Wiadomości Statystyczne. The Polish Statistician, 2020, vol. 65, 6, 69–87 DOI: 10.5604/01.3001.0014.2347
Edukacja statystyczna / Statistical education

Wyznaczanie współczynników korelacji liniowej – podstawy

Zbigniew Śleszyński[a]

Streszczenie. Celem artykułu jest przedstawienie podstawowych miar współzależności cech ilościowych, które wykorzystuje się w modelowaniu ekonometrycznym, oraz ich wybranych zastosowań. Omówiono współczynniki: korelacji liniowej Pearsona, korelacji wielowymiarowej, determinacji, korelacji cząstkowej i semikorelacji cząstkowej. Zaprezentowane w artykule podejście do wymienionych miar jest jednorodne. Każdą zdefiniowano jako współczynnik korelacji liniowej odpowiednich wektorów otrzymanych na podstawie równań regresji. Podano też wzajemne zależności między współczynnikami. W obliczeniach wykorzystano macierze brzegowe, co znacznie ułatwiło ten proces. W celu sprawdzenia poprawności obliczeń posłużono się programem Statistica 13.3 PL. Zagadnienie zilustrowano na przykładzie modelu regresji wzrostu płac w Polsce w latach 2001–2019 zawierającego cztery zmienne objaśniające, szacowanego metodą najmniejszych kwadratów.

Słowa kluczowe: współczynnik korelacji liniowej, współczynnik korelacji wielowymiarowej, współczynnik determinacji, współczynnik korelacji cząstkowej, współczynnik semikorelacji cząstkowej, macierze brzegowe

JEL: C10, C20, C40, C50

The basics of determining the coefficients of a linear correlation

Abstract. The aim of the paper is to present the basic measures related to the analysis of relationships between quantitative variables used in econometric modelling and their selected applications. The following measures are discussed: the Pearson correlation coefficient, the multivariate correlation coefficient, coefficient of determination, partial correlation coefficient and semi-partial correlation coefficient. A homogeneous approach is applied to the measures presented. Each is defined as a linear correlation coefficient of relevant vectors derived from regression equations. Additionally, mutual relations between the coefficients are described. Bordered matrices have been applied to the calculations, which significantly simplified the process, while the Statistica 13.3 PL program was used to verify the correctness of the calculations. The issue is illustrated in the model of regression of salary growth in Poland in the years 2001–2019 with four covariates, estimated using the least squares method.

Keywords: correlation coefficient, multivariate correlation coefficient, coefficient of determination, partial correlation coefficient, semi-partial correlation coefficient, bordered matrices

[a] Uniwersytet Technologiczno-Humanistyczny im. Kazimierza Pułaskiego w Radomiu, Wydział Ekonomii i Finansów. ORCID: https://orcid.org/0000-0002-8388-351X.

1. Wprowadzenie

Podstawową miarą współzależności liniowej cech ilościowych jest współczynnik korelacji liniowej Pearsona. W badaniach ekonometrycznych wykorzystuje się go m.in. do doboru zmiennych, sprawdzania przydatności zmiennych objaśniających w modelu, badania współliniowości, koincydencji i jakości modelu. Jego różne zastosowania wynikają z obliczania współczynników korelacji różnych wektorów.

Celem artykułu jest przedstawienie podstawowych miar współzależności cech ilościowych wykorzystywanych w modelowaniu ekonometrycznym oraz ich wybranych zastosowań. Skoncentrowano się na zdefiniowaniu tych miar jako współczynników korelacji liniowej różnych wektorów, danych lub otrzymanych w trakcie modelowania. Podano wzory macierzowe służące do ich wyznaczania. Takie podejście wydaje się uzasadnione, gdyż ułatwia interpretację poszczególnych współczynników i ich wzajemnych zależności. Jednocześnie różni się od powszechnie stosowanego w większości opracowań (np. Dziechciarz, 2015; Ezekiel i Fox, 1959; Zeliaś, 2000). W szczególności dotyczy to współczynników korelacji cząstkowej. W literaturze najczęściej podawane są wzory analityczne jedynie dla modelu regresji liniowej z dwiema lub trzema zmiennymi niezależnymi albo wzory przedstawiające dany współczynnik korelacji jako iloraz wyznaczników pewnych macierzy. Ponadto w polskiej literaturze współczynnik semikorelacji cząstkowej jest w zasadzie pomijany. Przedstawienie mało znanych współczynników korelacji cząstkowej i semikorelacji cząstkowej nadaje pracy dodatkowy walor edukacyjny. Wykorzystanie macierzy brzegowych do obliczeń wymienionych współczynników ułatwia proces obliczeniowy i pozwala zauważyć wzajemne zależności.

2. Współczynnik korelacji liniowej

W celu zbadania współzależności liniowej dwóch cech ilościowych X oraz Y zwykle obliczamy współczynnik korelacji liniowej. Dla n wartości tych cech (odpowiednio x_t oraz y_t, $t = 1, 2, \dots, n$) współczynnik ten będzie oznaczany $r(X, Y)$ i obliczany zgodnie ze wzorem:

$$r(X,Y) = \frac{\sum_{t=1}^{n}(x_t - \overline{x})\,(y_t - \overline{y})}{\sqrt{\sum_{t=1}^{n}(x_t - \overline{x})^2 \sum_{t=1}^{n}(y_t - \overline{y})^2}} \tag{1}$$

gdzie $\overline{x}$ oraz $\overline{y}$ – odpowiednio średnie arytmetyczne wartości cechy X oraz Y.

Jak wiadomo, współczynnik korelacji liniowej jest symetryczny, przyjmuje wartości z przedziału [−1, 1], jego bezwzględna wartość świadczy o sile związku liniowego, a znak – o kierunku tego związku.

W przypadku zależności nieliniowych należy posługiwać się współczynnikiem korelacji nieliniowej $R(Y,X)$ zmiennej Y względem zmiennej X, zgodnie ze wzorem (Zeliaś, 2000):

$$R(Y,X) = \sqrt{1 - \frac{\sum_{t=1}^{n}(y_t - \hat{y}_t)^2}{\sum_{t=1}^{n}(y_t - \overline{y})^2}} \tag{2}$$

gdzie $\hat{y}_t$ – wartość teoretyczna zmiennej Y w okresie t oszacowana na podstawie odpowiedniego modelu.

Warto nadmienić, że współczynnik korelacji nieliniowej oraz współczynnik korelacji wielowymiarowej są pierwiastkami ze współczynnika determinacji (na ogół oznaczanego R^2) w odpowiednim modelu regresyjnym. Model liniowy można traktować jako szczególny przypadek modelu nieliniowego lub modelu liniowego wielowymiarowego (regresja wielokrotna).

Współczynnik korelacji nieliniowej nie jest symetryczny, przyjmuje wartości z przedziału [0, 1], a im bardziej jego wartość zbliża się do 1, tym związek korelacyjny między rozpatrywanymi zmiennymi jest mocniejszy. W celu zilustrowania tych zależności obliczono współczynnik korelacji liniowej $r(X,Y)$ dla danych pięciu wartości cech X i Y. Niezbędne obliczenia przedstawiono w tabl. 1.

Tabl. 1. Wyznaczanie współczynnika korelacji linowej cech X oraz Y

t	x_t	y_t	$y_t - \bar{y}$	$(x_t - \bar{x})(y_t - \bar{y})$	$(x_t - \bar{x})^2$	$(y_t - \bar{y})^2$
1	–2	4	2	–4	4	4
2	–1	1	–1	1	1	1
3	0	0	–2	0	0	4
4	1	1	–1	–1	1	1
5	2	4	2	4	4	4
Σ	0	10	0	0	10	14

Źródło: obliczenia własne na podstawie danych umownych.

Stąd na podstawie (1) otrzymujemy:

$$r(X,Y) = \frac{\sum_{t=1}^{5}(x_t - \overline{x})(y_t - \overline{y})}{\sqrt{\sum_{t=1}^{5}(x_t - \overline{x})^2 \sum_{t=1}^{5}(y_t - \overline{y})^2}} = \frac{0}{\sqrt{10 \cdot 14}} = 0 \tag{3}$$

Można zauważyć, że wartości zmiennej Y są kwadratami wartości zmiennej X. Współczynnik korelacji nieliniowej obliczony na podstawie wzoru (2) jest równy 1.

Inne współczynniki wykorzystywane w modelowaniu ekonometrycznym: korelacji wielowymiarowej, korelacji cząstkowej i semikorelacji cząstkowej – które określono jako współczynniki korelacji liniowej odpowiednich wektorów – mają takie same zalety i wady co omówiony współczynnik korelacji liniowej. Należy pamiętać, że są one miarą zależności liniowej cech ilościowych, więc ich niska wartość bezwzględna nie musi świadczyć o braku innej zależności (co pokazały zaprezentowane wcześniej obliczenia). Niewzięcie tego pod uwagę jest najczęstszym błędem przy interpretacji wartości tych współczynników.

3. Współczynnik korelacji wielowymiarowej

W dalszej części artykułu przyjęto, że dany jest model postaci:

$$Y = \alpha_1 Z_1 + \alpha_2 Z_2 + \cdots + \alpha_k Z_k + \xi \tag{4}$$

Załóżmy, że mamy n obserwacji wartości zmiennych i że zmienne występujące w modelu (4) są standaryzowane (wartości średnie obserwacji są równe 0, a odchylenia standardowe wynoszą 1). Otrzymujemy zatem:

$$\mathbf{Z} = \left[z_{tj}\right]_{n\times k} \quad \text{oraz} \quad \boldsymbol{y} = [y_t]_{n\times 1} \tag{5}$$

Oznaczmy przez $(\mathbf{R}(k)\ \boldsymbol{R}_0(k))$ parę korelacyjną określającą model (4). Jak wiadomo (zob. np. Kolupa i Śleszyński, 2010):

$$\begin{gathered} \mathbf{R}(k) = \frac{1}{n}\mathbf{Z}^T\mathbf{Z} \\ \boldsymbol{R}_0(k) = \frac{1}{n}\mathbf{Z}^T\boldsymbol{y} \end{gathered} \tag{6}$$

Jeśli oznaczymy:

$$\boldsymbol{A}^T = [a_1\ a_2 \ldots a_{\mathrm{k}}]_{k\times 1} \tag{7}$$

wektor oszacowań parametrów strukturalnych modelu (4) uzyskany za pomocą metody najmniejszych kwadratów, to:

$$\boldsymbol{A} = (\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T\boldsymbol{y} = \mathbf{R}^{-1}\boldsymbol{R}_0 \tag{8}$$

Wektor wartości teoretycznych zmiennej objaśnianej Y ma postać:

$$\boldsymbol{y}^* = \mathbf{Z} \cdot \boldsymbol{A} = \mathbf{Z}(\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T\boldsymbol{y} = \mathbf{Z} \cdot \mathbf{R}^{-1}\boldsymbol{R}_0 \tag{9}$$

Z kolei wektor reszt $\boldsymbol{u}$ wyznaczamy zgodnie ze wzorem:

$$\boldsymbol{u} = \boldsymbol{y} - \boldsymbol{y}^* = \boldsymbol{y} - \mathbf{Z} \cdot \boldsymbol{A} = \boldsymbol{y} - \mathbf{Z}(\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T\boldsymbol{y} = \boldsymbol{y} - \mathbf{Z} \cdot \mathbf{R}^{-1}\boldsymbol{R}_0 \tag{10}$$

Oznaczmy przez R współczynnik korelacji liniowej wyznaczony dla wektora wartości empirycznych $\boldsymbol{y}$ oraz wektora wartości teoretycznych $\boldsymbol{y}^*$ (Charemza i Deadman, 1997). Jest to współczynnik korelacji wielowymiarowej (niekiedy nazywany współczynnikiem korelacji wielokrotnej lub wielorakiej). Przyjmuje on wartości z przedziału [0, 1] (Theil, 1979) i określa siłę oddziaływania zmiennych objaśniających modelu (4) na jego zmienną objaśnianą. Warto dodać, że współczynnik korelacji nieliniowej $R(Y, X)$ dany wzorem (2), obliczony dla modelu (4), jest współczynnikiem korelacji wielowymiarowej dla tego modelu. Ponieważ dalsze rozważania dotyczą modelu (4), dla tych dwóch współczynników przyjęto identyczne oznaczenia.

Kwadrat współczynnika korelacji wielowymiarowej nazywamy współczynnikiem determinacji. Wyraża on procent zmienności zmiennej zależnej wytłumaczony przez model. Dla modeli szacowanych na podstawie danych w postaci szeregów czasowych przyjmuje zwykle wysoką wartość (większą od 0,9), natomiast w przypadku równań szacowanych przy wykorzystaniu danych przekrojowych często akceptowane są modele o znacznie niższej wartości współczynnika determinacji (większej od 0,3) (Gajda, 2004).

Mając parę korelacyjną, współczynnik determinacji można wyznaczyć zgodnie ze wzorem:

$$R^2 = \boldsymbol{R}_0^T\mathbf{R}^{-1}\boldsymbol{R}_0 \tag{11}$$

W celu wyznaczenia wartości R^2 można posłużyć się macierzą brzegową $\mathbf{K}$ postaci:

$$\mathbf{K} = \begin{bmatrix} \mathbf{R} & \boldsymbol{R}_0 \\ -\boldsymbol{R}_0^T & 0 \end{bmatrix} \tag{12}$$

Jeśli na macierzy $\mathbf{K}$ danej wzorem (12) dokonamy przekształceń elementarnych, w wyniku których w miejscu macierzy wewnętrznej $\mathbf{R}$ otrzymamy macierz górną trójkątną z jedynkami na głównej przekątnej $\mathbf{R}^*$, a w miejscu wektora $-\boldsymbol{R}_0^T$ otrzymamy wektor zerowy, to w miejscu liczby 0 otrzymamy wartość współczynnika determinacji R^2. W efekcie macierz $\mathbf{K}$ zostanie przekształcona w macierz $\mathbf{K}^*$ postaci:

$$\mathbf{K}^* = \begin{bmatrix} \mathbf{R}^* & \boldsymbol{R}_0^* \\ 0, \dots, 0 & R^2 \end{bmatrix} \tag{13}$$

Liczbę $1 - R^2$ nazywamy współczynnikiem zbieżności (niekiedy tolerancją) i oznaczamy przez φ^2. Określa on, jaki procent zmienności zmiennej zależnej nie jest wyjaśniony przez zmienne objaśniające modelu, zgodnie ze wzorem:

$$\varphi^2 = 1 - \boldsymbol{R}_0^T \mathbf{R}^{-1} \boldsymbol{R}_0 \tag{14}$$

Jeśli mając daną macierz **P** postaci:

$$\mathbf{P} = \begin{bmatrix} \mathbf{R} & \boldsymbol{R}_0 \\ \boldsymbol{R}_0^T & 1 \end{bmatrix} \tag{15}$$

dokonamy przekształceń elementarnych, w wyniku których w miejscu macierzy wewnętrznej **R** otrzymamy macierz górną trójkątną z jedynkami na głównej przekątnej $\mathbf{R}^*$, a w miejscu wektora $\boldsymbol{R}_0^T$ otrzymamy wektor zerowy, to w miejscu liczby 1 otrzymamy wartość współczynnika zbieżności φ^2. W efekcie macierz **P** zostanie przekształcona w macierz $\mathbf{P}^*$ postaci:

$$\mathbf{P}^* = \begin{bmatrix} \mathbf{R}^* & \boldsymbol{R}_0^* \\ 0, \dots, 0 & \varphi^2 \end{bmatrix} \tag{16}$$

4. Współczynnik korelacji cząstkowej

Aby wprowadzić definicję współczynnika korelacji cząstkowej r_i^* zmiennej objaśnianej Y ze zmienną objaśniającą Z_i w modelu (4), nazwijmy model (4) modelem pierwotnym, a model powstały z niego przez usunięcie i-tej zmiennej objaśniającej – modelem uciętym. Jeśli model pierwotny jest określony przez parę korelacyjną $(\mathbf{R}(k)\ \boldsymbol{R}_0(k))$ daną wzorem (6), to model ucięty jest określony przez parę $(\mathbf{R}_{ii}, \boldsymbol{R}_{0i})$, gdzie macierz $\mathbf{R}_{ii}$ powstaje z macierzy $\mathbf{R}(k)$ przez skreślenie wiersza i kolumny o numerze i, a wektor $\boldsymbol{R}_{0i}$ powstaje z wektora $\boldsymbol{R}_0(k)$ przez skreślenie i-tej składowej.

Oznaczmy jako $\boldsymbol{u}_i$ wektor reszt dla modelu uciętego. Przez analogię do wzoru (10) mamy:

$$\boldsymbol{u}_i = \boldsymbol{y} - \mathbf{Z}(-i)\mathbf{R}_{ii}^{-1}\boldsymbol{R}_{0i} \tag{17}$$

gdzie macierz $\mathbf{Z}(-i)$ powstaje z macierzy obserwacji $\mathbf{Z}$ danej wzorem (5) przez odrzucenie i-tej kolumny.

Modelem wewnętrznym dla modelu pierwotnego (6) nazywamy model, w którym zmienną objaśnianą jest i-ta zmienna objaśniająca modelu pierwotnego, natomiast zbiór zmiennych objaśniających jest identyczny jak w modelu uciętym. Model wewnętrzny określa para korelacyjna $(\mathbf{R}_{ii}, \boldsymbol{\rho}_i)$, gdzie $\boldsymbol{\rho}_i$ oznacza i-tą kolumnę macierzy $\mathbf{R}(k)$ bez i-tego elementu. Oznaczmy jako $\boldsymbol{v}_i$ wektor reszt dla modelu wewnętrznego. Przez analogię do (17) otrzymujemy:

$$\boldsymbol{v}_i = \boldsymbol{z}_i - \mathbf{Z}(-i)\mathbf{R}_{ii}^{-1}\boldsymbol{\rho}_i \tag{18}$$

gdzie $\boldsymbol{z}_i$ – i-ta kolumna macierzy $\mathbf{Z}$ danej wzorem (5).

Współczynnikiem korelacji cząstkowej r_i^* zmiennej objaśnianej Y ze zmienną objaśniającą Z_i w modelu (4) nazywamy współczynnik korelacji liniowej pomiędzy wektorami reszt modeli wewnętrznego i uciętego, a zatem:

$$r_i^* = r(\boldsymbol{u}_i, \boldsymbol{v}_i) \tag{19}$$

Stąd:

$$r_i^* = \frac{\boldsymbol{u}_i^T \boldsymbol{v}_i}{\sqrt{\boldsymbol{u}_i^T \boldsymbol{u}_i}\sqrt{\boldsymbol{v}_i^T \boldsymbol{v}_i}} \tag{20}$$

Ponieważ spełnione są równości (Kolupa, Pacholewicz i Śleszyński, 1992):

$$\begin{aligned} \boldsymbol{u}_i^T \boldsymbol{v}_i &= n\left(r_i - \boldsymbol{\rho}_i^T \mathbf{R}_{ii}^{-1} \boldsymbol{R}_{0i}\right) \\ \boldsymbol{u}_i^T \boldsymbol{u}_i &= n\left(1 - \boldsymbol{R}_{0i}^T \mathbf{R}_{ii}^{-1} \boldsymbol{R}_{0i}\right) \\ \boldsymbol{v}_i^T \boldsymbol{v}_i &= n\left(1 - \boldsymbol{\rho}_i^T \mathbf{R}_{ii}^{-1} \boldsymbol{\rho}_i\right) \end{aligned} \tag{21}$$

gdzie:

r_i – współczynnik korelacji liniowej zmiennej objaśnianej Y ze zmienną objaśniającą Z_i,

n – liczba obserwacji,

to wzór (20) przyjmuje postać:

$$r_i^* = \frac{r_i - \boldsymbol{\rho}_i^T \mathbf{R}_{ii}^{-1} \boldsymbol{R}_{0i}}{\sqrt{(1 - \boldsymbol{R}_{0i}^T \mathbf{R}_{ii}^{-1} \boldsymbol{R}_{0i})(1 - \boldsymbol{\rho}_i^T \mathbf{R}_{ii}^{-1} \boldsymbol{\rho}_i)}} \tag{22}$$

Wartości współczynnika korelacji cząstkowej zawierają się w przedziale [−1, 1]. Jest on miarą korelacji pomiędzy daną zmienną objaśniającą Z_i, z uwzględnieniem jej skorelowania z pozostałymi zmiennymi objaśniającymi modelu, a zmienną objaśnianą Y, z uwzględnieniem jej skorelowania z pozostałymi zmiennymi objaśniającymi.

Kwadrat współczynnika korelacji cząstkowej nazywamy współczynnikiem determinacji cząstkowej.

Zauważmy, że jeśli we wzorze (6) przyjmiemy $k = 2$, to otrzymamy wyniki analogiczne do podanych przez Zeliasia (2000). Mamy bowiem:

$$\mathbf{R}(2) = \begin{bmatrix} 1 & r_{12} \\ r_{12} & 1 \end{bmatrix}, \boldsymbol{R}_0(2) \begin{bmatrix} r_1 \\ r_2 \end{bmatrix} \tag{23}$$

Wtedy przykładowo dla $i = 1$:

$$\boldsymbol{\rho}_1 = [r_{12}], \mathbf{R}_{11} = [1], \boldsymbol{R}_{01} = [r_2] \tag{24}$$

$$r_i^* = \frac{r_1 - \boldsymbol{\rho}_1^T \mathbf{R}_{11}^{-1} \boldsymbol{R}_{01}}{\sqrt{(1 - \boldsymbol{R}_{01}^T \mathbf{R}_{11}^{-1} \boldsymbol{R}_{01})(1 - \boldsymbol{\rho}_1^T \mathbf{R}_{11}^{-1} \boldsymbol{\rho}_1)}} = \frac{r_1 - r_{12} r_2}{\sqrt{(1 - r_2^2)(1 - r_{12}^2)}} \tag{25}$$

Z kolei dla $i = 2$ otrzymujemy:

$$r_2^* = \frac{r_2 - \boldsymbol{\rho}_2^T \mathbf{R}_{22}^{-1} \boldsymbol{R}_{02}}{\sqrt{(1 - \boldsymbol{R}_{02}^T \mathbf{R}_{22}^{-1} \boldsymbol{R}_{02})(1 - \boldsymbol{\rho}_2^T \mathbf{R}_{22}^{-1} \boldsymbol{\rho}_2)}} = \frac{r_2 - r_{12} r_1}{\sqrt{(1 - r_1^2)(1 - r_{12}^2)}} \tag{26}$$

Warto nadmienić, że aby wyznaczyć współczynnik korelacji cząstkowej r_i^* zgodnie ze wzorem (22), wygodnie jest posłużyć się podwójną macierzą brzegową Δ_i postaci:

$$\boldsymbol{\Delta}_i = \begin{bmatrix} \mathbf{R}_{ii} & \boldsymbol{\rho}_i & \boldsymbol{R}_{0i} \\ \boldsymbol{\rho}_i^T & 1 & r_i \\ \boldsymbol{R}_{0i}^T & r_i & 1 \end{bmatrix} \tag{27}$$

Jeśli na elementach macierzy $\boldsymbol{\Delta}_i$ wykonamy przekształcenia elementarne, otrzymamy macierz $(\boldsymbol{\Delta}_i)^*$, w której w miejsce macierzy wewnętrznej $\mathbf{R}_{ii}$ będziemy mieć macierz górną trójkątną z jedynkami na głównej przekątnej (oznaczymy ją $(\mathbf{R}_{ii})^*$, natomiast w miejsce wektorów $\boldsymbol{\rho}_i^T$ oraz $\boldsymbol{R}_{0i}^T$ – wektory zerowe, czyli:

$$\mathbf{\Delta}_i = \begin{bmatrix} \mathbf{R}_{ii} & \boldsymbol{\rho}_i & \boldsymbol{R}_{0i} \\ \boldsymbol{\rho}_i^T & 1 & r_i \\ \boldsymbol{R}_{0i}^T & r_i & 1 \end{bmatrix} \sim (\mathbf{\Delta}_i)^* = \begin{bmatrix} (\mathbf{R}_{ii})^* & (\boldsymbol{\rho}_i)^* & (\boldsymbol{R}_{0i})^* \\ 0, \dots, 0 & a_i & b_i \\ 0, \dots, 0 & b_i & c_i \end{bmatrix} \tag{28}$$

Wówczas na podstawie (22) otrzymamy:

$$r_i^* = \frac{b_i}{\sqrt{a_i \cdot c_i}} \tag{29}$$

Warto zwrócić uwagę, że wykorzystując macierz (28), możemy dla dowolnego $i = 1, 2, \dots, k$ wyznaczyć współczynnik zbieżności φ^2 dla modelu pierwotnego (4). Spełniona jest bowiem równość:

$$\varphi^2 = c_i - \frac{b_i^2}{a_i} \tag{30}$$

Spełnione są także następujące zależności (Theil, 1979):

$$\varphi^2 = \varphi_{(-i)}^2 (1 - (r_i^*)^2) \tag{31}$$

gdzie $\varphi_{(-i)}^2$ – współczynnik zbieżności dla modelu uciętego.

Jeśli przez $R_{(-i)}^2$ oznaczymy współczynnik determinacji dla modelu uciętego, to równość (31) można zapisać w nieco innej postaci:

$$R^2 - R_{(-i)}^2 = (r_i^*)^2 \left(1 - R_{(-i)}^2\right) \tag{32}$$

Z równości (32) wynika, że dodatkowy udział i-tej zmiennej objaśniającej jest równy iloczynowi kwadratu współczynnika korelacji cząstkowej tej zmiennej ze zmienną objaśnianą w danym modelu i współczynnika zbieżności modelu uciętego. Jeśli współczynnik zbieżności modelu uciętego jest mały, to nie ma już miejsca na kolejną zmienną objaśniającą, dlatego jej dodatkowy udział w wyjaśnianiu zmiennej zależnej będzie niewielki.

5. Współczynnik semikorelacji cząstkowej

Podobny do współczynnika korelacji cząstkowej jest współczynnik semikorelacji cząstkowej $r_i^{(*)}$, będący miarą korelacji pomiędzy daną zmienną niezależną Z_i, z uwzględnieniem jej skorelowania z pozostałymi zmiennymi objaśniającymi modelu, a zmienną objaśnianą Y, bez uwzględnienia jej skorelowania z pozostałymi zmiennymi objaśniającymi modelu. Można zatem powiedzieć, że jest to współczynnik korelacji liniowej pomiędzy wektorem reszt modelu wewnętrznego a wektorem wartości zmiennej objaśnianej Y. Wobec tego:

$$r_i^{(*)} = r(\boldsymbol{v}_i, \boldsymbol{y}) \tag{33}$$

Ponieważ zmienne w modelu pierwotnym (a więc i w wewnętrznym) są standaryzowane, możemy zapisać:

$$r_i^{(*)} = \frac{\boldsymbol{v}_i^T \boldsymbol{y}}{\sqrt{\boldsymbol{y}^T \boldsymbol{y}} \sqrt{\boldsymbol{v}_i^T \boldsymbol{v}_i}} \tag{34}$$

Wykorzystując (18), otrzymujemy:

$$\begin{aligned} \boldsymbol{v}_i^T \boldsymbol{y} &= (z_i - \mathbf{Z}(-i)\mathbf{R}_{ii}^{-1}\boldsymbol{\rho}_i)^T \boldsymbol{y} = \\ &= nr_i - \boldsymbol{\rho}_i^T \mathbf{R}_{ii}{}^{-1}\left(\mathbf{Z}(-i)\right)^T \boldsymbol{y} = \\ &= nr_i - n\boldsymbol{\rho}_i^T \, \mathbf{R}_{ii}{}^{-1}\boldsymbol{R}_{0i} \end{aligned} \tag{35}$$

Na podstawie (21) – ponieważ zmienna zależna Y jest standaryzowana – wzór (34) przyjmuje postać:

$$r_i^{(*)} = \frac{\boldsymbol{v}_i^T \boldsymbol{y}}{\sqrt{\boldsymbol{y}^T \boldsymbol{y}} \sqrt{\boldsymbol{v}_i^T \boldsymbol{v}_i}} = \frac{r_i - \boldsymbol{\rho}_i^T \mathbf{R}_{ii}^{-1} \boldsymbol{R}_{0i}}{\sqrt{1 - \boldsymbol{\rho}_i^T \mathbf{R}_{ii}^{-1} \boldsymbol{\rho}_i}} \tag{36}$$

Warto zauważyć, że w celu obliczenia współczynnika semikorelacji cząstkowej możemy również posłużyć się podwójną macierzą brzegową (27). Na podstawie (28) otrzymujemy:

$$r_i^{(*)} = \frac{b_i}{\sqrt{a_i}} \tag{37}$$

Współczynnik semikorelacji cząstkowej jest liczbą z przedziału $[-1, 1]$.

Kwadrat współczynnika semikorelacji cząstkowej – współczynnik determinacji semicząstkowej – oznacza stosunek wyłącznej zmienności danej zmiennej objaśniającej Z_i do całkowitej zmienności zmiennej objaśnianej Y. Przyjmuje wartości z przedziału [0, 1]. Podobnie jak w wypadku współczynnika determinacji cząstkowej jego wartość bliska 0 oznacza, że zmienna objaśniająca niesie bezużyteczną informację o zmiennej objaśnianej w danym modelu, a zatem jest nadmiarowa (StatSoft, 2006).

Zauważmy, że:

$$r_i^* = \frac{r_i^{(*)}}{\sqrt{1 - \boldsymbol{R}_{0i}^T \mathbf{R}_{\bar{i}\bar{i}}^{-1} \boldsymbol{R}_{0i}}} = \frac{r_i^{(*)}}{\sqrt{\varphi_{(-i)}^2}} \tag{38}$$

Na podstawie (38) stwierdzamy, że współczynnik semikorelacji cząstkowej jest zawsze co najwyżej równy współczynnikowi korelacji cząstkowej oraz że oba mają identyczne znaki.

Równość (38) można zapisać też w innej, równoważnej postaci:

$$\left(r_i^{(*)}\right)^2 = (r_i^*)^2 \varphi_{(-i)}^2 \tag{39}$$

gdzie $\varphi_{(-i)}^2$ – współczynnik zbieżności (tolerancja) odpowiedniego modelu uciętego.

6. Porównanie omówionych współczynników korelacji liniowej

Zanim przejdziemy do przykładu prezentującego wyznaczanie i interpretacje omówionych w artykule współczynników korelacji liniowej, podsumujmy ich wybrane własności.

Współczynnik korelacji wielowymiarowej wyznacza się dla wektorów wartości empirycznej i teoretycznej zmiennej objaśnianej modelu. Jest zawsze nieujemny, a jego wartość jest nie mniejsza od największej wartości bezwzględnej współczynnika korelacji liniowej zmiennej objaśnianej i objaśniającej. Jego kwadrat to współczynnik determinacji, wyrażający procent zmienności zmiennej zależnej wyjaśniony przez model.

Współczynnik korelacji cząstkowej danej zmiennej objaśniającej ze zmienną objaśnianą modelu, czyli współczynnik korelacji liniowej wektorów reszt modeli uciętego i wewnętrznego, jest miarą korelacji liniowej pomiędzy zmienną objaśnianą i objaśniającą z uwzględnieniem ich skorelowania z pozostałymi zmiennymi objaśniającymi. Jego wartość może być mniejsza lub większa od wartości odpowiedniego współczynnika korelacji liniowej r_i, a znak jest zgodny ze znakiem oszacowania parametru strukturalnego przy danej zmiennej objaśniającej modelu. Iloczyn kwadratu współczynnika korelacji cząstkowej i współczynnika zbieżności modelu uciętego wyraża dodatkowy udział danej zmiennej objaśniającej w modelu mierzony wzrostem wartości współczynnika determinacji w stosunku do współczynnika determinacji modelu uciętego. Mały dodatkowy udział świadczy o tym, że zmienna jest zbędna w modelu.

Współczynnik semikorelacji cząstkowej jest współczynnikiem korelacji liniowej pomiędzy wektorem reszt modelu wewnętrznego a wektorem wartości empirycznych zmiennej objaśnianej Y. Stanowi on miarę korelacji liniowej pomiędzy daną zmienną niezależną Z_i, z uwzględnieniem jej skorelowania z pozostałymi zmiennymi objaśniającymi modelu, a zmienną objaśnianą Y, bez uwzględnienia jej skorelowania z pozostałymi zmiennymi objaśniającymi modelu. Jest zawsze co najwyżej równy współczynnikowi korelacji cząstkowej (oba zawsze mają identyczne znaki) i jednocześnie nie większy od odpowiedniego współczynnika korelacji liniowej r_i.

Współczynnik determinacji semicząstkowej (czyli kwadrat współczynnika semikorelacji cząstkowej) jest iloczynem współczynnika zbieżności modelu uciętego oraz kwadratu współczynnika korelacji cząstkowej. Jego duża wartość bezwzględna świadczy o tym, że zmienna objaśniająca, po wyeliminowaniu z niej wpływu pozostałych zmiennych objaśniających modelu, ma istotny wpływ na zmienną zależną modelu.

Każdy z omówionych współczynników korelacji eksponuje inne własności zmiennych w modelu, dlatego wskazane jest najpierw wyznaczenie wartości ich wszystkich i dopiero na tej podstawie wyciąganie wniosków o jakości zmiennej czy modelu.

7. Przykład praktyczny

Na zakończenie prezentujemy przykład wyznaczania omówionych współczynników korelacji na podstawie modelu wzrostu płac w Polsce, szacowanego metodą najmniejszych kwadratów, z wykorzystaniem danych za lata 2001–2019 zaczerpniętych z Banku Danych Makroekonomicznych GUS. Wykonane obliczenia mają wyłącznie charakter ilustrujący wcześniejsze rozważania.

Rozpatrzmy następujący model:

$$y_t = \alpha_0 z_{0t} + \alpha_1 z_{1t} + \alpha_2 z_{2t} + \alpha_3 z_{3t} + \alpha_4 z_{4t} + \xi_t \tag{40}$$

gdzie:

y – wzrost przeciętnego rocznego wynagrodzenia brutto w %,

z_0 – zmienna tożsamościowo równa 1,

z_1 – zmienna czasowa w latach,

z_2 – zmiana cen towarów i usług konsumpcyjnych w %,

z_3 – zmiana PKB brutto *per capita* w %,

z_4 – stopa bezrobocia rejestrowanego w %.

W modelu wykorzystano dane empiryczne za 19 lat, a zatem liczba obserwacji $n = 19$. Dane zamieszczono w tabl. 2.

Tabl. 2. Wyniki obserwacji zmiennych modelu (40)

t	y_t	z_{0t}	z_{1t}	z_{2t}	z_{3t}	z_{4t}
2001	2,5	1	1	5,5	4,44	17,5
2002	0,7	1	2	1,9	3,99	19,0
2003	3,4	1	3	0,8	4,43	20,0
2004	0,7	1	4	3,5	10,35	19,0
2005	1,8	1	5	2,1	6,20	17,6
2006	4,0	1	6	1,0	8,10	14,8
2007	5,5	1	7	2,5	11,04	11,2
2008	5,9	1	8	4,2	8,23	9,5
2009	2,0	1	9	3,5	6,61	12,1
2010	1,4	1	10	2,6	4,34	12,4
2011	1,4	1	11	4,3	8,38	12,5
2012	0,1	1	12	3,7	4,01	13,4
2013	2,8	1	13	0,9	1,79	13,4
2014	3,2	1	14	0,0	3,88	11,4
2015	4,5	1	15	–0,9	4,75	9,7
2016	4,3	1	16	–0,6	3,39	8,2
2017	3,7	1	17	2,0	6,89	6,6
2018	5,4	1	18	1,6	6,39	5,8
2019	4,8	1	19	2,3	7,13	5,2

Źródło: Bank Danych Makroekonomicznych GUS.

Na podstawie danych z tabl. 2, po dokonaniu standaryzacji zmiennych, zgodnie ze wzorem (6) wyznaczono parę korelacyjną określającą model (40):

$$\mathbf{R}(4) = \begin{bmatrix} 1 & -0{,}411 & -0{,}160 & -0{,}900 \\ -0{,}411 & 1 & 0{,}395 & 0{,}215 \\ -0{,}160 & 0{,}395 & 1 & -0{,}083 \\ -0{,}900 & 0{,}215 & -0{,}083 & 1 \end{bmatrix}, \quad \boldsymbol{R}_0(4) = \begin{bmatrix} 0{,}403 \\ -0{,}339 \\ 0{,}237 \\ -0{,}631 \end{bmatrix} \tag{41}$$

Z analizy uzyskanych wartości współczynników korelacji liniowej zmiennej endogenicznej z poszczególnymi zmiennymi objaśniającymi modelu (40) wynika, że w badanym okresie z roku na rok następował wzrost przeciętnego wynagrodzenia, ale nie nadążał on za wzrostem cen (współczynnik korelacji liniowej tych cech wynosił –0,339). Analogicznie zwiększeniu wzrostu wydajności pracy (mierzonej wzrostem PKB *per capita*) towarzyszył wzrost wynagrodzeń. Wzrost bezrobocia powodował zaś spadek wartości zmiennej objaśnianej.

Za pomocą metody najmniejszych kwadratów wyznaczono oszacowania parametrów modelu o zmiennych standaryzowanych, określonego przez parę korelacyjną $(\mathbf{R}(4), \boldsymbol{R}_0(4))$ daną wzorem (41)[1]:

[1] W celu uproszczenia zapisu zmienne i parametry modelu o zmiennych standaryzowanych (41) oraz modelu wyjściowego (39) oznaczono identycznie.

$$y_t = \alpha_1 z_{1t} + \alpha_2 z_{2t} + \alpha_3 z_{3t} + \alpha_4 z_{4t} + \xi_t \tag{42}$$

Wykorzystując (8), otrzymano:

$$\boldsymbol{A} = \mathbf{R}^{-1}\boldsymbol{R}_0 = \begin{bmatrix} -1{,}408 \\ -0{,}574 \\ 0{,}093 \\ -1{,}767 \end{bmatrix} \tag{43}$$

Warto zwrócić uwagę na rozbieżność znaków oszacowania parametru stojącego przy pierwszej zmiennej objaśniającej modelu (czas) oraz współczynnika korelacji liniowej tej zmiennej ze zmienną objaśnianą, a zatem brak koincydencji zmiennej czasowej. Można stąd wysnuć wniosek, że analizowany model wskazuje na zbyt wolne tempo wzrostu płac.

Do wyznaczenia współczynnika korelacji cząstkowej r_1^* zmiennej objaśnianej Y ze zmienną objaśniającą Z_1 w modelu (40), zgodnie ze wzorem (22), wykorzystano podwójną macierz brzegową (27). Stąd:

$$\boldsymbol{\Delta}_1 = \begin{bmatrix} \mathbf{R}_{11} & \boldsymbol{\rho}_1 & \boldsymbol{R}_{01} \\ \boldsymbol{\rho}_1^T & 1 & r_1 \\ \boldsymbol{R}_{01}^T & r_1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0{,}395 & 0{,}215 & -0{,}411 & -0{,}339 \\ 0{,}395 & 1 & -0{,}083 & -0{,}160 & 0{,}237 \\ 0{,}215 & -0{,}083 & 1 & 0{,}900 & -0{,}631 \\ -0{,}411 & -0{,}160 & 0{,}900 & 1 & 0{,}403 \\ -0{,}339 & 0{,}237 & -0{,}631 & 0{,}403 & 1 \end{bmatrix} \tag{44}$$

W wyniku przekształceń elementarnych wykonanych na macierzy (44) zgodnie ze wzorem (28) otrzymano:

$$(\boldsymbol{\Delta}_1)^* = \begin{bmatrix} (\mathbf{R}_{11})^* & (\boldsymbol{\rho}_1)^* & (\boldsymbol{R}_{01})^* \\ 0,\dots,0 & \alpha_1 & b_1 \\ 0,\dots,0 & b_1 & c_1 \end{bmatrix} =$$

$$= \begin{bmatrix} 1 & -0{,}395 & 0{,}215 & -0{,}411 & -0{,}339 \\ 0 & 1 & -0{,}199 & 0{,}003 & 0{,}439 \\ 0 & 0 & 1 & -0{,}881 & -0{,}526 \\ 0 & 0 & 0 & 0{,}117 & -0{,}164 \\ 0 & 0 & 0 & -0{,}164 & 0{,}467 \end{bmatrix} \tag{45}$$

Na podstawie (29), wykorzystując (45), można już wyznaczyć współczynnik korelacji cząstkowej r_1^*:

$$r_1^* = \frac{b_1}{\sqrt{a_1 \cdot c_1}} = \frac{-0{,}164}{\sqrt{0{,}117 \cdot 0{,}467}} = -0{,}704 \tag{46}$$

Współczynnik r_1^* ma wartość ujemną – znak współczynnika korelacji cząstkowej jest zawsze zgodny ze znakiem oszacowania parametru stojącego przy danej zmiennej. Wysoka wartość bezwzględna tego współczynnika świadczy o tym, że pierwsza zmienna objaśniająca modelu ma istotny wpływ na część zmiennej zależnej niewyjaśnioną przez pozostałe zmienne objaśniające. Zgodnie ze wzorem (30) współczynnik zbieżności dla modelu pierwotnego (42) jest równy:

$$\varphi^2 = c_1 - \frac{b_1^2}{a_1} = 0{,}467 = \frac{(-0{,}164)^2}{0{,}117} = 0{,}236 \tag{47}$$

Współczynnik zbieżności $\varphi^2_{(-1)}$ dla modelu uciętego wynosi:

$$\varphi^2_{(-1)} = c_1 = 0{,}467 \tag{48}$$

Następnie wyznaczono dodatkowe udziały zmiennej Z_1 modelu pierwotnego. Zgodnie ze wzorem (32) otrzymano:

$$\begin{gathered} \varphi^2_{(-1)} - \varphi^2 = (r_1^*)^2\left(1 - R^2_{(-1)}\right) = (r_1^*)^2\varphi^2_{(-1)} \\ \varphi^2_{(-1)} - \varphi^2 = (-0{,}704)^2 \cdot 0{,}467 = 0{,}232 \end{gathered} \tag{49}$$

Dodatkowy udział zmiennej Z_1 jest wysoki. Zmienna ta w sposób istotny podnosi jakość modelu (40), mierzoną wartością współczynnika zbieżności.

Współczynnik semikorelacji cząstkowej wyznaczono na podstawie (37), wykorzystując (45). Otrzymano:

$$r_1^{(*)} = \frac{b_1}{\sqrt{a_1}} = \frac{-0{,}164}{\sqrt{0{,}117}} = -0{,}481 \tag{50}$$

Duża wartość bezwzględna tego współczynnika świadczy o przydatności zmiennej Z_1 w modelu. Pierwsza zmienna objaśniająca modelu, po wyeliminowaniu z niej wpływu pozostałych zmiennych niezależnych, istotnie wpływa na zmienną zależną.

Na zakończenie rozważań dotyczących pierwszej zmiennej zauważmy, że liczba a_1, otrzymana w macierzy (45), jest współczynnikiem zbieżności dla modelu wewnętrznego (w którym zmienną objaśnianą jest zmienna Z_1, a zbiór zmiennych objaśniających jest identyczny ze zbiorem w modelu uciętym). Jeśli więc przez $R^2_{(w1)}$ oznaczymy współczynnik determinacji dla tego modelu wewnętrznego, to:

$$R^2_{(w1)} = 1 - a_1 = 1 - 0{,}117 = 0{,}883 \tag{51}$$

Dla zmiennej Z_2 otrzymano:

$$\mathbf{\Delta}_2 = \begin{bmatrix} \mathbf{R}_{22} & \boldsymbol{\rho}_{12} & \boldsymbol{R}_{02} \\ \boldsymbol{\rho}_2{}^T & 1 & r_2 \\ \boldsymbol{R}_{02}{}^T & r_2 & 1 \end{bmatrix} =$$

$$= \begin{bmatrix} 1 & -0{,}160 & -0{,}900 & -0{,}411 & 0{,}403 \\ -0{,}160 & 1 & -0{,}083 & 0{,}395 & 0{,}237 \\ -0{,}900 & -0{,}083 & 1 & 0{,}215 & -0{,}631 \\ -0{,}411 & 0{,}395 & 0{,}215 & 1 & -0{,}339 \\ 0{,}403 & 0{,}237 & -0{,}631 & -0{,}339 & 1 \end{bmatrix} \tag{52}$$

W wyniku przekształceń elementarnych wykonanych na macierzy (52), zgodnie ze wzorem (28) mamy:

$$(\mathbf{\Delta}_2)^* = \begin{bmatrix} (\mathbf{R}_{22})^* & (\boldsymbol{\rho}_2)^* & (\boldsymbol{R}_{02})^* \\ 0,\dots,0 & \alpha_2 & b_2 \\ 0,\dots,0 & b_2 & c_2 \end{bmatrix} =$$

$$= \begin{bmatrix} 1 & -0{,}160 & -0{,}900 & -0{,}411 & 0{,}403 \\ 0 & 1 & -0{,}232 & 0{,}338 & 0{,}309 \\ 0 & 0 & 1 & -0{,}565 & -1{,}443 \\ 0 & 0 & 0 & 0{,}676 & -0{,}388 \\ 0 & 0 & 0 & -0{,}388 & 0{,}458 \end{bmatrix} \tag{53}$$

$$r_2{}^* = \frac{b_2}{\sqrt{a_2 . c_2}} = \frac{-0{,}388}{\sqrt{0{,}676 \cdot 0{,}458}} = -0{,}697 \tag{54}$$

Druga zmienna objaśniająca modelu, po wyeliminowaniu z niej wpływu pozostałych trzech zmiennych niezależnych, istotnie wpływa na zmienną zależną:

$$\varphi^2 = c_2 - \frac{b_2{}^2}{a_2} = 0{,}458 - \frac{(-0{,}388)^2}{0{,}676} = 0{,}236 \tag{55}$$

$$\varphi^2_{(-2)} = c_2 = 0{,}458 \tag{56}$$

Dodatkowe udziały zmiennej Z_2 modelu pierwotnego wynoszą:

$$\begin{gathered} \varphi^2_{(-2)} - \varphi^2 = (r_2^*)^2\left(1 - R^2_{(-2)}\right) = (r_2^*)^2 \varphi^2_{(-2)} \\ \varphi^2_{(-2)} - \varphi^2 = (-0{,}697)^2 \cdot 0{,}458 = 0{,}223 \end{gathered} \tag{57}$$

Jak widać, dodatkowy udział zmiennej Z_2 również jest wysoki i zmienna ta jest przydatna w modelu.

Współczynnik semikorelacji cząstkowej wynosi:

$$r_2^{(*)} = \frac{b_2}{\sqrt{a_2}} = \frac{-0{,}388}{\sqrt{0{,}676}} = -0{,}472 \tag{58}$$

Jego duża wartość bezwzględna świadczy o tym, że zmienna Z_2 wnosi istotne informacje do modelu. Druga zmienna objaśniająca modelu, po wyeliminowaniu z niej wpływu pozostałych zmiennych niezależnych, istotnie wpływa na zmienną zależną.

Współczynnik determinacji dla odpowiedniego modelu wewnętrznego jest równy:

$$R^2_{(w2)} = 1 - a_2 = 1 - 0{,}676 = 0{,}324 \tag{59}$$

Analogiczne obliczenia należałoby przeprowadzić dla dwóch ostatnich zmiennych modelu (42). W tym miejscu ograniczymy się do podania w postaci tabelarycznej wyników końcowych, które zostały uzyskane z wykorzystaniem programu Statistica 13.3 PL (tabl. 3). Zauważmy, że dla dwóch pierwszych zmiennych są one identyczne z wynikami uzyskanymi w wyniku zastosowania macierzy brzegowej (27), co potwierdza poprawność przeprowadzonych obliczeń.

Tabl. 3. Udziały zmiennych wyróżnionych w modelu (40)

Zmienne	Oszacowania parametrów	Współczynnik			
		korelacji cząstkowej	semikorelacji cząstkowej	zbieżności modeli wewnętrznych $\varphi^2_{(wi)}$ (tolerancja)	determinacji modeli wewnętrznych $R^2_{(wi)}$
Z_1	–1,408	–0,704	–0,481	0,117	0,883
Z_2	–0,574	–0,697	–0,472	0,676	0,324
Z_3	0,093	0,155	0,076	0,675	0,325
Z_4	–1,767	–0,795	–0,635	0,129	0,871

Źródło: obliczenia własne z wykorzystaniem programu Statistica 13.3 PL.

Analizując wyniki zawarte w tabl. 3, a także wektor korelacji dany wzorem (41), należy stwierdzić, że mając model regresji liniowej, warto wyznaczyć współczynniki korelacji liniowej i cząstkowej oraz semikorelacji cząstkowej i dopiero na tej podstawie wyciągać wnioski, a nie opierać się tylko na jednym z tych współczynników, każdy bowiem dostarcza innych informacji. Dopiero niska wartość bezwzględna wszystkich trzech współczynników może świadczyć o nieprzydatności danej zmiennej w modelu (w omawianym przykładzie taką zmienną była Z_3 – zmiana PKB brutto *per capita*, która miała przybliżać zmianę wydajności pracy). Z kolei zmienna Z_2 (indeks cen towarów i usług) ma niską wartość bezwzględną współczynnika korelacji liniowej ze zmienną objaśnianą (0,335), ale wysokie wartości bezwzględne współczynników korelacji cząstkowej i semikorelacji cząstkowej. Świadczy to o wyjaśnieniu znacznej części zmiennej objaśnianej, która nie została wyjaśniona przez inne zmienne egzogeniczne modelu, i przydatności tej zmiennej. Podobnie duży zasób nowych informacji o zmiennej endogenicznej wnoszą dwie pozostałe zmienne i dlatego powinny pozostać w omawianym modelu.

8. Podsumowanie

Przedstawiona w artykule definicja współczynników korelacji wielowymiarowej, korelacji cząstkowej i semikorelacji cząstkowej jako współczynników korelacji liniowej Pearsona odpowiednich wektorów pozwala na ujednolicenie ich procesu obliczeniowego oraz ich interpretacji. Z kolei zastosowanie podwójnej macierzy brzegowej znacznie ułatwia wyznaczanie tych współczynników, a obliczenia są dzięki temu przejrzystsze. Zaprezentowany przez autora sposób postępowania odbiega od powszechnie spotykanego w literaturze, który w wielu przypadkach polega na podawaniu wzorów na poszczególne współczynniki korelacji wykorzystujących wyznaczniki odpowiednich macierzy.

Zaproponowane podejście ma znaczenie praktyczne i edukacyjne. Zastosowanie metody obliczania z wykorzystaniem macierzy brzegowych nawet w przypadku stosunkowo licznej próby nie nastręcza większych trudności – wystarczy w tym celu napisać kilka formuł w arkuszu kalkulacyjnym. Obliczenia są przejrzyste i zrozumiałe, dzięki temu bardziej intuicyjne, a jednocześnie cechuje je duża dokładność. Jest to warte polecenia, szczególnie dla początkujących badaczy oraz studentów przed zapoznaniem się z pakietami statystycznymi. Należy przy tym podkreślić, że stosując przedstawione w artykule podejście, definicje i wzory, można wyznaczyć współczynniki korelacji wielowymiarowej, cząstkowej i semicząstkowej za pomocą arkusza kalkulacyjnego i funkcji współczynnika korelacji liniowej odpowiednich wektorów, nawet bez posługiwania się macierzami brzegowymi czy pakietem statystycznym.

Bibliografia

Charemza, W. W., Deadman, D. F. (1997). *Nowa ekonometria*. Warszawa: Polskie Wydawnictwo Ekonomiczne.

Dziechciarz, J. (red.). (2015). *Ekonometria: metody, przykłady, zadania.* Wrocław: Wydawnictwo Uniwersytetu Ekonomicznego.

Ezekiel, M., Fox, K. A. (1959). *Methods of Correlation and Regression Analysis: Linear and Curvilinear.* New York, London: Wiley & Sons.

Gajda, J. B. (2004). *Ekonometria*. Warszawa: Wydawnictwo C.H. Beck.

Kolupa, M., Pacholewicz, E., Śleszyński, Z. (1992). *Metody algebry liniowej w ekonometrii.* Radom: Wydawnictwo Wyższej Szkoły Inżynierskiej.

Kolupa, M., Śleszyński, Z. (2010). *Metody ekonometryczne*. Radom: Wydawnictwo Politechniki Radomskiej.

StatSoft. (2006) *Elektroniczny Podręcznik Statystyki PL*. Pobrane z: https://www.statsoft.pl/textbook/stathome.html.

Theil, H. (1979). *Zasady ekonometrii*. Warszawa: Państwowe Wydawnictwo Naukowe.

Zeliaś, A. (2000). *Metody statystyczne*. Warszawa: Polskie Wydawnictwo Ekonomiczne.