FOLIA POMERANAE UNIVERSITATIS TECHNOLOGIAE STETINENSIS
Folia Pomer. Univ. Technol. Stetin. 2011, Oeconomica 285 (62), 97–104

Maria Szmuksta-Zawadzka, Jan Zawadzki [1]

MODELE HIERARCHICZNE DLA DANYCH SEZONOWYCH W ODNIESIENIU DO ANALIZY WARIANCJI

HIERARCHICAL MODELS FOR SEASONAL DATA IN THE LIGHT OF VARIANCE ANALYSIS

Studium Matematyki, Zachodniopomorski Uniwersytet Technologiczny w Szczecinie
al. Piastów 48, 71-371 Szczecin, e-mail: mszmuksta@zut.edu.pl
[1] Katedra Zastosowań Matematyki w Ekonomii, Zachodniopomorski Uniwersytet Technologiczny w Szczecinie
ul. Klemensa Janickiego 31, 71-270 Szczecin, e-mail: Jan.Zawadzki@zut.edu.pl

Summary. The work demonstrates that the parameter estimation of a regular two-stage hierarchical model of a time series for the seasonal data deprived of a linear trend are the same as the estimations obtained on the basis of a two-factor hierarchical model of a variance analysis, in which each of the factors of a higher order has corresponding lower-order factors of the same value. Theoretical considerations are illustrated by an empirical example.

Słowa kluczowe: analiza szeregów czasowych, analiza wariancji, modele hierarchiczne.
Key words: hierarchical models, time series analysis, variance analysis.

METODA

Steczkowski i Zeliaś (1982) przedstawiają model hierarchicznej analizy wariancji:

$$Y_{ijr} = \mu + \alpha_i + \beta_{ij} + u_{ijr} \tag{1}$$

$$(i = 1,2,..., k\ ; \quad j = 1,2,..., p_i \quad ; \ r = 1,2,..., m)$$

$$\sum_{i=1}^{k} \alpha_i = 0\,, \qquad \sum_{j=1}^{p_i} \beta_{ij} = 0$$

gdzie:

- μ – średnia ogólna,
- α_i – wpływ i-tego poziomu czynnika *A*,
- β_{ij} – wpływ poziomów czynnika *B* w obrębie i-tego poziomu czynnika *A*,
- u_{ijr} – niezależne zmienne losowe o rozkładzie normalnym $N(0,\sigma^2)$ dla wszystkich i, j, r.

Następnie ww. autorzy wyprowadzają wzory na estymatory MNK parametrów $\mu, \alpha_i, \beta_{ij}$:

$$\hat{\mu} = \overline{y}_{...} \tag{2}$$

$$\hat{\alpha}_i = \overline{y}_{i..} - \overline{y}_{...} \tag{3}$$

$$\hat{\beta}_{ij} = \overline{y}_{ij.} - \overline{y}_{i..} \tag{4}$$

Występujące po prawej stronie wzorów (2)–(4) wartości średnie wyznacza się ze wzorów:

$$\hat{\mu} = \bar{y}_{...} = \frac{1}{n} \sum_{i=1}^{k} \sum_{j=1}^{p_i} \sum_{r=1}^{m} y_{ijr} \tag{5}$$

$$\bar{y}_{i..} = \frac{1}{p_i m} \sum_{j=1}^{p_i} \sum_{r=1}^{m} y_{ijr} \tag{6}$$

$$\bar{y}_{ij.} = \frac{1}{m} \sum_{r=1}^{m} y_{ijr} \tag{7}$$

Swoje wywody autorzy ilustrują przykładem dotyczącym kształtowania się skupu mleka w województwie krakowskim według miesięcy, weryfikując hipotezy o występowaniu trendu i wahań sezonowych w szeregu czasowym.

W rozważaniach przedstawionych w niniejszej pracy zakładamy, że każdy z czynników poziomu *B* przyjmuje te same wartości (oznaczające na przykład numery miesięcy w każdym z trzech okresów czteromiesięcznych), co oznacza, że $p_1 = p_2 = \ldots = p_k = p$ oraz że ogólna wartość średnia jest równa zeru $(\mu = 0)$. Wówczas model (1) przyjmie postać:

$$Y_{ijr} = \alpha_i + \beta_{(i)j} + u_{ijr} \quad (j = 1, 2, ..., p) \tag{8}$$

Tym samym wzory na estymatory parametrów powyższego równania ulegną uproszczeniu:

$$\hat{\alpha}_i = \bar{y}_{i..} \qquad i = 1,2 \ldots k \tag{9}$$

$$\hat{\beta}_{(i)j} = \hat{\beta}_j = \frac{1}{p} \sum_{j=1}^{p} \hat{\beta}_{ij} \quad (j = 1, 2, ..., p) \tag{10}$$

Celem pracy jest wykazanie, że parametry dwuczynnikowego hierarchicznego modelu analizy wariancji dla danych, z których wyeliminowano trend liniowy, są takie same jak w odpowiadającym mu regularnym dwustopniowym modelu hierarchicznym szeregu czasowego. Regularność oznacza, że każdemu z czynników wyższego stopnia hierarchii są przyporządkowane takie same wartości czynnika należącego do stopnia niższego.

Liczba regularnych modeli hierarchicznych równa jest permutacji i permutacji z powtórzeniami podzielników p_i długości cyklu wahań okresowych m (Szmuksta-Zawadzka i Zawadzki 2000).

Podzielniki p_i spełniają jednocześnie dwa warunki:

$$\begin{gathered} 2 \le p_i \le \frac{m}{2} \\ \prod_{i=1}^{r} p_i = m \end{gathered} \tag{11}$$

W przypadku dwunastomiesięcznego cyklu wahań $(m = 12)$ otrzymamy 7 modeli, w tym 4 modele dwustopniowe oraz 3 modele trójstopniowe.

Zapis analityczny dowolnego dwustopniowego modelu hierarchicznego dla danych niezawierających trendu jest następujący:

$$Y^*_{srt} = \sum_{s=1}^{p_1} b_{0s} Q_{st} + \sum_{r=1}^{p_2} b_{0sr} Q_{(s)rt} + U_{srt} \quad (t = 1, 2, ..., n) \tag{13}$$

gdzie:

Y^*_{srt} – wartości zmiennej endogenicznej niezawierające trendu;

Q_{st} – zmienna zero-jedynkowa przyjmująca wartość jeden dla s-tego poziomu pierwszego czynnika (stopnia hierarchii), np. dla miesięcy wchodzących w skład danego okresu czteromiesięcznego, i zero dla pozostałych;

Q_{srt} – zmienna zero-jedynkowa przyjmująca wartość jeden dla r-tego poziomu drugiego czynnika, np. miesiąca w kwartale, i zero dla pozostałych;

p_1 i p_2 – podzielniki długości cyklu wahań sezonowych m spełniające warunki dane wzorem (11).

Parametry b_{0s} oraz $b_{0(s)r}$ spełniają warunki sumowalności do zera:

$$\sum_{s=1}^{p_1} b_{0s} = \sum_{r=1}^{p_2} b_{0(s)r} = 0 \tag{14}$$

WYNIKI I DYSKUSJA

Przedstawimy wyniki obliczeń związane z oszacowaniami parametrów α_i oraz $\beta_{(i)j}$ dla przykładowego dwuczynnikowego hierarchicznego modelu analizy wariancji dla odchyleń od trendu liniowego. Oceny parametrów trendu zostały wzięte z oszacowanego wcześniej MNK modelu szeregu czasowego z liniowym trendem i periodycznym składnikiem sezonowym:

$$Y_t = \alpha_1 t + \alpha_0 + \sum_{k=1}^{12} d_{0k} Q_{kt} + U_t \tag{15}$$

przy czym: $\sum_{k=1}^{12} d_{ok} = 0$

Pierwszy stopień hierarchii obejmować będzie 3 okresy czteromiesięczne. Drugi stopień hierarchii obejmować będzie odpowiednio: pierwsze, drugie, trzecie i czwarte miesiące wchodzące w skład każdego z tych okresów. Oznacza to, że powyższemu modelowi analizy wariancji odpowiadać będzie hierarchiczny model szeregu czasowego **H34***, przy czym gwiazdka (*) oznacza, że odnosi się on do danych sezonowych, z których wyeliminowano trend.

Badaną zmienną będzie produkcja energii elektrycznej w Polsce w latach 2003–2006 według miesięcy.

Przed przystąpieniem do obliczeń został wyeliminowany, w sposób podany wyżej, trend z oryginalnego szeregu wyrażający się równaniem:

$$\hat{f}(t) = 32{,}767t + 1161{,}71$$

W obliczeniach związanych z oszacowaniami parametrów modelu (8) zostaną wykorzystane dane liczbowe zawarte w tab. 1. W kolumnach od 2 do 13 podano dla poszczególnych miesięcy wartości szeregu, z którego wyeliminowano trend liniowy. Natomiast w wierszach od trzeciego do szóstego zawarto odchylenia odpowiadające kolejnym latom. Następne dwa wiersze zawierają odpowiednio sumy kolumn oraz średnie $\overline{Y}_{ij.}$ wyznaczone dla każdego miesiąca ze wzoru (7).

Tabela 1. Tablica robocza do wyznaczania ocen parametrów modelu hierarchicznej analizy wariancji

Nr roku	Miesiąc											
	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
1	2208,523	117,755	360,988	–772,780	–1713,547	–1922,314	–1668,082	–1641,849	–753,617	1174,616	1049,849	2469,081
2	2372,314	898,546	1092,779	–51,988	–1347,756	–2029,523	–998,291	–1204,058	–864,825	1268,407	1016,640	2425,872
3	2330,105	966,338	1412,570	–311,197	–1191,965	–1599,732	–1366,499	–1194,267	–1038,034	178,198	268,431	864,664
4	528,896	673,129	1544,361	–611,406	–1132,173	–1785,941	–1534,708	–1652,476	–1032,243	339,990	870,222	1786,455
5	4404,687	565,920	1554,153	–365,615	–1298,382	–1461,150	–1424,917	–1374,684	–849,452	176,781	554,013	719,246
$\sum$	11 844,525	3221,688	5964,851	–2112,986	–6683,823	–8798,660	–6992,497	–7067,334	–4538,171	3137,992	3759,155	8265,318
$\overline{y}_{ij.}$	2368,905	644,338	1192,970	–422,597	–1336,765	–1759,732	–1398,499	–1413,467	–907,634	627,598	751,831	1653,064
$\overline{y}_{i..}$		945,9039				–1477,1157				531,2147		
$\hat{\beta}_{ij}$	1423,0011	–301,566	247,066	–1368,501	140,351	–282,616	78,616	63,649	–1438,849	96,384	220,616	1121,849
$\hat{\beta}_{(i)j}$	41,501	–162,600	182,100									

Wartości średnie, odpowiadające kolejnym okresom czteromiesięcznym, zostały wykorzystane do obliczenia średnich $\overline{Y}_{i.}$. Jak wynika ze wzoru (9), odpowiadają one estymatorom parametrów α_i zmodyfikowanego dwustopniowego modelu analizy wariancji danego wzorem (8).

Oceny parametrów α_i są następujące:

$$\hat{\alpha}_i = \overline{y}_{i..} \quad i = 1,2,3,4 \tag{15}$$

$$\hat{\alpha}_1 = \overline{y}_{1..} = \frac{1}{3}(2368{,}905 + 664{,}338 + 192{,}970 - 422{,}597) = = \frac{1}{4}3783{,}616 = 945{,}904 \tag{16}$$

$$\hat{\alpha}_2 = \overline{y}_{2..} = \frac{1}{4}(-1336{,}765 - 1759{,}732 - 1398{,}499 - 1413{,}467) = = \frac{1}{4}(-5908{,}463) = -1477{,}116 \tag{17}$$

$$\hat{\alpha}_3 = \overline{y}_{3..} = \frac{1}{4}(-907{,}634 + 627{,}598 + 751{,}831 + 1653{,}064) = = \frac{1}{4}2124{,}859 = 531{,}215 \tag{18}$$

Zamieszczone one zostały w 8 wierszu wspomnianej tabeli.

Znajomość ocen parametrów α_i pozwala na wyznaczenie ocen parametrów β_{ij}:

$$\hat{\beta}_{ij} = \overline{y}_{1j.} - \overline{y}_{1..}(i = 1, 2, 3, 4; j = 1, 2, 3 \tag{19}$$

Podstawiając kolejne wartości i, otrzymujemy:

$$\hat{\beta}_{1j} = \overline{y}_{1j.} - \overline{y}_{1..} \tag{20}$$

$$\hat{\beta}_{2j} = \overline{y}_{2j.} - \overline{y}_{2..} \quad (j = 1, 2, 3) \tag{21}$$

$$\hat{\beta}_{3j} = \overline{y}_{3j.} - \overline{y}_{3..} \tag{22}$$

Szczegółowo przebieg obliczeń, ze względu na ich powtarzalność, pokażemy jedynie dla $i = 1$:

$$\hat{\beta}_{11} = \overline{y}_{11.} - \overline{y}_{1..} = 2368{,}905 - 945{,}904 = 1423{,}001 \tag{23}$$

$$\hat{\beta}_{12} = \overline{y}_{12.} - \overline{y}_{1..} = 644{,}338 - 945{,}904 = -301{,}566 \tag{24}$$

$$\hat{\beta}_{13} = \overline{y}_{13.} - \overline{y}_{1..} = 1192{,}970 - 945{,}904 = 247{,}066 \tag{25}$$

$$\hat{\beta}_{14} = \overline{y}_{14.} - \overline{y}_{1..} = -422{,}597 - 945{,}904 = -1368{,}501 \tag{26}$$

Oceny te zostały zestawione w kolumnach od 2 do 5 w przedostatnim wierszu tab. 1. Wyniki obliczeń dla dwóch pozostałych okresów czteromiesięcznych zawierają kolumny od 6 do 13.

Oceny te posłużą do wyznaczenia estymatorów parametrów $\beta_j = \beta_{(i)j}$:

$$\hat{\beta}_1 = \hat{\beta}_{(i)1} = \frac{1}{3}\left(\hat{\beta}_{11} + \hat{\beta}_{21} + \hat{\beta}_{31}\right) = \frac{1}{3}\left(1423{,}00 + 140{,}3511 - 1438{,}849\right) = \frac{1}{3}124{,}503 = 41{,}501 \tag{27}$$

$$\hat{\beta}_2 = \hat{\beta}_{(i)2} = \frac{1}{3}\left(\hat{\beta}_{12} + \hat{\beta}_{22} + \hat{\beta}_{32}\right) = \frac{1}{3}\left(-301{,}5663 - 282{,}6163 + 96{,}3837\right) = \frac{1}{3}\left(-487{,}7989\right) = 162{,}600 \tag{28}$$

$$\hat{\beta}_3 = \hat{\beta}_{(i)3} = \frac{1}{3}\left(\hat{\beta}_{13} + \hat{\beta}_{23} + \hat{\beta}_{33}\right) = \frac{1}{3}\left(247{,}0663 + 78{,}6163 + 220{,}6163\right) = \frac{1}{3}\left(546{,}2989\right) = 182{,}010 \tag{29}$$

$$\hat{\beta}_4 = \hat{\beta}_{(i)4} = \frac{1}{3}\left(\hat{\beta}_{14} + \hat{\beta}_{24} + \hat{\beta}_{34}\right) = \frac{1}{3}\left(-1368{,}5011 + 63{,}6489 + 1121{,}8489\right) = \frac{1}{3}\left(-183{,}0033\right) = 61{,}001 \tag{30}$$

Przed wynikami oszacowań parametrów regularnego modelu hierarchicznego szeregu czasowego **H34*** przedstawiono postać macierzy pozwalającą uzyskać estymatory parametrów b_{0s} oraz $b_{0(s)r}$.

Macierz **H34*** uwzględniająca sumowalność do zera parametrów pierwszego i drugiego stopnia hierarchii:

$$\mathbf{H34}^* = \begin{array}{c} \begin{array}{ccccc} \mathrm{CZ1} & \mathrm{CZ2} & \mathrm{MCZ1} & \mathrm{MCZ2} & \mathrm{MCZ3} \end{array} \\ \begin{bmatrix} 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & -1 & -1 & -1 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & -1 & -1 & -1 \\ -1 & -1 & 1 & 0 & 0 \\ -1 & -1 & 0 & 1 & 0 \\ -1 & -1 & 0 & 0 & 1 \\ -1 & -1 & -1 & -1 & -1 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ . & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \end{bmatrix} \end{array} \tag{31}$$

W drugiej kolumnie tab. 2 zamieszczone zostały oceny parametrów oraz tworzących podmacierze CZ i MCZ, przy czym oceny $\hat{b}_{03}$ oraz $\hat{b}_{0(s)4}$ wyznaczone zostały z warunków sumowalności do zera. W kolejnych wierszach podano oceny parametrów struktury stochastycznej – współczynnika determinacji $-R^2$ oraz odchylenia standardowego składnika losowego. W kolumnach trzeciej i czwartej zestawione zostały odpowiednio oceny i oznaczenia estymatorów parametrów α_i i $\beta_{(i)j}$ modelu hierarchicznego analizy wariancji.

Tabela 2. Oceny parametrów modeli hierarchicznego i analizy wariancji

Zmienna	Oceny parametrów dla		Estymator parametru
	modelu hierarchicznego	modelu analizy wariancji	
Wyraz wolny	0,000	0,000	$\hat{\mu}$
CZ1	945,900	945,900	$\hat{\alpha}_1$
CZ2	–1477,120	–1477,120	$\hat{\alpha}_2$
CZ3	531,220	531,220	$\hat{\alpha}_3$
MCZ1	41,501	41,501	$\hat{\beta}_1 = \hat{\beta}_{(s)1}$
MCZ2	–162,600	–162,600	$\hat{\beta}_2 = \hat{\beta}_{(s)2}$
MCZ3	182,100	182,100	$\hat{\beta}_3 = \hat{\beta}_{(s)3}$
MCZ4	–61,000	–61,000	$\hat{\beta}_4 = \hat{\beta}_{(s)4}$

Porównując oceny parametrów otrzymane dla modelu hierarchicznego **H34*** i hierarchicznej analizy wariancji, należy stwierdzić, że są one takie same. Oznacza to, że została wykazana prawdziwość sformułowanej hipotezy.

Wykonanie obliczeń tego rodzaju dla pozostałych dwustopniowych modeli hierarchicznych nie nastręcza trudności. Rozszerzenie na trzystopniowe modele analizy wariancji nie jest konieczne, ponieważ wystarczy oszacowanie parametrów trójstopniowych hierarchicznych modeli szeregu czasowego.

Niżej przedstawiono oszacowanie parametrów modelu hierarchicznego **H34** dla danych oryginalnych (bez eliminacji trendu) – ich oszacowania zestawione zostały w tab. 3.

Tabela 3. Oceny parametrów modeli hierarchicznych

Zmienna	Oceny parametrów dla danych	
	oryginalnych	oczyszczonych
Wyraz wolny	11616,71	0,00
CZ1	945,90	945,90
CZ2	–1477,12	–1477,12
CZ3	531,22	531,22
MCZ1	41,501	41,501
MCZ2	–162,600	–162,600
MCZ3	182,100	182,100
MCZ4	–61,00	–61,00
t	32,767	–
R^2	0,6239	0,5698
S	985,15	975,99

Z informacji zawartych w tab. 3 wynika, że oceny parametrów, występujących przy zmiennych tworzących podmacierze CZ oraz MCZ, są identyczne.

Zróżnicowanie w ocenach współczynników determinacji wynika z różnic w mianownikach. W pierwszym przypadku jest to suma kwadratów od średniej. Natomiast w drugim przypadku jest to suma kwadratów zmiennej po eliminacji trendu. W przypadku ocen odchyleń standardowych składników losowych są one pochodną różnic w liczbie stopni swobody.

PODSUMOWANIE

1. Oceny parametrów dwustopniowego modelu hierarchicznego dla danych oczyszczonych z trendu są takie same jak w dwuczynnikowym modelu hierarchicznym analizy wariancji. Uogólnienie na modele wieloczynnikowe nie nastręcza większych trudności.

2. Zaletą modelowania hierarchicznego jest możliwość bezpośredniego testowania statystycznej istotności parametrów poszczególnych składowych każdego stopnia hierarchii.

3. Oceny parametrów modeli hierarchicznych dla danych oczyszczonych z trendu oraz modeli dla danych oryginalnych (z trendami) są identyczne. Ponieważ modele dla danych oryginalnych są szacowane bezpośrednio, możliwa jest budowa prognoz na ich podstawie.

PIŚMIENNICTWO

Steczkowski J., Zeliaś A. 1982. Analiza wariancyjna i kowariancyjna w badaniach ekonomicznych. Warszawa, PWN.

Szmuksta-Zawadzka M., Zawadzki J. 2000. On hierarchic models of time series with seasonal fluctuation. Dynamic Econometric Models 4, 25–30.