

Kazimierz DAŚAL*, Tomasz POPLAWSKI**

Problemy związane z prognozowaniem zużycia energii elektrycznej w Polsce

STRESZCZENIE. Mówiąc o prognozowaniu zużycia energii elektrycznej można rozważać ten problem w różnych kategoriach. W podziale czasowym będą to najczęściej prognozy krótko, średnio- i długookresowe. Chociaż podkreślić należy wyraźnie względność tego podziału. Horyzont roczny dla potrzeb Giełdy będzie prognozą długoterminową, dla innego odbiorcy – krótkookresową. Można wyróżnić prognozy w podziale strukturalnym, napięciowym, taryfowym, terytorialnym czy też w węzłach sieci na różnych poziomach napięć [2, 6]. Każdy z tych podziałów ma swoją specyfikę, wymaga różnych modeli, różnych danych wejściowych, różnić może się oceną dokładności prognoz, odmienne na ogół są też cele prognoz krótkoterminowych i długoterminowych. Znajomość prognoz długoterminowych jest niezbędna głównie dla celów racjonalnego prowadzenia eksploatacji systemu elektroenergetycznego, a także do planowania jego rozwoju. W artykule skoncentrowano się na jeszcze jednym ważnym aspekcie modelowania w ogóle, a w szczególności energii elektrycznej. Chodzi tu o podjęcie decyzji: modelować z wykorzystaniem jedynie procesu, czyli szeregu czasowego, czy też budować model przyczynowo-skutkowy.

SŁOWA KLUCZOWE: krótkoterminowe i długoterminowe prognozowanie w elektroenergetyce, szeregi czasowe, model prognostyczny

* Dr inż., ** Dr hab. inż. – Instytut Elektroenergetyki, Politechnika Częstochowska, Zakład Inżynierii Materiałów Elektrotechnicznych i Gospodarki Elektroenergetycznej, Częstochowa.

1. Charakterystyka problemu

Rozważając problem modelowania przebiegu zużycia energii elektrycznej możliwe są dwa rozwiązania. Można modelować przebieg jedynie w oparciu o szereg czasowy lub budować model przyczynowo-skutkowy. Są to dwa zasadniczo różniące się podejścia do problemu modelowania zjawisk technicznych czy szerzej, gospodarczych [1]. Każde z nich ma pozytywne jak i negatywne cechy. Pokróćce można te cechy scharakteryzować następująco.

Model szeregu czasowego nie wymaga dodatkowych zmiennych objaśniających, których zebranie może być kosztowne i trudne, a niezbędne do wykonania modelu. Te własności takiego modelowania można uznać za pozytywne. Jednakże ze względu na określone wymogi stawiane modelowaniu szeregów czasowych, należy dokładniej wniknąć w strukturę procesu, czyli stacjonarność, własności funkcji korelacyjnych, funkcje częstości, funkcje gęstości. Wskazane jest, aby wyodrębnić z procesu ewentualne składowe takie jak trend, wahania okresowe lub cykliczne [4]. Zbadanie tych własności i uwzględnienie ich w procesie implementacji modelu jest trudne i wymaga dużej wiedzy oraz doświadczenia.

W modelowaniu przyczynowo-skutkowym zderzamy się z trudnym problemem typowania zbioru potencjalnych zmiennych objaśniających badany proces zużycia energii, ich dostępnością na etapie budowy modelu i później w czasie eksploatacji modelu.

Następnym, jednym z najtrudniejszych zagadnień, a jednocześnie bardzo istotnym dla jakości modelu, jest zagadnienie doboru merytorycznego składowych spośród zbioru potencjalnego [3]. Bardzo ważną własnością modelowania takiego modelowania jest możliwość zbadania wpływu innych czynników na przebieg badanego procesu, wykonanie wielu symulacji umożliwiających wnikliwą analizę wpływu różnych wielkości na badany proces. Takich możliwości nie mają modele szeregów czasowych. Problemem jest to, że do wykonania prognozy badanego procesu, już po zbudowaniu modelu, konieczna jest znajomość prognoz lub scenariuszy zmiennych objaśniających. Dokładność scenariuszy, czy też tych prognoz, w sposób oczywisty wpływa na dokładność prognoz końcowych.

Do modelowania i sporządzania prognoz zużycia energii elektrycznej w horyzontach długoterminowych, w zależności od celu, dostępnego materiału statystycznego, komplikacji zadania, używa się różnych technik i modeli prognostycznych.

Przykładowo prognozy rozwoju sektora energii opracowane przez Ministerstwo Gospodarki [8] oraz Instytut Ochrony Środowiska [9] wykonano modelem MAED (*Model of Analysis Energy Demand*). Model ten korzysta z metodologii opracowanej i wprowadzonej przez Międzynarodową Agencję Energii Atomowej (IAEA). Jest szeroko stosowana w długookresowych (powyżej 10 lat) badaniach energetycznych. Zakłada się w niej, że najistotniejszą przyczyną wzrostu zapotrzebowania na energię w ogóle, w tym na energię elektryczną jest ogólny wzrost gospodarczy opisany za pomocą wielu zmiennych makroekonomicznych z różnych sektorów gospodarczych, a także uwzględniająca założenia polityki energetycznej. Wyniki tego modelu stanowią wejście do modelu energetyczno-ekologicznego BALANCE, który wyznacza zapotrzebowanie na energię finalną w podziale na

poszczególne nośniki oraz wyznacza krajowe bilanse energii oraz wielkości emisji zanieczyszczeń przy założonych cenach paliw pierwotnych, postępie technologicznym, ograniczeniach w dostępie do nośników energii, ograniczeniach inwestycyjnych. Prognozy rozwoju systemu elektroenergetycznego wykonano modelem WASP. Model wyznacza koszty produkcji, koszty energii niedostarczonej, poziom niezawodności, ekonomiczny rozdział obciążeń i inne wielkości. Z tego krótkiego opisu wynika, że jest to modelowanie o bardzo dużej złożoności, elastyczne w sensie symulacji, jednak wymagające wielu szczegółowych danych wejściowych różnego typu, co jest kosztowne i kłopotliwe. Inny sposób modelowania to modele ekstrapolacji trendów, modele hybrydowe, (np. AI+ szeregi czasowe) [5] i wiele innych.

W pracy zaprezentowano dwa, wcześniej wymienione sposoby modelowania zużycia energii elektrycznej tzn. modelowanie szeregami oraz model przyczynowo-skutkowy. Przedstawiono rezultaty i wnioski wynikające z wykonanych badań. W badaniach użyto następujących danych, obejmujących okres od 1973 do 1997 roku:

Y – zużycie energii elektrycznej w Polsce przez gospodarstwa domowe [TW·h], jako badany proces, oraz potencjalne zmienne objaśniające:

X₁ – liczba ludności w Polsce [mln],

X₂ – import energii elektrycznej [TW·h],

X₃ – straty w sieci elektrycznej [TW·h],

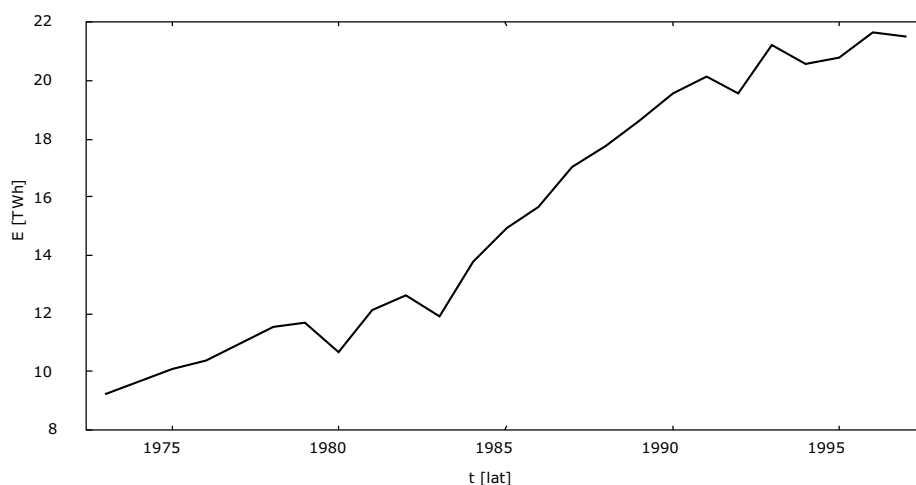
X₄ – przeciętne zatrudnienie w przemyśle energetycznym [tys.].

Wybór zbioru zmiennych objaśniających w rozważanym przykładzie wynikał z możliwości pozyskania danych. Autorzy mając świadomość, o mało fortunnym wyborze zmiennych X₂ oraz X₃, uważali, że procedura doboru wyeliminuje je, a jednocześnie uzmysłowi nawet na tak prostym przykładzie, licznosc kombinacji, dla których należy estymować parametry modeli w przypadku rezygnacji z istniejących procedur doboru zmiennych (tab. 3). Problemem doboru zmiennych jest fakt wykorzystywania prawie we wszystkich znanych metodach współczynnika korelacji liniowej w mniej lub bardziej jawnej postaci. Jak wiadomo, fakt istnienia silnej korelacji liniowej świadczy jedynie o liniowej zależności, zaś jej brak nie wyklucza jednak zależności nieliniowej. Inne zmienne zawierające informacje przykładowo o powierzchni mieszkań, zużyciu energii na oświetlenie czy w coraz większym stopniu na ogrzewanie w gospodarstwach, liczba urządzeń takich jak pralki, zmywarki, lodówki, urządzenia klimatyzacyjne w połączeniu dodatkowo z energochłonnością, byłyby z pewnością bardziej adekwatne dla prezentowanego przypadku.

Do modelowania, wykonania testów, analiz i obliczeń wykorzystano program Gretl. Posiada on licencję GNU (GNU *General Public License*, GPL) pozwalającą na swobodny dostęp każdemu użytkownikowi oraz możliwość dokonania własnych zmian bezpośrednio w kodzie źródłowym aplikacji. Oprogramowanie autorstwa Allin'a Cottrell'a z Uniwersytetu Wake Forest w Północnej Karolinie (USA) w pełnej wersji dostępne jest na stronie internetowej: <http://gretl.sourceforge.net>.

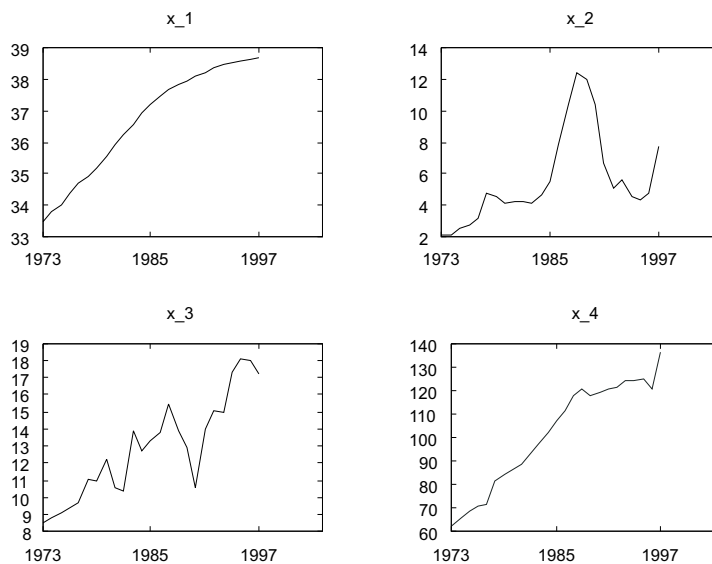
2. Model szeregu czasowego

Wykres badanego zużycia energii elektrycznej w gospodarstwach domowych przedstawia rysunek 1, a przebiegi potencjalnych zmiennych objaśniających rysunek 2. Modelując



Rys. 1. Przebieg zmiennej Y zużycie energii elektrycznej w Polsce w gospodarstwach domowych

Fig. 1. The dependence of variable Y – consumption of electric energy in Poland in households



Rys. 2. Wykresy ilustrujące przebiegi zmiennych objaśniających

Fig. 2. The charts depicting the dependencies for explanatory variables

szereg czasowy nie korzystamy z wiedzy ekonomicznej, a tylko analizujemy własności statystyczne szeregu czasowego. Nie ma w tym przypadku zmiennych niezależnych. Założono zastosowanie dynamicznego modelu autoregresyjnego. Założenie jest uzasadnione, gdy występuje korelacja pomiędzy wartościami zmiennej prognozowanej, a jej wartościami opóźnionymi w czasie. Model taki oznacza się symbolicznie AR(p) (*AutoRegressive*), p jest rzędem opóźnienia. W modelowaniu autoregresyjnym można także uwzględnić proces średniej ruchomej będącej uogólnieniem białego szumu oznaczany symbolicznie MA(q) (*Moving Average*), q jest operatorem opóźnienia dla składników resztowych. Popaczenie tych dwóch modeli tworzy model ARMA(p, q), który można zapisać ogólnie w następujący sposób:

$$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \alpha_3 Y_{t-3} + \dots + \alpha_p Y_{t-p} + \quad \{\text{AR}(p)\}$$

$$+ \beta_1 \varepsilon_{t-1} + \beta_2 \varepsilon_{t-2} + \beta_3 \varepsilon_{t-3} + \dots + \beta_q \varepsilon_{t-q} + \varepsilon_t \quad \{\text{MA}(q)\}$$

Modele AR, MA czy ARMA mogą być stosowane do procesów stacjonarnych. Jeżeli proces jest niestacjonarny (najczęściej), wówczas musi być przekształcony do stacjonarnego. Sprowadzanie do stacjonarności najczęściej wykonywane jest poprzez d -krotne różnicowanie szeregu wyrazów szeregu źródłowego. Tak tworzone modele określa się odpowiednio mianem zintegrowanych modeli ARI, IMA oraz w połączeniu ARIMA. Wszystkie te modele mogą być opisane przy użyciu uniwersalnej notacji ARIMA (p, d, q), w której p oznacza rząd autoregresji (wielkość opóźnienia), d – krotność różnicowania, q – liczbę parametrów (wielkość opóźnienia) średniej ruchomej. Wybór odpowiedniego modelu dokonuje się w oparciu o testy oraz badanie funkcji autokorelacji ACF i autokorelacji cząstkowej PACF.

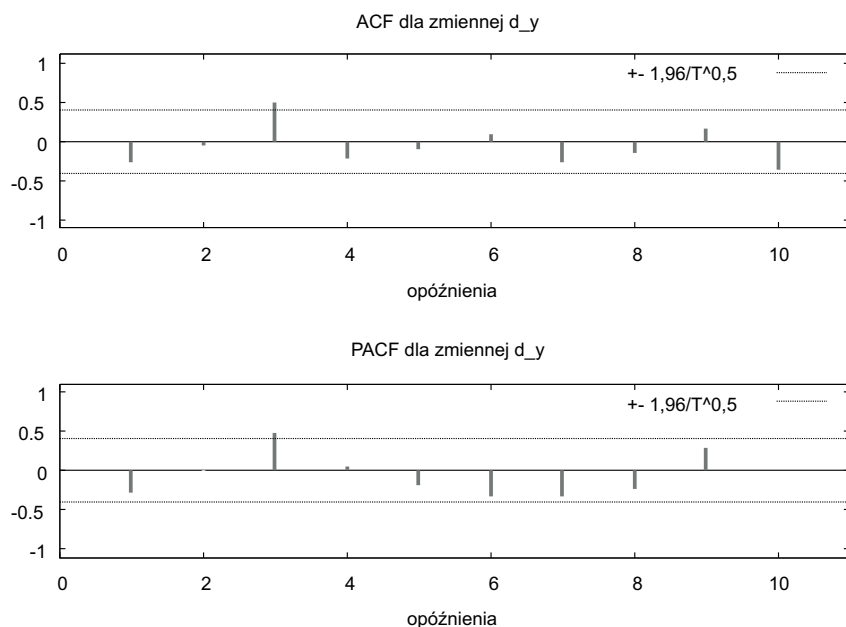
W przedstawianym przykładzie, aby stwierdzić, czy będzie to model ARMA, czy ARIMA wykonano testy stacjonarności. Test Dickeya_Fullera wykazał stacjonarność dla różnic pierwszego rzędu, wobec czego zastosowano model w ogólnej postaci ARIMA ($p, 1, q$) z różnicowaniem pierwszego rzędu, czyli $d = 1$. Dla zdefiniowania parametrów p i q modelu wyznaczono funkcje autokorelacji, ACF oraz autokorelacji cząstkowej PACF. Wykresy tych funkcji przedstawiono na rysunku 3.

Analizując przebiegi funkcji z rysunku 3 zauważa się, że wartości współczynników autokorelacji, zarówno funkcji ACF jak i funkcji PACF, urywają się po 3 opóźnieniu, co wskazuje na następującą postać modelu:

$$Y_t = \text{ARIMA}(3,1,3) \quad (1)$$

Niestety przebiegi funkcji są jedynie wskazówką, która nie zawsze zapewnia sukces, w sensie zbudowania poprawnego modelu. Dla modelu przyjętego w wyniku analizy funkcji korelacyjnych, estymowane parametry okazały się nieistotne, wobec tego model należy odrzucić. Z przeprowadzonych wielu prób, modelem, który może być zaakceptowany okazał się model w postaci:

$$Y_t = \text{ARIMA}(1,1,0) \quad (2)$$



Rys. 3. Wykresy funkcji autokorelacji ACF oraz autokorelacji cząstkowej PACF dla szeregu pierwszych różnic zmiennej Y

Fig. 3. The charts of autocorrelation function ACF and partial autocorrelation PACF for the series of first differences of variable Y

Wyznaczając parametry modelu pominięto 3 ostatnie obserwacje (lata) z szeregu. Wykorzystano je w ocenie dokładności modelu wyznaczając prognozę sprawdzają. Wyniki estymacji modelu ARIMA(1,1,0) są następujące:

Błędy standardowe na bazie Outer Products matrix

| Zmienna | Współczynnik | Błąd stand. | Statystyka t | Wartość p |
|---------|--------------|-------------|----------------|-------------|
| const | 0,555888 | 0,119895 | 4,6364 | <0,00001 |
| phi_1 | -0,315134 | 0,148648 | -2,1200 | 0,001550 |

Średnia arytmetyczna zmiennej zależnej = 0,540714

Odchylenie standardowe zmiennej zależnej = 0,769505

Średnia z zaburzeń losowych = -0,0018591

Wariancja z zaburzeń losowych = 0,50974

Kryterium informacyjne Akaike'a = 51,5491

Kryterium bayesowskie Schwarz = 54,6826

Kryterium infor. Hannana-Quinna = 52,2291

Błąd MAPE (*Mean Absolute Percentage Error*) zdefiniowany jest następująco:

$$MAPE_t = \frac{|y_t - \hat{y}_t|}{y_t} 100\%$$

gdzie: y_t – wartość rzeczywista procesu w chwili t ,
 \hat{y}_t – wartość procesu w chwili t uzyskana z modelu

Średni błąd dla n wartości:

$$\text{MAPE} = \frac{\sum \text{MAPE}_t}{n}$$

Rezultaty otrzymane z powyższego modelu zamieszczono w tabeli 1 i 2.

TABELA 1. Błędy MAPE prognozy wygasłej dla zużycia energii elektrycznej obciążeń gospodarstwach domowych według modelu wzór (2)

TABLE 1. MAPE errors of extinct forecast for the electric energy consumption in households according to the model, equation (2)

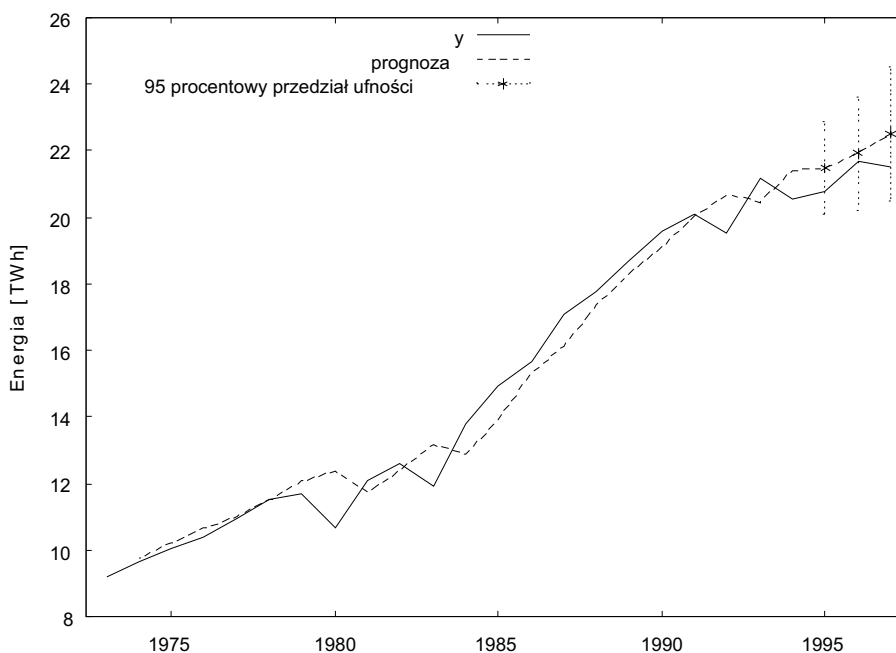
| Rok | Zużycie energii elektrycznej przez gospodarstwa domowe w Polsce [TW·h] | Prognoza wygasła zużycia energii elektrycznej przez gospodarstwa domowe w Polsce [TW·h] | MAPE _t [%] |
|----------------------------|--|---|-----------------------|
| 1973 | 9,20 | | |
| 1974 | 9,63 | 9,75 | 1,29% |
| 1975 | 10,06 | 10,22 | 1,65% |
| 1976 | 10,40 | 10,65 | 2,43% |
| 1977 | 10,94 | 11,02 | 0,76% |
| 1978 | 11,52 | 11,50 | 0,14% |
| 1979 | 11,70 | 12,07 | 3,11% |
| 1980 | 10,70 | 12,37 | 15,64% |
| 1981 | 12,10 | 11,75 | 2,92% |
| 1982 | 12,60 | 12,39 | 1,67% |
| 1983 | 11,93 | 13,17 | 10,42% |
| 1984 | 13,77 | 12,87 | 6,52% |
| 1985 | 14,94 | 13,92 | 6,82% |
| 1986 | 15,64 | 15,30 | 2,16% |
| 1987 | 17,06 | 16,15 | 5,33% |
| 1988 | 17,79 | 17,34 | 2,51% |
| 1989 | 18,66 | 18,29 | 1,98% |
| 1990 | 19,60 | 19,12 | 2,46% |
| 1991 | 20,12 | 20,04 | 0,42% |
| 1992 | 19,55 | 20,69 | 5,82% |
| 1993 | 21,20 | 20,46 | 3,49% |
| 1994 | 20,55 | 21,41 | 4,19% |
| Wartość średnia błędu MAPE | | | 3,89% |

TABELA 2. Błędy MAPE prognozy sprawdzającej dla zużycia energii elektrycznej obciążeń gospodarstw domowych według modelu wzór (2)

TABLE 2. MAPE errors of the checking forecast for the consumption of electric energy in households, according to the model equation (2)

| Rok | Zużycie energii elektrycznej w gospodarstwach domowych w Polsce [TW·h] | Wartość średnia prognozy wygasłej zużycia energii elektrycznej w gospodarstwach domowych w Polsce [TW·h] | Błąd ex ante 95% przedział ufności | Dolna wartość prognozy | Górna wartość prognozy | MAPE _t [%] |
|----------------------------|--|--|------------------------------------|------------------------|------------------------|-----------------------|
| 1995 | 20,78 | 21,49 | 0,71 | 20,09 | 22,89 | 3,40% |
| 1996 | 21,69 | 21,92 | 0,87 | 20,23 | 23,62 | 1,07% |
| 1997 | 21,54 | 22,52 | 1,03 | 20,50 | 24,54 | 4,53% |
| Wartość średnia błędu MAPE | | | | | | 3,00% |

Na rysunku 4 pokazano przebieg rzeczywisty zużycia energii elektrycznej w gospodarstwach domowych oraz przebieg wyznaczony przez model. Ostatnie 3 lata na wykresie są prognozą wraz 95% przedziałem ufności, będącym odpowiednikiem wyznaczenia prognozy dolnej i górnej lub inaczej prognozy pesymistycznej oraz optymistycznej.



Rys. 4. Wykres wartości rzeczywistych i uzyskanych z modelu ARIMA

Fig. 4. The depiction of true values and obtained from ARIMA model

3. Model ekonometryczny przyczynowo-skutkowy

W modelowaniu tym jednym z trudniejszych, a może najtrudniejszym zadaniem jest określenie potencjalnego zbioru zmiennych, a następnie wybór z tego zbioru zmiennych objaśniających. W prezentowanym przykładzie dysponowano 4 zmiennymi objaśniającymi. W [1, 2] podano opis wielu metod doboru składowych do modeli ekonometrycznych przyczynowo-skutkowych. Najpełniej opracowane są metody dla modeli liniowych lub sprowadzalnych do liniowych, jednorodnaniowych. Podstawowy postulat dotyczący zmiennych – silna korelacja między każdą zmienną objaśniającą a zmienną objaśnianą i jednocześnie słaba korelacja między zmiennymi objaśniającymi sprawia, iż zdecydowana większość metod doboru zmiennych wykorzystuje własności współczynnika korelacji liniowej. Rozwiązaniem uznawanym za optymalne jest jak najmniejsza liczba nieskorelowanych wzajemnie składowych objaśniających, a w możliwie maksymalnym stopniu skorelowanych ze zmienną objaśnianą, które z wymaganą dokładnością wyjaśniają wariację badanego procesu. Ta ogólna definicja przekłada się w literaturze omawiającej tę problematykę na zbiór kilkunastu, a z różnymi modyfikacjami, zbiór kilkudziesięciu metod doboru zmiennych. Liczba zmiennych objaśniających może być różna, zwykle jest ich kilka do kilkunastu. Pojawia się trudny problem wyboru zmiennych objaśniających. Dla modeli liniowych bardzo często stosowana jest metoda Hellwiga [4] polegająca na wyborze kombinacji zmiennych według maksymalnego wskaźnika pojemności integralnej spośród wszystkich możliwych kombinacji nośników informacji. Nawet dla modeli, dla których takie postępowanie jest możliwe, metoda ta nie gwarantuje wyboru optymalnego, w sensie błędu, dopasowania. Ponieważ żadna z istniejących metod tej gwarancji nie daje, pozostaje do rozważenia wariant polegający na wykonaniu modeli dla wszystkich możliwych kombinacji. Jest to droga niezwykle praco- i czasochłonna. Istnieją także modele, dla których liczba możliwych kombinacji wynika nie z liczby kombinacji bez powtórzeń, ale z liczby wariacji bez powtórzeń, a to oznacza znaczący wzrost liczby potencjalnych rozwiązań. W poniższej tabeli zamieszczono liczbowe dane ilustrujące omawiany.

Stosując metodę wyboru zmiennych Hellwiga możliwe są następujące kombinacje:

$K_1 = \{x_1\}$, $K_2 = \{x_2\}$, $K_3 = \{x_3\}$, $K_4 = \{x_4\}$, $K_5 = \{x_1x_2\}$, $K_6 = \{x_2x_3\}$, $K_7 = \{x_3x_4\}$, $K_8 = \{x_1x_3\}$, $K_9 = \{x_1x_4\}$, $K_{10} = \{x_2x_4\}$, $K_{11} = \{x_1x_2x_4\}$, $K_{12} = \{x_1x_2x_3\}$, $K_{13} = \{x_1x_3x_4\}$, $K_{14} = \{x_2x_3x_4\}$, $K_{15} = \{x_1x_2x_3x_4\}$.

Zakres wartości wyznaczonych pojemności integralnych dla wszystkich 15 kombinacji zawiera się w przedziale 0,32 dla kombinacji $K_2 = \{x_2\}$, do 0,91 dla kombinacji $K_9 = \{x_1x_4\}$. Wskazanie według metody Hellwiga pozostaje więc w zgodzie z intuicyjną oceną możliwego wpływu badanych zmiennych na zużycie energii przez gospodarstwa domowe. Model, dla którego należy estymować parametry strukturalne ma postać:

$$y = \alpha_0 + \alpha_1x_1 + \alpha_2x_4 + \varepsilon \quad (3)$$

TABELA 3. Liczba wariantów

TABLE 3. The number of options

| Liczba potencjalnych zmiennych | Liczba wariantów – kombinacje bez powtórzeń | Liczba wariantów – wariacje bez powtórzeń |
|--------------------------------|---|---|
| 1 | 1 | 1 |
| 2 | 3 | 4 |
| 3 | 7 | 15 |
| 4 | 15 | 64 |
| 5 | 31 | 325 |
| . | . | . |
| . | . | . |
| . | . | . |
| 10 | 1 023 | 9 864 100 |

Do wyznaczenia parametrów modelu także pominięto trzy ostatnie lata z szeregu obserwacji. Te ostatnie trzy obserwacje wykorzystano do wyznaczenia prognozy sprawdzającej. Wyniki estymacji modelu przyczynowo-skutkowego (3) są następujące:

Estymacja KMNK z wykorzystaniem 22 obserwacji z lat 1973–1994

Zmienna zależna: y

| Zmienna | Współczynnik | Błąd stand. | Statystyka t | Wartość p |
|---------|--------------|-------------|----------------|-------------|
| X_1 | -0,111399 | 0,045505 | -2,4481 | 0,02371 |
| X_4 | 0,18954 | 0,0165495 | 11,4529 | <0,00001 |

Średnia arytmetyczna zmiennej zależnej = 14,5292

Odchylenie standardowe zmiennej zależnej = 4,07998

Suma kwadratów reszt = 33,0427

Błąd standardowy reszt = 1,28535

Wsp. determinacji $R^2 = 0,993383$

Skorygowany $R^2 = 0,993052$

Statystyka $F(2, 20) = 1501,3$ (wartość $p < 0,00001$)

Statystyka testu Durбина-Watsona = 0,410626

Autokorelacja reszt rzędu pierwszego = 0,792033

Kryterium informacyjne Akaike'a = 75,382

Kryterium bayesowskie Schwarz'a = 77,564

Kryterium infor. Hannana-Quinna = 75,896

Rezultaty otrzymane z modelu przyczynowo-skutkowego zamieszczono w tabelach 4 i 5.

TABELA 4. Błędy MAPE prognozy wygasłej dla zużycia energii elektrycznej obciążeń gospodarstwach domowych według regresji wzór (3)

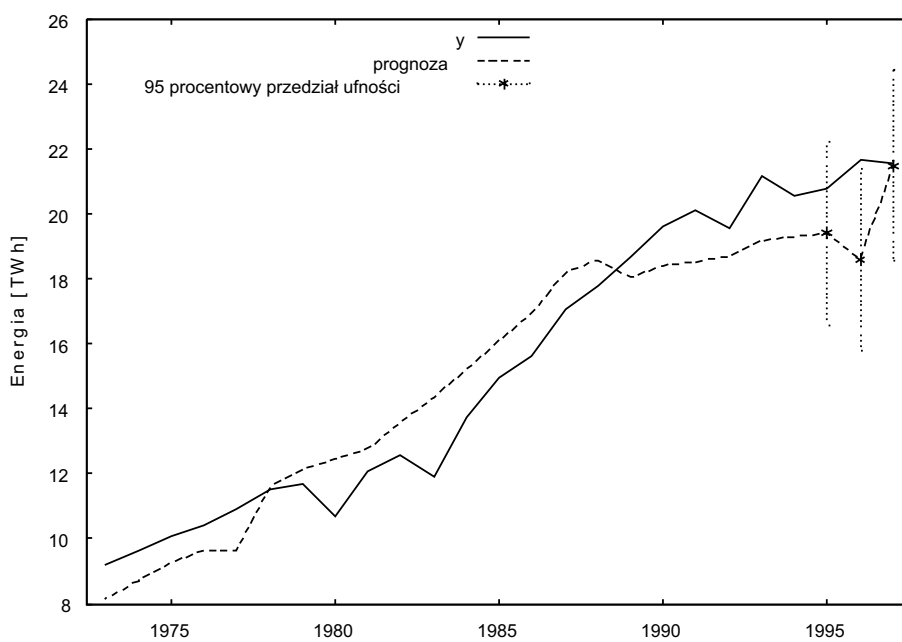
TABLE 4. MAPE errors of the extinct forecast for the consumption of electric energy in households, according to the regression equation (3)

| Rok | Zużycie energii elektrycznej przez gospodarstwa domowe w Polsce [TW·h] | Prognoza wygasła życia energii elektrycznej przez gospodarstwa domowe w Polsce [TW·h] | MAPE _t [%] |
|----------------------------|--|---|-----------------------|
| 1973 | 9,20 | 8,16 | 11,3% |
| 1974 | 9,63 | 8,71 | 9,5% |
| 1975 | 10,06 | 9,27 | 7,8% |
| 1976 | 10,40 | 9,66 | 7,2% |
| 1977 | 10,94 | 9,66 | 11,7% |
| 1978 | 11,52 | 11,59 | 0,7% |
| 1979 | 11,70 | 12,13 | 3,6% |
| 1980 | 10,70 | 12,47 | 16,5% |
| 1981 | 12,10 | 12,78 | 5,7% |
| 1982 | 12,60 | 13,59 | 7,9% |
| 1983 | 11,93 | 14,37 | 20,5% |
| 1984 | 13,77 | 15,20 | 10,4% |
| 1985 | 14,94 | 16,12 | 7,9% |
| 1986 | 15,64 | 16,94 | 8,3% |
| 1987 | 17,06 | 18,20 | 6,7% |
| 1988 | 17,79 | 18,58 | 4,5% |
| 1989 | 18,66 | 18,06 | 3,2% |
| 1990 | 19,60 | 18,42 | 6,0% |
| 1991 | 20,12 | 18,54 | 7,8% |
| 1992 | 19,55 | 18,70 | 4,4% |
| 1993 | 21,20 | 19,20 | 9,4% |
| 1994 | 20,55 | 19,30 | 6,1% |
| Wartość średnia błędu MAPE | | | 8,05% |

TABELA 5. Błędy MAPE prognozy sprawdzającej dla zużycia energii elektrycznej obciążeń gospodarstwach domowych według modelu regresji modelu wzór (3)

TABLE 5. MAPE errors of the checking forecast for the consumption of electric energy in households, according to the regression equation (3)

| Rok | Zużycie energii elektrycznej przez gospodarstwa domowe w Polsce [TW·h] | Wartość średnia prognozy wygasłej żyżycia energii elektrycznej w gospodarstwach domowych w Polsce [TW·h] | Błąd ex ante 95% przedział ufności | Dolna wartość prognozy | Górna wartość prognozy | MAPE [%] |
|----------------------------|--|--|------------------------------------|------------------------|------------------------|----------|
| 1995 | 20,78 | 19,41 | 1,36 | 16,58 | 22,25 | 6,6% |
| 1996 | 21,69 | 18,61 | 1,34 | 15,81 | 21,42 | 14,2% |
| 1997 | 21,54 | 21,50 | 1,42 | 18,55 | 24,46 | 0,2% |
| Wartość średnia błędu MAPE | | | | | | 6,98% |



Rys. 5. Wykres wartości rzeczywistych i uzyskanych z modelu ekonometrycznego

Fig. 5. The depiction of true values and obtained from econometric model

Podsumowanie i wnioski

W artykule zaprezentowano dwa odmienne podejścia do modelowania procesu zużycia energii elektrycznej. W przykładzie obliczeniowym wykorzystano dane dotyczące zużycia

energii elektrycznej w gospodarstwach domowych, ale rozważania można oczywiście uogólnić. Panuje dość powszechne przekonanie, że modele autoregresyjne wykazują swoje zalety w prognozowaniu krótkoterminowym. Zaprezentowany przykład osłabia ten pogląd. Prognozy sprawdzające wykonano na okres 3 lat, jest to zatem prognoza średnioterminowa. Model przyczynowo-skutkowy w porównaniu z modelem ARIMA, okazał się zdecydowanie gorszym, przynajmniej w sensie dokładności liczonej błędem MAPE. Z punktu widzenia statystycznego zmienne objaśniające są wybrane poprawnie, spełniają stawiane wymagania. Merytorycznie można mieć zastrzeżenia do zmiennych X_2 oraz X_3 , przy czym jedynie zmienna X_2 wykazuje niski (ok. 0,5) współczynnik korelacji ze zużyciem energii. Dobór opóźnień w modelach autoregresyjnych na podstawie oceny funkcji autokorelacyjnych nie jest metodą pewną – potwierdził to wykonany przykład. Podobną, krytyczną uwagę można wypowiedzieć na temat metod doboru zmiennych. Trafność wyboru wariantu metodą Hellwiga w rozważanym przykładzie nie jest regułą. W obecnej sytuacji mając różnorodne narzędzia programistyczne, w wielu przypadkach można wykonać modele dla wszystkich możliwych kombinacji zmiennych i wybrać optymalny, według przyjętego kryterium optymalności, którym najczęściej z punktu widzenia użytkownika jest średni błąd prognozy sprawdzającej określony wraz z przedziałem ufności. Oczywiście próby poszukiwania nowych metod wyznaczania opóźnień dla modeli dynamicznych oraz metod doboru zmiennych należy kontynuować choćby z racji poznawczych. W przypadkach dużej liczby zmiennych (opóźnień) wykonanie obliczeń dla wszystkich kombinacji może stanowić numeryczną i czasową przeszkodę. Wniosek natury ogólniejszej, który nasuwa się po analizie uzyskanych prognoz, wskazuje, że modelowanie procesów stochastycznych w bardzo dużym stopniu zależy od zastosowanej metodologii. W prezentowanym przykładzie ten sam proces modelowany dwiema metodami dał dwie prognozy znacznie różniące się dokładnością. Na zakończenie warto podkreślić, że udział zużycia energii w gospodarstwach domowych, w całkowitym zużyciu w kraju, wzrasta w ostatnich latach. Wzrost zużycia energii elektrycznej w kraju w ciągu 4 lat od 2003 do 2007 wyniósł 12%, czyli średnio 3% rocznie. Na najbliższe lata do roku 2010 ale także i dalsze, tempo może być jeszcze większe. W tegorocznym Raportcie [7] przedstawiono wyniki analizy wpływu pełnej realizacji polityki ograniczenia wzrostu emisji CO_2 , proponowanej w regulacjach UE na poziom cen energii dla polskiej gospodarki, w tym na skutki dla gospodarstw domowych. Ocenia się w nim, wzrost udziału kosztów za energię elektryczną w budżecie z 3,7% w 2005 roku do 6,2% w roku 2030. Nasuwa się wniosek, że uwzględnienie tej informacji może mieć istotne znaczenie w modelowaniu przyszłego zużycia energii elektrycznej w gospodarstwach domowych.

Dokładne prognozy cząstkowe, w tym zużycie energii w gospodarstwach domowych mogą być użyteczne w opracowywaniu prognoz globalnych dla kraju, których znaczenie jest coraz większe choćby w kontekście ostatnich przydziałów limitów emisji CO_2 dla Polski.

Literatura

- [1] CIEŚLAK M. i in., 2000 – Prognozowanie Gospodarcze. Metody i Zastosowania, PWN, Warszawa.
- [2] DAŚAL K., 2003 – Metoda doboru wejść w prognozowaniu krótkoterminowym obciążeń systemu elektroenergetycznego dla modelu rozkładu kanonicznego wektora losowego. Praca doktorska Politechnika Częstochowska.
- [3] GRABIŃSKI T., WYDYMUS S., ZELIAŚ A., 1982 – Metody doboru zmiennych w modelach ekonometrycznych. PWN, Warszawa.
- [4] HELLWIG Z., 1998 – Elementy rachunku prawdopodobieństwa i statystyki matematycznej. PWN, Warszawa.
- [5] MALKO J., 1995 – Wybrane zagadnienia prognozowania w elektroenergetyce. Oficyna Wydawnicza Politechniki Wrocławskiej, Wrocław.
- [6] DOBRZAŃSKA I., DAŚAL K., ŁYP J., POPŁAWSKI T., SOWIŃSKI J., 2002 – Prognozowanie w elektroenergetyce. Zagadnienia wybrane. Wydawnictwo Politechniki Częstochowskiej. Częstochowa.
- [7] Badania Systemowe „EnergSys” Sp. z o.o.: RAPORT 2030 Wpływ proponowanych regulacji unijnych w zakresie wprowadzenia europejskich strategii rozwoju energetyki wolnej od emisji CO₂ na bezpieczeństwo energetyczne Polski, a w szczególności możliwości odbudowy mocy wytwórczych wykorzystujących paliwa kopalne oraz poziom cen energii elektrycznej. SYNTEZA. Warszawa Wersja z 10.06.2008.
- [8] Ministerstwo Gospodarki.: Polityka Energetyczna Polski do 2030 Roku. Raport Projekt, wersja 3.2. Warszawa, wrzesień 2007.
- [9] Instytut Ochrony Środowiska.: Czwarty Raport Rządowy Dla Konferencji Stron Ramowej Konwencji Narodów Zjednoczonych w Sprawie Zmiana Klimatu. Warszawa 2006.

Kazimierz DAŚAL, Tomasz POPŁAWSKI

Problems related to forecasting of electric energy consumption in Poland

Abstract

Talking about forecasting of electric energy consumption we shall consider the issue in different categories. In terms of forecasting period, the short-term, medium-term and long-term forecasts can be distinguished. The forecasts may be classified in the terms of structure, voltage, tariff, area or in the nodes of the network for different voltage levels [2, 6]. Each of these specifications has its own peculiarities, requires different models, different input data, the accuracy assessment may be different as well as the aims of the short-term and long-term forecasts. The knowledge of long-term forecasts is necessary for reasonable use of resources in the electric power engineering system as well as for

planning of its development. In the paper the attention was focused on another important aspect of modeling, in particular modeling of electric energy consumption. The decision to be taken is: to model using the process data only, i.e. time series, or to develop an econometric model.

KEY WORDS: short- and long term forecasting in electric power engineering, time series, forecasting model

