

Maria Szmuksta-Zawadzka

Politechnika Szczecińska

Jan Zawadzki

Akademia Rolnicza w Szczecinie

ZASTOSOWANIE MODELI WYRÓWNYWANIA WYKŁADNICZEGO W PROGNOZOWANIU BRAKUJĄCYCH DANYCH DEKADOWYCH

1. Wstęp

Wykorzystywanie modeli szeregu czasowego ze stałą postacią analityczną i stałymi parametrami w prognozowaniu zmiennych ekonomicznych podlegających szybko zachodzącym zmianom o charakterze strukturalnym dość często nie pozwala na otrzymanie prognoz dopuszczalnych. Stąd w literaturze proponuje się wykorzystanie do tego celu modeli wyrównania wykładniczego Holta–Wintersa. Modele te dla kompletnych danych dają na ogół prognozy dopuszczalne. W pracy [4] przedstawiono propozycję wykorzystania omawianej klasy modeli do prognozowania brakujących danych dla szeregów miesięcznych.

W niniejszej pracy dokonamy rozszerzenia rozważań na zmienne ekonomiczne z wahaniami o cyklu 36-dekadowym. Ich rozszerzenie na dane dekadowe nie jest tylko prostą kontynuacją rozważań dla danych miesięcznych, ponieważ istotną rolę w prognozowaniu odgrywa struktura harmoniczna zmiennych.

2. Rozważania teoretyczne

Do najczęściej wykorzystywanych modeli w prognozowaniu zmiennych z wahaniami sezonowymi należą modele Holta–Wintersa (HW) w postaci addytywnej lub multiplikatywnej oraz model Harrisona. Opisy tych modeli i przykłady ich zastosowań w prognozowaniu na podstawie kompletnych szeregów czasowych można znaleźć m.in. w pracach [1; 3; 5].

Modele te składają się z 3 równań opisujących kolejno: operator rzędu pierwszego, parametr kierunkowy trendu oraz efekt sezonowy (składniki lub wskaźniki sezonowości).

Model addytywny Holta–Wintersa jest rozwiniętą – o równanie efektu periodycznego C_t – wersją 2-parametrowego modelu Holta.

Równania tego modelu są następujące:

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

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

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

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

Predyktor oparty na tym modelu wyraża się wzorem:

$$\Pi = m_{t_0} + \delta_{it_0}h + C_{t_0+1+h}. \quad (4)$$

W modelu powyższym zakłada się liniowe zmiany trendu.

Natomiast multiplikatywna postać modelu Holta–Wintersa jest następująca:

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

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

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

Oceny C_t są wskaźnikami sezonowości. Predyktor oparty na modelu przybiera postać:

$$\Pi = (m_{t_0} + \delta_{it_0}h)C_{t_0-m+h}.$$

Wartości α, β, γ stałych wygładzenia α, β, γ wybiera się najczęściej metodą kolejnych empirycznych przybliżeń. Może to być jednak kłopotliwe ze względu na dość dużą liczbę kombinacji parametrów. Jeżeli przyjmiemy, że każdy z parametrów zmienia się od 0,1 do 0,9 co 0,1, to otrzymamy $9^3 = 729$ kombinacji. Na rynku pojawiły się programy komputerowe wyznaczające optymalne wartości stałych wygładzenia w sposób automatyczny, jednym z nich jest Statistica. Jedynym ograniczeniem jest konieczność posiadania pełnych danych statystycznych obejmujących minimum 5 cykli (lat).

Rozważania prowadzone dotąd dotyczyły kompletnych szeregów czasowych. Modele wyrównywania wykładniczego Holta–Wintersa mogą być wykorzystywane w prognozowaniu niekompletnych szeregów czasowych z lukami niesystematycznymi. Występowanie luk niesystematycznych oznacza dysponowanie przynajmniej jedną informacją o każdym sezonie. W przypadku omawianej klasy modeli odnosi się to do $m + 1$ wartości początkowych. Znajomość ich jest konieczna do wyznaczenia wartości początkowych (startowych) dla $t = 1$ w okresie m_0 , δ_0 oraz C_{1-m} . To znaczy, że zastosowanie automatycznych procedur wyszukiwania optymalnych wartości stałych wygładzenia (parametrów wyrównywania) α , β oraz δ nie jest możliwe. Dzieje się tak dlatego, że w module pakietu Statistica brakujące dane zastępuje się jedną z czterech metod, które są wskazane w programie. W tej sytuacji 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 miernik przybierze wartość najniższą.

3. Wyniki empiryczne

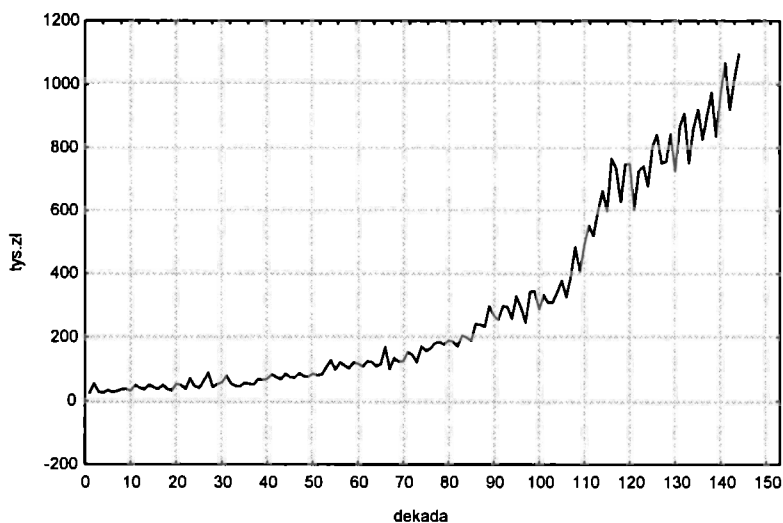
W części empirycznej pracy przedstawione zostaną wyniki zastosowania multiplikatywnego modelu Holta–Wintersa do prognozowania dekadowych stanów depozytów na rachunkach oszczędnościowo-rozliczeniowych w oddziale Banku Z (zmienna Y w tys. zł). Depozyty te charakteryzują się występowaniem zmiennych wahań sezonowych. Kształtowanie się badanej zmiennej przedstawiono na rys. 1.

Z rysunku wynika, że dekadowe stany depozytów (Y) wykazują także wahania o cyklu krótszymi niż rok (36 dekad). Informację o tym, z jaką długością cyklu mamy do czynienia, można uzyskać, badając strukturę harmoniczną zmiennej prognozowanej.

W tabeli 1 przedstawione zostały udziały wariancji składowych sinuso- i cosinusoidalnych oraz harmonik w wyjaśnieniu ogólnej wariancji sezonowej zmiennej prognozowanej. Z informacji zawartych w tabeli wynika, że zdecydowanie dominującymi są harmoniki: pierwsza o cyklu 36-dekadowym i dwunasta o cyklu 3-dekadowym (miesięcznym).

Udziały ich w wyjaśnianiu łącznej wariancji sezonowej wynoszą odpowiednio: 42,36 i 36,20%. Kilka słów komentarza wymaga sprawa występowania wahań o cyklu 3-dekadowym (miesięcznym). Ich występowanie jest związane z dopływem lub brakiem dopływu środków na rachunki *a vista*. W dekadzie pierwszej przeważają wpływy związane z wypłatą wynagrodzeń, czyli występuje dodatnie odchylenie od trendu. Natomiast w dekadzie drugiej przeważają wypłaty związane z koniecznością regulowania comiesięcznych opłat za mieszkanie, telefony oraz związane z bieżącymi potrzebami depozytariuszy. Z kolei pod koniec dekady

trzeciej na konta części ich właścicieli przelewane są wynagrodzenia za pracę, zachodzi więc także dodatnie odchylenie od trendu.



Rys. 1. Kształtowanie się dekadowych stanów depozytów na ROR-ach

Źródło: opracowanie własne.

Tabela 1. Struktura harmoniczna zmiennej Y (w %)

Składowe sinusoidalne	Udział w wyjaśn. war. sezonowej	Składowe kosinusoidalne	Udział w wyjaśn. war. sezonowej	Udziały harmonik
S1TS1	36,98	C1TC1	5,36	42,34
S2TS2	6,16	C2TC2	0,33	6,49
S3TS3	0,42	C3TC3	0,59	1,01
S4TS4	0,24	C4TC4	1,06	1,3
S5TS5	0,81	C5TC5	0,11	0,92
S6TS6	0,13	C6TC6	0,12	0,26
S7TS7	0,72	C7TC7	0,74	1,46
S8TS8	0,76	C8TC8	0,58	1,34
S9TS9	0,52	C9TC9	0,44	0,97
S10TS10	1,4	C10TC10	0,53	1,93
S11TS11	2,17	C11TC11	0,74	2,91
S12TS12	20,56	C12TC12	15,64	36,2
S13TS13	0,52	C13TC13	0,11	0,63
S14TS14	0,09	C14TC14	0,17	0,26
S15TS15	0,65	C15TC15	0,07	0,71
S16TS16	0,02	C16TC16	0,86	0,88
S17TS17	0,11	C17TC17	0,29	0,4
		C18TC18	0,05	0,05

Źródło: obliczenia własne.

Zawarte w kolumnach pierwszej i trzeciej symbole SiTSi oraz CiTCi oznaczają składowe sinuso- i cosinusoidalne dla zmiennej sezonowości. Przykładowo SiTSi oznacza: $\sin \omega_1 t$ i $t \sin \omega_1 t$.

Odsetkami powyżej 5% charakteryzuje się jeszcze harmonika druga o cyklu półrocznym. Te 3 harmoniki wyjaśniają prawie 85% zmienności sezonowej. Ponadto jeszcze odsetki dla 6 harmonik przekraczają 1%.

Z punktu widzenia celu badań najistotniejszy jest dominujący wpływ wymienionych wyżej 2 harmonik. Jego konsekwencją jest to, że rozpatrywane będą 2 wersje modelu Holta–Wintersa różniące się długością cyklu wahań okresowych. W pierwszej wersji długość tego cyklu wynosi 36 dekad ($m = 36$), a w drugiej 3 dekady ($m = 3$).

Najpierw przeprowadzone zostaną obliczenia dla zmiennej bez luk obejmujące 144 dekady. Prognozy zostały wyznaczone na kolejne 9 dekad. Te 9 dekad stanowiło okres empirycznej weryfikacji prognoz. Dla każdej wersji szacowanych było $9^3 = 729$ modeli różniących się kombinacjami parametrów wygładzania. Dla każdej wersji wybrano 25 modeli charakteryzujących się najniższymi ocenami średnich względnych błędów wartości wyrównanych – $\text{MAPE}^{\alpha, \beta, \gamma}$. Uwzględnienie stosunkowo dużej liczby kombinacji związane było z tym, że oceny $\text{MAPE}^{\alpha, \beta, \gamma}$ różniły się nie więcej niż o 0,2-0,3 punktu procentowego. Stąd też wybór mniejszej liczby modeli, a zwłaszcza jednego, byłby zbyt arbitralny. Ponadto, podobnie jak w przypadku modelowania klasycznego, nie zawsze model o najlepszych własnościach predyktywnych będzie dawał najlepsze prognozy. Analiza mierników dokładności wartości wyrównanych oraz prognoz *ex post* pozwala ocenić, czy oczekiwana dokładność pokrywała się z rzeczywistością. Ponadto w przypadku rozpatrywania szeregow z lukami w danych będzie można stwierdzić, czy najniższe błędy względne prognoz interpolacyjnych i ekstrapolacyjnych odpowiadają tym samym kombinacjom parametrów α, β, γ .

Tabela 2. Błędy względne dla wartości wyrównanych i średnie błędy względne prognoz *ex post* dla szeregów bez luk – wersje P i PP

Model	$\text{MAPE}^{\alpha, \beta, \gamma}$	Śr. błąd wzgl.	Model	$\text{MAPE}^{\alpha, \beta, \gamma}$	Śr. błąd wzgl.
P219	0,1201	0,0437	PP324	0,0945	0,0928
P127	0,1258	0,0450	PP424	0,0957	0,0937
P117	0,1257	0,0486	PP523	0,0992	0,0944
P218	0,1203	0,0528	PP425	0,0958	0,0952
P118	0,1248	0,0556	PP524	0,0969	0,0963
P128	0,1236	0,0590	PP525	0,0966	0,0978
P217	0,1235	0,0622	PP535	0,0991	0,0979
P129	0,1242	0,0800	PP416	0,0941	0,0990
P529	0,1430	0,0827	PP413	0,0942	0,0990
P229	0,1233	0,0869	PP513	0,0967	0,0995

Źródło: obliczenia własne.

Dla 25 modeli obliczono następnie średnie błędy względne prognoz *ex post*. Wyniki obliczeń dla 10 modeli o najniższych ocenach tych błędów zestawiono w tab. 2. Wersję pierwszą oznaczono literą P ($m = 36$), wersję drugą literami PP ($m = 3$). Dla każdej kombinacji podane zostały liczby dziesiętne odpowiadającym wartościom stałych wygładzania – np. P421 oznacza, że $\alpha = 0,4$, $\beta = 0,2$, $\gamma = 0,1$.

Z tabeli wynika, że niższe o ok. 3 punkty procentowe oceny parametrów $MAPE^{\alpha, \beta, \gamma}$ odpowiadają wersji z wahaniami 3-dekadowymi (PP). Wyjątkiem jest model P529, dla którego różnica ta wynosi prawie 5 punktów procentowych.

Natomiast dla pierwszych 7 modeli wersji i 36-dekadowej otrzymano zdecydowanie niższe oceny średnich względnych prognoz *ex post*. Oceny te są zazwyczaj o 3-5 punktów procentowych niższe od ocen otrzymanych dla wersji drugiej. Dla 3 ostatnich modeli różnice te nie przekraczają 2 punktów procentowych. Widoczne jest zdecydowanie większe zróżnicowanie tych błędów dla wersji pierwszej wynoszące aż 4,32 punktu procentowego. Różnice ocen $MAPE^{\alpha, \beta, \gamma}$ nie przekraczały 2,3 punktu procentowego, a w przypadku wyłączenia modelu P529 tylko 0,54 punktu. W przypadku wersji PP oceny obydwu parametrów są zbliżone i jednocześnie bardzo stabilne i kształtują się w granicach 9,28-9,95%. Natomiast w ramach tych samych modeli nie przekraczają one 0,48 punktu procentowego. Oznacza to, że spodziewana dokładność prognoz pokrywała się z rzeczywistością.

Dla wersji pierwszej rzeczywistość dokładność prognoz była zdecydowanie wyższa od spodziewanej. Oceny przeciętnych względnych błędów prognoz *ex post* były przynajmniej o 4 punkty procentowe niższe od błędów dla wartości wyrównanych.

Obecnie przechodzimy do omówienia wyników prognozowania inter- i ekstrapolacyjnego otrzymanych dla szeregu z lukami w danych. Rozpatrzony zostanie jeden wariant niesystematycznych luk w danych. Luki będą występować na przemian co 3 dekady.

Ze względu na warunek dotyczący minimalnej liczby obserwacji początkowych dla $m = 36$ wynoszący 37 przyjęto założenie, że luki występują od $t = 40$, 41 i 42. Jednocześnie założono także, że dostępna jest wartość zmiennej w ostatniej, 144. dekadzie przedziału czasowego „próby”. Wiązało się to z wcześniejszymi doświadczeniami dotyczącymi prognozowania zmiennych ekonomicznych dla danych miesięcznych. Brak ostatniej obserwacji wpływał nie tylko na wydłużenie okresu prognozowanego ale przede wszystkim na zmniejszenie dokładności prognoz zarówno inter- jak i ekstrapolacyjnych.

Przyjęto następujące oznaczenia wersji modeli dla szeregów z lukami w danych:

A – luki występują na przemian co 3 dekady od $t = 40$, 41, 42, a okres wahań periodycznych wynosi 36 dekad ($m = 36$),

B – luki występują na przemian co 3 dekady od $t = 40$, 41, 42, a okres wahań periodycznych wynosi 3 dekady ($m = 3$),

C – luki występują na przemian co 3 dekady od $t = 10$, 11, 12, a okres wahań periodycznych wynosi 3 dekady ($m = 3$).

Włączenie wersji C wynikało z faktu, że dla $m = 3$ minimalna liczba obserwacji na początku szeregu wynosi 4. Rozpatrzenie tej wersji pozwoli na zbadanie wpływu na dokładność prognoz inter- i ekstrapolacyjnych zwiększenia liczby luk z 51 do 67.

W tabeli 3 zestawiono błędy względne wartości wyrównanych ($\text{MAPE}^{\alpha, \beta, \gamma}$) i prognoz *ex post* (ekstrapolacyjnych). Wersja A ($m = 36$) charakteryzuje się zarówno najwyższymi ocenami błędów $\text{MAPE}^{\alpha, \beta, \gamma}$, jak i ocenami błędów prognoz *ex post*. Oznacza to, że w odróżnieniu od szeregu pełnego, dla cyklu 36-dekadowego otrzymano mniej dokładne prognozy niż w przypadku cyklu 3-dekadowego. Różnice te wynoszą przynajmniej 3 punkty procentowe.

Tablica 3. Średnie błędy względne dla wartości wyrównanych i średnie błędy względne prognoz *ex post* dla wersji A, B i C z lukami w danych

Model	$\text{MAPE}^{\alpha, \beta, \gamma}$	Śr. błąd wzgl.	Model	$\text{MAPE}^{\alpha, \beta, \gamma}$	Śr. błąd wzgl.	Model	$\text{MAPE}^{\alpha, \beta, \gamma}$	Śr. błąd wzgl.
A439	15,61	10,57	B323	12,86	8,42	C425	10,85	7,94
A519	15,46	10,69	B324	12,58	8,58	C626	11,46	8,07
A429	15,05	11,63	B515	12,41	9,03	C628	11,29	8,32
A619	16,06	11,91	B617	12,67	9,16	C324	10,59	8,58
A529	15,94	12,11	B516	12,44	9,17	C716	11,46	8,96
A641	18,56	12,15	B414	12,31	9,26	C516	10,41	9,17
A437	16,36	12,27	B829	14,83	9,29	C618	10,59	9,24
A318	14,58	13,68	B718	13,3	9,3	C517	10,43	9,25
A719	17,93	14,17	B415	12,27	9,4	C518	10,51	9,3
A319	14,08	15,07	B416	12,38	9,47	C719	10,98	9,38

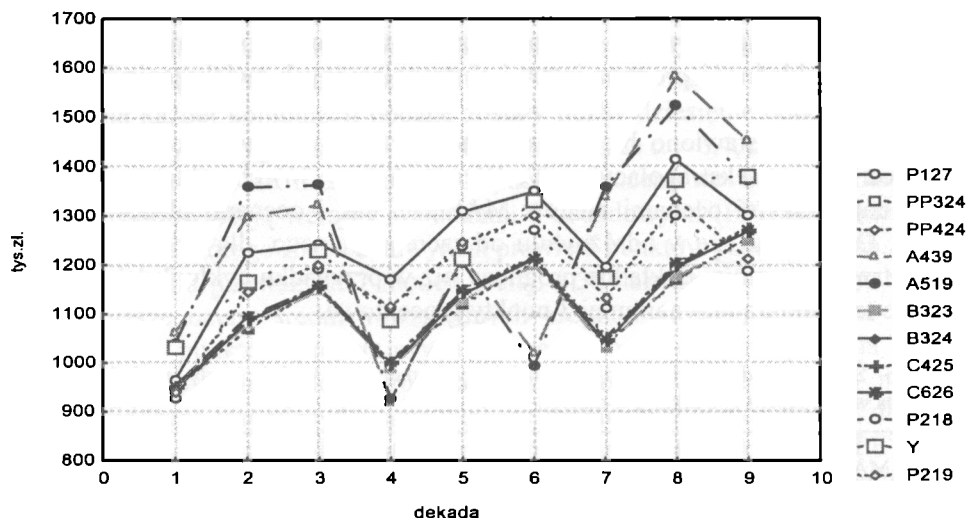
Źródło: obliczenia własne.

Cyfry występujące po literach A, B lub C odnoszą się do miejsc dziesiętnych stałych wygładzania. Przykładowo A439 oznacza: $\alpha = 0,4$, $\beta = 0,3$, $\gamma = 0,9$.

Rzeczywista dokładność prognoz w każdym przypadku była wyższa od oczekiwanej mierzonej ocenami $\text{MAPE}^{\alpha, \beta, \gamma}$.

Oceny błędów prognoz *ex post* dla wariantu B były nawet nieco niższe od analogicznych błędów otrzymanych dla pełnego szeregu (wersja PP). Dla wariantu C zwiększenie liczb luk wpłynęło pozytywnie do dokładności ocen $\text{MAPE}^{\alpha, \beta, \gamma}$ – były one zazwyczaj o ok. 2 punkty procentowe niższe. Dla 3 pierwszych kombinacji parametrów otrzymano także niższe błędy względne prognoz *ex post*.

Kształtowanie się prognoz ekstrapolacyjnych otrzymanych na podstawie 2 najlepszych predyktorów dla każdej wersji oraz realizacji zmiennej Y w okresie empirycznej weryfikacji prognoz przedstawione zostało w sposób graficzny na rys. 2.



Rys. 2. Prognozy ekstrapolacyjne i realizacje dekadowych stanów depozytów (Y)
Źródło: opracowanie własne.

W przypadku prognoz interpolacyjnych także wybierano 10 spośród 25 modeli charakteryzujących się najniższymi ocenami przeciętnych błędów względnych. Ich oceny zostały zestawione w tab. 4. Półgrubym drukiem zaznaczone zostały te modele, które znalazły się w dziesiętce modeli o najniższych ocenach błędów prognoz ekstrapolacyjnych. Nastąpiło to tylko w trzech przypadkach: A318, A319 i C517. Oznacza to, że na ogół inne modele były „optymalne” dla prognoz ekstrapolacyjnych i inne dla prognoz interpolacyjnych.

Tabela 4. Średnie błędy względne prognoz interpolacyjnych dla wersji A, B i C (w %)

Model	Błąd wzgl.	Model	Błąd wzgl.	Model	Błąd wzgl.
A239	17,77	B214	7,61	C126	9,74
A159	17,81	B215	7,64	C216	9,81
A148	17,92	B315	7,67	C316	9,91
A49	17,94	B216	7,74	C215	9,93
A229	17,94	B316	7,76	C226	9,94
A228	18,06	B125	7,78	C318	9,96
A319	18,51	B213	7,80	C315	10,00
A318	18,57	B126	7,84	C225	10,05
A329	18,57	B225	7,88	C517	10,07
A138	18,62	B116	8,00	C418	10,08

Źródło: obliczenia własne.

Najwyższymi oceny błędów prognoz interpolacyjnych były w wersji A luk w danych, a najniższe – w wersji B. Zwiększenie liczby luk w danych (wersja C) skutkowało zmniejszeniem dokładności prognoz interpolacyjnych, przeciętnie o ok. 2 punkty procentowe.

Zachodzą jednak zasadnicze różnice w ich dokładności. Zdecydowanie najmniej dokładne są prognozy interpolacyjne dla wersji A – są one o ok. 10 punktów procentowych wyższe, niż dla wersji B oraz o ok. 8 punktów procentowych wyższe niż dla wersji C.

4. Podsumowanie

Z przeprowadzonych w pracy badań empirycznych wynika, że w przypadku prognozowania w warunkach braku pełnej informacji zmiennej z wahaniami sezonowymi dla danych dekadowych na dokładność prognoz wpływa struktura harmoniczna. Mimo wyższego o kilka procent udziału harmoniki o okresie 36-dekadowym, wyższą dokładnością charakteryzowały się prognozy (zarówno inter- jak i ekstrapolacyjne) otrzymane dla cyklu 3-dekadowego. Wydłużenie dla $m = 3$ okresu występowania luk o 34 dekady wpłynęło pozytywnie na dokładność prognoz ekstrapolacyjnych. Prognozy interpolacyjne charakteryzowały się nieco niższą dokładnością.

Przeprowadzone badanie wykazało przydatność tej klasy modeli do prognozowania brakujących danych dekadowych zmiennych wykazujących wahania periodyczne.

Jednak w wyprowadzaniu wniosków uogólniających należy wykazać daleko idącą ostrożność. Uogólnianie będzie możliwe po przeprowadzeniu szeroko zakrojonych badań obejmujących m.in. większą liczbę zmiennych, różne warianty luk w danych oraz szeregi czasowe o zróżnicowanej długości.

Literatura

- [1] Chatfield C., *The Holt-Winters Forecasting Procedure*, „Applied Statistics” 1978, 27 (3), 264-279.
- [2] Reid D.J., *A Review of Short-term Projection Techniques*, [w:] *Practical Aspects of Forecasting*, red. H. Gordon, London 1975.
- [3] Winters P.R., *Forecasting Sales by Exponentially Weighted Moving Averages*, „Management Science” 1960.
- [4] Zawadzki J., *O wykorzystaniu modeli adaptacyjnych w prognozowaniu brakujących danych*, PS, Szczecin 2003.
- [5] Zawadzki J., *Modelowanie predyktywne i prognozowanie zjawisk w skali mikroekonomicznej*, US, Szczecin 1996.

THE APPLICATION OF EXPONENTIAL SMOOTHING MODELS IN FORECASTING OF THE DECADE VARIABLES WITH MISSING DATA

Summary

The paper presents an extended analysis of application of Holt–Winters' exponential smoothing models in forecasting the variables with seasonal fluctuations with missing data. The paper proves that this class of models can be used in forecasting when there is $m + 1$ observations at the beginning of time series (where m is the length of seasonal cycle).

The theoretical considerations were illustrated by empirical example referring to decade state of the *a vista* deposits.