



AKADEMIA GÓRNICZO-HUTNICZA IM. STANISŁAWA STASZICA W KRAKOWIE

Wydział Zarządzania

Ekonometria Przestrzenna

Projekt 4

Krzysztof Księżyc, Karol Kuciński, Maciej Nagły

Wstęp:

Projekt 4 stanowi rozszerzenie poprzedniej analizy rynku nieruchomości o elementy ekonometrii przestrzennej. Model regresji wielorakiej zostanie uzupełniony o aspekty zależności przestrzennych między powiatami.

Cel projektu:

1. Identyfikacja zależności przestrzennych w cenach nieruchomości
2. Weryfikacja wpływu lokalizacji na ceny mieszkań
3. Porównanie modeli przestrzennych (SAR, SEM) z modelem klasycznym

Przygotowanie danych:

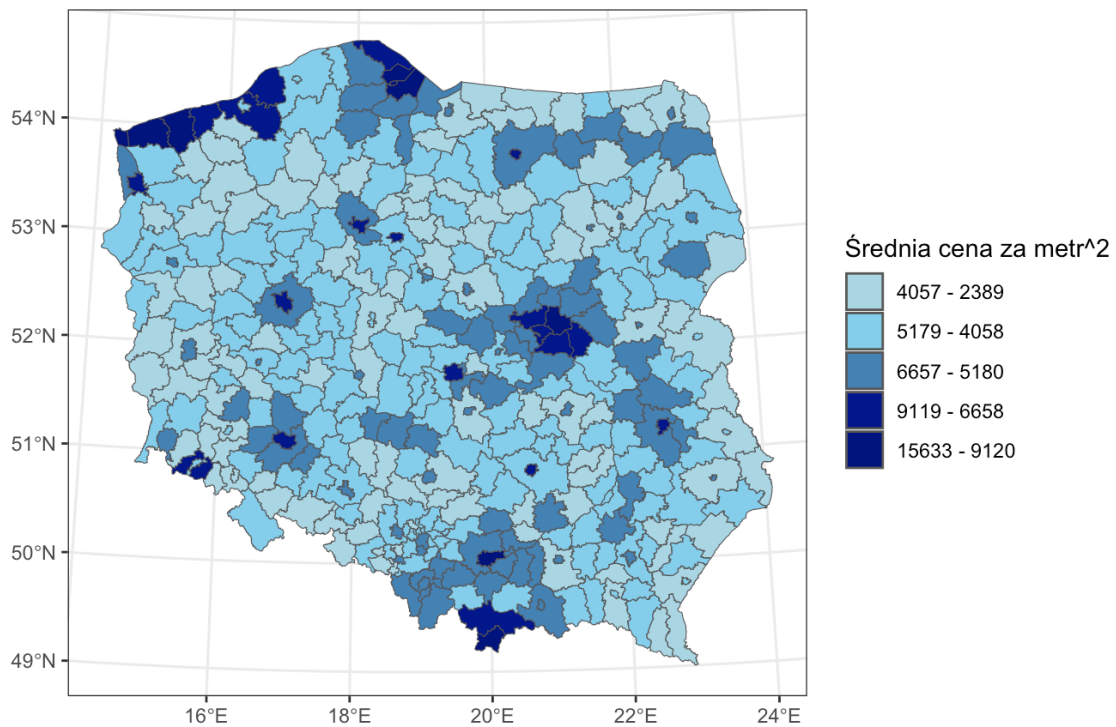
Wykorzystano dane z projektu 3 (ceny za m², wynagrodzenia, liczba oddanych mieszkań, liczba małżeństw). Następnie wykorzystano macierz wag przestrzennych w oparciu o sąsiedztwo powiatów. Wykonano wizualizacje przestrzenne dla wszystkich zmiennych w postaci map, które pozwoliły na wstępną identyfikację skupień przestrzennych. Przeprowadzono standaryzację macierzy wag do dalszej analizy- typ W- list2nb.

W stosunku do poprzedniego projektu, dokonano modyfikacji zmiennej 'małżeństwa' poprzez jej zlogarytmowanie. Transformacja ta miała na celu poprawę dopasowania modelu i linearyzację zależności między zmiennymi. Logarytmowanie zmiennej 'małżeństwa' pozwoliło na redukcję skośności rozkładu i lepsze uchwycenie relacji z cenami nieruchomości.

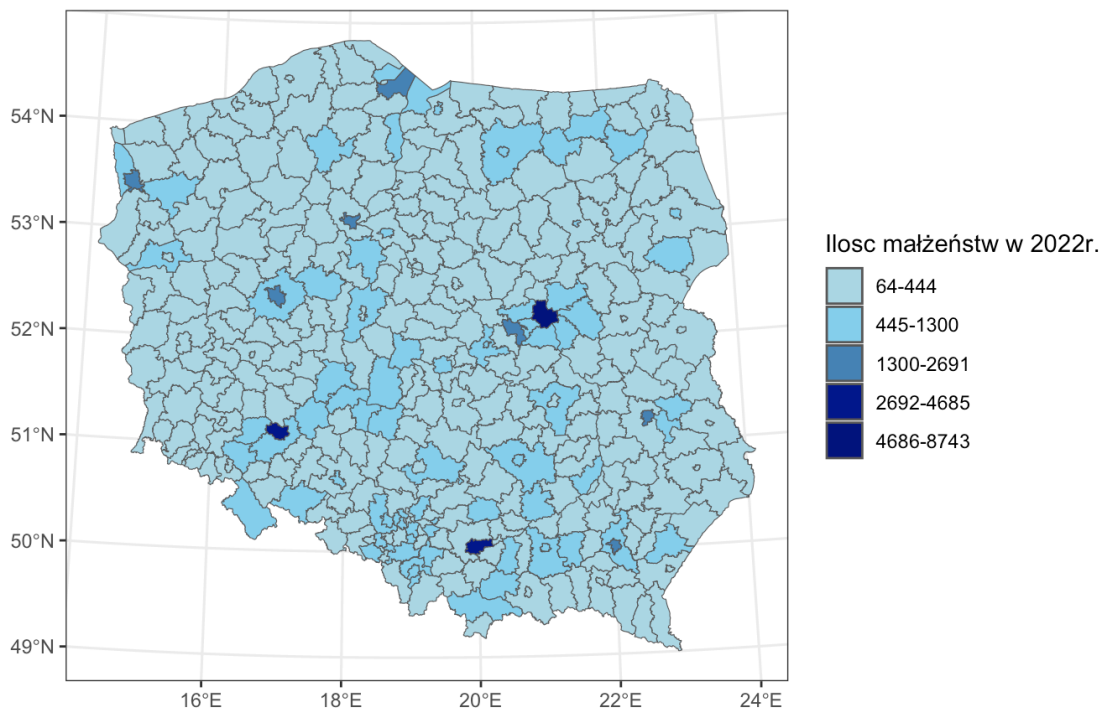
Model klasyczny:

$$\text{metr} = 484.03 + 0.4007 \text{ wynagrodzenie} + 0.42965 \text{ oddane} + 312.08 \log(\text{małżeństwa}) + \text{epsilon}$$

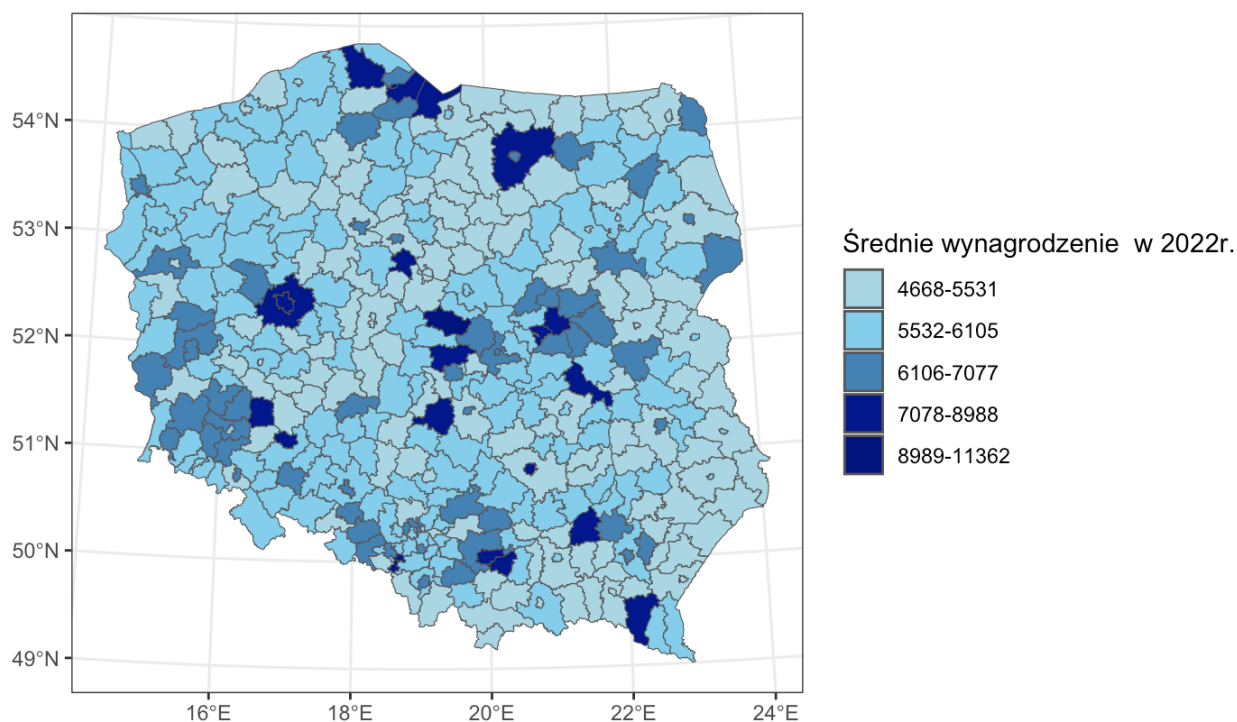
Model liniowy wykazuje, że wszystkie zmienne objaśniające są istotne statystycznie, co wskazuje na ich znaczący wpływ na zmienną objaśnianą metr. Współczynniki regresji wskazują, że wzrost wynagrodzenia o jednostkę zwiększa metr średnio o 0.401, wzrost oddane o jednostkę zwiększa metr o 0.430, a wzrost logarytmu małżeństwa o jednostkę zwiększa metr o 312.08. Wskaźnik dopasowania $R^2 = 0.3097$ oznacza, że model wyjaśnia około 31% zmienności zmiennej objaśnianej, co wskazuje na umiarkowaną siłę modelu. Reszty modelu wykazują standardowy błąd wynoszący 1294, co wskazuje na pewne rozproszenie wyników wokół przewidywanej linii regresji.



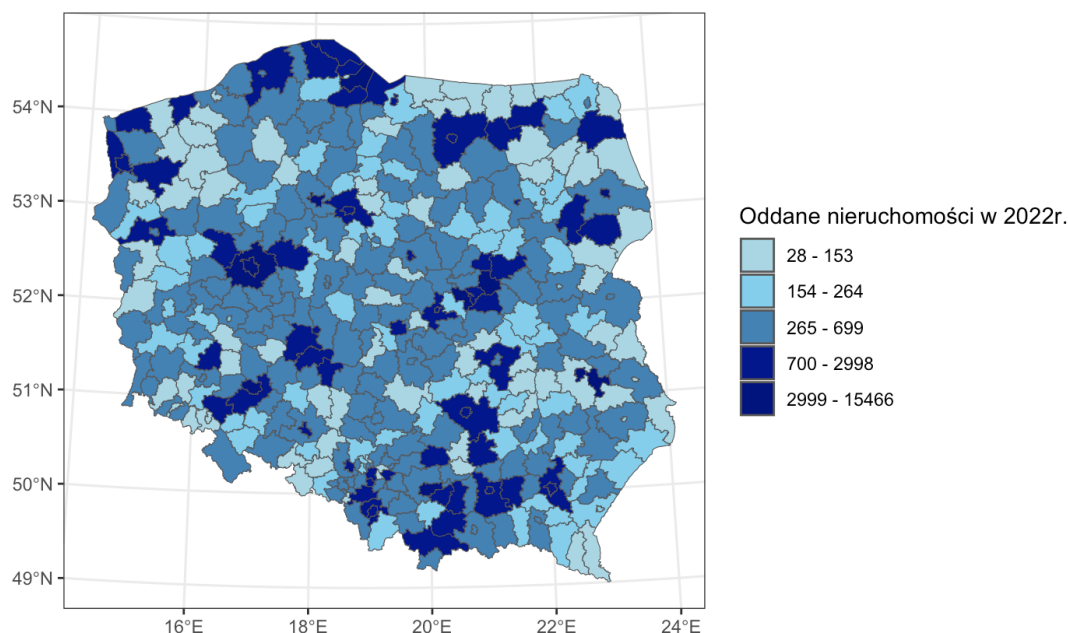
Mapa przedstawia przestrzenny rozkład średnich cen za metr kwadratowy nieruchomości w 2022 roku. Najwyższe ceny (ciemnoniebieski) występują w Warszawie i okolicach, dużych miastach wojewódzkich oraz niektórych obszarach nadmorskich. Najniższe ceny (jasnoniebieski) dominują w regionach wiejskich i mniej rozwiniętych gospodarczo, szczególnie w Polsce wschodniej.



Mapa pokazuje liczbę zawartych małżeństw w 2022 roku w podziale na powiaty. Najwięcej małżeństw (ciemnoniebieski) zawarto w dużych aglomeracjach miejskich, takich jak Warszawa, Kraków i Wrocław. Regiony o mniejszej liczbie ludności, szczególnie wiejskie i oddalone od dużych miast, cechują się mniejszą liczbą małżeństw (jasnoniebieski).



Mapa ilustruje przestrzenny rozkład średnich wynagrodzeń w powiatach Polski. Najwyższe wynagrodzenia (ciemnoniebieski) występują w dużych miastach wojewódzkich (Warszawa, Wrocław, Gdańsk) oraz niektórych bogatych powiatach (np. powiaty górnicze). Obszary wiejskie i mniejsze powiaty charakteryzują się niższymi wynagrodzeniami, szczególnie w Polsce wschodniej i północno-wschodniej.



Mapa przedstawia liczbę nieruchomości oddanych do użytku w 2022 roku w podziale na powiaty. Najwięcej nieruchomości (ciemnoniebieski) oddano w Warszawie, Trójmieście, Krakowie oraz w powiatach otaczających duże miasta. Najmniej nieruchomości (jasnoniebieski) oddano w regionach wiejskich i peryferyjnych, gdzie rozwój budownictwa jest mniej intensywny.

Model przestrzenny

Statystyka Morana(I)

Test Type	Moran I Statistic	Expectation	Variance	Z-score	p-value	Alternative
Randomisation	0.00961	-0.00264	0.00110	0.3696	0.3558	greater
Global	0.00961	-0.00266	0.00116	0.3608	0.7183	two-sided

Interpretacja wyników

Moran I test under randomisation bardzo blisko zera, brak wyraźnej autokorelacji, a p-value nieistotne, brak podstaw do odrzucenia hipotezy zerowej o braku autokorelacji

Global Moran I for regression residuals również blisko zera p-value znów pokazuje że wartość nieistotne, brak podstaw do odrzucenia hipotezy zerowej.

Oba testy wskazują, że autokorelacja przestrzenna w resztach modelu jest nieistotna. Różnice w wynikach testów wynikają z przyjętej metody estymacji wariancji (randomizacja vs. globalna metoda), ale w obu przypadkach Moran I jest bliski zera. Wartość p-value dla testu

Morana w modelu z logarytmowaną zmienną jest wyższa co sugeruje lepszą eliminację autokorelacji przestrzennej w resztach. Model z log(małżeństwa) lepiej pasuje do danych.

SAR Model Coefficients	Estimate	Std. Error	z value	p-value
(Intercept)	494.8160	984.1083	0.5028	0.6151
wynagrodzenie	0.4012	0.1033	3.8824	0.0001
oddane	0.4294	0.0754	5.6950	<0.0001
log(malzenstwa)	312.4020	147.2574	2.1215	0.0339

Model SAR

Wszystkie zmienne objaśniające są statystycznie istotne:

wynagrodzenie: Współczynnik 0.401 , istotny ($p = 0.0001$).

oddane: Współczynnik 0.429 , bardzo istotny ($p < 0.0001$).

log(malzenstwa): Współczynnik 312.40 , istotny ($p = 0.0339$).

Współczynnik przestrzenny $\rho = -0.0031$, $p = 0.9601$:

Współczynnik ρ jest odpowiedzialny za występowanie zależności przestrzennych: pokazuje on brak istotnych zależności przestrzennych w zmiennej objaśnianej. Model SAR nie przynosi korzyści względem klasycznego modelu liniowego.

Dopasowanie modelu:

Log-likelihood: -3260.155.

AIC: 6532.3 , wyższe niż w modelu klasycznym (6530.3), co wskazuje na brak przewagi modelu SAR.

Test LM dla reszt: $p = 0.5471$, brak istotnych zależności przestrzennych w resztach.

SEM Model Coefficients	Estimate	Std. Error	z value	p-value
(Intercept)	509.5856	940.3005	0.5419	0.5879
wynagrodzenie	0.3983	0.1033	3.8559	0.0001
oddane	0.4313	0.0753	5.7250	<0.0001
log(malzenstwa)	309.7478	147.3656	2.1019	0.0356

Model SEM

Współczynniki:

Wszystkie zmienne objaśniające są statystycznie istotne:

wynagrodzenie: Współczynnik 0.398 , istotny ($p = 0.0001$).

oddane: Współczynnik 0.431 , bardzo istotny ($p < 0.0001$).

log(malzenstwa): Współczynnik 309.75 , istotny ($p = 0.0356$).

Współczynnik przestrzenny - $\lambda = 0.0135$, $p = 0.8595$:

Lambda pokazuje brak istotnych zależności przestrzennych w resztach. Model SEM również nie przynosi istotnej poprawy względem klasycznego modelu.

Dopasowanie modelu:

Log-likelihood: -3260.132, bardzo zbliżone do SAR.

AIC: 6532.3 , wyższe niż w modelu klasycznym (6530.3).

Test LM dla reszt: brak istotnych zależności przestrzennych ($p = 0.82397$).

Wnioski ekonometryczne z modelu SEM i SAR:

Modele przestrzenne (SAR, SEM): Współczynniki zmiennych objaśniających są istotne, ale modele nie poprawiają dopasowania względem klasycznego modelu liniowego. Brak istotnych zależności przestrzennych (ρ i λ są nieistotne).

Model klasyczny (LM): AIC niższe niż w SAR/SEM. Reszty z modelu klasycznego nie wykazują istotnej autokorelacji przestrzennej (test LM dla reszt).

Klasyczny model liniowy jest wystarczający i lepiej dopasowany niż modele przestrzenne. Modele SAR/SEM nie są potrzebne w tej analizie.

Test	Value	p-value	Interpretation
LMerr (RSerr)	0.0785	0.7793	No significant spatial error dependence
LMlag (RSlag)	0.0048	0.9449	No significant spatial lag dependence
RLMerr (adjRSerr)	0.3369	0.5616	No robust evidence of error dependence
RLMlag (adjRSlag)	0.2631	0.6080	No robust evidence of lag dependence
SARMA	0.3417	0.8430	No evidence of combined spatial effects

Test LMerr -Testuje występowanie zależności przestrzennej w składniku błędu. Wysoka wartość p wskazuje na brak istotnej autokorelacji przestrzennej w błędach.

Test LMlag -Bada brakującą przestrzennie opóźnioną zmienną zależną. P-value znacznie powyżej 0.05 sugeruje brak potrzeby uwzględnienia opóźnionej zmiennej zależnej w modelu.

Test RLMerr Odporna wersja testu na zależność przestrzenną błędów, uwzględniająca potencjalną obecność opóźnionej zmiennej zależnej. Wynik potwierdza brak autokorelacji przestrzennej błędów.

Test RLMlag Odporna wersja testu na brakującą opóźnioną zmienną zależną przy możliwej obecności autokorelacji błędów. Wynik wskazuje na brak potrzeby modyfikacji modelu o składnik przestrzenny.

Test SARMA Test łączny badający jednocześnie oba efekty przestrzenne. Wysoka wartość p wskazuje na brak jakichkolwiek istotnych efektów przestrzennych w modelu.

Podsumowanie testów zależności autokorelacji

Przeprowadzona analiza zależności przestrzennych, mimo zastosowania różnorodnych testów (LMerr, LMlag, RLMerr, RLMlag, SARMA), nie wykazała istotnych efektów przestrzennych w badanym modelu. Wszystkie testy diagnostyczne wskazują na brak autokorelacji przestrzennej, zarówno w składniku błędu, jak i w zmiennej zależnej. Sugeruje to, że klasyczny model regresji wielorakiej może być wystarczający do opisu badanego zjawiska, a ceny nieruchomości w danym powiecie nie wykazują systematycznego związku z cenami w powiatach sąsiednich.

Pseudo R²:

Pseudo- R² w modelach przestrzennych takich jak SAR (Spatial Autoregressive Model) i SEM (Spatial Error Model) to miara, która pozwala ocenić dopasowanie modelu, podobnie jak klasyczne R² w modelach liniowych. W modelach przestrzennych klasyczne R² nie jest bezpośrednio dostępne ze względu na dodatkowe komponenty (np. autoregresję przestrzenną lub zależności w resztach), dlatego stosuje się jego odpowiedniki.

Pseudo-R² dla SAR: 0.3015913

Pseudo-R² dla SEM: 0.3019328

Wartości Pseudo-R² dla obu modeli przestrzennych są bardzo zbliżone do R² modelu klasycznego (0.3064), co dodatkowo potwierdza brak istotnych efektów przestrzennych. Model SEM wykazuje minimalnie lepsze dopasowanie (0.3019) niż model SAR (0.3016), jednak różnica jest znikoma i nie uzasadnia komplikacji modelu o składnik przestrzenny.

Podsumowanie projektu

W projekcie oszacowano model liniowy (OLS) opisujący średnią cenę za metr kwadratowy nieruchomości w Polsce w zależności od takich zmiennych, jak średnie wynagrodzenie, liczba oddanych nieruchomości oraz logarytm liczby małżeństw. Analiza reszt modelu OLS, przeprowadzona za pomocą testu Morana, nie wykazała istotnej autokorelacji przestrzennej, co sugeruje brak globalnych zależności przestrzennych w danych.

Oszacowano również modele przestrzenne: SAR (uwzględniający autoregresję przestrzenną zmiennej objaśnianej) oraz SEM (uwzględniający przestrzenne zależności w resztach). Współczynniki przestrzenne ρ dla SAR i λ dla SEM okazały się statystycznie nieistotne, co wskazuje na brak potrzeby uwzględniania zależności przestrzennych w modelu.

Porównanie modeli na podstawie AIC i log-likelihood potwierdziło, że model klasyczny (OLS) jest równie dobrze dopasowany, co modele przestrzenne, przy jednocześniej prostszej specyfikacji. Wszystkie zmienne objaśniające w modelach były istotne statystycznie, co wskazuje na ich wpływ na ceny nieruchomości.

Wnioski potwierdzają, że w analizowanych danych przestrzenne zależności są znikome, a model OLS jest wystarczający do opisu zależności między zmiennymi. Modele SAR i SEM nie wnoszą istotnej poprawy w dopasowaniu i są zbędne w przypadku tego zbioru danych. Warto się jednak zastanowić, czy na pewno dane i model pozwala nam w 100 procentach zidentyfikować model jako nieistotny przestrzennie, mając świadomość populacji oraz potencjału środowiska największych powiatów-miast.