

Marek Kulczycki\*, Marcin Ligas\*

## Regresja ważona geograficznie jako narzędzie analizy rynku nieruchomości

### 1. Wprowadzenie

Świadomość istnienia lokalnych rynków nieruchomości nie zawsze idzie w parze z wykorzystywaniem odpowiednich metod statystycznych stosowanych w analizie rynku nieruchomości oraz estymacji parametrów modeli wyceny. Choć wiadomo, że rynki nieruchomości wykazują przestrzenną heterogeniczność (przestrzenna heteroskedastyczność) oraz przestrzenną autokorelację, na gruncie teorii wyceny nieruchomości (szczególnie w Polsce) daje się zaobserwować brak wykorzystania narzędzi statystyki oraz ekonometrii umożliwiających rozwiązywanie problemów istniejących w zbiorze informacji rynkowych.

Klasyczne modele regresyjne stosowane do modelowania wartości rynkowej nieruchomości nie uwzględniają potencjalnych interakcji, jakie mogą zachodzić pomiędzy jednostkami obserwacji (nieruchomościami) – autokorelacja przestrzenna, oraz zakładają „stałość” procesu związanego z kształtowaniem cen w przestrzeni geograficznej.

Takie podejście owocuje jednym globalnym modelem dla całego rynku nieruchomości potencjalnie niosącym ze sobą nieuwzględnione właściwości zbioru informacji rynkowych. Powoduje, że interpretacja wyników estymacji na podstawie tak zbudowanego modelu może okazać się problematyczna.

Dane przestrzenne, a takimi z całą pewnością są dane o nieruchomościach, powinny być analizowane metodami adekwatnymi do ich charakteru. Celem niniejszego artykułu jest wprowadzenie w technikę eksploracyjną statystyki przestrzennej zwaną regresją ważoną geograficznie (GWR – *Geographically Weighted Regression*). Technika ta umożliwia bezpośrednie modelowanie przestrzennej heterogeniczności poprzez sekwencyjne lokalne dopasowywanie linii regresji.

---

\* Wydział Geodezji Górniczej i Inżynierii Środowiska, Akademia Górniczo-Hutnicza, Kraków

## 2. Klasyczne (globalne) modele wyceny

Klasyczne metody zakładają stacjonarność procesu generującego dane dla całego obszaru opracowania, nie biorąc pod uwagę przestrzennego charakteru danych. Bardzo często zależność wartości nieruchomości od jej cech charakterystycznych wyraża się za pomocą addytywnej funkcji liniowej, co może być zapisane w postaci równania (*hedonic price model*)

$$y_i = \beta_0 + \sum_{k=1}^u x_{ik} \cdot \beta_k + \varepsilon_i \quad (2.1)$$

lub w postaci macierzowej (układ równań złożony z równań postaci (2.1)):

$$y = X \cdot \beta + \varepsilon \quad (2.2)$$

gdzie:

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} \text{ – wektor obserwowanych jednostkowych cen nieruchomości,}$$

$$X = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1u} \\ 1 & x_{21} & \cdots & x_{2u} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{N1} & \cdots & x_{Nu} \end{bmatrix} \text{ – macierz zmiennych objaśniających (cech nieruchomości),}$$

$$\beta = \begin{bmatrix} \beta \\ \beta_1 \\ \vdots \\ \beta_u \end{bmatrix} \text{ – wektor szacowanych parametrów strukturalnych modelu,}$$

$$\varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{bmatrix} \text{ – wektor zakłóceń modelu (składnik losowy).}$$

Estymację parametrów modelu (2.2) przeprowadza się za pomocą metody najmniejszej sumy kwadratów (MNK), zgodnie z warunkiem

$$\sum_{i=1}^N \left[ y_i - \left( \beta_0 + \sum_{k=1}^u x_{ik} \cdot \beta_k \right) \right]^2 = \varepsilon^T \varepsilon = (y - X\beta)^T (y - X\beta) = \min \quad (2.3)$$

Przy spełnionych założeniach Gaussa–Markowa:

$$E(\varepsilon) = 0 \Leftrightarrow \varepsilon(y) = X\beta,$$

$$\text{cov}(\varepsilon) = \sigma^2 I \Leftrightarrow \text{cov}(y) = \sigma^2 I.$$

Parametry modelu (2.2) szacowane MNK są najlepszymi nieobciążonymi estymatorami liniowymi (BLUE – *Best Linear Unbiased Estimator*).

Rozwiązanie układu równań (2.2), według MNK prowadzi do następujących zależności

$$\hat{\beta} = (X^T X)^{-1} X^T y \quad (2.4)$$

Wariancja reszt oraz odchylenie standardowe reszt zwane również błędem szacunku równania, wyznaczane są odpowiednio według wzorów:

$$\hat{\sigma}_o^2 = \frac{\hat{\varepsilon}^T \hat{\varepsilon}}{N - u - 1} = \frac{(y - X\hat{\beta})^T (y - X\hat{\beta})}{N - u - 1} \quad (2.5)$$

$$\sigma_o = \sqrt{\frac{\hat{\varepsilon}^T \hat{\varepsilon}}{N - u - 1}} = \sqrt{\frac{(y - X\hat{\beta})^T (y - X\hat{\beta})}{N - u - 1}}$$

gdzie:

$N$  – liczba nieruchomości służących do konstrukcji modelu,

$u$  – liczba rozpatrywanych cech nieruchomości.

Macierz kowariancji wyznaczonych parametrów modelu regresji określa się według następującego wzoru

$$\text{cov}(\hat{\beta}) = \hat{\sigma}_o^2 \cdot (X^T X)^{-1} \quad (2.6)$$

Elementy diagonalne macierzy kowariancji (2.6) są oszacowaniami wariancji odpowiednich parametrów strukturalnych modelu, zaś pierwiastki z elementów diagonalnych są standardowymi błędami szacunku parametrów.

W oparciu o elementy macierzy kowariancji dla współczynników regresji, określonej wzorem (2.6), można również określić macierz kowariancji dla prognozowanych wartości nieruchomości ustalających model regresji. W tym celu trzeba wykonać następujące mnożenia macierzowe

$$\text{cov}(\hat{y}) = \hat{\sigma}_o^2 \cdot X \cdot (X^T X)^{-1} \cdot X^T \quad (2.7)$$

### 3. Regresja ważona geograficznie

Regresja ważona geograficznie należy do technik eksploracyjnych statystyki przestrzennej umożliwiających bezpośrednio modelowanie przestrzennej heterogeniczności, przestrzennej niestacjonarności modelowanego zjawiska. Istotą tej techniki jest konstrukcja lokalnych modeli regresji (dla każdej lokalizacji) opartych na

macierzy sąsiedztwa, kluczowego pojęcia statystyki oraz ekonometrii przestrzennej. Model GWR pozornie wygląda identycznie jak model klasyczny. Ważną różnicę stanowi fakt, że parametry modelu GWR są związane z lokalizacją, co przedstawia poniższe równanie

$$y_i = \beta_0(x_i, y_i) + \sum_{k=1}^p \beta_k(x_i, y_i) \cdot x_{ik} + \varepsilon_i \quad (3.1)$$

Parametry modelu (3.1) są estymowane przy użyciu ważonej metody najmniejszych kwadratów, zgodnie z poniższą formułą:

$$\sum_{i=1}^N w_{i(i)} \left[ y_i - \left( \beta_0(x_i, y_i) + \sum_{k=1}^p \beta_k(x_i, y_i) \cdot x_{ik} \right) \right]^2 = \varepsilon^T W_{(i)} \varepsilon = (y - X\beta)^T W_{(i)} (y - X\beta) = \min.$$

System wag dla każdej lokalizacji jest funkcją odległości między lokalizacją  $i$ -tą a każdą lokalizacją z bazy. Macierz wag jest macierzą diagonalną

$$W_{(i)} = \begin{bmatrix} w_{i1} & 0 & 0 \\ 0 & \ddots & \vdots \\ 0 & \cdots & w_{in} \end{bmatrix},$$

której elementy najczęściej wyrażane są jako:

$$w_{ij} = e^{\left( \frac{-d_{ij}^2}{h^2} \right)},$$

$$w_{ij} = \begin{cases} \left[ 1 - \left( \frac{d_{ij}}{h} \right)^2 \right]^2, & \text{dla } d_{ij} \leq h \\ 0, & \text{dla } d_{ij} > h \end{cases}$$

gdzie:

$w_{ij}$  – element „ $ij$ ” macierzy wag,

$d_{ij}$  – odległość między nieruchomością  $i$ -tą i  $j$ -tą,

$h$  – parametr opóźnienia (*distance decay parameter, bandwidth*).

Dla tak zdefiniowanego modelu dla każdej lokalizacji dostajemy zestaw współczynników regresji w postaci

$$\beta(x_i, y_i) = (X^T W_{(i)} X)^{-1} X^T W_{(i)} y \quad (3.2)$$

Dla każdej lokalizacji otrzymujemy również wartości przewidywane przez model dla nieruchomości ustalających model w postaci

$$\hat{y}_{(i)} = x_{(i)} \cdot \beta(x_i, y_i) \quad (3.3)$$

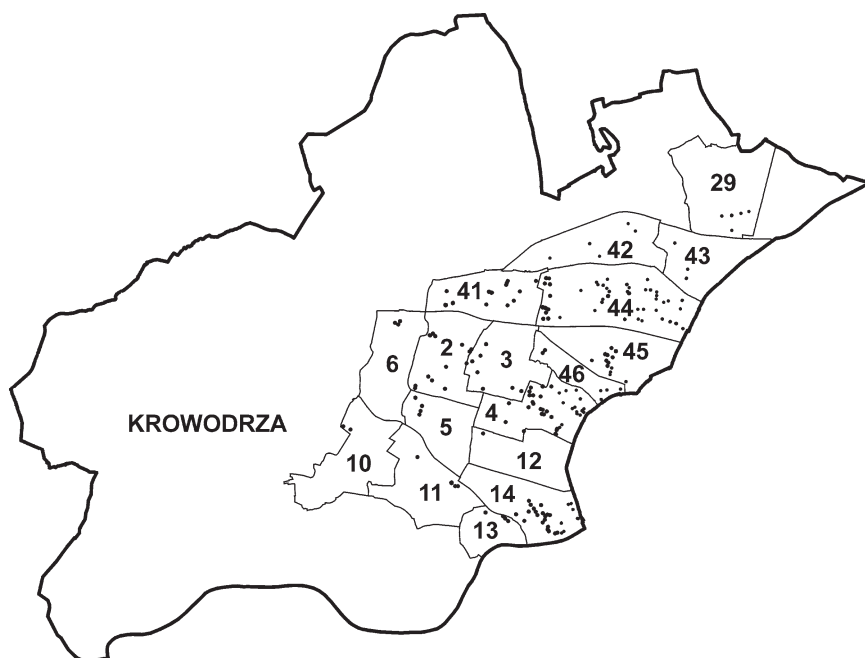
oraz zestaw reszt do modelu w postaci

$$\hat{\varepsilon}_{(i)} = y_i - \hat{y}_{(i)} \quad (3.4)$$

Na podstawie tak wyznaczonych parametrów oraz wielkości opisujących model można budować standardowe miary dopasowania modelu, także jak: współczynnik determinacji, skorygowany współczynnik determinacji oraz wariancję reszt.

#### 4. Opis bazy nieruchomości

Materiał badawczy wykorzystany przy pisaniu niniejszego artykułu dotyczy bazy nieruchomości lokalowych. Zgromadzone dane obejmują informacje o nieruchomościach znajdujących się w granicach byłej dzielnicy Krowodrza (obecnie dzielnice IV–VII) miasta Krakowa (rys. 1). Baza obejmuje 276 transakcji zawartych między październikiem 2004 a październikiem 2005 r.



Rys. 1. Obszar opracowania (była dzielnica Krowodrza, 17 obrębów ewidencyjnych oznaczonych cyframi arabskimi; punkty reprezentują położenie nieruchomości)

Każda ze sprzedanych nieruchomości została opisana siedmioma atrybutami, których nazwy oraz skale liczbowe i opisowe zaprezentowano poniżej:

- 1) *czas* – przesunięcie transakcji w czasie wyrażone liczbą miesięcy względem daty pierwszej transakcji w bazie;
- 2) *powierzchnia* – powierzchnia lokalu wyrażona w [m<sup>2</sup>];
- 3) *otoczenie* – wpływ otoczenia nieruchomości:
  - bardzo korzystne (2),
  - korzystne (1),
  - przeciętne (0),
  - niekorzystne (-1);
- 4) *standard* – standard wyposażenia lokalu w media, urządzenia sanitarne i użytkowe:
  - super komfort (2),
  - komfort (1),
  - przeciętny (0),
  - prymitywny (-1);
- 5) *komunikacja* – dostępność komunikacji:
  - bardzo dobra (2),
  - dobra (1),
  - przeciętna (0),
  - utrudniona (-1);
- 6) *zużycie* – zużycie łączne lokalu wyrażone procentowo;
- 7) *X, Y* – współrzędne *X, Y* centroidu działki.

W bazie informacji rynkowych dotyczących nieruchomości lokalowych, dla każdej transakcji obliczono również cenę jednostkową sprzedaży jako stosunek ceny transakcyjnej do pola powierzchni przedmiotu transakcji.

Opisane w ten sposób nieruchomości były podstawą do analiz przestrzennej heterogeniczności na rynku nieruchomości w dzielnicy Krowodrza.

## 5. Wyniki oraz analiza estymacji liniowego modelu globalnego oraz liniowych modeli lokalnych (regresja ważona geograficznie)

Przedstawiony przykład w formie tabelarycznej oraz graficznej ujmuje porównanie wyników oszacowania klasycznego modelu regresji liniowej wielorakiej oraz regresji ważonej geograficznie dla zbioru informacji rynkowych zawartych w opisie bazy w treści niniejszego artykułu. Tabela 1 zawiera parametry charakteryzujące model globalny obejmujący cały obszar opracowania. Poniżej wartości współ-

czynników w nawiasach podane zostały poziomy ich istotności. Współczynnik  $\beta_0$  to wyraz wolny natomiast pozostałe współczynniki  $\beta$  odpowiadają odpowiednio atrybutom przedstawionym w opisie bazy.

**Tabela 1.** Wyniki z estymacji globalnego modelu regresji liniowej wielorakiej dla zbioru informacji rynkowych „Krowodrza”

Wsp.	$\beta_0$	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$	$\beta_5$	$\beta_6$	$R^2$	$\bar{R}^2$	$\sigma_0$
	2732,49 (0,00)	68,90 (0,00)	-4,30 (0,04)	335,97 (0,00)	290,24 (0,00)	207,56 (0,00)	-2,18 (0,57)	0,26	0,24	700,14

Powyższy model przedstawia „słabe” dopasowanie do danych empirycznych ( $\bar{R}^2 = 0,24$ ). W przypadku wszystkich współczynników za wyjątkiem  $\beta_6$  uzyskano statystyczna istotność (poziom istotności 0,05).

W tabeli 2 zestawiono wyniki estymacji dla modelu GWR dla każdej nieruchomości z bazy, jednakże ze względu na ograniczoną objętość niniejszego opracowania zostało przedstawionych jedynie kilka rekordów.

**Tabela 2.** Wyniki z estymacji regresji ważonej geograficznie dla zbioru informacji rynkowych „Krowodrza”

Lp.	$\beta_0$	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$	$\beta_5$	$\beta_6$	$\sigma_0$
1	2891,92 (0,00)	33,25 (0,07)	-3,35 (0,19)	271,86 (0,00)	423,84 (0,00)	394,76 (0,00)	-7,66 (0,09)	156,24
2	3123,68 (0,00)	43,45 (0,02)	-8,52 (0,00)	416,28 (0,00)	432,81 (0,00)	451,37 (0,00)	-11,60 (0,02)	149,22
3	3122,76 (0,00)	43,52 (0,02)	-8,52 (0,00)	416,65 (0,00)	432,07 (0,00)	451,77 (0,00)	-11,62 (0,02)	149,21
4	2936,51 (0,00)	54,38 (0,00)	-4,59 (0,07)	400,55 (0,00)	291,08 (0,00)	449,74 (0,00)	-14,15 (0,00)	147,35
5	2934,85 (0,00)	57,34 (0,00)	-4,60 (0,07)	410,14 (0,00)	273,19 (0,00)	450,52 (0,00)	-14,76 (0,00)	144,77
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
257	3329,10 (0,00)	96,68 (0,00)	-9,63 (0,00)	490,10 (0,00)	120,38 (0,23)	203,49 (0,01)	-16,45 (0,00)	160,26

Całkowite dopasowanie modelu GWR może zostać scharakteryzowane następującymi wielkościami:  $R^2 = 0,62$ ,  $\bar{R}^2 = 0,61$  oraz  $\sigma_0 = 498,58$ .

Dodatkowo w tabeli 3 zestawiono zakres zmienności współczynników modelu GWR dla bazy nieruchomości „Krowodrza”, dane liczbowe w rozpatrywanej tabeli świadczą, że proces generujący dane nie może zostać uznany za stały w przestrzeni, to raczej wiele lokalnych rynków nieruchomości wchodzących w skład rynku „Krowodrza”.

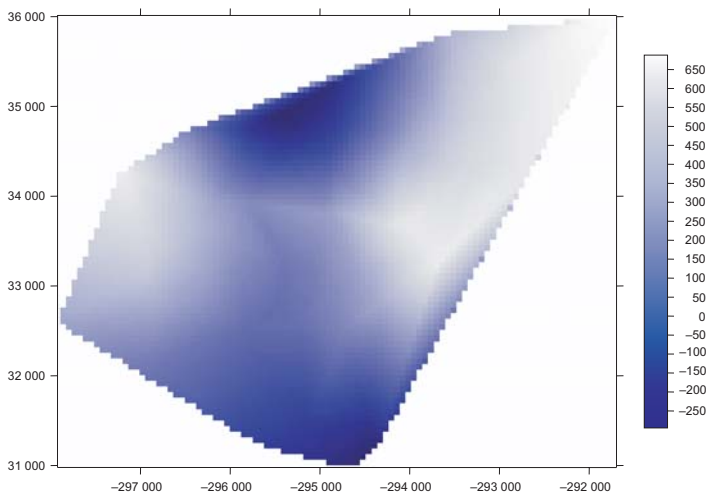
**Tabela 3.** Porównanie wartości współczynników modeli dla OLS oraz GWR

Wsp.	OLS*	GWR**		
		min.	max.	średnia
$\beta_0$	2732,49	2200,60	4845,30	3165,77
$\beta_1$	68,90	5,3276	121,29	71,99
$\beta_2$	-4,30	-16,42	2,2933	-6,31
$\beta_3$	335,97	-300,27	688,03	205,57
$\beta_4$	290,24	-249,66	733,22	279,22
$\beta_5$	207,56	-356,23	516,77	70,06
$\beta_6$	-2,18	-22,189	7,16	-4,61

\* *Ordinary Least Squares* (klasyczny model regresji liniowej szacowany metodą najmniejszych kwadratów)

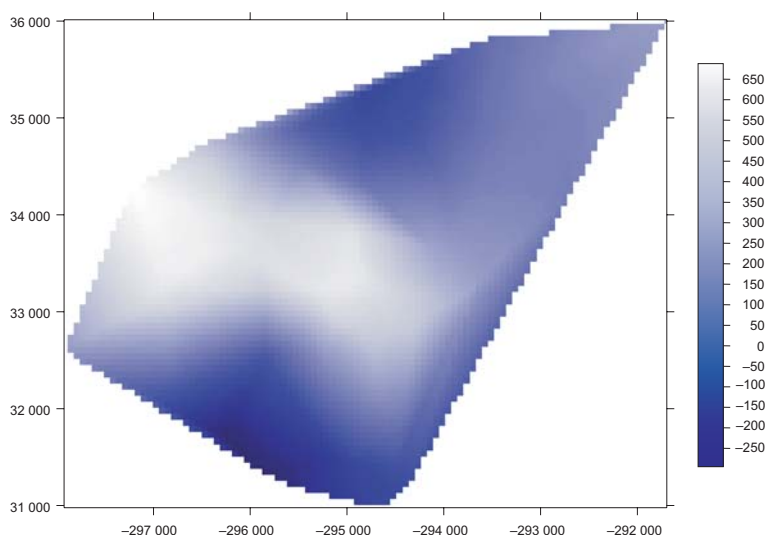
\*\* *Geographically Weighted Regression* (regresja ważona geograficznie)

Niestacjonarność wpływu poszczególnych charakterystyk na wartość nieruchomości, mierzonych wartościami współczynników regresji ( $\beta$ ) może zostać przedstawiona w postaci graficznej. Poniżej przedstawione zostały przykładowo zależności wartości parametru *beta* od położenia w przestrzeni geograficznej (rys. 2 i 3).



**Rys. 2.** Zależność współczynnika  $\beta_3$  od położenia w przestrzeni geograficznej





Rys. 3. Zależność współczynnika  $\beta_5$  od położenia w przestrzeni geograficznej

## 6. Wnioski płynące z zastosowania regresji ważonej geograficznie na gruncie wyceny nieruchomości

Według autorów regresja ważona geograficznie może stanowić dobre narzędzie przy powszechnej taksacji nieruchomości, prognozowaniu skutków różnego rodzaju działań związanych z zmianą zagospodarowania – między innymi zmian planu miejscowego, lokalizacji zakładów przemysłowych itp. – może również służyć jako narzędzie wspomagające podejmowanie decyzji o lokalizacji inwestycji. Z dotychczas uważanych za niewystarczające metod liniowych poprzez zaaplikowanie odpowiednich metod opartych o macierze sąsiedztwa, można wyciągać wnioski porównywalne, a nawet lepsze, niż te uzyskiwane ze skomplikowanych w stosowaniu modeli nieliniowych.

### Literatura

- [1] Anselin L.: *Advances in Spatial Econometrics*. Springer, Berlin 2004.
- [2] Anselin L.: *A Workbook for Using SpaceStat in the Analysis of Spatial Data*. Technical Report S – 92 – 1 of the National Center for Geographic Information and Analysis, University of California 1992.
- [3] Czaja J.: *Metody szacowania wartości rynkowej i katastralnej nieruchomości*. Wyd. KOMP-SYSTEM, Kraków 2001.

- 
- [4] Fotheringham A.S., Brunson C., Charlton M.: *Geographically Weighted Regression – the Analysis of Spatially Varying Relationships*. Wiley, Chichester 2002.
- [5] LeSage J.P.: *Spatial Econometrics. The Web Book of Regional Science*, Regional Research Institute, West Virginia University, Morgantown 1999.
- [6] Morrison D.F.: *Wielowymiarowa analiza statystyczna*. PWN, Warszawa 1990.
- [7] Zeliaś A. (red.): *Ekonometria przestrzenna*. PWE, Warszawa 1991.