

<i>Autor/Autorzy</i>	Monika Kaczan, 410998 Michał Sękowski, 411399
<i>Który temat?</i>	Regresja kwantylowa

Informacje o artykule będącym inspiracją dla minimodelu

<i>Tytuł</i>	Wykorzystanie regresji kwantylowej w analizie zróżnicowania cen mieszkań
<i>Autor/Autorzy</i>	Marta Widłak, Natalia Nehrebecka
<i>Journal/Miejsce publikacji</i>	Wiadomości Statystyczne
<i>Rok</i>	2011
<i>Zakres stron</i>	17-46
<i>Tematyka, problemy i cele badawcze</i>	Artykuł został poświęcony analizie zróżnicowania cen mieszkań w Warszawie z wykorzystaniem regresji kwantylowej. Celem autorek było zbudowanie hedonicznych modeli cen mieszkań dla różnych przedziałów cenowych i hedonicznych indeksów cen mieszkań. Autorki chciały zbadać, czy wycena pewnych charakterystyk mieszkania może się różnić w zależności od badanego segmentu cenowego. Dodatkowo w artykule porównano różne metody szacowania hedonicznych cen mieszkań.
<i>Główne wnioski</i>	Wyniki analizy wskazały, że w badanym okresie dla warszawskiego rynku wtórnego jedynie standard wykończenia oraz lokalizacja były istotnie różnie wyceniane dla różnych segmentów cenowych. Między wycenami pozostałych cech nie było statycznie istotnych różnic. Autorki zwróciły również uwagę na zalety regresji kwantylowej w porównaniu do innych metod szacowania hedonicznych cen mieszkań.
<i>Metodyka badawcza</i>	W celu sprawdzenia, czy wyceny poszczególnych mieszkań różnią się w poszczególnych segmentach cenowych w badaniu wykorzystano metodę regresji kwantylowej oraz międzykwantylowej. Estymatory błędów standardowych współczynników regresji uzyskano za pomocą metody bootstrap. Ponadto, otrzymane wyniki porównano z tymi otrzymanymi za pomocą standardowego MNK.
<i>Dane</i>	W badaniu wykorzystano dane dotyczące warszawskiego rynku wtórnego mieszkań zebrane na podstawie transakcji biur pośrednictwa sprzedaży nieruchomości. Dane pochodzą z Bazy Rynku Nieruchomości tworzonej w NBP. Dane mają charakter przekrojowo-czasowy i dotyczą transakcji zawartych pomiędzy III kw. 2006 roku a II kw. 2010 roku.
<i>Dlaczego wybrano właśnie ten artykuł?</i>	Tematyka rynku nieruchomości wydaje nam się aktualna i interesująca. Jednocześnie na podstawie własnych obserwacji zauważamy, że relacja pomiędzy ceną mieszkania a jego charakterystykami może być różna dla różnych przedziałów cenowych, a użycie regresji kwantylowej pozwala na lepsze zrozumienie tych zależności oraz zaadresowanie problemów związanych z poprawnością budowanych modeli. Wybrany artykuł dotyczył lokalnego rynku warszawskiego, dlatego uznaliśmy za ciekawe zestawienie wyników ze stworzonym przez nas modelem opartym na danych pochodzących z innej części świata.

Podstawowe informacje o minimodelu

<i>Tytuł minimodelu</i>	Analiza zróżnicowania cen mieszkań w Pekinie
<i>Tematyka, problemy i cele badawcze</i>	Badanie zostało poświęcone analizie zróżnicowania cen mieszkań w Pekinie. Chcieliśmy sprawdzić, czy wycena charakterystyk nieruchomości różni się w zależności od segmentu cenowego mieszkania. Kierując się artykułem Widłak i Nehrebeckiej (2011), który był inspiracją do niniejszego badania, spodziewamy się, że przynajmniej niektóre z cech będą różnie wyceniane w różnych segmentach cenowych.
<i>Metodyka badawcza</i>	W celu poznania i zbadania zróżnicowania wyceny charakterystyk nieruchomości w różnych segmentach cenowych zbudowaliśmy modele hedoniczne cen mieszkań w oparciu o regresje kwantylowe dla pierwszego, drugiego i trzeciego kwartyla. Przeprowadziliśmy również regresje międzykwantylowe między poszczególnymi kwartylami oraz analizę graficzną wykresów rozkładów. We wszystkich przypadkach estymatory błędów standardowych współczynników uzyskaliśmy za pomocą bootstrapu. Otrzymane wyniki zostały także porównane do wyników standardowej MNK.
<i>Dane</i>	W badaniu wykorzystaliśmy zbiór danych dotyczący ofert sprzedaży mieszkań w Pekinie zamieszczonych na stronie Lianjia.com. Dane pobrano z serwisu Kaggle ¹ , a następnie wyczyszczono je i odpowiednio przekształcono na potrzeby analizy. Ostateczne dane mają charakter przekrojowo-czasowy i dotyczą 6603 ofert (obserwacji) zamieszczonych w serwisie w III kwartale 2015 roku.
<i>Główne wnioski</i>	Wyniki naszych badań wskazują, że wycena powierzchni mieszkania, stanu renowacji, typu budynku, wieku budynku, liczby pięter, bliskości metra, posiadania mieszkania przez właściciela dłużej niż 5 lat oraz lokalizacji różni się statystycznie w różnych segmentach cenowych wyznaczonych na podstawie kwartyli. Ponadto, analiza graficzna pokazuje, że zdecydowana większość cech ma widocznie różną wycenę w różnych punktach rozkładu. Jest to spowodowane różnymi preferencjami klientów o różnej zamożności.

¹ <https://www.kaggle.com/datasets/ruiqurm/lianjia>

Ekonometryczny opis badania

1. Wstęp

Pekin jest jednym z największych i najludniejszych miast na świecie. Jednocześnie jest również jednym z najdroższych miast na świecie do zakupu nieruchomości, z cenami wynoszącymi średnio ok. 5 tys. dolarów za metr kwadratowy w 2019 roku². Tym bardziej konsumenci, inwestorzy oraz firmy pośredniczące w obrocie nieruchomościami są zainteresowani odpowiednią wyceną sprzedawanych mieszkań. Nie wszystkie nieruchomości są wyceniane tak samo, a różni klienci mogą być w różnym stopniu skłonni zapłacić za konkretne charakterystyki nabywanej nieruchomości.

W niniejszej pracy chcielibyśmy zbudować hedoniczne modele cen mieszkań w Pekinie dla różnych segmentów cenowych. Modele hedoniczne to modele ekonometryczne opisujące cenę dobra za pomocą kombinacji jego istotnych charakterystyk i ich indywidualnych wycen³. Chcielibyśmy zbadać, czy wycena poszczególnych charakterystyk mieszkań różni się w zależności od segmentu cenowego. W tym celu wykorzystamy metodę regresji kwantylowej, która pozwala na zbadanie zależności całego rozkładu zmiennej objaśnianej od zmiennych objaśniających.

Nasza praca jest inspirowana badaniem Widłak i Nehrebeckiej (2011) „*Wykorzystanie regresji kwantylowej w analizie zróżnicowania cen mieszkań*”. Przypuszczamy, że podobnie jak w analizie Widłak i Nehrebeckiej dla warszawskiego rynku nieruchomości, niektóre z cech takie jak standard wykończenia czy lokalizacja mogą być różnie wyceniane w różnych segmentach cenowych. Jest to spowodowane tym, że osoby kupujące mieszkania z różnych segmentów mają inne preferencje. Przykładowo, osoby zamożne cenią luksusowość i prestiżową lokalizację, podczas gdy osobom mniej zamożnym zależy na opłacalności i dostępności do podstawowych udogodnień.

2. Opis danych i użytych zmiennych

W badaniu wykorzystano zbiór danych dotyczący ofert sprzedaży mieszkań w Pekinie zamieszczonych na stronie Lianjia.com. Lianjia jest chińską firmą zajmującą się pośrednictwem sprzedaży nieruchomości. Oryginalne dane zostały pobrane z serwisu Kaggle⁴, a następnie wyczyszczono je i odpowiednio przekształcono na potrzeby niniejszej analizy.

Oryginalne dane pochodziły z lat 2011-2017, jednak poszczególne okresy nie były reprezentowane równomiernie. W celu uzyskania jednolitej próby i ograniczenia wpływu wahań rynkowych wykraczających poza zakres modelu zdecydowano się na pozostawienie

² <https://www.ceicdata.com/en/china/nbs-property-price-monthly/property-price-ytd-avg-beijing>, dane z grudnia 2019

³ Król, A. i Targaszewska, M. (2016). Zastosowanie klasyfikacji do wyodrębniania homogenicznych grup dóbr w modelowaniu hedonicznym. *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu*, 426.

⁴ <https://www.kaggle.com/datasets/ruiqurm/lianjia>

obserwacji tylko z jednego kwartału. Dodatkowo usunięto obserwacje (ok. 0.1% próby) podejrzane o bycie obserwacjami błędnymi, a także obserwacje z brakującymi danymi. Wybrane kategorie zmiennych nominalnych zostały ujednolicone. Ostateczny zbiór danych zawierał 6603 obserwacji dotyczących ofert sprzedaży mieszkań w III kwartale 2015 roku.

Tabela 1. przedstawia zmienne użyte ostatecznie w modelach. Ich wybór został podyktowany dostępnością danych, analizą badania Widłak i Nehrebeckiej oraz specyfiką chińskiego rynku nieruchomości. Podobnie jak w pracy Widłak i Nehrebeckiej jako zmienną objaśnianą zdecydowano się przyjąć logarytm naturalny ceny metra kwadratowego mieszkania. Zmienne objaśniające reprezentują natomiast różne charakterystyki sprzedawanej nieruchomości.

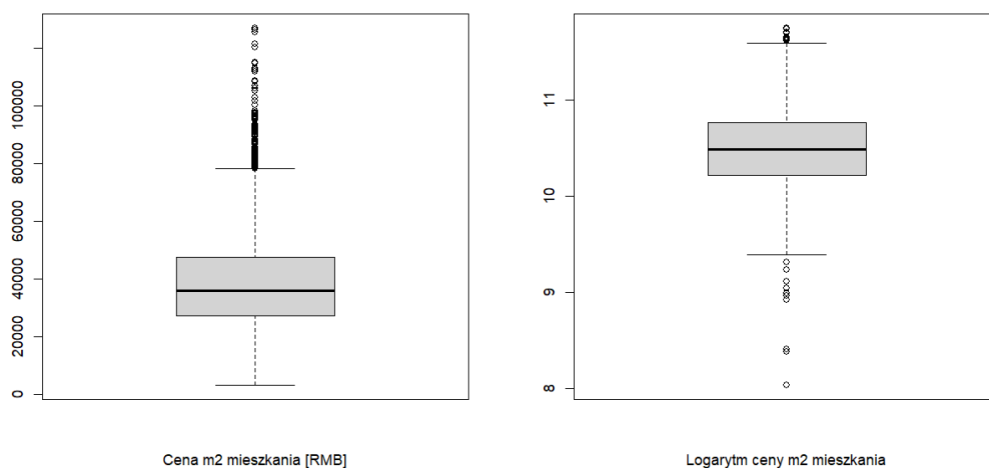
Tabela 1. Zmienne użyte w modelach

Zmienna	Opis
<i>ln_pricem2</i>	Logarytm naturalny ceny metra kwadratowego mieszkania w RMB
<i>aream2</i>	Powierzchnia mieszkania w metrach kwadratowych
<i>bathroom</i>	Liczba łazienek
<i>kitchen</i>	Oddzielna kuchnia (zmienna zerojedynkowa: 0 – nie, 1 – tak)
<i>district</i>	Dzielnica Pekinu (zmienna nominalna: <i>DongCheng, FengTai, DaXing, FaXing, FangShang, ChangPing, ChaoYang, HaiDian, ShiJingShan, XiCheng, TongZhou, ShunYi, MenTouGou</i>)
<i>subway</i>	Stacja metra w pobliżu budynku (zmienna zerojedynkowa: 0 – nie, 1 – tak)
<i>buildingType</i>	Typ budynku (zmienna nominalna: <i>plate, plate/tower, tower</i>)
<i>buildingStructure</i>	Materiał, z którego wykonano budynek (zmienna nominalna: <i>brick/concrete, mixed, steel/concrete</i>)
<i>renovationCondition</i>	Stan renowacji mieszkania (zmienna nominalna: <i>rough, simplicity, hardcover, other</i>)
<i>floor</i>	Liczba kondygnacji budynku
<i>elevator</i>	Budynek posiada windę (zmienna zerojedynkowa: 0 – nie, 1 – tak)
<i>buildingAge</i>	Wiek budynku
<i>fiveYearsProperty</i>	Obecny właściciel mieszkania posiada go krócej niż 5 lat (zmienna zerojedynkowa: 0 – nie, 1 – tak) ⁵

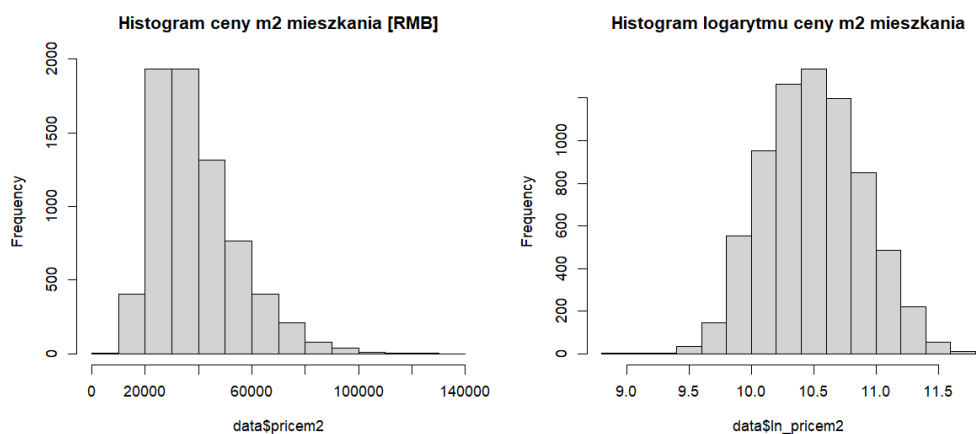
Rysunek 1. przedstawia wykresy pudełkowe odpowiednio ceny m2 mieszkania oraz logarytmu ceny m2 kwadratowego mieszkania, natomiast Rysunek 2. przedstawia histogramy rozkładu tych zmiennych. Widzimy, że nawet po zlogarytmowaniu zmiennej objaśnianej, mamy do czynienia z obserwacjami nietypowymi. Wskazuje to na możliwe problemy z estymacją modelu metodą najmniejszych kwadratów, które może pomóc zaadresować regresja kwantylowa.

⁵ Uwzględnienie tej zmiennej jest związane z chińskimi przepisami dotyczącymi kupowania mieszkań na cele inwestycyjne, które nakładają na właścicieli dodatkowy podatek od zysków kapitałowych przy sprzedaży mieszkań posiadanych krócej niż 5 lat.

Rysunek 1. Wykresy pudełkowe ceny m2 mieszkania i logarytmu ceny m2 mieszkania⁶



Rysunek 2. Histogramy ceny m2 mieszkania i logarytmu ceny m2 mieszkania



3. Modele ekonometryczne

3.1. Metoda badawcza

W celu budowy modeli hedonicznych i zbadania zróżnicowania wyceny charakterystyk mieszkań w Pekinie zdecydowaliśmy się przeprowadzić regresję kwantylową oraz regresję międzykwantylową na przedstawionych danych. Regresja kwantylowa pozwala na zbadanie zależności całego rozkładu zmiennej objaśnianej od zmiennych objaśniających. Regresja kwantylowa przyjmuje postać:

$$Q_p(y_i | x_i) = \alpha_p + x_i \beta_p + F_{\varepsilon_i}^{-1}(p)$$

⁶ Wszystkie rysunki i tabele w niniejszej pracy zostały wykonane na podstawie opracowań własnych.

gdzie: $Q_p(y_i | x_i)$ – warunkowa wartość oczekiwana y względem x dla p -tego kwantyla; α_p , β_p – parametry regresji dla p -tego kwantyla; $F_{\varepsilon_i}^{-1}(p)$ – p -ty kwantyl składnika losowego.

Oszacowania parametrów regresji kwantylowej znajdujemy poprzez minimalizację ważonej sumy wartości bezwzględnych reszt. W najprostszym przypadku mediany uzyskany estymator jest więc równoważny estymatorowi metody najmniejszych odchyleń bezwzględnych (LAD).

Zastosowanie regresji kwantylowej pozwala na uzyskanie pełniejszego obrazu analizowanej zależności, ponieważ możemy otrzymać informacje o całym rozkładzie zmiennej objaśnianej. Dzięki temu będziemy mogli porównać oszacowania parametrów regresji w różnych segmentach cenowych mieszkań bez konieczności dzielenia danych na podróby. Ponadto, w porównaniu do MNK regresja kwantylowa jest bardziej odporna na występowanie obserwacji nietypowych – które są obecne w naszej próbie – oraz na występowanie silnie nieliniowego składnika losowego. W niektórych przypadkach estymacja regresji kwantylowej pozwala także na uzyskanie efektywniejszych parametrów.

3.2. Regresja kwantylowa

Podobnie jak w artykule Widłak i Nehrebeckiej (2011), zdecydowaliśmy się przeprowadzić regresję dla trzech kwantyli (segmentów cenowych mieszkań): pierwszego, drugiego i trzeciego kwantyla. Estymatory błędów standardowych współczynników uzyskaliśmy za pomocą bootstrapu z 999 powtórzeniami. Dodatkowo w celach porównawczych przeprowadziliśmy również modelowanie metodą najmniejszych kwadratów.

Tabela 2. przedstawia porównanie wyników modeli. Wszystkie modele regresji kwantylowej charakteryzują się poprawną formą funkcyjną (spełniają testy RESET) oraz brakiem współliniowości (wyznaczonej na podstawie macierzy korelacji i testów VIF). Każda z badanych charakterystyk mieszkań okazała się statystycznie istotna w przynajmniej jednym modelu, a w przypadku zmiennych kategorycznych, zmienne w nie wchodzące były łącznie istotne we wszystkich modelach na poziomie 5%.

Tabela 2. Wyniki regresji kwantylowej i międzykwantylowej

Zmienna	Regresja kwantylowa				Regresja międzykwantylowa		
Zmienna objaśniana: $\ln_pricem2$	MNK	Q25	Q50	Q75	Q25-Q50	Q50-Q75	Q25-Q75
aream2	-.0010077 ***	-.0013498 ***	-.0008689 ***	-.0008210 ***	**		*
kitchen	.0761445 **	.1588203 *	.1092562 ***	0.040724		.	
bathroom	.0674957 ***	.0622564 ***	.0588858 ***	.0724638 ***			
renovation (base level: hardcover)							
other	-.0393441 ***	-.0362343 ***	-.0356996 **	-.0228429 *			
rough	-.0477224 **	-.0582359 *	-.0735919 **	-.0387274 *			
simplicity	-.0559080 ***	-.0560551 ***	-.0630607 ***	-.0472443 ***		*	
buildingType (base level: plate)							
plate/tower	-.0459759 ***	-.0580125 ***	-.0616012 ***	-.0403575 **			
tower	-.1215377 ***	-.1013577 ***	-.1271465 ***	-.1439881 ***	.		**
buildingStructure (base level: brick/concrete)							
mixed	.0033808	.0087048	-.0096117	-.0001531			
steel/concrete	.0544009 **	.0854371 **	.0641772 **	.0413770 *			
buildingAge	.0049577 ***	.0063946 ***	.0055641 ***	.0037586 ***	*	**	***
floor	.0023576 ***	.0008345	.0022035 *	.0039144 ***		*	***
elevator	.0760834 ***	.0801654 ***	.0699888 ***	.0686271 ***			
fiveYearsProperty	-.0039688	-.0188069 .	-.0101348	-.0198334 *		.	***
subway	.0955506 ***	.1038833 ***	.1040583 ***	.0756387 ***	**	***	*
disctrict (base level: ChangPing)							
ChaoYang	.3374495 ***	.2595023 ***	.3108246 ***	.3734830 ***	***	***	***
DaXing	.0615254 .	.0347505	.1013460 *	.0591172 .			
DongCheng	.6433964 ***	.5812228 ***	.6397156 ***	.6811535 ***	**	**	***
FangShang	-.2702617 ***	-.2559429 ***	-.3185747 ***	-.2796334 ***			
FaXing	.0404318 *	-.0177648	.0189894	.0153629	.		
FengTai	.2523608 ***	.2277784 ***	.2222266 ***	.2346267 ***			
HaiDian	.6224918 ***	.5653764 ***	.6007672 ***	.6501794 ***	**	***	***
MenTouGou	-.1563468 ***	-.1391570 ***	-.1792951 ***	-.1949091 ***	*		*
ShiJingShan	.2077667 ***	.1459025 ***	.1817174 ***	.1981159 ***	**		
ShunYi	-.0788283 .	-.0606012	-.0949535 ***	-.1579475 ***		**	**
TongZhou	-.0230495	-.0555362 **	-.0555202 **	-.0310190.			
XiCheng	.8383775 ***	.7111767 ***	.7988574 ***	.9133055 ***	***	***	***
const.	9.8956428 ***	9.7003524 ***	9.8929033 ***	10.087509 ***	***	***	***
R² / adjusted R²	0.6577 / 0.6562						
pseudo R²		0.4220	0.4387	0.4526			

Oznaczenia: *** - poziom istotności 0.1%; ** - poziom istotności 1%; * - poziom istotności 5%; . – poziom istotności 10%

Poszczególne kategorie zmiennych kategorycznych zostały zaprezentowane oddzielnie, jednak w przypadku każdej z nich były one łącznie istotne na poziomie 5%.

Znaki parametrów dla poszczególnych charakterystyk nieruchomości są generalnie zgodne z oczekiwaniami. Możemy jednak zauważyć, że chociaż znaki przy (statystycznie istotnych) oszacowaniach parametrów są takie same we wszystkich modelach, to przy niektórych zmiennych wartości oszacowań znacząco się różnią. Przykładowo, dla niskiego segmentu cenowego (pierwszego kwartyła) obecność metra w pobliżu zwiększa cenę metra kwadratowego mieszkania o 10,3%, podczas gdy dla wysokiego segmentu cenowego (trzeciego kwartyła) o 7,6%. Wyniki regresji MNK również różnią się więc od oszacowań dla poszczególnych kwantyli.

3.3. Regresja międzykwantylowa

Aby formalnie sprawdzić, czy parametry regresji dla poszczególnych kwantyli są statystycznie różne, przeprowadziliśmy regresje międzykwantylowe w trzech wariantach: między 0,25 a 0,50; 0,50 a 0,75 oraz 0,25 a 0,75 kwantylem. Estymatory błędów standardowych współczynników ponownie uzyskaliśmy za pomocą bootstrapu z 999 powtórzeniami. Wyniki regresji międzykwantylowej zostały przedstawione w Tabeli 2.

Widzimy, że wycena powierzchni mieszkania, stanu renowacji, typu budynku, wieku budynku, liczby pięter, bliskości metra, posiadania mieszkania przez właściciela dłużej niż 5 lat oraz lokalizacji różni się statystycznie w różnych segmentach cenowych wyznaczonych na podstawie kwantyli.

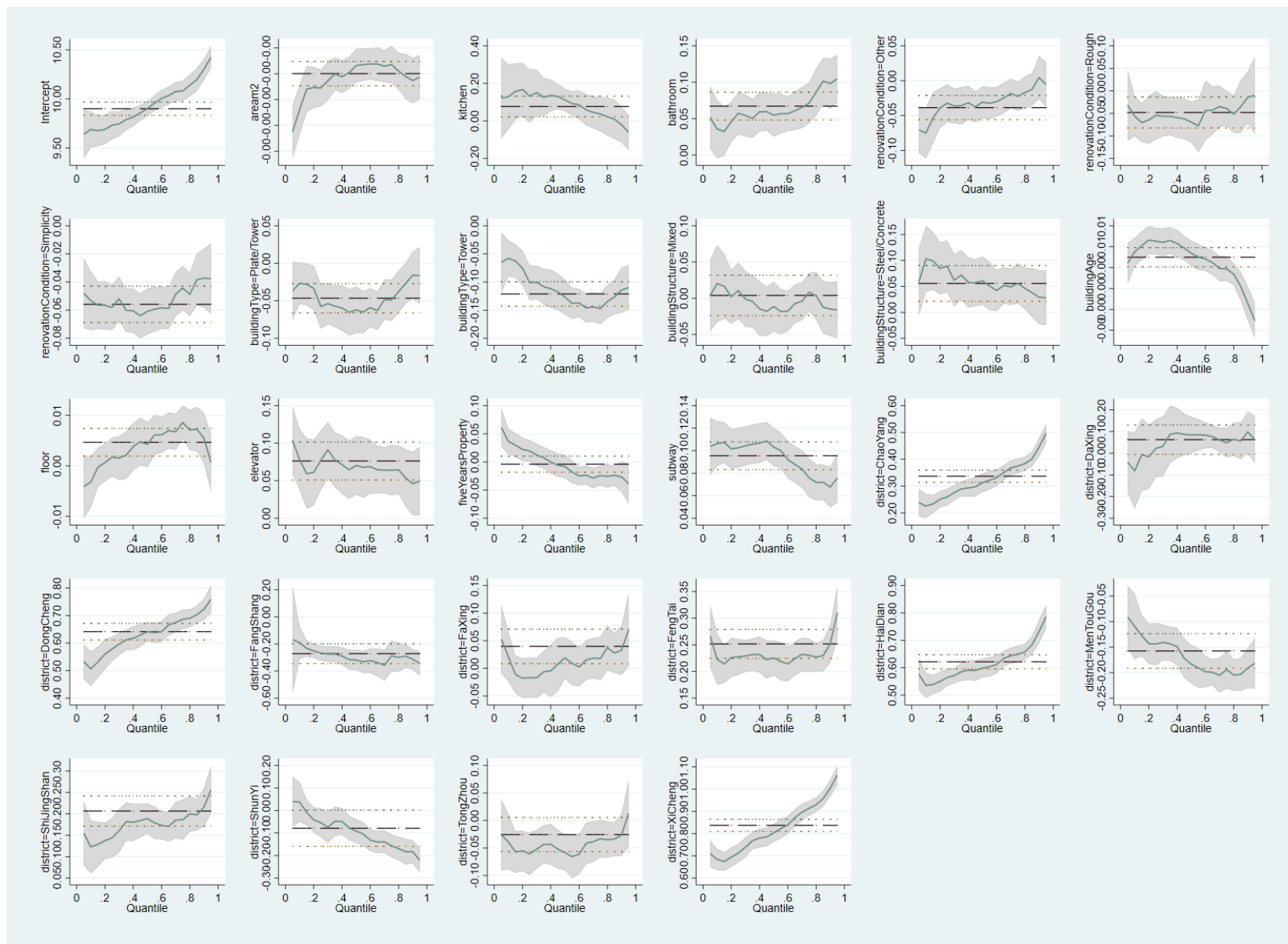
W niskich segmentach cenowych klienci przykładają relatywnie większą wagę do opłacalności mieszkania, którą można mierzyć spadkiem ceny metra kwadratowego wraz ze wzrostem powierzchni. Tanio wyceniane mieszkania są również starsze i niższe.

Z drugiej strony, osoby bogate często kupują drugie czy trzecie mieszkania pod cele inwestycyjne, dlatego ma dla nich znaczenie, czy zapłacą dodatkowy podatek od zysków kapitałowych przy mieszkaniach posiadanych przez właścicieli krócej niż 5 lat. Osoby bogate bardziej cenią również mieszkanie w prestiżowych dzielnicach blisko centrum takich jak XiCheng i DongCheng obejmujących stare miasto Pekinu. Mniejsze znaczenie przykładają natomiast do np. obecności metra w pobliżu, ponieważ mogą pozwolić sobie na poruszanie się innymi środkami transportu.

3.4. Analiza graficzna

Dotychczas analizowaliśmy jedynie różnice między pierwszym, drugim i trzecim kwantylem. Można jednak przyjrzeć się także zróżnicowaniu zmiennej w całym rozkładzie, co umożliwia graficzna analiza wyników regresji kwantylowej przedstawiona na Rysunku 3. Przy generowaniu wykresów przedziały ufności uzyskaliśmy za pomocą bootstrapu z 499 powtórzeniami.

Rysunek 3. Wykresy regresji kwantylowej



Linie przerywane – wyniki MNK z 95% przedziałami ufności. Linia zielona i szary obszar – wyniki regresji kwantylowej z 95% przedziałami ufności.

Na podstawie analizy wykresów możemy zaobserwować, że większość zmiennych jest widocznie różnie wyceniana w różnych punktach rozkładu. W szczególności wycena wybranych dzielnic znacząco spada lub rośnie wraz ze wzrostem ceny metra kwadratowego. Jest to dosyć naturalne, ponieważ różne lokalizacje są preferowane przez osoby o różnej zamożności, które chcą mieszkać w podobnym środowisku.

Interesująca jest też relacja między ceną metra kwadratowego mieszkania a wysokością (liczbą pięter) budynku dla różnych kwartyli. Generalnie im bogatsza osoba, tym wyżej wycenia mieszkanie w kompleksie wysokich apartamentów – w przeciwieństwie do niskich budynków bardziej przypominającego slumsy. Jednak najbogatsze osoby mogą sobie pozwolić na mieszkanie w niskich ekskluzywnych apartamentowcach.

Z drugiej strony, jedynie kilka zmiennych np. obecność windy czy oddzielnej kuchni wydaje się mieć w miarę podobną wycenę w całym rozkładzie.

4. Podsumowanie

Celem niniejszej pracy była analiza zróżnicowania wycen charakterystyk mieszkań w Pekinie. Za pomocą regresji kwantylowej zbudowaliśmy modele hedoniczne cen metra kwadratowego nieruchomości dla pierwszego, drugiego oraz trzeciego kwartyli. Następnie na podstawie regresji międzykwantylowej oraz analizy graficznej wykresów zbadaliśmy różnice między wycenami cech mieszkań dla różnych segmentów cenowych.

Wyniki naszych badań wskazują, że wycena powierzchni mieszkania, stanu renowacji, typu budynku, wieku budynku, liczby pięter, bliskości metra, posiadania mieszkania przez właściciela dłużej niż 5 lat oraz lokalizacji różni się statystycznie w różnych segmentach cenowych wyznaczonych na podstawie kwartyli. Ponadto, analiza graficzna pokazuje, że zdecydowana większość cech ma widocznie różną wycenę w różnych punktach rozkładu. Jest to spowodowane różnymi preferencjami klientów o różnej zamożności.

Niniejsza analiza może mieć znaczenie dla agentów nieruchomości oraz innych podmiotów zajmujących się wyceną mieszkań. Na jej podstawie można powiedzieć, w jaki sposób powinny być wyceniane mieszkania kierowane do różnych grup klientów.