

Ekonometryczny model zużycia paliw ciekłych

Econometric model of country consumption of liquid fuels

Tadeusz Kwilosz, Bogdan Filar

Institut Nafty i Gazu – Państwowy Instytut Badawczy

STRESZCZENIE: Jednym z priorytetowych zagadnień dla państwa jest zapewnienie stabilnego rozwoju gospodarczego, który jest ściśle związany z bezpieczeństwem energetycznym kraju. Problem zagwarantowania bezpieczeństwa energetycznego Polski staje się bardzo istotny z powodu dużego importu ropy naftowej i gotowych produktów. W 2017 roku przerób ropy naftowej przez rafinerie krajowe wyniósł 26,8 mln ton. Krajowe wydobycie ropy naftowej w 2017 roku wynosiło około 1 mln ton i stanowiło zaledwie 3,7% przerobu. W związku z tym w Polsce występuje bardzo wysoki stopień uzależnienia od importu ropy naftowej, wynoszący 96,3%. Drugi niekorzystny czynnik zwiększający ryzyko związane z dostawami ropy naftowej do Polski to bardzo duże uzależnienie od jednego kierunku dostaw (Paliński i Łucki, 2006). Zgodnie z publikacją Polskiej Organizacji Przemysłu i Handlu Naftowego (POPiHN) aż 81,4% importu ropy naftowej pochodzi z Rosji. W sytuacji tak dużego uzależnienia od jednego kierunku dostaw bardzo istotne staje się zgromadzenie rezerw interwencyjnych ropy naftowej i produktów naftowych, gwarantujących dostawy przez wymagane 90 dni. W związku z tym, że budowa magazynów paliw ciekłych trwa co najmniej 2–3 lata, odpowiednie prace analityczne powinny wyprzedzać prace budowlane również co najmniej o kilka lat. W zaprezentowanym artykule przedstawiono wyniki zastosowania zbudowanego modelu matematycznego służącego do prognozowania zużycia paliw w kolejnych latach. Zaprojektowano i skalibrowano w oparciu o wybrany zestaw atrybutów (zmiennych objaśniających) niejednorodny liniowy model ekonometryczny. Dokonano statystycznej weryfikacji oszacowanych parametrów modelu. Do budowy modelu wykorzystano podstawowe wskaźniki makroekonomiczne. Zakres danych historycznych obejmował lata 2000–2017. Ze względu na dużą liczbę kombinacji, jaką należy przebadać dla każdego z modeli ($2^{17} - 1$) opisane wyżej algorytmy zostały zaimplementowane w postaci programu komputerowego, za pomocą którego dokonano niezbędnych obliczeń.

Słowa kluczowe: rynek paliw ciekłych, model ekonometryczny, prognozowanie.

ABSTRACT: Ensuring the stable development of the country is one of the most important tasks facing the government. The stable development of the country is closely related to its energy security. The problem of ensuring Poland's energy security becomes very important due to the large import of crude oil and liquid fuels. In 2017, the refining of crude oil by domestic refiners amounted to 26.8 million tones. Domestic crude oil production in 2017 amounted to approximately 1 million tons and accounted for only 3.7% of processing. Therefore, in Poland there is a high dependence on crude oil imports (96.3%). The second unfavorable factor increasing the risk of crude oil supplies to Poland lies in the very large dependence on one direction of supply (Paliński and Łucki, 2006). According to POPiHN, up to 81.4% of crude oil imports come from Russia. In the situation of such a large dependence on one direction of supply, it becomes very important to collect emergency reserves of crude oil and petroleum products, guaranteeing deliveries for the required 90 days. Due to the fact that the construction of liquid fuel storage facilities lasts at least 2–3 years, appropriate analytical works should precede the construction works also at least by several years. The article presents the results of the application of the mathematical model used to predict fuel consumption in subsequent years. A non-uniform linear econometric model was designed and calibrated based on a selected set of attributes (explanatory variables). A statistical verification of estimated model parameters was carried out. Basic macroeconomic indicators have been used to build the model. The historical data covered the years 2000–2017. Due to the large number of combinations to be tested for each of the models ($2^{17} - 1$), the algorithms described above were implemented in the form of a computer program which was used to make all the necessary calculations.

Keywords: liquid fuels market, econometric model, forecasting.

Wybór modelu i podstawowe założenia

Zasadniczym celem zaprezentowania przedmiotowego artykułu było opisanie procesu budowy modelu matematycznego pozwalającego na prognozowanie zużycia paliw ciekłych w Polsce. Zużycie tych paliw w kraju charakteryzuje się pewną zmiennością i w związku z tym poznanie i opisanie zależności mających wpływ na ten sektor gospodarki nastęrcza sporych trudności analitykom rynku energii (Gajda, 2001; Paliński i Łucki, 2006). Istnieje wiele czynników wpływających na ten stan rzeczy, chociażby rosnące zainteresowanie źródłami odnawialnymi (Rogowska, 2017, 2018). Kwestia krajowego zużycia paliw ciekłych jest zagadnieniem z dziedziny makroekonomii. W modelowaniu tego typu problemów dobre efekty przynosi zastosowanie modeli ekonometrycznych (Łucki, 1988; Cieślak, 1996; Szmigiel i Mercik, 2000; Osińska, 2006; Ye et al., 2006; Zagaglia, 2010). Modele ekonometryczne opisują wzajemne zależności między badanymi cechami, które umożliwiają lepsze zrozumienie mechanizmów rządzących analizowanym fragmentem rzeczywistości, a także przewidywanie zachowania modelowanych procesów (Gładysz i Mercik, 2007; Łucki et al., 2008; Zhang et al., 2008).

Skonstruowano cztery modele ekonometryczne opisujące kolejno: zużycie benzyny silnikowej, oleju napędowego, oleju opałowego lekkiego i oleju opałowego ciężkiego i dokonano ich weryfikacji. Jako krok czasowy wybrano jeden rok, w związku z czym nie uwzględniono sezonowych wahań opisywanego zjawiska.

Jednym z podstawowych etapów budowy modelu ekonometrycznego jest dobór tzw. zmiennych objaśniających, czyli zmiennych zależnych modelu. Na tym etapie należy zdecydować o tym, jakie zmienne ilościowe (związane z badanym zjawiskiem) najlepiej je opisują i jaka ma być ich liczba. Jeśli chodzi o pierwsze zagadnienie, to literatura przedmiotu (Welfe, 1998; Zeliaś et al., 2003) sugeruje, że do opisu zjawisk makroekonomicznych najlepiej zastosować podstawowe wskaźniki związane z tą dziedziną, takie jak: parametry rozwoju gospodarczego, ceny surowców, zużycie konsumpcyjne, podaż pieniądza, inwestycje itp.

W przypadku rozstrzygnięcia o liczbie zmiennych modelu (szczególnie w przypadku modeli liniowych) ich liczba nie może przekraczać liczebności serii danych. W opisywanym przypadku dane historyczne obejmują lata: 2000–2017 (Berent-Kowalska et al., 2017a, 2017b; Polska Organizacja Przemysłu i Handlu Naftowego, 2017; Siewierski et al., 2017; Tkaczyk et al., 2017; Portal Geostatystyczny, 2018), czyli zmiennych objaśniających nie może być więcej niż 18. Faktycznie powinno ich być istotnie mniej, gdyż różnica pomiędzy liczbą danych (n) a liczbą zmiennych (k), zwana liczbą stopni swobody modelu, ma istotny wpływ na

statystyczną weryfikację poprawności oszacowanych parametrów (Wydimus, 1984). W opisywanych modelach wybrano wstępnie do analizy 18 zmiennych objaśniających.

Związki ilościowe występujące w zjawiskach z dziedziny makroekonomii mają w przeważającej liczbie przypadków charakter liniowy (Zieliński, 1995; Gawlik, 2008). Z tego względu, budując model ekonometryczny, wybrano liniowy model niejednorodny z wieloma zmiennymi objaśniającymi.

Niejednorodny liniowy model ekonometryczny z wieloma zmiennymi objaśniającymi można zapisać w postaci (Goryl et al., 2009):

$$y_t = \alpha_0 + \alpha_1 X_{t1} + \alpha_2 X_{t2} + \dots + \alpha_k X_{tk} + \varepsilon_t$$

gdzie:

y_t – zmienna objaśniana o wartościach empirycznych w okresie t ,

X_{ij} – zmienna objaśniająca o wartościach empirycznych w okresie t ,

α_j – nieznane parametry modelu,

ε_t – składnik losowy.

Powyższy model można zapisać w postaci macierzowej:

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad \mathbf{X} = \begin{bmatrix} 1 & x_{11} & \dots & x_{1k} \\ 1 & x_{21} & \dots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \dots & x_{nk} \end{bmatrix} \quad \boldsymbol{\alpha} = \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_k \end{bmatrix} \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

lub

$$\mathbf{y} = \mathbf{X}\boldsymbol{\alpha} + \boldsymbol{\varepsilon}$$

Po oszacowaniu parametrów strukturalnych model przyjmuje postać:

$$\hat{y}_t = \alpha_0 + \alpha_1 X_{t1} + \alpha_2 X_{t2} + \dots + \alpha_k X_{tk}$$

gdzie:

\hat{y}_t – teoretyczna wartość zmiennej objaśnianej w okresie t ,

α_j – wyliczony estymator nieznanego parametru α_j .

Wyrażenie:

$$e_t = y_t - \hat{y}_t$$

nazywane jest resztą modelu.

Oszacowanie nieznanych parametrów modelu przeprowadza się metodą najmniejszych kwadratów, polegającą na wyznaczeniu minimum sumy kwadratów reszt modelu:

$$\sum_{t=1}^n e_t^2 \rightarrow \min$$

Funkcja ta osiąga minimum w punkcie:

$$\boldsymbol{a} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$$

Po wstępnym wyborze zmiennych objaśniających modelu należy dokonać ich oceny i ewentualnej eliminacji. Zmienne

objaśniające należy tak dobierać, aby: (i) ich wzajemna korelacja liniowa była jak najmniejsza oraz (ii) ich korelacja ze zmienną objaśniającą była istotna. Jako miarę owej korelacji zastosowano współczynnik Pearsona. Drugim czynnikiem oceny jakości zmiennych objaśnianych jest ich współczynnik zmienności:

$$V = \frac{s}{\bar{X}} 100\%$$

gdzie:

\bar{s} – odchylenie standardowe z próby,

\bar{X} – średnia z próby.

Przyjmuje się arbitralnie, że wartość ta powinna przekraczać 10%.

Kolejnym kryterium doboru zmiennych objaśniających modelu jest metoda pojemności informacyjnej Hellwiga. Polega ona na:

- wyznaczeniu wszystkich $k = 2^{n-1}$ podzbiorów zbioru zmiennych objaśniających,
- przypisaniu każdemu z tak wyznaczonych podzbiorów (k) integralnej pojemności nośników informacyjnych:

$$H_k = \sum_j h_{kj}$$

gdzie:

$h_{ki} = \sum_j h_{kj} \frac{r_{0j}^2}{|r_{ij}|}$ – indywidualna pojemność j -tej zmiennej

w k -tej kombinacji (podzbiórze),

r_{0j} – współczynnik korelacji j -tej zmiennej objaśniającej ze zmienną objaśnianą,

r_{ij} – współczynnik korelacji j -tej zmiennej objaśniającej z i -tą zmienną występującą w tej kombinacji (zbiorze).

Parametr H_k przyjmuje wartość z przedziału $<0, 1>$. Im jest ona większa, tym wytypowana kombinacja zmiennych lepiej opisuje modelowane zjawisko. Na użytek tej pracy przyjęto, że wartość H_k nie może być mniejsza od 0,4.

Po wyznaczeniu parametrów strukturalnych modelu należy dokonać jego weryfikacji. Służy do tego celu szereg parametrów:

- Miarą dopasowania oszacowanego modelu do danych empirycznych jest współczynnik determinacji:

$$R^2 = 1 - \frac{\sum_{t=1}^n e_t^2}{\sum_{t=1}^n (y_t - \bar{y})^2}$$

gdzie:

\bar{y} – średnia arytmetyczna z serii danych empirycznych zmiennej objaśnianej. Współczynnik ten przyjmuje wartości z przedziału $<0, 1>$ i informuje, jaka część całkowitej zaobserwowanej zmienności zmiennej objaśnianej została wyjaśniona przez model.

- Nieobciążonym estymatorem wariancji składnika losowego jest wariancja resztowa:

$$S_e^2 = \frac{\sum_{t=1}^n e_t^2}{n - k - 1}$$

Pierwiastek kwadratowy z wariancji resztowej S_e nazywany jest odchyleniem resztowym modelu. Informuje on, o ile średnio wartości teoretyczne modelu różnią się od wartości empirycznych zmiennej objaśnianej.

- Współczynnik zmienności resztowej informuje, o ile średnio wartości zmiennej objaśnianej stanowią odchylenia losowe:

$$V_e = \frac{S_e}{\bar{y}} 100\%$$

Zakłada się, że estymatory a_j parametrów α_j mają rozkład normalny:

$$a_j: N(\alpha_j, D(a_j))$$

Estymatorem nieznanego odchylenia standardowego $D(a_j)$ jest:

$$S(a_j) = S_e \sqrt{c_{jj}}$$

gdzie:

c_{jj} – elementy głównej przekątnej macierzy $(X^T X)^{-1}$. Mając oszacowane średnie błędy estymatorów, można zapisać:

$$\alpha_j = a_j \pm S(a_j)$$

- Miarą oceny trafności doboru poszczególnych zmiennych objaśniających jest test istotności pojedynczej zmiennej objaśnianej przeprowadzany w formie weryfikacji hipotezy statystycznej (Domański, 1990; Szmigiel i Mercik, 2000; Gajda, 2001).

Zakładamy hipotezę zerową $H_0: \alpha_j = 0$, oznaczającą, że zmienna x_j nie ma wpływu na wartość zmiennej y , przeciw hipotezie alternatywnej $H_1: \alpha_j \neq 0$, oznaczającej, że zmienna x_j ma wpływ na wartość zmiennej y . Sprawdzianem hipotezy jest ocena statystyki:

$$t(a_j) = \frac{a_j}{S(a_j)}$$

W przypadku gdy:

$$|t(a_j)| > t_{\alpha, Q}$$

odrzucaamy hipotezę zerową (H_0) na rzecz hipotezy alternatywnej (H_1), tzn. zmienna x_j ma istotny (statystycznie) wpływ na zmienną y . W przeciwnym razie nie ma przesłanek do odrzucenia hipotezy zerowej.

Parametr $t_{\alpha, Q}$ jest wartością krytyczną rozkładu t -Studenta dla poziomu istotności α oraz dla $Q = n - k - 1$ stopni swobody. Dla celów tego opracowania przyjęto $\alpha = 0,05$.

Realizacja modeli

Skonstruowano i dokonano weryfikacji czterech modeli ekonometrycznych opisujących kolejno: zużycie benzyny silnikowej, oleju napędowego, oleju opałowego lekkiego i oleju opałowego ciężkiego. Wymienione atrybuty są, kolejno, zmiennymi objaśnianymi konstruowanych modeli. Jako krok czasowy wybrano jeden rok, w związku z czym nie uwzględniono sezonowych wahań opisywanego zjawiska. Wstępnie wybrano 19 zmiennych (18 objaśniających i jedną objaśnianą):

- y/x_0 – zużycie benzyny silnikowej [tys. t],
- x_1 – zużycie (przerób) ropy naftowej [tys. t],
- y/x_2 – zużycie oleju napędowego [tys. t],
- y/x_3 – zużycie oleju opałowego ciężkiego [tys. t],
- y/x_4 – zużycie oleju opałowego lekkiego [tys. t],
- x_5 – średnie roczne ceny ropy Brent [PLN/baryłka],
- x_6 – nominalny PKB – ceny stałe [tys. PLN],
- x_7 – podaż pieniądza (stan z końca roku) [mln PLN],
- x_8 – stopa oprocentowania kredytu lombardowego (stan z końca roku) [%],
- x_9 – stopa referencyjna (stan z końca roku) [%],
- x_{10} – sprzedaż detaliczna towarów ogółem (ceny stałe) [rok 1999 = 100],
- x_{11} – transport samochodowy [tys. t],
- x_{12} – przewozy pasażerów ogółem [tys. osób],
- x_{13} – produkcja sprzedana przemysłu ogółem (ceny stałe) [rok 1999 = 100],

- x_{14} – nakłady inwestycyjne (ceny bieżące) [mln PLN],
- x_{15} – nakłady inwestycyjne na środki transportu (ceny stałe) [rok 1999 = 100],
- x_{16} – dochód realny do dyspozycji brutto w sektorze gospodarstw domowych ogółem [rok 1999 = 100],
- x_{17} – przeciętne miesięczne wynagrodzenia realne brutto w gospodarce narodowej [rok 1999 = 100],
- x_{18} – transport w zakresie spożycia indywidualnego (z dochodów osobistych) w sektorze gospodarstw domowych (ceny bieżące) [rok 1999 = 100].

Należy zauważyć, że zmienne: x_0 , x_2 , x_3 i x_4 w jednych modelach pełnią rolę zmiennych objaśnianych, a w innych – objaśniających.

W celu oszacowania wymaganej zmienności zmiennych wyliczono współczynniki zmienności V poszczególnych zmiennych. Tylko w jednym przypadku (benzyny silnikowej) współczynnik V ma wartość (8,6%) poniżej 10%. Ze względu na istotne znaczenie merytoryczne (zużycie jednego z nośników energii) nie usunięto tej zmiennej z modelu.

Dla sprawdzenia korelacji liniowych między poszczególnymi zmiennymi wyliczono współczynniki korelacji Pearsona (tabela 7). Największą wartość współczynnika korelacji ($r = 0,998$) wyliczono dla pary zmiennych x_8 i x_9 . Zmienna x_8 została usunięta z modelu. Mimo usunięcia zmiennej zachowano numerację pozostałych zmiennych.

Pomimo występujących korelacji nie zdecydowano się (na tym etapie) na usunięcie większej liczby zmiennych

Tabela 1. Wartości empiryczne (GUS) zmiennych modelu

Table 1. Empirical values of the model variables (GUS)

Rok	x_0	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}	x_{17}	x_{18}
2000	5080	18080	5173	2809	2222	124	684982	300757	23,0	19,0	101	1006705	1319972	107	133160	122	99	101	101
2001	4740	18113	4724	2583	2289	100	750786	329705	15,5	11,5	101	996517	1236583	107	121363	100	101	104	99
2002	4170	17878	4866	2356	2355	102	781112	326125	8,8	6,8	103	931190	1124940	108	109266	84	101	104	103
2003	4040	17947	5700	2163	2375	112	816081	345145	6,8	5,3	107	911997	1112533	117	110860	85	102	108	102
2004	4198	18068	6748	2354	2649	140	923948	377535	8,0	6,5	110	956939	1085509	132	120467	100	103	109	106
2005	4065	18191	7405	1975	2297	176	983302	427125	6,0	4,5	107	1079761	1046930	137	131055	106	104	111	103
2006	4158	20050	8527	1943	1742	202	1060031	495310	5,5	4,0	115	1113880	1024413	153	154880	136	109	115	106
2007	4144	20024	9530	1839	1346	200	1176737	561624	6,5	5,0	123	1213246	1006369	169	191714	179	114	121	122
2008	4416	21036	10441	1804	1252	233	1275432	666231	6,5	5,0	129	1339473	966001	175	217260	191	120	128	141
2009	4242	20282	10810	1661	1140	192	1344037	720233	5,0	3,5	132	1424883	902176	168	218581	158	126	131	142
2010	4141	22839	11687	1681	1145	240	1416447	783649	5,0	3,5	130	1491253	838024	183	217287	173	128	133	146
2011	3946	24169	12235	1406	1080	330	1528127	881496	6,0	4,5	135	1596209	807141	196	243346	186	130	135	156
2012	3867	25153	12018	1212	923	364	1596378	921413	5,8	4,3	135	1493386	779797	197	237627	165	131	135	158
2013	3659	24202	11171	986	789	343	1662678	978908	4,0	2,5	137	1553050	739556	201	231155	178	132	139	161
2014	3578	24132	11001	868	729	312	1719146	1059015	3,0	2,0	142	1547883	709698	209	250776	209	136	143	172
2015	3777	26108	11937	1118	658	197	1799321	1154993	2,5	1,5	150	1505719	653835	222	271839	227	140	149	183
2016	4026	25777	14471	1049	627	172	1788832	1265675	2,5	1,5	159	1546572	652433	230	244429	193	149	156	189
2017	4156	26797	17393	1084	643	205	1982080	1324383	2,5	1,5	167	1747266	642135	245	252298	229	162	161	196
X_{ref}	5080	26797	17393	2809	2649	364	1982080	1324383	23	19	167	1747266	1319972	245	271839	229	162	161	196

Tabela 2. Unormowane wartości empiryczne zmiennych modelu

Table 2. Standard empirical values of the model variables

Rok	x_0	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}	x_{17}	x_{18}
2000	1,00	0,67	0,30	1,00	0,84	0,34	0,35	0,23	1,00	1,00	0,60	0,58	1,00	0,44	0,49	0,53	0,61	0,63	0,51
2001	0,93	0,68	0,27	0,92	0,86	0,28	0,38	0,25	0,67	0,61	0,61	0,57	0,94	0,44	0,45	0,44	0,62	0,64	0,51
2002	0,82	0,67	0,28	0,84	0,89	0,28	0,39	0,25	0,38	0,36	0,62	0,53	0,85	0,44	0,40	0,37	0,63	0,65	0,52
2003	0,80	0,67	0,33	0,77	0,90	0,31	0,41	0,26	0,29	0,28	0,64	0,52	0,84	0,48	0,41	0,37	0,63	0,67	0,52
2004	0,83	0,67	0,39	0,84	1,00	0,38	0,47	0,29	0,35	0,34	0,66	0,55	0,82	0,54	0,44	0,44	0,63	0,67	0,54
2005	0,80	0,68	0,43	0,70	0,87	0,48	0,50	0,32	0,26	0,24	0,64	0,62	0,79	0,56	0,48	0,46	0,64	0,69	0,53
2006	0,82	0,75	0,49	0,69	0,66	0,56	0,53	0,37	0,24	0,21	0,69	0,64	0,78	0,62	0,57	0,59	0,67	0,71	0,54
2007	0,82	0,75	0,55	0,65	0,51	0,55	0,59	0,42	0,28	0,26	0,74	0,69	0,76	0,69	0,71	0,78	0,71	0,75	0,62
2008	0,87	0,79	0,60	0,64	0,47	0,64	0,64	0,50	0,28	0,26	0,77	0,77	0,73	0,72	0,80	0,83	0,74	0,80	0,72
2009	0,84	0,76	0,62	0,59	0,43	0,53	0,68	0,54	0,22	0,18	0,79	0,82	0,68	0,68	0,80	0,69	0,78	0,81	0,72
2010	0,82	0,85	0,67	0,60	0,43	0,66	0,71	0,59	0,22	0,18	0,78	0,85	0,63	0,75	0,80	0,76	0,79	0,82	0,74
2011	0,78	0,90	0,70	0,50	0,41	0,91	0,77	0,67	0,26	0,24	0,81	0,91	0,61	0,80	0,90	0,81	0,80	0,84	0,79
2012	0,76	0,94	0,69	0,43	0,35	1,00	0,81	0,70	0,25	0,22	0,81	0,85	0,59	0,81	0,87	0,72	0,81	0,84	0,80
2013	0,72	0,90	0,64	0,35	0,30	0,94	0,84	0,74	0,17	0,13	0,82	0,89	0,56	0,82	0,85	0,78	0,81	0,86	0,82
2014	0,70	0,90	0,63	0,31	0,28	0,86	0,87	0,80	0,13	0,11	0,85	0,89	0,54	0,85	0,92	0,91	0,84	0,89	0,88
2015	0,74	0,97	0,69	0,40	0,25	0,54	0,91	0,87	0,11	0,08	0,90	0,86	0,50	0,91	1,00	0,99	0,87	0,93	0,93
2016	0,79	0,96	0,83	0,37	0,24	0,47	0,90	0,96	0,11	0,08	0,95	0,89	0,49	0,94	0,90	0,84	0,92	0,97	0,96
2017	0,82	1,00	1,00	0,39	0,24	0,56	1,00	1,00	0,11	0,08	1,00	1,00	0,49	1,00	0,93	1,00	1,00	1,00	1,00
X_{ref}	5080	18080	5173	2809	2222	124	684982	300757	23,0	19,0	101	1006705	1319972	107	133160	122	99	101	101

z modelu, aby nie utracić istotnych informacji opisujących zmienne objaśniane. Ze względu na duże różnice w rzędach wielkości wartości poszczególnych zmiennych, w celu uniknięcia błędów wynikających z obliczeń numerycznych, wartości wszystkich zmiennych unormowano, przyjmując wartość 1 dla maksymalnej wartości z serii danych danej zmiennej. W celu korzystania z opracowanych modeli w wartościach realnych zachowano informacje o wartościach referencyjnych x_{ref} (odpowiadających unormowanej wartości 1) dla każdej zmiennej.

W trakcie testowania każdego z czterech modeli poddano testom wszystkie $2^{17} - 1$ kombinacji zmiennych objaśniających opisujących konstruowane modele. W pracach nad oceną poprawności i dopasowania modeli przyjęto (dla każdej testowanej kombinacji zmiennych) następujące założenia:

- współczynnik integralnej pojemności nośników informacyjnych $H_k > 0,4$;
- współczynnik determinacji $R^2 > 0,6$;
- wszystkie oszacowane parametry strukturalne modelu dla danej kombinacji zmiennych muszą spełniać kryterium istotności ($|t(a_j)| > t_{\alpha, q}$).

Spośród modeli z kombinacjami danych spełniających powyższe kryteria wybrano te o wysokich współczynnikach determinacji i na ich podstawie wyliczono wartości zmiennych teoretycznych (y) w zakresie obejmującym dostępne dane empiryczne i w zakresie objętym prognozą

(lata 2018–2025). Wartości zmiennych objaśniających uzyskano, ekstrapolując (do prognozowanego zakresu) dopasowane funkcje. Posłużono się funkcjami liniowymi i wykładniczymi. Ze względu na dużą liczbę kombinacji, jaką należy przebadać dla każdego z modeli ($2^{17} - 1$), opisane wyżej algorytmy zostały zaimplementowane w postaci programu komputerowego, za pomocą którego dokonano niezbędnych obliczeń.

Model zużycia benzyny silnikowej

W tym modelu zmienną objaśnianą jest zużycie benzyny silnikowej. Jako początkowy zestaw zmiennych objaśniających wybrano zestaw zmiennych zaprezentowany wcześniej z zastrzeżeniem, że zmienna oznaczona y/x_0 pełni rolę zmiennej objaśnianej y . Zatem pierwsze cztery zmienne objaśniające oznaczono jako:

- x_1 – zużycie (przerób) ropy naftowej [tys. t],
- x_2 – zużycie oleju napędowego [tys. t],
- x_3 – zużycie oleju opałowego ciężkiego [tys. t],
- x_4 – zużycie oleju opałowego lekkiego [tys. t].

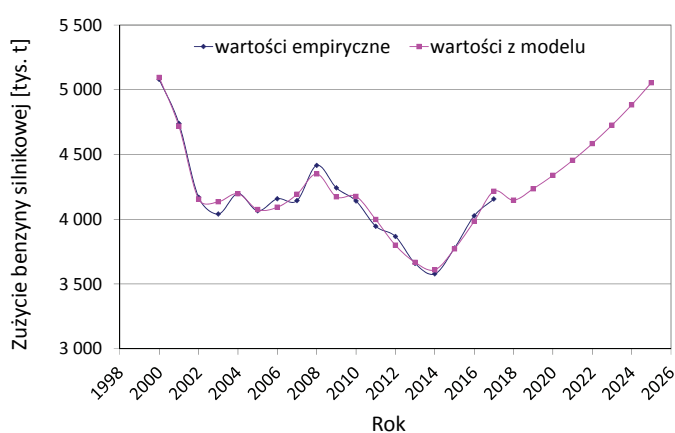
Dla wszystkich wygenerowanych kombinacji (podzbiorów zmiennych objaśniających) dokonano oszacowania parametrów strukturalnych modelu i poddano je ocenie poprawności i istotności. Wszystkie założone kryteria spełniają

kombinacje zawierające: 4, 5, 6, 7 i 8 zmiennych objaśniających. Kombinacje zawierające więcej niż 8 zmiennych nie spełniają założonych kryteriów oceny. Dla każdej liczebności zmiennych wybrano modele charakteryzujące się największym współczynnikiem determinacji R^2 . Uzyskane wyniki zaprezentowano w tabeli 2. Numerem „0” oznaczono parametr wolny modelu (model niejednorodny). Każdy z zaprezentowanych modeli jest statystycznie poprawny i może być stosowany do prognozowania zużycia benzyny silnikowej. W celu jego wykorzystania do prognozowania zmiennej y , należy wykonać prognozy każdej zmiennej objaśniającej z osobna. Wybór modelu może zależeć od dostępności i jakości prognoz tych zmiennych. Jako przykład zaprezentowano (rys. 1, tab. 3) prognozę dla 6 zmiennych objaśniających $\{x_2, x_3, x_8, x_{10}, x_{12}, x_{17}\}$.

Tabela 3. Wyniki oszacowania i oceny parametrów strukturalnych modelu zużycia benzyny silnikowej

Table 3. The results of estimation and assessment of the structural parameters of the petrol consumption model

Model z 6 zmiennymi objaśniającymi			
H	R^2	S_e	V_e
0,534	0,981	0,012	0,015
X_i	a_i	$t(ai)$	$ta_{,q}$
0	-0,774	-3,377	2,228
2	0,169	2,991	2,228
3	0,266	3,987	2,228
8	0,085	2,783	2,228
10	-0,841	-2,234	2,228
12	0,760	4,514	2,228
17	1,794	3,661	2,228



Rys. 1. Prognoza zużycia benzyny silnikowej w latach 2018–2025

Fig. 1. Petrol consumption forecast in the years 2018–2025

Model zużycia oleju napędowego

W tym modelu zmienną objaśnianą y jest zużycie oleju napędowego (zmienna x_2 z pierwotnego zestawu danych). Zatem pierwsze cztery zmienne objaśniające oznaczono jako:

- x_1 – zużycie (przerób) ropy naftowej [tys. t],
- x_2 – zużycie benzyny silnikowej [tys. t],
- x_3 – zużycie oleju opałowego ciężkiego [tys. t],
- x_4 – zużycie oleju opałowego lekkiego [tys. t].

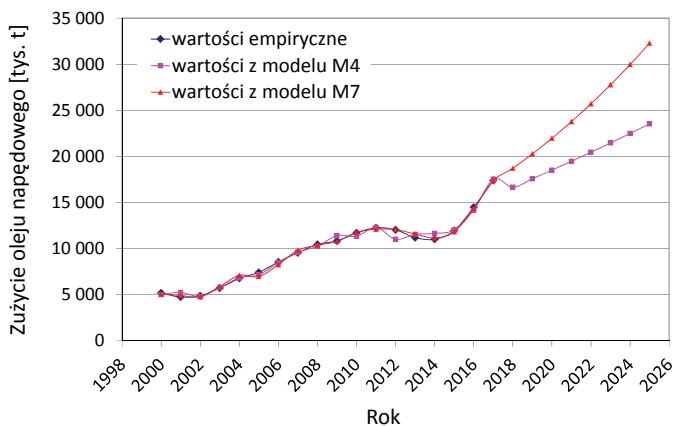
Przeprowadzono procedurę jak dla wcześniej opisanego modelu. Wszystkie założone kryteria spełniają kombinacje zawierające: 4, 5, 6, 7 i 8 zmiennych objaśniających. Kombinacje zawierające więcej niż 8 zmiennych nie spełniają założonych kryteriów oceny.

W celu zademonstrowania, w jaki sposób ekstrapolacja modelu do prognozy zależy od oszacowania prognozowanych zmiennych objaśniających, zaprezentowano (rys. 2, tab. 4) prognozę dla dwóch modeli. Pierwszego – oznaczonego jako M4 – dla kombinacji 4 zmiennych objaśniających $\{x_2, x_{10}, x_{11}, x_{18}\}$ oraz dla modelu oznaczonego jako M7 – dla 7 zmiennych objaśniających $\{x_7, x_8, x_{11}, x_{13}, x_{15}, x_{16}, x_{17}\}$. Na rysunku widać, że obydwa modele dobrze odtwarzają wartość zmiennej empirycznej i znacząco się różnią w obszarze prognozowanym.

Tabela 4. Wyniki oszacowania i oceny parametrów strukturalnych modelu zużycia oleju napędowego

Table 4. The results of estimation and assessment of the structural parameters of the diesel fuel consumption model

Model z 4 zmiennymi objaśniającymi			
H	R^2	S_e	V_e
0,914	0,988	0,025	0,045
X_i	a_i	$t(ai)$	$ta_{,q}$
0	-1,394	-7,894	2,179
4	0,208	2,475	2,179
10	2,595	9,078	2,179
11	1,004	6,560	2,179
18	-1,248	-5,567	2,179
Model z 7 zmiennymi objaśniającymi			
H	R^2	S_e	V_e
0,932	0,996	0,016	0,029
X_i	a_i	$t(ai)$	$ta_{,q}$
0	-2,089	-8,875	2,262
7	-1,529	-9,325	2,262
8	0,185	3,771	2,262
11	0,238	2,627	2,262
13	1,793	8,882	2,262
15	-0,439	-4,960	2,262
16	1,608	4,404	2,262
17	1,395	2,292	2,262



Rys. 2. Prognoza zużycia oleju napędowego w latach 2018–2025
Fig. 2. Diesel oil consumption forecast in the years 2018–2025

Model zużycia oleju opałowego ciężkiego

W tym modelu zmienną objaśnianą jest zużycie oleju opałowego ciężkiego (zmienna x_3 z pierwotnego zestawu danych). Zatem pierwsze cztery zmienne objaśniające oznaczono jako:

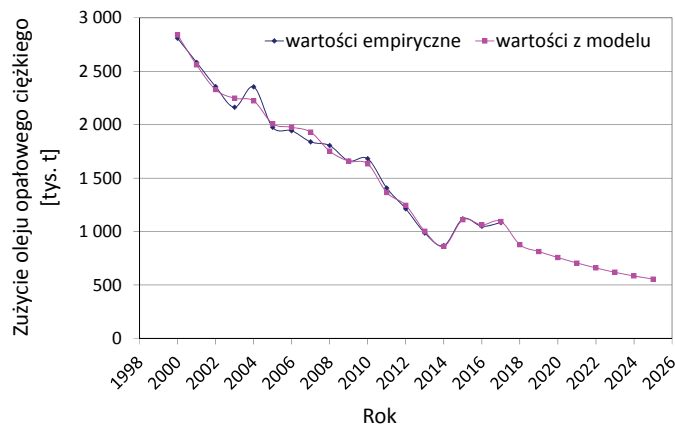
- x_1 – zużycie (przerób) ropy naftowej [tys. t],
- x_2 – zużycie benzyny silnikowej [tys. t],
- x_3 – zużycie oleju napędowego [tys. t],
- x_4 – zużycie oleju opałowego lekkiego [tys. t].

Przeprowadzono procedurę jak dla wcześniej opisanych modeli. Wszystkie założone kryteria spełniają kombinacje zawierające: 4, 5, 6, 7, 8 i 9 zmiennych objaśniających. Kombinacje zawierające więcej niż 9 zmiennych nie spełniają założonych kryteriów oceny. Jako przykład zastosowania modelu do prognozowania wybrano kombinację 5 zmiennych $\{x_1, x_3, x_5, x_7, x_8\}$ (rys. 3, tab. 5). Charakterystyczne dla tego zestawu zmiennych jest to, że zawiera on zmienną x_1 , oznaczającą zużycie (przerób) ropy. Zmienna ta rzadko pojawia się w kombinacjach spełniających oceny poprawności i istotności.

Tabela 5. Wyniki oszacowania i oceny parametrów strukturalnych modelu zużycia oleju opałowego ciężkiego

Table 5. The results of estimation and assessment of the structural parameters of the heavy heating oil consumption model

Model z 5 zmiennymi objaśniającymi			
H	R^2	S_e	V_e
0,946	0,992	0,022	0,036
X_i	ai	$t(ai)$	$ta_{,0}$
0	0,524	4,000	2,201
1	0,649	2,558	2,201
3	0,352	4,393	2,201
5	-0,315	-8,577	2,201
7	-0,988	-7,769	2,201
8	0,276	7,561	2,201



Rys. 3. Prognoza zużycia oleju opałowego ciężkiego w latach 2018–2025

Fig. 3. Heavy heating oil consumption forecast in the years 2018–2025

Model zużycia oleju opałowego lekkiego

W tym modelu zmienną objaśnianą jest zużycie oleju opałowego lekkiego (zmienna x_4 z pierwotnego zestawu danych). Zatem pierwsze cztery zmienne objaśniające oznaczono jako:

- x_1 – zużycie (przerób) ropy naftowej [tys. t],
- x_2 – zużycie benzyny silnikowej [tys. t],
- x_3 – zużycie oleju napędowego [tys. t],
- x_4 – zużycie oleju opałowego ciężkiego [tys. t].

Przeprowadzono procedurę jak dla wcześniej opisanych modeli. Wszystkie założone kryteria spełniają kombinacje zawierające: 4, 5, 6 i 7 zmiennych objaśniających. Kombinacje zawierające więcej niż 7 zmiennych nie spełniają założonych kryteriów oceny. Jako przykład zastosowania modelu do prognozowania wybrano kombinację 7 zmiennych $\{x_3, x_5, x_6, x_{14}, x_{16}, x_{17}, x_{18}\}$ (rys. 4, tab. 6). Charakterystyczne dla tego

Tabela 6. Wyniki oszacowania i oceny parametrów strukturalnych modelu zużycia oleju opałowego lekkiego

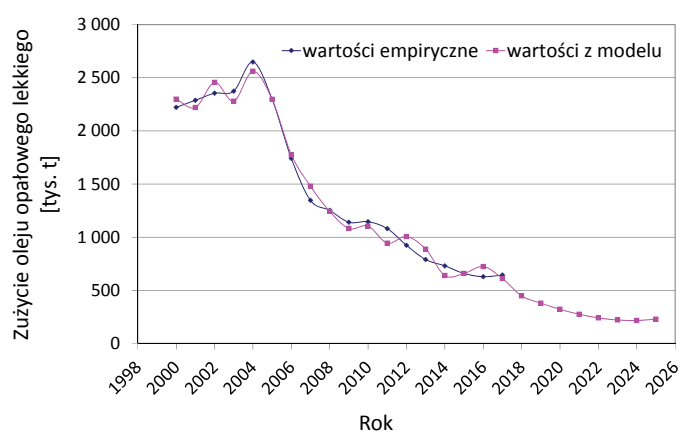
Table 6. The results of estimation and assessment of the structural parameters of the light heating oil consumption model

Model z 7 zmiennymi objaśniającymi			
H	R^2	S_e	V_e
0,932	0,987	0,040	0,073
X_i	ai	$t(ai)$	$ta_{,0}$
0	4,403	6,104	2,262
3	1,178	3,603	2,262
5	-0,407	-2,862	2,262
6	1,575	2,501	2,262
14	-1,194	-5,553	2,262
16	-3,241	-3,053	2,262
17	-4,624	-3,000	2,262
18	2,276	3,471	2,262

Tabela 7. Współczynniki korelacji liniowej Pearsona dla zmiennych modelu**Table 7.** Pearson's correlation coefficients for the model variables

	Y	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈	X ₉	X ₁₀	X ₁₁	X ₁₂	X ₁₃	X ₁₄	X ₁₅	X ₁₆	X ₁₇	X ₁₈
X ₁	-0,59	1,00	0,92	-0,92	-0,93	0,67	0,97	0,98	-0,63	-0,63	0,95	0,94	-0,95	0,96	0,94	0,89	0,96	0,96	0,97
X ₂	-0,49	0,92	1,00	-0,87	-0,90	0,60	0,95	0,94	-0,70	-0,69	0,97	0,95	-0,92	0,97	0,90	0,89	0,97	0,96	0,93
X ₃	0,80	-0,92	-0,87	1,00	0,92	-0,75	-0,96	-0,93	0,81	0,82	-0,90	-0,90	0,98	-0,94	-0,90	-0,84	-0,90	-0,93	-0,92
X ₄	0,55	-0,93	-0,90	0,92	1,00	-0,70	-0,95	-0,94	0,63	0,64	-0,93	-0,96	0,91	-0,95	-0,98	-0,94	-0,92	-0,94	-0,94
X ₅	-0,63	0,67	0,60	-0,75	-0,70	1,00	0,68	0,59	-0,48	-0,48	0,53	0,74	-0,66	0,65	0,73	0,62	0,54	0,57	0,58
X ₆	-0,66	0,97	0,95	-0,96	-0,95	0,68	1,00	0,99	-0,74	-0,74	0,98	0,96	-0,98	0,99	0,95	0,92	0,97	0,99	0,98
X ₇	-0,59	0,98	0,94	-0,93	-0,94	0,59	0,99	1,00	-0,68	-0,68	0,98	0,94	-0,97	0,98	0,93	0,90	0,99	0,99	0,99
X ₈	0,86	-0,63	-0,70	0,81	0,63	-0,48	-0,74	-0,68	1,00	1,00	-0,70	-0,62	0,83	-0,74	-0,62	-0,57	-0,67	-0,72	-0,65
X ₉	0,86	-0,63	-0,69	0,82	0,64	-0,48	-0,74	-0,68	1,00	1,00	-0,70	-0,63	0,83	-0,74	-0,62	-0,57	-0,67	-0,72	-0,65
X ₁₀	-0,54	0,95	0,97	-0,90	-0,93	0,53	0,98	0,98	-0,70	-0,70	1,00	0,93	-0,95	0,98	0,92	0,91	0,99	1,00	0,98
X ₁₁	-0,53	0,94	0,95	-0,90	-0,96	0,74	0,96	0,94	-0,62	-0,63	0,93	1,00	-0,92	0,95	0,96	0,92	0,94	0,95	0,95
X ₁₂	0,76	-0,95	-0,92	0,98	0,91	-0,66	-0,98	-0,97	0,83	0,83	-0,95	-0,92	1,00	-0,97	-0,92	-0,86	-0,94	-0,97	-0,95
X ₁₃	-0,62	0,96	0,97	-0,94	-0,95	0,65	0,99	0,98	-0,74	-0,74	0,98	0,95	-0,97	1,00	0,95	0,94	0,97	0,99	0,97
X ₁₄	-0,55	0,94	0,90	-0,90	-0,98	0,73	0,95	0,93	-0,62	-0,62	0,92	0,96	-0,92	0,95	1,00	0,96	0,91	0,94	0,94
X ₁₅	-0,45	0,89	0,89	-0,84	-0,94	0,62	0,92	0,90	-0,57	-0,57	0,91	0,92	-0,86	0,94	0,96	1,00	0,89	0,92	0,91
X ₁₆	-0,52	0,96	0,97	-0,90	-0,92	0,54	0,97	0,99	-0,67	-0,67	0,99	0,94	-0,94	0,97	0,91	0,89	1,00	0,99	0,98
X ₁₇	-0,58	0,96	0,96	-0,93	-0,94	0,57	0,99	0,99	-0,72	-0,72	1,00	0,95	-0,97	0,99	0,94	0,92	0,99	1,00	0,99
X ₁₈	-0,56	0,97	0,93	-0,92	-0,94	0,58	0,98	0,99	-0,65	-0,65	0,98	0,95	-0,95	0,97	0,94	0,91	0,98	0,99	1,00

zestawu zmiennych jest to, że zawiera on zmienne x_6 (nominalny PKB) oraz x_{16} (dochód realny do dyspozycji brutto w sektorze gospodarstw domowych). Zestawy z tymi zmiennymi nie były wcześniej prezentowane (jako przykłady zastosowania do prognozowania) dla modeli innych zmiennych objaśnianych.

**Rys. 4.** Prognoza zużycia oleju opałowego lekkiego w latach 2018–2025**Fig. 4.** Light heating oil consumption forecast in the years 2018–2025

Wnioski

1. Zbudowane i zaprezentowane modele ekonometryczne w wystarczającym stopniu wyjaśniają zależności kształtowania się zużycia paliw ciekłych w latach 2000–2017.

Świadczą o tym uzyskane współczynniki determinacji z przedziału 0,961–0,998.

- Zbiór zaproponowanych zmiennych objaśnianych daje możliwości wyboru wielu kombinacji zmiennych, dla których spełnione jest kryterium dostatecznej integralnej pojemności informacyjnej i jednocześnie spełniają kryteria istotności.
- Pomimo dobrego dopasowania wartości teoretycznych do empirycznych w okresach historycznych – jakość wykonanych prognoz przy użyciu modeli w zasadniczy sposób zależy od jakości oszacowania prognozowanych zmiennych objaśniających.
- Zastosowanie liniowych modeli ekonometrycznych w modelowaniu zużycia paliw ciekłych jest obiecującym rozwiązaniem i daje dobre podstawy do dalszych prac związanych z modelowaniem tego zjawiska.

Artykuł powstał na podstawie pracy statutowej pt.: *Opracowanie modelu matematycznego dla prognozowania zużycia paliw płynnych w Polsce* – praca INiG – PIB na zlecenie MNiSW; nr zlecenia: 14/KW/18, nr archiwalny: DK-4100-6/2018/01.

Literatura

- Berent-Kowalska G., Gilecki R., Kacprowska J., Peryt Sz., Żarek E., Boczek R., Brasse J., 2017a. Zużycie paliw i nośników energii w 2016 r. Główny Urząd Statystyczny, Warszawa.
- Berent-Kowalska G., Wnuk R., Peryt S., Gilewski P., 2017b. Efektywność wykorzystania energii w latach 2005–2015. Informacje i opracowania statystyczne. Główny Urząd Statystyczny, Warszawa.

- Cieślak M., 1996. Prognozowanie gospodarcze. Wydawnictwo Akademii Ekonomicznej im. Oskara Langego, Wrocław.
- Domański C., 1990. Testy statystyczne. PWE, Warszawa.
- Gajda J., 2001. Prognozowanie i symulacje a decyzje gospodarcze. Wydawnictwo C.H. Beck, Warszawa.
- Gawlik L., 2008. Budowa i weryfikacja modelu ekonometrycznego dla określenia liniowej zależności pomiędzy kosztami pozyskania węgla a wielkością wydobycia. *Gospodarka Surowcami Mineralnymi*, 24(1): 27–44.
- Gładysz B., Mercik J., 2007. Modelowanie ekonometryczne: studium przypadku. Oficyna Wydawnicza Politechniki Wrocławskiej, Wrocław.
- Goryl A., Jędrzejczyk Z., Kukuła K., Osiewalski J., Walkosz A., 2009. Wprowadzenie do ekonometrii. Wydawnictwo Naukowe PWN, Warszawa.
- Łucki Z., 1988. Modele ekonometryczne do zarządzania górnictwem naftowym. *Zeszyty Naukowe AGH. Zagadnienia Techniczno-Ekonomiczne*, 41.
- Łucki Z. (red.), Byrska-Rapała A., Kozarkiewicz A., 2008. Statystyka i ekonometria. AGH Uczelniane Wydawnictwa Naukowo-Dydaktyczne, Kraków.
- Osińska M., 2006. Ekonometria finansowa. PWE, Warszawa.
- Paliński A., Łucki Z., 2006. Problemy zarządzania finansowego i zarządzania ryzykiem w sektorze handlu energią. *Polityka Energetyczna*, 2: 129–152.
- Polska Organizacja Przemysłu i Handlu Naftowego, 2017. Przemysł i Handel Naftowy – Raport Roczny 2016. POPiHN, Warszawa 2017.
- Portal Geostatystyczny, 2018. <<https://geo.stat.gov.pl>> (dostęp: 2.07.2018).
- Rogowska D., 2017. Wykorzystanie OZE w energetyce a zrównoważony rozwój. *Nafta-Gaz*, 8: 616–623. DOI: 10.18668/NG.2017.08.09.
- Rogowska D., 2018. Produkcja biopaliw jako element gospodarki o obiegu zamkniętym. *Nafta-Gaz*, 2: 156–163. DOI: 10.18668/NG.2018.02.10.
- Siewierski J. et al., 2017. Raport Polski Rynek Paliw Płynnych 2016. Polska Izba Paliw Płynnych, Warszawa.
- Szmigiel C., Mercik J., 2000. Ekonometria. Wydawnictwo Wyższej Szkoły Zarządzania i Finansów we Wrocławiu, Wrocław.
- Tkaczyk W., Kacprowska J., Gilecki R., Berent-Kowalska G. et al., 2017. Gospodarka paliwowo-energetyczna w latach 2015 i 2016. Informacje i opracowania statystyczne. Główny Urząd Statystyczny, Warszawa.
- Wydymus S., 1984. Zmienne syntetyczne w modelowaniu ekonometrycznym. *Ruch Prawniczy, Ekonomiczny i Socjologiczny*, 1: 187–200.
- Welfe A., 1998. Ekonometria. Metody i ich zastosowania. PWE, Warszawa.
- Ye M., Zyren J., Shore J., 2006. Forecasting short-run crude oil price using high-and low-inventory variables. *Energy Policy*, 34: 2736–2743.
- Zagaglia P., 2010. Macroeconomic factors and oil future prices: a data-rich model. *Energy economics*, 32(2): 409–417.
- Zhang X., Lai K.K., Wang S.Y., 2008. A new approach for crude oil price analysis based on empirical mode decomposition. *Energy Economics*, 30(3): 905–918.
- Zeliaś A., Pawełek B., Wanat S., 2003. Prognozowanie ekonomiczne. Teoria, przykłady, zadania. PWN, Warszawa.
- Zieliński Z., 1995. Liniowe modele zgodnie opisujące zależności symulacyjnych (zintegrowanych) procesów ekonomicznych. Wydawnictwo Akademii Ekonomicznej, Kraków.



Dr Tadeusz KWIŁOSZ
Adiunkt w Zakładzie Podziemnego
Magazynowania Gazu
Instytut Nafty i Gazu – Państwowy Instytut Badawczy
ul. Lubicz 25 A
31-503 Kraków
E-mail: tadeusz.kwilosz@inig.pl



Mgr inż. Bogdan FILAR
Kierownik Zakładu Podziemnego
Magazynowania Gazu
Instytut Nafty i Gazu – Państwowy Instytut Badawczy
ul. Lubicz 25 A
31-503 Kraków
E-mail: bogdan.filar@inig.pl