Sylwester Smolik
Katedra Zastosowań Matematyki SGGW

Modelowanie statystyczne zjawisk z trendem i sezonowością

Wprowadzenie

Ekonometrycy analizując szeregi czasowe zjawisk ekonomicznych wyróżniają w nich następujące składowe: trend, wahania sezonowe, wahania koniunkturalne oraz wahania losowe. Wszystkie z wymienionych wahań nie muszą jednocześnie występować w opisywanym procesie. Przez trend rozumie się pewną wypośrodkowującą linię wyznaczającą zasadniczy kierunek rozwojowy badanego zjawiska. Nie wyjaśnia on, dlaczego badana wielkość ulegała zmianom, jednak posługiwanie się trendem do opisu przeszłości lub przewidywania przyszłości jest rozpowszechnione i celowe. Wahania sezonowe obrazują zmiany zjawiska w okresie rocznym. Mogą być stałe, tzn. systematycznie z roku na rok w odpowiadających okresach zakłócają trend w ten sam sposób. Wahania sezonowe mogą też być zmienne, tzn. w każdym następnym roku rośnie (lub maleje) ich nasilenie – zachowując charakter zakłóceń. Najczęściej objawia się to zjawisko rosnącą amplitudą zakłóceń wraz ze wzrostem wartości trendu. Sprawcą sezonowości zjawisk ekonomicznych są czynniki klimatyczno-przyrodnicze oraz powtarzające się systematycznie zdarzenia kalendarzowe (święta). Do najważniejszych z nich należą: czas trwania dnia, temperatura powietrza, opady, oraz zjawiska wtórne w stosunku do wymienionych (grubość pokrywy śnieżnej lub grubość pokrywy lodowej na rzekach). Wszystkie tzw. podstawowe czynniki sezonowości można wyrazić ilościowo za pomocą liczb i charakteryzują się one następującymi własnościami:

1° są funkcjami okresowymi czasu o okresie rocznym;

2° ich wartości, jak i kształt są zróżnicowane przede wszystkim w zależności od położenia geograficznego.

Podstawowe czynniki kształtujące sezonowość nie muszą oddziaływać na rozpatrywane zjawisko ekonomiczne w sposób bezpośredni, mogą występować przesunięcia czasowe i łańcuchy powiązań wieloogniwowych.

Wahania koniunkturalne nie muszą być ściśle okresowymi, trwają od kilku do kilkunastu lat. Zwykle wahania dotyczące rozwoju społeczno--gospodarczego krajów mają dłuższy okres, podczas gdy okresy dotyczące poszczególnych zjawisk są zwykle krótsze (np. cykl świński). Wahania losowe traktujemy jako wynik działania dużej liczby przyczyn ubocznych, mających charakter podobny do błędów przypadkowych, których analizą zajmuje się statystyka matematyczna. Zdarza się także, że w szeregu czasowym daje się wyodrębnić wahania katastrofalne. Występują one sporadycznie i powodują znacznie większe odchylenia od trendu w porównaniu z wahaniami losowymi. Są one spowodowane przez zdarzenia historyczne, takie jak wojna, kryzys, katastrofy żywiołowe, epidemie itp. Mając dany szereg czasowy, chcemy racjonalnie rozłożyć go na opisane wyżej składowe oraz poprawnie oszacować parametry opisującego go modelu.

Estymacja parametrów proponowanego modelu

Z wahaniami sezonowymi spotykamy się bardzo często w rolnictwie, które jest uzależnione od przebiegu pór roku. Zmianom tym podlega zapotrzebowanie na siłę roboczą, siłę pociągową, skup mleka, żywca, płodów rolnych i związany z nimi transport kolejowy towarów. Na potrzeby rolnictwa prowadzi się systematyczne obserwacje meteorologiczne. Ich efektem są publikowane średnie miesięczne temperatury powietrza i gruntu, opady, usłonecznienie itp. wybranych miejscowości. W powszechnie dostępnych periodykach są publikowane dane dotyczące miesięcznej produkcji najważniejszych towarów i różnych parametrów opisujących stan gospodarki. Opracujemy model opisujący tego typu zjawiska. Zakładamy, że wartości opisywanej zmiennej zależnej Y_{it} charakteryzują się składową i-tego roku i mają wyraźną sezonowość zależną od t--tego miesiąca, ponadto są obarczone błędem losowym. Mamy zatem:

$$Y_{it} = L(i) + M(t) + \varepsilon_{it} \qquad (i = 1,2,\dots,\ l;\ t = 1,2,\dots,\ 12). \tag{1}$$

Bez względu na postać funkcji L, zawsze $L(i) = L_i$ dla każdego i. Natomiast zakładamy, że sezonowość daje się opisać jedną harmoniką. Zauważmy, że w naszym przypadku pulsacja harmoniki $w = 2\pi/T = 2\pi/12 = \pi/6$, wtedy proponowany model przyjmuje postać:

$$y_{it} = B_i + A \cdot \sin\left(\frac{\pi}{6}t + \theta\right) + \varepsilon_{it}, \tag{2}$$

w którym A nazywamy amplitudą, a θ fazą początkową harmoniki i razem z B_i ($i = 1, 2,..., l$) stanowią szacowane parametry modelu (2), na podstawie wyników obserwacji zebranych w macierz liczb o l wierszach i $t = 12$ kolumnach. Zakładamy a priori, że ε_{it} są niezależnymi zmiennymi losowymi o rozkładzie normalnym z wartością oczekiwaną zero i wariancją σ^2 dla wszystkich i, t. Oszacujemy parametry modelu (2) metodą najmniejszych kwadratów (MNK). Mamy więc:

$$S(B_i, \theta, A) = \sum_{i=1}^{l} \sum_{t=1}^{12} \varepsilon_{it}^2 = \sum_{i=1}^{l} \sum_{t=1}^{12} \left[y_{it} - B_i - A \sin\left(\frac{\pi}{6} t + \theta\right)\right]^2 = \min. \tag{3}$$

Korzystamy z warunku koniecznego ekstremum funkcji $S(B_i, \theta, A)$:

$$\frac{\partial S}{\partial B_i} = 2 \cdot \sum_{t=1}^{12} \left[y_{it} - B_i - A \sin\left(\frac{\pi}{6} t + \theta\right)\right] \cdot (-1) = 0, \qquad \text{dla } i = 1, 2, ..., l;$$

$$\frac{\partial S}{\partial \theta} = 2 \cdot \sum_{i=1}^{l} \sum_{t=1}^{12} \left[y_{it} - B_i - A \sin\left(\frac{\pi}{6} t + \theta\right)\right] \cdot \left[-A \cos\left(\frac{\pi}{6} t + \theta\right)\right] = 0; \tag{4}$$

$$\frac{\partial S}{\partial A} = 2 \cdot \sum_{i=1}^{l} \sum_{t=1}^{12} \left[y_{it} - B_i - A \sin\left(\frac{\pi}{6} t + \theta\right)\right] \cdot [-\sin\left(\frac{\pi}{6} t + \theta\right) = 0.$$

Po prostych przekształceniach otrzymujemy układ równań normalnych dla modelu (2):

$$\begin{cases} \sum_{t=1}^{12} \hat{B}_i + \hat{A} \cdot \sum_{t=1}^{12} \sin\left(\frac{\pi}{6} t + \hat{\theta}\right) = \sum_{t=1}^{12} y_{it}, \qquad dla \;\; i = 1, 2, ..., l; \\ \sum_{i=1}^{l} \sum_{t=1}^{12} \hat{B}_i \cos\left(\frac{\pi}{6} t + \hat{\theta}\right) + \frac{\hat{A}}{2} \cdot \sum_{i=1}^{l} \sum_{t=1}^{12} \sin\left(\frac{\pi}{3} t + 2\hat{\theta}\right) = \sum_{i=1}^{l} \sum_{t=1}^{12} y_{it} \cos\left(\frac{\pi}{6} t + \hat{\theta}\right); \\ \sum_{i=1}^{l} \sum_{t=1}^{12} \hat{B}_i \sin\left(\frac{\pi}{6} t + \hat{\theta}\right) + \hat{A} \cdot \sum_{i=1}^{l} \sum_{t=1}^{12} \sin^2\left(\frac{\pi}{6} t + \hat{\theta}\right) = \sum_{i=1}^{l} \sum_{t=1}^{12} y_{it} \sin\left(\frac{\pi}{6} t + \hat{\theta}\right). \end{cases} \tag{5}$$

Korzystając z definicji sumy i zmieniając kolejność sumowania, w rezultacie otrzymujemy:

$$\begin{cases} \hat{B}_i \cdot 12 + \hat{A} \cdot \sum\limits_{t=1}^{12} \sin\left(\frac{\pi}{6}t + \hat{\theta}\right) = \sum\limits_{t=1}^{12} y_{it}, \quad dla \quad i = 1, 2, \ldots, l; \\ \sum\limits_{i=1}^{l} \hat{B}_i \cdot \sum\limits_{t=1}^{12} \cos\left(\frac{\pi}{6}t + \hat{\theta}\right) + \hat{A} \cdot \frac{l}{2} \cdot \sum\limits_{t=1}^{12} \sin\left(\frac{\pi}{3}t + 2\hat{\theta}\right) = \sum\limits_{i=1}^{l} \sum\limits_{t=1}^{12} y_{it} \cos\left(\frac{\pi}{6}t + \hat{\theta}\right); \\ \sum\limits_{i=1}^{l} \hat{B}_i \cdot \sum\limits_{t=1}^{12} \sin\left(\frac{\pi}{6}t + \hat{\theta}\right) + \hat{A} \cdot l \cdot \sum\limits_{t=1}^{12} \sin^2\left(\frac{\pi}{6}t + \hat{\theta}\right) = \sum\limits_{i=1}^{l} \sum\limits_{t=1}^{12} y_{it} \sin\left(\frac{\pi}{6}t + \hat{\theta}\right). \end{cases} \tag{6}$$

W pracy [5] udowodniono, że

$$\sum_{t=1}^{12} \sin\left(\frac{\pi}{6}t + \hat{\theta}\right) = \sum_{t=1}^{12} \cos\left(\frac{\pi}{6}t + \hat{\theta}\right) = \sum_{t=1}^{12} \sin\left(\frac{\pi}{3}t + 2\hat{\theta}\right) = 0, \text{ natomiast}$$

$$\sum_{t=1}^{12} \sin^2\left(\frac{\pi}{6}t + \hat{\theta}\right) = 6. \tag{7}$$

Podstawiając związki (7) do (6), po prostych przekształceniach ostatecznie otrzymujemy:

$$\begin{cases} \hat{B}_i + \frac{\hat{A}}{12} \cdot 0 = \frac{1}{12} \cdot \sum\limits_{t=1}^{12} y_{it}, \quad dla \quad i = 1, 2, \ldots, l; \\ \frac{1}{l} \cdot \sum\limits_{i=1}^{l} \hat{B}_i \cdot 0 + \frac{\hat{A}}{2} \cdot 0 = \frac{1}{l} \cdot \sum\limits_{i=1}^{l} \sum\limits_{t=1}^{12} y_{it} \cos\left(\frac{\pi}{6}t + \hat{\theta}\right); \\ \frac{1}{l} \cdot \sum\limits_{i=1}^{l} \hat{B}_i \cdot 0 + \hat{A} \cdot 6 = \frac{1}{l} \cdot \sum\limits_{i=1}^{l} \sum\limits_{t=1}^{12} y_{it} \sin\left(\frac{\pi}{6}t + \hat{\theta}\right). \end{cases} \tag{8}$$

Z układu równań (8) obliczamy wartości ocen parametrów modelu (2):

$$\begin{cases} \hat{B}_i = \bar{y}_i, \quad dla \quad i = 1, 2, \ldots, l; \\ \hat{\theta} = \operatorname{arctg}\left[\sum\limits_{t=1}^{12} \bar{y}_{\bullet t} \cos\frac{\pi}{6}t \Big/ \sum\limits_{t=1}^{12} \bar{y}_{\bullet t} \sin\frac{\pi}{6}t\right]; \\ \hat{A} = \frac{1}{6}\left[\cos\hat{\theta} \cdot \sum\limits_{t=1}^{12} \bar{y}_{\bullet t} \sin\frac{\pi}{6}t + \sin\hat{\theta} \cdot \sum\limits_{t=1}^{12} \bar{y}_{\bullet t} \cos\frac{\pi}{6}t\right]; \end{cases} \tag{9}$$

w którym $\bar{y}_{i\bullet}$ jest średnią miesięczną i-tego roku, natomiast $\bar{y}_{\bullet t}$ jest średnią t--tego miesiąca z l lat $(t = 1, 2, \ldots, 12)$.

Łatwo zauważyć, że w modelu (2) trend wyraża się funkcją schodkową podaną w równaniu (9).

W przypadku, gdy lata mają jednakowy wpływ na opisywaną zmienną zależną, tzn. w modelu (2) $B_i = s$, wtedy jego oszacowanie z (9) wynosi:

$$\hat{B}_i = \hat{s} = \bar{y}_{\bullet\bullet} = \frac{1}{12l} \sum_{i=1}^{l} \sum_{t=1}^{12} y_{it} \tag{10}$$

i nazywa się średnią globalną z danych empirycznych. Inaczej mówiąc, jest to średnia miesięczna, obliczona ze wszystkich lat, z których korzystaliśmy przy opracowywaniu modelu.

Model produkcji miesięcznej energii elektrycznej w Polsce

Racjonalna i skuteczna polityka energetyczna stanowi jedną z podstawowych przesłanek rozwoju społecznego i wzrostu ekonomicznego. Dążenie do oszczędnego gospodarowania energią jest obecnie oczywistą koniecznością, uzasadnioną pogłębiającym się wyczerpywaniem naturalnych zasobów energetycznych Ziemi i potrzebą zmniejszenia zanieczyszczenia środowiska naturalnego. Z wydobyciem i użytkowaniem każdego z surowców energetycznych wiąże się emisja niepożądanych substancji, które przyczyniają się do pogłębienia efektu cieplarnianego, zagrażającego zmianami klimatu. Najbardziej pożądana jest energia elektryczna. Wynika to z korzyści ekonomicznych, jakie daje jej stosowanie, oraz z komfortu, który zapewnia jej odbiorcom. Z tej przyczyny produkcja energii elektrycznej jest jednym z najszybciej rosnących elementów gospodarki rozwijających się państw. Stanowi ona zarazem jeden z najbardziej istotnych wskaźników obrazujących stopień ich rozwoju gospodarczego. Od niej też w głównej mierze zależy rozwój przemysłu, transportu, a nawet rolnictwa. Powszechnie uznanym wskaźnikiem zagospodarowania kraju jest zużycie energii elektrycznej na jednego mieszkańca. Natomiast o poziomie życia jego mieszkańców między innymi świadczy zużycie energii elektrycznej przez gospodarstwa domowe. Opiszemy i zbadamy kształtowanie się produkcji energii elektrycznej w Polsce, na podstawie danych miesięcznych z lat 1974–1993, podanych w tabeli 1 i zinterpretowanych graficznie na rysunku 1.

Aby oszacować parametry harmoniki, przytoczymy dane wyjściowe z tabeli 1 uwzględniające wyżej wymienione dwudziestolecie.

t	1	2	3	4	5	6	7	8	9	10	11	12
$\bar{y}_{\bullet t}$	12,07	11,12	11,48	10,18	9,45	8,83	8,83	9,04	9,46	10,93	11,53	12,23

Wykonujemy obliczenia pomocnicze:

$$\sum_{t=1}^{12} \bar{y}_{\bullet t} \cos\frac{\pi}{6}t = 9{,}422\,; \quad \sum_{t=1}^{12} \bar{y}_{\bullet t} \sin\frac{\pi}{6}t = 3{,}752\,.$$

Wykorzystując wyprowadzone wzory (9), po zaokrągleniach ostatecznie otrzymujemy poszukiwany model:

$$\hat{y}_{it} = \bar{y}_{i\bullet} + 1{,}69\sin\left(\frac{\pi}{6}t + 1{,}192\right) \quad \text{w TW} \cdot \text{h}, \quad (t = 1, 2, \dots, 12)\,, \qquad (11)$$

w którym $\bar{y}_{i\bullet}$ dla każdego $i = 1, 2, \dots, 20$ należy odczytać z tabeli 1.
Jego dopasowanie do danych empirycznych kształtuje się następująco: współczynnik zbieżności

$$\varphi^2 = \sum_{i=1}^{20} \sum_{t=1}^{12} (y_{it} - \hat{y}_{it})^2 \Big/ \sum_{i=1}^{20} \sum_{t=1}^{12} (y_{it} - \bar{y}_{\bullet\bullet})^2 = 66{,}9575/832{,}8316 = 0{,}0804\,;$$

współczynnik determinacji $R^2 = 1 - \varphi^2 = 0{,}9196$ – wnioskujemy z niego, że oszacowany model (11) tłumaczy około 92% zmienności badanej zmiennej zależnej; odchylenie standardowe reszt

$$s_e = \sqrt{\sum_{i=1}^{20} \sum_{t=1}^{12} \varepsilon_{it}^2 / (n-k)} = \sqrt{66{,}9575/(240-2)} = \; = 0{,}56 \text{ TW} \cdot \text{h}\,;$$ współczynnik zmienności resztowej $v = 100\, s_e / \bar{y}_{\bullet\bullet} = 56/10{,}43 = 5{,}4\%$. Na rysunku 1 cienką linią zaznaczono wykres wyprowadzonej zależności (11). Używając tak prostych środków do opisu dominującej w tym zjawisku sezonowości, uzyskaliśmy dobrą zgodność modelu z danymi empirycznymi. Tym samym cel nasz został osiągnięty.

Podsumowanie

W zagadnieniach ekonomicznych przyszłość nie zawsze jest w pełni zdeterminowana przez stan aktualny i charakter podejmowanych działań. Z tego powodu wnioskowanie o przyszłym kształtowaniu się takich zjawisk nie jest niezawodne. Sprawą nauki jest opracowanie takich metod wnioskowania, aby obszar tej zawodności ograniczyć do minimum. Wieloletnie obserwacje różnych zjawisk upewniają nas, że przyszły stan większości zjawisk ekonomicznych i przyrodniczych zależy stochastycznie od ich stanu teraźniejszego oraz od przeszłego ich przebiegu w czasie. Stwarza to możliwość wykorzystania metod statystycznych do opisu i mierzenia zaobserwowanych prawidłowości rozwoju. Metody te nazywamy ekonometrycznym modelowaniem pewnego wycinka sfery zjawisk ekonomicznych. Ten sam problem można jednak rozwiązywać w różny sposób. Przyjęło się już, że budowanie modelu ekonometrycznego jest połączeniem nauki ze sztuką, podobnie jak w pracy architekta – ograniczonego wytycznymi ogólnymi zleceniodawcy. Propozycje rozwiązań mogą być różne, ale muszą bazować na metodach naukowych i mieć walory praktycznej użyteczności. W jednym artykule nie można rozwiązać wszystkich problemów związanych z tak trudnym zagadnieniem jak jednoczesna estymacja trendu i sezonowości. Dlatego problem ten rozwiązujemy etapami i przytaczamy praktyczne zastosowanie prezentowanych wzorów.

Literatura

BOX G.E.P., JENKINS G.M., 1983: *Analiza szeregów czasowych. Prognoza i sterowanie*. PWN, Warszawa. [1]

CHOW G.C., 1995: *Ekonometria*. PWN, Warszawa. [2]

Metody prognozowania. Zbiór zadań. 2000, red. B. RADZIKOWSKA, Wydawnictwo AE we Wrocławiu, Wrocław. [3]

Prognozowanie gospodarcze. Metody i zastosowania. 1997. Red. naukowa M. CIEŚLAK, PWN, Warszawa. [4]

SMOLIK S., 1995: *Uproszczona procedura estymacji modelu wahań okresowych*. „Przegląd Statystyczny", R. XLII, z. 3–4, s. 449–457. [5]

SMOLIK S., 1998: *Oszczędne modele dla okresowych szeregów czasowych*. Przestrzenno-czasowe modelowanie i prognozowanie zjawisk gospodar-

czych. Red. A. ZELIAŚ, Akademia Ekonomiczna w Krakowie, s. 177–188. [6]
SMOLIK S., 1997: *Sezonowość w opisie procesów rolniczych.* „Wiadomości Statystyczne", nr 4, s. 10–14. [7]
STECZKOWSKI J., ZELIAŚ A., 1982: *Analiza wariancyjna i kowariancyjna w badaniach ekonomicznych.* PWN, Warszawa. [8]
ZELIAŚ A., 1997: *Teoria prognozy.* Wyd. 3 zmienione, PWE, Warszawa. [9]
ZIELIŃSKI Z., 1969: *Ekonometryczne metody analizy wahań sezonowych.* Zeszyty Naukowe Politechniki Szczecińskiej, nr 112 (rozprawa habilitacyjna). [10]

Statistic modelling of trend-based and season-related phenomena

Abstract

In quite a considerable number of economic and natural issues, during research, observation series are obtained with a very distinct seasonal dependence and trend, whether constant or vaguely outlined. Chance variations (randomness) of this type is encountered in climatology, when it comes to describing average monthly temperatures of air and the ground, and in temporal series describing production of, not infrequently elementary, commodities. For description of these phenomena, the following model has been proposed:

$$y_{it} = B_i + A\sin(\pi t/6 + \theta) + \varepsilon_{it}, \text{ for } i = 1, 2, \ldots, l \text{ years; } t = 1, 2, \ldots, 12 \text{ months.}$$

It has been proved that evaluation of its parameters with the use of the least square (minimum chi-square) method, based on the empirical data table with l lines and 12 columns, assumes the following form:

$\hat{B}_i = \overline{y}_i\bullet$, for each $i = 1, 2, \ldots, l$; whereas $\overline{y}_i\bullet$ is the monthly average of an i-th year,

$$\hat{\theta} = \operatorname{arctg}\left[\sum_{t=1}^{12} \overline{y}_{\bullet t} \cos\frac{\pi}{6}t \Big/ \sum_{t=1}^{12} \overline{y}_{\bullet t} \sin\frac{\pi}{6}t\right],$$

$$\hat{A} = \frac{1}{6}\left[\cos\hat{\theta} \cdot \sum_{t=1}^{12} \overline{y}_{\bullet t} \sin\frac{\pi}{6}t + \sin\hat{\theta} \cdot \sum_{t=1}^{12} \overline{y}_{\bullet t} \cos\frac{\pi}{6}t\right],$$

whereas $\overline{y}_{\bullet t}$ is the average of a t-th month $(t = 1, 2, \ldots, 12)$, calculated basing on l years.

The thus proposed model has been used in a description of power (electric energy) production in Poland, employing monthly empirical data from the period of 1974–1993.