PRZEGLĄD STATYSTYCZNY
R. LX – ZESZYT 3 – 2013

ALICJA OLEJNIK

WYBRANE METODY TESTOWANIA MODELI REGRESJI PRZESTRZENNEJ

1.WPROWADZENIE

Regresja liniowa jest jedną z najpopularniejszych technik eksploracji danych. Zakłada ona jednakże, iż obserwacje w próbie są niezależne, co niekoniecznie jest spełnione w przypadku danych przestrzennych, często charakteryzujących się autokorelacją przestrzenną. Współcześnie w literaturze światowej wiele miejsca poświęca się problemom analizy i modelowania zjawisk o charakterze przestrzennym, marginalizując jednak często aspekt weryfikacji poprawności założonej struktury autokorelacji przestrzennej. Powszechne podejście do modelowania zjawisk przestrzennych opiera się na technicznej eliminacji problemu autokorelacji przestrzennej z modelu ekonometrycznego. Nietrudno jednak zauważyć, że potraktowanie autokorelacji przestrzennej, jako dodatkowego źródła informacji, pomaga lepiej modelować rzeczywiste zjawiska o charakterze przestrzennym. Takie podejście wymaga jednak skutecznych narzędzi testujących i weryfikujących poprawność wyboru postaci modelu przestrzennego.

Celem niniejszego artykułu jest zatem przybliżenie wybranych metod pozwalających na wybór modeli najlepiej opisujących zjawiska o charakterze przestrzennym. W niniejszej pracy opisane zostaną podstawowe testy dla modeli regresji przestrzennej ze wskazaniem ich pozytywnych i negatywnych stron oraz zalecenia dotyczące ich stosowalności. Zaprezentowane metody pozwalają na testowanie występowania autokorelacji przestrzennej w modelu oraz wybór najlepszej specyfikacji modelu przestrzennego. W artykule wskazano również propozycję odpowiedniej procedury testującej przestrzenną niestacjonarność, mającej istotny wpływ na wybór ostatecznej postaci modelu.

2. NIESTANDARDOWY TEST MORANA *I* W MODELU REGRESJI LINIOWEJ Z AUTOKORELACJĄ PRZESTRZENNĄ SKŁADNIKA LOSOWEGO

Przypomnijmy, że autokorelacja przestrzenna danych oznacza stopień skorelowania obserwowanej wartości zmiennej z wartością tej samej zmiennej w innej lokalizacji w przestrzeni (np. regionie). Wyróżniamy dwa rodzaje autokorelacji przestrzennej: dodatnią i ujemną. Dla autokorelacji dodatniej obiekty, które są położone blisko siebie mają tendencję do przyjmowania podobnych wartości, dla ujemnej autokorelacji przestrzennej, obserwujemy, sąsiedztwo wysokich wartości z niskimi (i odwrotnie). Jeśli autokorelacja przestrzenna występuje w modelu ekonometrycznym, należy uwzględnić ją w postaci modelu, aby zminimalizować jej negatywny wpływ na jakość oszacowań parametrów modelu[1]. Zauważmy również, że przy dużej autokorelacji przestrzennej efektywna wielkość próby jest mniejsza niż obserwowana.

Rozważmy model z przestrzennie autoskorelowanym składnikiem losowym, który uwzględnia przestrzenną zależność w rozkładzie składnika losowego (por. Anselin, 1988). Wówczas poszczególne błędy u_i dla różnych lokalizacji są skorelowane z błędami w innych lokalizacjach:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}, \quad \mathbf{u} = \lambda\mathbf{W}\mathbf{u} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim \mathrm{N}(\mathbf{0}, \sigma^2\mathbf{I}), \tag{1}$$

gdzie **y** jest wektorem zmiennych losowych, z których każda reprezentuje odpowiedni region ekonomiczny (województwo, państwo, itp.), **W** oznacza przestrzenną macierz wag, o wymiarach $N\times N$, a przez macierz **X**, o wymiarach $N\times K$ rozumieć będziemy macierz zmiennych objaśniających, z uwzględnieniem wyrazu wolnego, wnoszącą do modelu dodatkową, zewnętrzną informację o badanym procesie. Zmienna **u** to N-wymiarowy wektor skorelowanych przestrzennie składników losowych, zaś **ε** to proces gaussowskiego białego szumu z zerową wartością oczekiwaną i macierzą wariancji-kowariancji $\sigma^2\mathbf{I}$. Parametr λ jest współczynnikiem przestrzennej korelacji reszt, dla którego przyjmujemy założenie: $|\lambda|<1$.

Test Morana opracowany został przez Patricka Morana w roku 1950 i jest najpowszechniej stosowanym testem autokorelacji przestrzennej. Poniżej przedstawimy test Morana jednak w nieco odmiennym od standardowego ujęciu. W odróżnieniu od podejścia klasycznego nie będziemy zakładać *a priori* normalności rozkładu składnika losowego, a raczej rozważymy asymptotyczny rozkład statystyk testowych. Kelejian i Prucha (2001) argumentują, że możemy przyjąć ogólne założenia, przy których statystyka Morana I jest asymptotycznie równoważna formie kwadratowej składnika losowego oraz podają ogólną postać Centralnego Twierdzenia Granicznego dla takiej formy kwadratowej. Te własności statystyki I pozwalają na uzależnienie elementów macierzy formy kwadratowej od wielkości próby, tak jak to zazwyczaj ma miejsce w przypadku modeli typu Cliffa-Orda (1981). Takie podejście, dopuszcza heteroskedastyczność składnika losowego. Należy podkreślić, iż wcześniejsze rozważania dotyczące asymptotycznego rozkładu statystyki I w teście Morana, lub równoważnych testów opartych na mnożniku Lagrange'a, nie uwzględniały przypadku, w którym elementy przestrzennej macierzy wag zależą od wielkości próby. Podobnie, nie dopuszczano heteroskedastyczności składnika losowego, przyjmując *a priori* jego rozkład normalny $\mathrm{N}(\mathbf{0}, \sigma^2\mathbf{I})$.

---

[1] Przypomnijmy, że w przypadku występowania zjawiska autokorelacji przestrzennej estymator metody najmniejszych kwadratów może nie być zgodny, ani nawet nieobciążony (por. Anselin, 1988).

Poniżej zdefiniujemy statystykę testową I dla modelu regresji liniowej bez opóźnień przestrzennych w zmiennej zależnej. Następnie pokażemy, że test Morana oraz test mnożników Lagrange'a (*LM*) przestrzennej autokorelacji składnika losowego są tożsame (por. Burridge, 1980). Okaże się, iż statystyka testu Morana jest zdefiniowana w postaci formy kwadratowej składnika losowego (lub reszt) znormalizowanych przez szacowany błąd standardowy formy kwadratowej. Cliff i Ord (1973, 1981) w swojej definicji statystyki Morana I zastosowali czynnik normalizujący będący niezgodnym estymatorem odchylenia standardowego formy kwadratowej. Wymagał on dalszej normalizacji w celu uzyskania statystyki asymptotycznie standaryzowanej. W kolejnych publikacjach błąd ten był powtarzany przez wielu autorów.

2.1. DEFINICJA STATYSTYKI *MORANA I*

Niech **e** oznacza wektor reszt modelu z estymacji MNK. Wówczas, przy hipotezie zerowej $H_0: \lambda = 0$ o braku autokorelacji przestrzennej reszt modelu (z hipotezą alternatywną: $H_1: \lambda \neq 0$) statystyka Morana I przyjmuje postać:

$$I = \frac{\mathbf{e}^T \mathbf{W} \mathbf{e}}{N^{-1} \mathbf{e}^T \mathbf{e} [tr((\mathbf{W}^T + \mathbf{W})\mathbf{W})]^{1/2}}, \tag{2}$$

gdzie $tr(.)$ oznacza ślad macierzy.

Wyniki prac opublikowane przez Kelejiana i Pruchę (2001) dowodzą, iż przy założeniu prawdziwości hipotezy H_0 statystyka I dąży względem dystrybuanty do rozkładu normalnego, tj.:

$$I \xrightarrow{D} \mathrm{N}(0,1).$$

Na uwagę zasługuje fakt, iż w większości pozycji literaturowych, we wzorze (2) dzielnik, będący pierwiastkiem śladu macierzy zastępowany jest sumą elementów macierzy **W**. Ponieważ w ogólności:

$$[tr((\mathbf{W}^T + \mathbf{W})\mathbf{W})]^{1/2} \neq \Sigma_i \Sigma_j w_{ij}, \tag{3}$$

to powszechnie stosowane podejście jest niepoprawne. Należy nadmienić, że w swoim artykule z 1950 roku Moran rozważał przestrzenne macierze o wagach przyjmujących tylko wartości równe zero lub jeden. Wówczas zachodzi: $w_{ij} = w^2_{ij}$, co we wzorze (3) daje równość[2].

---

[2] Zauważmy, że własności testu takie jak np. moc, czy nieobciążoność zależą od stopnia w jakim macierz wag przestrzennych wykorzystana do wyliczenia statystyki Morana właściwie odwzorowuje rzeczywiste powiązania pomiędzy obiektami przestrzennymi. Problem ten dotyczy wszystkich testów autokorelacji przestrzennej składnika losowego i jest zauważany oraz aktualnie dyskutowany przez specjalistów (por. Drukker, Prucha, 2013).

2.2. TEST OPARTY NA MNOŻNIKACH LAGRANGE'A

Przy przyjętych powyżej oznaczeniach oraz założeniach łatwo zauważyć, że statystyka *LM* przyjmuje postać:

$$LM = \left[\frac{\mathbf{e}^T \mathbf{W} \mathbf{e}}{N^{-1} \mathbf{e}^T \mathbf{e} [tr((\mathbf{W}^T + \mathbf{W})\mathbf{W})]^{1/2}} \right]^2 \sim \chi^2(1). \tag{4}$$

Zatem statystyka testowa *LM* jest kwadratem statystyki Morana *I*. Choć fakt ten został zaobserwowany już w 1980 roku przez Burridge'a (1980), do dziś pozostaje w cieniu szeroko rozpowszechnionego poglądu o rozłączności tychże dwóch testów.

2.3. STANDARYZACJA STATYSTYKI MORANA *I* DLA MAŁYCH PRÓB

Oznaczmy reszty modelu (1) przez: $\mathbf{e} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}$, gdzie $\hat{\boldsymbol{\beta}}$ jest estymatorem postaci: $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$. Wprowadźmy oznaczenie: $\mathbf{M}_\mathbf{X} = \mathbf{I} - \mathbf{P}_\mathbf{X}$, gdzie $\mathbf{P}_\mathbf{X} = \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$. Wówczas:

$$\begin{aligned}
\mathbf{e} &= \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} \\
&= \mathbf{y} - \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \\
&= (\mathbf{X}\boldsymbol{\beta} + \mathbf{u}) - \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T (\mathbf{X}\boldsymbol{\beta} + \mathbf{u}) \\
&= [\mathbf{X}\boldsymbol{\beta} + \mathbf{u} - \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{X}\boldsymbol{\beta} - \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{u}] \\
&= \mathbf{u} - \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{u} \\
&= \mathbf{M}_\mathbf{X} \mathbf{u}.
\end{aligned}$$

Przy założeniu o normalności składnika losowego można pokazać, że:

$$E(I) = \frac{-N}{N-K} \frac{1}{[tr((\mathbf{W}^T + \mathbf{W})\mathbf{W})]^{1/2}} tr(\mathbf{P}_\mathbf{X} \mathbf{W}),$$

$$Var(I) = C\left[tr(\mathbf{M}_\mathbf{X} \mathbf{W} \mathbf{M}_\mathbf{X} \mathbf{W}^T) + tr(\mathbf{M}_\mathbf{X} \mathbf{W} \mathbf{M}_\mathbf{X} \mathbf{W}) + [tr(\mathbf{M}_\mathbf{X} \mathbf{W})]^2 \right] - E^2(I),$$

gdzie: $C = \dfrac{N^2}{(N-K)(N-K+2)} \dfrac{1}{[tr((\mathbf{W}^T + \mathbf{W})\mathbf{W})]}$.

W przypadku małych prób, wartość oczekiwana statystyki Morana I może być zatem różna od zera, a jej wariancja różna od jedności. Dla małych prób można jednak zaproponować standaryzację, która umożliwi testowanie obecności autokorelacji przestrzennej.

Warto zauważyć, iż w literaturze spotyka się liczne uproszczenia postaci wariancji statystyki Morana I, przy przyjęciu pewnych dodatkowych warunków uszczegóławiających, np.: $\mathbf{W}^T = \mathbf{W}$.

Przy pewnych założeniach dotyczących macierzy $\mathbf{W}$ zachodzi, iż:

$$\lim_{N\to\infty} E(I) = 0, \quad \lim_{N\to\infty} Var(I) = 1.$$

Ponieważ $I \xrightarrow{D} \mathrm{N}(0,1)$ możemy wnioskować, że:

$$\frac{I - E(I)}{[Var(I)]^{1/2}} \xrightarrow{D} \mathrm{N}(0,1).$$

I tak na przykład, dla testu Morana hipoteza zerowa o braku autokorelacji przestrzennej, przy poziomie istotności 0,05, odrzucana jest gdy:

$$\frac{I - E(I)}{[Var(I)]^{1/2}} \geq 1{,}96$$

3. NIESTANDARDOWY TEST MORANA *I* W MODELU AUTOREGRESJI PRZESTRZENNEJ Z AUTOKORELACJĄ PRZESTRZENNĄ SKŁADNIKA LOSOWEGO

Rozważmy teraz model autoregresji przestrzennej z autokorelacją przestrzenną składnika losowego:

$$\begin{aligned} \mathbf{y} &= \rho\mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \mathbf{u}, \\ \mathbf{y} &= \mathbf{Z}\boldsymbol{\delta} + \mathbf{u}, \\ \mathbf{u} &= \lambda\mathbf{W}\mathbf{u} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim \mathrm{N}(\mathbf{0}, \sigma^2\mathbf{I}), \end{aligned} \tag{5}$$

gdzie $|\rho|<1$ – współczynnikiem przestrzennym, $|\lambda|<1$ – współczynnikiem przestrzennej autokorelacji składnika losowego oraz $\mathbf{Z} = [\mathbf{X}, \mathbf{W}\mathbf{y}]$, $\boldsymbol{\delta} = [\boldsymbol{\beta}^T, \rho]^T$.

Przy standardowych założeniach możemy rozważać estymator Metody Zmiennych Instrumentalnych dla parametru $\boldsymbol{\delta}$:

$$\hat{\boldsymbol{\delta}} = (\hat{\mathbf{Z}}^T\hat{\mathbf{Z}})^{-1}\hat{\mathbf{Z}}^T\mathbf{y},$$

gdzie: $\hat{\mathbf{Z}} = \mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{Z}$, $\mathbf{H}$ – macierz instrumentów, oraz $\mathbf{e} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\delta}}$, oznacza wektor reszt modelu. Stawiamy hipotezę zerową H_0: $\lambda \neq 0$, przy hipotezie alternatywnej H_1: $\lambda \neq 0$ Statystyka Morana I dla modelu autoregresji przestrzennej z autokorelacją przestrzenną składnika losowego przyjmuje postać:

$$I = \frac{\mathbf{e}^T\mathbf{W}\mathbf{e}}{\left[(N^{-1}\mathbf{e}^T\mathbf{e})^2\,[tr((\mathbf{W}^T + \mathbf{W})\mathbf{W})] + (N^{-1}\mathbf{e}^T\mathbf{e})\hat{b}^T\hat{b}\right]^{1/2}}, \tag{6}$$

gdzie: $\hat{b} = -\mathbf{H}(\hat{\mathbf{Z}}^T\hat{\mathbf{Z}})^{-1}\mathbf{Z}^T\mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{Z}^T(\mathbf{W} + \mathbf{W}^T)\mathbf{e}$, oraz $I \xrightarrow{D} \mathrm{N}(0,1)$[3].

Zatem przy dwustronnym teście na poziomie istotności 5% hipoteza zerowa H_0: $\lambda = 0$ jest odrzucana dla wartości $|\boldsymbol{I}| > 1{,}96$.

4. TEST F DLA MODELU PRZESTRZENNEGO

Analogicznie jak dla modeli klasycznych, do porównywania oszacowanych przestrzennych modeli ekonometrycznych można wykorzystać test F. Rozważmy zatem dwa przykładowe modele przestrzenne, przestrzenny model autoregresyjny SAR oraz przestrzenny model SADL (Spatially autoregressive-distributed lag model) (Olejnik, 2008):

$$\text{SAR:} \quad \mathbf{y} = \rho\mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim \mathrm{N}(\mathbf{0}, \sigma^2\mathbf{I}),$$
$$\text{SADL:} \quad \mathbf{y} = \rho\mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \mathbf{W}\mathbf{X}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim \mathrm{N}(\mathbf{0}, \sigma^2\mathbf{I}),$$

gdzie $\mathbf{X}$ to macierz zmiennych objaśniających o wymiarach $(N \times K)$, $\boldsymbol{\beta}$ to odpowiadający jej k-elementowy wektor parametrów, natomiast $\boldsymbol{\gamma}$ jest k-elementowym wektorem parametrów dla ważonej przestrzennie macierzy zmiennych objaśniających. Przyjmijmy założenie, że $\boldsymbol{\beta} \neq \mathbf{0}$. W celu zbadania istotności wektora przestrzennego $\boldsymbol{\gamma}$ stawiamy hipotezę: H_0: $\gamma = \mathbf{0}$ przy hipotezie alternatywnej H_1: $\gamma \neq \mathbf{0}$. Statystyka F przyjmuje wówczas postać:

$$F = \frac{N - (2K+1)}{K+1}\frac{[RSS_{\mathrm{SAR}} - RSS_{\mathrm{SADL}}]}{RSS_{\mathrm{SADL}}} \sim F(K+1, N-(2K+1)),$$

[3] Rozumowanie pozwalające stwierdzić, że statystyka o postaci (6) dąży według dystrybuanty do rozkładu normalnego można znaleźć w opracowaniu Kelejian i Prucha (2001).

gdzie RSS_{SADL} oznacza sumę kwadratów reszt z modelu SADL, zaś RSS_{SAR} – sumę kwadratów reszt modelu SAR.

Za pomocą testu *F* istnieje również możliwość wykonania analizy porównawczej dla dowolnych dwóch modeli przestrzennych. Zauważmy, że analogiczne rozumowanie możemy przeprowadzić również dla modelu przestrzennego i klasycznej regresji liniowej. Taka analiza w istotny sposób może pomóc w podjęciu decyzji o zasadności, bądź nie zastosowania modelu przestrzennego. W ocenie autora test ten wydaje się być niedoceniany przez badaczy pomimo swej prostej konstrukcji i istotnej roli dla wyboru właściwej postaci modelu. Za pomocą powyższego testu można dokonywać porównań nie tylko pomiędzy modelami o takiej samej postaci funkcyjnej, ale również różniących się macierzami wag przestrzennych **W**.

5. TEST J DLA SPECYFIKACJI MODELU PRZESTRZENNEGO

Poniżej zaprezentujemy test oparty na procedurze testu *J*, będącej znanym algorytmem testowania modeli niezagnieżdżonych (por. Kelejian, 2008). Przedstawiony tu test, umożliwia porównanie wybranego modelu przestrzennego z pojedynczym lub wieloma niezagnieżdżonymi modelami alternatywnymi. Rozważane modele mogą, choć nie muszą zawierać opóźnienia przestrzenne, zarówno w zmiennej zależnej, jak i składniku losowym. Idea testu sprowadza się do odpowiedzi na pytanie, czy model alternatywny wnosi dodatkową informację w stosunku do modelu wyjściowego.

Rozważmy autoregresyjny model przestrzenny z autokorelacją składnika losowego. Jako hipotezę zerową, zaproponujmy model wyjściowy:

$$H_0: \left| \begin{array}{l} \mathbf{y} = \rho \mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \\ \quad = \mathbf{Z}\boldsymbol{\delta} + \mathbf{u} \\ \mathbf{u} = \lambda \mathbf{W}\mathbf{u} + \boldsymbol{\varepsilon} \quad \boldsymbol{\varepsilon} \sim \mathrm{N}(\mathbf{0}, \sigma^2 \mathbf{I}) \end{array} \right|. \tag{8}$$

Jako alternatywę, przyjmujemy *G* przeciwstawnych modeli:

$$H_i: \left| \begin{array}{l} \mathbf{y} = \rho_i \mathbf{W}_i\mathbf{y} + \mathbf{X}_i\boldsymbol{\beta}_i + \mathbf{u}_i \\ \quad = \mathbf{Z}_i\boldsymbol{\delta}_i + \mathbf{u}_i \\ \mathbf{u}_i = \lambda_i \mathbf{W}_i\mathbf{u}_i + \boldsymbol{\varepsilon}_i \quad \boldsymbol{\varepsilon} \sim \mathrm{N}(\mathbf{0}, \sigma^2 \mathbf{I}) \end{array} \right|, \tag{9}$$

gdzie $i = 1, ..., G$. Zauważmy, że w szczególnym przypadku, model wyjściowy od modeli alternatywnych może różnić się tylko postacią przestrzennej macierzy wag, co z punktu widzenia analiz przestrzennych jest kwestią bardzo istotną.

W wyniku przemnożenia obustronnego przez macierz $(\mathbf{I}-\lambda\mathbf{W})$ modelu (8) możemy go zapisać w postaci:

$$(\mathbf{I}-\lambda\mathbf{W})\mathbf{y}=(\mathbf{I}-\lambda\mathbf{W})\mathbf{Z}\boldsymbol{\delta}+\boldsymbol{\varepsilon}. \tag{10}$$

Dla skrócenia zapisu, przyjmijmy następujące oznaczenia: $\mathbf{y}(\lambda)=(\mathbf{I}-\lambda\mathbf{W})\mathbf{y}$, $\mathbf{Z}(\lambda)=(\mathbf{I}-\lambda\mathbf{W})\mathbf{Z}$, $\mathbf{Z}_i(\lambda_i)=(\mathbf{I}-\lambda_i\mathbf{W}_i)\mathbf{Z}_i$. Wówczas model (10) przyjmie postać:

$$\mathbf{y}(\lambda)=\mathbf{Z}(\lambda)\boldsymbol{\delta}+\boldsymbol{\varepsilon}. \tag{11}$$

Przez $\hat{\boldsymbol{\delta}}_i$ oznaczmy pewien estymator parametru $\boldsymbol{\delta}_i$. Wówczas modyfikacja klasycznego testu *J* (por. Davidson i MacKinnon, 1981; Pesaran i Weeks, 2001) na przypadek przestrzenny wymaga rozszerzenia modelu (11) do postaci:

$$\begin{aligned}\mathbf{y}(\lambda)&=\mathbf{Z}(\lambda)\boldsymbol{\delta}+\sum_{i=1}^{G}\alpha_i[\mathbf{Z}_i(\lambda_i)\hat{\boldsymbol{\delta}}_i]+\boldsymbol{\varepsilon}\\&=\mathbf{Z}(\lambda)\boldsymbol{\delta}+\sum_{i=1}^{G}\alpha_i[\mathbf{Z}_i\hat{\boldsymbol{\delta}}_i]+\sum_{i=1}^{G}\phi_i[\mathbf{W}_i\mathbf{Z}_i\hat{\boldsymbol{\delta}}_i]+\boldsymbol{\varepsilon}.\end{aligned} \tag{12}$$

gdzie α_i, ϕ_i, są parametrami oraz spełniony jest warunek: $\phi_i=-\alpha_i\lambda_i$. Zauważmy, że dla modelu (8), z hipotezy zerowej mamy: $\alpha_i = 0$. Na mocy powyższego możemy przeformułować testowaną hipotezę:

$$H_0\colon \alpha_i=\phi_i=0,\ \text{dla } i=1,\ \ldots,\ G \tag{13}$$

W swojej pracy Kelejian (2008), jako procedurę estymacyjną, zaproponował metodę opartą na zastosowaniu zmiennych instrumentalnych. Rozważmy zatem następujące macierze instrumentów:

$$\mathbf{H}=[\mathbf{X},\mathbf{W}\mathbf{X},\mathbf{W}^2\mathbf{X},\ \ldots,\ \mathbf{W}^r\mathbf{X}]_{LNZ},$$

$$\mathbf{H}_i=[\mathbf{X}_i,\mathbf{W}_i\mathbf{X}_i,\mathbf{W}_i^2\mathbf{X}_i,\ \ldots,\ \mathbf{W}_i^r\mathbf{X}_i]_{LNZ},$$

$$\overline{\mathbf{H}}_i=[\overline{\mathbf{X}},\mathbf{W}\overline{\mathbf{X}},\mathbf{W}^2\overline{\mathbf{X}},\ \ldots,\ \mathbf{W}^r\overline{\mathbf{X}}]_{LNZ},$$

gdzie $\overline{\mathbf{X}}=[\mathbf{X},\mathbf{X}_1,\mathbf{X}_2,\ \ldots,\ \mathbf{X}_G]$, natomiast indeks *LNZ*, oznacza liniowo niezależne kolumny rozważanej macierzy[4]. Zakłada się, iż przyjęte macierze instrumentów zawierają co najmniej wszystkie liniowo niezależne kolumny (w sensie rozpina-

[4] W większości przypadków nie rozważa się instrumentów z macierzą wag w potędze wyższej niż dwa.

nej przestrzeni) odpowiednich macierzy: $[\mathbf{X}, \mathbf{WX}]$, $[\mathbf{X}_i, \mathbf{W}_i\mathbf{X}_i]$, oraz $[\overline{\mathbf{X}}, \mathbf{W}\overline{\mathbf{X}}]$, dla $i = 1, \ldots, G$. Przy powyższych oznaczeniach, możemy przedstawić etapy procedury postępowania w przypadku testu J dla modeli przestrzennych.

Etap 1

Estymujemy model (8) metodą 2MNK (dwustopniową MNK) przy użyciu macierzy instrumentów $\mathbf{H}$, skąd otrzymujemy reszty modelu, który oznaczmy przez $\hat{\mathbf{u}}$. Równolegle, estymujemy G modeli alternatywnych postaci (9), metodą 2MNK, przy użyciu macierzy instrumentów $\mathbf{H}_i$, uzyskując oszacowania parametrów $\boldsymbol{\delta}_i$.

Etap 2

Stosujemy procedurę UMM (Uogólnionej Metody Momentów) dla modelu (8), w celu estymacji parametru λ z wykorzystaniem wyliczonych wcześniej reszt modelu $\hat{\mathbf{u}}$. Podstawiając oszacowany parametr $\hat{\lambda}$ do modelu (11) należy oszacować tak otrzymany model metodą 2MNK przy zastosowaniu macierzy instrumentów $\mathbf{H}$. Następnie, otrzymany z estymacji wektor reszt $\hat{\boldsymbol{\varepsilon}}$, wykorzystujemy do estymacji wariancji składnika losowego:

$$\hat{\sigma}_{\boldsymbol{\varepsilon}}^2 = N^{-1}\hat{\boldsymbol{\varepsilon}}^T\hat{\boldsymbol{\varepsilon}}. \tag{14}$$

Etap 3

W rozszerzonym modelu (12) następujemy parametr λ jego oszacowaniem $\hat{\lambda}$, otrzymanym w poprzednim kroku.

Przyjmijmy oznaczenia:

$$\Upsilon = [\mathbf{Z}_1\hat{\boldsymbol{\delta}}_1, \ldots, \mathbf{Z}_G\hat{\boldsymbol{\delta}}_G, \mathbf{W}_1\mathbf{Z}_1\hat{\boldsymbol{\delta}}_1, \ldots, \mathbf{W}_G\mathbf{Z}_G\hat{\boldsymbol{\delta}}_G], \psi = [\alpha_1, \ldots, \alpha_G, \phi_1, \ldots, \phi_G]^T.$$

Wówczas empirycznym odpowiednikiem rozszerzonego modelu (12) będzie model postaci:

$$\mathbf{y}(\hat{\lambda}) \approx \mathbf{Z}(\hat{\lambda})\boldsymbol{\delta} + \Upsilon\boldsymbol{\psi} + \boldsymbol{\varepsilon}. \tag{15}$$

Etap 4

W kolejnym kroku należy oszacować parametry modelu (15) metodą 2MNK z zastosowaniem macierzy instrumentów $\overline{\mathbf{H}}$. Oznaczmy macierz zmiennych objaśniających w modelu (15) przez $\Xi \equiv [\mathbf{Z}(\hat{\lambda}), \Upsilon]$, parametry regresji zaś przez $\boldsymbol{\gamma}^T \equiv [\boldsymbol{\delta}^T, \boldsymbol{\psi}^T]$. Zauważmy, że dla modelu określonego w hipotezie zerowej mamy $\boldsymbol{\gamma}_0^T = [\boldsymbol{\delta}^T, \mathbf{0}^T]$. Niech $\hat{\Xi} = \mathbf{P}_{\overline{\mathbf{H}}}\Xi \equiv [\hat{\mathbf{Z}}(\hat{\lambda}), \hat{\Upsilon}]$, gdzie operator rzutu na macierz Ξ jest postaci: $\mathbf{P}_{\overline{\mathbf{H}}} = \overline{\mathbf{H}}(\overline{\mathbf{H}}^T\overline{\mathbf{H}})^{-1}\overline{\mathbf{H}}^T$

Wówczas estymator 2MNK parametru γ przyjmuje postać:

$$\hat{\boldsymbol{\gamma}} = (\hat{\Xi}^T\hat{\Xi})^{-1}\hat{\Xi}^T\mathbf{y}(\hat{\lambda}). \tag{16}$$

Twierdzenie

Przy pewnych standardowych założeniach (por. Kelejian, 2008) i przyjętych powyżej oznaczeniach, mamy:

$$N^{1/2}(\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}_0)\xrightarrow{D} N(\mathbf{0},\hat{\sigma}^2_{\varepsilon} \operatorname*{plim}_{N\to\infty}(\hat{\Xi}^T\hat{\Xi})^{-1}),$$

$$\sigma^2_{\varepsilon} = \operatorname*{plim}_{N\to\infty} \hat{\sigma}^2_{\varepsilon},$$

gdzie plim oznacza zbieżność względem prawdopodobieństwa[5].

Na mocy powyższego twierdzenia, wnioskowania dla małych prób mogą być oparte na założeniu:

$$\hat{\boldsymbol{\gamma}} \simeq N\left(\boldsymbol{\gamma},\hat{\sigma}^2_{\varepsilon}(\hat{\Xi}^T\hat{\Xi})^{-1}\right). \tag{17}$$

Niech $\overline{K}\equiv K+1+2G$, $\hat{\boldsymbol{\gamma}}^T\equiv[\hat{\boldsymbol{\delta}}^T,\hat{\boldsymbol{\psi}}^T]$ oraz niech $\widehat{\mathbf{VC}}_{\hat{\psi}}$ reprezentuje oszacowanie macierzy wariancji-kowariancji estymatora $\hat{\boldsymbol{\psi}}$ opartego na małej próbie. Wówczas $\widehat{\mathbf{VC}}_{\hat{\psi}}$ stanowi podmacierz o wymiarach 2G×2G macierzy wariancji-kowariancji parametru $\hat{\boldsymbol{\gamma}}$ – macierzy $\hat{\sigma}^2_{\varepsilon}(\hat{\Xi}^T\hat{\Xi})^{-1}$ o wymiarach $\overline{K}\times\overline{K}$. Wówczas test Walda (por. Godfrey i Pesaran, 1983) odrzuca hipotezę zerową H_0: $\boldsymbol{\psi} = \mathbf{0}$ na korzyść hipotezy alternatywnej H_1: $\boldsymbol{\psi} \neq \mathbf{0}$ na poziomie istotności α, gdy:

$$\left(\hat{\boldsymbol{\psi}}^T \widehat{\mathbf{VC}}_{\hat{\psi}}^{-1} \hat{\boldsymbol{\psi}}\right) > \chi^2_{1-\alpha}(2G), \tag{18}$$

gdzie $\chi^2_{(1-\alpha)}(2G)$ oznacza kwantyl rozkładu χ^2 o 2G stopniach swobody, rzędu 1-α.

6. TESTOWANIE STACJONARNOŚCI PRZESTRZENNEJ

Problem niestacjonarności przestrzennej jako pierwszy rozważał Fingleton (1999). W swej pracy zauważył on, że przestrzenna niestacjonarność, analogicznie jak w przypadku szeregów czasowych, może prowadzić do wystąpienia regresji pozornej. Poniżej zaprezentowano procedurę testową umożliwiającą wykluczenie występowania problemu niestacjonarności przestrzennej.

Jako punkt wyjścia przyjmijmy model postaci:

$$\mathbf{y} = \boldsymbol{\rho}\mathbf{W}\mathbf{y} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim N(\mathbf{0},\sigma^2\mathbf{I}). \tag{19}$$

---

[5] Dowód powyższego twierdzenia można znaleźć w pracy Kelejian (2008).

W powyższym równaniu spełnienie warunku $|\rho|<1$ zapewnia co najmniej asymptotyczną stacjonarność danych generujących proces przestrzenny. W tym wypadku możemy powiedzieć, że **y** jest przestrzennie zintegrowanym procesem rzędu zero – SI(0).

Rozważmy model postaci:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} = \lambda \mathbf{W}\boldsymbol{\varepsilon} + \boldsymbol{\mu}, \quad \boldsymbol{\mu} \sim \mathrm{N}(\mathbf{0}, \sigma^2\mathbf{I}). \tag{20}$$

Kosfeld i Lauridsen (2004) zaproponowali dwustopniowy test wskazujący na istnienie przestrzennej stacjonarności. Ich strategia, oparta jest na dwustopniowym teście mnożników Lagrange'a dla przestrzennie skorelowanego składnika losowego[6]. Zatem dla modelu , przy H_0: $\lambda = 0$ możemy zastosować procedurę testu *LM* dla reszt modelu (LME). Jeśli statystyka testowa okaże się nieistotnie różna od zera, powiemy, że nie mamy tu do czynienia z autokorelacją przestrzenną ($\lambda=0$). W przeciwnym wypadku możemy wnioskować o występowaniu autokorelacji przestrzennej lub też o istnieniu przestrzennej niestacjonarności.

W drugim etapie, należy zweryfikować, czy parametr przestrzennej autoregresji w równaniu (20) wynosi jeden, czy jest mniejszy od jedności. W tym celu należy zastosować test *LM* dla przestrzennie zróżnicowanego modelu (DLME) postaci:

$$\Delta\mathbf{y} = \Delta\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\mu} \quad \boldsymbol{\mu} \sim \mathrm{N}(\mathbf{0}, \sigma^2\mathbf{I}), \tag{21}$$

gdzie: $\mu=(1-\lambda\mathbf{W})\boldsymbol{\varepsilon}$. Wówczas hipoteza zerowa jest równoważna: $H_0: |\lambda|=1$, przy hipotezie alternatywnej $H_1: |\lambda|<1$.

Analogiczna procedura może zostać zastosowana w celu badania przestrzennej niestacjonarności każdej ze zmiennych rozważanego modelu. Wtedy należy wykonać regresję danej zmiennej na czynnik stały. Bardziej szczegółową analizę problemu przestrzennej niestacjonarności wraz z przykładem badania przestrzennej stacjonarności dla procesu konwergencji w UE można znaleźć w publikacji autora (Olejnik, 2008).

---

[6] Własności testu zostały dokładnie opisane w pracy Kosfelda i Lauridsena (2006). Alternatywne metody testowania można znaleźć w pracy Kosfelda i Lauridsena (2004), gdzie zastosowana procedura oparta została na teście Walda oraz w opracowaniu Lauridsena (1999), w którym wykorzystano podejście Dickey-Fullera.

7. PODSUMOWANIE

W artykule przedstawiono metody testowania autokorelacji przestrzennej w najpopularniejszych modelach przestrzennych. Jako narzędzie do porównywania dwóch ekonometrycznych modeli przestrzennych, lub modelu przestrzennego z klasycznym zaproponowano zastosowanie znanego testu F. Z bardziej zaawansowanych narzędzi wspomagających dokonanie wyboru właściwego modelu zaprezentowano test J. Mimo dość skomplikowanej konstrukcji testu, jego rola dla oceny poprawności przyjętej struktury przestrzennej jest bardzo duża. Zdaniem autora, w chwili obecnej stanowi on najlepsze narzędzie porównawcze nieposiadające porównywalnej alternatywy. Ostatnia część pracy przybliża problem przestrzennej niestacjonarności oraz prezentuje odpowiednią procedurę testującą.

Mimo szerokich możliwości przedstawionych metod można zauważyć konieczność dalszego rozszerzania i modyfikacji zaprezentowanej metodologii. W dalszych opracowaniach planowane jest przedstawienie zatem konstrukcji testów dla przestrzennych modeli wielowymiarowych. W szczególności planuje się opracowanie wielowymiarowego testu J dla modeli klasy WAMP – Wielowymiarowych autoregresyjnych modeli przestrzennych (por. Olejnik, 2013).

Uniwersytet Łódzki

LITERATURA

[1] Anselin L., (1988), *Spatial Econometrics: Methods and Models*, Kluwer Academic Publications, Dordrecht.
[2] Burridge P., (1980), On the Cliff–Ord Test for Spatial Correlation, *Journal of the Royal Statistical Society*, B, 42 (1), 107–108.
[3] Cliff A., Ord J. K., (1981), *Spatial Processes: Models and Applications*, Pion, London.
[4] Drukker D. M., Prucha I. R., (2013), Finite Sample Properties of the $I^2(q)$ Test Statistic for Spatial Dependence, *Spatial Economic Analysis*, 8, 271–292.
[5] Fingleton B., (1999), Spurious Spatial Regression: Some Monte Carlo Results with Spatial Unit Root and Spatial Cointegration, *Journal of Regional Science*, 39, 1–19.
[6] Godfrey L., Pesaran H., (1983), Tests of Nonnested Regression Models After Estimation by Instrumental Variables or Least Squares, *Journal of Econometrics*, 21, 133–154.
[7] Kelejian H. H., (2008), A Spatial Test J for Model Specification Against a Single or a Set of Non–Nested Alternatives, *Letters in Spatial and Resource Sciences*, 1 (1), 3–11.
[8] Kelejian H. H., Prucha I. R., (2001), On the Asymptotic Distribution of the Moran I Test Statistic with Applications, *Journal of Econometrics*, 104, 219–257.
[9] Kosfeld R., Lauridsen J., (2004), Dynamic Spatial Modeling of Regional Convergence Processes, *Empirical Economics*, 29, 705–722.
[10] Kosfeld R., Lauridsen J., (2006), A Test Strategy for Spurious Regression, Spatial Nonstationarity, and Spatial Cointegration, *Papers in Regional Science*, 85 (3), 363–377.
[11] Lauridsen J., (1999), Spatial Cointegration Analysis in Econometric Modelling, *ERSA conference papers ersa99pa181*, European Regional Science Association.
[12] Moran P., (1950), Notes on Continuous Stochastic Phenomena, *Biometrika*, 37, 17–23.

[13] Olejnik A., (2008), Using the Spatial Autoregressively Distributed Lag Model in Assessing the Regional Convergence of per–capita Income in the EU25, *Papers in Regional Science*, 87 (3), 371–384.
[14] Olejnik A., (2013), Spatial Autoregressive Model – a Multidimensional Perspective with an Example Study of the Spatial Income Process in the EU 25, PREPARE, working papers, wysłane do recenzji do *Geographical Analysis*.
[15] Pesaran H., Weeks M., (2001), Nonnested Hypothesis Testing: An Overview, w: Baltagi B., (red.), *A Companion to Theoretical Econometrics*, Blackwell, Oxford.

WYBRANE METODY TESTOWANIA MODELI REGRESJI PRZESTRZENNEJ

Streszczenie

Celem niniejszej pracy jest przybliżenie nowoczesnych metod testowania modeli regresji przestrzennej pozwalających wybrać poprawną postać struktury zależności przestrzennych. Metody te umożliwiają testowanie występowania autokorelacji przestrzennej, a w dalszym etapie pozwalają na właściwą specyfikację modelu przestrzennego. W artykule wskazano wady i zalety prezentowanych metod oraz przedstawiono pewne rekomendacje dotyczące ich stosowania. W opracowaniu przedstawiono również problem przestrzennej niestacjonarności oraz zaproponowano właściwą procedurę weryfikacyjną.

Słowa kluczowe: modele regresji przestrzennej, test *J*, test *F*, Statystyka Morana *I*, przestrzenna niestacjonarność

SOME TESTING METHODS FOR SPATIAL REGRESSION MODELS

Abstract

The aim of this paper is to introduce modern testing methods for Spatial Regression Models used for selection of the correct structure of spatial dependencies. Presented methodology allows one to test for the presence of spatial autocorrelation and at a further stage to make the correct model specification. The paper presents some merits and drawbacks of selected statistical test widely used in spatial modelling with some recommendations. The problem of spatial non-stationarity with an appropriate testing procedure is also introduced.

Keywords: spatial regression models, *J* test, *F* test, Moran *I*, spatial non-stationarity