liniowego na trzy sposoby: Metodą Najmniejszych Kwadratów, Metodą Największej Wiarygodności i Metodą Momentów. Postać równania wykorzystana w badaniu jest równa:

$$\ln(cena_m2_i) = \beta_0 + \beta_1wiek_budynku + \beta_2winda_i + \beta_3powierzchnia_i + \beta_4pietro_i + \beta_5pietro_max_i + \beta_6balkon_i + \beta_7odleglosc_metro_i + \beta_8odleglosc_centrum_i + \beta_9agencja_i + \beta_{10}spoldzielcze_i + \beta_{11}remont_i + \beta_{12}wykonczenie_i + \beta_{13}taras_i + \varepsilon_i,$$

Wyniki oszacowań parametrów zgodnie z teorią ujawniły dokładnie takie same wartości dla wszystkich badanych metod modelowania (Tabele 3-5). Cechy, które okazały się mieć dodatni wpływ na cenę ofertową nieruchomości mieszkaniowych to: winda w budynku, taras, niewielka powierzchnia użytkowa, wysoki budynek oraz bliski dystans do centrum miasta i najbliższej stacji metra. Ujemnym wpływem na cenę mieszkania charakteryzuje się: spółdzielcze własnościowe prawo własności, wysoki wiek budynku, balkon, pogorszony stan fizyczny oraz brak wykończenia. Pośrednictwo nieruchomości oraz piętro nie ma istotnego wpływu na cenę. Okazuje się, że mieszkańcy Warszawy bardzo cenią sobie metro, ponieważ wpływ bliskiej odległości od stacji metra okazał się odgrywać kluczową rolę przy wycenie mieszkania oraz jest zbliżony do efektu skracania dystansu od centrum. Współczynnik -0.0426 przy zmiennej odleglosc_centrum wskazuje na spadek ceny mieszkania o 4.17% za każdy dodatkowy kilometr odległości od centrum. Z kolei współczynnik -0.0258 przy zmiennej odleglosc metro oznacza zmniejszenie wartości nieruchomości o 2.55%, jeżeli odległość od najbliższej stacji metra wzrośnie o 1 km. Zmienne te są niezależne od siebie, a więc efekt kumuluje się, gdy mieszkanie oddala się jednocześnie od centrum oraz linii metra.

Błędy standardowe w modelach nieco się różnią, ponieważ zostały oszacowane za pomocą różnych metod. Różnice w modelu MNK i MNW są nieznaczne, natomiast w przypadku Metody Momentów wpłynęło to wyraźnie na wartość p-value doprowadzając do zmiany interpretacji istotności zmiennych „powierzchnia” oraz „pietro_max”. Metoda numeryczna, taka jak Metoda Momentów, może prowadzić do różnic w p-value w porównaniu do Metody Najmniejszych Kwadratów. Zmienne te jednak powinny być uwzględnione, gdyż zarówno MNW, jak i MNK ujawniają istotność, a MNK jest teoretycznie najbardziej efektywne w sensie najmniejszej wariancji.

Tabela 3. Wyniki estymacji modelu Metodą Najmniejszych Kwadratów

Nazwa zmiennej	Oszacowanie	Błąd standardowy	Statystyka Z	p-value
constant	10,22198151	0,020086	508,915	0,000000
remont	-0,14235484	0,013509	-10,538	0,000000
wykonczenie	-0,09285503	0,012358	-7,514	0,000000
winda	0,05934988	0,009337	6,356	0,000000
balkon	-0,04000363	0,008115	-4,93	0,000001
taras	0,03635170	0,010783	3,371	0,000757
wiek_budynku	-0,00308352	0,000213	-14,463	0,000000
spoldzielcze	-0,09063164	0,012926	-7,011	0,000000
powierzchnia	-0,00020729	0,000099	-2,099	0,035909
pietro	-0,00048455	0,001799	-0,269	0,787657
pietro_max	-0,00293354	0,001424	-2,061	0,039417
odleglosc_metro	-0,02578898	0,002204	-11,703	0,000000
odleglosc_centrum	-0,04257560	0,001576	-27,01	0,000000
agencja	-0,01346051	0,008653	-1,556	0,119910

Źródło: Opracowanie własne na podstawie danych z: <https://www.otodom.pl/>

Tabela 4. Wyniki estymacji modelu Metodą Największej Wiarygodności

Nazwa zmiennej	Oszacowanie	Błąd standardowy	Statystyka Z	p-value
constant	10,22198151	0,02004	510,025	0,000000
remont	-0,14235484	0,01348	-10,5605	0,000000
wykonczenie	-0,09285503	0,012331	-7,53029	0,000000
winda	0,05934988	0,009316	6,370489	0,000000
balkon	-0,04000363	0,008097	-4,94057	0,000001
taras	0,03635170	0,010759	3,37857	0,000729
wiek_budynku	-0,00308352	0,000213	-14,4946	0,000000
spoldzielcze	-0,09063164	0,012899	-7,02652	0,000000
powierzchnia	-0,00020729	0,000099	-2,10341	0,035430
pietro	-0,00048455	0,001795	-0,26997	0,787187
pietro_max	-0,00293354	0,001421	-2,06515	0,038909
odleglosc_metro	-0,02578898	0,002199	-11,7289	0,000000
odleglosc_centrum	-0,04257560	0,001573	-27,0682	0,000000
agencja	-0,01346051	0,008634	-1,55896	0,119005

Źródło: Opracowanie własne na podstawie danych z: <https://www.otodom.pl/>

Tabela 5. Wyniki estymacji modelu Metodą Momentów

Nazwa zmiennej	Oszacowanie	Błąd standardowy	Statystyka Z	p-value
constant	10,22198151	0,02373	430,8287	0,000000
remont	-0,14235484	0,011319	-12,5768	0,000000
wykonczenie	-0,09285503	0,013377	-6,9413	0,000000
winda	0,05934988	0,009783	6,066501	0,000000
balkon	-0,04000363	0,00859	-4,6568	0,000003
taras	0,03635170	0,012976	2,801431	0,005118
wiek_budynku	-0,00308352	0,00022	-14,0152	0,000000
spoldzielcze	-0,09063164	0,012074	-7,50628	0,000000
powierzchnia	-0,00020729	0,000218	-0,95136	0,341496
pietro	-0,00048455	0,001823	-0,26584	0,790379
pietro_max	-0,00293354	0,001646	-1,78251	0,074760
odleglosc_metro	-0,02578898	0,002554	-10,0973	0,000000
odleglosc_centrum	-0,04257560	0,00171	-24,8729	0,000000
agencja	-0,01346051	0,007916	-1,70046	0,089141

Źródło: Opracowanie własne na podstawie danych z: <https://www.otodom.pl/>

3. Wnioski

Przeprowadzone badanie miało na celu przeanalizowanie jak poszczególne charakterystyki mieszkania oddziałują na poziom jego ceny w ogłoszeniu ofertowym. Udało się określić wpływ poszczególnych charakterystyk na cenę mieszkania. Metody estymacji nie wpływają na oszacowanie parametrów, ale mogą wpływać na istotność zmiennych. Dalsze prace nad tym modelem powinny skupić się na wyodrębnieniu większej ilości zmiennych, a także na wykorzystaniu danych panelowych oraz zastosowaniu metod ekonometrii przestrzennej.