ANNALES UNIVERSITATIS MARIAE CURIE-SKŁODOWSKA

LUBLIN-POLONIA

VOL.XXIV,28 SECTIO H 1990

Zakład Nauk Ekonomicznych
Filii UMCS w Rzeszowie

Elżbieta MAKSYMIAK

# Macierze brzegowe w niektórych metodach wykrywania oraz mierzenia natężenia współliniowości

## The Borderline Matrix in Selected Methods of Detecting and Measuring the Intenity of Collineation

Niech $\chi$ oraz $X$ oznaczają zbiór zmiennych $X_1, X_2, ..., X_k$ oraz macierz (o wymiarach n x k), której element $x_{ij}$ jest wartością j - tej zmiennej dla i - tej obserwacji (i=1,2,...,n; j=1,2,...,k), czyli

$$\chi = \{X_1, X_2, ...X_k\}$$

zaś

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1k} \\ x_{21} & x_{22} & \cdots & x_{2k} \\ x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix}$$

W literaturze przedmiotu rozróżnia się dwa rodzaje współliniowości; współliniowość algebraiczną oraz statystyczną.

Definicja 1

Zmienne $X_1, X_2, ...X_k$ są współliniowe algebraicznie wtedy i tylko wtedy, gdy $rz(X) < k$, gdzie $rz(X)$ oznacza rząd macierzy X. Definicja ta jest równoważna stwierdzeniu, że zjawisko współliniowości algebraicznej w zbiorze $\chi$ występuje wtedy i tylko wtedy, gdy

$$\bigvee_{\substack{\beta \in R^k \\ \beta \neq 0}} \quad \bigwedge_{i\in\{1,2,...,n\}} \quad \sum_{j=1}^{k} \beta_j x_{ij} = 0$$

gdzie $\beta_j$ jest j-tą współrzędną wektora $\beta$.

Definicja 2

Zmienne $X_1, X_2, ..., X_k$ są współliniowe statystycznie wtedy i tylko wtedy, gdy co najmniej jedna z nich jest silnie skorelowana z pozostałymi.

Specyficzny charakter zmiennych ekonomicznych sprawia, że w ekonometrii przydatna jest przede wszystkim definicja współliniowości statystycznej. Dane empiryczne najczęściej mają postać szeregów czasowych. Dla takich danych współliniowość jest zjawiskiem typowym, gdyż często mają one tendencję do podobnego kształtowania się w czasie. W przypadku danych przekrojowych występowanie współliniowości tłumaczy się najczęściej tendencję do dość proporcjonalnego zmieniania się obserwowanych wielkości na skutek zmian jednostki obserwacji.

W niniejszym artykule przedstawiamy zastosowanie macierzy brzegowych w wybranych metodach współliniowości, jak metoda Farrara-Glaubera, Schipsa-Stiera oraz Theila.

Niech $R = [r_{ij}]$ $(i, j = 1, 2, ..., k)$ będzie macierzą korelacji dla zmiennych zbioru $\{X_1, X_2, ..., X_k\}$.

D. E. Farrar i R. R. Glauber[1] proponują taką metodę wykrywania współliniowości w której określają miarę $\vartheta_j$ postaci:

$$\vartheta_j = R_{j.1...j-1\,j+1...k} - R_{0.1...k} \quad (j = 1, 2, ..., k), \tag{1}$$

gdzie $R_{j.1...j-1\,j+1...k}$ oznacza współczynnik korelacji wielorakiej między zmienną $X_j$ a pozostałymi zmiennymi zbioru $\chi$, zaś $R_{0.1...k}$ jest współczynnikiem korelacji wielorakiej między zmienną objaśnianą $Y$ a zmiennymi objaśniającymi $X_1, X_2, ...X_k$. . Jeżeli

$$\bigvee_{j\in\{1,2,...,k\}} \; ; \quad \vartheta < 0,$$

[1]D.E. Farrar, R.R Glauber: Multicollineary in Regresion Analysis: The Problem Revised "Review of Economics and Statistics 1967, vol. 49; Theil H.: Principles of Econometries North - Holland Publishing Company, Amsterdam - London 1971.

to w zbiorze $\chi$ nie występuje zjawisko współliniowości. Jeżeli

$$\bigwedge_{j\in\{1,2,\dots,k\}} \vartheta_j \geq,$$

to zmienne zbiory $\chi$ są współliniowe.

Wiadomo, że współczynniki $R_{j.1\dots j-1\,j+1\dots,k} - R_{0.1\dots k}$ wyrażają się odpowiednio wzorami:

$$R_{j.1\dots j-1\,j_1\dots k} = \sqrt{(R_j)^T R_{jj}^{-1} R_j} \; (j = 1,2,\dots,k) \tag{2}$$

$$R_{0.1\dots k} = \sqrt{(R^0)^T R^{-1} R^0}, \tag{3}$$

gdzie $R_{jj}$ jest podmacierzą macierzy $R$ powstałą przez skreślenie j-tego wiersza oraz j-tej kolumny,

$$(R)^T = [r_{j1}\dots r_{jj-1} r_{jj+1}\dots r_{jk}],$$

$$(R^0)^T = [r_{YX_1} r_{YX_2}\dots r_{YX_k}]$$

zaś $r_{YX_j}$ oznacza współczynnik korelacji między zmienną objaśniającą $Y$ i zmienną $X_j$.

Oblicznie współczynnika $R_{j.1\dots j-1\,j+1\dots k}$ oraz $R_{0.1\dots k}$ odpowiednio ze wzorów (2), (3) sprowadza się przede wszystkim do wyznaczenia macierzy odwrotnych $(R_{jj})^{-1}$ i $R^{-1}$.

Aby tego uniknąć współczynniki te można w prosty sposób wyznaczyć korzystając z macierzy brzegowych w oparciu o następujące dwa twierdzenia.

Twierdzenie 1[2]

Jeżeli macierz wewnętrzna A macierzy brzegowej $A_1$ zdefiniowanej następująco

$$A_1 = \begin{bmatrix} A & f \\ g & z \end{bmatrix}_{(k+1)x(k+1)}, \tag{4}$$

gdzie $A = [a_{ij}]$ $(i, j = 1,2,\dots,k)$,
$g = [g_1, g_2, \dots, g_k]$,
$f^T = [f_1, f_2, \dots, f_k]$,
$z \in R$ (R - zbiór liczb rzeczywistych)
jest nieosobliwa, to

$$\frac{\det A_1}{\det A} = z - gA^{-1}f. \tag{5}$$

[2]M. Kolupa: Macierze brzegowe w badaniach ekonometrycznych, PWE Warszawa 1982.

Twierdzenie 2[3]

Jeżeli macierz brzegową $A_1$ daną wzorem (4), której macierz wewnętrzna A jest nieosobliwa, sprowadzamy do górnej macierzy trójkątnej $D$, to

$$\frac{\det A_1}{\det A} = d, \tag{6}$$

gdzie $d$ jest elementem macierzy $D$ stojącym w jej prawym dolnym rogu.

Na mocy twierdzenia 1 oraz wzorów (2) i (3) współczynniki: $R_{j.1\ldots j-1\, j_1 \ldots k}$, $R_{0.1\ldots k}$ wyrażają się wzorami:

$$R_{j.1\ldots j-1\, j_1 \ldots k} = \sqrt{\frac{\det R^{11}}{\det R_{jj}}} \quad (j = 1, 2, \ldots, k), \tag{7}$$

$$R_{0.1\ldots k} = \sqrt{\frac{\det R'}{\det R}}, \tag{8}$$

gdzie

$$R^{11} = \begin{bmatrix} R_{jj} & R_j \\ -(R_j)^T & 0 \end{bmatrix}, \tag{9}$$

$$R' = \begin{bmatrix} R & R^0 \\ -(R^0)^T & 0 \end{bmatrix}. \tag{10}$$

Obliczanie ilorazu $\sqrt{\frac{\det R^{11}}{\det R_{jj}}}$ i $\sqrt{\frac{\det R'}{\det R}}$ wykonujemy zgodnie z twierdzeniem 2. W tym celu należy sprowadzić macierze $R^{11}$, $R'$ do górnych macierzy trójkątnych wykonując odpowiednie przekształcenia elementarne. Wówczas elementy stojące w prawych dolnych rogach tak otrzymanych macierzy są równe wspomnianym wyżej ilorazom, czyli współczynnikowi $R^2_{j.1\ldots j-1\, j_1 \ldots k}$ oraz $R^2_{0.1\ldots k}$.

B. Schips i W. Stier[4] jako miarę natężenia współliniowości proponują współczynnik $\lambda$ określony wzorem

$$\lambda = R^2_{0.1\ldots k} - \sum_{j=1}^{k} r^2_{YX_j}. \qquad 11$$

[3]Ibidem.

[4]B. Scips, W. Stier: Bestimmung der Answirkung von Multikollinearität zwischen den erklärenden Variablen in linearen Regressionsmodellen auf Kleinst - Quadrate Schätzwerte durch Simulation, "Statistische Hefte" 1971, nr 2.

Współczynnik ten przyjmuje wartości z przedziału $< -k+1, 0 >$. Schips i Stier twierdzą, że jeżeli $\lambda = 0$, to zmienne $X_1, X_2, \ldots, X_k$ nie są skorelowane. W przeciwnym wypadku, gdy $\lambda \neq 0$ zmienne $X_1, X_2, \ldots, X_k$ są skorelowane. Natężenie współliniowości w zbiorze $\chi$ jest tym większe, im bardziej $\lambda$ różni się od zera. W powyższej metodzie współcznnik $R_{0.1\ldots k}$ również można obliczyć na podstawie wzoru (8) i twierdzenia 2.

W pracach Gruszczyńskiego[5] i Theila, jako miarę natężenia współliniowości proponuje się współczynnik

$$\vartheta = R^2_{0.1\ldots k} - \sum_{j=1}^{k}(R^2_{0.1\ldots k} - R^2_{0.1\ldots j-1\, j+1\ldots k}).$$

Współczynnik $\vartheta$ podobnie $\lambda$ jest miarą należącą do przedziału $< -k+1, 0 >$ i jego mała wartość świadczy o silnej współliniowości w zbiorze $\chi$. Do obliczania współczynnika $\vartheta$ można również wykorzystać twierdzenie 1 i 2.

Wiadomo, że

$$R_{0.1\ldots j-1\, j+1\ldots k} = (R^0_j)^T (R_{jj})^{-1} R^0_j \quad (j = 1, 2, \ldots, k)$$

stąd na mocy twierdzenia 1 współczynnik $R_{0.1\ldots j-1\, j+1\ldots k}$ wyraża się wzorem

$$R_{0.1\ldots j-1\, j+1\ldots k} = \sqrt{\frac{det R^{11}}{det R_{jj}}},$$

gdzie

$$R'_{jj} = \begin{bmatrix} R_{jj} & R^0_j \\ -(R_j)^T & 0 \end{bmatrix},$$

przy czym $R^0_j$ oznacza wektor postaci

$$[r_{YX_1} \ldots r_{YX_{j-1}}\ [r_{YX_{j+1}} \ldots r_{YX_k}].$$

Z kolei aby wyznaczyć iloraz $\frac{det R_{jj}}{det R_{jj}}$ należy wykorzystać twierdzenie 2 i sprowadzić macierz $R'_{jj}$ (wykonując przekształcenia elementarne) do postaci górnej macierzy trójkątnej. Element stojący w prawym dolnym rogu tak otrzymanej macierzy będzie równy kwadratowi współczynnika $R_{0.1\ldots j-1\, j+1\ldots k}$. Współczynnik $R_{0.1\ldots k}$ można również obliczyć korzystając z wzoru (8) i twierdzenia 2.

[5]H. Gruszczyński: Współliniowość zmiennych i jej wpływ na estymację modeli ekonometrycznych. Praca doktorska, maszynopis, SGPiS, Warszawa 1977.

Poniżej przedstawiamy przykład liczbowy ilustrujący zastosowanie macierzy brzegowych w metodzie Farrara-Glaubera.

Przykład:

Niech $\chi = \{X_1, X_2, X_3\}$ oraz

$$R = \begin{bmatrix} 1 & 0.3 & 0.5 \\ 0.3 & 1 & 0.8 \\ 0.5 & 0.8 & 1 \end{bmatrix}, R^0 = \begin{bmatrix} 0.6 \\ 0.2 \\ 0.3 \end{bmatrix}. \tag{12}$$

Najpierw obliczmy $\vartheta_j$ dla $j = 2$ przy pomocy wzorów (2) i (3) a później stosując macierze brzegowe, tzn. korzystając z twierdzenia 1 i 2. Na mocy (1) mamy równość:

$$\vartheta_2 = R_{2.13} - R_{0.123}$$

Aby wyznaczyć $R_{2.13}$ i $R_{0.123}$ przy pomocy wzorów (2) i (3) należy obliczyć $(R_{22}^{-1}$ i $R^{-1}$. Po odpowiednich przekształceniach otrzymujemy, że:

$$(R_{22}^{-1} = \begin{bmatrix} 4/3 & -2/3 \\ -2/3 & 4/3 \end{bmatrix} \quad R^{-1} = \begin{bmatrix} 18/13 & 5/13 & -1 \\ 5/13 & 75/26 & -5/2 \\ -1 & 5/2 & 91/26 \end{bmatrix}.$$

Wtedy

$$(R_{2.13}^2 = [3/10\ 8/10] \cdot \begin{bmatrix} 4/3 & -2/3 \\ -2/3 & 4/3 \end{bmatrix} \cdot \begin{bmatrix} 3/5 \\ 2/5 \\ 3/10 \end{bmatrix} = 0.36$$

$$(R_{0.123}^2 = [5/3\ 1/5\ 3/10] \cdot \begin{bmatrix} 18/13 & 5/13 & -1 \\ 5/13 & 75/26 & -5/2 \\ -1 & -5/2 & 91/26 \end{bmatrix} \cdot \begin{bmatrix} 3/5 \\ 2/5 \\ 3/10 \end{bmatrix} = 0.36$$

czyli $(\vartheta_2 = \sqrt{0.65} - \sqrt{0.36} \sim 0.8 - 0.6 = 0.2$

Z kolei obliczamy $\vartheta_2$ wykorzystując macierze brzegowe.

Korzystając z wzorów (9), (10) i (12) otrzymujemy macierze $R^{11}$, $R'$ następującej postaci:

$$R^{11} = \begin{bmatrix} 1 & 0.5 & 0.3 \\ 0.5 & 1 & 0.8 \\ -0.3 & -0.8 & 0 \end{bmatrix},$$

$$R' = \begin{bmatrix} 1 & 0.3 & 0.5 & 0.6 \\ 0.3 & 1 & 0.8 & 0.2 \\ 0.5 & 0.8 & 1 & 0.3 \\ -0.6 & -0.2 & -0.3 & 0 \end{bmatrix}$$

Zgodnie z twierdzeniem 2 przekształcamy macierz $R^{11}$ oraz $R'$ do postaci górnej macierzy trójkątnej. W wyniku tych przekształceń otrzymujemy:

$$R^{11} = \begin{bmatrix} 1 & 0.5 & 0.3 \\ 0 & 0.75 & 0.65 \\ 0 & -0.65 & 0.09 \end{bmatrix} \sim \begin{bmatrix} 1 & 0.5 & 0.3 \\ 0 & 0.75 & 0.65 \\ 0 & 0 & 0.65 \end{bmatrix},$$

$$R' = \begin{bmatrix} 1 & 0.3 & 0.5 & 0.6 \\ 0 & 0.91 & 0.65 & 0.02 \\ 0 & 0.65 & 0.75 & 0 \\ 0 & -0.02 & 0 & 0.36 \end{bmatrix} \sim \begin{bmatrix} 1 & 0.3 & 0.5 & 0.6 \\ 0 & 0.91 & 0.65 & 0.02 \\ 0 & 0 & 0.3 & -0.01 \\ 0 & 0 & 0.01 & 0.36 \end{bmatrix} \sim$$

$$\begin{bmatrix} 1 & 0.3 & 0.5 & 0.6 \\ 0 & 0.91 & 0.65 & 0.02 \\ 0 & 0 & 0.3 & -0.01 \\ 0 & 0 & 0 & 0.36 \end{bmatrix},$$

czyli $R_{2.13} = \sqrt{0.65} \sim 0.8$ oraz $R_{0.123} = \sqrt{0.36} \sim 0.6$. Stąd $\vartheta_2 = 0.2$.

Podsumowując należy stwierdzić, że macierze brzegowe wykorzystane w wyżej omawianych metodach badania współliniowości zmiennych objaśniających w znacznym stopniu ułatwiają obliczenie występujących w nich współczynników.

## SUMMARY

The present article is devoted to the application of the borderline matrix in three methods of studying collineation; the methods of Farrar-Glauber, Schips-Stier and Theil. Certain properties of these matrixes make it possible to calculate the values of different coefficient applied in these methods.