METODY ILOŚCIOWE W BADANIACH EKONOMICZNYCH
Tom XII/2, 2011, str. 222–231

STATYSTYKA MORANA W ANALIZIE ROZKŁADU CEN NIERUCHOMOŚCI

Dorota Kozioł-Kaczorek
Katedra Ekonomiki Rolnictwa i Międzynarodowych Stosunków Gospodarczych
Szkoła Główna Gospodarstwa Wiejskiego w Warszawie
e-mail: dorota_koziol@sggw.pl

Streszczenie: W pracy przedstawiono sposób wykorzystania statystyki Morana do analizy rozkładu cen nieruchomości. Podstawowym zagadnieniem, rozważanym w niniejszej publikacji, jest rozkład cen w zależności od lokalizacji i techniki jego identyfikacji. Prezentowane podejście zostało zilustrowane na przykładzie analizy cen z wtórnego rynku nieruchomości z dzielnicy Praga Południe w Warszawie.

Słowa kluczowe: autokorelacja przestrzenna, statystyka Morana I, lokalna statystyka Morana I_i, ceny nieruchomości

WPROWADZENIE

Zasadniczy wpływ na kształtowanie się wysokości cen nieruchomości oraz ich zróżnicowanie mają tak zwane cechy rynkowe nieruchomości. Do zbioru tych cech zalicza się między innymi cechy dotyczące lokalizacji, cechy techniczne, fizyczne oraz użytkowe[1]. Z punktu widzenia wyceny nieruchomości najważniejszą z nich jest lokalizacja i to właśnie ona w najwyższym stopniu kształtuje cenę nieruchomości. Istotnym wobec tego zagadnieniem jest badanie cen w kontekście położenia nieruchomości, a co za tym idzie wykorzystanie analizy przestrzennej[2] [Pietrzykowski 2010].

[1] Autorka rozważa wyłącznie rynek lokalny i bezpośrednie determinanty ceny rynkowej 1 m^2 nieruchomości dlatego też w analizie nie uwzględnia wpływu czynników makroekonomicznych takich jak np. kryzys.

[2] Metody ekonometrii przestrzennej wykorzystywane w badaniach dotyczących rynku nieruchomości opisali między innymi Can [1998], Anselin [1998], LaSage i Pace [2004]. W Polsce publikacje na ten temat można znaleźć między innymi w Studiach i Materiałach Towarzystwa Naukowego Nieruchomości.

Podstawowym problemem, rozważanym w niniejszej publikacji, jest rozkład cen w zależności od lokalizacji nieruchomości. Zależność ta została zbadana w zakresie globalnym, a więc określono zależność ceny w danej lokalizacji od cen w innych lokalizacjach badanego obszaru. Zbadano również skorelowanie lokalne cen, czyli zależności przestrzenne ceny w danej lokalizacji z otoczeniem, czyli cenami w lokalizacjach sąsiednich.

Rozważane zagadnienie zostało przedstawione na przykładzie analizy cen 1 m^2 nieruchomości na wtórnym rynku nieruchomości lokalowych (lokale mieszkalne) dzielnicy Praga Pd. w Warszawie.

Celem pracy jest prezentacja sposobu wykorzystania statystyki Morana do analizy przestrzennego rozkładu cen na danym rynku nieruchomości. W trakcie badań sformułowano następujące hipotezy badawcze:

1. Występuje dodatnia autokorelacja przestrzenna cen nieruchomości (nieruchomości z „podobnymi" cenami sąsiadują ze sobą).
2. Wraz ze wzrostem odległości lokalizacji od centrum Warszawy ceny 1 m^2 nieruchomości maleją.

W procesie weryfikacji sformułowanych hipotez badawczych wykorzystano współczynnik korelacji przestrzennej Morana (globalny i lokalny) oraz korelogram.

Niezbędne do osiągnięcia postawionego celu analizy i obliczenia wykonano w pakiecie statystycznym R 2.11.1 [R Development Core Team 2005] oraz w programie MS Excel 2007.

AUTKORELACJA PRZESTRZENNA

Globalna autokorelacja przestrzenna – statystyka Morana *I*

Stopień skorelowania wartości badanej zmiennej w danej lokalizacji z wartościami tej samej zmiennej w innych lokalizacjach jest określany poprzez globalną autokorelację przestrzenną. Występowanie takiej zależności oznacza przestrzenne grupowanie się wartości. Jeżeli jest to autokorelacja dodatnia, to tworzą się klastry (grupy) podobnych wartości (wysokich lub niskich) obserwowanej zmiennej. Autokorelacja ujemna to odwrotność autokorelacji dodatniej, czyli wysokie wartości obserwowanych zmiennych sąsiadują z niskimi wartościami tych zmiennych, a niskie sąsiadują z wysokimi [Suchecki 2010].

Istnieje kilka metod testowania grupowania się danych w skali globalnej. Ich konstrukcja opiera się na ogólnej koncepcji współczynnika korelacji liniowej Pearsona oraz na statystyce gamma. Statystyka ta jest połączeniem dwóch rodzajów informacji dotyczących podobieństwa obserwacji w przestrzeni. Jeden rodzaj to informacje dotyczące odległości między analizowanymi obiektami, czyli macierz odległości pomiędzy obserwacjami zapisana w postaci macierzy wag **W**. Drugi rodzaj informacji dotyczy zależności pomiędzy wartościami badanej zmiennej w poszczególnych lokalizacjach. Jest on przedstawiany w postaci macierzy korelacji **A**. W ogólnej postaci statystykę gamma zapisujemy jako

$$\Gamma = \mathbf{WA} = \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij} a_{ij}. \tag{1}$$

Jest ona funkcją iloczynów krzyżowych miary odległości w przestrzeni (w_{ij}) i miary współzależności (podobieństwa) wartości badanej zmiennej (a_{ij}). W przypadku statystyki Morana elementy macierzy **A** = $\{a_{ij}\}$ przyjmują postać iloczynów krzyżowych

$$a_{ij} = z_i^s z_j^s, \tag{2}$$

gdzie z_i^s oraz z_j^s oznaczają standaryzowane wartości badanej zmiennej. Statystyka Morana ma wobec tego postać

$$I = \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}^{*} z_i^s z_j^s, \tag{3}$$

gdzie $w_{ij}^{*} = \frac{w_{ij}}{\sum_j w_{ij}}$ jest elementem standaryzowanej macierzy wag, $w_{ij} = \frac{1}{d_{ij}}$ jest elementem macierzy wag, d_{ij} jest elementem macierzy odległości oznaczającym odległość pomiędzy dwiema lokalizacjami *i* oraz *j* [Suchecki 2010].

Istotność statystyczną statystyki Morana I weryfikuje się poprzez następujący zbiór hipotez:

H_0: obserwowane wartości badanej zmiennej są rozmieszczone w sposób losowy pomiędzy poszczególnymi lokalizacjami – brak autokorelacji przestrzennej;

H_1: obserwowane wartości badanej zmiennej nie są rozmieszczone w sposób losowy pomiędzy poszczególnymi lokalizacjami – występuje autokorelacja przestrzenna.

Weryfikację powyższych hipotez przeprowadza się w oparciu o unormowaną statystykę

$$Z_i = \frac{I - E(I)}{\sqrt{\mathrm{var}(I)}} \sim N(0,1), \tag{4}$$

gdzie $E(I) = -\frac{1}{n-1}$ oznacza wartość oczekiwaną statyki *I*, $\mathrm{var}(I) = \frac{n^2 S_1 - n S_2 + 3 S_0^2}{(n^2-1) S_0^2} - \frac{1}{(n-1)^2}$ jest wariancją statystyki *I* oraz

$S_0 = \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}$, $S_1 = 2\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}$, $S_2 = 4\left(\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\right)^2$ pochodzą z symetrycznej macierzy wag. [Suchecki 2010].

Przyjmuje się, że w przypadku braku autokorelacji przestrzennej statystyka Morana I ma tendencje do przyjmowania wartości $\mathrm{I} \approx \mathrm{E(I)}, \mathrm{Z(I)} \approx 0$. W przeciwnym razie występuje:

- autokorelacja dodatnia, gdy $\mathrm{I} > \mathrm{E(I)}, \mathrm{Z(I)} > 0,$
- autokorelacja ujemna, gdy $\mathrm{I} < \mathrm{E(I)}, \mathrm{Z(I)} < 0.$

W celu przeprowadzenia bardziej szczegółowej analizy należy sporządzić wykres rozproszenia[3], który pozwala na dokładniejszą identyfikację typu autokorelacji w przestrzeni. Wykres ten przedstawia relację wartości badanej zmiennej w danej lokalizacji z wartościami tej zmiennej w innych lokalizacjach. Na osi odciętych odkładane są wartości zmiennej standaryzowanej, natomiast na osi rzędnych jej wartości „przestrzennie” opóźnione poprzez zastosowanie wybranej macierzy wag. Skupienia punktów w I i III ćwiartce oznaczają występowanie dodatniej autokorelacji przestrzennej, czyli wysokie wartości sąsiadują z wysokimi wartościami (I ćwiartka) i niskie wartości sąsiadują z niskimi wartościami (III ćwiartka). Punkty występujące w II i IV ćwiartce traktowane są jako obserwacje nietypowe [Suchecki 2010].

Lokalna autokorelacja przestrzenna – statystyki Morana I_i

Stopień skorelowania wartości badanej zmiennej w danej lokalizacji z jej sąsiadami. Statystyki lokalne autokorelacji przestrzennej wskazują na statystycznie istotne skupienia podobnych wartości w sąsiadujących lokalizacjach. Umożliwiają one ocenę założeń stacjonarności, identyfikację obszarów niestacjonarności, identyfikację obserwacji odstających, skupień dużych i małych wartości oraz jednorodnych podobszarów. Statystyki lokalne pozwalają ponadto na identyfikację maksymalnego dystansu dostrzegalnych współzależności w przestrzeni.

Statystyka autokorelacji lokalnej również może być oparta na funkcji gamma iloczynów krzyżowych. W tym przypadku jednakże funkcja ta jest zredukowana do sumy ważonej miar podobieństwa wartości badanej zmiennej

$$\Gamma_i = \sum_j w_{ij} a_{ij} \tag{5}$$

W lokalnej statystyce Morana I_i elementy macierzy $\mathbf{A} = \{a_{ij}\}$ przyjmują postać

$$a_{ij} = (x_i - \bar{x})(x_j - \bar{x}), \tag{6}$$

gdzie x_i oraz x_j oznaczają wartości badanej zmiennej, a $\bar{x}$ oznacza średnią arytmetyczną badanej zmiennej. Lokalna statystyka Morana ma postać

[3] inaczej: moranowski wykres rozproszenia Anselina, wykres Morana

$$I_i = z_i^s \sum_{j=1}^{n} w_{ij}^* z_j^s , \tag{7}$$

gdzie z_i^s i z_j^s oznaczają standaryzowane wartości badanej zmiennej, $w_{ij}^* = \frac{w_{ij}}{\sum_j w_{ij}}$ jest elementem standaryzowanej macierzy wag, $w_{ij} = \frac{1}{d_{ij}}$ jest elementem macierzy wag, d_{ij} jest elementem macierzy odległości oznaczającym odległość pomiędzy dwiema lokalizacjami *i* oraz *j* [Suchecki 2010].

Istotność statystyczną lokalnej statystyki Morana I_i weryfikuje się poprzez następujący zbiór hipotez:
H_0: brak skupienia małych lub dużych wartości badanej zmiennej w sąsiedztwie *i* -tej lokalizacji – brak lokalnej autokorelacji przestrzennej;
H_1: występują skupienia małych lub dużych wartości badanej zmiennej w sąsiedztwie *i* -tej lokalizacji - występuje lokalna autokorelacja przestrzenna.

Weryfikację powyższych hipotez przeprowadza się podobnie jak w przypadku autokorelacji globalnej w oparciu o unormowaną statystykę

$$Z_{Ii} = \frac{I_i - E(I_i)}{\sqrt{\operatorname{var}(I_i)}} \sim N(0,1), \tag{8}$$

gdzie $E(I_i) = -\frac{\sum_j w_{ij}}{n-1}$ oznacza wartość oczekiwaną statyki I_i. Wariancją statystyki I_i jest $\operatorname{var}(I_i) = \frac{(n-b_2)\sum_{j \neq i} w_{ij}^2}{n-1} + \frac{2(2b_2 - n)\sum_{k \neq i}\sum_{h \neq j} w_{ik} w_{ih}}{(n-1)(n-2)} - \frac{\left(-\sum_j w_{ij}\right)^2}{(n-1)^2}$,

gdzie $b_2 = \frac{\frac{1}{n}\sum_i z_i^4}{\left(\frac{1}{n}\sum_{i=1}^{n} z_i^2\right)^2}$ [Suchecki 2010].

Przyjmuje się, że w przypadku braku autokorelacji przestrzennej statystyka Morana I_i ma tendencje do przyjmowania wartości $\mathrm{I_i} \approx \mathrm{E(I_i)}, \mathrm{Z(I_i)} \approx 0$. W przeciwnym razie występuje:

- autokorelacja dodatnia, gdy $\mathrm{I_i} > \mathrm{E(I_i)}, \mathrm{Z(I_i)} > 0$,
- autokorelacja ujemna, gdy $\mathrm{I_i} < \mathrm{E(I_i)}, \mathrm{Z(I_i)} < 0$.

Występowanie dodatniej autokorelacji oznacza, iż dana lokalizacja jest otoczona podobnymi pod względem badanej cechy lokalizacjami. Jeżeli natomiast występuje ujemna (negatywna) autokorelacja to dana lokalizacja otoczona jest lokalizacjami znacząco różniącymi się pod względem wartości badanej zmiennej. Taki region uważa się za lokalizację nietypową [Kopczewska 2007].

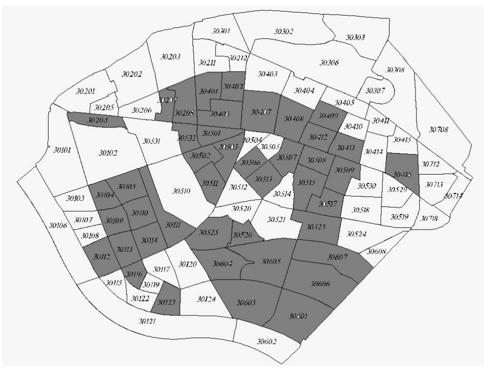
ANALIZA PRZESTRZENNEGO ROZKŁADU CEN

Materiał badawczy

Zaprezentowane w pracy badania przeprowadzono w oparciu o dane dotyczące transakcji nieruchomościami lokalowymi (lokale mieszkalne), które miały miejsce na rynku wtórnym na terenie dzielnicy Praga Południe w Warszawie. Okres monitorowania rynku obejmuje rok 2009 (od 02.01.2009 do 30.12.2009). Baza danych obejmowała 485 danych, które zawierały informacje dotyczące obrębu, w którym znajduje się nieruchomość, adresu (z dokładnością do ulicy i numeru budynku), powierzchni, kondygnacji, liczby pomieszczeń oraz ceny 1 m^2.

Celem niniejszej publikacji była charakterystyka przestrzennego rozkładu cen nieruchomości w dzielnicy Praga Południe w Warszawie, wobec tego ograniczono się wyłącznie do informacji dotyczących cen i obrębów, w których znajdowały się nieruchomości będące przedmiotem transakcji. Ponieważ nieruchomości położone w danym obrębie ewidencyjnym wykazywały takie same wartości cech dotyczących lokalizacji (położenia, otoczenia, sąsiedztwa, komunikacji z centrum miasta itp.) oraz podobny poziom cen to w analizie ograniczono się do poziomu obrębów. Przestrzenne rozmieszczenie nieruchomości będących przedmiotem transakcji z dokładnością do obrębu przedstawia rysunek 1. Zaznaczono na nim zarys dzielnicy Praga Południe oraz zarysy obrębów ewidencyjnych w niej występujących. Szary kolor oznacza obręby, w których odnotowano transakcje na rynku wtórnym w badanym 2009 roku.

Rysunek 1. Obręby Pragi Południe, w których zaobserwowano transakcje.



Źródło: obliczenia własne

W badanym okresie zaobserwowano transakcje w 43 obrębach. W pozostałych obrębach obserwowano transakcje wyłącznie na rynku pierwotnym, który nie był przedmiotem badania lub nie odnotowano żadnych transakcji.

Obręby, w których zaobserwowano transakcje zostały scharakteryzowane za pomocą średniej ceny 1 m^2 nieruchomości w danym obrębie oraz za pomocą współrzędnych środka obrębu.

Wyniki

Do badania występowania globalnej autokorelacji przestrzennej niezbędne było ustalenie macierzy odległości oraz macierzy wag. W przypadku omawianych badań skonstruowana została macierz odległości pomiędzy środkami obrębów ewidencyjnych (w metrach). Na jej podstawie ustalono macierz wag (odwrotność odległości) oraz standaryzowaną macierz wag. Następnie obliczono wartość statystyki Morana *I* oraz jej wartość oczekiwaną i wariancję. Wyniki przeprowadzonych obliczeń przedstawiono w tabeli 1.

Tabela 1. Statystyka Morana I

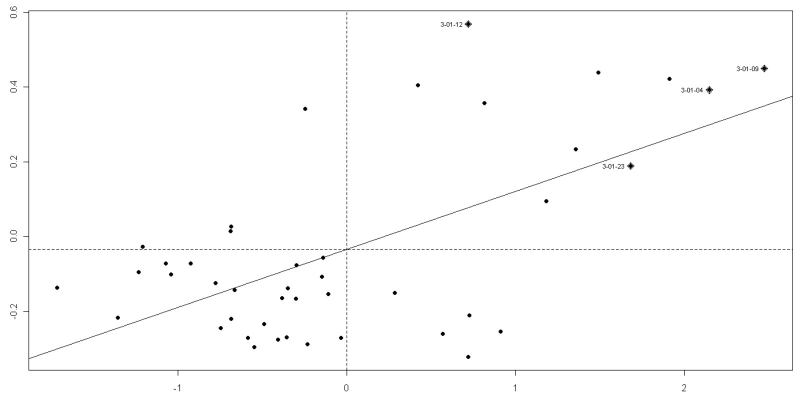
I	E(I)	var (I)	Z(I)	p-value
0.15506	-0.02381	0.00055	7.61258	0.00000

Źródło: opracowanie własne

Na podstawie uzyskanych wyników, hipotezę o braku globalnej autokorelacji przestrzennej, na poziomie istotności $\alpha = 0.05$, odrzucono. Można wobec tego uznać, że istnieje globalna autokorelacja przestrzenna. Oznacza to, że ceny 1 m^2 nieruchomości determinują i jednocześnie są determinowane przez ceny 1 m^2 w innych obrębach. Ponieważ $I > E(I)$ oraz $Z(I) > 0$, to można wyciągnąć wniosek, iż w przypadku badanych cen występuje autokorelacja dodatnia, czyli ceny wysokie sąsiadują z wysokimi, a ceny niskie z niskimi.

W celu czytelnego zobrazowania uzyskanych wyników zostały one przedstawione na wykresie Morana, który pokazuje rozproszenie cen nieruchomości ze względu na lokalizację.

Rysunek 2. Wykresy rozrzutu Morana dla średnich cen nieruchomości mieszkaniowych w warszawskiej dzielnicy Praga Południe

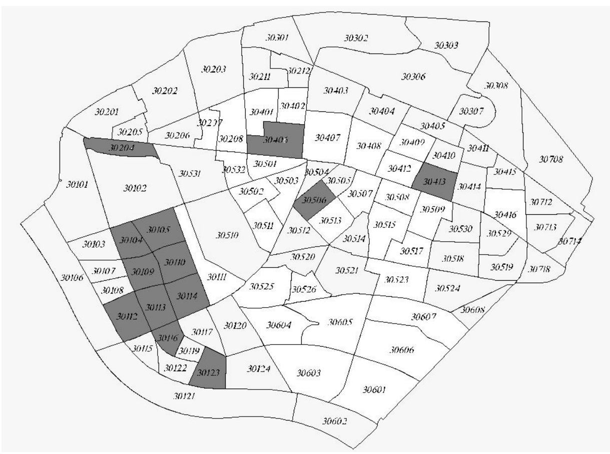


Źródło: obliczenia własne

Na przedstawionym wykresie wyraźnie widać występującą dodatnią autokorelację przestrzenną – obserwacje zgromadzone są głównie w pierwszej (ceny wysokie znajdują się obok siebie) i trzeciej ćwiartce (ceny niskie są zlokalizowane obok siebie). Nieliczne obserwacje występujące w II i IV ćwiartce to obserwacje nietypowe.

W celu identyfikacji skupień dużych i małych wartości badanej zmiennej oraz wyodrębnienia lokalizacji nietypowych zbadano występowanie istotnej autokorelacji lokalnej. W tym celu oszacowano wartość lokalnej statystyki Morana I_i. Wyniki przeprowadzonych obliczeń przedstawiono na Rysunku 3.

Rysunek 3. Rozmieszczenie obrębów ze względu na wartości lokalnych statystyk Morana I_i



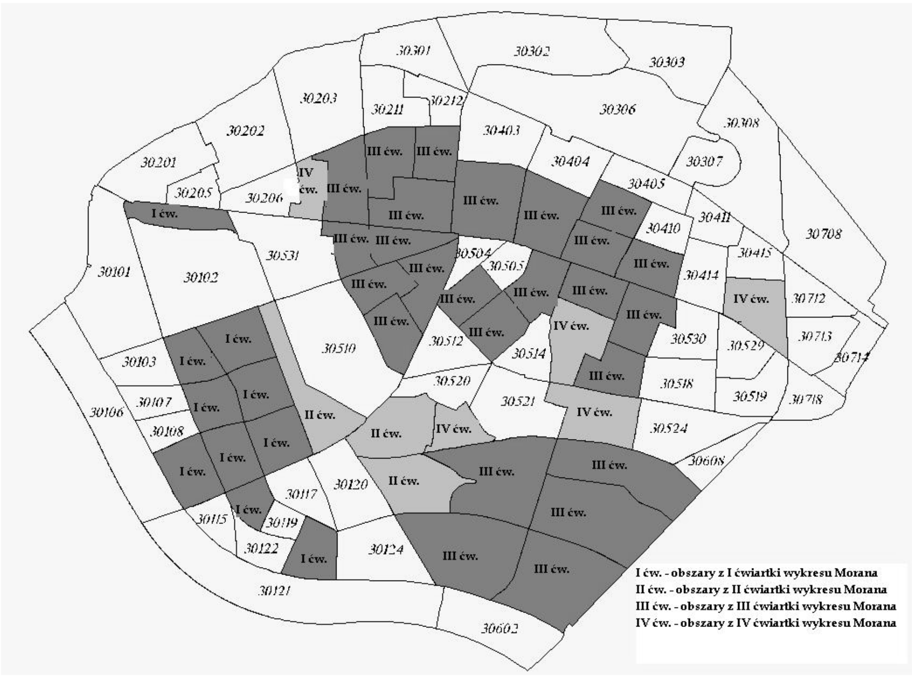
Źródło: obliczenia własne

Obręby zaznaczone szarym kolorem to obręby z lokalną autokorelacja dodatnią. Oznacza to, że są one otoczone okręgami o podobnych cenach.

Wnioski

Przeprowadzone analizy wykazały prawdziwość sformułowanych na wstępie hipotez badawczych. Na podstawie uzyskanych wyników wyraźnie widać zależność wartości cen 1 m^2 nieruchomości od lokalizacji. Na obserwowanym rynku nieruchomości lokalowych można zaobserwować wyraźne klastry obrębów z cenami wyższymi i klastry obrębów z cenami niższymi. Ponadto, wraz ze wzrostem odległości lokalizacji od centrum Warszawy ceny 1 m^2 nieruchomości maleją.

Rysunek 4. Rozkład cen nieruchomości lokalowych na Pradze Południe wg obrębów



Źródło: obliczenia własne

Obręby z pierwszej ćwiartki wykresu Morana to obręby z wysokimi cenami 1 m^2 nieruchomości. Są to obręby leżące w modnej części Pragi Południe (i całej Warszawy) czyli na Saskiej Kępie. Bez wątpienia jest to atrakcyjna okolica ze względu na łatwy dostęp do komunikacji miejskiej, centrum Warszawy, centrów handlowych, placówek oświatowych itp.

Obręby z trzeciej ćwiartki wykresu Morana to obręby z niskimi cenami 1 m^2 nieruchomości. Obejmują one środkową cześć dzielnicy Praga Południe (Grochów

(Grochów). Są to przeważnie „blokowiska” z lat 60-tych i 70-tych ubiegłego wieku. Otoczenie tych obrębów stanowią obszary przemysłowe, węzły kolejowe.

Obręby z drugiej i czwartej ćwiartki wykresu Morana to obręby uznane za nietypowe. Nie stanowią one zbyt licznej reprezentacji.

BIBLIOGRAFIA

Kopczewska K. (2007) Ekonometria i statystyka przestrzenna.CEDEWU.PL. Warszawa.

Kulczyki M., Ligas M. (2007) Zastosowanie analizy przestrzennej do modelowania danych pochodzących z rynku nieruchomości. Studia i Materiały Towarzystwa Naukowego Nieruchomości vol. 15, nr 3-4, str. 145 – 153.

Pietrzykowski R. (2010) Przestrzenne ujęcie rynku nieruchomości mieszkaniowych w latach 2007 – 2010, Zeszyty Naukowe Uniwersytetu Szczecińskiego. Finanse. Rynki Finansowe. Ubezpieczenia, 616 (29), 97-107

R Development Core Team 2005 R: A language and environment for statistical computing. R Foundation for Statistical Computing, Viena, Austria, ISBN3-900051-07-0, URL: http://www.R-project.org

Suchecki B. (2010) Ekonometria przestrzenna. Metody i modele analizy danych przestrzennych. C.H. Beck. Warszawa.

A MORAN’S STATISTIC IN ANALYSIS OF DISTRIBUTION OF PRICES ON REAL ESTATE MARKET

Abstract: In the paper a problem of the distribution of prices of the real estate market is considered. The paper contain metod for testing this distribution. This solution is Moran I. The presented method is illustrated by the example of data from the secondary property market in the Praga Pd. district of Warsaw.

Key words: spatial autocorrelation, Moran’s statistics, prices on real estate market