

Jolanta ŻUROWSKA¹

ANALIZA SEZONOWOŚCI KOLEJOWYCH PRZEWÓZÓW KWALIFIKOWANYCH

Streszczenie. W artykule zaprezentowano metodykę badania sezonowości na przykładzie miesięcznych danych o kształtowaniu się kolejowych przewozów kwalifikowanych w ostatnich trzech latach. Rozpoznano prawidłowości w zakresie tendencji rozwojowych i miesięcznych wahań przewozów.

ANALYSIS OF THE SEASONALITY CHARACTER ON QUALIFIED RAIL CARRIAGES

Summary. The methodics of the seasonality testing on case of monthly data on qualified rail carriages in last three years have been presented in the paper. The development and monthly fluctuations of carriages have been identified.

1. WPROWADZENIE

Zmienność trendów, nierównomierność i okresowe wahania przewozów mają swoje źródła, przede wszystkim po stronie popytu na przewozy i są spowodowane wieloma czynnikami. Dobowe, miesięczne lub kwartalne wahania popytu na przewozy są poważnym problemem ekonomicznym i organizacyjno-technicznym przewoźników, bowiem pociągają za sobą okresowe wahania przychodów, kosztów oraz wykorzystania potencjału przewozowego.

Zjawisko to powinno być systematycznie analizowane. Wyniki analizy sezonowości mogą i powinny być wykorzystywane w działalności marketingowej w strategiach cenowych i promocyjnych oraz powinny być przydatne w planowaniu i organizacji przewozów.

2. METODYKA BADANIA SEZONOWOŚCI

Analizę sezonowości występującej w szeregach czasowych przeprowadza się w sześciu następujących etapach:

- badanie istotności tendencji rozwojowej (trendu),
- aproksymacja tendencji rozwojowej za pomocą modeli trendu,
- eliminacja tendencji rozwojowej z szeregu czasowego (wyznaczenie odchyleń od trendu),
- badanie istotności wahań sezonowych,

¹ Zakład Organizacji i Ekonomiki Transportu, Politechnika Krakowska 31-155 Kraków, ul. Warszawska 24, tel. (+48 12) 6283292, jmzur@pk.edu.pl

- eliminacja wahań przypadkowych (obliczenie surowych składników/wskaźników sezonowych),
- obliczenie czystych składników/wskaźników sezonowych.

Do badania istotności trendu można wykorzystać parametryczny test współczynnika korelacji Pearsona. W teście tym formuluje się hipotezy:

$$H_0: \rho = 0,$$

$$H_1: \rho \neq 0.$$

Sprawdzianem hipotezy jest statystyka t określana ze wzoru:

$$t = \frac{|r| \cdot \sqrt{n-2}}{\sqrt{1-r^2}}, \quad (1)$$

gdzie:

r – oznacza współczynnik korelacji dla zmiennej czasowej oraz zmiennej analizowanej Y .

Wartość krytyczną statystyki (t_α) odczytuje się z tablic rozkładu t -Studenta dla przyjętego poziomu istotności α oraz dla $n-2$ stopni swobody. Jeżeli $|t| \leq t_\alpha$, to nie ma podstaw do odrzucenia hipotezy H_0 , co oznacza, że nie udało się udowodnić występowania w szeregu czasowym statystycznie istotnego trendu, gdy zaś $|t| > t_\alpha$, wówczas hipotezę H_0 odrzuca się, przyjmując, że w szeregu czasowym występuje trend.

Wyodrębnienie tendencji rozwojowej polega na określeniu modelu trendu i wyznaczeniu na jego podstawie wartości teoretycznych $\hat{y}_t$, a następnie wartości odchyień od trendu, które w przypadku wahań addytywnych określa się jako różnice rzeczywistych i teoretycznych wartości zmiennej ($y_t - \hat{y}_t$), a w przypadku wahań multiplikatywnych jako ich ilorazy ($y_t / \hat{y}_t$). Do opisu tendencji rozwojowej najczęściej wykorzystuje się analityczne modele trendu liniowego lub wielomianowego. Parametry tych modeli szacuje się klasyczną metodą najmniejszych kwadratów (KMNK) na podstawie wzoru, którego postać w zapisie macierzowym jest następująca:

$$\mathbf{a} = (\mathbf{X}^T \cdot \mathbf{X})^{-1} \cdot \mathbf{X}^T \cdot \mathbf{y}, \quad (2)$$

gdzie:

- $\mathbf{a}$ - wektor ocen parametrów o wymiarach $(m+1) \times 1$, przy czym m oznacza stopień wielomianu (w przypadku trendu liniowego $m = 1$),
- $\mathbf{X}$ - macierz o wymiarach $n \times (m+1)$, której kolejnymi kolumnami są odpowiednie potęgi zmiennej czasowej t , przy czym w pierwszej kolumnie występuje zmienna sztuczna równa 1 dla każdego t ,
- $\mathbf{X}^T$ - transponowana macierz $\mathbf{X}$ o wymiarach $(m+1) \times n$,
- $\mathbf{y}$ - wektor wartości empirycznych o wymiarach $n \times 1$.

Po uzyskaniu ocen parametrów trendu szacuje się wartości teoretyczne trendu (wektor $\hat{\mathbf{y}}$), korzystając z zależności:

$$\hat{\mathbf{y}} = \mathbf{X} \cdot \mathbf{a}. \quad (3)$$

W przypadku gdy w szeregu czasowym występuje duża nieregularność, można wykorzystać model trendu petzającego. W modelu tym spośród realizacji y_t ($t = 1, \dots, n$) analizowanej zmiennej, dla arbitralnie ustalonej wartości k ($1 < k < n$), zwanej stałą wygładzania, wybiera się kolejne fragmenty o długości k :

$$\begin{array}{ccccccc}
 y_1 & , & y_2 & , & \dots & , & y_k, \\
 y_2 & , & y_3 & , & \dots & , & y_{k+1}, \\
 \vdots & & & & \dots & & \vdots \\
 y_{n-k+1} & , & y_{n-k+2} & , & \dots & , & y_n.
 \end{array} \quad (4)$$

Na ich podstawie szacuje się parametry $j = n-k+1$ liniowych modeli trendu. Dla dowolnego t wartościom y_t odpowiadają wygładzone wartości (wartości teoretyczne) $\hat{y}_{jt}$ otrzymane na podstawie tych spośród podanych modeli, które odnoszą się do tego samego momentu lub okresu t . Są to mianowicie modele, dla których:

$$d_t \leq j \leq g_t, \quad (5)$$

gdzie:

$$\begin{array}{ll}
 d_t = 1 & \text{dla } 1 \leq t \leq k, \\
 d_t = t - k + 1 & \text{dla } k + 1 \leq t \leq n, \\
 g_t = t & \text{dla } 1 \leq t \leq n - k + 1, \\
 g_t = n - k + 1 & \text{dla } n - k + 2 \leq t \leq n.
 \end{array} \quad (6)$$

Dla każdego t otrzymuje się zatem L_t wartości teoretycznych, gdzie:

$$\begin{array}{ll}
 L_t = t & \text{dla } 1 \leq t \leq k, \\
 L_t = k & \text{dla } k + 1 \leq t \leq n - k + 1, \\
 L_t = n - t + 1 & \text{dla } t > n - k + 1.
 \end{array} \quad (7)$$

Elementom y_t ($t = 1, \dots, n$) szeregu czasowego przyporządkowuje się średnie wartości wszystkich takich wygładzeń, to znaczy wartości:

$$\tilde{y}_t = \frac{1}{L_t} \sum_{j=d_t}^{g_t} \hat{y}_{jt}. \quad (8)$$

Łącząc kolejne punkty $(t, \tilde{y}_t)$ odcinkami liniowymi, otrzymuje się wykres tendencji rozwojowej szeregu czasowego w postaci funkcji segmentowej, zwanej **trendem pelzającym**.

W modelu trendu pelzającego istotnym problemem jest dobór stałej wygładzania k . Jej wartość zależy od szybkości zmian poziomu zmiennej w czasie. Jeśli występują duże różnice w poziomach zmiennej w krótkich okresach, należy wybrać dość małą wartość stałej wygładzania. Jeżeli zaś zauważa się powolne zmiany, wtedy można przyjąć stałą wygładzania o wyższej wartości. Wyższa wartość stałej wygładzania powoduje większe wygładzenie szeregu, a w związku z tym słabsze reagowanie na zmiany zachodzące w szeregu czasowym. W praktyce, dobiera się taką wartość k , przy której uzyskuje się najlepsze dopasowanie trendu pelzającego do danych empirycznych, tj. najmniejszy współczynnik zbieżności φ^2 .

Do badania istotności wahań sezonowych wykorzystuje się nieparametryczny test zgodności Kendalla dla wielu zmiennych. Podstawą wyjściową jest macierz odchyień od trendu $U(N \times L)$. W wierszach tej macierzy umieszcza się wartości odchyień od trendu dla poszczególnych faz cyklu sezonowego, tak więc macierz ta ma postać:

$$\mathbf{U}=[u_{ij}], \quad i=1,2,\dots,N, \quad j=1,2,\dots,L, \quad (9)$$

gdzie:

u_{ij} - odchylenie od trendu w j -tej fazie i -tego cyklu,

L - długość cyklu sezonowego (liczba faz),

N - liczba cykli w badanym n -elementowym szeregu ($N = n/L$).

Elementom u_{ij} poszczególnych wierszy macierzy przypisuje się rangi r_{ij} od 1 do L . Otrzymuje się w ten sposób macierz rang $\mathbf{R}(N \times L)$:

$$\mathbf{R}=[r_{ij}], \quad i=1,2,\dots,N, \quad j=1,2,\dots,L. \quad (10)$$

Na podstawie macierzy rang $\mathbf{R}$ oblicza się wartość współczynnika korelacji rangowej Kendalla W (zwanego współczynnikiem zgodności), korzystając ze wzoru:

$$W = \frac{12 \cdot \sum_{j=1}^L \left\{ R_j - \left[\frac{N \cdot (L+1)}{2} \right] \right\}^2}{N^2 \cdot L \cdot (L^2 - 1)}, \quad (11)$$

gdzie:

$$R_j - \text{oznacza sumę rang dla } j\text{-tej fazy cyklu, czyli: } R_j = \sum_{i=1}^N r_{ij}.$$

W celu zbadania istotności wahań sezonowych formułuje się hipotezy:

$H_0: W = 0$ (układ N wierszy odchyłeń od trendu nie jest skorelowany, czyli wahania sezonowe są nieistotne),

$H_1: W \neq 0$ (układ N wierszy odchyłeń od trendu jest skorelowany, czyli wahania sezonowe są istotne).

Weryfikacja tej hipotezy zależy od liczebności próby (L). Dla danych miesięcznych ($L = 12$) wyznacza się wartość statystyki χ^2 na podstawie wzoru:

$$\chi^2 = N \cdot (L-1) \cdot W. \quad (12)$$

Dla danego poziomu istotności α i $(L-1)$ stopni swobody odczytuje się z tablicy rozkładu χ^2 wartość krytyczną χ_{α}^2 . Jeżeli $\chi^2 \geq \chi_{\alpha}^2$, wówczas hipotezę H_0 należy odrzucić, co oznacza, że wahania są istotne, w przeciwnym przypadku nie ma podstaw do odrzucenia hipotezy H_0 .

Chcąc wyeliminować wahania przypadkowe, należy obliczyć tzw. surowe składniki i wskaźniki sezonowe. Stanowią je wartości średnich odchyłeń dotyczących tej samej fazy cyklu sezonowego, czyli wartości średnie w poszczególnych kolumnach macierzy $\mathbf{U}$, a więc:

$$S_j^{(1)} = P_j^{(1)} = \frac{1}{N} \sum_{i=1}^N u_{ij} \quad j=1,2,\dots,L. \quad (13)$$

Korzystając z założenia, że suma składników addytywnych powinna być równa zeru, a wskaźników multiplikatywnych liczbie faz cyklu sezonowego, określa się poprawkę korygującą p jako średnią arytmetyczną składników/wskaźników surowych. Czyste składniki addytywne oblicza się z zależności:

$$S_j = S_j^{(1)} - p \quad j=1,2,\dots,L. \quad (14)$$

natomiast czyste wskaźniki multiplikatywne z zależności:

$$P_j = \frac{P_j^{(1)}}{p} \quad j = 1, 2, \dots, L. \quad (15)$$

Wahania addytywne nakładają się na trend przez dodawanie, a multiplikatywne przez mnożenie, tak więc wartości teoretyczne wyznacza się z zależności:

$$y_t^* = \hat{y}_t + S_j, \quad (16)$$

$$y_t^* = \hat{y}_t \cdot P_j, \quad (17)$$

gdzie:

$$t = 1, 2, \dots, n,$$

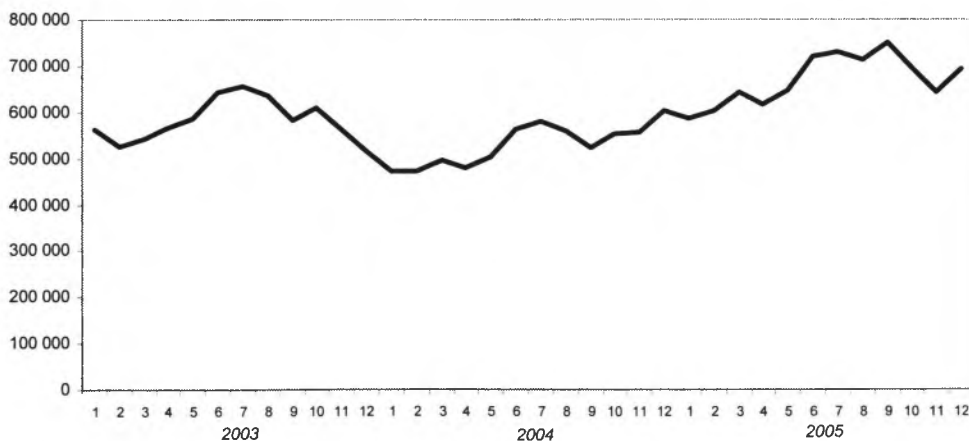
$$j = 1, 2, \dots, L.$$

Składniki addytywne są wyrażone w tych samych jednostkach miary co analizowana zmienna. Składniki dodatnie oznaczają sezonowy wzrost powyżej trendu, składniki ujemne - sezonowe zmniejszenie poniżej trendu. Wskaźniki multiplikatywne są niemianowane. Wskaźniki większe od jedności oznaczają sezonowy wzrost powyżej trendu, wskaźniki mniejsze od jedności oznaczają spadek poniżej trendu.

Wyboru rodzaju wahań sezonowych można dokonać na podstawie analizy wykresów odchyłań od trendu pogrupowanych według poszczególnych faz cyklu lub w oparciu o współczynnik zbieżności φ^2 , badający zgodność pomiędzy danymi empirycznymi i danymi teoretycznymi (im bliższa zero wartość współczynnika, tym większa zgodność).

3. ANALIZA KOLEJOWYCH PRZEWÓZÓW KWALIFIKOWANYCH

Analiza została oparta na udostępnionych przez Spółkę PKP Intercity danych o przewozach pasażerów pociągami kwalifikowanymi w poszczególnych miesiącach lat 2003-2005. Dane te ilustruje wykres na rys.1.



Rys. 1. Kolejowe przewozy kwalifikowane w latach 2003-2005

Fig. 1. Qualified rail carriages in years 2003-2005

Ocena wzrokowa wykresu wskazuje, że w badanym szeregu czasowym występuje tendencja wzrostowa. Dla potwierdzenia istotności trendu wyznaczono współczynnik korelacji dla zmiennej czasowej t oraz zmiennej analizowanej Y . Wynosi on 0,570. Wyznaczona na podstawie (1) statystyka $t = 4,044$ jest większa od wartości krytycznej $t_{\alpha} = 2,032$, wyznaczonej z rozkładu t -Studenta dla $\alpha = 0,05$ i 34 stopni swobody. Potwierdza to występowanie w analizowanym szeregu czasowym statystycznie istotnego trendu.

Aproksymację tendencji rozwojowej przeprowadzono za pomocą analitycznych modeli trendu liniowego oraz wielomianowego stopnia 2 i 3. Ze względu na to, że w przypadku modelu trendu wielomianowego stopnia 3 okazały się one statystycznie nieistotne, dalszej analizie poddano pierwsze dwa modele trendu. Wyniki przedstawia tabl.1.

Tablica 1

Analityczne modele trendu

Trend		Wahania			Trend z wahaniami
typ	φ^2	χ^2	istotność	typ	φ^2
liniowy	0,675	22,333	tak	multiplikatywne	0,487
wielomianowy st.2	0,420	21,821	tak	multiplikatywne	0,196

Źródło: obliczenia własne

Jak wynika z tablicy 1, w obu przypadkach aproksymacji tendencji rozwojowej wahania sezonowe okazały się istotne. Lepsze wyniki uzyskano dla wahań multiplikatywnych. Jednak zawarte w ostatniej kolumnie wartości współczynnika zbieżności φ^2 wskazują na nienajlepsze dopasowanie do danych empirycznych wartości teoretycznych uwzględniających tendencję rozwojową i wahania. Ponieważ metoda iteracyjna polegająca na wielokrotnym szacowaniu parametrów analitycznej funkcji trendu i wskaźników sezonowych nie dała większej poprawy wyników (uzyskano wartości $\varphi^2 = 0,484$ dla modelu trendu liniowego i $\varphi^2 = 0,189$ dla modelu trendu wielomianowego stopnia 2), do opisu tendencji rozwojowej wykorzystano model trendu pełzającego. Wyniki analizy dla różnych wartości stałej wygładzania przedstawia tabl.2.

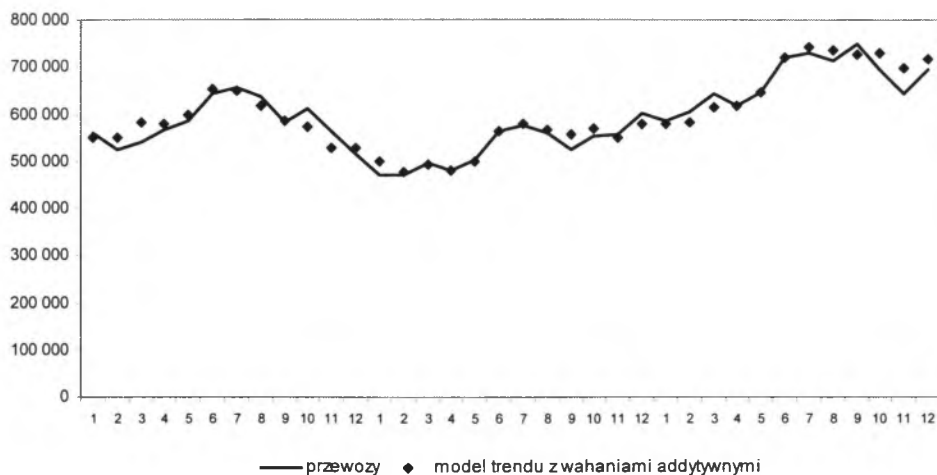
Tablica 2

Model trendu pełzającego

Trend pełzający		Wahania			Trend z wahaniami
stała wygładzania	φ^2	χ^2	istotność	typ	φ^2
3	0,900	23,154	tak	addytywne	0,713
4	0,886	22,538	tak	addytywne	0,702
5	0,852	23,051	tak	addytywne	0,670
6	0,815	23,256	tak	addytywne	0,635
7	0,775	23,308	tak	addytywne	0,597
8	0,736	24,077	tak	addytywne	0,560
9	0,114	18,026	nie		
10	0,136	17,615	nie		
11	0,167	18,641	nie		
12	0,203	19,821	tak	addytywne	0,080
13	0,237	20,333	tak	addytywne	0,089
14	0,264	21,051	tak	addytywne	0,097

Źródło: obliczenia własne

Najlepsze dopasowanie ($\varphi^2 = 0,08$) uzyskano dla modelu trendu pełzającego o stałej wygładzania równej 12 i wahaniami addytywnych (rys.2).



Rys. 2. Ilustracja dopasowania modelu

Fig. 2. Illustration of adjusting of model

Wartości czystych składników addytywnych przedstawia tabl. 3.

Tablica 3

Wartości czystych składników sezonowych

Miesiąc											
I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
-21339	-32127	-13644	-25648	-11380	42347	48751	27175	8454	8771	-22694	-8667

Źródło: obliczenia własne

Największe odchylenia poniżej trendu występują w miesiącach: luty (ponad 32 tys. pasażerów), kwiecień (ponad 25 tys. pasażerów) i listopad (ponad 22 tys. pasażerów), natomiast największe odchylenia powyżej trendu występują w lipcu i czerwcu (odpowiednio ponad 48 i 32 tys. pasażerów).

4. PODSUMOWANIE

W artykule przedstawiono analizę sezonowości kolejowych przewozów kwalifikowanych, której podstawą informacyjną były dane o liczbie pasażerów w pociągach kwalifikowanych przewożonych w poszczególnych miesiącach w latach 2003-2005, udostępnione przez Spółkę PKP Intercity. Ze względu na duże nierównomierności przewozów tendencję rozwojową opisano modelem trendu pełzającego. Analiza wykazała występowanie bezwzględnie stałych wahań sezonowych.

Literatura

1. Dittmann P.: Prognozowanie w przedsiębiorstwie. Oficyna Ekonomiczna, Kraków 2003.
2. Kuźnia S.B., Żurowska J.: Analiza sezonowości kolejowych przewozów towarowych. Problemy Ekonomiki Transportu, Zeszyt 4, Warszawa 1999.

3. Cieślak M. (red.): Prognozowanie gospodarcze. Metody i zastosowania. Redakcja Cieślak M. Wydawnictwo Naukowe PWN, Warszawa 2002, wydanie drugie rozszerzone i poprawione.
4. Żurowska J.: Prognozowanie przewozów. Modele, metody, przykłady. Politechnika Krakowska, Kraków 2005.
5. Żurowska J.: Współczynnik zgodności i jego zastosowanie w prognozowaniu dla potrzeb transportu. Zeszyty Naukowe Politechniki Śląskiej, seria Transport, z.52, Gliwice 2004.