

DYNAMICZNE MODELE EKONOMETRYCZNE

VIII Ogólnopolskie Seminarium Naukowe, 9-11 września 2003 w Toruniu
Katedra Ekonometrii i Statystyki, Uniwersytet Mikołaja Kopernika w Toruniu

Elżbieta Szulc

Uniwersytet Mikołaja Kopernika w Toruniu

Odległość ekonomiczna w badaniu struktury zależności w procesach gospodarczych

1. Wprowadzenie

Ekonometrycy badając obiekty ekonomiczne posługują się informacjami bądź to w postaci danych przekrojowych lub szeregów czasowych, bądź to danych przekrojowo – czasowych. Dane te stanowią próby losowe, co do których zakłada się niezależność tworzących je obserwacji.

Klasyczne założenie niezależności pomiędzy obserwacjami w analizach opartych na danych przekrojowych często okazuje się niezgodne z teorią ekonomiczną oraz obserwacją współzależności zjawisk w różnych przekrojach.

Pomijając oczywiste i powszechnie rozważane w ekonometrii współzależności w czasie, należy podkreślić, że jest to sytuacja podobna również do tej obserwowanej w danych przestrzennych, a którą geografowie streścili w postaci tzw. pierwszego prawa geografii: „Wszystkie rzeczy są zależne od wszystkich innych, lecz rzeczy bliskie są bardziej zależne niż odległe”¹.

Na wzór autokorelacji czasowej pojawia się pojęcie autokorelacji przestrzennej² a naturalnym uzupełnieniem powyższych wydaje się pojęcie autokorelacji przekrojowej³.

Obiekty ekonomiczne, takie jak: gospodarstwa domowe, przedsiębiorstwa, gospodarki regionów, krajów itp. działają w taki sposób, że pojawiają się mię-

¹ Patrz, W.Tobler (1970).

² Patrz, A.D.Cliff, J.K.Ord (1973).

³ Takie określenie nie jest powszechnie stosowane. Użyto go po to, aby podkreślić konkretny aspekt analizy. Ogólnie zaś, mówiąc „autokorelacja przestrzenna” można mieć na myśli przestrzeń w znaczeniu szerokim, nie ograniczonym do jej wyrażenia w ujęciu geograficznym.

dzy nimi określone powiązania. Źródłem tych powiązań mogą być bezpośrednio nie obserwowalne zmienne, które są przekrojowo lub przestrzennie skorelowane i przez to wytwarzają korelacje w błędach równań opisujących zachowania ekonomiczne.

Na przykład, popyt gospodarstw domowych na określone dobra może być skorelowany z popytem na takie dobra innych gospodarstw (sąsiadów). Korelacja ta może być związana z przestrzenną korelacją w takich zmiennych, jak np.: dostępność dóbr substytucyjnych, klimat, jakość powietrza, jakość gleby. Ponadto, gospodarstwa domowe mogą czerpać użyteczność z faktu konsumowania dóbr podobnych do tych konsumowanych przez sąsiadów (tzw. naśladownictwo).

Inny przykład, dotyczący jednostek z definicji przestrzennych, to wielkość bezrobocia na obszarze kraju podzielonego na mniejsze jednostki administracyjne. Należy spodziewać się korelacji przestrzennej pomiędzy jednostkami, z uwagi na wpływ na bezrobocie wielu czynników o charakterze ekonomicznym, które decydując o potencjale gospodarczym jednostek, mają następnie wpływ na migrację ludności.

Z kolei, w literaturze wzrostu gospodarczego, badając zależności między tempem wzrostu GNP a czynnikami go opisującymi w przekroju krajów, obserwuje się korelacje między składnikami losowymi dla różnych krajów w odpowiednich modelach regresji.

Założenia niezależności przeważające w ekonometrii przekrojowej ułatwiają estymację i dalsze wnioskowanie, ale mogą niedokładnie odzwierciedlać rzeczywistość. Alternatywą dla niezależności są założenia o naturze zależności pomiędzy obserwacjami. Podejście to wymaga przyjęcia określonych założeń o rozkładach danych i błędów.

Ustalenie postaci zależności w przypadku danych przekrojowych jest trudniejsze niż w analizach szeregów czasowych, gdzie całkowitoliczbowe (lub naturalne) indeksowanie obserwacji nadaje danym naturalny porządek i strukturę.

Celem referatu jest dyskusja na temat możliwości ustalenia struktury danych przekrojowych podobnej do struktury szeregów czasowych i jej użyteczności w modelowaniu ekonometrycznym.

W punkcie 2 referatu wskazuje się na przykładowe miary odległości ekonomicznej. W kolejnych punktach 3 i 4 definiuje się procesy losowe będące przedmiotem analizy oraz procesy zarządzające, które decydują o wyborze jednostek do badania. Punkt 5 sygnalizuje znaczenie kowariancji w badaniu struktury zależności przekrojowych. Uwagi na temat liniowego modelowania ekonometrycznego zależności między procesami w warunkach zależności w danych przekrojowych zawarto w punkcie 6. Referat kończą wnioski z przedstawionej dyskusji.

2. Odległość ekonomiczna jako alternatywa odległości fizycznej

Struktura danych w szeregach czasowych jest prosta, ponieważ tworzą ją ciągi obserwacji dla kolejnych momentów czasu lub okresów, przy czym odległość między obserwacjami jest odległością między punktami na osi czasu.

Strukturę danych przekrojowych tworzy się rozważając odległość między obiektami, która niekoniecznie jest rozumiana jako odległość fizyczna. W wielu sytuacjach użycie odległości fizycznej będzie jednak uzasadnione. Może być tak we wspomnianych wcześniej przykładach.

Korelacja w popycie gospodarstw domowych będzie dotyczyć prawdopodobnie gospodarstw zlokalizowanych blisko siebie w sensie fizycznym. Ale w innym podejściu odległość między gospodarstwami zlokalizowanymi przestrzennie może być wyrażona, na przykład, w kategoriach czasu i kosztów podróży między punktami dla użycia lokalnych dóbr publicznych⁴.

Ze względu na ograniczoną mobilność siły roboczej, autokorelacja w procesie bezrobocia może być rozważana w ujęciu ściśle przestrzennym a to oznacza, że wykorzystywać się będzie odległość w sensie geograficznym. Nie wyklucza to jednak weryfikowania zależności z uwzględnieniem dodatkowej informacji o kondycji ekonomicznej jednostek przestrzennych, konstruując odległość ekonomiczną między nimi na podstawie charakterystyk ekonomicznych.

Wskazuje się na wiele potencjalnych miar ekonomicznej odległości pomiędzy dwoma krajami, jak np. wolumen handlu i koszty transportu kapitału fizycznego lub ludzkiego⁵.

Każda z wymienionych odległości ma swoje uzasadnienie w kontekście konkretnego zastosowania. Nie są to miary ani doskonałe ani uniwersalne. Szczególnie w analizie danych przekrojowych nie wystarcza odwoływanie się tylko do odległości fizycznej.

Pojęcie odległości ekonomicznej jest pewnym uogólnieniem mającym swoje uzasadnienie w świetle stwierdzenia, iż na przebieg zjawisk ekonomicznych mają wpływ czynniki wyrażające się położeniem obiektów w wielowymiarowej przestrzeni cech. Wzajemne położenie obiektów przedstawia się za pomocą odpowiedniej „odległości”. Dane przekrojowe poddawane analizie z uwzględnieniem owej odległości pozwalają określić strukturę „przestrzennych” powiązań, które rodzą korelację w ujęciu przekrojowym.

Mówienie o wykorzystaniu odległości ekonomicznej do konstrukcji struktury zależności między obiektami ekonomicznymi będzie możliwe w terminologii pól losowych.

⁴ Patrz: Conley (1999).

⁵ Tamże.

3. Podstawowy model pola losowego

Zakłada się, że populacja jednostek znajduje się w przestrzeni euklidesowej R^2 . Każda jednostka zlokalizowana jest w określonym punkcie p tej przestrzeni. Populacja potencjalnie obserwowanych lokalizacji tworzy siatkę H , która w ogólnym przypadku jest nieregularna. Z każdą pozycją p w przestrzeni R^2 związana jest zmienna X_p . Zmienna X_p jako funkcja argumentu p nazywa się polem losowym.

Zbiór jednostek i poddawanych modelowaniu ekonometrycznemu tworzy próbę losową, pobraną z populacji. Każda jednostka (obiekt) i zlokalizowana jest w punkcie p_i . Zbiór lokalizacji $\{p_i\}$ tworzy region próby. Można założyć dalej, że każda z osi współrzędnych rozważanej przestrzeni jest całkowitoliczbową i H jest regularna. Obserwacje zmiennej X na siatce H stanowią realizację procesu losowego – pola losowego.

Próba ekonometryczna składa się z dwóch części. Pierwsza część, to wspomniana realizacja pola losowego X_{p_i} we wszystkich punktach p_i regionu próby. Drugą część danych stanowić będzie symetryczna macierz D , której elementy d_{ij} są miarami odległości pomiędzy indeksami p_i, p_j . Przy założeniu, że wyznaczono N lokalizacji p_i , macierz D ma wymiary $N \times N$.

Zależność pomiędzy jednostkami i jest rozważana w odniesieniu do odległości pomiędzy pozycjami jednostek, które to odległości odpowiadają odległościom ekonomicznym. Ta funkcja odległości rozumiana jest następująco. Jeśli dwie lokalizacje p_i oraz p_j obiektów i, j są bliskie, wtedy X_{p_i} oraz X_{p_j} mogą być silnie skorelowane. Natomiast, w miarę jak odległość pomiędzy p_i oraz p_j rośnie, korelacja między X_{p_i}, X_{p_j} staje się coraz mniejsza.

Podwójne indeksowanie wartości X_{p_i} , w odróżnieniu od pojedynczego indeksowania wartości szeregu czasowego, jest potrzebne również po to, by umożliwić uwzględnienie w badaniu wielostronnej natury zależności w danych przekrojowych.

4. Generowanie danych

Przyjmując, że określone są dokładne pomiary odległości ekonomicznej pomiędzy jednostkami, ustalenie położenia obiektów na płaszczyźnie może być dokonane w drodze skalowania wielowymiarowego, które umożliwia wyznaczenie układu obiektów w sytuacji, gdy znana jest kolejność odstępów w macierzy odległości D . Przyjmuje się również, że proces generujący dane opisany jest przez regularną siatkę, zatem X_{p_i} podporządkowane jest polu losowemu X_p z indeksem $p \in Z^2$, gdzie Z oznacza zbiór liczb całkowitych.

W sytuacji, gdy obserwacje pochodzą z modelu nieregularnej siatki⁶, można skonstruować regularną siatkę kwadratową w następujący sposób⁷. Powierzchnię R^2 dzieli się na kwadraty o przekątnej nie większej niż minimalna odległość d_0 pomiędzy jednostkami zlokalizowanymi w R^2 . Ponieważ odległość dowolnego obiektu od każdego innego wynosi przynajmniej d_0 , w każdym kwadracie konstruowanej siatki znajdzie się najwyżej jeden obiekt. Każdy kwadrat może być oznaczony przez dwie całkowitoliczbowe współrzędne p_1, p_2 , tj. $p=[p_1, p_2]$. Normalizuje się wymiary kwadratów na 1×1 i p będzie oznaczać współrzędne lewego dolnego rogu każdego kwadratu.

5. Macierz kowariancji

Podstawą analizy zależności w danych przekrojowych traktowanych w podany wyżej sposób jest odpowiednio zdefiniowana macierz kowariancji. Kowariancja jako funkcja odległości stanowi użyteczne narzędzie badania struktury zależności.

Odnosząc się bezpośrednio do dwóch współrzędnych lokalizacji p , tj. ustalając $p = [p_1, p_2]$ oraz zakładając, że $p_1 = 1, 2, \dots, m$ i $p_2 = 1, 2, \dots, n$ można zaproponować następujący estymator macierzy kowariancji⁸:

$$\hat{C} = \frac{1}{m \times n} \sum_{j=0}^{m-j} \sum_{k=0}^{n-k} \sum_{p_1=j+1}^m \sum_{p_2=k+1}^n (X_{p_1, p_2} X'_{p_1-j, p_2-k} + X_{p_1-j, p_2-k} X'_{p_1, p_2}) - \frac{1}{m \times n} \sum_{p_1=1}^m \sum_{p_2=1}^n X_{p_1, p_2} X'_{p_1, p_2} \quad (1)$$

Macierz obliczona według (1) jest punktem wyjścia do ustalenia struktury autoregresyjnej pola losowego X_p oraz wyboru odpowiedniego modelu⁹.

Badanie zależności w przekroju odległości między obiektami w pojedynczym procesie losowym X_p jest interesujące niosąc określoną interpretację ekonomiczną rozkładu procesu. Nie mniej ciekawe jest badanie korelacji w

⁶ Jest tak często w przypadku, gdy pomiar odległości odbywa się w zbiorze liczb rzeczywistych.

⁷ Patrz: Conley (1999).

⁸ Por. tamże.

⁹ Identyfikacja pola losowego X_p jako autoregresyjnego przebiega, do pewnego stopnia, podobnie jak w przypadku procesu stochastycznego X_t . Należy jednak pamiętać, że pojawiają się dodatkowe problemy, do których należy się odnieść, jak np. wybór pomiędzy modelem jednostronnym a wielostronnym. Pewne próby analizy na ten temat zawarto w: Szulc (2001).

resztach regresji, w której pominięto efekty przynależności obiektu do miejsca na siatce H .

6. Oddziaływania przekrojowe w modelach regresji liniowej

Zakłada się, że proces Y_p zależy od L procesów $X_{(l)p}$ ($l = 1, 2, \dots, L$). Wszystkie procesy określone są w przestrzeni Z^2 , tj. $p = [p_1, p_2]$, $p_1, p_2 \in Z$.

Model regresji liniowej dla powyższych procesów przyjmuje następującą postać:

$$Y_p = \sum_{l=1}^L \alpha_l X_{(l)p} + \eta_p. \quad (2)$$

Jeśli $E(\eta_p \eta_p') \neq \sigma^2 I$, to może wskazywać na potrzebę zmiany struktury modelu i uwzględnienie autoregresji¹⁰.

Zapewnienie odpowiednich własności składnika losowego gwarantuje modelowanie zgodne. Identyfikując Y_p oraz $X_{(l)p}$ jako procesy autoregresyjne, otrzymuje się następujące modele podstawowe¹¹:

$$Y_{p_1, p_2} = \sum_{j=-r}^r \sum_{k=-s}^s \beta_{jk} Y_{p_1-j, p_2-k} + \varepsilon_{y_{p_1, p_2}}, \quad (3)$$

$(j, k) \neq (0, 0)$

$$X_{(l)p_1, p_2} = \sum_{j=-r_l}^{r_l} \sum_{k=-s_l}^{s_l} \gamma_{(l)jk} X_{(l)p_1-j, p_2-k} + \varepsilon_{x_{(l)p_1, p_2}}. \quad (4)$$

$(j, k) \neq (0, 0)$

Zgodny model zależności między rozważanymi procesami otrzymuje się wychodząc od równania postaci:

$$\varepsilon_{y_{p_1, p_2}} = \sum_{l=1}^L \rho_l \varepsilon_{x_{(l)p_1, p_2}} + \varepsilon_{p_1, p_2}. \quad (5)$$

Podstawiając (3) i (4) do (5) oraz porządkując odpowiednie wyrazy, otrzymuje się model:

¹⁰ Patrz, np. modele proponowane w: Case (1991). W cytowanej pracy identyfikuje się przestrzenne szablony oddziaływań przestrzennych między gospodarstwami domowymi. Proponuje się modele uwzględniające te oddziaływania i nadaje im ciekawą interpretację. Wskazuje się na możliwość użycia w takich modelach tzw. socjalnej bliskości między gospodarstwami.

¹¹ Przywołane zostają wielostronne modele autoregresyjne, ponieważ to właśnie one będą raczej regułą niż wyjątkiem w przypadku zależności w przekroju odległości ekonomicznej.

$$\begin{aligned}
 Y_{p_1, p_2} &= \sum_{j=-r, k=-s}^r \sum_{(j,k) \neq (0,0)}^s \beta_{jk} Y_{p_1-j, p_2-k} + \sum_{l=1}^L \rho_l X_{(l)p_1, p_2} + \\
 &+ \sum_{l=1}^L \sum_{j=-r, k=-s}^{r_1} \sum_{(j,k) \neq (0,0)}^{s_1} \gamma_{(l)jk}^* X_{(l)p_1-j, p_2-k} + \varepsilon_{p_1, p_2}
 \end{aligned} \tag{6}$$

gdzie $\gamma_{(l)jk}^* = \rho_l \gamma_{(l)jk}$,

który rozważa się zamiast modelu (2).

Główną zaletą modelu (6) jest przejrzysta interpretacja parametrów strukturalnych. Parametry ρ_l utożsamiają czysty wpływ zjawisk X na Y . Poziom zjawiska Y w jednostce i zależy również od poziomu tego zjawiska w jednostkach „podobnych” (bliskich). Miarą tej zależności są parametry β_{jk} . Ponadto, wyodrębnia się wpływ zmian czynników egzogenicznych X , obserwowanych w „podobnych” jednostkach na zmiany Y w danej jednostce i (parametry $\gamma_{(l)jk}^*$).

7. Wnioski

- 1) Odległość ekonomiczna pomiędzy jednostkami badania jest ważną charakterystyką, którą powinno się uwzględnić w modelowaniu zjawisk i zależności między tymi zjawiskami na podstawie danych przekrojowych.
- 2) Ustalenie struktury powiązań w przekroju odległości jest możliwe, przy założeniu, że badane procesy traktuje się jako pola losowe, tj. funkcje losowe określone na wielowymiarowej, w szczególności dwuwymiarowej, przestrzeni argumentów.
- 3) Odpowiednia specyfikacja powiązań pomiędzy jednostkami pozwala na zbudowanie modelu, w którym zagregowany wpływ egzogenicznych czynników rozbija się na bezpośredni wpływ oraz wpływy pośrednie w tym sensie, że realizowane są najpierw w jednostkach podobnych. Ponadto, uwzględnia się tzw. autozależności w procesie egzogenicznym.

Literatura

- Case, A. (1991), Spatial patterns in household demand, *Econometrica*, 59, s. 953–965.
- Cliff, A.D., Ord, J.K. (1973), *Spatial autocorrelation*, Pion, London.
- Conley, T.G. (1999), GMM estimation with cross sectional dependence, *Journal of Econometrics*, 92, s. 1–45.
- Szulc, E. (2001), Identyfikacja kierunków zależności w procesach ekonomicznych. Przykładowe rozwiązania modelowe, *Dynamiczne modelowanie ekonometryczne*, Materiały na VII Ogólnopolskie Seminarium Naukowe, 4–6 września 2001, UMK, Toruń.
- Tobler, W. (1970), A computer movie simulating urban growth in the Detroit Region, *Economic Geography*, 46, s. 234–240.