

**METODA HELLWIGA JAKO METODA ILOŚCIOWA
DOBORU ZMIENNYCH DO MODELOWANIA PROCESÓW LOGISTYCZNYCH**

**HELLWIG METHOD AS A QUANTITATIVE METHOD
SELECTION OF VARIABLES FOR MODELING LOGISTIC PROCESSES**

Agnieszka DUDA

Akademia Sztuki Wojennej, Wydział Zarządzania i Dowodzenia, Instytut Logistyki

Streszczenie

Temat artykułu dotyczy doboru zmiennych w modelowaniu procesów logistycznych jedną z wybranych metod ilościowych, jaką jest metoda Hellwiga. Dokonując analizy teoretycznej i empirycznej badanego problemu, opracowano precyzyjny algorytm doboru zmiennych objaśniających, spośród zmiennych kandydujących, do modelu ekonometrycznego metodą Hellwiga. Przedstawiony algorytm pozwala na jednoznaczny wybór, na podstawie pojemności nośników informacji w oparciu o współczynniki korelacji liniowej Pearsona, najwłaściwszych zmiennych do tego modelu.

Algorytm ten, posłużył do praktycznego obliczenia i na tej podstawie wyboru najlepszych zmiennych objaśniających na bazie fizycznego przykładu opartego na danych empirycznych firmy Wawel S.A. Dla właściwej egzemplifikacji, podjęto próbę zbudowania empirycznego modelu ekonometrycznego. Zidentyfikowano cztery potencjale zmienne kandydujące do tego modelu poprzez analizę i opracowanie danych empirycznych pozyskanych ze sprawozdań finansowych firmy. Następnie wykorzystując opracowany algorytm metody Hellwiga, otrzymano rzeczywiste wyniki doboru prawdziwych zmiennych do realnego modelu ekonometrycznego opisującego produkcję sprzedaną w firmie Wawel S.A.

Słowa kluczowe: metoda Hellwiga, model ekonometryczny, dobór zmiennych do modelu, metoda ilościowa, procesy logistyczne

Abstract

The subject of the article concerns the selection of variables in the modeling of logistic processes, one of the quantitative methods chosen, which is the Hellwig method. Performing a theoretical and empirical analysis of the examined problem, a precise algorithm for the selection of explanatory variables was developed, among the candidate variables, to the econometric model using the Hellwig method. The presented algorithm allows for an unambiguous selection, based on the capacity of information carriers based on Pearson's linear correlation coefficients, the most appropriate variables for this model.

This algorithm was used for practical calculation and on this basis the selection of the best explanatory variables on the basis of a physical example based on empirical data of the Wawel SA company. For the proper exemplification, an attempt to build an empirical econometric model. Four potential variables candidate for this model were identified by analyzing and developing empirical data obtained from the company's financial reports. Next, using the developed algorithm of the Hellwig method, real results of the selection of real variables were obtained to the real econometric model describing the production sold in the company Wawel S.A.

Key words: Hellwig's method, econometric model, selection of variables to the model,

Wstęp

Dobór zmiennych objaśniających¹ w modelowaniu procesów logistycznych metodą Hellwiga, podobnie jak w innych metodach, takich jak metoda analizy grafów, oparty jest na realizacji opracowanego algorytmu działań. W literaturze przedmiotu badań występuje wiele metod doboru zmiennych do modeli ekonometrycznych. Występują metody jakościowe, ilościowe, a także heurystyczne. Metody jakościowe polegają na analizie i ocenie badanego zjawiska i na bazie obserwacji empirycznej prowadzonej przez badacza zostają wybrane stosowne zmienne do modelu ekonometrycznego. Metody ilościowe to zbiór metod matematycznych, statystycznych, ekonometrycznych czy badań operacyjnych², które wykorzystuje się do obliczeń w efekcie, których zostają dobrane zmienne do modelu, natomiast metody heurystyczne polegają na analizie, krok po kroku, badanego zjawiska poprzez odkrywanie kolejnych, lepszych rozwiązań.

Celem podjętych badań jest właściwy, na podstawie pojemności nośników informacji w oparciu o współczynniki korelacji liniowej Pearsona, dobór zmiennych objaśniających do modelu ekonometrycznego metodą Hellwiga, spośród wszystkich analizowanych zmiennych kandydujących do tego modelu. Natomiast **problem badawczy** został postawiony w formie pytania: w jaki sposób dobrać zmienne objaśniające do modelu ekonometrycznego, wykorzystując metodę Hellwiga?

Na podstawie tak postawionego celu i problemu badawczego sformułowano **hipotezę badawczą** w brzmieniu: dobór zmiennych objaśniających do modelu ekonometrycznego metodą Hellwiga zgodnie z przyjętym algorytmem, wykorzystując zależności funkcyjne współczynników korelacji liniowej Pearsona ujętych w indywidualnych i integralnych pojemnościach nośników informacji dla kombinacji zmiennych, eliminują te zmienne kandydujące, które nie są istotne dla badanego procesu logistycznego.

Takie sformułowanie celu badań, problemu i hipotezy badawczej wymusiło wykorzystanie stosownej teoretycznej i empirycznej metody badawczej. W metodzie teoretycznej zastosowano krytyczną analizę literatury przedmiotu badań. Natomiast metodę empiryczną zastosowano do praktycznego wyboru (obliczenia) najlepszych zmiennych objaśniających spośród zmiennych kandydujących do modelu ekonometrycznego, który w swym założeniu opisuje wielkość produkcji sprzedanej. Dane empiryczne (tabela 1), wykorzystane w obliczeniach, pochodzą ze sprawozdań finansowych firmy Wawel S.A. z lat 2004÷2018.

W budowaniu modeli ekonometrycznych w fazie początkowej niezmiernie ważnym jest właściwe pod względem merytorycznym i ilościowym dobranie stosownych zmiennych objaśniających do tworzonego modelu. Niewłaściwy dobór zmiennych często skutkuje zbudowaniem słabego modelu, który zazwyczaj nie spełnia w kolejnych etapach modelowania nawet podstawowych wymagań statystycznych testów i weryfikacji.

¹ Zmienna objaśniająca w modelach matematycznych i ekonometrycznych w literaturze przedmiotu badań przez różnych autorów nazywana jest także zmienną niezależną, egzogeniczną, regresorem czy zmienną kontrolującą. B. R. Górecki, *Ekonometria podstawy teorii i praktyki*, Wydawnictwo Kay Text, Warszawa 2010, s. 15.

² Tamże, s. 13.

Metoda Hellwiga polega w swej istocie na ograniczeniu liczby zmiennych objaśniających spośród wielu kandydujących do tego modelu. Ponadto, opiera się na doborze zmiennych objaśniających, które słabo są skorelowane między sobą, a jednocześnie silnie skorelowane ze zmienną objaśnianą³.

Dobór zmiennych objaśniających do modelu ekonometrycznego wymaga policzenia współczynników korelacji pomiędzy zmienną objaśnianą a kandydującymi zmiennymi objaśniającymi oraz współczynników korelacji pomiędzy kandydującymi zmiennymi objaśniającymi. Na tej podstawie można policzyć indywidualne i integralne pojemności nośników informacji we wszystkich możliwych kombinacjach tych zmiennych, mogących wystąpić w modelu ekonometrycznym. Najlepszą kombinacją zmiennych, które wybierzemy do modelu, jest ta kombinacja, dla której obliczono największą wartość integralnej pojemności nośników informacji.

Przesłanką do podjęcia badań, dotyczących doboru zmiennych do modeli ekonometrycznych metodą Hellwiga, są występujące w literaturze przedmiotu luki dotyczące kompletności algorytmu i praktycznych przykładów zastosowań.

Dobór zmiennych objaśniających metodą Hellwiga

Niezmiernie ważnym elementem modelowania procesów logistycznych jest wybór właściwej metody ilościowej, dzięki której możliwe będzie dobranie odpowiednich zmiennych objaśniających do modelu ekonometrycznego ze wszystkich zmiennych kandydujących. „Modelowanie procesów logistycznych polega na tworzeniu modeli matematycznych nazywanych w konkretnych przypadkach modelami ekonometrycznymi, które otrzymujemy dzięki realizacji z góry przyjętego algorytmu oraz dokonanych pośrednich czynności i obliczeń⁴”.

Istotą metody Hellwiga doboru zmiennych objaśniających w początkowej, wyjściowej postaci modelu empirycznego jest analiza mierników pojemności nośników informacji dotyczących wielkości powiązań (korelacji), pomiędzy zmiennymi objaśniającymi. Badanie pojemności nośników informacji realizowane jest na bazie rozstrzygnięć na podstawie analizy prostej zależności funkcyjnej, obliczonych współczynników korelacji liniowej Pearsona dla wszystkich zmiennych kandydujących (objaśniających i objaśnianej) do modelu. Zmiennymi kandydującymi możemy nazwać wszystkie możliwe zmienne objaśniające, które utworzą zbiór opisujący badany proces logistyczny.

Natomiast, zmienna objaśniana to taka zmienna, której określone liczbowo wartości rosną lub maleją pod wpływem działania zmiennych objaśniających⁵. Różne procesy logistyczne, rozpatrywane przez badacza w początkowej fazie, mogą zawierać wiele kandydujących zmiennych objaśniających, dlatego niezmiernie ważnym jest wybór odpowiedniej metody ilościowej, w naszym rozpatrywanym przypadku metody Hellwiga, w celu ograniczenia liczby tych zmiennych do najbardziej istotnych.

³ Zmienna objaśniana w literaturze przedmiotu badań nazywana jest także zmienną zależną, endogeniczną, regresantem czy też zmienną kontrolowaną. B. R. Górecki, *Ekonometria...*, wyd. cyt., s. 15.

⁴ M. Mróz, *Dobór zmiennych niezależnych w modelowaniu procesów logistycznych metodą analizy grafów*, „Zeszyty Naukowe Akademii Sztuki Wojennej”, nr 3(108)/2017, s. 129.

⁵ J. D. Łaniec, *Elementy statystyki dla pedagogów*, Wydawnictwo Uniwersytetu Warmińsko-Mazurskiego, Olsztyn 1999, s. 17.

Cechą charakterystyczną metod ilościowych mających na celu ograniczenie zbioru wstępnie wytypowanych zmiennych objaśniających jest przyjęcie zasady, że w modelu ekonometrycznym powinny znaleźć się zmienne objaśniające silnie skorelowane ze zmienną objaśnianą i jednocześnie słabo skorelowane między sobą. Nieprzestrzeganie tej zasady lub tolerowanie zjawiska współliniowości zmiennych objaśniających powoduje wiele niekorzystnych efektów podczas modelowania ekonometrycznego⁶.

Najczęściej wykorzystywaną metodą doboru zmiennych do modelu ekonometrycznego jest metoda Hellwiga, która została opracowana 1969 roku przez Zdzisława Hellwiga. Metoda ta „polega na wyborze takiej kombinacji zmiennych objaśniających ze zbioru zmiennych potencjalnych, których współczynnik korelacji liniowej ze zmienną objaśnianą jest jak największy, natomiast współczynnik korelacji liniowej między zmiennymi objaśniającymi występującymi w danej kombinacji jest jak najmniejszy⁷”. Jednak nie jest to jedyne kryterium doboru tych zmiennych. Ważnym elementem tej metody jest także kryterium liczbowe, czyli pojemności integralnej nośników informacji. Każda zmienna objaśniająca posiada odpowiedni nośnik informacji.

Aby zastosować metodę Hellwiga, należy przeprowadzić działania i obliczenia zgodnie z przyjętym poniżej algorytmem:

- 1) Dobór merytorycznych zmiennych objaśniających kandydujących do modelu ekonometrycznego;
- 2) Obliczenie współczynników korelacji liniowej Pearsona pomiędzy zmienną objaśnianą y_t a zmiennymi objaśniającymi x_{it} oraz współczynników korelacji pomiędzy zmiennymi objaśniającymi x_{it} ;
- 3) Wyznaczenie wektora głównego R_0 współczynników korelacji pomiędzy zmienną objaśnianą y_t a zmiennymi objaśniającymi x_{it} oraz macierzy głównej R współczynników korelacji między zmiennymi objaśniającymi x_{it} ;
- 4) Ustalenie liczby kombinacji L zmiennych objaśniających x_{it} ;
- 5) Wypisanie wszystkich kombinacji k_k zmiennych objaśniających x_{it} ;
- 6) Dla wszystkich wypisanych kombinacji k_k , utworzenie odpowiadających im cząstkowych wektorów R_{0k} współczynników korelacji r_{0j} , pomiędzy zmienną objaśnianą y_t a odpowiednimi zmiennymi objaśniającymi x_{it} oraz cząstkowych macierzy R_k współczynników korelacji r_{ij} pomiędzy zmiennymi objaśniającymi x_{it} ;
- 7) Obliczanie indywidualnych pojemności nośników informacji $h_{k,j}$ oraz na ich podstawie integralnych pojemności nośników informacji H_k ;
- 8) Wybór zmiennych objaśniających x_{it} do modelu ekonometrycznego, spośród wszystkich kombinacji, na podstawie maksymalnej wartości integralnej pojemności nośników informacji H_k i utworzenie na tej podstawie empirycznego modelu ekonometrycznego z wybranymi zmiennymi objaśniającymi x_{it} .

⁶ K. Kukuła (red.), Wprowadzenie do ekonometrii w przykładach i zadaniach, PWN, Warszawa 2009, s. 28.

⁷ A. Duda, Modelowanie i prognozowanie ekonometryczne w logistyce przedsiębiorstwa, [w:] Systemy Logistyczne Wojsk, „Zeszyty Naukowe Instytutu Logistyki Wydziału Logistyki Wojskowej Akademii Technicznej”, nr 44/2016, s. 27.

Dla lepszego wyjaśnienia powyższego algorytmu został on przedstawiony zarówno w formie teoretycznej jak i w empirycznej. Forma teoretyczna opierała się na wyselekcjonowaniu odpowiednich wzorów, ich objaśnieniu oraz pokazaniu metodycznego użycia. Przykład zastosowania powyższego algorytmu w ujęciu empirycznym polegał na dobraniu zmiennych objaśniających do modelu ekonometrycznego, który w późniejszym etapie mógłby posłużyć do analizy przyczynowo-skutkowej wybranego procesu logistycznego, a także do prognozowania. Do przykładu wykorzystano zgromadzone dane empiryczne (tabela 1) zawierające 57 kwartałów (obserwacji) pochodzących ze wszystkich dostępnych sprawozdań finansowych⁸ spółki Wawel S.A. od pierwszego kwartału 2004 roku do pierwszego kwartału 2018 roku⁹.

Przedsiębiorstwo Wawel to jedna z najstarszych fabryk cukierniczych w Polsce szczycąca się tradycjami sięgającymi końca XIX wieku. Momentem przełomowym dla firmy było przekształcenie jej w spółkę publiczną. Akcje Zakładu Przemysłu Cukierniczego Wawel S.A. zadebiutowały 11 marca 1998 roku na Warszawskiej Giełdzie Papierów Wartościowych. W 2005 roku firma zmieniła nazwę na Wawel S.A.¹⁰.

Dobór merytoryczny zmiennych objaśniających kandydujących do modelu ekonometrycznego

Zmienną objaśnianą y_t w analizowanym przypadku jest wielkość produkcji sprzedanej przedsiębiorstwa Wawel S.A. w latach 2004÷2018 (tabela 1).

Po merytorycznej analizie sprawozdań finansowych firmy Wawel S.A. w aspekcie wielkości produkcji sprzedanej, autor za najbardziej reprezentatywne zmienne, wpływające na wielkość tej produkcji, uznał: koszty wytworzenia, koszty sprzedaży, koszty zarządu oraz pozostałe koszty. Trzy pierwsze wielkości, występują w sprawozdaniach finansowych pod tymi samymi nazwami, natomiast za pozostałe koszty autor przyjął sumę strat z działalności operacyjnej i finansowej. Reasumując, jako zmienne kandydujące do modelu ekonometrycznego w rozważanym przykładzie wybrano cztery zmienne objaśniające, którym przypisano następujące oznaczenia:

- x_{1t} – koszty wytworzenia,
- x_{2t} – koszty sprzedaży,
- x_{3t} – koszty zarządu,
- x_{4t} – pozostałe koszty.

Na tej podstawie, poniżej zbudowano tabelę główną (tabela 1) z danymi empirycznymi dotyczącymi produkcji sprzedanej i jej kosztów w przedsiębiorstwie Wawel S.A. w latach 2004÷2018.

⁸ Spółki akcyjne notowane na Giełdzie Papierów Wartościowych mają obowiązki informacyjne obejmujące swoim zakresem konieczność przekazywania do publicznej wiadomości informacji poufnych, raportów bieżących i okresowych. <https://www.gpw.pl/obowiazki-informacyjne> (dostęp: 21.05.2018).

⁹ <https://www.gpw.pl/spolka?isin=PLWAWEL00013#reportsTab1> (dostęp: 20.05.2018).

¹⁰ <https://www.wawel.com.pl/o-firmie/historia/> (dostęp: 25.05.2018).

Tabela 1. Dane empiryczne produkcji sprzedanej i jej kosztów firmy Wawel S.A. w latach 2004-2018

t	y_t produkcja sprzedana	x_{1t} koszty wytworzenia	x_{2t} koszty sprzedaży	x_{3t} koszty zarządu	x_{4t} pozostałe koszty
1	50619,0	32915,0	8782,0	5339,0	285,0
2	39734,0	25269,0	8862,0	3924,0	0,0
3	46731,0	29909,0	7719,0	5166,0	388,0
4	61968,0	37040,0	10671,0	5759,0	0,0
5	52113,0	30324,0	10808,0	6315,0	349,0
6	44997,0	26138,0	10724,0	4781,0	314,0
7	52298,0	29271,0	11358,0	4685,0	314,0
8	68213,0	35374,0	16067,0	5005,0	568,0
9	60238,0	32704,0	13108,0	6201,0	548,0
10	50612,0	27673,0	13524,0	6061,0	69,0
11	56182,0	33604,0	12997,0	5681,0	0,0
12	72190,0	39861,0	17208,0	6022,0	292,0
13	64603,0	33712,0	15333,0	5934,0	291,0
14	46134,0	26474,0	12495,0	4773,0	99,0
15	60372,0	33659,0	13980,0	6226,0	299,0
16	74223,0	38404,0	17855,0	7005,0	295,0
17	64861,0	34229,0	16062,0	6072,0	113,0
18	46340,0	26151,0	13145,0	4492,0	151,0
19	63412,0	33232,0	15598,0	5854,0	450,0
20	81067,0	40090,0	19785,0	7254,0	612,0
21	71881,0	40317,0	16763,0	5915,0	107,0
22	58869,0	35476,0	15884,0	4652,0	143,0
23	76052,0	42380,0	17835,0	5962,0	165,0
24	107522,0	58945,0	23926,0	6772,0	0,0
25	89330,0	47933,0	19329,0	7511,0	105,0
26	71964,0	40557,0	18189,0	5304,0	242,0
27	89994,0	52942,0	17192,0	6529,0	266,0
28	126571,0	69150,0	25729,0	9651,0	267,0
29	108208,0	60538,0	21125,0	7543,0	439,0
30	85319,0	51415,0	20010,0	6002,0	772,0
31	120923,0	72349,0	22963,0	7625,0	242,0
32	159983,0	96945,0	30623,0	8696,0	282,0
33	149978,0	90155,0	27927,0	9721,0	511,0
34	109191,0	70479,0	21122,0	5868,0	1042,0
35	133026,0	81075,0	23739,0	8274,0	640,0
36	163702,0	101847,0	23592,0	11309,0	427,0
37	148037,0	88875,0	23995,0	8808,0	0,0
38	108703,0	68557,0	19198,0	6163,0	351,0
39	143941,0	89433,0	21128,0	9114,0	872,0
40	194964,0	120259,0	29352,0	12629,0	1313,0
41	149352,0	90054,0	22865,0	10638,0	0,0
42	105952,0	66511,0	18702,0	8956,0	440,0
43	134022,0	77699,0	21933,0	10532,0	57,0
44	204835,0	116979,0	31231,0	16109,0	0,0
45	170511,0	99881,0	22515,0	12928,0	947,0
46	108447,0	65310,0	19236,0	9922,0	0,0
47	141271,0	85774,0	20963,0	11388,0	0,0
48	210959,0	129117,0	29890,0	16082,0	48,0
49	177979,0	112696,0	23396,0	10793,0	133,0
50	110001,0	67677,0	21418,0	12954,0	67,0
51	151133,0	92738,0	23386,0	9451,0	165,0

t	y_t produkcja sprzedana	x_{1t} koszty wytworzenia	x_{2t} koszty sprzedaży	x_{3t} koszty zarządu	x_{4t} pozostałe koszty
52	206783,0	125340,0	31398,0	15528,0	79,0
53	174829,0	94489,0	11571,0	11050,0	0,0
54	92316,0	69372,0	35825,0	8622,0	179,0
55	146042,0	84277,0	24508,0	11852,0	205,0
56	205269,0	117905,0	29942,0	21195,0	551,0
57	159277,0	95857,0	13619,0	17689,0	163,0

Źródło: opracowanie własne na podstawie <https://www.gpw.pl> (dostęp: 20.05.2018).

Zebrane dane empiryczne umieszczone w tabeli 1, obrazują w tysiącach złotych kwartalne: wielkości produkcji sprzedanej – przychody ze sprzedaży, koszty wytworzenia, koszty sprzedaży, koszty zarządu oraz pozostałe koszty w latach 2004÷2018.

Obliczenie współczynników korelacji liniowej Pearsona pomiędzy zmienną objaśnianą y_t a zmiennymi objaśniającymi x_{it} oraz współczynników korelacji pomiędzy zmiennymi objaśniającymi x_{it}

Przy doborze zmiennych objaśniających do modelu ekonometrycznego najważniejsze jest to, aby współczynniki korelacji liniowej Pearsona¹¹ były jak najmniejsze pomiędzy zmiennymi objaśniającymi x_{it} , a najlepiej, aby dążyły do zera, i jak największe w swej wartości pomiędzy zmienną objaśnianą y_t a zmiennymi objaśniającymi x_{it} .

Należy dodać, że współczynnik korelacji liniowej Pearsona jest liczbą¹², która wskazuje, w jakim stopniu dwa zdarzenia są ze sobą powiązane i z jaką siłą wywołują wzajemne zmiany¹³. Dlatego w analizie statystycznej można przyjąć niższą skalę (tabela 2), do określania i opisu charakteryzowanej zależności korelacyjnej pomiędzy dwiema analizowanymi zmiennymi. Ponadto, cechą charakterystyczną tego współczynnika jest jego symetryczność¹⁴. Oznacza to, że jego wartość nie zależy od tego, którą zmienną uznamy za objaśnianą, a którą za objaśniającą.

Tabela 2. Określenia wartości siły związku – korelacji

Wartość siły związku lub wartość współczynnika korelacji	Siła związku między zmiennymi
$r = 0$	Brak
$0 < r < 0,1$	Nikła
$0,1 \leq r < 0,3$	Słaba
$0,3 \leq r < 0,5$	Przeciętna
$0,5 \leq r < 0,7$	Wysoka
$0,7 \leq r < 0,9$	Bardzo wysoka
$0,9 \leq r < 1,0$	Niemal pełna

¹¹Rozważa się obliczanie współczynnika korelacji liniowej Pearsona przede wszystkim wtedy, gdy dwie cechy są ilościowe oraz zależność między nimi jest zbliżona do liniowej. M. Piłatowska, *Repetitorium ze statystyki*, Wydawnictwo Naukowe PWN, Warszawa 2008, s. 97.

¹² Współczynnik korelacji liniowej Pearsona przyjmuje wartości od **-1** do **1**, czyli z przedziału zamkniętego **(-1, 1)**.

¹³ J. D. Łaniec, *Elementy...*, wyd. cyt., s. 161.

¹⁴ J. B. Buttolph-Johnson, H. T. Reynolds, J. D. Mycoff, *Metody badawcze w naukach politycznych*, Wydawnictwo Naukowe PWN, Warszawa 2010, s. 518.

Wartość siły związku lub wartość współczynnika korelacji	Siła związku między zmiennymi
$ r = 1,0$	Pełna

Źródło: opracowanie własne na podstawie A. Góralski, *Metody opisu i wnioskowania statystycznego w psychologii i pedagogice*, PWN, Warszawa 1987, s. 38.

„Należy pamiętać, że współczynniki korelacji liniowej Pearsona można obliczać tylko pomiędzy dwiema zmiennymi mierzalnymi typu ilościowego, tzn. tylko zmiennymi, którym można przypisać jednostki miary takie jak np.: kg, metry, sekundy, zł itp. Obliczając współczynniki korelacji Pearsona dla szeregów liczbowych utworzonych z wielu zmiennych, należy sprawdzić czy¹⁵:

- wszystkie zmienne są ilościowe,
- zmienne posiadają charakterystykę zbliżoną do rozkładu normalnego,
- związek między zmiennymi zachowuje liniową zależność”.

Spełnienie powyższych warunków, umożliwia zastosowanie do obliczeń współczynnika korelacji liniowej Pearsona. Współczynnik ten można zdefiniować, jako iloraz kowariancji dwóch zmiennych do iloczynu odchyłeń standardowych tych zmiennych¹⁶. Wzór definicyjny przedstawia się następująco (wzór 1):

$$r_{x_i x_j} = \frac{cov(x_i, x_j)}{s_{x_i} \cdot s_{x_j}} \quad i, j = 1, 2, \dots, m \quad (1)$$

gdzie:

$r_{x_i x_j}$ – współczynnik korelacji liniowej Pearsona pomiędzy zmiennymi x_i, x_j ,

$cov(x_i, x_j)$ – kowariancja zmiennych x_i, x_j ,

s_{x_i}, s_{x_j} – odchylenie standardowe odpowiednio zmiennych x_i, x_j .

Przystępując do oceny zależności korelacyjnej pomiędzy zmienną objaśnianą y_t , a kolejnymi potencjalnymi zmiennymi objaśniającymi $x_{1t}, x_{2t}, x_{3t}, x_{4t}$ oblicza się współczynniki korelacji liniowej Pearsona¹⁷ wykorzystując do tego celu wzór skalarny¹⁸ (wzór 2):

$$r_{yx_j} = r_{0j} = \frac{\sum_{t=1}^n (y_t - \bar{y})(x_{jt} - \bar{x}_j)}{\sqrt{\sum_{t=1}^n (y_t - \bar{y})^2 \sum_{t=1}^n (x_{jt} - \bar{x}_j)^2}} \quad j = 1, 2, \dots, m \quad (2)$$

gdzie:

r_{0j} – współczynnik korelacji liniowej Pearsona zmiennej objaśnianej y ze zmiennymi objaśniającymi x_j ,

y_t – zmienna objaśniana y , dla kolejnych obserwacji t ,

\bar{y} – średnia arytmetyczna zmiennych objaśnianych y_t ,

x_{jt} – zmienna objaśniająca x_j , dla kolejnych obserwacji t ,

¹⁵ M. Mróz, *Dobór...*, wyd. cyt., s. 131-132.

¹⁶ Tamże, s. 132.

¹⁷ Duda A., *Modelowanie...*, wyd. cyt., s. 27.

¹⁸ M. Mróz, *Dobór...*, wyd. cyt., s. 133.

\bar{x} – średnia arytmetyczna zmiennych objaśniających x_j ,

n – liczba wszystkich obserwacji t .

Aby przystąpić do praktycznych obliczeń współczynników korelacji liniowej Pearsona pomiędzy zmienną objaśnianą y_t a zmiennymi objaśniającymi x_{1t} , x_{2t} , x_{3t} , x_{4t} , dla ułatwienia tych obliczeń oraz metodycznego zaprezentowania procedury obliczeń pomiędzy poszczególnymi zmiennymi, należy utworzyć na bazie tabeli głównej z danymi empirycznymi (tabela 1) oraz wzoru 2 tabelę pomocnicze (tabela 3 i 4). Po utworzeniu tych tabeli pomocniczych, przystępujemy do obliczania kolejnych współczynników korelacji liniowej Pearsona. Obliczone poniżej współczynniki korelacji posłużą do utworzenia, w późniejszym etapie, głównego wektora R_0 (wzór 5).

Współczynnik korelacji liniowej Pearsona (wzór 2) między zmienną objaśnianą y_t a zmienną objaśniającą x_{1t} .

$$r_{01} = \frac{\sum_{t=1}^n (y_t - \bar{y})(x_{1t} - \bar{x}_1)}{\sqrt{\sum_{t=1}^n (y_t - \bar{y})^2 \sum_{t=1}^n (x_{1t} - \bar{x}_1)^2}} = \frac{8616790879,3}{\sqrt{14116768099,0 \cdot 5347373094,9}} = 0,99$$

Otrzymana wartość współczynnika korelacji r_{01} pomiędzy zmienną objaśnianą y_t a zmienną objaśniającą x_{1t} wynosi **0,99**, czyli można stwierdzić, że ta korelacja jest niemal pełna, bo zbliżona do jedności (tabela 2).

Współczynnik korelacji liniowej Pearsona (wzór 2) między zmienną objaśnianą y_t a zmienną objaśniającą x_{2t} .

$$r_{02} = \frac{\sum_{t=1}^n (y_t - \bar{y})(x_{2t} - \bar{x}_2)}{\sqrt{\sum_{t=1}^n (y_t - \bar{y})^2 \sum_{t=1}^n (x_{2t} - \bar{x}_2)^2}} = \frac{1411946830,0}{\sqrt{14116768099,0 \cdot 23705958010}} = 0,77$$

Otrzymana wartość współczynnika korelacji r_{02} pomiędzy zmienną objaśnianą y_t a zmienną objaśniającą x_{2t} wynosi **0,77**. Można stwierdzić, że ta korelacja jest bardzo wysoka (tabela 2).

Współczynnik korelacji liniowej Pearsona (wzór 2) między zmienną objaśnianą y_t a zmienną objaśniającą x_{3t} .

$$r_{03} = \frac{\sum_{t=1}^n (y_t - \bar{y})(x_{3t} - \bar{x}_3)}{\sqrt{\sum_{t=1}^n (y_t - \bar{y})^2 \sum_{t=1}^n (x_{3t} - \bar{x}_3)^2}} = \frac{91586181748}{\sqrt{14116768099,0 \cdot 7661935336}} = 0,88$$

Wartość współczynnika korelacji r_{03} pomiędzy zmienną objaśnianą y_t a zmienną objaśniającą x_{3t} wynosi **0,88**. Można stwierdzić, że ta korelacja jest bardzo wysoka (tabela 2).

Współczynnik korelacji liniowej Pearsona (wzór 2) między zmienną objaśnianą y_t a zmienną objaśniającą x_{4t} .

$$r_{04} = \frac{\sum_{t=1}^n (y_t - \bar{y})(x_{4t} - \bar{x}_4)}{\sqrt{\sum_{t=1}^n (y_t - \bar{y})^2 \sum_{t=1}^n (x_{4t} - \bar{x}_4)^2}} = \frac{93541726,4}{\sqrt{14116768079,0 \cdot 4579354,0}} = 0,12$$

Otrzymana wartość współczynnika korelacji r_{04} pomiędzy zmienną objaśnianą y_t a zmienną objaśniającą x_{4t} wynosi **0,12**, czyli można stwierdzić, że ta korelacja jest słaba (tabela 2).

Na podstawie przeprowadzonej powyżej cząstkowej analizy obliczonych współczynników korelacji można stwierdzić, że do celów ekonometrycznego wykorzystania najlepsze wyniki posiadają współczynniki korelacji r_{01} , r_{03} i r_{02} , ponieważ mają najsilniej skorelowaną zależność pomiędzy zmienną objaśnianą y_t a odpowiednimi zmiennymi objaśniającymi x_{it} .

Po obliczeniu wszystkich współczynników korelacji liniowej Pearsona pomiędzy zmienną objaśnianą y_t a zmiennymi objaśniającymi x_{1t} , x_{2t} , x_{3t} , x_{4t} , należy przystąpić do obliczenia współczynników korelacji pomiędzy kandydującymi zmiennymi objaśniającymi x_{1t} , x_{2t} , x_{3t} , x_{4t} (wzór 3¹⁹). Obliczenie tych współczynników korelacji posłuży do utworzenia, w późniejszym etapie, macierzy głównej R (wzór 7).

$$r_{x_i x_j} = r_{ij} = \frac{\sum_{t=1}^n (x_{it} - \bar{x}_i)(x_{jt} - \bar{x}_j)}{\sqrt{\sum_{t=1}^n (x_{it} - \bar{x}_i)^2 \sum_{t=1}^n (x_{jt} - \bar{x}_j)^2}} \quad i, j = 1, 2, \dots, m \quad (3)$$

gdzie:

- r_{ij} – współczynnik korelacji liniowej Pearsona pomiędzy zmiennymi objaśniającymi x_i , x_j ,
- x_{it} – zmienna objaśniająca x_i , dla kolejnych obserwacji t ,
- x_{jt} – zmienna objaśniająca x_j , dla kolejnych obserwacji t ,
- \bar{x}_i – średnia arytmetyczna zmiennych objaśniających x_i ,
- \bar{x}_j – średnia arytmetyczna zmiennych objaśniających x_j ,
- n – liczba wszystkich obserwacji t .

Współczynnik korelacji liniowej Pearsona (wzór 3) między zmiennymi objaśniającymi x_{1t} i x_{2t} .

¹⁹Tamże.

$$r_{12} = \frac{\sum_{t=1}^n (x_{1t} - \bar{x}_1)(x_{2t} - \bar{x}_2)}{\sqrt{\sum_{t=1}^n (x_{1t} - \bar{x}_1)^2 \sum_{t=1}^n (x_{2t} - \bar{x}_2)^2}} = \frac{885226382\cancel{3}}{\sqrt{534737309\cancel{4},9 \cdot 237059580\cancel{10}}} = 0,79$$

Otrzymana wartość współczynnika korelacji r_{12} pomiędzy zmiennymi objaśniającymi x_{1t} i x_{2t} wynosi **0,79**, czyli można stwierdzić, że ta korelacja jest bardzo wysoka (tabela 2).

Współczynnik korelacji liniowej Pearsona (wzór 3) między zmiennymi objaśniającymi x_{1t} i x_{3t} .

$$r_{13} = \frac{\sum_{t=1}^n (x_{1t} - \bar{x}_1)(x_{3t} - \bar{x}_3)}{\sqrt{\sum_{t=1}^n (x_{1t} - \bar{x}_1)^2 \sum_{t=1}^n (x_{3t} - \bar{x}_3)^2}} = \frac{553919224\cancel{87}}{\sqrt{534737309\cancel{4},9 \cdot 7661935336}} = 0,87$$

Wartość współczynnika korelacji r_{13} pomiędzy zmiennymi objaśniającymi x_{1t} i x_{3t} wynosi **0,87**, czyli można stwierdzić, że ta korelacja jest bardzo wysoka (tabela 2).

Współczynnik korelacji liniowej Pearsona (wzór 3) między zmiennymi objaśniającymi x_{1t} i x_{4t} .

$$r_{14} = \frac{\sum_{t=1}^n (x_{1t} - \bar{x}_1)(x_{4t} - \bar{x}_4)}{\sqrt{\sum_{t=1}^n (x_{1t} - \bar{x}_1)^2 \sum_{t=1}^n (x_{4t} - \bar{x}_4)^2}} = \frac{63786022,5}{\sqrt{534737309\cancel{4},9 \cdot 4579354,0}} = 0,13$$

Otrzymana wartość współczynnika korelacji r_{14} pomiędzy zmiennymi objaśniającymi x_{1t} i x_{4t} wynosi **0,13**, czyli można stwierdzić, że ta korelacja jest słaba (tabela 2).

Współczynnik korelacji liniowej Pearsona (wzór 3) między zmiennymi objaśniającymi x_{2t} i x_{3t} .

$$r_{23} = \frac{\sum_{t=1}^n (x_{2t} - \bar{x}_2)(x_{3t} - \bar{x}_3)}{\sqrt{\sum_{t=1}^n (x_{2t} - \bar{x}_2)^2 \sum_{t=1}^n (x_{3t} - \bar{x}_3)^2}} = \frac{85865224\cancel{38}}{\sqrt{237059580\cancel{10} \cdot 7661935336}} = 0,64$$

Wartość współczynnika korelacji r_{23} pomiędzy zmiennymi objaśniającymi x_{2t} i x_{3t} wynosi **0,64**, czyli można stwierdzić, że ta korelacja jest wysoka (tabela 2).

Współczynnik korelacji liniowej Pearsona (wzór 3) między zmiennymi objaśniającymi x_{2t} i x_{4t} .

$$r_{24} = \frac{\sum_{t=1}^n (x_{2t} - \bar{x}_2)(x_{4t} - \bar{x}_4)}{\sqrt{\sum_{t=1}^n (x_{2t} - \bar{x}_2)^2 \sum_{t=1}^n (x_{4t} - \bar{x}_4)^2}} = \frac{15046417,4}{\sqrt{237059580\cancel{10} \cdot 4579354,0}} = 0,14$$

Otrzymana wartość współczynnika korelacji r_{24} pomiędzy zmiennymi objaśniającymi x_{2t} i x_{4t} wynosi **0,14**, czyli można stwierdzić, że ta korelacja jest słaba (tabela 2).

Współczynnik korelacji liniowej Pearsona (wzór 3) między zmiennymi objaśniającymi x_{3t} i x_{4t} .

$$r_{34} = \frac{\sum_{t=1}^n (x_{3t} - \bar{x}_3)(x_{4t} - \bar{x}_4)}{\sqrt{\sum_{t=1}^n (x_{3t} - \bar{x}_3)^2 \sum_{t=1}^n (x_{4t} - \bar{x}_4)^2}} = \frac{736677,9}{\sqrt{7661935336 \cdot 4579354,0}} = 0,01$$

Wartość współczynnika korelacji r_{34} pomiędzy zmiennymi objaśniającymi x_{3t} i x_{4t} wynosi **0,01**, czyli można stwierdzić, że ta korelacja jest nikła (tabela 2).

Na podstawie przeprowadzonej cząstkowej analizy obliczonych poszczególnych współczynników korelacji można stwierdzić, że do celów ekonometrycznego wykorzystania najlepsze wyniki posiadają współczynniki korelacji r_{34}, r_{14} i r_{24} , ponieważ zmienne są najsłabiej skorelowane między sobą.

Posiadając obliczone wszystkie współczynniki korelacji pomiędzy zmiennymi objaśniającymi $x_{1t}, x_{2t}, x_{3t}, x_{4t}$ oraz posiadając wszystkiewcześniej obliczone współczynniki korelacji między zmienną objaśnianą y_t a zmiennymi objaśniającymi $x_{1t}, x_{2t}, x_{3t}, x_{4t}$, możemy przystąpić do kolejnego etapu jakim jest wyznaczenie głównego wektora R_0 (wzór 5) oraz macierzy głównej R (wzór 7).

Tabela 3. Tabela pomocnicza do obliczenia współczynników korelacji liniowej Pearsona cz. 1

t	y_t	x_{1t}	x_{2t}	x_{3t}	x_{4t}	$y_t - \bar{y}$	$(y_t - \bar{y})^2$	$x_{1t} - \bar{x}_1$	$(x_{1t} - \bar{x}_1)^2$	$x_{2t} - \bar{x}_2$	$(x_{2t} - \bar{x}_2)^2$	$x_{3t} - \bar{x}_3$	$(x_{3t} - \bar{x}_3)^2$	$x_{4t} - \bar{x}_4$	$(x_{4t} - \bar{x}_4)^2$
1	50619,0	32915,0	8782,0	5339,0	285,0	-56820,4	3228552273,8	-30547,0	933121352,7	-10658,4	113600443,4	-3122,2	9747869,9	-7,2	52,2
2	39734,0	25269,0	8862,0	3924,0	0,0	-67705,4	4584014537,4	-38193,0	1458707929,2	-10578,4	111901507,3	-4537,2	20585801,8	-292,2	85397,2
3	46731,0	29909,0	7719,0	5166,0	388,0	-60708,4	3685503866,2	-33553,0	1125806163,6	-11721,4	137390066,4	-3295,2	10858065,6	95,8	9172,3
4	61968,0	37040,0	10671,0	5759,0	0,0	-45471,4	2067643750,6	-26422,0	698123938,2	-8769,4	76901514,8	-2702,2	7301657,3	-292,2	85397,2
5	52113,0	30324,0	10808,0	6315,0	349,0	-55326,4	3061005101,4	-33138,0	1098129369,5	-8632,4	74517481,7	-2146,2	4605993,7	56,8	3223,1
6	44997,0	26138,0	10724,0	4781,0	314,0	-62442,4	3899047183,1	-37324,0	1393083595,2	-8716,4	75974772,6	-3680,2	13543562,1	21,8	474,0
7	52298,0	29271,0	11358,0	4685,0	314,0	-55141,4	3040568576,6	-34191,0	1169026880,4	-8082,4	65324395,7	-3776,2	14259368,4	21,8	474,0
8	68213,0	35374,0	16067,0	5005,0	568,0	-39226,4	1538706603,1	-28088,0	788937715,1	-3373,4	11379496,1	-3456,2	11945027,4	275,8	76050,2
9	60238,0	32704,0	13108,0	6201,0	548,0	-47201,4	2227967524,6	-30758,0	946056722,5	-6332,4	40098667,6	-2260,2	5108313,7	255,8	65419,3
10	50612,0	27673,0	13524,0	6061,0	69,0	-56827,4	3229347807,7	-35789,0	1280855032,5	-5916,4	35003207,7	-2400,2	5760757,9	-223,2	49830,8
11	56182,0	33604,0	12997,0	5681,0	0,0	-51257,4	2627316018,9	-29858,0	891502259,3	-6443,4	41516770,5	-2780,2	7729277,9	-292,2	85397,2
12	72190,0	39861,0	17208,0	6022,0	292,0	-35249,4	1242516737,3	-23601,0	557008857,2	-2232,4	4983390,4	-2439,2	5949491,2	-0,2	0,1
13	64603,0	33712,0	15333,0	5934,0	291,0	-42836,4	1834952956,5	-29750,0	885064587,7	-4107,4	16870331,2	-2527,2	6386527,0	-1,2	1,5
14	46134,0	26474,0	12495,0	4773,0	99,0	-61305,4	3758346046,2	-36988,0	1368114739,7	-6945,4	48237898,8	-3688,2	13602508,7	-193,2	37337,1
15	60372,0	33659,0	13980,0	6226,0	299,0	-47067,4	2215335518,6	-29803,0	888220900,4	-5460,4	29815431,7	-2235,2	4995930,8	6,8	45,9
16	74223,0	38404,0	17855,0	7005,0	295,0	-33216,4	1103325965,6	-25058,0	627905122,5	-1585,4	2513337,4	-1456,2	2120395,8	2,8	7,7
17	64861,0	34229,0	16062,0	6072,0	113,0	-42578,4	1812915963,4	-29233,0	854570340,4	-3378,4	11413254,6	-2389,2	5708075,4	-179,2	32122,7
18	46340,0	26151,0	13145,0	4492,0	151,0	-61099,4	3733130677,6	-37311,0	139211339,3	-6295,4	39631442,7	-3969,2	15754214,4	-141,2	19945,4
19	63412,0	33232,0	15598,0	5854,0	450,0	-44027,4	1938407625,3	-30230,0	913855021,4	-3842,4	14763660,3	-2607,2	6797272,3	157,8	24892,0
20	81067,0	40090,0	19785,0	7254,0	612,0	-26372,4	695500890,8	-23372,0	546252024,1	344,6	118783,0	-1207,2	1457230,2	319,8	102254,1
21	71881,0	40317,0	16763,0	5915,0	107,0	-35558,4	1264396317,1	-23145,0	535692649,2	-2677,4	7168207,7	-2546,2	6482920,0	-185,2	34309,4
22	58869,0	35476,0	15884,0	4652,0	143,0	-48570,4	2359078984,3	-27986,0	783218159,9	-3556,4	12647631,6	-3809,2	14509683,9	-149,2	22269,0
23	76052,0	42380,0	17835,0	5962,0	165,0	-31387,4	985165795,1	-21082,0	444452203,4	-1605,4	2577151,4	-2499,2	6245790,2	-127,2	16187,0
24	107522,0	58945,0	23926,0	6772,0	0,0	82,6	6830,9	-4517,0	20403606,0	4485,6	20121048,1	-1689,2	2853254,4	-292,2	85397,2
25	89330,0	47933,0	19329,0	7511,0	105,0	-18109,4	327948589,2	-15529,0	241150930,8	-111,4	12399,0	-950,2	902800,0	-187,2	35054,4
26	71964,0	40557,0	18189,0	5304,0	242,0	-35475,4	1258500519,9	-22905,0	524640632,4	-1251,4	1565879,0	-3157,2	9967646,0	-50,2	2522,9
27	89994,0	52942,0	17192,0	6529,0	266,0	-17445,4	304340267,2	-10520,0	110671138,2	-2248,4	5055081,7	-1932,2	3733234,1	-26,2	687,9
28	126571,0	69150,0	25729,0	9651,0	267,0	19131,6	366019998,2	5688,0	32352944,8	6288,6	39547107,8	1189,8	1415724,2	-25,2	636,5
29	108208,0	60538,0	21125,0	7543,0	439,0	768,6	590821,5	-2924,0	8549981,2	1684,6	2838042,7	-918,2	843013,9	146,8	21542,0
30	85319,0	51415,0	20010,0	6002,0	772,0	-22120,4	489309922,9	-12047,0	145131054,4	569,6	324500,1	-2459,2	6047457,6	479,8	230181,1
31	120923,0	72349,0	22963,0	7625,0	242,0	13483,6	181808793,7	8887,0	78978145,4	3522,6	12409056,8	-836,2	699160,0	-50,2	2522,9
32	159983,0	96945,0	30623,0	8696,0	282,0	52543,6	2760835063,1	33483,0	1121108939,3	11182,6	125051641,4	234,8	55150,8	-10,2	104,6
33	149978,0	90155,0	27927,0	9721,0	511,0	42538,6	1809536669,2	26693,0	712514375,8	8486,6	72023213,3	1259,8	1587202,1	218,8	47861,2
34	109191,0	70479,0	21122,0	5868,0	1042,0	1751,6	3068274,6	7017,0	49237796,6	1681,6	2827943,8	-2593,2	6724467,9	749,8	562157,9
35	133026,0	81075,0	23739,0	8274,0	640,0	25586,6	654676613,3	17613,0	310216533,0	4298,6	18478384,3	-187,2	35028,1	347,8	120945,3
36	163702,0	101847,0	23592,0	11309,0	427,0	56262,6	3165485686,3	38385,0	1473405531,3	4151,6	17236190,4	2847,8	8110204,7	134,8	18163,5
37	148037,0	88875,0	23995,0	8808,0	0,0	40597,6	1648169114,3	25413,0	645818785,6	4554,6	20744828,6	346,8	120299,4	-292,2	85397,2
38	108703,0	68557,0	19198,0	6163,0	351,0	1263,6	1596809,1	5095,0	25958667,5	-242,4	58733,9	-2298,2	5281529,7	58,8	3454,1
39	143941,0	89433,0	21128,0	9114,0	872,0	36501,6	1332370388,7	25971,0	674491018,5	1687,6	2848159,6	652,8	426202,8	579,8	336135,5
40	194964,0	120259,0	29352,0	12629,0	1313,0	87524,6	7660564204,1	56797,0	3225895223,2	9911,6	98240788,3	4167,8	17370907,8	1020,8	1041975,3
41	149352,0	90054,0	22865,0	10638,0	0,0	41912,6	1756670156,5	26592,0	707132597,9	3424,6	11728221,6	2176,8	4738641,6	-292,2	85397,2
42	105952,0	66511,0	18702,0	8956,0	440,0	-1487,4	2212212,6	3049,0	9296187,0	-738,4	545162,0	494,8	244868,7	147,8	21836,5

t	y_t	x_{1t}	x_{2t}	x_{3t}	x_{4t}	$y_t - \bar{y}$	$(y_t - \bar{y})^2$	$x_{1t} - \bar{x}_1$	$(x_{1t} - \bar{x}_1)^2$	$x_{2t} - \bar{x}_2$	$(x_{2t} - \bar{x}_2)^2$	$x_{3t} - \bar{x}_3$	$(x_{3t} - \bar{x}_3)^2$	$x_{4t} - \bar{x}_4$	$(x_{4t} - \bar{x}_4)^2$
43	134022,0	77699,0	21933,0	10532,0	57,0	26582,6	706637234,4	14237,0	202691169,9	2492,6	6213299,6	2070,8	4288387,0	-235,2	55332,2
44	204835,0	116979,0	31231,0	16109,0	0,0	97395,6	9485912468,1	53517,0	2864065533,4	11790,6	139019406,7	7647,8	58489488,9	-292,2	85397,2
45	170511,0	99881,0	22515,0	12928,0	947,0	63071,6	3978032923,1	36419,0	1326341005,3	3074,6	9453467,2	4466,8	19952678,4	654,8	428726,3
46	108447,0	65310,0	19236,0	9922,0	0,0	1007,6	1015356,8	1848,0	3414974,3	-204,4	41759,3	1460,8	2134059,7	-292,2	85397,2
47	141271,0	85774,0	20963,0	11388,0	0,0	33831,6	1144580482,4	22312,0	497823778,2	1522,6	2318460,4	2926,8	8566404,7	-292,2	85397,2
48	210959,0	129117,0	29890,0	16082,0	48,0	103519,6	10716317754,5	65655,0	4310574417,6	10449,6	109195166,8	7620,8	58077234,4	-244,2	59647,4
49	177979,0	112696,0	23396,0	10793,0	133,0	70539,6	4975842098,4	49234,0	2423983301,0	3955,6	15647160,0	2331,8	5437487,6	-159,2	25353,6
50	110001,0	67677,0	21418,0	12954,0	67,0	2561,6	6562046,2	4215,0	17765929,2	1977,6	3911096,1	4492,8	20185630,2	-225,2	50727,7
51	151133,0	92738,0	23386,0	9451,0	165,0	43693,6	1909134973,7	29276,0	857082121,5	3945,6	15568147,0	989,8	979787,4	-127,2	16187,0
52	206783,0	125340,0	31398,0	15528,0	79,0	99343,6	9869160621,0	61878,0	3828882541,7	11957,6	142985372,5	7066,8	49940257,3	-213,2	45466,2
53	174829,0	94489,0	11571,0	11050,0	0,0	67389,6	4541364808,9	31027,0	962672551,7	-7869,4	61926683,2	2588,8	6702103,4	-292,2	85397,2
54	92316,0	69372,0	35825,0	8622,0	179,0	-15123,4	228715741,8	5910,0	34927685,3	16384,6	268456726,9	160,8	25870,2	-113,2	12820,6
55	146042,0	84277,0	24508,0	11852,0	205,0	38602,6	1490164519,3	20815,0	433262764,3	5067,6	25681067,6	3390,8	11497810,2	-87,2	7608,7
56	205269,0	117905,0	29942,0	21195,0	551,0	97829,6	9570640247,5	54443,0	2964036428,4	10501,6	110284634,3	12733,8	162150734,8	258,8	66962,9
57	159277,0	95857,0	13619,0	17689,0	163,0	51837,6	2687141866,6	32395,0	1049433751,7	-5821,4	33888126,0	9227,8	85153069,9	-129,2	16699,9
Σ	6124043,0	3617336,0	1108100,0	482286,0	16657,0	0,0	141167680799,0	0,0	53473730945,9	0,0	2370595801,0	0,0	766193533,6	0,0	4579354,0

Tabela 4. Tabela pomocnicza do obliczenia współczynników korelacji liniowej Pearsona cz. 2

t	$y_t - \bar{y}$	$x_{1t} - \bar{x}_1$	$x_{2t} - \bar{x}_2$	$x_{3t} - \bar{x}_3$	$x_{4t} - \bar{x}_4$	$(y_t - \bar{y})(x_{1t} - \bar{x}_1)$	$(y_t - \bar{y})(x_{2t} - \bar{x}_2)$	$(y_t - \bar{y})(x_{3t} - \bar{x}_3)$	$(y_t - \bar{y})(x_{4t} - \bar{x}_4)$	$(x_{1t} - \bar{x}_1)(x_{2t} - \bar{x}_2)$	$(x_{1t} - \bar{x}_1)(x_{3t} - \bar{x}_3)$	$(x_{1t} - \bar{x}_1)(x_{4t} - \bar{x}_4)$	$(x_{2t} - \bar{x}_2)(x_{3t} - \bar{x}_3)$	$(x_{2t} - \bar{x}_2)(x_{4t} - \bar{x}_4)$	$(x_{3t} - \bar{x}_3)(x_{4t} - \bar{x}_4)$
1	-56820,4	-30547,0	-10658,4	-3122,2	-7,2	1735693251,9	605611236,6	177402107,1	410701,5	325581018,2	95372666,8	220796,1	33277054,3	77039,3	22567,2
2	-67705,4	-38193,0	-10578,4	-4537,2	-292,2	2585872841,7	716210957,8	307189867,2	19785404,0	404019326,2	173287830,7	11161076,9	47995648,2	3091291,1	1325884,9
3	-60708,4	-33553,0	-11721,4	-3295,2	95,8	2036949427,1	711583881,8	200043601,7	-5814155,9	393286897,3	110562548,5	-3213438,9	38623701,9	-1122576,4	-315583,6
4	-45471,4	-26422,0	-8769,4	-2702,2	-292,2	1201445628,4	398754230,7	122870769,8	13288005,1	231704096,6	71396510,7	7721260,3	23696170,7	2562650,5	789646,4
5	-55326,4	-33138,0	-8632,4	-2146,2	56,8	1833406556,6	477596473,5	118739084,7	-3140983,7	286059146,3	71119455,6	-1881310,2	18526388,0	-490075,2	-121841,5
6	-62442,4	-37324,0	-8716,4	-3680,2	21,8	2330600495,1	544269439,8	229797710,5	-1359490,5	325329386,0	137358342,4	-812616,3	32077547,5	-189771,8	-80124,1
7	-55141,4	-34191,0	-8082,4	-3776,2	21,8	1885339862,6	445671745,6	208222447,4	-1200533,6	276343942,4	129110747,1	-744404,8	30520233,1	-175968,4	-82214,2
8	-39226,4	-28088,0	-3373,4	-3456,2	275,8	1101791119,8	132324245,1	135572462,3	-10817526,5	94750797,8	97076684,2	-7745891,6	11658833,3	-930275,5	-953111,3
9	-47201,4	-30758,0	-6332,4	-2260,2	255,8	1451820806,5	298895515,6	106682505,8	-12072780,6	194770670,5	69518015,8	-7867042,0	14312112,8	-1619637,6	-578084,9
10	-56827,4	-35789,0	-5916,4	-2400,2	-223,2	2033796054,5	336210547,2	136394614,8	12685459,9	211740489,1	85899335,1	7989117,2	14200176,3	1320695,6	535782,6
11	-51257,4	-29858,0	-6443,4	-2780,2	-292,2	1530443781,0	330269096,7	142503528,7	14978836,7	192385796,6	83010052,0	8725356,0	17913532,8	1882928,0	812440,2
12	-35249,4	-23601,0	-2232,4	-2439,2	-0,2	831921166,9	78688919,4	85978732,5	8039,3	52685791,4	57566651,1	5382,7	5445056,3	509,1	556,3
13	-42836,4	-29750,0	-4107,4	-2527,2	-1,2	1274382941,6	175943923,4	108254222,3	52606,0	122193832,7	75183036,0	36535,1	10379924,2	5044,1	3103,5
14	-61305,4	-36988,0	-6945,4	-3688,2	-193,2	2267564469,3	425787172,5	226103813,8	11845914,6	256894881,9	136417713,6	7147126,6	25615550,7	1342036,7	51655,6
15	-47067,4	-29803,0	-5460,4	-2235,2	6,8	1402749909,7	257004250,6	105202960,9	-318736,8	162735028,8	66614489,2	-201824,1	12204746,4	-36977,1	-15136,3
16	-33216,4	-25058,0	-1585,4	-1456,2	2,8	832336485,8	52659571,0	48368251,6	-92073,4	39725777,9	36488455,6	-69459,1	2308521,2	-4394,5	-4036,4
17	-42578,4	-29233,0	-3378,4	-2389,2	-179,2	1244694425,2	143844609,0	101726403,1	7631235,7	98759449,7	69842336,6	5239380,5	8071413,7	605495,3	428204,2
18	-61099,4	-37311,0	-6295,4	-3969,2	-141,2	2279680024,4	384641852,1	242512970,9	862894,4	234886057,5	148093389,5	5269365,5	24987241,6	889080,3	560556,5
19	-44027,4	-30230,0	-3842,4	-2607,2	157,8	1330948361,8	169168530,3	114786255,4	-6946280,1	116154401,8	78814474,6	-4769451,0	10017615,4	-606215,1	-411336,3
20	-26372,4	-23372,0	344,6	-1207,2	319,8	616375510,0	-9089207,6	31835591,6	-8433137,5	-8055151,4	28213736,7	-7473720,8	-416045,9	110209,1	-386015,2
21	-35558,4	-23145,0	-2677,4	-2546,2	-185,2	822999278,7	95202181,9	90537175,8	6586404,7	61967380,0	58930913,8	4287110,2	6816958,1	495920,5	471619,9
22	-48570,4	-27986,0	-3556,4	-3809,2	-149,2	1359291543,9	172733209,9	185012135,5	7248059,7	99528160,4	106603226,5	4176302,0	13546702,0	530707,4	568433,3
23	-31387,4	-21082,0	-1605,4	-2499,2	-127,2	661709232,5	50387711,3	78441945,7	3993352,1	33844063,5	52687334,4	2682226,6	4012025,3	204245,7	317963,0

Gospodarka Materiałowa i Logistyka nr 11/2018

t	$y_t - \bar{y}$	$x_{1t} - \bar{x}_1$	$x_{2t} - \bar{x}_2$	$x_{3t} - \bar{x}_3$	$x_{4t} - \bar{x}_4$	$(y_t - \bar{y})(x_{1t} - \bar{x}_1)$	$(y_t - \bar{y})(x_{2t} - \bar{x}_2)$	$(y_t - \bar{y})(x_{3t} - \bar{x}_3)$	$(y_t - \bar{y})(x_{4t} - \bar{x}_4)$	$(x_{1t} - \bar{x}_1)(x_{2t} - \bar{x}_2)$	$(x_{1t} - \bar{x}_1)(x_{3t} - \bar{x}_3)$	$(x_{1t} - \bar{x}_1)(x_{4t} - \bar{x}_4)$	$(x_{2t} - \bar{x}_2)(x_{3t} - \bar{x}_3)$	$(x_{2t} - \bar{x}_2)(x_{4t} - \bar{x}_4)$	$(x_{3t} - \bar{x}_3)(x_{4t} - \bar{x}_4)$
24	82,6	-4517,0	4485,6	-1689,2	-292,2	-373329,0	370735,0	-139607,4	-24152,4	-20261834,5	7629985,5	1320004,4	-7576969,6	-1310832,6	493619,4
25	-18109,4	-15529,0	-111,4	-950,2	-187,2	281220745,2	2016492,1	17206742,7	3390578,8	1729171,7	14755035,3	2907471,3	105800,9	20848,0	177896,2
26	-35475,4	-22905,0	-1251,4	-3157,2	-50,2	812564156,6	44392111,4	112001284,1	1781858,4	28662235,7	72314812,4	1150475,7	3950712,3	62852,9	158577,9
27	-17445,4	-10520,0	-2248,4	-1932,2	-26,2	183525703,3	39223269,9	33707172,4	457557,9	23652730,1	20326368,8	275920,2	4344168,9	58969,9	50676,8
28	19131,6	5688,0	6288,6	1189,8	-25,2	108820148,9	120312228,5	22763641,7	-482654,6	35769615,6	6767780,1	-143496,4	7482499,5	-158650,5	-30017,4
29	768,6	-2924,0	1684,6	-918,2	146,8	-2247557,0	1294904,1	-705741,3	112816,1	-4925973,1	2684725,9	-429166,3	-1546773,9	247259,2	-134759,8
30	-22120,4	-12047,0	569,6	-2459,2	479,8	266484643,2	-12600838,5	54397435,5	-10612723,4	-6862583,0	29625561,4	-5779829,3	-1400857,1	273301,7	-1179834,9
31	13483,6	8887,0	3522,6	-836,2	-50,2	119828716,6	47498164,8	-11274459,7	-677257,7	31305659,2	-7430905,9	-446375,1	-2945490,9	-176935,9	41998,6
32	52543,6	33483,0	11182,6	234,8	-10,2	1759317159,9	587577191,8	12339461,2	-537420,1	374428248,2	7863210,0	-342466,1	2626156,9	-114376,9	-2402,0
33	42538,6	26693,0	8486,6	1259,8	218,8	1135482668,5	361010589,3	53591981,3	9306262,4	226533827,3	33628921,1	5839671,4	10691837,9	1856640,6	275618,1
34	1751,6	7017,0	1681,6	-2593,2	749,8	12291260,4	2945659,2	-4542302,8	1313337,3	11800072,9	-18196098,0	5261123,3	-4360781,7	1260853,3	-1944277,0
35	25586,6	17613,0	4298,6	-187,2	347,8	450656753,2	109988026,8	-4788743,4	8898318,3	75711956,2	-3296405,4	6125294,8	-804526,1	1494949,5	-65088,3
36	56262,6	38385,0	4151,6	2847,8	134,8	2159639812,5	233582777,9	160227141,1	7582625,8	159360905,9	109314319,3	5173215,8	11823241,2	559525,8	383809,2
37	40597,6	25413,0	4554,6	346,8	-292,2	1031706632,7	184908047,0	14080974,1	-11863772,7	115747138,3	8814286,3	-7426381,7	1579744,1	-1330996,3	-101357,0
38	1263,6	5095,0	-242,4	-2298,2	58,8	6438247,9	-306246,5	-2904065,2	74267,1	-1234769,2	-11709033,8	299440,9	556960,6	-14243,4	-135067,2
39	36501,6	25971,0	1687,6	652,8	579,8	947983048,6	61601976,1	23829813,5	21162631,6	43829876,2	16954939,4	15057236,4	1101768,4	978451,6	378499,5
40	87524,6	56797,0	9911,6	4167,8	1020,8	4971134425,2	867513611,7	364788917,9	89342705,0	562951587,5	236720781,8	57976747,5	41310188,5	10117533,2	4254416,2
41	41912,6	26592,0	3424,6	2176,8	-292,2	1114539694,9	143536117,1	91237219,4	-12248052,6	91068149,3	57886508,9	-7770918,6	7454920,4	-1000778,6	-636134,4
42	-1487,4	3049,0	-738,4	494,8	147,8	-4534880,6	1098186,8	-736003,8	-219788,7	-2251205,9	1508756,2	450551,4	-365367,1	-109107,5	73123,8
43	26582,6	14237,0	2492,6	2070,8	-235,2	378456242,8	66261217,0	55048469,1	-6252985,3	35487758,1	29482506,4	-3348933,8	5161882,8	-586341,0	-487120,2
44	97395,6	53517,0	11790,6	7647,8	-292,2	5212319536,7	1148357924,9	744866546,2	-28461742,6	630999755,4	409289297,6	-15639159,4	90173022,8	-3445558,6	-2234914,1
45	63071,6	36419,0	3074,6	4466,8	654,8	2297004176,4	193923190,6	281731098,0	41297545,4	111975538,5	162677765,9	23846115,9	13733972,2	2013193,9	2924762,8
46	1007,6	1848,0	-204,4	1460,8	-292,2	1862100,2	-205914,0	1472016,3	-294463,4	-377633,3	2699585,0	-540027,2	-298524,4	59717,1	-426899,1
47	33831,6	22312,0	1522,6	2926,8	-292,2	754850568,2	51513730,9	99019895,1	-9886557,5	33973293,8	65303598,4	-6520182,4	4456553,6	-444960,8	-855305,4
48	103519,6	65655,0	10449,6	7620,8	-244,2	6796578930,9	1081744010,6	788906900,8	-25282404,1	686071346,5	500346121,0	-16034785,4	79635126,0	-2552097,6	-1861223,6
49	70539,6	49234,0	3955,6	2331,8	-159,2	3472946609,8	279030101,2	164487323,9	-11231892,2	194752290,1	114805832,4	-7839429,2	9223949,2	-629850,4	-371294,7
50	2561,6	4215,0	1977,6	4492,8	-225,2	10797261,2	5066043,1	11509085,0	-576955,3	8335721,7	18937171,8	-949328,4	8885265,2	-445422,1	-1011914,2
51	43693,6	29276,0	3945,6	989,8	-127,2	1279173738,6	172399808,3	43249813,6	-5559058,7	115512685,3	28978582,7	-3724724,5	3905569,6	-501997,3	-125935,7
52	99343,6	61878,0	11957,6	7066,8	-213,2	6147182834,7	1187916498,8	702045882,5	-21182854,6	739914992,9	437281807,8	-13194119,0	84502818,3	-2549706,4	-1506849,1
53	67389,6	31027,0	-7869,4	2588,8	-292,2	2090896278,8	-530312794,4	174461161,1	-19693147,1	-244162073,5	80323913,2	-9066950,1	-20372506,9	2299645,2	-756532,3
54	-15123,4	5910,0	16384,6	160,8	-113,2	-89378473,0	-247790797,7	-2432471,6	1712387,8	96832701,4	950571,2	-669173,9	2635341,5	-1855202,2	-18211,8
55	38602,6	20815,0	5067,6	3390,8	-87,2	803512787,0	195624681,0	130895488,0	-3367234,6	105482938,7	70580259,4	-1815649,2	17183598,0	-442041,3	-295776,6
56	97829,6	54443,0	10501,6	12733,8	258,8	5326136154,6	1027372648,9	1245747305,1	25315567,1	571740914,7	693268118,9	14088311,1	133726341,8	2717532,0	3295160,9
57	51837,6	32395,0	-5821,4	9227,8	-129,2	1679278824,5	-301765144,2	478349641,2	-6698879,4	-188582457,4	298935621,2	-4186638,8	-53718506,7	752281,9	-1192496,2
Σ	0,0	0,0	0,0	0,0	0,0	86167908797,3	14119468304,0	9158618174,8	93541726,4	8852263822,3	5539192248,7	63786022,5	858652243,8	15046417,4	736677,9

Wyznaczenie wektora głównego R_0 współczynników korelacji pomiędzy zmienną objaśnianą y_t a zmiennymi objaśniającymi x_{it} oraz macierzy głównej R współczynników korelacji między zmiennymi objaśniającymi x_{it}

Tworzenie głównej jednokolumnowej macierzy, czyli głównego wektora²⁰ R_0 (wzór 4), rozpoczynamy od wyznaczenia tego wektora na ogólnych oznaczeniach współczynników korelacji (zależności między zmienną objaśnianą a zmiennymi objaśniającymi), jako elementów składowych tego wektora. Następnie, po uproszczeniu oznaczeń (indeksów dolnych tych współczynników), podstawiamy wartości już wcześniej obliczonych współczynników korelacji (wzór 5).

$$R_0 = \begin{matrix} & y \\ x_1 & r_{yx_1} \\ x_2 & r_{yx_2} \\ x_3 & r_{yx_3} \\ x_4 & r_{yx_4} \end{matrix} \quad (4)$$

Dla uproszczenia przyjmuje się zapis, w którym zmienna objaśniana y będzie reprezentowana we współczynniku korelacji r_{0j} przez indeks dolny 0 , zaś poszczególne zmienne objaśniające x_j w indeksie dolnym będą reprezentowane kolejnymi liczbami $j = 1, 2, 3, 4$.

$$R_0 = \begin{bmatrix} r_{01} \\ r_{02} \\ r_{03} \\ r_{04} \end{bmatrix} = \begin{bmatrix} 0,99 \\ 0,77 \\ 0,88 \\ 0,12 \end{bmatrix} \quad (5)$$

gdzie:

R_0 – wektor współczynników korelacji zmiennej objaśnianej i kandydujących zmiennych objaśniających,

r_{01} – współczynnik korelacji zmiennej objaśnianej y_t ze zmienną objaśniającą x_{1t} ,

r_{02} – współczynnik korelacji zmiennej objaśnianej y_t ze zmienną objaśniającą x_{2t} ,

r_{03} – współczynnik korelacji zmiennej objaśnianej y_t ze zmienną objaśniającą x_{3t} ,

r_{04} – współczynnik korelacji zmiennej objaśnianej y_t ze zmienną objaśniającą x_{4t} .

Na podobnej zasadzie, jak powyżej opisano, tworzymy główną macierz R współczynników korelacji między kandydującymi zmiennymi objaśniającymi, rozpoczynając w rozpatrywanym przykładzie od wyznaczenia macierzy kwadratowej, symetrycznej²¹ ($r_{ij} = r_{ji}$) o rozmiarze 4×4 , dla czterech zmiennych kandydujących. Jej elementami składowymi są współczynniki korelacji z ogólnie oznaczonymi indeksami dolnymi (wzór 6):

²⁰ Wektorem w matematyce nazywamy jednokolumnową lub jednowierszową macierz.

²¹ Duda A., *Modelowanie...*, wyd. cyt., s. 28.

$$R = \begin{matrix} & x_1 & x_2 & x_3 & x_4 \\ \begin{matrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{matrix} & \begin{bmatrix} r_{x_1x_1} & r_{x_1x_2} & r_{x_1x_3} & r_{x_1x_4} \\ r_{x_2x_1} & r_{x_2x_2} & r_{x_2x_3} & r_{x_2x_4} \\ r_{x_3x_1} & r_{x_3x_2} & r_{x_3x_3} & r_{x_3x_4} \\ r_{x_4x_1} & r_{x_4x_2} & r_{x_4x_3} & r_{x_4x_4} \end{bmatrix} \end{matrix} \quad (6)$$

Następnie po uproszczeniu oznaczeń współczynników korelacji (wzór 7), gdzie zmienne objaśniające i we współczynnikach korelacji reprezentowane są kolejnymi liczbami $i, j = 1$, podstawiamy wartości już wcześniej obliczonych współczynników korelacji:

$$R = \begin{bmatrix} r_{11} & r_{12} & r_{13} & r_{14} \\ r_{21} & r_{22} & r_{23} & r_{24} \\ r_{31} & r_{32} & r_{33} & r_{34} \\ r_{41} & r_{42} & r_{43} & r_{44} \end{bmatrix} = \begin{bmatrix} 1,00 & 0,79 & 0,87 & 0,13 \\ 0,79 & 1,00 & 0,64 & 0,14 \\ 0,87 & 0,64 & 1,00 & 0,01 \\ 0,13 & 0,14 & 0,01 & 1,00 \end{bmatrix} \quad (7)$$

gdzie:

R – macierz współczynników korelacji między zmiennymi objaśniającymi x_{it} ,

r_{11} – współczynnik korelacji pomiędzy zmienną objaśniającą x_{1t} a x_{1t} ,

$r_{12} = r_{21}$ – współczynniki korelacji pomiędzy zmiennymi objaśniającymi x_{1t} i x_{2t} ,

$r_{13} = r_{31}$ – współczynniki korelacji pomiędzy zmiennymi objaśniającymi x_{1t} i x_{3t} ,

$r_{14} = r_{41}$ – współczynniki korelacji pomiędzy zmiennymi objaśniającymi x_{1t} i x_{4t} ,

r_{22} – współczynnik korelacji pomiędzy zmienną objaśniającą x_{2t} a x_{2t} ,

$r_{23} = r_{32}$ – współczynniki korelacji pomiędzy zmiennymi objaśniającymi x_{2t} i x_{3t} ,

$r_{24} = r_{42}$ – współczynniki korelacji pomiędzy zmiennymi objaśniającymi x_{2t} i x_{4t} ,

r_{33} – współczynnik korelacji pomiędzy zmienną objaśniającą x_{3t} a x_{3t} ,

$r_{34} = r_{43}$ – współczynniki korelacji pomiędzy zmiennymi objaśniającymi x_{3t} i x_{4t} ,

r_{44} – współczynnik korelacji pomiędzy zmienną objaśniającą x_{4t} a x_{4t} .

Należy zauważyć, że w kwadratowych macierzach symetrycznych współczynników korelacji, główna diagonalna (lewo skośna przekątna) przyjmuje zawsze wartości równe jedności, a pozostałe elementy macierzy są odpowiednio sobie równe po przeciwnej stronie tej diagonalnej.

Reasumując, spośród zmiennych objaśniających należy wybrać te, które w końcowym efekcie wniosą najwięcej informacji o modelowanym procesie logistycznym.

Ustalenie liczby kombinacji L zmiennych objaśniających x_{it} oraz wypisanie wszystkich kombinacji k_k zmiennych objaśniających x_{it}

Do ustalenia liczby wszystkich możliwych kombinacji zmiennych objaśniających x_{it} mogących wystąpić w przyszłym modelu ekonometrycznym, wykorzystujemy na podstawie założeń kombinatoryki zmodyfikowany wzór 8:

$$L = 2^m - 1 \quad (8)$$

gdzie:

L – liczba wszystkich możliwych kombinacji zmiennych objaśniających x_{it} ,

m – liczba zmiennych objaśniających x_{it} .

W matematyce, a konkretnie w kombinatoryce rozpatruje się wszystkie możliwe kombinacje przyjętych zmiennych rozpoczynając od pustego zbioru, a kończąc na pełnym zbiorze zmiennych. Natomiast w ekonometrii, aby uniknąć rozpatrywania pustego zbioru, we wzorze zastosowano poprawkę polegającą na dodaniu liczby -1 .

Z powyższego wynika, że dla czterech zaproponowanych zmiennych kandydujących do modelu ekonometrycznego można utworzyć $2^m - 1$ kombinacji, czyli zestawów zmiennych objaśniających. Gdy liczba rozpatrywanych zmiennych wynosi $m = 4$ otrzymujemy $L = 2^m - 1 = 2^4 - 1 = 15$ możliwych kombinacji k_k ²²:

$$\begin{array}{ll} k_1 = \{x_1\} & k_9 = \{x_2, x_4\} \\ k_2 = \{x_2\} & k_{10} = \{x_3, x_4\} \\ k_3 = \{x_3\} & k_{11} = \{x_1, x_2, x_3\} \\ k_4 = \{x_4\} & k_{12} = \{x_1, x_2, x_4\} \\ k_5 = \{x_1, x_2\} & k_{13} = \{x_1, x_3, x_4\} \\ k_6 = \{x_1, x_3\} & k_{14} = \{x_2, x_3, x_4\} \\ k_7 = \{x_1, x_4\} & k_{15} = \{x_1, x_2, x_3, x_4\} \\ k_8 = \{x_2, x_3\} & \end{array}$$

Należy zaobserwować, że z zaprezentowanych powyżej piętnastu kombinacji zmiennych objaśniających, możemy utworzyć alternatywnie cztery różne modele ekonometryczne z jedną zmienną, sześć różnych modeli ekonometrycznych z dwiema zmiennymi, cztery różne modele ekonometryczne z trzema zmiennymi lub jeden model ekonometryczny z czterema zmiennymi.

Dla wszystkich kombinacji k_k , utworzenie odpowiadających im cząstkowych wektorów R_{0k} oraz cząstkowych macierzy R_k oraz obliczanie indywidualnych pojemności nośników informacji $h_{k,j}$ na ich podstawie integralnych pojemności nośników informacji H_k

Obliczanie indywidualnych pojemności nośników informacji $h_{k,j}$ oraz na ich podstawie integralnych pojemności nośników informacji H_k dokonano na podstawie poniższych wzorów 9 i 10. „Dla każdej zmiennej objaśniającej, w każdej kombinacji oblicza się indywidualną pojemność nośnika informacji²³ $h_{k,j}$ według wzoru:

$$h_{k,j} = \frac{r_{0j}^2}{\sum |r_{ij}|} \quad (9)$$

gdzie:

²² Tamże.

²³ Tamże.

- indywidualna pojemność nośników informacji zmiennej w kombinacji,
- współczynnik korelacji zmiennej objaśnianej ze zmiennymi objaśniającymi ,
- współczynnik korelacji pomiędzy zmiennymi objaśniającymi ,
- suma bezwzględnych wartości współczynników korelacji .

Należy pamiętać, że według ogólnego wzoru 9 na indywidualną pojemność nośników informacji w mianowniku zawsze występuje suma wartości bezwzględnych odpowiednich współczynników korelacji. Niezmiernie ważnym jest przestrzeganie tego formalnego zapisu, ponieważ współczynniki korelacji liniowej Pearsona przyjmują wartości z przedziału , w związku z tym w mianowniku mogą występować także wartości ujemne tych współczynników. Nie stosując wartości bezwzględnych w powyższym wzorze 9 możemy uzyskać niewłaściwe wyniki pojemności indywidualnych, a co za tym idzie integralnych pojemności nośników informacji.

W ramach każdej kombinacji należy obliczyć integralną pojemność nośników informacji (wzór 10), jako sumy poszczególnych indywidualnych pojemności nośników informacji :

$$H_k = \sum h_{k,j} \quad (10)$$

gdzie:

- integralna pojemność nośników informacji kombinacji,
- indywidualna pojemność nośników informacji zmiennej w kombinacji.

Przystępując do obliczeń indywidualnych $h_{k,j}$ i integralnych H_k pojemności nośników informacji, dla wszystkich wypisanych powyżej kombinacji k_k , rozpoczynamy od utworzenia odpowiadających im cząstkowych wektorów R_{0_k} współczynników korelacji r_{0_j} oraz cząstkowych macierzy R_k współczynników korelacji r_{ij} .

Cząstkowy wektor R_{0_k} tworzymy na podstawie wektora głównego R_0 w ten sposób, że wykreślamy z niego wiersze współczynników korelacji r_{0_j} zmiennej y_t ze zmiennymi x_{it} , które nie występują w danej kombinacji. Zatem, w wektorze cząstkowym R_{0_k} pozostają jedynie współczynniki korelacji r_{0_j} , tych zmiennych objaśniających x_{it} , które zawiera dana kombinacja.

Macierz cząstkową R_k tworzymy na podstawie macierzy głównej R w ten sposób, że wykreślamy z niej wiersze i kolumny współczynników korelacji r_{ij} pomiędzy zmiennymi x_{it} , które nie występują w danej kombinacji. Zatem, w macierzy cząstkowej pozostają jedynie współczynniki korelacji r_{ij} , tych zmiennych objaśniających x_{it} , które zawiera dana kombinacja.

Poniżej przedstawiono obliczenia dla kombinacji od do , w których utworzono wektory cząstkowe i macierze cząstkowe (wykorzystując wektor główny – wzór 5

oraz macierz główną – wzór 7) i na ich podstawie obliczono indywidualne i integralne pojemności nośników informacji. Wyniki tych obliczeń, (a dokładniej obliczone wielkości integralnych pojemności nośników informacji), w późniejszym etapie posłużyły do wyboru, ze wszystkich kombinacji zmiennych kandydujących, najlepszych zmiennych objaśniających do tworzonego modelu ekonometrycznego.

Obliczenia rozpoczęto od pierwszej kombinacji:

$$k_1 = \{x_1\}$$

W kombinacji występuje zmienna, lecz nie występują zmienne objaśniające. Dlatego, w celu utworzenia cząstkowego wektora należy z wektora głównego wykreślić drugi, trzeci i czwarty wiersz odpowiadający korelacjom tych zmiennych ze zmienną objaśnianą, które nie biorą udziału w tej kombinacji. Po wykreśleniu, pozostaje wektor jednoelementowy:

$$R_0 = \begin{matrix} & y \\ x_1 & [r_{01}] \\ x_2 & [r_{02}] \\ x_3 & [r_{03}] \\ x_4 & [r_{04}] \end{matrix} \rightarrow R_{0_1} = [r_{01}]$$

Tworząc macierz cząstkową, należy wykreślić z głównej macierzy drugi, trzeci i czwarty wiersz oraz drugą, trzecią i czwartą kolumnę, które odpowiadają korelacjom zmiennych objaśniających, niebiorących udziału w tej kombinacji. Po wykreśleniu, pozostaje macierz jednoelementowa:

$$R = \begin{matrix} & x_1 & x_2 & x_3 & x_4 \\ x_1 & [r_{11}] & r_{12} & r_{13} & r_{14}] \\ x_2 & r_{21} & [r_{22}] & r_{23} & r_{24}] \\ x_3 & r_{31} & r_{32} & [r_{33}] & r_{34}] \\ x_4 & r_{41} & r_{42} & r_{43} & [r_{44}] \end{matrix} \rightarrow R_1 = [r_{11}]$$

Według podanej powyżej zasady tworzenia cząstkowego wektora i macierzy cząstkowej, uzyskano na podstawie obliczonych współczynników korelacji, z danych empirycznych (wzór 5 i 7) dla kombinacji, następujące wielkości:

$$R_{0_1} = [r_{01}] = [0,99]$$

$$R_1 = [r_{11}] = [1,00]$$

Dzięki wyznaczeniu wektora cząstkowego oraz macierzy cząstkowej, można przejść do obliczania indywidualnej (wzór 9) i integralnej (wzór 10) pojemności nośników informacji. Dużym ułatwieniem w tworzeniu wzoru szczegółowego indywidualnej pojemności nośników informacji (wzór 9) jest budowanie go według następującej zasady. W liczniku tego wzoru, wpisujemy symbol ogólny współczynnika korelacji (a później

jego wartość) z wektora k_1 . Natomiast w mianowniku wpisujemy symbol współczynnika korelacji r_{01} (a następnie jego wartość) z macierzy R_{01} . Należy pamiętać, że według ogólnego wzoru na indywidualną pojemność nośników informacji (wzór 9) w mianowniku zawsze występuje suma odpowiednich współczynników korelacji zapisanych w wartościach bezwzględnych. W tym przypadku, w kombinacji jednoelementowej, w mianowniku występuje wartość bezwzględna jednego współczynnika korelacji.

$$h_{1,1} = \frac{r_{01}^2}{|r_{11}|} = \frac{0,99^2}{|1,00|} = \frac{0,980}{1} = 0,980 \quad H_1 = h_{1,1} = 0,980$$

Na podstawie dokonanych obliczeń wartość integralnej pojemności nośnika informacji wynosi H_1 dla kombinacji k_1 .

Dla pozostałych kombinacji jednoelementowych k_2, k_3, k_4 wyznaczamy odpowiednio wektory k_2, k_3, k_4 , oraz cząstkowe macierze R_{02}, R_{03}, R_{04} , analogicznie jak wyznaczono dla k_1 , a także w podobny sposób obliczamy indywidualne i integralne pojemności nośników informacji.

$$k_2 = \{x_2\}$$

$$R_{0_2} = [r_{02}] = [0,77] \quad R_2 = [r_{22}] = [1,00]$$

$$h_{2,2} = \frac{r_{02}^2}{|r_{22}|} = \frac{0,77^2}{|1,00|} = \frac{0,593}{1} = 0,593 \quad H_2 = h_{2,2} = 0,593$$

Na podstawie dokonanych obliczeń wartość integralnej pojemności nośnika informacji wynosi H_2 .

$$k_3 = \{x_3\}$$

$$R_{0_3} = [r_{03}] = [0,88] \quad R_3 = [r_{33}] = [1,00]$$

$$h_{3,3} = \frac{r_{03}^2}{|r_{33}|} = \frac{0,88^2}{|1,00|} = \frac{0,774}{1} = 0,774 \quad H_3 = h_{3,3} = 0,774$$

Na podstawie dokonanych obliczeń wartość integralnej pojemności nośnika informacji wynosi H_3 .

$$k_4 = \{x_4\}$$

$$R_{0_4} = [r_{04}] = [0,12] \quad R_4 = [r_{44}] = [1,00]$$

$$h_{4,4} = \frac{r_{04}^2}{|r_{44}|} = \frac{0,12^2}{|1,00|} = \frac{0,014}{1} = 0,014 \quad H_4 = h_{4,4} = 0,014$$

Na podstawie dokonanych obliczeń, wartość integralnej pojemności nośnika informacji wynosi .

$$k_5 = \{x_1, x_2\}$$

W kombinacji występują zmienne , , lecz nie występują zmienne objaśniające , . Dlatego w celu utworzenia cząstkowego wektora należy z głównego wektora wykreślić trzeci i czwarty wiersz odpowiadający korelacjom tych zmiennych ze zmienną objaśnianą, które nie biorą udziału w tej kombinacji. Po wykreśleniu, pozostaje wektor dwuelementowy:

$$R_0 = \begin{matrix} & y \\ x_1 & r_{01} \\ x_2 & r_{02} \\ x_3 & r_{03} \\ x_4 & r_{04} \end{matrix} \rightarrow R_{0_5} = \begin{bmatrix} r_{01} \\ r_{02} \end{bmatrix}$$

Tworząc macierz cząstkową należy usunąć z głównej macierzy trzeci i czwarty wiersz oraz trzecią i czwartą kolumnę, które odpowiadają korelacjom zmiennych objaśniających , , które nie biorą udziału w tej kombinacji. Po wykreśleniu, pozostaje macierz kwadratowa o rozmiarze :

$$R = \begin{matrix} & x_1 & x_2 & x_3 & x_4 \\ x_1 & r_{11} & r_{12} & r_{13} & r_{14} \\ x_2 & r_{21} & r_{22} & r_{23} & r_{24} \\ x_3 & r_{31} & r_{32} & r_{33} & r_{34} \\ x_4 & r_{41} & r_{42} & r_{43} & r_{44} \end{matrix} \rightarrow R_5 = \begin{bmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \end{bmatrix}$$

Według podanej powyżej zasady tworzenia cząstkowego wektora i macierzy cząstkowej , uzyskano na podstawie obliczonych współczynników korelacji, z danych empirycznych (wzór 5 i 7) dla kombinacji $k_5 = \{x_1, x_2\}$, następujące wielkości:

$$R_{0_5} = \begin{bmatrix} r_{01} \\ r_{02} \end{bmatrix} = \begin{bmatrix} 0,99 \\ 0,77 \end{bmatrix} \quad R_5 = \begin{bmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \end{bmatrix} = \begin{bmatrix} 1,00 & 0,79 \\ 0,79 & 1,00 \end{bmatrix}$$

Dzięki wyznaczeniu wektora cząstkowego oraz macierzy cząstkowej , można przejść do obliczania indywidualnych (wzór 9) i integralnej (wzór 10) pojemności nośników informacji. Dużym ułatwieniem w tworzeniu dwóch wzorów szczegółowych indywidualnych pojemności nośników informacji (wzór 9) jest budowanie ich według następującej zasady. Dla pierwszej integralnej pojemności , w liczniku tego wzoru wpisujemy symbol ogólny współczynnika korelacji (a później jego wartość) z pierwszego wiersza wektora , natomiast w mianowniku wpisujemy symbole współczynników korelacji i (a następnie ich wartości) z pierwszego wiersza macierzy . Natomiast dla drugiej pojemności integralnej

w liczniku tego wzoru wpisujemy symbol ogólny współczynnika korelacji (a później jego wartość) z drugiego wiersza wektora , a w mianowniku wpisujemy symbole współczynników korelacji i (a następnie ich wartości) z drugiego wiersza macierzy . Należy pamiętać, że według ogólnego wzoru na indywidualną pojemność nośników informacji (wzór 9) w mianowniku zawsze występuje suma odpowiednich współczynników korelacji zapisanych w wartościach bezwzględnych.

$$h_{5,1} = \frac{r_{01}^2}{|r_{11}| + |r_{12}|}$$

$$h_{5,1} = \frac{0,99^2}{|1,00| + |0,79|} = \frac{0,980}{1,79} = 0,547$$

$$H_5 = h_{5,1} + h_{5,2}$$

$$h_{5,2} = \frac{r_{02}^2}{|r_{21}| + |r_{22}|}$$

$$H_5 = 0,547 + 0,331 = 0,878$$

$$h_{5,2} = \frac{0,77^2}{|0,79| + |1,00|} = \frac{0,583}{1,79} = 0,331$$

Na podstawie dokonanych obliczeń, wartość integralnej pojemności nośnika informacji wynosi , dla kombinacji $k_5 = :$

Dla pozostałych kombinacji dwuelementowych k_i wyznaczamy odpowiednio wektory , , , , oraz cząstkowe macierze , , , , , analogicznie jak wyznaczono dla , a także w podobny sposób obliczamy indywidualne i integralne pojemności nośników informacji.

$$k_6 = \{x_1, x_3\}$$

$$R_{0_6} = \begin{bmatrix} r_{01} \\ r_{03} \end{bmatrix} = \begin{bmatrix} 0,99 \\ 0,88 \end{bmatrix}$$

$$R_6 = \begin{bmatrix} r_{11} & r_{13} \\ r_{31} & r_{33} \end{bmatrix} = \begin{bmatrix} 1,00 & 0,87 \\ 0,87 & 1,00 \end{bmatrix}$$

$$h_{6,1} = \frac{r_{01}^2}{|r_{11}| + |r_{13}|}$$

$$h_{6,1} = \frac{0,99^2}{|1,00| + |0,87|} = \frac{0,980}{1,87} = 0,524$$

$$H_6 = h_{6,1} + h_{6,3}$$

$$h_{6,3} = \frac{r_{03}^2}{|r_{31}| + |r_{33}|}$$

$$H_6 = 0,524 + 0,414 = 0,938$$

$$h_{6,3} = \frac{0,88^2}{|0,87| + |1,00|} = \frac{0,774}{1,87} = 0,414$$

Na podstawie dokonanych obliczeń wartość integralnej pojemności nośnika informacji wynosi .

$$k_7 = \{x_1, x_4\}$$

$$R_{0_7} = \begin{bmatrix} r_{01} \\ r_{04} \end{bmatrix} = \begin{bmatrix} 0,99 \\ 0,12 \end{bmatrix}$$

$$R_7 = \begin{bmatrix} r_{11} & r_{14} \\ r_{41} & r_{44} \end{bmatrix} = \begin{bmatrix} 1,00 & 0,13 \\ 0,13 & 1,00 \end{bmatrix}$$

$$h_{7,1} = \frac{r_{01}^2}{|r_{11}| + |r_{14}|}$$

$$h_{7,1} = \frac{0,99^2}{|1,00| + |0,13|} = \frac{0,980}{1,13} = 0,867$$

$$H_7 = h_{7,1} + h_{7,4}$$

$$h_{7,4} = \frac{r_{04}^2}{|r_{41}| + |r_{44}|}$$

$$H_7 = 0,867 + 0,012 = 0,879$$

$$h_{7,4} = \frac{0,12^2}{|0,13| + |1,00|} = \frac{0,014}{1,13} = 0,012$$

Na podstawie dokonanych obliczeń wartość integralnej pojemności nośnika informacji wynosi .

$$k_8 = \{x_2, x_3\}$$

$$R_{0_8} = \begin{bmatrix} r_{02} \\ r_{03} \end{bmatrix} = \begin{bmatrix} 0,77 \\ 0,88 \end{bmatrix}$$

$$R_8 = \begin{bmatrix} r_{22} & r_{23} \\ r_{32} & r_{33} \end{bmatrix} = \begin{bmatrix} 1,00 & 0,64 \\ 0,64 & 1,00 \end{bmatrix}$$

$$h_{8,2} = \frac{r_{02}^2}{|r_{22}| + |r_{23}|}$$

$$h_{8,2} = \frac{0,77^2}{|1,00| + |0,64|} = \frac{0,593}{1,64} = 0,362$$

$$H_8 = h_{8,2} + h_{8,3}$$

$$h_{8,3} = \frac{r_{03}^2}{|r_{32}| + |r_{33}|}$$

$$H_8 = 0,362 + 0,472 = 0,834$$

$$h_{8,3} = \frac{0,88^2}{|0,64| + |1,00|} = \frac{0,774}{1,64} = 0,472$$

Na podstawie dokonanych obliczeń wartość integralnej pojemności nośnika informacji wynosi .

$$k_9 = \{x_2, x_4\}$$

$$R_0 = \begin{matrix} & & & y \\ & & & x_1 \\ & & & x_2 \\ & & & x_3 \\ & & & x_4 \end{matrix} \begin{bmatrix} r_{01} \\ r_{02} \\ r_{03} \\ r_{04} \end{bmatrix} \rightarrow R_{0_{11}} = \begin{bmatrix} r_{01} \\ r_{02} \\ r_{03} \end{bmatrix}$$

Tworząc macierz cząstkową należy usunąć z głównej macierzy czwarty wiersz oraz czwartą kolumnę, które odpowiadają korelacjom zmiennej objaśniającej, która nie bierze udziału w tej kombinacji. Po wykreśleniu, pozostaje macierz kwadratowa o rozmiarze :

$$R = \begin{matrix} & x_1 & x_2 & x_3 & x_4 \\ x_1 & r_{11} & r_{12} & r_{13} & r_{14} \\ x_2 & r_{21} & r_{22} & r_{23} & r_{24} \\ x_3 & r_{31} & r_{32} & r_{33} & r_{34} \\ x_4 & r_{41} & r_{42} & r_{43} & r_{44} \end{matrix} \rightarrow R_{11} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix}$$

Według podanej powyżej zasady tworzenia cząstkowego wektora i macierzy cząstkowej, uzyskano na podstawie obliczonych współczynników korelacji, z danych empirycznych (wzór 5 i 7) dla kombinacji $k_{11} = \{x_1$, następujące wielkości:

$$R_{0_{11}} = \begin{bmatrix} r_{01} \\ r_{02} \\ r_{03} \end{bmatrix} = \begin{bmatrix} 0,99 \\ 0,77 \\ 0,88 \end{bmatrix} \quad R_{11} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} = \begin{bmatrix} 1,00 & 0,79 & 0,87 \\ 0,79 & 1,00 & 0,64 \\ 0,87 & 0,64 & 1,00 \end{bmatrix}$$

Dzięki wyznaczeniu wektora cząstkowego oraz macierzy cząstkowej, można przejść do obliczania indywidualnych (wzór 9) i integralnej (wzór 10) pojemności nośników informacji. Dużym ułatwieniem w tworzeniu trzech wzorów szczegółowych indywidualnych pojemności nośników informacji (wzór 9) jest budowanie ich według następującej zasady. Dla pierwszej integralnej pojemności, w liczniku tego wzoru wpisujemy symbol ogólny współczynnika korelacji (a później jego wartość) z pierwszego wiersza wektora, natomiast w mianowniku wpisujemy symbole współczynników korelacji, i (a następnie ich wartości) z pierwszego wiersza macierzy. Natomiast dla drugiej pojemności integralnej w liczniku tego wzoru wpisujemy symbol ogólny współczynnika korelacji (a później jego wartość) z drugiego wiersza wektora, a w mianowniku wpisujemy symbole współczynników korelacji, i (a następnie ich wartości) z drugiego wiersza macierzy. Także dla trzeciej pojemności integralnej w liczniku tego wzoru wpisujemy symbol ogólny współczynnika korelacji (a później jego wartość) z trzeciego wiersza wektora, a w mianowniku wpisujemy symbole współczynników korelacji,

i (a następnie ich wartości) z trzeciego wiersza macierzy . Należy pamiętać, że według ogólnego wzoru na indywidualną pojemność nośników informacji (wzór 9) w mianowniku zawsze występuje suma odpowiednich współczynników korelacji zapisanych w wartościach bezwzględnych.

$$h_{11,1} = \frac{r_{01}^2}{|r_{11}| + |r_{12}| + |r_{13}|}$$

$$h_{11,1} = \frac{0,99^2}{|1,00| + |0,79| + |0,87|} = \frac{0,980}{2,66} = 0,368$$

$$h_{11,2} = \frac{r_{02}^2}{|r_{21}| + |r_{22}| + |r_{23}|}$$

$$h_{11,2} = \frac{0,77^2}{|0,79| + |1,00| + |0,64|} = \frac{0,593}{2,43} = 0,244$$

$$h_{11,3} = \frac{r_{03}^2}{|r_{31}| + |r_{32}| + |r_{33}|}$$

$$h_{11,3} = \frac{0,88^2}{|0,87| + |0,64| + |1,00|} = \frac{0,774}{2,51} = 0,308$$

$$H_{11} = h_{11,1} + h_{11,2} + h_{11,3}$$

$$H_{11} = 0,368 + 0,244 + 0,308$$

$$H_{11} = 0,920$$

Na podstawie dokonanych obliczeń wartość integralnej pojemności nośnika informacji wynosi dla kombinacji $k_{11} = \{x_1\}$.

Dla pozostałych kombinacji trzejelementowych k_1 , wyznaczamy odpowiednio wektory , , oraz cząstkowe macierze , , , analogicznie jak wyznaczono dla , a także w podobny sposób obliczamy indywidualne i integralne pojemności nośników informacji.

$$k_{12} = \{x_1, x_2, x_4\}$$

$$R_{0,12} = \begin{bmatrix} r_{01} \\ r_{02} \\ r_{04} \end{bmatrix} = \begin{bmatrix} 0,99 \\ 0,77 \\ 0,12 \end{bmatrix}$$

$$R_{12} = \begin{bmatrix} r_{11} & r_{12} & r_{14} \\ r_{21} & r_{22} & r_{24} \\ r_{41} & r_{42} & r_{44} \end{bmatrix}$$

$$R_{12} = \begin{bmatrix} 1,00 & 0,79 & 0,13 \\ 0,79 & 1,00 & 0,14 \\ 0,13 & 0,14 & 1,00 \end{bmatrix}$$

$$h_{12,1} = \frac{r_{01}^2}{|r_{11}| + |r_{12}| + |r_{14}|}$$

$$H_{12} = h_{12,1} + h_{12,2} + h_{12,4}$$

$$H_{12} = 0,510 + 0,307 + 0,011$$

$$h_{12,1} = \frac{0,99^2}{|1,00| + |0,79| + |0,13|} = \frac{0,980}{1,92} = 0,510$$

$$H_{12} = 0,828$$

$$h_{12,2} = \frac{r_{02}^2}{|r_{21}| + |r_{22}| + |r_{24}|}$$

$$h_{12,2} = \frac{0,77^2}{|0,79| + |1,00| + |0,14|} = \frac{0,593}{1,93} = 0,307$$

$$h_{12,4} = \frac{r_{04}^2}{|r_{41}| + |r_{42}| + |r_{44}|}$$

$$h_{12,4} = \frac{0,12^2}{|0,13| + |0,14| + |1,00|} = \frac{0,014}{1,27} = 0,011$$

Na podstawie dokonanych obliczeń wartość integralnej pojemności nośnika informacji wynosi .

$$k_{13} = \{x_1, x_3, x_4\}$$

$$R_{0_{13}} = \begin{bmatrix} r_{01} \\ r_{03} \\ r_{04} \end{bmatrix} = \begin{bmatrix} 0,99 \\ 0,88 \\ 0,12 \end{bmatrix}$$

$$R_{13} = \begin{bmatrix} r_{11} & r_{13} & r_{14} \\ r_{31} & r_{33} & r_{34} \\ r_{41} & r_{43} & r_{44} \end{bmatrix}$$

$$R_{13} = \begin{bmatrix} 1,00 & 0,87 & 0,13 \\ 0,87 & 1,00 & 0,01 \\ 0,13 & 0,01 & 1,00 \end{bmatrix}$$

$$h_{13,1} = \frac{r_{01}^2}{|r_{11}| + |r_{13}| + |r_{14}|}$$

$$h_{13,1} = \frac{0,99^2}{|1,00| + |0,87| + |0,13|} = \frac{0,980}{2,00} = 0,490$$

$$h_{13,3} = \frac{r_{03}^2}{|r_{31}| + |r_{33}| + |r_{34}|}$$

$$h_{13,3} = \frac{0,88^2}{|0,87| + |1,00| + |0,01|} = \frac{0,774}{1,88} = 0,412$$

$$h_{13,4} = \frac{r_{04}^2}{|r_{41}| + |r_{43}| + |r_{44}|}$$

$$h_{13,4} = \frac{0,12^2}{|0,13| + |0,01| + |1,00|} = \frac{0,014}{1,14} = 0,012$$

$$H_{13} = h_{13,1} + h_{13,3} + h_{13,4}$$

$$H_{13} = 0,490 + 0,412 + 0,012$$

$$H_{13} = 0,914$$

Na podstawie dokonanych obliczeń wartość integralnej pojemności nośnika informacji wynosi .

$$k_{14} = \{x_2, x_3, x_4\}$$

$$R_{0,14} = \begin{bmatrix} r_{02} \\ r_{03} \\ r_{04} \end{bmatrix} = \begin{bmatrix} 0,77 \\ 0,88 \\ 0,12 \end{bmatrix}$$

$$R_{14} = \begin{bmatrix} r_{22} & r_{23} & r_{24} \\ r_{32} & r_{33} & r_{34} \\ r_{42} & r_{43} & r_{44} \end{bmatrix}$$

$$R_{14} = \begin{bmatrix} 1,00 & 0,64 & 0,14 \\ 0,64 & 1,00 & 0,01 \\ 0,14 & 0,01 & 1,00 \end{bmatrix}$$

$$h_{14,2} = \frac{r_{02}^2}{|r_{22}| + |r_{23}| + |r_{24}|}$$

$$h_{14,2} = \frac{0,77^2}{|1,00| + |0,64| + |0,14|} = \frac{0,593}{1,78} = 0,333$$

$$h_{14,3} = \frac{r_{03}^2}{|r_{32}| + |r_{33}| + |r_{34}|}$$

$$h_{14,3} = \frac{0,88^2}{|0,64| + |1,00| + |0,01|} = \frac{0,774}{1,65} = 0,469$$

$$h_{14,4} = \frac{r_{04}^2}{|r_{42}| + |r_{43}| + |r_{44}|}$$

$$h_{14,4} = \frac{0,12^2}{|0,14| + |0,01| + |1,00|} = \frac{0,014}{1,15} = 0,012$$

$$H_{14} = h_{14,2} + h_{14,3} + h_{14,4}$$

$$H_{14} = 0,333 + 0,469 + 0,012$$

$$H_{14} = 0,814$$

Na podstawie dokonanych obliczeń wartość integralnej pojemności nośnika informacji wynosi .

$$k_{15} = \{x_1, x_2, x_3, x_4\}$$

Dla kombinacji nie ma potrzeby wykreślenia wierszy i kolumn z głównego wektora i głównej macierzy , ponieważ zawiera ona wszystkie kandydujące zmienne objaśniające:

$$R_{0,15} = R_0 = \begin{bmatrix} r_{01} \\ r_{02} \\ r_{03} \\ r_{04} \end{bmatrix} = \begin{bmatrix} 0,99 \\ 0,77 \\ 0,88 \\ 0,12 \end{bmatrix}$$

$$R_{15} = R = \begin{bmatrix} r_{11} & r_{12} & r_{13} & r_{14} \\ r_{21} & r_{22} & r_{23} & r_{24} \\ r_{31} & r_{32} & r_{33} & r_{34} \\ r_{41} & r_{42} & r_{43} & r_{44} \end{bmatrix}$$

$$R_{15} = R = \begin{bmatrix} 1,00 & 0,79 & 0,87 & 0,13 \\ 0,79 & 1,00 & 0,64 & 0,14 \\ 0,87 & 0,64 & 1,00 & 0,01 \\ 0,13 & 0,14 & 0,01 & 1,00 \end{bmatrix}$$

Po wyznaczeniu wektora cząstkowego oraz macierzy cząstkowej , można przejść do obliczania indywidualnych (wzór 9) i integralnej (wzór 10) pojemności nośników

informacji. Dużym ułatwieniem w tworzeniu czterech wzorów szczegółowych indywidualnych pojemności nośników informacji (wzór 9) jest budowanie ich według następującej zasady. Dla pierwszej integralnej pojemności $H_{15,1}$, w liczniku tego wzoru wpisujemy symbol ogólny współczynnika korelacji r_{01} (a później jego wartość) z pierwszego wiersza wektora r_0 , natomiast w mianowniku wpisujemy symbole współczynników korelacji r_{11} , r_{12} , r_{13} i r_{14} (a następnie ich wartości) z pierwszego wiersza macierzy R_{15} . Natomiast dla drugiej pojemności integralnej $H_{15,2}$ w liczniku tego wzoru wpisujemy symbol ogólny współczynnika korelacji r_{02} (a później jego wartość) z drugiego wiersza wektora r_0 , a w mianowniku wpisujemy symbole współczynników korelacji r_{21} , r_{22} , r_{23} i r_{24} (a następnie ich wartości) z drugiego wiersza macierzy R_{15} . Także dla trzeciej pojemności integralnej $H_{15,3}$ w liczniku tego wzoru wpisujemy symbol ogólny współczynnika korelacji r_{03} (a później jego wartość) z trzeciego wiersza wektora r_0 , a w mianowniku wpisujemy symbole współczynników korelacji r_{31} , r_{32} , r_{33} i r_{34} (a następnie ich wartości) z trzeciego wiersza macierzy R_{15} . Dla czwartej pojemności integralnej $H_{15,4}$ w liczniku tego wzoru wpisujemy symbol ogólny współczynnika korelacji r_{04} (a później jego wartość) z czwartego wiersza wektora r_0 , a w mianowniku wpisujemy symbole współczynników korelacji r_{41} , r_{42} , r_{43} i r_{44} (a następnie ich wartości) z czwartego wiersza macierzy R_{15} . Należy pamiętać, że według ogólnego wzoru na indywidualną pojemność nośników informacji (wzór 9) w mianowniku zawsze występuje suma odpowiednich współczynników korelacji zapisanych w wartościach bezwzględnych.

$$h_{15,1} = \frac{r_{01}^2}{|r_{11}| + |r_{12}| + |r_{13}| + |r_{14}|}$$

$$h_{15,1} = \frac{0,99^2}{|1,00| + |0,79| + |0,87| + |0,13|}$$

$$h_{15,1} = \frac{0,980}{2,79} = 0,351$$

$$h_{15,2} = \frac{r_{02}^2}{|r_{21}| + |r_{22}| + |r_{23}| + |r_{24}|}$$

$$h_{15,2} = \frac{0,77^2}{|0,79| + |1,00| + |0,64| + |0,14|}$$

$$H_{15} = h_{15,1} + h_{15,2} + h_{15,3} + h_{15,4}$$

$$H_{15} = 0,351 + 0,231 + 0,307 + 0,011$$

$$H_{15} = 0,900$$

$$h_{15,2} = \frac{0,593}{2,57} = 0,231$$

$$h_{15,3} = \frac{r_{03}^2}{|r_{31}| + |r_{32}| + |r_{33}| + |r_{34}|}$$

$$h_{15,3} = \frac{0,88^2}{|0,87| + |0,64| + |1,00| + |0,01|}$$

$$h_{15,3} = \frac{0,774}{2,52} = 0,307$$

$$h_{15,4} = \frac{r_{04}^2}{|r_{41}| + |r_{42}| + |r_{43}| + |r_{44}|}$$

$$h_{15,4} = \frac{0,12^2}{|0,13| + |0,14| + |0,01| + |1,00|}$$

$$h_{15,4} = \frac{0,014}{1,28} = 0,011$$

Na podstawie dokonanych obliczeń wartość integralnej pojemności nośnika informacji wynosi dla kombinacji $k_{15} = \{x_1, x_2\}$.

Wybór zmiennych objaśniających do modelu ekonometrycznego na podstawie maksymalnej wartości integralnej pojemności nośników informacji i utworzenie na tej podstawie empirycznego modelu ekonometrycznego z wybranymi zmiennymi objaśniającymi

Reasumując, przeprowadzone obliczenia w piętnastu kombinacjach pokazały, że kombinacja k_1 uzyskała największą wartość integralnej pojemności nośników informacji $H_1 =$. Kolejnymi kombinacjami były: $k_6 =$ o wartości $H_6 =$ oraz kombinacja $k_{11} = \{x_1\}$ z wynikiem $H_{11} =$. „Ostatecznie do modelu wybieramy zmienne objaśniające tej kombinacji, której integralna pojemność nośnika informacji ma wartość maksymalną ()²⁴”. Na tej podstawie budujemy empiryczny model ekonometryczny z jedną zmienną , ponieważ $H_1 = H_{max} =$, na przykład o postaci liniowej (wzór 11):

$$y_t = \alpha_0 + \alpha_1 x_{1t} + \xi_t \quad (11)$$

²⁴ Tamże, s. 29.

W rzeczywistych badaniach naukowych opartych na realnych danych empirycznych, tak jak w tym opisywanym przypadku, badacz powinien podjąć próbę, oprócz dokonanych obliczeń formalnych metodą Hellwiga, analizy merytorycznej zmiennych kandydujących. Okazuje się bowiem, że zmienna reprezentująca koszty wytworzenia oraz zmienna reprezentująca koszty zarządu są zmiennymi, które na podstawie analizowanych współczynników korelacji liniowej Pearsona najistotniej wpływają na zmienną objaśnianą, czyli na wielkość produkcji sprzedanej. Analiza ta uprawnia, na tym etapie modelowania ekonometrycznego, do wyboru kombinacji dwuelementowej $k_6 = \{1, 2\}$, zamiast jednoelementowej kombinacji k_1 . Dlatego też ostatecznie podjęto decyzję, że empiryczny model ekonometryczny zostanie zbudowany w oparciu o dwie zmienne objaśniające, a nie o jedną, mimo iż wartość pojemności $H_6 = 1$ nie była największa. Każdy badacz może bezpiecznie podjąć taką decyzję merytoryczną na tym początkowym etapie modelowania ekonometrycznego. Decyzja ta, będzie w kolejnych etapach weryfikowana podczas sprawdzania dobroci modelu, a w szczególności podczas weryfikacji istotności statystycznej parametrów strukturalnych modelu teoretycznego. Wzór 12 przedstawia postać liniową modelu z dwiema zmiennymi:

$$y_t = \alpha_0 + \alpha_1 x_{1t} + \alpha_3 x_{3t} + \xi_t \quad (12)$$

Postać funkcyjna modelu ekonometrycznego może być zarówno liniowa jak i nieliniowa. Poniżej przedstawiono dwa przykładowe nieliniowe modele empiryczne: potęgowy (wzór 13) i logarytmiczny (wzór 14) w ogólnej postaci, dla dwóch zmiennych.

$$y_t = \alpha_0 + x_{1t}^{\alpha_1} + x_{3t}^{\alpha_3} + 10^{\xi_t} \quad (13)$$

$$y_t = \alpha_0 + \alpha_1 \log x_{1t} + \alpha_3 \log x_{3t} + \xi_t \quad (14)$$

Każdy empiryczny model ekonometryczny zawiera czynnik losowy, który jest dopełnieniem tego modelu. Jest on niezbędny do pełnego opisanie badanego zjawiska, związków pomiędzy zmiennymi, które wynikają z następujących przyczyn²⁵:

- postępowanie podmiotów gospodarczych i człowieka nie jest zawsze deterministyczne,
- sposób mierzenia zjawisk ekonomicznych (w tym procesów logistycznych) nie jest doskonały,
- model ekonometryczny może być wadliwy ze względu na pominięcie niektórych zmiennych lub związków między zmiennymi, czy też uwzględnienie niewłaściwych zmiennych.

Dobór zmiennych objaśniających, spośród wszystkich zmiennych kandydujących do modelu ekonometrycznego za pomocą metody Hellwiga jest niezmiernie potrzebnym pierwszym etapem w konstruowaniu dobrego modelu ekonometrycznego. Podczas prac nad modelem ekonometrycznym może zmieniać się liczba (zestaw) zmiennych objaśniających²⁶. Najistotniejszym, w tej pracy jest weryfikacja przyjętych zmiennych, czyli sprawdzenie

²⁵ M. Gruszczyński, T. Kuszewski, M. Podgórska (red.), *Ekonometria i badania operacyjne*, Wydawnictwo Naukowe PWN, Warszawa 2009, s. 17.

²⁶ Tamże, s. 25.

istotności statystycznej parametrów strukturalnych stojących przy każdej z tych zmiennych, a co za tym idzie sprawdzenie istotności statystycznej samych zmiennych, na przykład za pomocą testu t-Studenta²⁷.

Podsumowanie

Analizowana metoda ilościowa, a konkretnie metoda Hellwiga doboru zmiennych w modelowaniu ekonometrycznym procesów logistycznych, oparta została na rzeczywistych danych empirycznych na przykładzie wyników finansowych firmy Wawel S.A. Pogłębiona i szczegółowa analiza procedury doboru zmiennych do modelu ekonometrycznego zaowocowała, na bazie rozpatrywanego przykładu empirycznego, opracowaniem szczegółowego algorytmu wyboru zmiennych objaśniających.

Podjęty na wstępie cel badań dotyczący właściwego, na podstawie pojemności nośników informacji w oparciu o współczynniki korelacji liniowej Pearsona, doboru zmiennych objaśniających do modelu ekonometrycznego metodą Hellwiga, spośród wszystkich analizowanych zmiennych kandydujących do tego modelu został osiągnięty. Przeprowadzenie teoretyczno-empirycznej analizy doboru zmiennych objaśniających do modeli ekonometrycznych opisujących wybrane procesy logistyczne oraz skonstruowanie algorytmu, który został wyjaśniony krok po kroku, pozwoliło na usystematyzowanie obszaru wiedzy w tym zakresie, niespójności elementów wiedzy i nie dookreśln ujętych we wzorach.

Postawiony we wstępie problem badawczy został rozwiązany poprzez pragmatyczne zastosowanie, zweryfikowanego teoretycznie, szczegółowego algorytmu opartego na metodzie Hellwiga. Doskonałym przykładem rozwiązania tego problemu: w jaki sposób dobrać zmienne objaśniające do modelu ekonometrycznego, wykorzystując metodę Hellwiga, jest praktyczne pokazanie eliminacji zmiennych kandydujących niespełniających kryteriów zależności korelacyjnej i informacyjnej (ujętych we wzorach), a pozostawienie tych, które najlepiej spełniają postawione kryteria?.

Natomiast, hipoteza badawcza została pozytywnie zweryfikowana dzięki wykorzystaniu opracowanego algorytmu i zastosowaniu go w rzeczywistym przykładzie doboru zmiennych objaśniających do modelu ekonometrycznego opisującego wielkość produkcji sprzedanej firmy Wawel S.A. Stosując metodę Hellwiga, a dokładniej rzecz biorąc zależności pomiędzy współczynnikami korelacji zawartymi w pojemnościach nośników informacji, wybrano w ramach metody ilościowej najlepszą zmienną objaśniającą x_1 (koszty wytworzenia). Na podstawie dokonanych obliczeń, można zbudować stosowny empiryczny model ekonometryczny z jedną zmienną:

$$y_t = \alpha_0 + \alpha_1 x_{1t} + \xi_t$$

Wykorzystując, w ramach zastosowanych teoretycznych metod i narzędzi badawczych, analizę, syntezę i wnioskowanie na podstawie zgromadzonego materiału empirycznego (tabela 1) autor merytorycznie dobrał do obliczonej zmiennej x_1 reprezentującej koszty wytworzenia, również zmienną x_3 reprezentującą koszty zarządu. Na tej podstawie

²⁷ M. Mróz, *Dobór...*, wyd. cyt., s. 154.

utworzono wyjściowy do dalszych badań, empiryczny liniowy model ekonometryczny z dwiema zmiennymi²⁸.

$$y_t = \alpha_0 + \alpha_1 x_{1t} + \alpha_3 x_{3t} + \xi_t$$

Metoda Hellwiga, podobnie jak metoda analizy grafów, służąca do początkowego doboru zmiennych do modelu ekonometrycznego podaje gotowe i jednoznaczne wyniki doboru tych zmiennych, jednak nie należy traktować tych wyników w sposób kategoriyczny i ostateczny. Należy natomiast uwzględnić w ostatecznej strukturze modelu ekonometrycznego na podstawie wybranych zmiennych, również analizę merytoryczną tych zmiennych w obszarze badanego procesu logistycznego. Wykonane obliczenia, na podstawie danych empirycznych uzupełnione o teoretyczną analizę rozpatrywanego problemu, powinny pomóc badaczowi zdecydować o ostatecznej budowie początkowego modelu ekonometrycznego.

Przeprowadzona analiza teoretyczno-empiryczna doboru zmiennych objaśniających w modelowaniu procesów logistycznych metodą Hellwiga stanowi przesłankę, biorąc pod uwagę różnorodne podejścia autorów do tego problemu, podjęcia kolejnych badań w zakresie identyfikacji i opisu różnych metod ilościowych mogących znaleźć zastosowanie w początkowym etapie modelowania ekonometrycznego tych procesów.

Bibliografia

1. Buttolph-Johnson J. B., Reynolds H. T., Mycoff J. D., *Metody badawcze w naukach politycznych*, Wydawnictwo Naukowe PWN, Warszawa 2010.
2. Duda A., *Modelowanie i prognozowanie ekonometryczne w logistyce przedsiębiorstwa*, [w:] Systemy Logistyczne Wojsk, „Zeszyty Naukowe Instytutu Logistyki Wydziału Logistyki Wojskowej Akademii Technicznej”, nr 44/2016.
3. Górecki B. R., *Ekonometria podstawy teorii i praktyki*, Wydawnictwo KayText, Warszawa 2010.
4. Gruszczyński M., Kuszewski T., Podgórska M. (red.), *Ekonometria i badania operacyjne*, Wydawnictwo Naukowe PWN, Warszawa 2009.
5. Kukuła K. (red.), *Wprowadzenie do ekonometrii w przykładach i zadaniach*, PWN, Warszawa 2009.
6. Łaniec J. D., *Elementy statystyki dla pedagogów*, Wydawnictwo Uniwersytetu Warmińsko-Mazurskiego, Olsztyn 1999.
7. Mróz M., *Dobór zmiennych niezależnych w modelowaniu procesów logistycznych metodą analizy grafów*, [w:] „Zeszyty Naukowe Akademii Sztuki Wojennej”, nr 3(108)/2017.
8. Piłatowska M., *Repetytorium ze statystyki*, Wydawnictwo Naukowe PWN, Warszawa 2008.
9. Sprawozdania finansowe firmy Wawel S.A. opublikowane na Warszawskiej Giełdzie Papierów wartościowych

²⁸Dobór postaci analitycznej modelu ekonometrycznego jest kolejnym etapem w modelowaniu ekonometrycznym, który wymaga szczególnej uwagi i właściwych badań teoretyczno-empirycznych.