

Jan Zawadzki

Akademia Rolnicza w Szczecinie

WYKORZYSTANIE MODELI HOLTA-WINTERSA W PROGNOZOWANIU BRAKUJĄCYCH DANYCH

W praktyce w prognozowaniu zmiennych ekonomicznych na podstawie modeli adaptacyjnych z wahaniami sezonowymi najczęściej wykorzystuje się modele Holta-Wintersa w postaci addytywnej lub multiplikatywnej. Modele te składają się z trzech równań opisujących kolejno operator rzędu pierwszego (m_t), parametr kierunkowy trendu (δ_{1t}) oraz efekt sezonowy (C_t). Modele Holta-Wintersa są w istocie rozwiniętymi o równanie efektu periodycznego wersjami dwuparametrowego modelu Holta.

Zapis modelu w postaci addytywnej jest następujący:

$$m_t = \alpha(Y_t - c_{t-l}) + (1 - \alpha)m_{t-1}, \quad (1)$$

$$\delta_{1t} = \beta(m_t - m_{t-1}) + (1 - \beta)\delta_{1t-1}, \quad (2)$$

$$C_t = \gamma(Y_t - m_t) + (1 - \gamma)C_{t-m}, \quad (3)$$

gdzie m oznacza długość okresu wahań periodycznych (4 kwartały lub 12 miesięcy), zaś α, β, γ są stałymi wygładzenia przyjmującymi wartości z przedziału $[0, 1]$.

Predyktor oparty na tym modelu wyraża się wzorem:

$$\Pi = m_{t_0} + \delta_{1t_0}h + C_{t_0+1+h}, \quad h = 1, 2, \dots, m, \quad (4)$$

gdzie h oznacza horyzont prognozy C_{t_0-m+h} .

Natomiast zapis postaci multiplikatywnej modelu Holta-Wintersa wygląda następująco:

$$m_t = \frac{\alpha Y_t}{C_{t-m}} + (1 - \alpha)(m_{t-1} + \delta_{1t-1}), \quad (5)$$

$$\delta_{1t} = \beta(m_t - m_{t-1}) + (1 - \beta)\delta_{1t-1}, \quad (6)$$

$$C_t = \frac{\delta Y_t}{m_t} + (1 - \delta)C_{t-m}. \quad (7)$$

Predyktor oparty na tym modelu przyjmuje postać:

$$\Pi = (m_{t_0} + \delta_{1t_0} h)C_{t_0-n+h}. \quad (8)$$

W literaturze proponuje się inną wersję modelu, różniącą się równaniem opisującym ocenę przyrostu trendu:

$$\delta_{1t} = \frac{\beta(m_t - m_{t-1})}{m_{t-1}} + (1 - \beta)\delta_{1t-1}. \quad (9)$$

Składniki sezonowości w każdym roku powinny się sumować do zera, natomiast iloczyn wskaźników powinien być równy jedności (lub ich suma powinna być równa m).

Wartości parametrów α , β , γ (stałych wygładzenia) najczęściej wybiera się metodą kolejnych empirycznych przybliżeń. Może to być jednak kłopotliwe z uwagi na liczbę kombinacji parametrów. Jeżeli przyjmiemy, że każdy z parametrów zmienia się od 0,1 do 0,9 co 0,1 otrzymamy 9^3 , czyli 729 kombinacji. Na rynku pojawiły się programy komputerowe wyznaczające optymalne wartości stałych wygładzenia w sposób automatyczny. Jednym z takich programów jest STATISTICA. Dla parametrów α , β , γ zmieniających się od 0,05 do 0,95 co 0,05 tworzą 19^3 , czyli 6859 kombinacji, czas obliczeń wynosi kilkanaście sekund. Jedynym ograniczeniem jest konieczność posiadania danych statystycznych obejmujących minimum 5 cykli (lat). Dla danych miesięcznych będzie to 60 obserwacji.

Rozważania prowadzone dotąd dotyczyły kompletnych szeregów czasowych. Dość często zdarza się jednak, że dane statystyczne o kształtowaniu się zmiennych ekonomicznych wykazujących wahania sezonowe nie są kompletne.

Wykorzystywanie modeli klasycznych lub hierarchicznych szeregu czasowego może, ze względu na szybko zachodzące zmiany o charakterze strukturalnym, nie dawać zadowalających wyników. Stąd w pracy proponujemy wykorzystanie do tego celu modeli wyrównania wykładniczego Holta-Wintersa.

Dla niekompletnych szeregów czasowych nie jest możliwe zastosowanie automatycznych procedur wyszukiwania optymalnych wartości stałych wygładzenia (parametrów wyrównywania) α , β i δ . Praktycznie jedyną dostępną metodą jest wspomniana wcześniej metoda polegająca na rozpatrywaniu wszystkich kombinacji parametrów i wyznaczeniu dla każdej z nich określonego miernika dokładności dla wartości wyrównanych obliczonych dla dostępnych danych. Optymalna będzie ta kombinacja, dla której przyjmie on najniższą wartość.

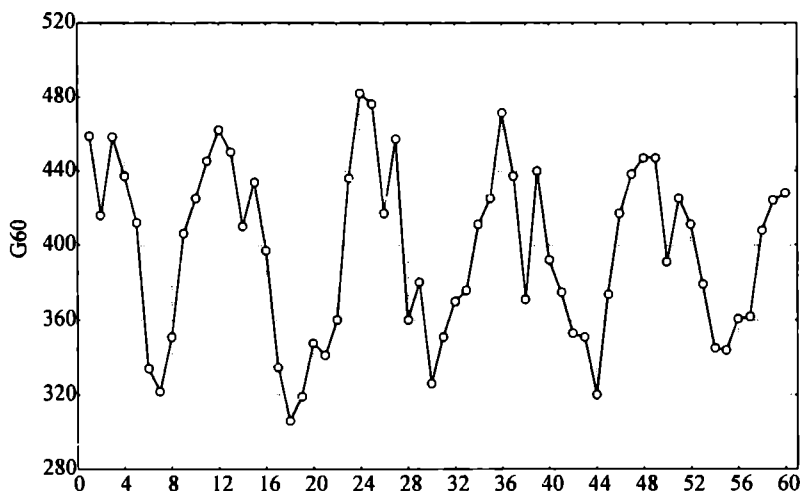
W prognozowaniu brakujących danych w szeregach czasowych dla danych sezonowych wyróżnia się dwa rodzaje luk: luki niesystematyczne i luki systematyczne.

Występowanie luk niesystematycznych oznacza dysponowanie przynajmniej jedną informacją o każdym sezonie. Natomiast z lukami systematycznymi będziemy mieć do

czynienia wtedy, gdy przynajmniej w jednym sezonie brakować będzie danych. Rozróżnienie to jest wystarczające, gdy wykorzystuje się w prognozowaniu modele „oszczędne” uwzględniające statystycznie istotne składowe harmoniczne oraz modele hierarchiczne.

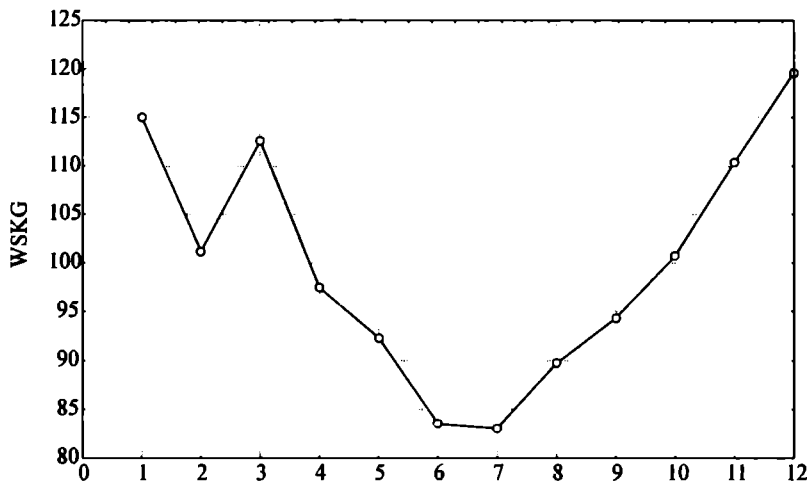
Natomiast w sytuacji, kiedy prognozy buduje się na podstawie modeli adaptacyjnych, czynnikiem decydującym o tym, z jakiego rodzaju lukami szeregi będą mogły być przedmiotem prognozowania będzie sposób wyznaczania wartości początkowych. Jest to konieczne, ponieważ we wzorach (1) - (3) oraz (5) - (7) występują opóźnienia jedno-okresowe bądź m -okresowe. Należy wyznaczyć wartości początkowe: m_0 , δ_0 oraz C_{1-m} . W literaturze proponuje się wykorzystanie do tego celu danych z pierwszego roku. Drugi sposób polega na wprowadzaniu informacji z „zewnątrz”. Mogą one być obliczone dla okresu bezpośrednio poprzedzającego przedział czasowy próby.

W drugim przypadku rodzaj luk nie ma znaczenia. Natomiast w pierwszym budowa prognoz jest możliwa jedynie dla szczególnego rodzaju szeregów z lukami niesystematycznymi, tylko takich, które zawierają przynajmniej m danych na początku szeregu. Jest to konieczne do oszacowania wyjściowych wartości m.in. wskaźników sezonowości. Próba zastosowania multiplikatywnego modelu Holta-Wintersa do prognozowania brakujących danych zostanie przedstawiona na przykładzie prognozowania dostaw gazu. Przedział czasowy „próby” obejmuje dane miesięczne za lata 2000-2003. Rok 2004 będzie okresem empirycznej weryfikacji prognoz. Kształtowanie się zmiennej prognozowanej oraz kształtowanie się ocen wskaźników sezonowości przedstawiono na rys. 1 i 2.



Rys. 1. Kształtowanie się produkcji gazu w latach 2000-2004

Źródło: opracowanie własne.



Rys. 2. Oceny wskaźników sezonowości produkcji gazu

Źródło: opracowanie własne.

Rozpatrywane będą cztery warianty niesystematycznych luk w danych (przy założeniu $m = 12$ wartości początkowych). Luki te otrzymano przez „wymazanie” obserwacji z pełnego szeregu. Warianty luk były następujące:

- A – luki w miesiącach nieparzystych,
- B – luki w miesiącach parzystych,
- C – luki obejmujące na przemian kolejne dwa miesiące, zaczynając od 13 i 14,
- D – luki obejmujące kwartały o numerach parzystych.

W wariantach A-D luki będą występować w 18 miesiącach, tzn. będą obejmować połowę obserwacji z okresu $m + 1$ do n . Symbolem P oznaczono wariant bez luk w danych. Mierniki dokładności wartości wyrównanych oraz prognoz ekstrapolacyjnych będą porównane z tego rodzaju miernikami otrzymanymi dla wariantów z lukami. Optymalne wartości stałych wygładzania ustalono metodą kolejnych przybliżeń („prób i błędów”). Przyjęto, że każda ze stałych wygładzania przyjmuje wartości od 0,1 do 0,9, zmieniające się co 0,1. Oznacza to, że dla każdego wariantu luk obliczono 9^3 , czyli 729 kombinacji tych parametrów. Kryterium wyboru „optymalnych” kombinacji dla pełnego szeregu były względne błędy wartości wyrównanych (prognoz dla przedziału czasowego „próby”):

$$\text{MAPE}^{\alpha, \beta, \gamma} = \frac{1}{n - m} \frac{\sum_{t=m+1}^n |y_t - \hat{y}_{tp}^{\alpha, \beta, \delta}|}{\sum_{t=m+1}^n y_t} \quad (10)$$

Dla wariantów z lukami sumowanie dotyczyło tylko tych okresów, dla których dane były dostępne. Dla każdego wariantu uszeregowano rosnąco oceny miernika danego wzorem (10).

Ponieważ przynajmniej kilka wartości początkowych tego miernika różniło się nieznacznie od oceny minimalnej, to dla każdego wariantu wybierano nie jedną, lecz dziesięć kombinacji.

W tab. 1 zestawiono błędy względne wartości wyrównanych.

Tabela 1. Oceny średnich względnych błędów wartości wyrównanych (w %)

Lp.	Wariant P		Wariant A		Wariant B		Wariant C		Wariant D	
1	P113	6,58	A111	6,51	B112	5,99	C112	5,14	D111	7,11
2	P213	6,60	A112	6,61	B113	5,91	C113	5,18	D112	7,17
3	P214	6,63	A121	6,75	B114	5,96	C114	5,21	D114	7,24
4	P313	6,65	A122	6,84	B111	6,11	C115	5,23	D115	7,30
5	P314	6,68	A113	6,84	B122	6,51	C116	5,24	D113	7,36
6	P414	6,72	A211	6,85	B115	6,12	C117	5,25	D124	7,42
7	P415	6,77	A212	6,89	B124	6,35	C118	5,29	D123	7,49
8	P416	6,82	A213	6,92	B121	6,61	C124	5,31	D213	7,55
9	P417	6,85	A123	6,94	B214	6,45	C125	5,34	D313	7,65
10	P425	6,88	A131	6,94	B314	6,65	C126	5,34	D413	7,67

Źródło: opracowanie własne.

Przy literach oznaczających warianty podane zostały trzy cyfry. Odpowiadają one wartościom parametrów α , β i γ , tworzących daną kombinację (np. P113 oznacza, że $\alpha = 0,1$, $\beta = 0,1$ i $\gamma = 0,3$).

Z informacji zawartej w kolumnie trzeciej wynika, że oceny miernika $MAPE^{\alpha,\beta,\gamma}$, dla pełnego szeregu (wariant P) wahają się od 6,58 do 6,88%, tzn. różnią się między sobą tylko o 0,3 punktu procentowego. W przypadku czterech wariantów luk w danych optymalne wartości stałych wygładzenia były liczone dla dostępnych informacji z okresu $m+1$ do n . Oceny odpowiadające poszczególnym wariantom są podane w następnych kolumnach. Zostały one także uporządkowane rosnąco. Widoczne jest ich na ogół wyższe zróżnicowanie niż w przypadku szeregu bez luk. Waha się ono od 0,43 punktu procentowego dla wariantu A do 0,74 punktu procentowego dla wariantu B. Jedynie dla wariantu C zróżnicowanie to jest niższe i wynosi 0,2 punktu procentowego. Minimalne wartości błędów względnych zostały zaznaczone „tłustym drukiem”. Z ich porównania wynika, że jedynie dla wariantu D otrzymano ocenę wyższą niż oceny dla wariantu P (bez luk w danych). Zwracają uwagę niższe na ogół wartości stałych wygładzenia α w wariantach z lukami w danych. Świadczy to o tym, że nadaje się niższe wagi obserwacjom najnowszym.

W tab. 2 w takim samym porządku podano oceny błędów względnych prognoz interpolacyjnych, obliczone dla danych, w których występowały luki.

Tabela 2. Oceny średnich względnych błędów prognoz interpolacyjnych (w %)

Lp.	Wariant P		Wariant A		Wariant B		Wariant C		Wariant D	
		INT		INT		INT		INT		INT
1	P113	6,58	A111	6,18	B112	5,97	C112	6,22	D111	6,06
2	P213	6,60	A112	6,16	B113	5,93	C113	6,22	D112	6,07
3	P214	6,63	A121	6,21	B114	5,89	C114	6,22	D114	6,09
4	P313	6,65	A122	6,19	B111	6,01	C115	6,23	D115	6,13
5	P314	6,68	A113	6,14	B122	6,14	C116	6,23	D113	6,08
6	P414	6,72	A211	6,07	B115	5,85	C117	6,23	D124	6,20
7	P415	6,77	A212	6,08	B124	6,02	C118	6,23	D123	6,19
8	P416	6,82	A213	6,10	B121	6,21	C124	6,40	D213	6,39
9	P417	6,85	A123	6,17	B214	5,71	C125	6,39	D313	0,00
10	P425	6,88	A131	6,20	B314	5,51	C126	6,38	D413	7,02

Źródło: opracowanie własne.

Druga i trzecia kolumna są takie same jak w tab. 1, ponieważ dla pełnych szeregów nie liczy się prognoz interpolacyjnych. Z porównania tych ocen z minimalnymi ocenami wartości wyrównanych podanych w tab. 1 wynika, że wszystkie są niższe od błędu wariantu P (bez luk). W przypadku porównań w ramach poszczególnych wariantów sytuacja jest zróżnicowana. Dla wariantów A, C i D minimalne oceny błędów są zbliżone. Jedynie dla wariantu B ocena ta jest o ponad pół punktu procentowego niższa. Niższą oceną, o blisko jeden punkt procentowy, charakteryzuje się wariant D (luki w kwartałach parzystych). Dla wariantu A jest ona także niższa o ponad 0,4 punktu procentowego. W przypadku wariantu C (luki występują co drugi miesiąc) miał miejsce wzrost o 1,08 punktu procentowego. Najniższą oceną charakteryzował się wariant B (luki w miesiącach parzystych), najwyższą zaś wariant C (luki dwumiesięczne).

Z informacji zawartych w tab. 2 wynika również, że nie zawsze minimalne oceny błędów prognoz interpolacyjnych odpowiadają kombinacjom o najniższych ocenach błędów wartości wyrównanych zawartych w tab. 1.

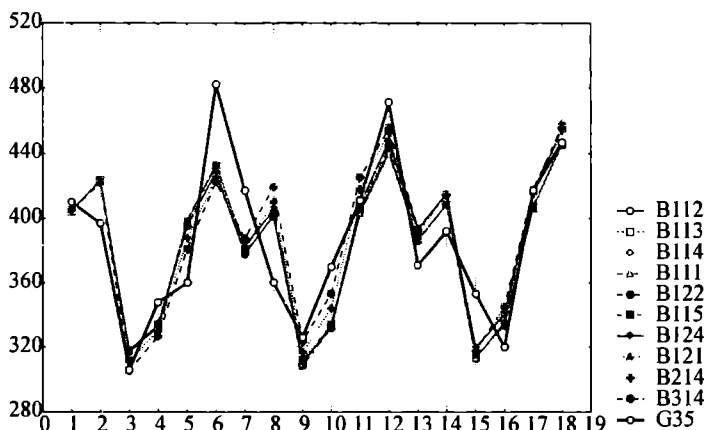
W tab. 3 podano oceny względnych błędów prognoz *ex post*.

Tabela zawiera także oceny błędów dla modelu bez luk. Nieznacznie niższe oceny minimalne błędów względnych, w porównaniu z wariantem P, otrzymano dla wariantu B (luki w miesiącach parzystych) oraz D (luki w kwartałach parzystych). Natomiast wyższe oceny, odpowiednio o 1,16 i 1,12 punktu procentowego, otrzymano dla wariantów A i C.

Tabela 3. Oceny średnich względnych błędów prognoz extrapolacyjnych (w %)

Lp.	Wariant P		Wariant A		Wariant B		Wariant C		Wariant D	
		Ex		Ex		Ex		Ex		Ex
1	P113	3,09	A111	4,10	B112	3,34	C112	4,06	D111	3,77
2	P213	3,07	A112	4,19	B113	3,00	C113	3,98	D112	2,78
3	P214	2,94	A121	4,53	B114	2,87	C114	3,97	D114	2,84
4	P313	3,12	A122	4,43	B111	3,79	C115	3,98	D115	3,30
5	P314	3,12	A113	4,23	B122	4,93	C116	4,02	D113	2,94
6	P414	3,43	A211	4,12	B115	2,75	C117	4,06	D124	2,84
7	P415	3,59	A212	4,13	B124	4,30	C118	4,10	D123	3,18
8	P416	3,79	A213	4,13	B121	5,39	C124	4,06	D213	2,99
9	P417	3,93	A123	4,32	B214	3,68	C125	3,96	D313	3,19
10	P425	4,61	A131	5,02	B314	3,80	C126	3,96	D413	3,29

Źródło: opracowanie własne.



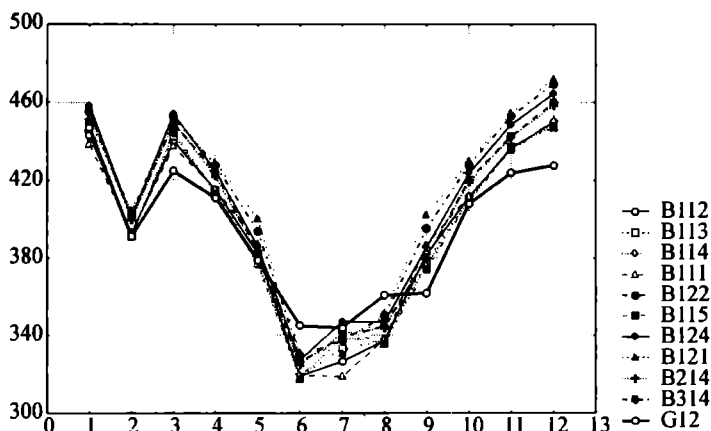
Rys. 3. Prognozy interpolacyjne dla wariantu B

Źródło: opracowanie własne.

Zwraca uwagę fakt, że błędy prognoz ekstrapolacyjnych są znacznie niższe (niekiedy nawet więcej niż o połowę) od błędów prognoz interpolacyjnych. Rozstęp błędów w ramach poszczególnych wariantów nie przekracza na ogół jednego punktu procentowego. Jedynie w przypadku wariantu B jest to więcej niż 2,5 punktu procentowego. Najwyższe zróżnicowanie błędów prognoz interpolacyjnych wynosiło 0,7 punktu procentowego. Jeżeli w ramach poszczególnych wariantów porówna się wersje modeli charakteryzujące się minimalnymi ocenami błędów prognoz ekstrapolacyjnych, to w żadnym przypadku nie pokrywają się one z wersjami dla prognoz interpolacyjnych. Je-

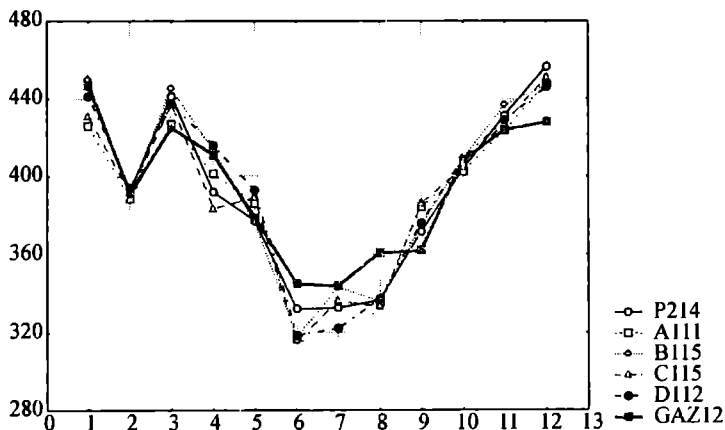
dynie dla wariantu A występuje zgodność z minimalną oceną błędu wartości wyrównanych. Jest to więc jeszcze dodatkowy, ważny argument przemawiający za tym, aby w prognozowaniu nie ograniczać się jedynie do modelu z optymalną kombinacją parametrów α , β i δ .

Na rys. 3 i 4 przedstawiono w formie graficznej odpowiednio wyniki prognozowania interpolacyjnego i ekstrapolacyjnego dla wariantu B.



Rys. 4. Prognozy ekstrapolacyjne dla wariantu B

Źródło: opracowanie własne.



Rys. 5. Prognozy ekstrapolacyjne dla najlepszych wersji wszystkich wariantów

Źródło: opracowanie własne.

Z kolei rys. 5 prezentuje wyniki prognozowania ekstrapolacyjnego dla najlepszych predyktorów z wszystkich pięciu wariantów.

Z przytoczonego przykładu wynika, że modele wyrównywania wykładniczego mogą być wykorzystywane z powodzeniem w prognozowaniu brakujących danych w sytuacji, gdy dysponujemy m obserwacjami początkowymi.

Literatura

- Chatfield C., *The Holt-Winters Forecasting Procedure*, „Applied Statistics” 1978, vol. 27, no 3.
 Reid D.J., *A Review of Short-Term Projection Techniques*, [w:] *Practical Aspects of Forecasting*, H. Gordon Ed., London 1975.
 Winters P.R., *Forecasting Sales by Exponentially Weighted Moving Averages*, „Management Science” 1960.
 Zawadzki J., *O wykorzystaniu modeli adaptacyjnych w prognozowaniu brakujących danych*, PS Szczecin 2003.
 Zawadzki J., *Modelowanie predykcyjne i prognozowanie zjawisk w skali mikroekonomicznej*, US Szczecin 1996.

APPLICATION OF HOLT-WINTERS MODELS IN FORECASTING OF MISSING DATA

Summary

The article presents exponential smoothing models in forecasting of economical variables with seasonal fluctuations in situation of lack of full information. The article proves that Holt-Winters models are useful in forecasting of times series with unsystematical gaps, which contained at least m observations at the beginning of the time series (m – length of cycle of periodical fluctuations). These observations are needed for estimate of output values of components for seasonality and trend. It attention was paid also to estimate “optimal” constant values of smoothing. Theoretical considerations in article were illustrated by empirical example.