

Hanna Dudek

Szkoła Główna Gospodarstwa Wiejskiego w Warszawie

WYKRYWANIE WSPÓŁLINIOWOŚCI ZA POMOCĄ SCENTROWANYCH, NIESCENTROWANYCH ORAZ UOGÓLNIONYCH CZYNNIKÓW INFLACJI WARIANCJI

1. Wstęp

Pojęcie czynnika inflacji wariancji VIF wprowadził w 1970 r. D.W. Marquardt [1970]. Od tego czasu przedmiotem licznych badań teoretycznych i praktycznych są różne rodzaje współczynników VIF.

W pracy opracowano syntetyczny wykaz możliwości zastosowań rozważanych współczynników do diagnozowania współliniowości. Ponadto zbadano, czy wykorzystanie metod doboru zmiennych objaśniających, powszechnie stosowanych w polskiej literaturze ekonometrycznej, eliminuje problem współliniowości. Rozpatrzono metodę eliminacji zmiennych quasi-stałych, metodę badania pojemności nośników informacji oraz metodę analizy grafów.

2. Pojęcie współliniowości w jednorównaniowym liniowym modelu ekonometrycznym

W pracy rozważono jednorównaniowy liniowy model ekonometryczny $Y = X\beta + \varepsilon$. Założono, że liczebność próby jest większa od liczby szacowanych parametrów.

Dokładna współliniowość dotyczy sytuacji, kiedy wektory wartości odpowiadające poszczególnym zmiennym objaśniającym i stałej są liniowo zależne. Macierz zaobserwowanych wartości zmiennych objaśniających X nie ma wtedy pełnego rzędu kolumnowego. W konsekwencji metodą najmniejszych kwadratów nie moż-

na wyznaczyć estymatorów parametrów strukturalnych modelu, ponieważ układ równań normalnych nie ma jednoznacznego rozwiązania.

W przypadku przybliżonej współliniowości kolumny macierzy X są „prawie” liniowo zależne. Macierz X ma pełny rząd kolumnowy, zatem wektor estymatorów parametrów strukturalnych można wyznaczyć ze wzoru $b = (X^T X)^{-1} X^T Y$. Jednakże w takim przypadku mogą wystąpić niekorzystne konsekwencje różnego typu. Do najpoważniejszych można zaliczyć:

- dużą wrażliwość wartości estymatorów parametrów strukturalnych na niewielkie zmiany wartości zmiennych,
- niewłaściwe znaki wartości estymatorów parametrów strukturalnych,
- duże (w stosunku do wartości estymatorów parametrów strukturalnych) standardowe błędy szacunku parametrów.

W celu wykrzycia współliniowości oraz pomiaru niektórych skutków jej wystąpienia konstruuje się różne współczynniki. W pracy tej analizie poddano współczynniki inflacji wariancji.

3. Scentrowany (zwykły) współczynnik inflacji wariancji

Najczęściej w literaturze przedmiotu (por. np. [Greene 2000; *Ekonometria...* 2003]) przez pojęcie współczynnika VIF rozumie się scentrowany czynnik inflacji wariancji zdefiniowany jako:

$$VIF_j = \frac{1}{1 - R_j^2}, \quad (1)$$

gdzie R_j^2 jest współczynnikiem determinacji w modelu liniowym ze stałą, w którym zmienną objaśnianą jest zmienna X_j , a zmiennymi objaśniającymi zmienne $X_1, X_2, \dots, X_{j-1}, X_{j+1}, \dots, X_k$, $j = 1, 2, \dots, k$, k oznacza liczbę zmiennych objaśniających uwzględnionych w modelu.

Z definicji (1) wynikają następujące zależności wyprowadzone m.in. w pracach: [Belsley 1991; Osiewalski 1992; 1988].

$$1. \quad VIF_j = r^{jj}, \quad (2)$$

gdzie r^{jj} jest j -tym elementem diagonalnym macierzy odwrotnej do macierzy R , R oznacza macierz współczynników korelacji między zmiennymi objaśniającymi.

$$2. \quad \frac{1}{\lambda_{\min}} \geq VIF_j \geq 1, \text{ gdzie } \lambda_{\min} - \text{minimalna wartość własna macierzy } R. \quad (3)$$

$VIF_j = 1$ wtedy i tylko wtedy, gdy zmienna X_j nie jest skorelowana z żadną inną zmienną objaśniającą. Osiąganie dużych wartości¹ VIF_j jest równoważne istnieniu

¹ Większość autorów podaje, że wartości $VIF_j > 10$ jest oznaką współliniowości (por. np. [Ekonometria... 2003]), niektórzy zaś przyjmują, że już $VIF_j > 5$ sygnalizuje poważne problemy (por. [Judge i in. 1988]).

przybliżonej liniowej zależności pomiędzy kolumną macierzy X odpowiadającą zmiennej X_j oraz co najmniej jedną z kolumn odpowiadających zmiennym $X_1, X_2, \dots, X_{j-1}, X_{j+1}, \dots, X_k$.

3. Wariancje estymatorów poszczególnych parametrów strukturalnych można wyrazić jako:

$$Var(b_j) = VIF_j \frac{\sigma^2}{\sum_{i=1}^n (x_{ij} - \bar{x}_j)^2}, j = 1, 2, \dots, k, \quad (4)$$

gdzie: σ^2 – wariancja składnika losowego, x_{ij} – i -ta, wartość zmiennej X_j , $i = 1, 2, \dots, n$, $\bar{x}_j$ – wartość średnia zmiennej X_j , $j = 1, 2, \dots, k$.

Z zależności (4) wynika, że skorelowanie zmiennej X_j z pozostałymi zmiennymi objaśniającymi pogarsza precyzję szacunku parametru β_j . Od zależności tej pochodzi nazwa współczynnika VIF (ang. *variance inflation factor*), co może być przetłumaczone jako czynnik inflacji wariancji (por. np. [Ekonometria... 2003]) lub jako współczynnik zwiększenia wariancji (por. [Osiewalski 1988]). Czynnik inflacji wariancji ma jako miernik współliniowości następujące wady (por. [Belsley 1991]).

1. Na podstawie VIF nie można określić liczby zależności łączących zmienne, np. jeśli są dwie silne liniowe zależności, pierwsza obejmująca X_1, X_2 , druga: X_3 i X_4 , to wszystkie wartości VIF_j będą duże, $j = 1, 2, 3, 4$.

2. Nie ma „wartości krytycznej” określającej, jakie wartości VIF wskazują na współliniowość. Zwykle stosowana wartość krytyczna równa 10 jest przyjęta *ad hoc*.

3. Wysokie wartości VIF są koniecznym, ale niewystarczającym warunkiem współliniowości. Jeśli kolumna wartości zmiennej X_j oraz kolumna jedynek odpowiadająca stałej są związane przybliżoną zależnością liniową, to VIF może przyjmować niskie wartości.

Pierwsza i druga wada dotyczy także innych rozważanych w tej pracy rodzajów czynników inflacji wariancji. Trzecia może zaś być wyeliminowana przez zastosowanie niescentrowanych czynników inflacji wariancji.

4. Niescentrowany czynnik inflacji wariancji

W pracy Belsleya [1991] zdefiniowano pojęcie niescentrowanego czynnika inflacji wariancji:

$$NVIF_j = \frac{1}{1 - NR_j^2}, \quad (5)$$

gdzie: NR_j^2 – niescentrowany współczynnik determinacji w modelu bez stałej, w którym zmienną objaśnianą jest X_j , zaś zmiennymi objaśniającymi są $X_0, X_1, X_2, \dots, X_{j-1}, X_{j+1}, \dots, X_k$, $j = 0, 1, 2, \dots, k$, X_0 odpowiada stałej.

$NVIF$ ma następujące własności (por. [Belsley 1991; Osiewalski 1992; 1988]).

$$1. \quad NVIF_j = Nr^{jj}, j = 0, 1, 2, \dots, k, \quad (6)$$

gdzie: Nr^{jj} oznacza j -ty element diagonalnym macierzy odwrotnej do $\mathbf{NR}$,
 $\mathbf{NR}$ jest macierzą niescentrowanych współczynników korelacji.

$$2. \quad \frac{1}{d_{\min}} \geq NVIF_j \geq 1, \text{ gdzie } d_{\min} - \text{minimalna wartość własna macierzy } \mathbf{NR}. \quad (7)$$

$NVIF_j$ osiąga duże wartości wtedy i tylko wtedy, gdy istnieje przybliżona liniowa zależność pomiędzy j -tą kolumną macierzy $\mathbf{X}$ oraz co najmniej jedną z pozostałych kolumn.

$$3. \quad Var(b_j) = NVIF_j \frac{\sigma^2}{\sum_{i=1}^n x_{ij}^2}, j = 0, 1, 2, \dots, k. \quad (8)$$

4. Własności 1-3 są zachowane także w modelu bez stałej.

5. Czynniki inflacji wariancji według Osiewalskiego

W pracach Osiewalskiego [1992; 1988] przedstawiono definicje czynników inflacji wariancji dowolnej liniowej funkcji MNK-estymatorów parametrów strukturalnych modelu oraz czynników inflacji wariancji błędu predykcji w przypadku predyktora MNK.

5.1. Scentrowane i niescentrowane czynniki inflacji wariancji dla liniowej funkcji estymatorów MNK parametrów strukturalnych

Niech $g = \mathbf{c}^T \mathbf{b}$, $\mathbf{c} \neq \mathbf{0}$, wtedy wariancja $Var(g) = \sigma^2 \mathbf{c}^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{c}$.

Scentrowany czynnik inflacji wariancji dla liniowej funkcji MNK-estymatorów parametrów strukturalnych definiuje się następująco (por. [Osiewalski 1988]):

$$VIF(g) = \frac{Var(g)}{Var_l(g)}, \quad (9)$$

gdzie: $Var_l(g)$ – wariancja g , w przypadku braku skorelowania między zmiennymi objaśniającymi, tj. gdy $\mathbf{R} = \mathbf{I}$. Zakłada się, że średnie i odchylenia standardowe zmiennych uwzględnionych w $Var_l(g)$ są takie same jak w $Var(g)$.

Niescentrowany czynnik inflacji wariancji dla liniowej funkcji estymatorów MNK parametrów strukturalnych określa się zaś jako (por. [Osiewalski 1992]):

$$NVIF(g) = \frac{Var(g)}{Var_{NI}(g)}, \quad (10)$$

gdzie: $Var_{NI}(g)$ – wariancja g wówczas, gdy macierz niescentrowanych współczynników korelacji $NR = I$. Zakłada się, że długości wektorów kolumnowych macierzy obserwacji X w $Var_{NI}(g)$ są takie same jak w $Var(g)$.

W pracach Osiewalskiego [1992; 1988] dowiedziono następujących własności czynników inflacji wariancji dla funkcji liniowej MNK-estymatorów parametrów strukturalnych.

1. Jeśli j -ty element wektora c , $c_j=1$, a pozostałe elementy są równe 0, to $VIF(g) = VIF_j$ oraz $NVIF(g) = NVIF_j$.

2. Zakres możliwych wartości $VIF(g)$ dla danej macierzy R i różnych wektorów c określony jest przez nierówności:
$$\frac{1}{\lambda_{\min}} \geq VIF(g) \geq \frac{1}{\lambda_{\max}} > \frac{1}{k}, \quad (11)$$

gdzie $\lambda_{\max}$, $\lambda_{\min}$ – odpowiednio maksymalna i minimalna wartość własna macierzy R .

3. Dla danej macierzy NR i różnych wektorów c :

$$\frac{1}{d_{\min}} \geq NVIF(f) \geq \frac{1}{d_{\max}} > \frac{1}{k+1}, \quad (12)$$

gdzie $d_{\max}$, $d_{\min}$ – odpowiednio maksymalna i minimalna wartość własna macierzy NR .

4. Dla każdej macierzy współczynników korelacji R istnieje wektor c , taki że $VIF(g) < 1$. Jeśli $c^T = [c_0, c_0\bar{x}_1, c_0\bar{x}_2, \dots, c_0\bar{x}_k]$, to $VIF(g) = 1$ dla każdej macierzy R . Skorelowanie zmiennych objaśniających nie ma wtedy żadnego wpływu na wariancję funkcji liniowej g .

5. Dla każdej macierzy NR istnieje taki wektor c , że $NVIF(g) < 1$.

5.2. Scentrowane i niescentrowane czynniki inflacji wariancji dla błędu predykcji

Niech $\hat{y}_* = \mathbf{x}_*^T \mathbf{b}$ przy ustalonym $\mathbf{x}_*$ będzie MNK-predyktorem zmiennej $y_* = \mathbf{x}_*^T \boldsymbol{\beta} + \varepsilon_*$, wówczas błąd predykcji $f = y_* - \hat{y}_*$. Wariancja błędu predykcji wyraża się wzorem: $Var(f) = \sigma^2 [1 + \mathbf{x}_*^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_*]$.

Scentrowany czynnik inflacji wariancji dla błędu predykcji ma postać:

$$VIF(f) = \frac{Var(f)}{Var_1(f)}, \quad (13)$$

gdzie $Var_1(f)$ – wariancja f , w przypadku braku skorelowania między zmiennymi objaśniającymi, tj. gdy $R = I$. Zakłada się, że średnie i odchylenia standardowe zmiennych uwzględnione w $Var(f)$ są takie same jak w $Var(f)$.

Niescentrowany czynnik inflacji wariancji dla błędu predykcji określa się jako:

$$NVIF(f) = \frac{Var(f)}{Var_{NI}(f)}, \quad (14)$$

gdzie $Var_{NR}(f)$ – wariancja f , w przypadku gdy macierz niescentrowanych współczynników korelacji $NR = I$. Zakłada się, długości wektorów kolumnowych macierzy obserwacji X w $Var_{NR}(f)$ są takie same jak w $Var(f)$.

W pracach Osiewalskiego [1992; 1988] dowiedziono następujących własności.

$$1. \text{ Dla danej macierzy } R \text{ i różnych } x_* \text{ zachodzi: } \frac{1}{\lambda_{\min}} \geq VIF(f) \geq \frac{1}{\lambda_{\max}} > \frac{1}{k}, \quad (15)$$

gdzie $\lambda_{\max}$, $\lambda_{\min}$ – odpowiednio maksymalna i minimalna wartość własna macierzy R .

$$2. \text{ Przy ustalonej macierzy } NR \text{ i różnych } x_* \quad \frac{1}{d_{\min}} \geq NVIF(f) \geq \frac{1}{d_{\max}} > \frac{1}{k+1}, \quad (16)$$

gdzie $d_{\max}$, $d_{\min}$ – odpowiednio maksymalna i minimalna wartość własna macierzy NR .

3. Dla każdej macierzy współczynników korelacji R istnieje takie x^* , że $VIF(f) < 1$. Jeśli $x_*^T = [1, \bar{x}_1, \bar{x}_2, \dots, \bar{x}_k]$, to $VIF(f) = 1$ dla każdej macierzy R .

4. Dla każdej macierzy współczynników korelacji NR istnieje takie x^* , że $NVIF(f) < 1$.

6. Uogólniony czynnik inflacji wariancji według Foxa i Monette'a

W niektórych modelach może być wskazany pomiar współliniowości dla zestawów zmiennych objaśniających.

Przez X_1 oznacza się część macierzy X odpowiadającą wyróżnionemu podzbiorowi zmiennych objaśniających, przez X_2 – część macierzy X odpowiadającą pozostałym zmiennym objaśniającym występującym w modelu, X_0 odpowiada zaś stałej w modelu. Model liniowy można wtedy zapisać w jako: $\hat{Y} = X_0 b_0 + X_1 b_1 + X_2 b_2$.

Wyróżnionym podzbiorem zmiennych objaśniających mogą być, np. zmienne zero-jedynkowe, odpowiadające zmiennej jakościowej. Ponieważ wybór zmiennej bazowej² wpływa na wartości współczynników korelacji pomiędzy zmiennymi zero-jedynkowymi a innymi zmiennymi, to od tego wyboru zależą także wartości czynników inflacji wariancji. Podobnie jest, gdy w modelu występuje część wielomianowa, np. gdy zestaw zmiennych objaśniających jest następujący: $X_1, X_1^2, X_1^3, X_2, X_3, \dots, X_k$. Wyróżnionym podzbiorem jest tu $\{X_1, X_1^2, X_1^3\}$. Wartości współczynników korelacji pomiędzy zmiennymi X_1^2, X_1^3 a pozostałymi zmiennymi objaśniającymi są w ogólnym przypadku różne od wartości współczynników korelacji pomiędzy zmiennymi $(X_1 + c)^2, (X_1 + c)^3$ a pozostałymi zmiennymi, gdzie c oznacza stałą. Stąd wartości czynników inflacji wariancji zależą od przyjęcia zmiennej bazowej³.

² Zmienną bazową w modelu ze zmienną jakościową jest pominięta zmienna zero-jedynkowa.

³ Wybór zmiennej bazowej polega tu na wyborze stałej c .

W celu pomiaru współliniowości dla podzbiorów zmiennych objaśniających, Fox i Monette [1992] rozważyli uogólniony czynnik inflacji wariancji zdefiniowany jako

$$GVIF_1 = \frac{\det \mathbf{R}_{11} \det \mathbf{R}_{22}}{\det \mathbf{R}}, \quad (17)$$

gdzie: $\mathbf{R}_{11}$ jest macierzą korelacji pomiędzy zmiennymi z wyróżnionego podzbioru zmiennych objaśniających, $\mathbf{R}_{22}$ – macierz korelacji pomiędzy pozostałymi zmiennymi objaśniającymi występującymi w modelu.

Uogólniony czynnik inflacji wariancji ma następujące własności (por. [Fox, Monette 1992]).

1. Jeśli wyróżniony zbiór $\mathbf{X}_1$ składa się z jednej zmiennej X_j , to $GVIF_1 = VIF_j$. W tym sensie $GVIF$ jest uogólnionym czynnikiem inflacji wariancji (ang. *generalized variance inflation factor*).

2. Wartość $GVIF_1$ nie zależy od sposobu przyjęcia zmiennej bazowej.

3. $GVIF_1 = 1$ odpowiada sytuacji, kiedy zbiór $\mathbf{X}_1$ zawiera zmienne nieskorelowane z pozostałymi zmiennymi występującymi w modelu.

7. Metody doboru zmiennych objaśniających a współliniowość

Kolejnym zagadnieniem tej pracy jest zbadanie, czy zastosowanie metod analizy grafów i badania pojemności nośników informacji eliminuje zawsze wystąpienie współliniowości. Metody te bazują na wartościach współczynników korelacji pomiędzy parami zmiennych. Zastosowanie ich nie wyklucza wyboru zmiennej objaśniającej, której wektor wartości jest związany przybliżoną zależnością z wektorem 1, odpowiadającym stałej w modelu. Dlatego też metoda eliminacji zmiennych quasi-stałych powinna być zawsze stosowana niezależnie od obu metod. Eliminuje ona bowiem zmienne charakteryzujące się małą zmiennością. Następnie zbadano, czy optymalny zestaw zmiennych objaśniających, uzyskanych metodami Bartosiewicz i Hellwiga, może zawierać zmienne związane ze sobą zależnością liniową.

W metodzie analizy grafów wychodzi się z założenia, że zmienne objaśniające w modelu liniowym powinny być słabo skorelowane między sobą oraz silnie skorelowane ze zmiennymi, które nie weszły do modelu. W metodzie tej zależności pomiędzy zmiennymi uważa się za nieistotne, jeśli wartości bezwzględne współczynników korelacji nie przekraczają wartości krytycznej r^* . Wartość krytyczna

określona jest wzorem:
$$r^* = \sqrt{\frac{(t^*)^2}{n-2+(t^*)^2}},$$

gdzie: t^* jest kwantylem rzędu $1 - \frac{\alpha}{2}$ w rozkładzie t-Studenta z $n-2$ stopniami swobody, n – liczebność próby.

Dobór zmiennych objaśniających odbywa się na podstawie analizy grafu, w którym wierzchołkami są poszczególne zmienne, a łuki reprezentują istotne powiązania pomiędzy zmiennymi. W grafie mogą wystąpić:

- zmienne izolowane, to znaczy niewykazujące istotnych powiązań z innymi zmiennymi,
- grupy zmiennych wzajemnie powiązanych.

Zmiennymi objaśniającymi zostają:

- zmienne izolowane,
- z każdej grupy zmiennych wzajemnie powiązanych zmienną reprezentującą staje się zmienna o maksymalnej liczbie powiązań. Jeśli w danej grupie istnieje kilka zmiennych o maksymalnej liczbie powiązań, to należy wybrać spośród nich zmienną najbardziej skorelowaną ze zmienną objaśnianą (por. [Bartosiewicz 1974]).

W metodzie analizie grafów wartość krytyczna współczynnika korelacji zależy od poziomu istotności oraz liczebności próby. Dla ustalonego poziomu istotności liczebność próby jest funkcją malejącą. W przypadku małego n r^* może przyjmować dość wysokie wartości, np. dla poziomu istotności $\alpha = 0,05$, gdy liczebność próby $n \leq 15$, to $r^* > 0,5$. Jeśli wartości bezwzględne współczynników korelacji pomiędzy zmiennymi nie przewyższają wartości krytycznej r^* , to zmienne uznaje się za izolowane. Zmienne te tym samym stają się zmiennymi objaśniającymi. W pracy [Dudek 2003] wykazano, że stosunkowo niskie wartości współczynników korelacji pomiędzy parami zmiennych objaśniających nie wykluczają wystąpienia współliniowości. W pracy tej m.in. podano przykład, w którym $X_3 = X_1 + X_2$ oraz wszystkie $|r_{ij}| = 0,5$, $i, j = 1, 2, 3$ $n = 5$. Zmienne X_1, X_2, X_3 są zmiennymi izolowanymi, czyli wchodzi do modelu. Metodą analizy grafów można zatem czasem dobrać zmienne objaśniające z liniowo zależnymi wektorami wartości.

W metodzie badania pojemności nośników informacji każdą zmienną objaśniającą traktuje się jak nośnik informacji o zmiennej objaśnianej. W metodzie tej rozpatruje się wszystkie niepuste kombinacje potencjalnych zmiennych objaśniających. Dla każdej kombinacji oblicza się współczynnik, zwany pojemnością integralną nośników informacji. Za optymalną kombinację uznaje się taki zestaw zmiennych, dla którego współczynnik ten osiąga wartość maksymalną (por. [Hellwig 1969]). Generalnie pojemności integralne nośników informacji przyjmują tym większe wartości, im zmienne objaśniające są silniej skorelowane ze zmienną objaśnianą oraz słabiej skorelowane między sobą. W pracy Guzika [1985] przedstawiono jednakże przykład, w którym pojemność integralna nośników informacji osiąga największą wartość dla zestawu zawierającego trzy zmienne, spośród których dwie są doskonale skorelowane. Dobór zmiennych objaśniających metodą Hellwiga nie zawsze zatem chroni przed problemem współliniowości.

8. Uwagi końcowe

Czynniki inflacji wariancji z pewnością nie mogą zastąpić procedury pomiaru współliniowości opartej na liczbie warunkowej. Jednakże ze względu na ich prostotę obliczeniową oraz interpretację warto jest je wyznaczyć. Powszechnie bowiem stosowane metody doboru zmiennych objaśniających nie zawsze wykluczają współliniowość. Odpowiednio duże wartości czynników inflacji wariancji powinny być sygnałem do wnikliwej analizy rozpatrywanych danych empirycznych.

Literatura

- Bartosiewicz S., *Prosta metoda wyboru zmiennych objaśniających w modelu ekonometrycznym*, Prace Naukowe WSE we Wrocławiu, nr 43 (65), Wrocław 1974, s. 93-101.
- Belsley D.A., *Conditioning Diagnostics, Collinearity and Weak Data in Regression*, John Wiley&Sons, New York 1991.
- Dudek H., *Wpływ współliniowości na wartości współczynników korelacji pomiędzy parami zmiennych objaśniających*, „Przegląd Statystyczny” 2003, t. 50, z. 2, s. 41-51.
- Ekonometria*, red. M. Gruszczyński, M. Podgórska, Oficyna Wydawnicza SGH, Warszawa 2003.
- Fox J., Monette G., *Generalized Collinearity Diagnostics*, „Journal of the American Statistical Association” 1992, vol. 87, nr 417, s. 178-183.
- Greene W.H., *Econometric Analysis*, Prentice Hall, Inc., New Jersey 2000.
- Guzik B., *Metoda Hellwiga w warunkach współliniowości par zmiennych objaśniających*, „Przegląd Statystyczny” 1985, t. 32, z. 1, s. 33-39.
- Hellwig Z., *Problem optymalnego wyboru predyktant*, Przegląd Statystyczny 1969, t. 16, z. 3-4, s. 221-237.
- Judge G.G., Hill C., Griffiths W.E., Lütkepohl H., Lee T., *Introduction to the Theory and Practice of Econometrics*, John Wiley&Sons, New York 1988.
- Marquardt D.W., *Generalized Inverses, Ridge Regression, Biased Linear Estimation, and Nonlinear Regression*, „Technometrics” 1970, vol. 12, s. 591-613.
- Osiewalski J., *Centered and Noncentered Variance Inflation Factors for the OLS Estimator of a Linear Function and for the OLS Prediction Error*, Acta Universitatis Lodzianensis, Folia Oeconomica, nr 123, Łódź 1992, s. 97-108.
- Osiewalski J., *Współczynniki zwiększenia wariancji estymatora MNK liniowej funkcji parametrów oraz błędu predykcji*, „Przegląd Statystyczny” 1988, t. 36, z. 4, s. 391-399.

DETECTING OF COLLINEARITY BY USING CENTERED, NONCENTERED AND GENERALIZED VARIANCE INFLATION FACTORS

Summary

In this article we analyze the concepts of variance inflation factors *VIF*. Centered *VIF* is misleading in the case of small variation of one of the regressors. In such cases noncentered *VIF* should be applied. In order to determine collinearity for a set of related regressors, such as dummy regressors corresponding to a categorical variable, generalized inflation factor is applied.

Moreover explanatory variables selection methods are analyzed in the context of collinearity. We demonstrate that by using Hellwig's method and Bartosiewicz's method it is possible to choose variables with linearly dependent vectors of observations.