EKONOMETRIA ECONOMETRICS 34 • 2011

ISSN 1507-3866

Beata Bal-Domańska

Uniwersytet Ekonomiczny we Wrocławiu

WYBRANE PROBLEMY AUTOKORELACJI W MODELOWANIU NA PODSTAWIE DANYCH PANELOWYCH

Streszczenie: Rosnące zainteresowanie analizami na poziomie regionalnym i lokalnym kieruje uwagę badaczy na problemy modelowania informacji o jednostkach przestrzennych. Coraz więcej miejsca poświęca się analizom z wykorzystaniem danych panelowych. Ich popularność wynika z możliwości prowadzenia wszechstronnych analiz i potencjalnych korzyści dla jakości i zakresu analiz. Wykorzystanie danych przekrojowych i panelowych wiąże się z trudnościami w weryfikacji podstawowych założeń Klasycznej Metody Najmniejszych Kwadratów – estymatora najczęściej wykorzystywanego w szacowaniu parametrów strukturalnych lub będącego podstawą pokrewnych estymatorów, a mianowicie założeń związanych z autokorelacją składnika losowego. W artykule przedstawiono wybrane problemy modelowania na podstawie danych panelowych z uwzględnieniem autokorelacji przestrzennej i w czasie.

Słowa kluczowe: dane panelowe, autokorelacja składnika losowego, korelacja przestrzenna.

1. Wstęp

Rozpoznanie charakteru zjawisk społeczno-gospodarczych i ich analiza wymaga zastosowania właściwych metod badawczych. W wieloletniej historii budowy modeli ekonometrycznych opracowane zostały różne techniki modelowania, dostosowujące formułę estymacji i weryfikacji do potrzeb specyficznych problemów analizowanych zjawisk. Rozwinięte zostały techniki estymacji m.in. modeli liniowych i nieliniowych, jedno- i wielorównaniowych, zbudowanych na podstawie czasowych, przekrojowych bądź panelowych szeregów danych. W przypadku tego ostatniego podziału szeroko wykorzystywane były modele zbudowane na podstawie czasowych szeregów danych, natomiast modele przekrojowe i panelowe cieszyły się mniejszym zainteresowaniem. Wraz ze wzrostem zainteresowania m.in. polityką i gospodarką regionalną wzrosło znaczenie analiz przekrojowych i panelowych, co sprzyja prowadzeniu dyskusji nad własnościami metod wykorzystywanych w analizach przestrzennych i panelowych.

2. Autokorelacja składnika losowego

Składnik losowy ε jest stochastycznym elementem modelu, który obrazuje nieprzewidywalność opisywanych zjawisk, wpływ pominiętych zmiennych lub błędy pomiaru. W liniowym modelu regresji z addytywnym składnikiem losowym można go uwzględnić w następujący sposób:

$$y = \alpha + \beta x + \varepsilon, \tag{1}$$

gdzie: y – zmienna zależna (objaśniana), $(\alpha + \beta x)$ – składnik deterministyczny, α – wyraz wolny, β – parametr strukturalny, x – zmienna niezależna (objaśniająca).

Poprawność oszacowań parametrów strukturalnych modelu zależy od stopnia spełnienia założeń przyjętej metody estymacji. Jedną z najpopularniejszych metod estymacji liniowych modeli regresji jest klasyczna metoda najmniejszych kwadratów (KMNK) (zob. m.in.: [Maddala 2006, s. 104 i nast.]). Pożądane własności tej metody można w skrócie opisać słowem BLUE (*Best Linear Unbiased Estymator* – najefektywniejszy liniowy estymator nieobciążony).

Jednym z kluczowych postulatów KMNK jest założenie sferyczności składnika losowego, w tym o niezależności jego wartości od wartości składnika losowego dla innych obserwacji. Niespełnienie tego założenia powoduje, że estymator parametrów strukturalnych jest nadal zgodny i nieobciążony, ale staje się mało efektywny, a błędy standardowe są obciążone. Stosowanie standardowych statystyk testowych w modelu z niesferycznym składnikiem losowym może skutkować błędnymi wnioskami.

Niezależność składnika losowego może wynikać z dwóch powodów. Pierwszy z nich dotyczy autokorelacji składnika losowego w szeregu czasowym, co ogólnie możemy zapisać jako $E(\varepsilon_t\varepsilon_{(t+q)}) = 0$, gdzie q oznacza liczbę całkowitą różną od zera. Oznacza to, że składnik losowy ε_t z okresu t może być skorelowany ze składnikami z okresów przeszłych: ε_{t-1}, ε_{t-2} itd. lub przyszłych ε_{t+1}, ε_{t+2} itd. Jeżeli $E(\varepsilon_t\varepsilon_{(t-1)}) = 0$, mamy do czynienia z autokorelacją pierwszego rzędu, gdy autokorelacji podlegają obserwacje oddalone o k okresów, mówimy o występowaniu autokorelacji k-tego rzędu $E(\varepsilon_t\varepsilon_{(t-k)}) = 0$. Do diagnozowania autokorelacji pierwszego rzędu powszechnie stosowany jest test Durbina-Watsona lub h-Durbina [Durbin, Watson 1950, s. 409--428; 1951, s. 159-179] (dla modelu z prawostronnymi zmiennymi endogenicznymi opóźnionymi w czasie). Do badania autokorelacji dowolnego rzędu wykorzystać można m.in.: test PACF (*Partial Autocorrelation Function*) [Box, Jenkins, Reinsel 1994], LM mnożników Lagrange'a oraz test Ljunga-Boxa [Osińska 2007, s. 82 i nast.]. Przyczyn autokorelacji w regresji na podstawie czasowych szeregów danych upatrywać można m.in. w: charakterze zjawisk gospodarczych, których skutki odczuwane są przez wiele kolejnych okresów, źle dobranych opóźnieniach zmiennych objaśniających, wzajemnym skorelowaniu pominiętych zmiennych objaśniających, niewłaściwej postaci analitycznej. Z autokorelacją dodatnią mamy do czynienia wtedy, gdy skutki zdarzeń losowych mogą trwać dłużej niż jeden okres, w efekcie otrzymujemy ciąg reszt o dużych wartościach. Zbyt częste zmiany wartości reszt świadczą o autokorelacji ujemnej.

Drugim przypadkiem jest sytuacja skorelowania składników losowych w szeregach przekrojowych. Jest to wynikiem powiązania ze sobą wybranych obiektów. Szczególnie dotyczy to obiektów (indywidualnych, jak gospodarstwa domowe, firmy, oraz ich grup: regionów, państw) umieszczonych w pewnym sąsiedztwie. Modelowanie z uwzględnieniem autokorelacji przestrzennej jest dużo trudniejsze niż czasowej, gdyż nie ma naturalnego uporządkowania obserwacji, a powiązania i wpływy są wielokierunkowe.

Problemy korelacji przestrzennej skłoniły G. Arbię [2006] do wyróżnienia (obok szeregów przekrojowych rozumianych jako obserwacje wielu obiektów, np. gospodarstw domowych, firm, regionów w wybranej jednostce czasu) danych przestrzennych (*spatial series*) definiowanych jako dane o poszczególnych obiektach wraz z dodatkowymi informacjami o ich położeniu w przestrzeni. Przy tym podejściu zakłada się istnienie zależności między sąsiadującymi obiektami przestrzeni (np. regionami), które przekładają się na niedoskonałości oszacowań. Podejście to umożliwia formułowanie hipotez podlegających weryfikacji przy formułowaniu modelu ekonometrycznego, szczególnie problemu sferyczności składnika losowego, co pozwala na uniknięcie ewentualnych, niepożądanych właściwości składnika losowego, a w konsekwencji niedoskonałości oszacowań modelu.

Z dodatnią autokorelacją przestrzenną mamy do czynienia w sytuacji, gdy obiekty sąsiednie mają zbliżone wartości. Autokorelacja ujemna to zdecydowanie różne wartości w obiektach położonych w swoim sąsiedztwie.

Podstawowymi przyczynami korelacji przestrzennej składników losowych są siła oddziaływania między jednostkami oraz niezgodność pomiędzy granicami analizowanych procesów a granicami jednostek administracyjnych. W tym pierwszym przypadku chodzi o zależność oddziaływania i odległości. Zakłada się, że siła oddziaływania między jednostkami maleje w miarę wzrostu odległości (warunki izotropowości) lub wraz ze zmianami odległości i kierunku (warunki anizotropowości). W drugim przypadku autokorelacja przestrzenna jest konsekwencją agregacji informacji według zdefiniowanych jednostek administracyjnych z pominięciem rzeczywistego oddziaływania w przestrzeni. W efekcie, jeśli oddziaływanie przekracza granice administracyjne, między jednostkami może zachodzić korelacja [Suchecki 2010]. W ekonometrycznych analizach przestrzennych mamy więc do czynienia z odległością ekonomiczną (a nie fizyczną) między obiektami [Szulc 2005]. Stanowi to duże utrudnienie w formułowaniu modelu oraz zdobywaniu informacji statystycznych. Pierwsze komplikacje napotykamy już na etapie zdefiniowania odległości ekonomicznej, gdyż możliwe jest skonstruowanie kilku wariantów zależności przestrzennych. Wykorzystanie różnych definicji odległości ekonomicznych może prowadzić do odmiennych rezultatów.

Zjawiskiem dodatkowo komplikującym autokorelację w przestrzeni jest opóźnione oddziaływanie zjawisk na obiekty sąsiadujące. Często ujawnienie się wpływu zjawisk w pobliskich jednostkach wymaga upływu określonego czasu. Zjawisko to określane jest mianem autokorelacji przestrzenno-czasowej [Szulc 2005].

Wykorzystanie danych panelowych powoduje, że spotykamy się z obydwoma problemami autokorelacji.

3. Identyfikacja autokorelacji w modelach dla danych panelowych

Tym, co wyróżnia dane panelowe od szeregów czasowych i przekrojowych, jest łączne uwzględnienie informacji o grupie obiektów i okresów. Są to przekrojowo-czasowe informacje zebrane dla tej samej grupy obiektów i ($i = 1, 2, \ldots, N$) w kolejnych okresach t ($t = 1, 2, \ldots, T$). Najogólniej liniowy model zbudowany na podstawie danych panelowych można zapisać jako:

$$y_{it} = \alpha_{it} + \beta_{it}^{T} x_{it} + \varepsilon_{it}, \quad \varepsilon_{it} \sim \mathrm{IID}(0,\ \sigma_{\varepsilon}^{2}). \tag{2}$$

Konstrukcja tak zapisanego modelu przewiduje możliwość wprowadzenia zmiennych parametrów. Stwarza to szerokie możliwości zastosowania modelu oraz umożliwia rozwiązanie niektórych problemów ekonometrycznych, jak np. pozwala na uniknięcie błędów wynikających z nieuwzględnienia ważnej zmiennej, użycia zmiennych zastępczych lub danych zagregowanych, nieuwzględnienia nieliniowego charakteru modelu [Dziechciarz 1993, s. 29-30].

Najpopularniejszymi metodami szacowania parametrów modeli panelowych są:

1) *pooled* estymator postaci:

$$y_{it} = \alpha_0 + \beta^{T} x_{it} + \varepsilon_{it}, \tag{3}$$

w którym zakłada się, że wpływ zmian w wartościach zmiennych objaśniających na zmienną objaśnianą jest taki sam dla wszystkich obiektów i okresów badania;

2) estymator klasycznej metody najmniejszych kwadratów ze zmiennymi sztucznymi[1], co ogólnie można zapisać jako:

$$y_{it} = \alpha_i + \beta^{T} x_{it} + \varepsilon_{it}, \tag{4}$$

gdzie: α_i – stałe w czasie efekty indywidualne dla każdego obiektu i ($i = 1, 2, \ldots, N$), przybliżone przez zbiór zmiennych zero-jedynkowych;

3) z efektami losowymi[2]:

$$y_{it} = \mu + \beta^{T} x_{it} + c_i + \varepsilon_{it}, \quad c_i \sim \mathrm{IID}(0,\ \sigma_{c}^{2}), \tag{5}$$

[1] Inne nazwy z efektami ustalonymi, z dekompozycją wyrazu wolnego: FE – *Fixed Effect,* LSDV – *Least Sqauares Dummy Variable.*

[2] Inne nazwy: modele z efektami losowymi, z dekompozycją składnika losowego, modele składowych błędu losowego, RE – *Random Effects*.

gdzie: c_i oznacza zmienną losową[3], wyrażającą zakłócenie losowe dla i-tego obiektu; $(c_i+\varepsilon_{it})$ – obserwowalny, łączny składnik losowy. W modelu tym przyjmujemy, że składniki losowe i indywidualne efekty losowe c_i pochodzące z różnych okresów i dla różnych jednostek nie są skorelowane. Ponadto chcemy, aby dla różnych jednostek składnik losowy ε_{it} oraz indywidualny efekt losowy c_i były nieskorelowane. Model ten pozwala uniknąć strat stopni swobody wynikających z wprowadzenia dużej liczby zmiennych sztucznych w modelu z efektami ustalonymi, jeżeli liczba jednostek N jest duża. Jest on szczególnie polecany, gdy jednostki badanej próby reprezentują większą populację, a ich liczba powinna zmierzać do nieskończoności.

Ze względu na charakter danych panelowych jako szeregu czasowego i przekrojowego analiza ekonometryczna obarczona jest ryzykiem wystąpienia zarówno czasowej, jak i przestrzennej autokorelacji składnika losowego. Stąd problem autokorelacji dotyczy (1) zależności w czasie między obserwacjami w każdej jednostce i (obiekcie) oraz (2) zależności przestrzennej między jednostkami i w każdym punkcie czasu.

3.1. Autokorelacja w czasie składnika losowego

W modelach z efektami ustalonymi (4) i losowymi (5) zakłada się, że może zachodzić autokorelacja składnika losowego w czasie, przy czym współczynnik autokorelacji ρ (zob. m.in.: [Verbeek 2000; Kelejian, Prucha 1999]) ma taką samą wartość dla każdej jednostki i. Proces autoregresyjny pierwszego rzędu możemy zapisać jako:

$$\varepsilon_{it} = \rho\varepsilon_{i,t-1} + \upsilon_{it},\ \upsilon_{it} \sim \mathrm{IID}(0,\ \sigma^2), \tag{6}$$

gdzie: $|\rho| < 1$ – oznacza współczynnik autoregresji pierwszego rzędu.

Ocenę występowania autokorelacji przeprowadza się dla hipotezy zerowej postaci H_0: $\rho = 0$ wobec hipotezy alternatywnej H_1: $\rho < 0$ lub H_1: $\rho > 0$.

Jednym ze sposobów testowania autokorelacji składnika losowego pierwszego rzędu [Baltagi 2005] w czasie w modelach panelowych z efektami ustalonymi jest statystyka Durbina-Watsona (d), uogólniona dla danych panelowych przez Bhargava, Franzini i Narendranathan [Bhargava, Franzini, Narendranathan 1982, s. 533-559] i zdefiniowana jako:

$$d = \frac{\sum_{i=1}^{N}\sum_{t=2}^{T}(e_{it} - e_{i,t-1})^2}{\sum_{i=1}^{N}\sum_{t=1}^{T}e_{it}^{2}}, \tag{7}$$

[3] Za Wooldridgem symbolu c użyto w modelach z efektami losowymi dla podkreślenia, że jest to nieobserwowalna zmienna losowa, a nie parametr jak w modelach z efektami ustalonymi.

gdzie: e_{it} oznacza reszty z modelu wewnątrzgrupowego[4] dla i-tego obiektu i t-tego okresu.

Wartość statystyki porównuje się z wartościami dolnymi d_l i górnymi d_u określonymi w zależności od liczby obiektów N, okresów T i zmiennych objaśniających K. Autorzy wskazują, że jest on najlepszy w przypadku, gdy ρ zbliża się do 0. Dla ρ bliskiego 1 autorzy polecają statystykę Berenblut-Webb opisaną wzorem:

$$g = \frac{\sum_{i=1}^{N}\sum_{t=2}^{T}\Delta u_{it}^{2}}{\sum_{i=1}^{N}\sum_{t=1}^{T}e_{it}^{2}}, \tag{8}$$

gdzie: u_{it} – oznacza reszty modelu na pierwszych różnicach oszacowanego klasyczną metodą najmniejszych kwadratów (KMNK) dla i-tego obiektu i t-tego okresu. Testy d i g są równoważne m.in. w przypadku prób o dużym N.

Innym testem na autokorelację składnika losowego w czasie jest opracowany w 2001 r. test Arellano-Bonda [Arellano, Bond 1991, s. 277-297]. Wykorzystywany jest on do weryfikacji hipotezy o braku autokorelacji dla dynamicznych modeli panelowych oszacowanych z wykorzystaniem estymatora uogólnionej metody momentów (UMM) dla pierwszych różnic lub systemowego estymatora UMM.

3.2. Autokorelacja w przestrzeni obiektów

Autokorelacja w odniesieniu do danych przekrojowych (panelowych) rozpatrywana jest w kontekście zależności przestrzennych między obiektami położonymi w pewnym sąsiedztwie. Przestrzenne zależności można uwzględnić w modelu przez [Suchecki 2010]:

1) autokorelację przestrzenną składnika losowego – gdy w modelu pominięto zmienne przestrzennie autoskorelowane,

2) autoregresję przestrzenną – gdy wartości zmiennej endogenicznej Y z jednostki l wpływają na kształtowania się zmiennej w jednostce i i innych,

3) krzyżową regresję przestrzenną – gdy na wartości zmiennej Y w jednostce i wpływają wartości zmiennych objaśniających X z innych jednostek.

Powstaje pytanie, która konstrukcja modelowa jest właściwa. Anselin [Anselin, Le Gallo, Jayet 2008, s. 901-969] wskazuje, że w sytuacji, gdy rozważamy model równowagi procesów przestrzennych lub społecznych, w których na kształtowanie się zmiennej objaśnianej mają wpływ także procesy zachodzące w sąsiednich jednostkach, właściwy jest model autoregresji przestrzennej. Jeżeli natomiast autokorelacja nie wynika z przestrzennych lub społecznych interakcji, a jest po prostu wyni-

[4] Oszacowanie parametrów otrzymuje się na podstawie zmiennych w postaci odchyleń od średniej grupowej: $(y_{it} - \overline{y}_i) = (x_{it} - \overline{x}_i)^T \beta + (\varepsilon_{it} - \overline{\varepsilon}_i)$, są one równe tym uzyskanym z wykorzystaniem estymatora ze zmiennymi sztucznymi (FE).

kiem niesferyczności składnika losowego, wtedy właściwy jest model z przestrzenną autokorelacją składnika losowego.

J.P. Elhorst [2003, s. 244-268] wskazuje, że model z efektami ustalonymi (4) poszerzony o przestrzennie opóźnioną zmienną zależną (*spatial leg model*) można zapisać jako:

$$y_{it} = \delta \mathbf{W} y_{jt} + \beta^T x_{it} + \alpha_i + \varepsilon_{it}, \tag{9}$$

gdzie: y_{it}, (x_{it}) – wartości zmiennej zależnej (niezależnej) dla każdego obiektu (jednostki) w jednym punkcie czasu, y_{jt} – wartości zmiennej zależnej (niezależnej) dla sąsiednich obiektów, **W** – macierz wag o wymiarach $N \times N$, której elementy diagonalne są równe zeru, δ – współczynnik autoregresji przestrzennej (skalar) przyjmujący zazwyczaj wartości z przedziału (–1,1), przy czym $1/\omega_{min} < \delta < 1/\omega_{max}$, gdzie ω_{min} i ω_{min} oznaczają najmniejszą (np. ujemną) i największą wartość pierwiastka macierzy **W**. W przypadku tego modelu zakłada się, iż zmienna objaśniająca *Y* zależy od wartości zmiennej objaśniającej *Y* w sąsiednich jednostkach i zmiennych charakteryzujących wybrane zjawiska (*x*) .

Wprowadzenie zależności przestrzennych możliwe jest także przez składnik losowy (model z autokorelacją przestrzenną składnika losowego; *spatial error model*). W tym przypadku model można zapisać jako:

$$y_{it} = \alpha_i + \beta^T x_{it} + \xi_{it}, \qquad \xi_{it} = \rho \mathbf{W} \xi_{it} + \varepsilon_{it}, \tag{10}$$

gdzie: ρ – współczynnik autokorelacji przestrzennej (skalar), przyjmujący zazwyczaj wartości z przedziału (–1,1), przy czym $1/\omega_{min} > \rho > 1/\omega_{max}$, ξ_{it} – przestrzennie skorelowany składnik losowy.

Możliwe jest także poszerzenie struktury modelu o przestrzennie opóźnione zmienne niezależne, wtedy model (9) przyjmuje postać:

$$y_{it} = \delta \mathbf{W} y_{jt} + \beta^T x_{it} + \mathbf{W} x_{ijt} \phi + \alpha_i + \varepsilon_{it}, \tag{11}$$

gdzie: φ – to wektor *k* parametrów. Jest to tzw. przestrzenny model Durbina (*Spatial Durbin Model* – SDM).

Estymacja tak sformułowanych modeli możliwa jest z wykorzystaniem metody największej wiarygodności (*Maximum Likehood* – ML), alternatywnie metody zmiennych instrumentalnych lub uogólnianej metody momentów (*Generalized Method of Moments* IV/GMM), która w przeciwieństwie do ML nie zakłada normalności rozkładu składnika losowego (zob. m.in. [Elhorst 2010]).

Możliwe jest także wykorzystanie obu powyższych podejść jednocześnie (model z opóźnioną przestrzennie zmienną zależną i opóźnionym przestrzennie składnikiem losowym) poszerzonych o uwzględnienie opóźnionych przestrzennie zmiennych niezależnych [Manski 1993, s. 531-542; Elhorst 2010]. Próba uwzględnienia wszystkich rodzajów interakcji przestrzennych (zmiennej zależnej, zmiennych niezależnych i składnika losowego) może skutkować kłopotami z identyfikacją parametrów autoregresji. Dlatego sugeruje się sprawdzenie zasadności wprowadzenia do modelu przestrzennie opóźnionych zmiennych niezależnych, a następnie udzielenie odpowiedzi na pytanie, czy model powinien być poszerzony o przestrzennie opóźnioną zmienną zależną i składnik losowy. J.P. Elhorst [Elhorst, Piras, Arbia 2006] proponuje, aby do oceny wykorzystać przestrzenny model Durbina (11). Sprawdzenie, kiedy model powinien być poszerzony o przestrzenną zmienną zależną, można zweryfikować, testując hipotezę zerową postaci: H_0: $\varphi = 0$, natomiast o przestrzennym opóźnieniu składnika losowego decydujemy na podstawie hipotezy postaci: H_0: $\varphi + \delta\beta = 0$.

Przedstawione powyżej rozwiązania nie wyczerpują możliwości estymacji modeli z efektami przestrzennymi. Efekty te mogą być uwzględnione – oprócz opisanego powyżej modelu z efektami ustalonymi – w modelach z efektami losowymi [Elhorst 2010], w modelach ze zmiennymi parametrami, w tym w modelach pozornie niezwiązanych równań regresyjnych SUR (*Seemingly Unrelated Regresions Model*) wprowadzonych przez Zellnera [1962, s. 348-368], składających się z układu równań oszacowanych na podstawie danych przestrzennych dla każdego okresu [Elhorst 2010]. Jednakże tematyka ta wykracza poza przewidziane ramy artykułu.

Kluczowym pojęciem dla autokorelacji przestrzennej jest pojęcie sąsiedztwa i definicja macierzy je opisującej. Do określenia odległości obiektów wykorzystywane są wagi przestrzenne w_{ij} (*spatial lag operator*) tworzące macierz wag **W**, której pozadiagonalne elementy charakteryzują przestrzenną strukturę zależności. Macierz ta budowana jest na podstawie macierzy sąsiedztwa lub odległości (więcej zob. m.in. [Suchecki 2010]). Wartości w_{ij} są miarami odległości ustalonymi dla wybranej jednostki i (wiersz macierzy) względem jednostki sąsiedniej j (kolumny). Elementy na głównej przekątnej przyjmują wartość zero ($w_{ii} = 0$), gdyż żadna jednostka nie może być sąsiedztwem dla samej siebie. Pozostałe elementy przyjmują wartości różne od zera wtedy, gdy dwie jednostki graniczą ze sobą lub są zlokalizowane w zdefiniowanym położeniu. Zazwyczaj macierz **W** jest standaryzowana według kolumn tak, że dla każdego i $\sum_j w_{ij} = 1$. Elementy macierzy są endogeniczne w odniesieniu do modelu i nie są stochastyczne.

4. Podsumowanie

Wykorzystanie danych przekrojowych i panelowych wiąże się z trudnościami w weryfikacji podstawowych założeń Klasycznej Metody Najmniejszych Kwadratów – estymatora najczęściej wykorzystywanego w szacowaniu parametrów strukturalnych modeli lub będącego podstawą pokrewnych estymatorów, a mianowicie założeń związanych z autokorelacją składnika losowego. Autokorelacja w modelu ekonometrycznym powoduje duże straty w efektywności estymatora, a co za tym idzie – oceny parametrów są niedoszacowanie lub przeszacowane.

W modelach panelowych mamy do czynienia zarówno z problemem autokorelacji skalnika losowego w czasie (zwłaszcza przy dużej liczbie okresów T), jak i z autokorelacją przestrzenną.

W analizie zależności przestrzennych najtrudniejszą kwestią jest skonstruowanie macierzy sąsiedztwa/odległości między obiektami. Jej poprawne skonstruowanie wymaga nie tylko rozpoznania sieci i kierunku powiązań, ale także zasięgu oddziaływania. W badaniach przestrzennych zakładamy, że zależność korelacyjna maleje w miarę wzrostu odległości. Założenie to może być niewystarczające, jeśli będziemy próbowali analizować zjawiska podatne na globalizację, kiedy pojawienie się wybranego zakłócenia (zależności) może być odczuwalne w wybranym regionie świata. Szybkie tempo powiększania się sieci zależności przestrzennych oraz nierównomierna siła oddziaływania wybranych regionów (np. aglomeracji) dodatkowo komplikują zdefiniowanie przestrzennych zależności. Powstające obszary metropolitalne tworzą własną sieć powiązań oddziałującą na znaczne odległości, często wykraczające poza granice pojedynczych jednostek administracyjnych i nieporównywalnych z innymi obszarami.

Przy analizie interakcji przestrzennych mamy więc do czynienia z problemem: określenia sieci powiązań, odległości (sąsiedztwa), kierunku oddziaływania, siły zależności, zmian charakteru relacji w czasie, dostosowania wielkości jednostki (regionu) do charakteru rozpatrywanego problemu.

Dla badań praktycznych szczególnie dotkliwy może być problem dostępności danych według jednostek ekonomicznych (a nie administracyjnych), co umożliwiłoby uwzględnienie rzeczywistych relacji łączących obiekty. Niestety dostępność danych dla małych jednostek przestrzennych jest ograniczona, a wykorzystanie informacji na wyższych poziomach agregacji może powodować rozmywanie się relacji.

Jednym ze sposobów minimalizowania problemu ewentualnej autokorelacji przestrzennej może być podział na klasy obiektów podobnych lub rozbudowa modelu o kluczowe czynniki łącznie z uwzględnieniem ewentualnych opóźnień tych zmiennych.

Literatura

Anselin L., Le Gallo J., Jayet H., *Spatial Panel Econometrics,* [w:] *The Econometrics of Panel Data, Fundamentals and Recent Developments in Theory and Practice*, L. Matyas, P. Sevestre (red.), 3rd ed. Kluwer, Dordrecht 2008.

Arbia G., *Spatial Econometrics,* Springer, Berlin Heidelberg 2006.

Arellano M., Bond S., *Some tests of specification for panel data: monte carlo evidence and an application to employment equation*, „The Review of Econometric Studies Ltd.", vol. 58, no 2, Apr., 1991.

Baltagi B.H., *Econometric Analysis of Panel Data,* Third edition, John Wiley & Sons, Ltd., Chichester 2005.

Bhargava A., Franzini L., Narendranathan W., *Serial correlation and the fixed effects model*, „Review of Economic Studies" 1982, no 49.

Box G.E.P., Jenkins G.M., Reinsel G.C., *Time Series Analysis, Forecasting and Control,* 3rd ed. Prentice Hall, Englewood Cliffs, NJ 1994.

Durbin J., Watson G.S., *Testing for serial correlation in least squares regression, II*, „Biometrika" 1951, no 38.

Durbin J., Watson G.S., *Testing for serial correlation in least squares regression, I*, „Biometrika" 1950, no 37.

Dziechciarz J., *Ekonometryczne modelowanie procesów gospodarczych. Modele ze zmiennymi i losowymi parametrami*, Monografie i Opracowania nr 95, Wrocław 1993.

Elhorst J.P., *Spatial Panel Data Models,* [w:] *Handbook of Applied Spatial Analysis*, M.M. Fischer, A. Getis (wydawcy), 2010, Part 3, 377-407, DOI: 10.1007/978-3-642-03647-7_19.

Elhorst J.P., *Specification and estimation of spatial panel data models*, „International Regional Science Review" 2003, no 26(3).

Elhorst J.P., Piras G., Arbia G., *Growth and convergence in a multi-regional model with space-time dynamics*, Paper presented at the Spatial Econometric Workshop, May 25-27, 2006, Rome.

Kelejian H.H., Prucha I.R., *A generalized moment estimator for the autoregressive parameter in spatial model,* „International Economic Review", vol. 40, no 2, May 1999.

Maddala G.S., *Ekonometria,* Wydawnictwo Naukowe PWN, Warszawa 2006.

Manski C.F., *Identification of endogenous social effects. The reflection problem*, „Review of Economic Studies" 1993, no 60(204).

Osińska M. (red.), *Ekonometria współczesna*, Dom Organizatora, Toruń 2007.

Suchecki B. (red.), *Ekonometria przestrzenna*, C.H. Beck, Warszawa 2010.

Szulc E., *Specyfikacja dynamicznego modelu z przestrzenną strukturą zależności*, „Dynamiczne modele ekonometryczne", IX Ogólnopolskie Seminarium Naukowe, 6-8 września 2005 r., Toruń, http://www.dem.umk.pl/DME/2005/17_szulc.pdf.

Verbeek M., *A Guide to Modern Econometric,* John Wiley & Sons, 2000.

Zellner A., *An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias*, „Journal of American Statistical Association" 1962, no 57.

SELECTED PROBLEMS OF AUTOCORRELATION IN MODELLING BASED ON PANEL DATA

Summary: Growing interest in analyses at regional and local level directs researchers' attention to problems related to information modelling about spatial units. This attention is focused especially on analyses applying panel data. Their popularity results from the possibility of conducting overall analyses and from potential advantages for both quality and scope of performed analyses. The application of cross-section and panel data brings about problems in verifying basic assumptions of Least Square Dummy Variable Model – the estimation most frequently used in estimating structural parameters, or as the basis for related estimations, namely the assumption referring to random component autocorrelation. The article presents selected problems of modelling with panel data and applying autocorrelation and spatial correlation.

Keywords: autocorrelation, spatial autocorrelation, panel data.