

LIDIA GAWLIK*

Budowa i weryfikacja modelu ekonometrycznego dla określenia liniowej zależności pomiędzy kosztami pozyskania węgla a wielkością wydobycia

Wprowadzenie

Poszukiwanie liniowej zależności między kosztami a wielkością wydobycia służy między innymi określeniu jak zachowują się te koszty przy zmianie skali produkcji. Znalezienie modelu ekonometrycznego opisującego tę zależność pozwala wnioskować o podziale kosztów na koszty stałe (niezależne od wielkości wydobycia) i koszty zmienne (zależne od wielkości wydobycia) (Badania... 2006). W artykule przedstawiono i opisano schemat budowy oraz weryfikacji takiego modelu, który został zastosowany do analizy kosztów pozyskania węgla kamiennego. Schemat opracowano w oparciu o literaturę tematu, uwzględniając jednocześnie specyfikę wykorzystanych narzędzi wspomagających proces modelowania – Program Statistica v. 6.0. Zastosowanie tego programu wspomaga proces weryfikacji modelu, gdyż program sam wyznacza szereg statystyk pozwalających na ominięcie procesu ich obliczania.

Przedstawiony schemat wsparto przykładem dotyczącym kosztów wynagrodzeń w jednej z kopalń węgla kamiennego.

1. Specyfikacja problemu

Liniowe modele ekonometryczne sprowadzają się do poszukiwania na podstawie danych *ex post* zależności pomiędzy wielkością którą chcemy objaśnić a wielkościami które ją kształtują w postaci równania o ogólnej postaci:

* Dr inż., Instytut Gospodarki Surowcami Mineralnymi i Energią PAN, Kraków;
e-mail: lidia.gawlik@min-pan.krakow.pl

$$y = a_0 + a_1x_1 + a_2x_2 + \dots + a_nx_n + \varepsilon \quad (1)$$

gdzie:

- y – zmienna objaśniana,
- x_1, x_2, \dots, x_n – zmienne objaśniające,
- $a_0, a_1, a_2, \dots, a_n$ – poszukiwane parametry modelu,
- ε – składnik losowy,
- N – liczba zmiennych objaśniających.

Skupiając się na przypadku analizy kosztów ponoszonych w kopalni można poszukiwać zależności danego rodzaju kosztu (materiałów, wynagrodzeń itp.) od wielkości wydobycia oraz innych wielkości, na przykład czasu czy też parametrów technicznych. Model tym lepiej odzwierciedla rzeczywistość, im więcej zmiennych „istotnie” wpływających na wysokość kosztu jest uwzględnionych w równaniu.

Korzystając z danych *ex post* dotyczących kosztów w określonym czasie oraz zestawu danych, które – jak podejrzewamy – mogą mieć wpływ na poziom tych kosztów można, posługując się metodami ekonometrycznymi, przybliżyć zależności między nimi funkcją określoną równaniem (1). Jedną z najpopularniejszych metod pozwalających na dopasowanie najkorzystniejszej funkcji wiążącej te wielkości jest metoda najmniejszych kwadratów.

Jeśli w analizowanym modelu ograniczymy ilość zmiennych objaśniających do jednej, to w przypadku poszukiwania zależności między kosztem a wielkością produkcji (w kopalni – wielkością wydobycia) mamy do czynienia z klasyczną funkcją kosztów. Takie ograniczenie z jednej strony powoduje, że jeśli inne wielkości niż wydobycie mają istotny wpływ na wysokość kosztu, to nie udaje się dopasować wystarczająco dokładnej linii regresji, z drugiej strony, jeśli już taki model wystarczająco dobrze przybliża badane wielkości, to znaleziona zależność może posłużyć na przykład do ustalenia udziału kosztów stałych i zmiennych w badanym koszcie, co było jednym z celów tej pracy.

Dla zbioru danych statystycznych składającego się z elementów (W_i, K_i) , gdzie i jest numerem zbioru o liczebności n , można dobrać najlepiej dopasowaną linię regresji kształtowania się kosztów K od wydobycia W w postaci równania:

$$K = aW + b + \varepsilon \quad (2)$$

gdzie:

- K – zmienna objaśniana (koszt),
- W – zmienna objaśniająca (wydobycie),
- a, b – nieznanne parametry strukturalne modelu,
- ε – składnik losowy.

Przy założeniu, że składnik losowy jest bliski zeru, można wyznaczyć liniową zależność kosztu od wielkości wydobycia, w którym parametry strukturalne modelu a i b wyznacza się z równań (Zając 1996):

$$a = \frac{\sum_{i=1}^n (W_i - \bar{W})(K_i - \bar{K})}{\sum_{i=1}^n (W_i - \bar{W})^2} \quad (3)$$

$$b = \bar{K} - \bar{W} \cdot a \quad (4)$$

gdzie:

- n – liczebność zbioru danych,
- $i = 1, \dots, n$ – oznaczenie elementu zbioru danych,
- W_i – wielkość wydobycia w i -tym elemencie zbioru danych,
- K_i – koszt całkowity w i -tym elemencie zbioru danych,

$$\bar{W} = \frac{\sum_{i=1}^n W_i}{n} \quad \text{– średnia wielkość wydobycia dla } n \text{ elementów zbioru danych,}$$

$$\bar{K} = \frac{\sum_{i=1}^n K_i}{n} \quad \text{– średni koszt całkowity dla } n \text{ elementów zbioru danych.}$$

Jeśli uzyskane równanie w sposób wystarczająco dobry opisuje badany zbiór danych to oznacza, że opisano zależność kosztu od wielkości wydobycia za pomocą funkcji liniowej, w której: a jest jednostkowym kosztem zmiennym (k_{jz}), b zaś jest całkowitym kosztem stałym (K_s) (Czopek 2003):

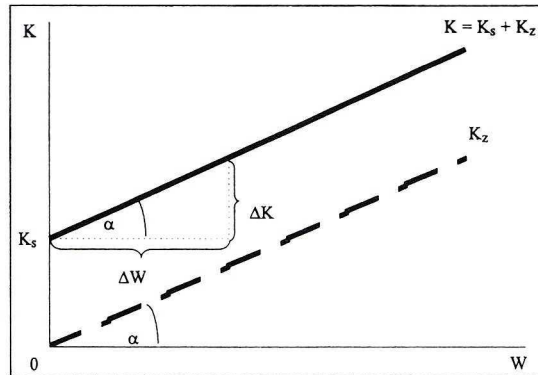
$$\bar{K} = k_{jz} \cdot \bar{W} + K_s \quad (5)$$

Ilustrację tego równania przedstawiono na rysunku 1, gdzie:

$$k_{jz} = \text{tg}\alpha = \frac{\Delta K}{\Delta W} \quad (6)$$

Po wyznaczeniu i weryfikacji modelu liniowego średni procentowy udział kosztów zmiennych (K_z [%]) dla badanej serii danych kosztowych wyznacza się jako:

$$Kz(\%) = \frac{k_{jz} \cdot \bar{W}}{k_{jz} \cdot \bar{W} + K_s} \cdot 100\% \quad (7)$$



Rys.1. Prostoliniowa funkcja kosztu

Fig. 1. Linear function of cost

2. Procedura wyznaczania i weryfikacji modelu ekonometrycznego

Program Statistica jest narzędziem wspomagającym proces budowy modelu ekonometrycznego. Dla serii dwóch danych wyznacza regresję liniową, której przykładowe wyniki przedstawiono na rysunku 2.

W przedstawionym przykładzie program Statistica wyznacza następujące wielkości:

- n – liczba przypadków = 12,
- k – liczba szacowanych parametrów modelu = 2,
- $(n - k)$ – liczba stopni swobody = 10,
- R – korelacja między analizowanymi seriami danych = 0,77675,
- R^2 – kwadrat korelacji = 0,60334,
- $\overline{R^2}$ – skorygowany R^2 = współczynnik determinacji = 0,563678,
- F – wartość statystyki F-Snedecora = 15,211,
- S_{ey} – standardowy błąd estymacji = 1971,5,
- a – parametr strukturalny modelu (współczynnik przy zmiennej) = 43,318,
- b – parametr strukturalny modelu (wyraz wolny) = 6138,920,
- S_{ea} – standardowa wartość błędu dla współczynnika a = 11,107,

Podsumowanie regresji zmiennej zależnej: Wynagrodzenia (Przykład 06-06)						
R= ,77675206 R2= ,60334376 Skoryg. R2= ,56367814						
F(1,10)=15,211 p<,00296 Błąd std. estymacji: 1971,5						
N=12	BETA	Błąd st. BETA	B	Błąd st. B	t(10)	poziom p
W. wolny			6136,920	2032,873	3,018841	0,012921
Wydobycie	0,776752	0,199162	43,318	11,107	3,900096	0,002961

Rys. 2. Wynik regresji kosztów wynagrodzeń od wydobywania w analizowanej kopalni

Fig. 2. Result of the regression function of salary costs and coal production in the analysed coal mine

- S_{eb} – standardowa wartość błędu dla współczynnika $b = 2032,873$,
 $t(10)_a$ – wartość statystyki t-Studenta dla współczynnika $a = 3,900$,
 $t(10)_b$ – wartość statystyki t-Studenta dla współczynnika $b = 3,019$,

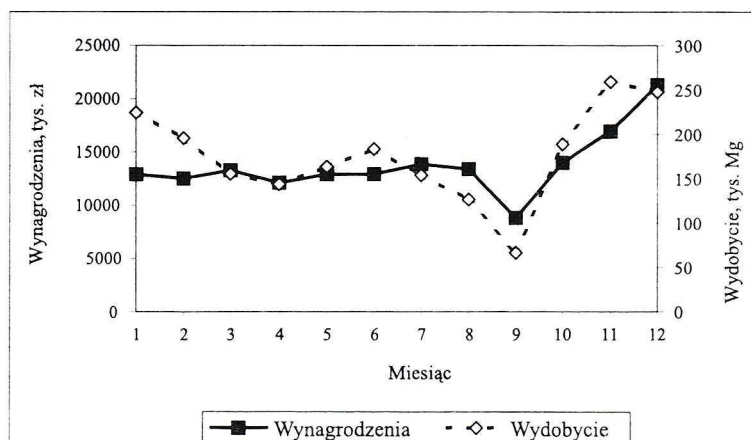
Te wielkości pozwalają w znacznym stopniu wnioskować o poprawności wyznaczonego równania regresji. Aby jednak upewnić się, że wyznaczone równanie linowe poprawnie odzwierciedla badane zjawisko należy model wszechstronnie zweryfikować zarówno ze względu na wyznaczone parametry strukturalne, jak również ze względu na charakter reszt modelu. Wymaga tego bowiem zastosowana metoda najmniejszych kwadratów, której poprawność warunkują przyjęte założenia (Brzychczy 2007; Ekonometria... 2003) dotyczące:

- braku współliniowości zmiennych niezależnych,
- nielosowości zmiennych objaśniających,
- zerowości sumy składników losowych,
- braku autokorelacji reszt między sobą oraz ze zmiennymi objaśniającymi.

Literatura dotycząca modeli ekonometrycznych (Ekonometria... 1998; Ekonometria 2003; Wefle 2003; Wprowadzenie... 1996; Zemke mat.internetowe) omawia szereg testów służących weryfikacji modelu. Na tej podstawie opracowany został schemat budowy i weryfikacji modelu liniowego przedstawiającego zależność kosztu od wielkości wydobycia w kopalni węgla kamiennego. Schemat ten przedstawiono w tabeli 1 i omówiono poniżej. W opisie pominięto wyznaczenie statystyk, które program Statistica oblicza automatycznie, skupiając się tylko na procedurach niezbędnych do oceny poprawności modelu.

Omówienie poszczególnych kroków poparto analizą kosztów wynagrodzeń w jednej z kopalń węgla kamiennego w roku 2006. W analogiczny sposób budowane są modele innych kosztów w oparciu o serie danych o różnej liczebności.

Kształtowanie się wielkości wydobycia w tej kopalni w poszczególnych miesiącach analizowanego okresu oraz miesięczne koszty wynagrodzeń przedstawiono na rysunku 3.



Rys. 3. Kształtowanie się wydobycia i wielkości wynagrodzeń w przykładowej kopalni w poszczególnych miesiącach 2006 r.

Fig. 3. Production and salary costs in particular months of 2006 in the analysed coal mine

TABELA 1

Schemat budowy i weryfikacji modelu zależności liniowej między kosztem a wielkością wydobycia

TABLE 1

Diagram of the construction and verification a model of linear dependence of cost as a function of production

Nr kolejny kroku	Procedura – opis	
KROK 1	Wyznaczenie korelacji pomiędzy wielkością objaśnianą a wielkością objaśniającą	
	Czy $R > R_{kr}$?	
	TAK	NIE
KROK 2	STOP Nie ma możliwości wyznaczenia modelu dla badanej serii danych	
KROK 2	Estymacja parametrów strukturalnych modelu a i b	
	Badanie istotności wyznaczonych parametrów strukturalnych	
	Czy a i b – istotne?	
KROK 3	TAK	NIE
	KROK 4	KROK 13
	Badanie możliwości fizycznej interpretacji modelu	
KROK 4	Czy $a > 0$ i $b \geq 0$?	
	TAK	NIE
	KROK 5	KROK 13
KROK 5	Badanie dopuszczalności relacji modelu ze względu na współczynnik determinacji	
	Czy $R^2 \geq 0,6$?	
	TAK	NIE
KROK 6	KROK 6	KROK 13
	Badanie dopuszczalności relacji modelu ze względu na współczynnik zmienności V	
	Czy $V \leq 0,1$?	
KROK 7	TAK	NIE
	KROK 7	KROK 13
	Badanie istotności układu współczynników (w oparciu o statystykę F-Snedecora)	
KROK 7	Czy $F > F_{kr}$?	
	TAK	NIE
	KROK 8	KROK 13
KROK 8	Badanie autokorelacji składnika losowego (w oparciu o statystykę Durbin-Watsona)	
	Czy $DW > d_U$?	
	TAK	NIE
KROK 9	KROK 9	KROK 13
	Badanie symetrii składnika losowego (dla małej próby liczba przypadków <30 – rozkład t-Studenta; dla dużej próby – rozkład normalny)	
	Czy $t \leq t_{kr}$?	
KROK 10	TAK	NIE
	KROK 10	KROK 13
	Badanie losowości składnika losowego (test serii)	
KROK 10	Czy $K > K_{kr}$?	
	TAK	NIE
	KROK 11	KROK 13
KROK 11	Badanie stacjonarności składnika losowego (statystyka t-Studenta)	
	Czy $t \leq t_{kr}$?	
	TAK	NIE
KROK 12	KROK 12	KROK 13
	Badanie normalności reszt	
	Czy $W > K_{kr}$? [lub] Czy $p > \alpha$?	
KROK 13	TAK	NIE
	STOP Model jest poprawny	
	Eliminacja danych odstających	
KROK 13	Czy liczba wyeliminowanych przypadków przekracza 20% pierwotnej liczby?	
	TAK	NIE
	STOP Nie ma możliwości wyznaczenia modelu dla badanej serii danych	
KROK 13	KROK 1	

W ciągu 12 miesięcy kopalnia zanotowała znaczące wahania wielkości miesięcznego wydobycia. Zmieniało się ono w przedziale od 259 tys. Mg (maksymalne miesięczne wydobycie osiągnięte w listopadzie) do 66,4 tys. Mg (minimalne wydobycie osiągnięte w wrześniu).

Wahania miesięcznych kosztów wynagrodzeń w ciągu analizowanych 12 miesięcy były również znaczne w przedziale od prawie 21,3 mln zł w grudniu do 8,8 mln zł we wrześniu.

Wynik analizy regresji wykonanej programem Statistica jest przedstawiony na rysunku 1.

Poniżej omówiono poszczególne kroki budowy weryfikacji modelu. Przyjęto poziom istotności dla przykładowego modelu $\alpha = 0,05$.

2.1. Określenie korelacji pomiędzy zmienną objaśniającą a wielkością objaśnianą (krok 1)

Krytyczny współczynnik korelacji wyznacza się z równania:

$$R_{kr} = \sqrt{\frac{t^2}{t^2 + n - 2}} \quad (8)$$

gdzie:

- t – wartość statystyki odczytana z tablic testu t-Studenta, przy zadanym poziomie istotności α ,
- n – liczba przypadków.

W przykładowej analizie mamy 12 przypadków (12 danych miesięcznych). Dla założonego poziomu istotności $\alpha = 0,05$ znajdujemy w tablicach $t = 2,228$, co oznacza, że $R_{kr} = 0,576$.

Tymczasem wyznaczony programem Statistica współczynnik korelacji pomiędzy wielkością wynagrodzeń a wielkością wydobycia wynosi 0,7767 i jest wyższy od krytycznego. Oznacza to, że można poszukiwać równania regresji między tymi wielkościami.

2.2. Estymacja parametrów strukturalnych modelu a i b (krok 2)

Wyznaczony model ma postać:

$$K = 43,318 \cdot W + 6\,136,920$$

Model ten w dalszej kolejności podlegać będzie weryfikacji.

2.3. Badanie istotności wyznaczonych parametrów strukturalnych (krok 3)

Do badania istotności parametrów strukturalnych modelu stosuje się test t-Studenta. Parametr jest istotny jeśli obliczona programem wartość statystyki t-Studenta jest wyższa od wartości krytycznej.

Wartość krytyczną odczytujemy z tablic t-Studenta dla zadanego poziomu istotności ($\alpha = 0,05$) i liczbie stopni swobody ($n - k = 10$): $t_{kr} = 2,228$.

Ponieważ:

$$t(10)_a = 3,019 > t_{kr} \quad \text{i} \quad t(10)_b = 3,900 > t_{kr}$$

można stwierdzić, że parametry analizowanego modelu są istotne.

2.4. Badanie możliwości fizycznej interpretacji modelu (krok 4)

Ponieważ analizowany model reprezentuje zależność kosztów od wielkości wydobycia, powinien zatem spełniać następujące założenia:

- 1) wraz ze wzrostem wydobycia koszt powinien rosnąć (współczynnik $a > 0$)
- 2) dla dowolnej wielkości wydobycia koszt powinien być nieujemny (współczynnik $b \geq 0$)

Jeśli warunki te nie są spełnione model nie odzwierciedla rzeczywistości i jest nieprzydatny (niepoprawny).

W analizowanym przykładzie oba współczynniki (a i b) są dodatnie, pozwalają więc na poprawną interpretację modelu.

2.5. Badanie dopuszczalności relacji modelu ze względu na współczynnik determinacji $\overline{R^2}$ (krok 5)

Współczynnik determinacji informuje w jakim procencie zmienność wielkości objaśnianej wyjaśniana jest przez model. Współczynnik ten przyjmuje wartości z przedziału (0,1) i im jest wyższy, tym model lepiej objaśnia badane zjawisko. Zwykle przyjmuje się w sposób arbitralny wartość krytyczną tego współczynnika. W praktyce często uznaje się, że przy minimalnej wartości $\overline{R^2} = 0,6$ możemy model uważać za poprawny. Badając zależność kosztów od wydobycia w kopalniach węgla kamiennego przyjęto taką wartość współczynnika determinacji za wartość krytyczną.

W analizowanym przykładzie wyznaczona wartość współczynnika determinacji wynosi 0,5637. **Jest zatem zbyt niska, by model uznać za poprawny.**

W takim przypadku należy przerwać dalsze testowanie modelu i uznać go za niepoprawny. Można poszukiwać innej zależności dla pełnej serii danych lub spróbować wyeliminować niektóre dane miesięczne, które zaburzają przebieg regresji liniowej.

W wykonywanej przykładowej analizie wykonano **krok 13** polegający na eliminacji niektórych danych miesięcznych, w których wielkości zaksięgowanych kosztów wynagrodzeń znacząco różniła się od miesięcy sąsiednich.

Po wyeliminowaniu 1 danej miesięcznej (wynagrodzeń wypłaconych w grudniu, które obciążone były wypłatą pensji barburkowej) otrzymano wynik regresji wykonanej programem Statistica przedstawiony na rysunku 4.

Podsumowanie regresji zmiennej zależnej: Wynagrodzenia (Przykład 06-06)						
R= ,80159566 R2= ,64255560 Skoryg. R2= ,60283956						
F(1,9)=16,179 p<,00301 Błąd std. estymacji: 1197,6						
N=11	BETA	Błąd st. BETA	B	Błąd st. B	t(9)	poziom p
W. wolny			7998,267	1310,058	6,105278	0,000178
Wydobycie	0,801596	0,199289	29,940	7,444	4,022281	0,003008

Rys. 4. Wynik regresji kosztów wynagrodzeń od wydobycia w analizowanej kopalni po wyeliminowaniu jednego przypadku

Fig. 4. Result of the regression function of salary costs and coal production in the analysed coal mine after rejection one set of data

Dla tej serii danych powtórzono analizę od początku otrzymując następujące wyniki:

Krok 1: Współczynnik korelacji $R = 0,80$ jest wyższy od krytycznego.

Krok 2: Wyznaczony model ma postać $K = 29,940 \cdot W + 7\,998,267$.

Krok 3: Wartość krytyczną odczytujemy z tablic t-Studenta dla zadanego poziomu istotności ($\alpha = 0,05$) i 9 stopni swobody. Wynosi on $t_{kr} = 2,262$.

Ponieważ:

$$t(9)_a = 4,022 > t_{kr} \quad \text{ i } \quad t(9)_b = 6,105 > t_{kr}$$

można stwierdzić, że parametry analizowanego modelu są istotne.

Krok 4: Oba współczynniki (a i b) są dodatnie, pozwalają więc na poprawną interpretację modelu.

Krok 5: Współczynnik determinacji modelu wynosi 0,603 i jest wyższy od krytycznego. Model można zatem uznać za poprawny ze względu na współczynnik determinacji.

2.6. Badanie dopuszczalności relacji modelu ze względu na współczynnik zmienności V (krok 6)

Współczynnik zmienności wyznacza się ze wzoru:

$$V = \frac{S_{ey}}{K} \cdot 100\% \quad (9)$$

gdzie:

\underline{S}_{ey} – standardowy błąd estymacji = 1197,6,
 \bar{K} – średnia wartość zmiennej K :

$$\bar{K} = \frac{\sum_{i=1}^n K_i}{n} \quad (10)$$

Współczynnik zmienności V powinien przyjmować jak najmniejsze wartości. W praktyce przyjmuje się, że model jest poprawny ze względu na współczynnik zmienności, jeśli współczynnik ten jest mniejszy od wartości przyjętej arbitralnie, zwykle na poziomie 10%.

W analizowanym przykładzie średnia wartość zmiennej K dla $n = 11$ przypadków wynosi 13064, zatem $V = 9,2\%$. Współczynnik ten jest niższy od wartości krytycznej, należy więc uznać, że relacja modelu jest wyznaczona poprawnie ze względu na współczynnik zmienności.

2.7. Badanie istotności układu współczynników – w oparciu o statystykę F-Snedecora (krok 7)

Testem F-Snedecora możemy stwierdzić, czy parametry stojące przy zmiennych objaśniających są nieistotne. A parametry te są istotne jeśli wyznaczona wartość F jest wyższa od wartości krytycznej: $F > F_{kr}$.

Wartość krytyczną odczytuje się z tablic rozkładu F-Snedecora dla danego α , $(k - 1)$ szacowanych parametrów modelu oraz $(n - k)$ stopni swobody.

A analizowanym przypadku $F(1,9) = 16,179$, zaś wartość krytyczna statystyki (odczytana z tablic) wynosi $F_{kr} = 3,36$.

Układ współczynników jest więc istotny.

Krok 7 kończy analizę parametrów strukturalnych modelu. Kolejne kroki mają za zadanie weryfikację reszt modelu.

2.8. Badanie autokorelacji składnika losowego – w oparciu o statystykę Durbina-Watsona (krok 8)

Autokorelacja składnika losowego może wystąpić zwłaszcza w modelach stochastycznych (budowanych w oparciu o szeregi czasowe) na przykład z powodu opóźnienia występowania zmiennej objaśnianej (kosztów) w stosunku do zmiennej objaśniającej (wydobywie).

Empiryczna wartość statystyki Durbina-Watsona wyrażona jest wzorem:

$$DW = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2} \quad (11)$$

DW wyznaczana jest przez program Statistica.

Wartość tej statystyki porównuje się z wartościami krytycznymi statystyki Durбина-Watsona odczytywanymi z tablic dla zadanego α , liczby analizowanych przypadków n oraz liczby zmiennych niezależnych w równaniu regresji (bez wyrazu wolnego). Tablice te podają dwie wielkości d_L oraz d_U .

Gdy $DW < d_L$ należy przyjąć hipotezę o autokorelacji składnika losowego, co oznacza, że model jest niepoprawny.

Gdy $d_L < DW < d_U$ test nie daje odpowiedzi. Nie możemy odrzucić hipotezy o autokorelacji składnika losowego, co w praktyce oznacza, że należy modyfikować model.

Gdy $DW \geq d_U$ brak autokorelacji składnika losowego – model jest poprawny.

W analizowanym przykładzie wyznaczona empirycznie wartość statystyki Durбина-Watsona to 1,504.

Z tablic wartości krytycznych rozkładu Durбина-Watsona odczytujemy dla $\alpha = 0,05$, $n = 11$ i $k = 1$ wartości krytyczne $d_L = 0,93$ oraz $d_U = 1,32$.

Ponieważ $DW > d_U$, brak zatem autokorelacji składnika losowego.

2.9. Badanie symetrii składnika losowego – dla małej próby liczba przypadków < 30 – rozkład t-Studenta; dla dużej próby – rozkład normalny (krok 9)

Podstawą analizy symetrii składnika losowego jest wartość empiryczna statystyki t o postaci:

$$t = \frac{\left| \frac{n1}{n} - \frac{1}{2} \right|}{\sqrt{\frac{\frac{n1}{n} \left(1 - \frac{n1}{n} \right)}{n-1}}} \quad (12)$$

gdzie:

n – liczba analizowanych przypadków (reszt),

$n1$ – liczba reszt dodatnich.

Reszty są symetryczne, a więc model jest poprawny jeśli $t < t_{kr}$.

Wartość krytyczną wyznacza się dla małej liczby przypadków ($n = <30$) z rozkładu t-Studenta, a dla $n > 30$ z rozkładu normalnego.

W analizowanym przykładzie reszty modelu zestawiono w tabeli 2.

Po zliczeniu $n = 11$ oraz $n1 = 6$, a zatem $t = 0,289$.

Odczytana z tablic rozkładu t-Studenta wartość krytyczna wynosi 2,262, więc $t < t_{kr}$, wobec czego nie ma podstaw do odrzucenia hipotezy, że reszty mają rozkład symetryczny. Model jest zatem poprawny.

TABELA 2

Zestawienie reszt analizowanego modelu dla sprawdzenia symetrii i losowości składnika losowego

TABLE 2

List of residual errors of the model for the purposes of checking if they are symmetrical and random

Nr przypadku	Reszta	Reszty dodatnie	Reszty ujemne	Nr serii
1	-1 803,65	0	1	1
2	-1 328,03	0	1	1
3	635,91	1	0	2
4	-180,36	0	1	3
5	14,30	1	0	4
6	-569,62	0	1	5
7	1 281,67	1	0	6
8	1 601,42	1	0	6
9	-1 178,82	0	1	7
10	335,40	1	0	8
11	1 191,78	1	0	8
Razem		6	5	8

2.10. Badanie losowości składnika losowego – test serii

(krok 10)

Test serii jest sprawdzianem losowości składnika losowego. Serię dla zjawiska stochastycznego stanowią podciągi składające się z reszt jednego znaku. Badanie polega na zliczeniu liczby serii K i porównaniu z odczytaną z tablic testu serii liczbą K_{kr} dla parametrów rozkładu ($n1, n2, \alpha$), gdzie $n1$ oznacza liczbę reszt jednego znaku (dodatnich), $n2$ liczbę reszt przeciwnego znaku (ujemnych) a α – przyjęty poziom istotności.

Jeżeli $K > K_{kr}$, to składnik losowy jest losowy i model jest poprawny.

W tabeli 2 zliczono liczbę serii dla reszt modelu w analizowanym przykładzie.

Liczba serii $K = 8$. Liczba reszt dodatnich $n_1 = 6$, zaś liczba reszt ujemnych $n_2 = 5$. Odczytana z tablic testu serii wartość $K_{kr} = 3$.

Ponieważ $K > K_{kr}$, można przyjąć hipotezę, że reszty modelu są losowe. Model jest zatem poprawny.

2.11. Badanie stacjonarności składnika losowego – statystyka t-Studenta (krok 11)

Składnik losowy jest stacjonarny jeśli $t \leq t_{kr}$, gdzie:

$$t = \frac{|r_{e,t}| \sqrt{n-2}}{\sqrt{1-r_{e,t}^2}} \quad (13)$$

a:

$$r_{e,t} = \frac{\sum_{t=1}^n (e_t - \bar{e})(t - \bar{t})}{\sqrt{\sum_{t=1}^n (e_t - \bar{e})^2 \sum_{t=1}^n (t - \bar{t})^2}} \quad (14)$$

Wartość t_{kr} odczytuje się z tablic t-Studenta o parametrach $(\alpha, n - 2)$.

W analizowanym przykładzie $t_{kr} = 2,262$.

W tabeli 3 zestawiono wielkości służące wyznaczeniu $r_{e,t}$ według wzoru (14).

Po wstawieniu wyliczonych wartości do wzoru (14) otrzymuje się $r_{e,t} = 0,5578$, a po zastosowaniu wzoru (13) wartość obliczonej statystyki $t = 2,016$.

Ponieważ $t < t_{kr}$, składnik losowy modelu jest zatem stacjonarny.

2.12. Badanie normalności składnika losowego – test Shapiro-Wilka (krok 12)

Normalność składnika losowego jest warunkiem poprawności zastosowania testów opartych na rozkładzie normalnym, takich jak t-Studenta i Fishera-Snedecora, które zostały użyte we wcześniejszych krokach testowania modelu.

Program Statistica wykonuje (między innymi) test Shapiro-Wilka. Test ten polega na obliczeniu statystyki dla n reszt ustawionych w kolejności niemalejącej:

Tabela pomocnicza do wyznaczenia $r_{e,t}$ dla zbadania stacjonarności składnia losowego w analizowanym przykładzie

TABLE 3

Auxiliary table for estimation $r_{e,t}$ for the purpose of checking if the residual errors are stationary

t	e_t	$e_t - \bar{e}$	$t - \bar{t}$	$(e_t - \bar{e})(t - \bar{t})$	$(e_t - \bar{e})^2$	$(t - \bar{t})^2$
1	-1 803,653	-1 803,65	-5,000	9 018,265	3 253 164,374	25,000
2	-1 328,034	-1 328,03	-4,000	5 312,136	1 763 674,474	16,000
3	635,912	635,91	-3,000	-1 907,736	404 384,118	9,000
4	-180,356	-180,35	-2,000	360,713	32 528,454	4,000
5	14,296	14,29	-1,000	-14,296	204,371	1,000
6	-569,618	-569,61	0,000	0,000	324 464,966	0,000
7	1 281,674	1 281,67	1,000	1 281,674	1 642 688,079	1,000
8	1 601,420	1 601,42	2,000	3 202,840	2 564 545,813	4,000
9	-1 178,817	-1 178,81	3,000	-3 536,451	1 389 609,670	9,000
10	335,402	335,40	4,000	1 341,609	112 494,660	16,000
11	1 191,775	1 191,77	5,000	5 958,875	1 420 327,499	25,000
$\bar{t} = 6,00$	$\bar{e} = 0,000064$			$\sum_{t=1}^n (e_t - \bar{e})(t - \bar{t}) =$ $= 21 017,629$	$\sum_{t=1}^n (e_t - \bar{e})^2 =$ $= 1 2908 086,48$	$\sum_{t=1}^n (t - \bar{t})^2 =$ $= 110,00$

$$W = \frac{\left[\sum_{i=1}^{[n/2]} a_{n,i} (e_{(n-i+1)} - e_i) \right]^2}{\sum_{t=1}^n (e_t - \bar{e})^2} \quad (15)$$

gdzie:

$[n/2]$ – część całkowita liczby $n/2$,

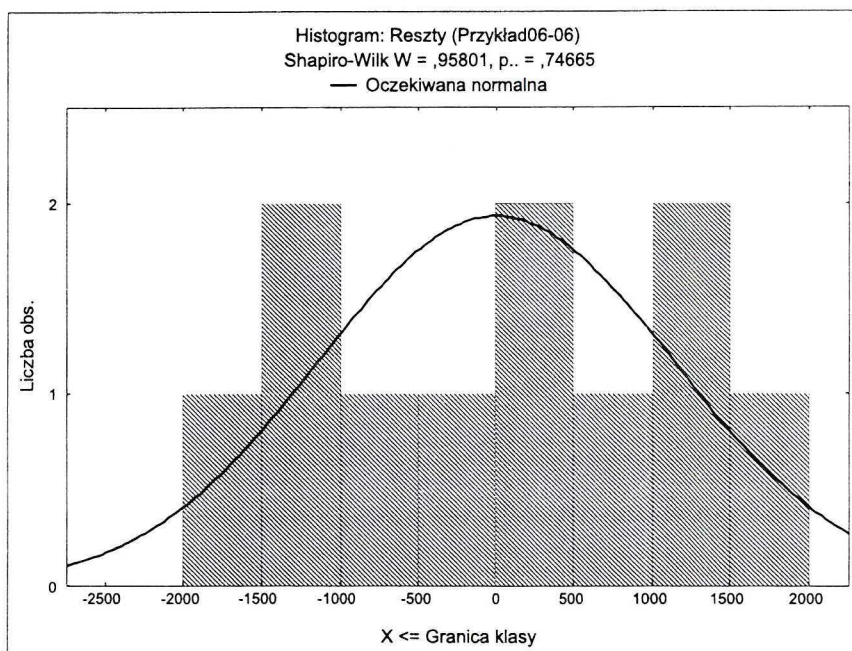
$a_{n,i}$ – współczynniki Shapiro-Wilka, które odczytuje się z tablic współczynników Shapiro-Wilka.

Wyznaczoną empirycznie wartość W porównuje się wartością krytyczną K_{kr} odczytywaną z tablicy testu Shapiro-Wilka dla zadanego poziomu istotności α i ilości analizowanych reszt n .

Jeśli $W > K_{kr}$, możemy przyjąć, że rozkład reszt jest normalny.

Program Statistica podaje również istotność rozkładu reszt p , który należy porównać z zadanym poziomem istotności dla modelu. Reszty mają rozkład normalny jeśli istotność rozkładu tych reszt jest wyższa od zadanego poziomu istotności α .

W analizowanym przykładzie histogram reszt wygląda jak przedstawiono na rysunku 5.



Rys. 5. Histogram reszt modelu w analizowanym przykładzie

Fig. 5. Histogram of the residual error of the model in the analysed example

W analizowanym przypadku odczytana z tablic Shapiro-Wilka wartość krytyczna $K_{kr} = 0,850$ i jest niższa od statystyki wyznaczonej empirycznie. Równocześnie wyznaczona programem Statistica istotność rozkładu $p = 0,7466$ jest znacząco wyższa od założonego poziomu istotności $\alpha = 0,05$. Oznacza to, że reszty mają rozkład normalny.

Krok ten kończy weryfikację modelu ekonometrycznego. Pozytywna odpowiedź na wszystkie postawione pytania pozwala na stwierdzenie, że model wyznaczono poprawnie. Negatywna odpowiedź na pytanie zadane w dowolnym kroku sugeruje, że badany model jest skonstruowany niepoprawnie. Można wówczas odstąpić od poszukiwania modelu lub dokonać analizy i weryfikacji danych do modelu.

2.13. Eliminacja danych odstających (krok 13)

Program Statistica posiada rozbudowane narzędzie wyróżniania danych, które odbiegają od innych. Reszty przedstawiane są na standaryzowanych wykresach rozrzutu, pozwalając

na łatwe wyszukanie przypadków zawierających dane odstające. Takie reszty sugerują wystąpienie zdarzeń zaburzających normalną pracę kopalni. Mogą to być gwałtowne spadki wydobywania spowodowane awarią albo nadzwyczajne okoliczności powodujące wzrost kosztu ponad zwykły poziom. Jeśli istnieje podejrzenie wystąpienia takiego zjawiska, to z serii analizowanych danych należy wyeliminować te pary danych, które są jego efektem.

Każda decyzja o eliminacji konkretnego przypadku powinna być poprzedzona merytoryczną oceną przyczyn jego wystąpienia. Należy również wprowadzić limit liczby wyeliminowanych przypadków, aby nie budować modelu „na siłę” poprzez pozostawienie tylko takich danych, które wpasowują się w założony model. W przedstawionym schemacie budowy modelu ekonometrycznego założono, że maksymalna liczba wyeliminowanych przypadków nie może przekraczać 20% pierwotnej liczby analizowanych serii danych.

Podsumowanie

Przedstawiona procedura budowy i weryfikacji modelu liniowego pomiędzy wielkością wydobywania a ponoszonymi kosztami jest procedurą uniwersalną mającą zastosowanie do poszukiwania zależności liniowych pomiędzy dowolnymi wielkościami. Procedura ta zawiera zarówno badanie uzyskanych parametrów strukturalnych modelu, jak również badanie reszt modelu w celu wyeliminowania wnioskowania na podstawie modeli uzyskanych przez przyjęcie nieprawidłowych założeń. Dopiero pełna weryfikacja modelu upoważnia na wnioskowania na jego podstawie.

Dla prowadzonej w tej pracy analizy kosztów wynagrodzeń w funkcji wydobywania w jednej z kopalń węgla kamiennego, po przeprowadzonej pozytywnie weryfikacji, można stwierdzić, że funkcją liniową opisującą tę zależność jest funkcja:

$$K = 29,940 \cdot W + 7\,998,267$$

Oznacza to, że w roku 2006 koszt stały w tej kopalni wynosił (średnio miesięcznie) $K_s = 7\,998,267$ tys. zł, a jednostkowy koszt zmienny $k_{jz} = 29,94$ zł/Mg.

Po zastosowaniu równania (7) można stwierdzić, że udział kosztów zmiennych w kosztach wynagrodzeń w tej kopalni wyniósł 38,8%.

Modele ekonometryczne, a w szczególności wnioski jakie z nich można wyciągnąć, mają szerokie zastosowanie do zarządzania produkcją (Gawlik 2007a, b; Łucki 1988; Przybyła 2004).

Ważne jest jednak, by modele te były budowane poprawnie. Wielu niedoświadczonych badaczy kończy poszukiwanie funkcji regresji w momencie uzyskania wystarczająco wysokiego współczynnika R^2 , nie do końca zdając sobie sprawę, że posłużenie się metodą najmniejszych kwadratów było nieuzasadnione z powodu niespełnienia podstawowych założeń tej metody.

Opracowany i przedstawiony schemat budowy i weryfikacji modelu liniowego stanowić może pomoc w ustaleniu kolejności działań w dziele tworzenia modelu. Kolejne kroki ustawione są bowiem w kolejności od najmniej do najbardziej pracochłonnych.

Przedstawiony tutaj model liniowy z jedną zmienną objaśniającą jest modelem najprostszy. Jego postać jest przydatna do określenia udziału kosztów stałych i zmiennych w analizowanym koszcie. Należy jednak zdawać sobie sprawę, że rzeczywiste zależności między badanymi wielkościami nie zawsze są tak proste. Zależność między kosztami i wydobyciem może przybierać inną niż liniowa postać. Koszty mogą zależeć od więcej niż jednej zmiennej objaśniającej. Postać analityczna modelu powinna zatem być dostosowana do celu jakiego badania modelowe mają służyć.

LITERATURA

- Badania kosztów pozyskania węgla kamiennego i brunatnego w celu określenia optymalnej struktury paliwowej produkcji energii elektrycznej. Red. L. Gawlik, Wyd. IGSMiE PAN, Kraków 2006.
- Brzychczy E., 2007 – Budowa modeli ekonometrycznych wybranych parametrów techniczno-ekonomicznych kopalni węgla kamiennego. *Wiadomości Górnicze* nr 11, s. 619–625.
- Brzychczy E., 2008 – Weryfikacja modeli ekonometrycznych opracowanych dla potrzeb analizy charakterystyk techniczno-ekonomicznych kopalń węgla kamiennego. *Wiadomości Górnicze* nr 3, s. 175–181.
- Czopek K., 2003 – Koszty stałe i zmienne. Teoria – Praktyka. Cz. I Funkcja prostoliniowa. Wyd. Art-Tekst, Kraków.
- Ekonometria. Metody i analiza problemów ekonomicznych. Red. K. Jajuga. Wyd. Akademii Ekonomicznej we Wrocławiu 1998.
- Ekonometria. Metody, przykłady, zadania. Red. J. Dziechciarz. Wyd. Akademii Ekonomicznej we Wrocławiu 2003.
- Gawlik L., 2007a – Koszty stałe i zmienne pozyskania węgla kamiennego jako element zarządzania produkcją, *Polityka Energetyczna* t. 10, z. spec. 2, s. 471–482.
- Gawlik L., 2007b – Zastosowanie kosztów stałych i zmiennych pozyskania węgla do oceny efektywności kopalni, [W:] *Szkoła ekonomiki i zarządzania w górnictwie 2007*, s. 115–124.
- Łucki Z., 1988 – Modele ekonometryczne do zarządzania w górnictwie naftowym. ZN AGH, Zagadnienia techniczno-ekonomiczne nr 1.
- Przybyła H., 2004 – Wykorzystanie modeli statystycznych w ocenie i prognozowaniu wyników produkcyjnych i ekonomicznych górnictwa węgla kamiennego, [W:] *Szkoła ekonomiki i zarządzania w górnictwie 2004*, s. 357–366.
- Welfe A., 2003 – *Ekonometria*. PWE, Warszawa.
- Wprowadzenie do ekonometrii w przykładach i zadaniach. Red. K. Kukuła. PWN, Warszawa 1996.
- Zajac K., 1976 – *Zarys metod statystycznych*. PWE Warszawa.
- Zemke J. – Weryfikacja modelu ekonometrycznego. Materiały internetowe.
<http://wzr.pl/~jz/pliki/Ekonometria/Weryfikacja.doc>

BUDOWA I WERYFIKACJA MODELU EKONOMETRYCZNEGO DLA OKREŚLENIA LINIOWEJ ZALEŻNOŚCI POMIĘDZY KOSZTAMI POZYSKANIA WĘGLA A WIELKOŚCIĄ WYDOBYCIA

Słowa kluczowe

Modele ekonometryczne, model liniowy, koszty pozyskania węgla

Streszczenie

W pracy przedstawiono i opisano kolejne kroki niezbędne do budowy modelu ekonometrycznego mającego za zadanie opisanie zależności pomiędzy kosztami a wielkością produkcji w kopalni węgla kamiennego. Zadanie sprowadza się do poszukiwania modelu liniowego z jedną zmienną objaśniającą. Przedstawiono matematyczny zapis problemu oraz możliwość zastosowania modelu do wyznaczenia udziału kosztów zmiennych w analizowanym koszcie produkcji. Opracowano schemat budowy i weryfikacji takiego modelu w oparciu o literaturę tematu, uwzględniając jednocześnie specyfikę wykorzystanych narzędzi wspomagających proces modelowania Program Statistica v. 6.0. Zastosowanie tego programu wspomaga proces weryfikacji modelu, gdyż program sam wyznacza szereg statystyk powalających na ominięcie procesu ich obliczania.

Do szacowania modelu zastosowano metodę najmniejszych kwadratów. Dlatego szczególną uwagę zwrócono na problem weryfikacji modelu, tak by spełniał on założenia leżące u podstaw stosowania tej metody. Zaproponowano weryfikację parametrów strukturalnych modelu oraz wszechstronną weryfikację reszt. Schemat pełnej weryfikacji modelu składa się z 13 kroków. W każdym z kroków opisano procedurę niezbędną dla zweryfikowania badanej cechy modelu oraz podano warunek jaki musi być spełniony, by można było przejść do kroku następnego. W przypadku braku pozytywnej weryfikacji w dowolnym z kroków należy albo zmodyfikować serię danych wejściowych, albo zrezygnować z modelu. Podano warunki eliminacji niektórych danych wejściowych.

Przedstawiony schemat zilustrowano budując model zależności kosztów wynagrodzeń od wydobycia dla jednej z kopalń węgla kamiennego. Model ten został zweryfikowany pozytywnie i został użyty do określenia udziału kosztów zmiennych w kosztach wynagrodzeń w tej kopalni.

CONSTRUCTION AND VALIDATION OF ECONOMETRIC MODEL OF LINEAR DEPENDENCE BETWEEN COSTS AND COAL PRODUCTION LEVEL

Key words

Econometric models, linear model, cost of coal production

Abstract

The paper presents the sequence of steps that are necessary to construct an econometric model aimed to describe the dependency between cost and production in coal mine. The task is delimited to linear model with one explanatory variable. The mathematic equation of the problem is given and the possibility of model application for evaluation the share of variable costs in analysed costs is shown. The diagram of the process of construction and validation of the model was worked out basing on a wide bibliography and it takes into account the usage of Statistica v. 6.0 program. The program facilitates the process of model verification as it estimates a range of statistics.

The least square method is applied for the model estimation. That is why the special attention is paid to the validation of the model to assure that the assumptions needed for application of the method are fulfilled. Verification of the structural parameters of the model is advised as well as multidimensional check out of residual errors. The scheme of full verification of the model consists of 13 steps. In each step the procedure of validation the particular feature of the model components is described and the criteria of positive verification are given. Only the positive result of validation entitles for going to the next step of the process. In case of negative verification at any step it is necessary to modify the set of input data or reformulate the model. The conditions of rejection of some input data are given.

The procedure is illustrated with the process of construction and validation of the model where the costs of salaries are dependent from coal production level in one of coal mines. The model was fully verified and finally is used for evaluation the share of variable costs in salary costs in the coal mine.