



SZKOŁA GŁÓWNA HANDLOWA W WARSZAWIE
WARSAW SCHOOL OF ECONOMICS

Studium Magisterskie

Kierunek: Metody ilościowe i systemy informacyjne

Nazwa przedmiotu: Ekonometria panelowa

Yauheni Semianiuk

Nr albumu: 82591

Prognozowanie przeciętnej ceny rynkowej 1 m^2 kwadratowego mieszkania wśród 15 największych miast Polski

Warszawa 2022

Spis treści

Wstęp	3
Rozdział I: Opis zjawiska	5
Rozdział II: Opis metody badawczej	8
2.1. Modele z efektami stałymi i losowymi	8
2.2. Dowody empiryczne	9
Rozdział III: Wyniki przeprowadzonych analiz	11
3.1. Analiza danych.....	11
3.2. Wybór modelu.....	14
3.3. Porównywanie wartości dopasowanych i prognozy z najlepszego modelu.....	17
Podsumowanie	19
Bibliografia	20
Spis rysunków.....	22
Spis tabeli.....	23

Wstęp

Cena nieruchomości stanowi jedną z najważniejszych wskaźników statystycznych, kształtując rozwój demograficzny i ekonomiczny zarówno rynku lokalnego, jak i globalnego kraju. Niska cena mieszkania stymuluje migrację zewnętrzną i wewnętrzną, co z kolei kumuluje dodatkowe zasoby ludnościowe do dyspozycji w danym obszarze geograficznym. Mówiąc o wpływie demograficznym, cena mieszkania jest faktorem kierującym decyzje o osiedlaniu, mianowicie dla osób w młodszym wieku produkcyjnym.¹ Przyczynia to do restrukturyzacji rozkładu wiekowego na terenie kraju. W końcu cena mieszkania oddziałuje na koszty prowadzenia działalności gospodarczej, wspierając tym samym ulokowanie przedsiębiorstw w obszarach o niskich kosztach pomieszczeń komercyjnych i zabudowy. W przypadku największych miast Polski, cena mieszkania ma zasadniczy wpływ również na rozwój całego kraju, będąc pewnym standardem rynkowym dla innych miast Rzeczypospolitej Polskiej.

Rynek nieruchomości w Polsce na przestrzeni lat ulega ciągłym przemianom, m.in. w wyniku zmieniającej się sytuacji demograficzno-społecznej, makro- i mikroekonomicznej kraju, jak również w wyniku wprowadzanych zmian regulacyjnych. Bodźcem do zmian jest także rozwój infrastruktury, który przyczynia się do powstawania inwestycji w dotychczas mało cenionych lokalizacjach. Zdolność do prognozowania zmienności ceny rynkowej mieszkania jest ważna dla podejmowania szeregu decyzji. Po pierwsze, jest to ważne dla zagospodarowania krajem, skoro pozwoli to na prognozowanie zmiany liczby ludności, liczby powstałych jednostek gospodarczych oraz na planowanie budżetu. Po drugie, wycena mieszkania jest niezbędna dla planowania stałych kosztów firm prywatnych, wynajmujących budynki komercyjne na terenie Polski. Po trzecie, analiza cen mieszkań jest przydatna dla jednostek fizycznych, planujących kupno, wynajęcie lub sprzedaż nieruchomości. Po czwarte, informacja o prognozowanej cenie mieszkania stanowi ważną część sektora bankowego, wpływając na zasób środków udzielanych dla celów kredytów hipotecznych.

¹ Peng C., The long- and short-run influences of housing prices on migration, s. 257

Celem niniejszej pracy jest ewaluacja wysokości średniej ceny jednego metra kwadratowego mieszkania wśród piętnastu największych miastach Polski oraz próba jej przybliżenia za pomocą odpowiednich modeli ekonometrycznych: modelu regresji liniowej, panelowego modelu o efektach stałych oraz panelowego modelu o efektach losowych. Najlepsze przybliżenie pozwoli dokładnie obliczyć wartość mieszkań komercyjnych i prywatnych, co z kolei pomoże jednostkom fizycznym, gospodarczym bądź instytucjom regulacyjnym podjąć prawidłowe decyzje związane z zarządzaniem nieruchomością oraz powiązanymi kosztami i ryzykiem.

W pierwszej części pracy zostaną opisane pojęcia rynku nieruchomości. Następnie przedstawione będą modele panelowe, które będą stanowiły metodologię niniejszej pracy. Rozdział III opisuje dane wykorzystane w badaniach oraz przedstawia wyniki obliczeń średniej ceny 1m. kwadratowego mieszkania metodami panelowymi wraz z prognozą na rok 2021. Ostatnia część dotyczy podsumowania najważniejszych aspektów niniejszego badania.

Rozdział I: Opis zjawiska

W niniejszej pracy za pojęcie rynku nieruchomości uznany jest zbiór stosunków zachodzących pomiędzy podmiotami, generujące popyt i podaż nieruchomości.² Opisując rynek nieruchomości i procesów w nim zachodzących, należy przedstawić kilka cech charakterystycznych dla tego rynku. Pierwszą ważną cechą opisującą rynek nieruchomości jest jego niejednorodność. Podział rynku nieruchomości jest w wielkim stopniu uzależniony od kryterium podziału. Do najczęściej wykorzystanych kryteriów należą:

- Kryterium według funkcji pełnionych przez nieruchomość,
- Kryterium przestrzenny,
- Kryterium cenowy,
- Kryterium jakościowy (szczególnie w odniesieniu do nieruchomości biurowych i hotelowych).

Ze względu na funkcje pełnione przez nieruchomość wyróżniają:

- Rynek nieruchomości mieszkaniowych, obejmujący domy oraz lokale mieszkalne poza nieruchomościami, przynoszącymi dochód,
- Rynek nieruchomości komercyjnych, obejmujący nieruchomości stanowiące źródło dochodów i zysku dla właścicieli,
- Rynek nieruchomości przemysłowych, zawierający głównie nieruchomości fabryczne,
- Rynek nieruchomości rolnych – dotyczy głównie gruntów rolnych, pastwisk i in.,
- Rynek nieruchomości specjalnych, do którego należą nieruchomości cmentarne, kościelne, pola golfowe itp.

Natomiast dzieląc rynek nieruchomości ze względu na kryterium przestrzenny odznaczają się:

- Rynek lokalny (zwykle, rynek na poziomie jednego miasta lub aglomeratu),
- Rynek regionalny (np. rynek województwa mazowieckiego),
- Rynek krajowy,

² Kucharska-Stasiak E., *Nieruchomości w gospodarce rynkowej*, PWN Warszawa 2006, str. 38-55

- Rynek międzynarodowy.

Jednak zwyczajnie rynek nieruchomości cechuje się mocnym charakterem lokalnym. Różnorodność podziału rynku nieruchomości pozwala wywnioskować, że rynek nieruchomości jest zbiorem wzajemnie powiązanych rynków. Wartość nieruchomości wyznaczana jest natomiast przez siły popytu i podaży, działające na poszczególnych rynkach.

Kolejną cechą rynku nieruchomości jest jego niedoskonałość. Niedoskonałość ta wynika z szeregu faktów. Po pierwsze, ceny mieszkań są zwyczajnie przedstawione na poziomie agregacji kwartalnej i dotyczą uśrednionej ceny mieszkań bez podziału na kategorie cenowe i jakościowe. Informacje dotyczące transakcji kupna sprzedaży mają często także charakter prywatny, co konstruuje pewne trudności w akumulacji cen dotyczących poszczególnych jednostek. Stanowi to tzw. problem asymetrii informacji na rynku nieruchomości, gdyż większość kupujących i sprzedających może nie posiadać wystarczających danych dotyczących cen nieruchomości w momencie zawierania transakcji. Po drugie, o niedoskonałości rynku nieruchomości decydują również wrażliwość wartości nieruchomości na zmiany w otoczeniu geograficznym, politycznym i socjalnym oraz wzajemne oddziaływanie na siebie mieszkań sąsiadujących. Po trzecie, transakcje wymagają posiadania dużego kapitału finansowego, stąd zmiany w warunkach rynkowych są trudno obserwowalne. Po czwarte, zarówno decyzji sprzedających, jak i kupujących są często oparte na czynnikach nieekonomicznych (np. prestiż lokalizacji, sąsiedztwo, tradycje w przypadku nabywców; stereotypy, poglądy osobiste w przypadku sprzedawców). Wreszcie, sposób użytkowania nieruchomości wpływa na wartość obiektu (w ten sposób np., wartość lokalu przeznaczonego na cele mieszkalne jest niższa, niż wartość tego samego lokalu przeznaczonego na cele biurowe).

Ostatnia istotna cecha rynku nieruchomości dotyczy małej elastyczności cenowej podaży i popytu. Niska elastyczność cenowa popytu wynika z relatywnej stałości podaży nieruchomości w krótkim okresie, wynikającej m in. z wysokiego początkowego kapitału inwestycyjnego. W konsekwencji wzrost popytu bardzo wolno przyczynia się do wzrostu podaży.

Wyżej wymienione cechy rynku nieruchomości są cechami stałymi. Z kolei, do podmiotów, biorących aktywny udział w tworzeniu rynku nieruchomości zaliczają się:

- Inwestorzy,
- Kredytodawcy,
- Dzierżawcy,
- Pośrednicy,
- Obsługa techniczna rynku.

Inwestorzy podejmują podstawowe decyzje kupna i sprzedaży nieruchomości na rynku, zwykle wykorzystując środki finansowe instytucji udzielających kredyty hipoteczne. Grupa pośredników w obrocie nieruchomościami dostarcza uczestnikom rynku informacji o nim, co ułatwia podejmowanie decyzji i częściowo rozwiązuje problem asymetrii informacji, zachodzący na rynku. Dzierżawcy z kolei pobudzają popyt na konkretne rodzaje nieruchomości.

Rozpatrując rynek nieruchomości jako całość, należy podkreślić, że pełni on bardzo istotne funkcje gospodarcze i społeczne. Do funkcji tych należą:

- Funkcja wymiany, która pozwala na alokację praw majątkowych i redystrybucję przestrzeni użytkowanych.
- Funkcja informacyjna, cechująca się tym, że rynek poprzez cenę informuje o wartości wymiennej różnych nieruchomości.
- Funkcja korygowania wykorzystania przestrzeni, występującej w przypadku, gdy niski dochód z zabudowanej przestrzeni skłania właściciela do poszukiwania rozwiązań, które pozwolą zmaksymalizować wartość nieruchomości.

W ten sposób przedstawiony został ogólny pogląd na rynek nieruchomości. W następnym rozdziale zostały omówione modele ekonometryczne do badania ceny mieszkania na rynku nieruchomości mieszkaniowej wśród piętnastu największych względem ludności miast Polski.

Rozdział II: Opis metody badawczej

2.1. Modele z efektami stałymi i losowymi

W danych panelowych, w których istnieją obserwacje podłużne dla tego samego podmiotu, efekty stałe reprezentują średnie specyficzne dla każdego badanego podmiotu osobno. W analizie danych panelowych termin estymator efektów stałych (znany również jako estymator wewnętrzny, (ang. within) jest używany w odniesieniu do estymatora dla współczynników w modelu regresji ze stałymi efektami stałymi. Takie modele pomagają w kontrolowaniu pominiętego odchylenia zmiennych z powodu nieobserwowanej heterogeniczności, gdy ta heterogeniczność jest stała w czasie. Ta heterogeniczność może zostać usunięta z danych poprzez różnicowanie, na przykład poprzez odjęcie średniej na poziomie grupy w czasie (transformacja within) lub poprzez wzięcie pierwszej różnicy, która usunie wszelkie niezmiennie w czasie komponenty modelu.

Liniowy model z efektami nieobserwowalnymi dla poszczególnych jednostek badawczych, dla N obserwacji i T okresów można przedstawić następująco:

$$y_{it} = X_{it}B + \alpha_i + u_{it} \text{ dla } t = 1, \dots, T \text{ oraz } i = 1, \dots, N,$$

gdzie:

- y_{it} – zmienna zależna (w niniejszej pracy, przeciętna cena 1m kwadratowego mieszkania),
- X_{it} – macierz $1 \times k$ regressorów zmiennych w czasie,
- B – macierz $1 \times k$ parametrów stojących przy odpowiednich regresorach,
- α_i – nieobserwowany niezmienny w czasie efekt indywidualny,
- u_{it} – składnik losowy.

W przeciwieństwie do X_{it} , α_i nie można być bezpośrednio zaobserwowana. Model z efektami stałymi eliminuje alfę, za pomocą odpowiedniej transformacji:

$$y_{it} - \bar{y}_i = (X_{it} - \bar{X}_i)B + (\alpha_i - \bar{\alpha}_i) + (u_{it} - \bar{u}_i) \Rightarrow \dot{y}_{it} = \dot{X}_{it}B + \dot{u}_{it},$$

gdzie:

- $\bar{y}_i = \frac{1}{T} \sum_{t=1}^T y_{it},$

- $\bar{X}_i = \frac{1}{T} \sum_{t=1}^T X_{it},$
- $\bar{u}_i = \frac{1}{T} \sum_{t=1}^T u_{it}.$

Skoro α_i jest stała, $\bar{\alpha}_i = \alpha_i$ a tym samym efekt jest wyeliminowany. Estymator efektów stałych \widehat{B}_{FE} jest zatem oszacowany przez estymator KNMK.

Założeniem modelu z efektami stałymi jest to, że efekty indywidualne są skorelowane ze zmiennymi niezależnymi. Przeciwnie, założenie modelu z efektami losowymi jest takie, że efekty indywidualne nie są skorelowane ze zmiennymi niezależnymi, czyli są częścią składnika losowego. Formalnie to można zapisać w następujący sposób:

$$y_{it} = X_{it}B + \alpha_i + v_{it} \text{ dla } t = 1, \dots, T \text{ oraz } i = 1, \dots, N,$$

$$v_{it} = u_{it} + \alpha_i$$

Dalej, w odróżnieniu od modelu z efektami stałymi, wykorzystany jest estymator uogólnionej metody najmniejszych kwadratów, UMNK³.

2.2. Dowody empiryczne

Na przestrzeni lat przeprowadzono mnóstwo badań na prognozowania cen nieruchomości na różnych rynkach światowych. Na przykład Evans i in. (1992) testowali dokładność predykcyjną sieci neuronowych do szacowania cen nieruchomości mieszkalnych w Wielkiej Brytanii; odkryli, że modele sieci neuronowych dobrze sprawdzają się przy wycenie własności. Crawford i Fratantoni (2003) zastosowali model przełącznikowy Markowa do wyceny domów w USA i porównali wyniki z autoregresyjną średnią ruchomą (ARMA) i uogólnionym modelem autoregresyjnej warunkowej heteroskedastyczności (GARCH). Odkryli, że chociaż model przełącznikowy był bardzo obiecujący z doskonałym dopasowaniem w próbie, jego prognozy poza próbą były generalnie gorsze od bardziej standardowych technik prognozowania. Miller i Peng (2006) wykorzystali dane statystyczne na poziomie obszaru metropolitalnego oraz model VAR do analizy zmienności aprecjacji wartości domu jednorodzinnego. Odkryli, że zmienność ceny nieruchomości odpowiada zmianom w tempie wzrostu populacji. Miles (2008) rozważył inne modele nieliniowe poza modelem przełącznikowym Markowa, a po ocenie alternatywnych zastosował model GAR.

Odkrył, że model GAR radzi sobie lepiej w prognozowaniu poza próbą niż modele ARMA i GARCH w wielu przypadkach stosowane przez Crawforda i Fratantoniego (2003), zwłaszcza na rynkach tradycyjnie kojarzonych z dużą zmiennością cen mieszkań.

Niemniej jednak, wszystkie powyższe metody są uważane za dobre narzędzia jedynie przy spełnieniu pewnych warunków odnośnie zbiorów danych; w rzeczywistości dokładność jest bardzo zależna od właściwego i wystarczającego zbioru danych. A w przypadkach, w których występuje brak odpowiednio długiego horyzontu czasowego, dokładność wyników ulegnie zmniejszeniu. Stosowanie modeli szeregów czasowych, takich jak ARIMA, VAR itp., oraz modeli sieci neuronowych nie jest odpowiednie dla małych próbek danych. Ponieważ w studium przypadku jedyne dostępne dane zawierają 11x15 obserwacji, zastosowanie powyższych modeli może nie dać dokładnych wyników. Niniejsze badanie wykorzystuje podejście oparte na danych panelowych i modelowaniem ekonometrycznym panelowym, aby poznać zalety danych panelowych w modelowaniu ekonometrycznym i uniknąć wspomnianych problemów.

W następnej części pracy będą przeciwstawione prosty model liniowy, model z efektami stałymi oraz model z efektami losowymi. Następnie najlepszy z tych modeli zostanie użyty do prognozowania ceny 1m kwadratowego mieszkania w roku 20221. Poprawność prognozy będzie oceniona za pomocą klasycznych miar precyzji.

Rozdział III: Wyniki przeprowadzonych analiz

3.1. Analiza danych

Niniejsza praca koncentruje się na estymacji przeciętnej ceny 1 metru kwadratowego mieszkania w piętnastu największych miastach Rzeczypospolitej Polskiej: Białystok, Bydgoszcz, Częstochowa, Gdańsk, Gdynia, Katowice, Kraków, Lublin, Łódź, Poznań, Radom, Rzeszów, Szczecin, Warszawa, Wrocław. Dane zostały pozyskane z Głównego Urzędu Statystycznego.⁴ Analizowany okres obejmuje 11 lat od 2011 do 2020 roku, kumulując 165 łącznie obserwacji. Zestaw danych zawiera 12 zmiennych objaśniających. Opis wykorzystanych zmiennych przedstawiony został w Tabeli 3.1:

Tabela 3.1. Zmienne wykorzystane do modelowania przeciętnej ceny 1 metru kwadratowego mieszkania

Kategoria	Zmienna	Nazwa zakodowana	Miara
Infrastruktura	Udział powierzchni terenów zieleni w powierzchni miasta	green_area	%
	Miejsca noclegowe na 1 tys. ludności	night_stays	szt.
	Drogi dla rowerów na 10 tys. ludności	bike_roads	km
	Mieszkania oddane do użytkowania na 1 tys. ludności	buildings_introduced	szt.
Demografia	Przyrost małżeństw na 1 tys. ludności	marriages_increase	szt.
	Liczba studentów na 1 tys. ludności	students	os.
	Przyrost naturalny na 1 tys. ludności	birth_rate	os.
	Saldo migracji na 1 tys. ludności	balance_of_migration	os.
Ekonomia	Wynagrodzenie przeciętne brutto	mean_gross_salary	zł

⁴ <https://bdl.stat.gov.pl/bdl/dane/teryt/jednostka>, dostęp 16.06.2022

PKB per capita	pkb_pc	zł
Wskaźnik cen towarów i usług konsumpcyjnych	cpi	%
Udział bezrobotnych zarejestrowanych w wieku produkcyjnym	unemployment_rate	%

Źródło: opracowanie własne

Jak wynika z tabeli, zmienne pochodzą z trzech najbardziej istotnych obszarów kształtujących ceny na mieszkania: demografia, ekonomia oraz infrastruktura. W celach porównywalności wyników, większość zmiennych została zskalowana do liczby ludności. Wybór zmiennych w większości opierał się na pokrewne badania, przeprowadzone głównie przez instytucje państwowe, wielkie banki komercyjne oraz firmy konsultingowe.

Zróźnicowanie wartości zmiennych dla różnych miast można łatwo zaobserwować za pomocą odpowiedniego wykresu liniowego (Rysunek 3.1). Os x zawiera nazwy miast, natomiast os y oznacza uśrednione wartości 11-letnie dla poszczególnych miast.

Rysunek 3.1. Zróźnicowanie wartości zmiennych dla różnych miast

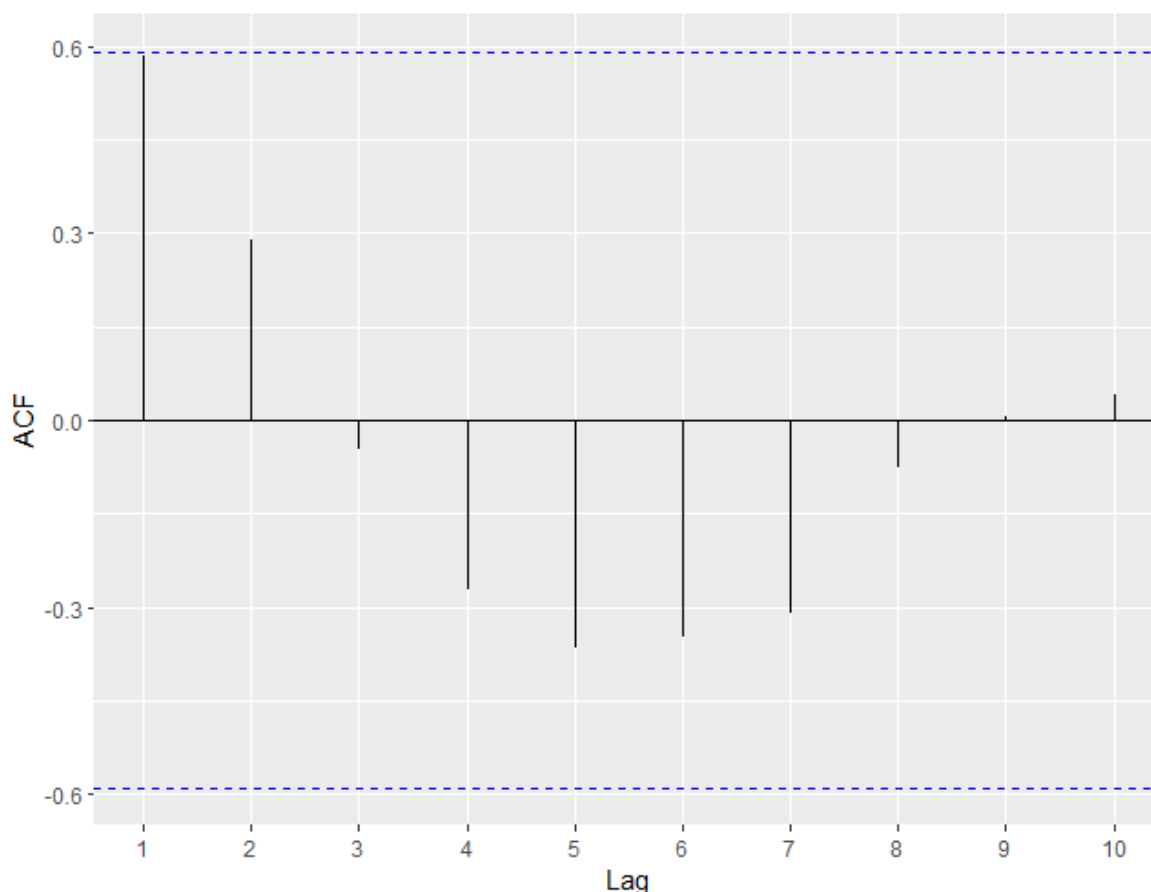


Źródło: opracowanie własne

Wykres wskazuje, że przyrost zawartych małżeństw na 1 tys. mieszkańców oraz liczba miejsc noclegowych na 1 tys. mieszkańców mogą okazać się zmiennymi nieistotnymi ze względu na swoją małą wariancję dla poszczególnych krajów, natomiast wszystkie pozostałe zmienne wykazują się dość dużym zróżnicowaniem.

Zanim przystąpić do wyboru finalnego modelu, należy omówić temat autokorelacji czasowej. Po pierwsze, zakres danych dotyczy tylko 11 lat obserwacji o częstotliwości rocznej, tzn. 11 obserwacji dla każdej jednostki badanej. Po drugie, jak wskazuje funkcja autokorelacji czasowej dla uśrednionej ceny mieszkania (z ang. autocorrelation function, ACF), przedstawiona na Rysunku 3.1, w danym zbiorze nie zachodzi autokorelacja zmiennej objaśnianej (słupki, odpowiadające za wartość funkcji autokorelacji dla poszczególnych opóźnień regresanta nie przekraczają poziomu ufności alfa równy 5%).

Rysunek 3.2. Autokorelacja czasowa uśrednionej ceny 1 metru kwadratowego mieszkania



Źródło: opracowanie własne

Powyższe faktory sprowadzą do konieczności zastosowania modeli klasy one-way do oszacowania przeciętnej ceny 1 metru kwadratowego mieszkania.

3.2. Wybór modelu

Wybór najlepszego modelu był prowadzony w trzy etapy. W pierwszej kolejności zostały zbudowane model zwykłej regresji liniowej estymowanej za pomocą metody najmniejszych kwadratów oraz model z efektami stałymi (ang. fixed effects model). Wyniki ocen estymatorów są przedstawione w Tabeli 3.2 oraz Tabeli 3.3.

Tabela 3.2. Wartości estymatorów modelu regresji liniowej

Zmienna	Estymator	Błąd standardowy	Statystyka t	Wartość p
(Intercept)	-370.79	3978.59	-0.09	0.926
green_area	4.52	23.77	0.19	0.850
night_stays	12.73	5.40	2.36	0.020
bike_roads	-255.94	59.53	-4.30	0.000
buildings_introduced	152.46	30.07	5.07	0.000
marriages_increase	-0.21	0.37	-0.57	0.571
students	1.48	1.37	1.08	0.281
birth_rate	14.27	32.91	0.43	0.665
balance_of_migration	-31.65	35.16	-0.90	0.370
mean_gross_salary	0.70	0.11	6.30	0.000
unemployment_rate	-237.63	42.80	-5.55	0.000
cpi	18.86	38.59	0.49	0.626
pkb_pc	0.01	0.01	2.21	0.029

Źródło: opracowanie własne

Tabela 3.3. Wartości estymatorów modelu z efektami stałymi

Zmienna	Estymator	Błąd standardowy	Statystyka t	Wartość p
green_area	-13.75	17.26	-0.80	0.427
night_stays	4.84	2.21	2.19	0.030
bike_roads	158.39	40.56	3.90	0.000
buildings_introduced	30.98	13.18	2.35	0.020
marriages_increase	-0.16	0.11	-1.46	0.146
students	11.16	2.20	5.07	0.000
birth_rate	-42.68	21.48	-1.99	0.049

balance_of_migration	-18.29	13.92	-1.31	0.191
mean_gross_salary	0.92	0.11	7.99	0.000
unemployment_rate	-26.48	16.49	-1.61	0.111
cpi	78.03	12.98	6.01	0.000
pkb_pc	-0.01	0.01	-1.68	0.095

Źródło: opracowanie własne

Jak wynika z oszacowań wartości p , w przypadku modeli regresji liniowej tylko 5 z 12 zmiennych objaśniających są traktowane jako zmienne statystycznie istotne dla poziomu ufności 5%, natomiast dla modelu z efektami stałymi takich zmiennych jest siedem, co między inn. może skutkować utratą istotnej informacji zawartą w zmiennych losowych, w przypadku gdy zastosowany zostałby model regresji liniowej. Jednak porównania empiryczne za pomocą współczynnika determinacji skorygowanego R^2_{adj} (0.7148 dla modelu liniowego przeciw 0.85498 z modelu z efektami stałymi), medialnej wartości reszt z modelu ϵ_{median} (-46,38 w modelu liniowym oraz -15,592 w modelu z efektami stałymi) oraz testu F na istotność efektów indywidualnych ze statystyką testową równą 124,84 przy stopniach swobody (14, 138) i wartością $p < 0.0002$ wskazują na sensowność stosowania modelu z efektami stałymi.

Następnie, oszacowany został model z efektami losowymi (z ang. random effects model). Oszacowania parametrów wraz z błędem standardowym estymatorów oraz istotnością zmiennych są przedstawione w Tabeli 3.4.

Tabela 3.4. Wartości estymatorów modelu z efektami losowymi

Zmienna	Estymator	Błąd standardowy	Statystyka z	Wartość p
(Intercept)	-8010.24	1443.31	-5.55	0.000
green_area	-13.99	18.90	-0.74	0.459
night_stays	6.91	2.60	2.66	0.008
bike_roads	67.92	41.32	1.64	0.100
buildings_introduced	32.35	14.92	2.17	0.030
marriages_increase	-0.15	0.13	-1.19	0.232
students	6.48	1.59	4.09	0.000
birth_rate	-59.24	23.64	-2.51	0.012
balance_of_migration	-18.39	16.26	-1.13	0.258

mean_gross_salary	0.75	0.11	6.83	0.000
unemployment_rate	-40.37	18.92	-2.13	0.033
cpi	84.58	14.53	5.82	0.000
pkb_pc	0.00	0.01	-0.06	0.954

Źródło: opracowanie własne

W przypadku modelu z efektami losowymi, istotnymi okazały się te same zmienne objaśniające. W celu podejmowania decyzji dotyczącej wyboru modelu najlepszej postaci wykorzystany został test Hausmana z hipotezą zerową mówiącą o tym, że oba estymatory (FE i RE) są nieobciążone, ale wówczas RE jest bardziej efektywny. Wynik testu ze statystyką testową $X^2 = 16.836$, przy dwunastu stopniach swobody i wartości p wynoszącej 0,156 hipoteza zerowa została odrzucona, co oznacza, że model z efektami losowymi jest efektywniejszy. Dodatkowym umocnieniem powyższego stwierdzenia posłużyły porównania mediany błędów z obu modeli, z czego mediana dla modelu z efektami stałymi (-46,38) była wyższa niż dla modelu z efektami losowymi (-41,52). W konsekwencji, najlepszym względem testów oraz precyzji oszacowań okazał się model z efektami losowymi.

Przed stworzeniem końcowego modelu zostały usunięte wszystkie zmienne nieistotne. Ostateczne parametry modelu oraz parametry rozkładu efektów losowych zostały przedstawione w Tabeli 3.5 i 3.6 odpowiednio.

Tabela 3.5. Wartości estymatorów modelu ostatecznego

Zmienna	Estymator	Błąd standardowy	Statystyka z	Wartość p
(Intercept)	-7878.35	1304.40	-6.04	0.000
night_stays	6.48	2.33	2.78	0.005
buildings_introduced	26.89	12.89	2.09	0.037
students	6.67	1.60	4.16	0.000
birth_rate	-60.01	20.56	-2.92	0.004
mean_gross_salary	0.81	0.06	14.11	0.000
unemployment_rate	-37.16	16.76	-2.22	0.027

cpi	80.95	13.40	6.04	0.000
------------	-------	-------	------	-------

Źródło: opracowanie własne

Tabela 3.6. Wartości rozkładu efektów losowych

Efekty	Wariancja	Odchylenie standardowe	Udział	Teta
idiosynkratyczne	55169.6	234.9	0.085	0.9084
indywidualne	592609.7	769.8	0.915	0.9084

Źródło: opracowanie własne

Wartości p wskazują na statystyczną istotność wszystkich użytych zmiennych objaśniających. Wartości estymatorów dla większości zmiennych nie załamały założeń teoretycznych. Przy wzroście liczby miejsc do noclegu na 1 tys. osób cena 1m kwadratowego też się wzrastała, co można zauważyć na przykładzie miasta turystycznego Gdańsku. Z interpretacji modeli wynika także, że wzrost liczby budynków oddanych do użytkowania podbija też przeciętne ceny 1m kwadratowego mieszkania, co oznacza, że wzrost ilościowy pociąga za sobą wzrost wartościowy. Patrząc na współczynnik przyrostu naturalnego na 1000 osób, należy wziąć pod uwagę ten fakt, że dla większości miast przyjmował on wartości ujemne. Wobec tego należy interpretować ten parametr ze znakiem odwrotnym: wzrost współczynnika przyrostu naturalnego na 1000 osób o jednostkę powoduje zwiększenie ceny mieszkania o 60 zł przeciętnie przy pozostałych czynnikach niezmiennych. Zasadniczo interesujący jest wpływ przeciętnego wynagrodzenia brutto. Skoro wartość estymatora jest mniejsza od jedności, oznacza to że przeciętna cena mieszkania wzrastała się wolniej niż przeciętne zarobki.

3.3. Porównywanie wartości dopasowanych i prognozy z najlepszego modelu

Na podstawie powyżej opisanego końcowego modelu z efektami losowymi porównywane zostały wartości prognozowane z modelu z wartościami rzeczywistymi. W tym celu zostały wykorzystane wcześniej wykluczone wartości za rok 2021, które stanowiły zbiór testowy. Jako miary dokładności wybrane zostały różnica wartościowa w jednostkach naturalnych (błąd średniokwadratowy, ang. RMSE, root mean square error) oraz różnica procentowa (ang. MAPE, mean absolute percentage error). Wyniki są przedstawione w Tabeli 3.7.

Tabela 3.7. Wartości rozkładu efektów losowych

Miasto	Wartość rzeczywista	Prognoza	Różnica procentowa	Różnica wartościowa
Białystok	8742	6431.51	26.43	2310.49
Bydgoszcz	6574	6755.95	2.77	181.95
Częstochowa	7058	7937.03	12.45	879.03
Gdańsk	10057	9067.61	9.84	989.39
Gdynia	9742	8906.53	8.58	835.47
Katowice	8533	7702.45	9.73	830.55
Kraków	10292	10099.15	1.87	192.85
Lublin	9135	7116.74	22.09	2018.26
Łódź	5992	7352.24	22.7	1360.24
Poznań	7772	8760.65	12.72	988.65
Radom	6240	5713.68	8.43	526.32
Rzeszów	7570.08	7528.53	0.55	41.56
Szczecin	9328	7518.7	19.4	1809.3
Warszawa	12060	10947.3	9.23	1112.7
Wrocław	8631	8617.11	0.16	13.89

Źródło: opracowanie własne

Na podstawie wyników porównywań widać, że dla dużej części (40%) miast różnica procentowa przewyższała 10 procent, co może świadczyć o dość słabej mocy predykcyjnej modelu dla wskazanych zmiennych. Natomiast w przypadku Wrocławia, Rzeszowa, Krakowa i Bydgoszczy wartość procentowa nie przekroczyła trzech procenty w ujęciu procentowym i 200 złotych w ujęciu absolutnym, co świadczy o dobrym dopasowaniu modelu do predykcji ceny średniej za metr kwadratowy w wybranych miastach.

Podsumowanie

Niniejsza praca koncentruje się na analizie rynku nieruchomości dla piętnastu największych miast Polski w latach 2010-2020. W tym celu oszacowane zostały modele regresji liniowej, model z efektami stałymi oraz model z efektami losowymi, gdzie w wyniku porównywań najlepszym modelem okazał się model z efektami losowymi. Wyniki empiryczne porównywań dla roku testowego 2021, dla którego była dokonana prognoza, świadczyły o dość dokładnej prognozie dla większości badanych miast. Jednak dla sześciu miast (Białystok, Częstochowa, Lublin, Łódź, Poznań, Szczecin) średni absolutny procentowy błąd prognozy wskazywał na słabą moc predykcyjną modelu dla wspomnianych miast. Może to świadczyć m.in. o czynnikach nieuwzględnionych, np. związanych z epidemią COVID-19. Ta hipoteza może się potwierdzić np. w przypadku Łodzi, gdzie nieoczekiwany wzrost ceny mieszkań mógł być spowodowany wzrostem liczby osób przebywających z Warszawy, gdyż praca zdalna umożliwiała na łatwą relokację.⁵ Natomiast uwzględnienie dodatkowych czynników wymaga zwielokrotnienie liczebności próby. W tym celu istotnym było by zwiększenie nie zakresu próby (dodając z kolei wartości dla lat wcześniejszych niż rok 2010), lecz zwiększenie częstotliwości próby, np. o częstotliwość miesięczną. Pozwoliło by to na pomnożenie, liczby obserwacji dla każdej jednostki badanej (miasta) z 11 do 132, tym samym potencjalnie zwiększając stabilność prognoz w czasie i moc testów statystycznych. Na koniec, warto było by uwzględnić czynniki infrastrukturalne na poziomie mikroekonomicznym za pomocą modeli przestrzenno-czasowych. W przypadku niektórych miast mogło by to okazać istotny wpływ na precyzję średniej ceny 1m. kwadratowego mieszkania mianowicie w przypadku miast o dużej heterogeniczności mieszkań.

Bibliografia

1. Campbell J.Y., Coco J.F., *How do house prices affect consumption? Evidence from micro data*, *Journal of Monetary Economics*, Wyd. 54 Nu. 3, 2006, str. 591-621.
2. Crawford G., Fratantoni M. *Assessing the forecasting performance of regime-switching, ARIMA, GARCH models of house prices*, *Real Estate Economics*, Wyd. 31, 2003, str. 223-43.
3. Evans A., James H., Collins A., *Artificial neural networks: an application to residential valuation in the UK*, *Journal of Property Valuation & Investment*, Wyd. 11, 1992, str. 195-204.
4. Follain J., Giertz S., *Predicting House Price Bubbles and Busts with Econometric Models: What We've Learned. What We Still Don't Know*, Lincoln Institute of Land Policy, 2012 Massachusetts.
5. Gardiner C., Luo Z., Lee A. *Fixed effects, random effects and GEE: What are the differences?*, *Statistics in Medicine*. Wyd. 28 Nu. 2, 2009. 221–239
6. Grabowiecka-Łaszek A., Krzysztofik P., Matysiak W., *Nieruchomości Mieszkaniowe*, PKO Bank Polski S.A, 2021 Warszawa.
7. Hacievliyagil N., Drachal K., Halil I., *Predicting House Prices Using DMA Method: Evidence from Turkey*, Faculty of Economics, University, 2014 Malatya.
8. Hadavandi E., Mirjani M., *An econometric panel data-based approach for housing price forecasting in Iran*, *International Journal of Housing Markets Analysis*, 2011 Wyd 4., str. 70-83.
9. Jadach-Sepiolo A., *Analiza polskiego rynku nieruchomości ze szczególnym uwzględnieniem nieruchomości przeznaczonych lub wykorzystywanych na cele mieszkaniowe za lata 2012-2014 z prognozą na rok 2015*, Instytut rozwoju miast, 2014 Warszawa.
10. Kucharska-Stasiak E., *Nieruchomości w gospodarce rynkowej*, PWN Warszawa 2006, str. 38-55.
11. McCluskey W., *Predictive accuracy of machine learning models for the mass appraisal of residential property*, *New Zealand Valuers Journal*, 1996, str. 41-7.

12. Miles, M., *Boom-bust cycles, the forecasting performance of linear, non-linear models of house prices*, Journal of Real Estate Finance & Economics, Wyd. 32, 2008, str. 49-264
13. Miller N., Peng L., *Exploring metropolitan housing price volatility*, Journal of Real Estate , Financial Economics, 2006, str. 5-18.
14. Peng C., *The long- and short-run influences of housing prices on migration*, National Taipei University, Taipei 2019, str. 253-262.
15. Yan Y., Song Y., Zhang, W., *Method for housing price forecasting based on TEI@I methodology*, Systems Engineering – Theory& Practice, Wyd. 27 Nu. 7, 2007 str. 1-9.
16. Yoo S., *A robust estimation of hedonic price functions: least absolute deviations estimation*, Applied Economics Letters, Wyd. 8, 2001, str. 55-58.

Spis rysunków

Rysunek 3.1. Zróźnicowanie wartości zmiennych dla różnych miast	12
.....	12
Rysunek 3.2. Autokorelacja czasowa uśrednionej ceny 1 metru kwadratowego mieszkania	
.....	13
.....	13

Spis tabeli

Tabela 3.1. Zmienne wykorzystane do modelowania przeciętnej ceny 1 metru kwadratowego mieszkania	11
Tabela 3.2. Wartości estymatorów modelu regresji liniowej	14
Tabela 3.3. Wartości estymatorów modelu z efektami stałymi	14
Tabela 3.4. Wartości estymatorów modelu z efektami losowymi	15
Tabela 3.5. Wartości estymatorów modelu ostatecznego	16
Tabela 3.6. Wartości rozkładu efektów losowych	17
Tabela 3.7. Wartości rozkładu efektów losowych	17