UNIWERSYTET SZCZECIŃSKI
WYDZIAŁ NAUK EKONOMICZNYCH I ZARZĄDZANIA

STUDIA I PRACE WYDZIAŁU NAUK EKONOMICZNYCH I ZARZĄDZANIA NR 36

**Mariusz Doszyń***
**Bartłomiej Pachis****
Uniwersytet Szczeciński

# BADANIE EFEKTYWNOŚCI PROGNOZ ZMIENNYCH OPISUJĄCYCH WYBRANE ASPEKTY FUNKCJONOWANIA PORTU SZCZECIN-ŚWINOUJŚCIE

## STRESZCZENIE

Celem artykułu jest wyznaczenie prognoz wartości zmiennych opisujących wybrane aspekty funkcjonowania portu Szczecin-Świnoujście. Analizowano własności predyktywne trzech rodzajów modeli: modeli trendu ze zmiennymi zero-jedynkowymi, modeli trendu w połączniu z analizą Fouriera oraz modeli adaptacyjnych (model Holta-Wintersa). Modele porównywano na podstawie skorygowanego współczynnika determinacji oraz kryteriów informacyjnych *AIC*, *BIC* i *HQC*. Efektywność prognoz badano na podstawie różnego rodzaju błędów *ex post*. Do efektywniejszych prognoz prowadziły modele trendu, zarówno ze zmiennymi zero-jedynkowymi, jak i modele z funkcjami harmonicznymi.

**Słowa kluczowe:** prognozowanie, modele trendu ze zmiennymi sezonowymi, analiza Fouriera, porównywanie modeli, modele Holta-Wintersa, efektywność prognoz, błędy prognoz.

* Adres e-mail: mariusz.doszyn@gmail.com.

** Adres e-mail: bpachis@o2.pl.

## Wprowadzenie

Ważnym elementem wspomagającym procesy podejmowania decyzji gospodarczych są prawidłowe prognozy wartości zmiennych, istotnych z punktu widzenia procesów zarządzania przedsiębiorstwem. W związku z tym, że zazwyczaj liczba istotnych zmiennych wpływających na procesy podejmowania decyzji jest duża, konieczne jest tworzenie systemów prognoz wspomaganych odpowiednimi technikami komputerowymi. Dysponowanie tego typu systemami pozwala na sprawne i – przede wszystkim – szybkie opracowywanie prognoz dla analizowanych zmiennych.

Jakie podstawowe własności powinien mieć tego typu system prognoz? Powinien odpowiednio klasyfikować szeregi statystyczne i do każdego z nich dopasowywać najlepszą metodę prognostyczną. Ważnym elementem sytemu prognoz powinno być porównywanie modeli ze względu na ich własności statystyczne, czyli na przykład ze względu na kryteria informacyjne lub odpowiednie miary dopasowania. Ponadto prognozy powinny być podawane z odpowiednimi błędami prognoz, które wskazują na rząd ich dokładności.

## 1. Metodyka badania

Prognozowanie wartości analizowanych zmiennych odbywało się dwutorowo. Pierwsze dwie grupy zastosowanych metod (grupy A i B) to modele szeregu czasowego z wahaniami sezonowymi[1]. Analiza graficzna oraz właściwości badanych szeregów czasowych mogą wskazywać na różne postacie funkcji trendu oraz rodzaj wahań sezonowych. W budowaniu modeli prognostycznych uwzględniono zarówno postać liniową, jak i funkcje logarytmiczne i kwadratowe. Prognozowanie na podstawie modeli z trendem opisywanym przez funkcję kwadratową jest uzasadnione wtedy, gdy funkcja ta jest monotoniczna, zarówno w empirycznym obszarze zmienności, jak i w okresie prognozowanym, co też miało miejsce w rozpatrywanych przypadkach.

[1] Przegląd wybranych metod prognozowania, w tym m.in. metod stosowanych w artykule, zawierają np. prace [Batóg, Foryś, 2009; Cieślak, 2002; Gnat, 2008; Hozer, 1997; Theil, 1961; Theil, 1996; Zeliaś, 1997].

W pierwszej grupie metod (grupa A) sezonowość uwzględniono przez wprowadzanie sztucznych zmiennych zero-jedynkowych. Rozpatrywano modele, w których amplituda wahań sezonowych jest stała, relatywnie stała lub zmienia się liniowo.

W drugiej grupie metod (grupa B) szacowano także modele tendencji rozwojowej, z tym że wahania sezonowe opisywano za pomocą wielomianu trygonometrycznego. Podobnie jak poprzednio, testowano modele z liniową, logarytmiczną i kwadratową funkcją trendu. W opisie wahań sezonowych przyjmowano, że amplituda wahań sezonowych jest stała lub zmienia się liniowo. Wahania sezonowe w tej klasie modeli opisywane są za pomocą zmiennych tworzonych na podstawie funkcji trygonometrycznych (sinus, cosinus).

Trzecia grupa (grupa C) to modele wygładzania wykładniczego z wahaniami sezonowymi (model addytywny i multiplikatywny Holta-Wintersa). Modele te są zaliczane do tak zwanych adaptacyjnych metod prognozowania i często stosowane w praktyce.

W przeprowadzonym badaniu dla każdego szeregu czasowego oszacowano różne rodzaje modeli.

**I. Grupa A.**

1. Model trendu liniowego z wahaniami sezonowymi o stałej amplitudzie:

$$y_t = \alpha_0 + \alpha_1 t + \sum_{k=1}^{m-1} d_{0k} Q_{kt} + u_t \left(t = 1, 2, ..., n\right). \tag{1}$$

gdzie:

$y_t$ – zmienna objaśniana (prognozowana),
$\alpha_0, \alpha_1, d_{0k}$ $k$ = 1, 2, …, $m$ – 1) – parametry strukturalne,
$m$ – liczba podokresów (miesięcy),
$t$ – zmienna czasowa,
$Q_{kt}$ – zmienne zero-jedynkowe,
$u_t$ – składnik losowy.

Model (1) jest szacowany z założeniem, że $\sum_{k=1}^{m} d_{0k} = 0$ czyli $d_{0m} = -\sum_{k=1}^{m-1} d_{0k}$. Parametry przy zmiennych sezonowych $Q_{kt}$ informują o tym, o ile poziom zjawiska różni się w danym podokresie (na przykład miesiącu) od poziomu trendu.

2. Model trendu liniowego z wahaniami sezonowymi o liniowo zmieniającej się amplitudzie:

$$y_t = \alpha_0 + \alpha_1 t + \sum_{k=1}^{m-1} d_{ok} Q_{kt} + \sum_{k=1}^{m-1} d_{1k} t Q_{kt} + u_t \tag{2}$$

Model (2) jest dodatkowo szacowany z założeniem, że $\sum_{k=1}^{m} d_{1k} = 0$ czyli analogicznie: $d_{1m} = -\sum_{k=1}^{m-1} d_{1k}$. W modelach z liniowo zmieniającą się sezonowością amplituda wahań nie jest stała (jej zmiany w czasie są liniowe). Modele te często cechują się lepszymi wartościami kryteriów dopasowania do wartości empirycznych, ale ich wadą jest to, że parametry przy zmiennych sezonowych nie poddają się prostej interpretacji.

3. Model trendu logarytmicznego z wahaniami sezonowymi o stałej amplitudzie:

$$y_t = \alpha_0 + \alpha_1 lnt + \sum_{k=1}^{m-1} d_{ok} Q_{kt} + u_t \tag{3}$$

4. Model trendu logarytmicznego z wahaniami sezonowymi o liniowo zmieniającej się amplitudzie:

$$y_t = \alpha_0 + \alpha_1 lnt + \sum_{k=1}^{m-1} d_{ok} Q_{kt} + \sum_{k=1}^{m-1} d_{1k} t Q_{kt} + u_t \tag{4}$$

5. Model trendu w postaci funkcji kwadratowej z wahaniami sezonowymi o stałej amplitudzie:

$$y_t = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \sum_{k=1}^{m-1} d_{ok} Q_{kt} + u_t \tag{5}$$

6. Model trendu w postaci funkcji kwadratowej ze wahaniami sezonowymi o liniowo zmieniającej się amplitudzie:

$$y_t = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \sum_{k=1}^{m-1} d_{ok} Q_{kt} + \sum_{k=1}^{m-1} d_{1k} t Q_{kt} + u_t \tag{6}$$

7. Model multiplikatywny (trend wykładniczy, relatywnie stała sezonowość):

$$y_t = \beta_0 \beta_1^t \prod_{k=1}^{m-1} \gamma_k^{Q_{kt}} \cdot e^{u_t} \tag{7}$$

lub

$$lny_t = \alpha_0 + \alpha_1 t + \sum_{k=1}^{m-1} d_{ok} Q_{kt} + u_t \tag{8}$$

gdzie $\alpha_0 = ln\beta_0$, $\alpha_1 = ln\beta_1$, $d_{ok} = ln\gamma_k$ ($k$ = 1, 2, …, $m$).

W modelach zaliczonych do grupy A wahania sezonowe uwzględniane są za pomocą sztucznych zmiennych zero-jedynkowych $Q_{kt}$. Innym sposobem uwzględnienia wahań sezonowych jest dodanie do zbioru zmiennych objaśniających tak zwanych harmonik, czyli zmiennych utworzonych na podstawie funkcji trygonometrycznych (sinus i cosinus). Tego typu modele przedstawiono w grupie B.

**II. Grupa B.**

8. Model trendu liniowego z wahaniami sezonowymi w postaci harmonik o stałej amplitudzie:

$$y_t = \alpha_0 + \alpha_1 t + \sum_{i=1}^{m/2} \left( a_{0i} sin\left( \frac{2\pi it}{m} \right) + b_{0i} cos\left( \frac{2\pi it}{m} \right) \right) + u_t \tag{9}$$

W modelu (9) do zbioru zmiennych objaśniających dodano – $m/2$ zmiennych dla funkcji sinus i tyle samo zmiennych dla funkcji cosinus, z tym że ostatnia harmonika dla funkcji sinus jest zawsze na wstępie eliminowana. Jest tak dlatego, że dla $t$ = 1, 2, …, $n$ $sin(\pi t) = 0$–

9. Model trendu liniowego z wahaniami sezonowymi w postaci harmonik o liniowo zmieniającej się amplitudzie:

$$y_t = \alpha_0 + \alpha_1 t + \sum_{j=0}^{1} \sum_{i=1}^{m/2} \left( a_{0i} t^j sin\left( \frac{2\pi it}{m} \right) + b_{0i} t^j cos\left( \frac{2\pi it}{m} \right) \right) + u_t \tag{10}$$

10. Model trendu logarytmicznego z wahaniami sezonowymi w postaci harmonik o stałej amplitudzie:

$$y_t = \alpha_0 + \alpha_1 lnt + \sum_{i=1}^{m/2} \left( a_{0i} sin\left( \frac{2\pi it}{m} \right) + b_{0i} cos\left( \frac{2\pi it}{m} \right) \right) + u_t \tag{11}$$

11. Model trendu logarytmicznego z wahaniami sezonowymi w postaci harmonik o liniowo zmieniającej się amplitudzie:

$$y_t = \alpha_0 + \alpha_1 lnt + \sum_{j=0}^{1}\sum_{i=1}^{m/2}\left( a_{0i}t^j sin\left(\frac{2\pi it}{m}\right) + b_{0i}t^j cos\left(\frac{2\pi it}{m}\right)\right) + u_t \qquad (12)$$

12. Model trendu kwadratowego z wahaniami sezonowymi w postaci harmonik o stałej amplitudzie:

$$y_t = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \sum_{i=1}^{m/2}\left( a_{0i} sin\left(\frac{2\pi it}{m}\right) + b_{0i} cos\left(\frac{2\pi it}{m}\right)\right) + u_t \qquad (13)$$

13. Model trendu kwadratowego z wahaniami sezonowymi w postaci harmonik o liniowo zmieniającej się amplitudzie:

$$y_t = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \sum_{j=0}^{1}\sum_{i=1}^{m/2}\left( a_{0i}t^j sin\left(\frac{2\pi it}{m}\right) + b_{0i}t^j cos\left(\frac{2\pi it}{m}\right)\right) + u_t \qquad (14)$$

W modelach ze zmiennymi zero-jedynkowymi zmienne $Q_{kt}$ nie są eliminowane z modeli nawet wtedy, gdy test t-Studenta wskazuje na ich odrzucenie. Postępuje się tutaj zgodnie z zasadą, że jeżeli występują wahania sezonowe, do modelu dodaje się cały zbiór zmiennych $Q_{kt}$ .W dużej mierze wynika to z tego, że modele ze zmiennymi zero-jedynkowymi są szacowane warunkową KMNK, z założeniem, że suma parametrów przy zmiennych $Q_{kt}$ jest równa zeru. Parametr przy pomijanej ze względu na współliniowość zmiennej zero-jedynkowej jest obliczany jako liczba przeciwna do sumy parametrów przy pozostałych zmiennych zero-jedynkowych. Pomijanie zmiennych zero-jedynkowych w ocenach parametrów nieistotnie różniących się od zera może prowadzić do obciążenia oceny parametru przy zmiennej pomijanej.

W modelach ze zmiennymi w postaci harmonik postępuje się odmiennie. Początkowo do zbioru zmiennych objaśniających dodaje się cały zestaw harmonik, a następnie eliminuje te zmienne, dla których empiryczny poziom istotności przekracza przyjęty poziom istotności (przyjęto, że $\alpha = 0{,}1$ ).

Prognozy wyznaczone za pomocą „klasycznych" metod ekonometrycznych opartych na funkcjach trendu i odpowiednich zmiennych sezonowych porównano również z prognozami wyznaczonymi za pomocą adaptacyjnych metod prognozowania, a dokładniej – za pomocą modelu Holta-Wintersa (grupa C).

**III. Grupa C.**

14. Model Holta-Wintersa:

a) wersja addytywna:

$$F_t = \alpha\left(y_t - C_{t-m}\right) + \left(1-\alpha\right)\left(F_{t-1} + S_{t-1}\right) \tag{15}$$

$$S_t = \beta\left(F_t - F_{t-1}\right) + \left(1-\beta\right)S_{t-1} \tag{16}$$

$$C_t = \gamma\left(y_t - F_t\right) + \left(1-\gamma\right)C_{t-m} \tag{17}$$

b) wersja multiplikatywna:

$$F_t = \alpha\left(y_t / C_{t-m}\right) + \left(1-\alpha\right)\left(F_{t-1} + S_{t-1}\right) \tag{18}$$

$$S_t = \beta\left(F_t - F_{t-1}\right) + \left(1-\beta\right)S_{t-1} \tag{19}$$

$$C_t = \gamma\left(y_t / F_t\right) + \left(1-\gamma\right)C_{t-m} \tag{20}$$

gdzie:

$F_t$ – wygładzona (bez wahań sezonowych) wartość zmiennej prognozowanej w okresie *t*,

$S_t$ – przyrost trendu w okresie *t*,

$C_t$ – składnik (wskaźnik) sezonowości w okresie *t*,

$\alpha, \beta, \gamma \in \langle 0:1 \rangle$ – parametry (stałe) wygładzania.

W modelach adaptacyjnych główną kwestią jest prawidłowy wybór wartości stałych wygładzania. Ich wartości wyznaczono tak, aby minimalizowany był średni błąd średniokwadratowy wyznaczony dla okresu, dla którego szacowano wszystkie modele (styczeń 2009–czerwiec 2012).

W modelu Holta-Wintersa prognozę dla okresu *t* wyznacza się za pomocą następujących zależności:

a) wersja addytywna:

$$y_{tp} = F_n + \left(t-n\right)S_n + C_{t-m} \tag{21}$$

b) wersja multiplikatywna:

$$y_{tp} = \left[ F_n + (t-n) S_n \right] C_{t-m}. \tag{22}$$

gdzie:

$n$ – liczba wyrazów prognozowanego szeregu czasowego,
$F_n, S_n$ – ostatnie z wygładzonych wartości,
$C_{t-m}$ – ostatni z wygładzonych składników (wskaźników) sezonowości.

Dla każdej grupy metod, czyli A, B i C, wybrano najlepszy model prognostyczny[2]. W kolejnym etapie na podstawie wybranego modelu wyznaczono prognozy wygasłe (prognozy *ex post*) i na ich podstawie określono dokładność każdej z zastosowanych metod.

Dokładność opisywanych metod prognozowania zweryfikowano na podstawie następujących błędów *ex post*:

a) średni błąd predykcji:

$$ME = \frac{\sum_{t=n+1}^{n+k} \left( y_t - y_{tp} \right)}{k} \tag{23}$$

gdzie:

$y_t$ – wartości rzeczywiste zmiennej w okresie empirycznej weryfikacji prognoz,
$y_{tp}$ – prognozy *ex post* wyznaczone dla okresu $t$,
$k$ – liczba prognoz wygasłych;

b) błąd średniokwadratowy:

$$MSE = \frac{\sum_{t=n+1}^{n+k} \left( y_t - y_{tp} \right)^2}{k} \tag{24}$$

c) pierwiastek błędu średniokwadratowego:

$$RMSE = \sqrt{\frac{\sum_{t=n+1}^{n+k} \left( y_t - y_{tp} \right)^2}{k}} \tag{25}$$

[2] W grupie C „wybór" oznacza przyjęcie addytywnej bądź multiplikatywnej wersji modelu Holta-Wintersa.

d) średni błąd absolutny:

$$MAE = \frac{\sum_{t=n+1}^{n+k} \left| y_t - y_{tp} \right|}{k}. \tag{26}$$

e) średni błąd procentowy:

$$MPE = \frac{\sum_{t=n+1}^{n+k} \left( y_t - y_{tp} \right) / y_t}{k} 100. \tag{27}$$

f) średni absolutny błąd procentowy:

$$MAPE = \frac{\sum_{t=n+1}^{n+k} \left| \left( y_t - y_{tp} \right) / y_t \right|}{k} 100. \tag{28}$$

g) współczynnik Theila:

$$U^2 = \frac{\sum_{t=n+1}^{n+k-1} \left( \frac{y_{p,t+1} - y_{t+1}}{y_t} \right)^2}{\sum_{t=n+1}^{n+k-1} \left( \frac{y_{t+1} - y_t}{y_t} \right)^2}. \tag{29}$$

Pierwiastek ze współczynnika Theila $U = \sqrt{U^2}$ to względny błąd predykcji *ex post*[3].

W grupach metod A i B porównano własności poszczególnych modeli. Wyboru najlepszego modelu dokonano na podstawie wartości skorygowanego współczynnika determinacji:

$$\overline{R}^2 = 1 - \frac{S_e^2}{S_y^2} \tag{30}$$

gdzie:

$S_e^2$ – wariancja reszt modelu,
$S_y^2$ – wariancja zmiennej objaśnianej,

[3] W literaturze najczęściej podawana jest wcześniejsza wersja współczynnika Theila [Theil, 1961], jednak powszechnie przyjmuje się, że współczynnik (29) wykazuje lepsze właściwości [Theil, 1996].

oraz kryteriów informacyjnych *AIC*, *BIC*, *HQC*:

$$AIC = -2l\left(\hat{\theta}\right) + 2k \tag{31}$$

$$BIC = -2l\left(\hat{\theta}\right) + klogn \tag{32}$$

$$HQC = -2l\left(\hat{\theta}\right) + 2kloglogn \tag{33}$$

gdzie:

$l\left(\hat{\theta}\right)$ – logarytm funkcji wiarygodności,

$k$ – liczba zmiennych objaśniających,

$n$ – liczba obserwacji.

## 2. Badanie empiryczne

W badaniu empirycznym wyznaczono prognozy *ex post* dla pięciu następujących zmiennych:

$y_{1t}$ – liczba zawinięć do portu (szt.),

$y_{2t}$ – liczba przewiezionych pasażerów (os.),

$y_{3t}$ – liczba przewiezionych samochodów osobowych (szt.),

$y_{4t}$ – liczba przewiezionych samochodów ciężarowych (szt.),

$y_{5t}$ – liczba przetransportowanych trailerów (szt.).

Analizowane szeregi czasowe obejmują okres od stycznia 2009 roku do końca 2012 roku (dane miesięczne). Poszczególne rodzaje modeli szacowano na podstawie danych od stycznia 2009 roku do czerwca 2012 roku. Obserwacje od czerwca do grudnia 2012 roku uwzględniono w badaniu efektywności prognoz *ex post*. Wszystkie obliczenia wykonano na podstawie skryptów napisanych w języku *hansl* w pakiecie do obliczeń ekonometrycznych *Gretl*.

W grupach A i B modele porównano ze względu na wartości skorygowanego współczynnika determinacji $\overline{R}^2$ oraz wartości kryteriów informacyjnych *AIC*, *BIC* i *HQC*. Odpowiednie wartości przedstawiono w tabelach 1–4.

Tabela 1. Wartości skorygowanego współczynnika determinacji wyznaczone dla poszczególnych modeli[4]

| Model/ zmienna | Grupa A | | | | | | | Grupa B | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
| $y_{1t}$ | 0,639 | 0,552 | **0,643** | 0,558 | 0,638 | 0,541 | 0,637 | 0,646 | 0,703 | 0,649 | **0,706** | 0,646 | 0,699 |
| $y_{2t}$ | 0,975 | 0,981 | 0,975 | **0,981** | 0,975 | 0,980 | 0,975 | 0,976 | **0,984** | 0,976 | **0,984** | 0,976 | **0,984** |
| $y_{3t}$ | **0,977** | 0,974 | 0,976 | 0,972 | 0,977 | 0,973 | 0,968 | 0,977 | 0,978 | 0,976 | 0,975 | 0,978 | **0,979** |
| $y_{4t}$ | 0,931 | 0,897 | 0,960 | 0,941 | **0,975** | 0,968 | 0,909 | 0,928 | 0,931 | 0,959 | 0,961 | 0,975 | **0,978** |
| $y_{5t}$ | 0,582 | 0,537 | 0,621 | 0,607 | 0,673 | **0,703** | 0,626 | 0,630 | 0,693 | 0,661 | 0,732 | 0,720 | 0,793 |

Boldem oznaczono maksymalne wartości w każdej grupie modeli.

Źródło: opracowanie własne.

Tabela 2. Wartości kryterium informacyjnego *AIC* wyznaczone dla poszczególnych modeli

| Model/ zmienna | Grupa A | | | | | | | Grupa B | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
| $y_{1t}$ | 292,9 | 304,0 | **292,4** | 303,4 | 293,6 | 304,6 | *–161,0* | 289,5 | 282,8 | 289,1 | **282,4** | 289,5 | 283,4 |
| $y_{2t}$ | 786,6 | 778,1 | 786,6 | **778,1** | 788,5 | 780,1 | *–111,7* | 784,4 | **771,0** | 784,4 | **771,0** | 784,4 | **771,0** |
| $y_{3t}$ | **677,8** | 685,1 | 680,8 | 688,5 | 679,1 | 686,5 | *–118,0* | 677,8 | 676,6 | 680,8 | 681,7 | 677,1 | **675,8** |
| $y_{4t}$ | 699,7 | 718,6 | 676,8 | 695,2 | **657,0** | 669,1 | *–119,1* | 698,8 | 697,3 | 676,5 | 674,8 | 657,0 | **652,6** |
| $y_{5t}$ | 512,4 | 518,8 | 508,3 | 511,8 | 502,6 | **499,7** | *–30,9* | 500,7 | 494,6 | 497,0 | 490,5 | 490,6 | **481,8** |

Boldem oznaczono minimalne wartości w każdej grupie modeli (z pominięciem wyników dla modelu 7).

Źródło: opracowanie własne.

[4] Numery przydzielone poszczególnym modelom odpowiadają numeracji w części metodycznej.

Tabela 3. Wartości kryterium informacyjnego *BIC* wyznaczone dla poszczególnych modeli

| Model/ zmienna | Grupa A | | | | | | | Grupa B | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
| $y_{1t}$ | 315,5 | 345,7 | **315,0** | 345,1 | 317,9 | 348,1 | *–138,4* | 305,2 | 300,2 | 304,8 | **299,8** | 305,2 | 300,8 |
| $y_{2t}$ | 809,2 | 819,8 | **809,2** | 819,8 | 812,9 | 823,5 | *–89,1* | 803,5 | **797,0** | 803,5 | **797,0** | 803,5 | **797,0** |
| $y_{3t}$ | **700,4** | 726,8 | 703,4 | 730,2 | 703,4 | 729,9 | *–95,4* | 700,4 | 699,2 | 703,4 | 702,5 | 699,7 | **698,4** |
| $y_{4t}$ | 722,3 | 760,3 | 699,4 | 736,9 | **681,3** | 712,5 | *–96,6* | 714,4 | 712,9 | 695,6 | 693,9 | 681,3 | **676,9** |
| $y_{5t}$ | 535,0 | 560,5 | 530,9 | 553,5 | **527,0** | 543,2 | *–8,3* | 507,7 | 505,1 | 504,0 | 504,4 | 501,1 | **500,9** |

Boldem oznaczono minimalne wartości w każdej grupie modeli (z pominięciem wyników dla modelu 7).

Źródło: opracowanie własne.

Tabela 4. Wartości kryterium informacyjnego *HQC* wyznaczone dla poszczególnych modeli

| Model/ zmienna | Grupa A | | | | | | | Grupa B | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
| $y_{1t}$ | 301,2 | 319,2 | **300,7** | 318,7 | 302,5 | 320,5 | *–152,8* | 295,3 | 289,1 | 294,9 | **288,8** | 295,3 | 289,8 |
| $y_{2t}$ | 794,9 | 793,4 | 794,9 | **793,4** | 797,4 | 796,0 | *–103,4* | 791,4 | **780,5** | 791,4 | **780,5** | 791,4 | **780,5** |
| $y_{3t}$ | **686,1** | 700,4 | 689,1 | 703,7 | 688,0 | 702,4 | *–109,7* | 686,1 | 684,9 | 689,1 | 689,3 | 685,4 | **684,1** |
| $y_{4t}$ | 708,0 | 733,9 | 685,1 | 710,5 | **665,9** | 685,0 | *–110,9* | 704,5 | 703,0 | 683,5 | 681,8 | 665,9 | **661,5** |
| $y_{5t}$ | 520,7 | 534,0 | 516,6 | 527,1 | **511,6** | 515,6 | *–22,6* | 503,3 | 498,5 | 499,6 | 495,6 | 494,5 | **488,8** |

Boldem oznaczono minimalne wartości w każdej grupie modeli (z pominięciem wyników dla modelu 7).

Źródło: opracowanie własne.

Wartości kryteriów informacyjnych wyznaczone dla modelu 7 nie były brane pod uwagę, ponieważ zmienna objaśniana w tej klasie modeli występuje w postaci po zlogarytmowaniu. Między innymi z tego powodu przy porównywaniu modeli prognostycznych wzięto pod uwagę przede wszystkim wartości $\overline{R}^2$.

Wnioski, do jakich prowadzą wartości skorygowanego współczynnika determinacji i kryteriów informacyjnych, są do siebie bardzo zbliżone. W poszczególnych grupach modeli wartości skorygowanego współczynnika determinacji osiągają wartości maksymalne w tych przypadkach, dla których kryterium *AIC* osiąga minimum. Podobnie jest w przypadku pozostałych kryteriów (*BIC* i *HQC*), z tym że występują tutaj pewne rozbieżności.

W grupie A (zmienna $y_{2t}$) kryterium *BIC* osiąga minimalny poziom, nie jak *AIC* dla modelu 4, lecz dla modelu 3 (tabela 3). W przypadku zmiennej $y_{5t}$ kryterium *BIC* jest natomiast najmniejsze nie dla modelu 6, lecz dla modelu 5 (tabela 3). Analizując *HQC,* można zauważyć jedną różnicę. W grupie A (zmienna $y_{5t}$) *HQC* osiąga minimum nie modelu 6 jak to ma miejsce w przypadku $\overline{R}^2$ i *AIC*, ale dla modelu 5 (tabela 4). Na podstawie wartości tych kryteriów wybrano modele, które w kolejnym etapie wykorzystano do prognozowania (tabela 5).

Tabela 5. Wybrane modele prognostyczne

| Zmienna | Grupa A | Grupa B |
|---|---|---|
| $y_{1t}$ | 3 – model trendu logarytmicznego z wahaniami sezonowymi o stałej amplitudzie | 11 – model trendu logarytmicznego z wahaniami sezonowymi w postaci harmonik o liniowo zmieniającej się amplitudzie |
| $y_{2t}$ | 4 – model trendu logarytmicznego z wahaniami sezonowymi o liniowo zmieniającej się amplitudzie | 13 – model trendu kwadratowego z wahaniami sezonowymi w postaci harmonik o liniowo zmieniającej się amplitudzie |
| $y_{3t}$ | 1 – model trendu liniowego z wahaniami sezonowymi o stałej amplitudzie | 13 – model trendu kwadratowego z wahaniami sezonowymi w postaci harmonik o liniowo zmieniającej się amplitudzie |
| $y_{4t}$ | 5 – model trendu w postaci funkcji kwadratowej z wahaniami sezonowymi o stałej amplitudzie: | 13 – model trendu kwadratowego z wahaniami sezonowymi w postaci harmonik o liniowo zmieniającej się amplitudzie |
| $y_{5t}$ | 6 – model trendu w postaci funkcji kwadratowej z wahaniami sezonowymi o liniowo zmieniającej się amplitudzie | 13 – model trendu kwadratowego z wahaniami sezonowymi w postaci harmonik o liniowo zmieniającej się amplitudzie |

Źródło: opracowanie własne.

W modelach z grupy A najlepszymi własnościami, z punktu widzenia przyjętych kryteriów, cechowały się modele 1, 3, 4, 5 i 6. W grupie B zdecydowanie najlepsze okazał się model 13. W jednym przypadku najkorzystniejsze własności wykazywał model 11. „Przewaga" modelu 13 nad pozostałymi modelami bierze się po części stąd, że model ten zawiera w punkcie wyjścia najwięcej zmiennych objaśniających (harmonik), które są następnie eliminowane metodą regresji krokowej wstecznej. W modelu 13 tendencja rozwojowa ma postać funkcji kwadratowej.

Ostatni etap to zbadanie efektywności wybranych modeli prognostycznych na podstawie wcześniej przedstawionych błędów prognoz *ex post*. Wartości poszczególnych błędów przedstawiono w tabeli 6.

Tabela 6. Błędy *ex post* wyznaczone dla poszczególnych grup modeli

| Zmienna | Grupa | ME | MSE | RMSE | MAE | MPE (%) | MAPE (%) | U |
|---|---|---|---|---|---|---|---|---|
| $y_{1t}$ | A | **1,46** | 67,23 | 8,20 | 6,87 | **0,46** | 3,12 | 0,40 |
| | B | 1,74 | **34,42** | **5,87** | **4,57** | 0,67 | **2,04** | **0,29** |
| | C | –8,04 | 120,70 | 10,99 | 9,56 | –3,83 | 4,48 | 0,46 |
| $y_{2t}$ | | **–1200,70** | **14 018 000,00** | **3744,00** | 3442,60 | –2,84 | 7,44 | **0,37** |
| | B | –2047,20 | 16 132 000,00 | 4016,50 | **3367,30** | –4,73 | 7,72 | 0,46 |
| | C | –3738,95 | 23 845 623,65 | 4883,20 | 3738,95 | –7,16 | **7,16** | 0,39 |
| $y_{3t}$ | | **–876,95** | **1 313 300,00** | **1146,00** | **980,08** | **–4,48** | **5,08** | **0,34** |
| | B | –1274,50 | 2 255 900,00 | 1502,00 | 1274,50 | –6,74 | 6,74 | 0,50 |
| | C | –1827,19 | 5 110 358,32 | 2260,61 | 1860,77 | –9,01 | 9,24 | 0,57 |
| $y_{4t}$ | | –296,15 | 934 880,00 | 966,89 | 740,99 | –1,74 | 3,51 | 0,26 |
| | B | **–217,31** | **775 420,00** | **880,58** | **700,77** | **–1,38** | **3,25** | **0,25** |
| | C | –1102,17 | 2 188 230,62 | 1479,27 | 1102,17 | –5,04 | 5,04 | 0,46 |
| $y_{5t}$ | | 169,69 | 37 929,00 | 194,75 | 170,44 | 20,35 | 20,43 | 1,54 |
| | B | 180,66 | 41 698,00 | 204,20 | 180,66 | 20,92 | 20,92 | 1,73 |
| | C | 127,61 | 18 045,73 | 134,33 | 127,61 | 14,14 | 14,14 | 1,05 |

Boldem oznaczono wartości minimalne w poszczególnych grupach modeli dla poszczególnych zmiennych.

Źródło: obliczenia własne.

Na podstawie wyznaczonych błędów *ex post* można stwierdzić, że do prognozowania rozpatrywanych szeregów najbardziej użyteczne są „klasyczne" ekonometryczne modele szeregów czasowych, a więc modele należące do grup A i B[5].

W przypadku zmiennych $y_{1t}, y_{4t}$ najlepsze właściwości predyktywne wykazywały modele z grupy B, a więc modele tendencji rozwojowej z wahaniami sezonowymi w postaci harmonik. Najlepszym modelem do prognozowania zmiennej $y_{1t}$ okazał się model trendu logarytmicznego z wahaniami sezonowymi w postaci harmonik o liniowo zmieniającej się amplitudzie. Prognozując wartości zmiennej $y_{4t}$, najlepiej zastosować model trendu kwadratowego z wahaniami sezonowymi w postaci harmonik o liniowo zmieniającej się amplitudzie. Modele zaliczone do grupy A cechowały się najlepszymi własnościami predyktywnymi w przypadku zmien-

[5] Nie scharakteryzowano szczegółowych wyników badań z powodu ograniczonej objętości artykułu.

nych $y_{2t}, y_{3t}$. Do prognozowania wartości zmiennej $y_{2t}$ najlepszy jest model trendu logarytmicznego z wahaniami sezonowymi o liniowo zmieniającej się amplitudzie, a do prognozowania zmiennej $y_{3t}$ – model trendu liniowego z wahaniami sezonowymi o stałej amplitudzie. Do wyznaczania prognoz dla zmiennej $y_{5t}$ najlepiej stosować multiplikatywny model adaptacyjny Holta-Wintersa.

## Podsumowanie

W artykule podjęto próbę systemowego podejścia do wyznaczania prognoz wartości zmiennych charakteryzujących wybrane aspekty działalności Portu Szczecin-Świnoujście w okresie od stycznia 2006 roku do grudnia 2012 roku. Analizowano własności predyktywne trzech rodzajów modeli: tendencji rozwojowej z wahaniami sezonowymi opisywanymi za pomocą zmiennych zero-jedynkowych, modeli tendencji rozwojowej z wielomianami trygonometrycznymi oraz modele wygładzania wykładniczego Holta-Wintersa. W pierwszym etapie na podstawie skorygowanego współczynnika determinacji i kryteriów informacyjnych *AIC*, *BIC*, *HQC* wybrano najlepsze modele w grupach A i B. Następnie zbadano efektywność prognoz *ex post* wyznaczonych na podstawie wcześniej dobranych modeli. Analizując własności predyktywne uwzględnionych modeli można stwierdzić, że w przypadku badanych szeregów czasowych najlepsze prognozy dają modele tendencji rozwojowej z wahaniami sezonowymi opisywanymi za pomocą zmiennych zero-jedynkowych lub harmonik.

## Literatura

Batóg B., Foryś I. (2009), *Prognozowanie zużycia ciepłej i zimnej wody w spółdzielczych zasobach mieszkaniowych*, w: *Metody ilościowe w ekonomii*, Studia i Prace WNEiZ nr 2, Szczecin.

Cieślak M. (red.) (2002), *Prognozowanie gospodarcze. Metody i zastosowania*, Wydawnictwo Naukowe PWN, Warszawa.

Gnat S. (2008), *Prognozowanie dochodów ze sprzedaży tygodników lokalnych – wybrane podejścia*, w: *Metody ilościowe w ekonomii*, Studia i Prace WNEiZ nr 2, Szczecin.

Hozer J. (red.) (1997), *Ekonometria*, Katedra Ekonometrii i Statystyki, Stowarzyszenie Pomoc i Rozwój, Szczecin.
Theil H. (1961), *Economic Forecasting and Policy*, North-Holland, Amsterdam.
Theil H. (1996), *Applied Economic Forecasting*, North-Holland, Amsterdam.
Zeliaś A. (1997), *Teoria prognozy*, PWE, Warszawa.

## STUDY OF THE FORECASTS EFFICIENCY OF VARIABLES DESCRIBING SELECTED ACTIVITIES OF SZCZECIN-ŚWINOUJŚCIE HARBOUR

### Abstract

In the article an attempt to compute forecasts of variables that characterize selected activities of Szczecin-Świnoujście Harbour in a systematic way was made. Analysed time series include data from January 2006 to December 2012. Predictive properties of three kind of models were considered: trend models with dummy variables as a seasonal variables, trend models with Fourier analysis and adaptive Holt-Winters models. At the first stage, model comparison was made on the basis of adjusted determination ratio as well as *AIC*, *BIC* and *HQC* information criteria. Then efficiency of forecast was investigated with using *ex post* forecasts errors. In most cases trend models with seasonal dummy variables and with Fourier analysis turned out to be the most effective.

*Translated by Mariusz Doszyń*

**Keywords:** forecasting, trend models with seasonal dummy variables, Fourier analysis, models' comparison, Holt-Winters models, forecasts efficiency, forecasts errors.
**Kod JEL:** C53, C12, C13.