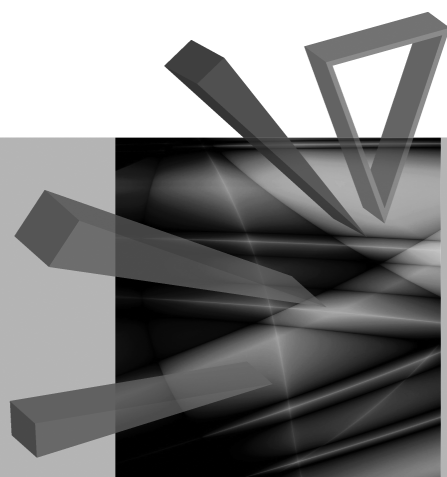


Prognozowanie w zarządzaniu firmą



Redaktorzy naukowi
Paweł Dittmann
Aleksandra Szpulak



Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu
Wrocław 2011

Senacka Komisja Wydawnicza

Zdzisław Pisz (przewodniczący),

Andrzej Bąk, Krzysztof Jajuga, Andrzej Matysiak, Waldemar Podgórski,

Mieczysław Przybyła, Aniela Styś, Stanisław Urban

Recenzenci

Włodzimierz Szkutnik, Jan Zawadzki

Redakcja wydawnicza

Barbara Majewska

Redakcja techniczna i korekta

Barbara Łopusiewicz

Skład i łamanie

Comp-rajt

Projekt okładki

Beata Dębska

Kopiowanie i powielanie w jakiegokolwiek formie
wymaga pisemnej zgody Wydawcy

© Copyright by Uniwersytet Ekonomiczny we Wrocławiu
Wrocław 2011

ISSN 1899-3192

ISBN 978-83-7695-141-6

Druk: Drukarnia TOTEM

Spis treści

Wstęp	7
Agnieszka Przybylska-Mazur: Optymalne zasady polityki pieniężnej w prognozowaniu wskaźnika inflacji	9
Alicja Wolny-Dominiak: Zmodyfikowana regresja Poissona dla danych ubezpieczeniowych z dużą liczbą zer	21
Andrzej Gajda: Doświadczenia i metody pozyskiwania danych eksperckich na potrzeby badań z wykorzystaniem metod foresight	30
Anna Gondek: Prognozy rozwoju gospodarczego Polski z użyciem metody analogii przestrzenno-czasowych	41
Bartosz Lawędziak: Sekurytyzacja papierów wartościowych opartych na hipotece odwrotnej	50
Filip Chybalski: Prakseologiczne aspekty prognozowania	59
Ireneusz Kuroпка, Paweł Lenczewski: Możliwość zastosowania modeli ekonometrycznych do prognozowania w przedsiębiorstwie Brenntag Polska	69
Jacek Szandula: Wyszukiwanie formacji w kursach giełdowych przy użyciu metod klasyfikacji danych	82
Joanna Perzyńska: Zastosowanie sztucznych sieci neuronowych do wyznaczania nieliniowych prognoz kombinowanych	94
Konstancja Poradowska, Tomasz Szkutnik, Mirosław Wójciak: Scenariusze rozwoju wybranych technologii oszczędności energii w życiu codziennym	102
Maciej Oesterreich: Wykorzystanie pakietu statystycznego R w prognozowaniu na podstawie danych w postaci szeregów czasowych z wahaniami sezonowymi	113
Marcin Błażejowski, Paweł Kufel, Tadeusz Kufel: Algorytm zgodnego modelowania i prognozowania procesów ekonomicznych jako pakiet funkcji <i>Congruent Specification</i> programu Gretl	125
Marcin Błażejowski: Stacjonarność szeregów czasowych o wysokiej częstotliwości obserwowania – implementacja testu stacjonarności Dickeya w programie Gretl	137
Mirosław Wójciak: Wpływ czynników i zdarzeń kluczowych na rozwój nowych technologii – wybrane metody korygowania prognoz na przykładzie technologii energooszczędnych	149
Monika Dyduch: Grupowanie produktów strukturyzowanych	159
Piotr Bernat: Planowanie działalności przedsiębiorstwa wspomagane prognozowaniem	170

Roman Pawlukowicz: Informacje prognostyczne w rynkowych sposobach wyceny nieruchomości – identyfikacja i pozyskiwanie	182
Wojciech Zatoń: Uwarunkowania psychologiczne w prognozowaniu	189

Summaries

Agnieszka Przybylska-Mazur: Optimal monetary policy rules in forecasting of inflation rate	20
Alicja Wolny-Dominiak: Zero-inflated Poisson Model for insurance data with a large number of zeros	29
Andrzej Gajda: Experience and methods of data collection from experts for research using foresight methods	40
Anna Gondek: Economic growth forecasts for Poland using the time-space analogy method	49
Bartosz Lawędziak: Securitization of survivor bonds based on the reverse mortgage	58
Filip Chybalski: Praxiological aspects of forecasting	68
Ireneusz Kuropka, Paweł Lenczewski: Econometric models usage feasibility in Brenntag Poland forecasting	81
Jacek Szandula: Searching for technical analysis formations in stock prices with the use of cluster analysis methods	93
Joanna Perzyńska: Application of artificial neural networks to build the nonlinear combined forecasts	101
Konstancja Poradowska, Tomasz Szkutnik, Mirosław Wójciak: The scenarios of development of selected technologies related to energy saving in everyday life	112
Maciej Oesterreich: The R application in forecasting unsystematic lacks in seasonal time series	124
Marcin Błażejowski, Paweł Kufel, Tadeusz Kufel: Congruent modelling and forecasting algorithm as function package Congruent Specification in GRETL	136
Marcin Błażejowski: Stationarity of high-frequency time series – implementation of Dickey’s stationarity test in GRETL	148
Mirosław Wójciak: The influence of key and events factors on the development of new technologies – selected methods of forecast correction on the example of energy-saving technologies	158
Monika Dyduch: Ranking of structured products	169
Piotr Bernat: Forecasting assisted business management planning	181
Roman Pawlukowicz: Prognostic data in market ways of property valuation – identification and acquisition	188
Wojciech Zatoń: Psychological aspects of forecasting	199

Maciej Oesterreich

Zachodniopomorski Uniwersytet Technologiczny w Szczecinie

WYKORZYSTANIE PAKIETU STATYSTYCZNEGO R W PROGNOZOWANIU NA PODSTAWIE DANYCH W POSTACI SZEREGÓW CZASOWYCH Z WAHANAMI SEZONOWYMI DLA LUK NIESYSTEMATYCZNYCH

Streszczenie: Celem pracy jest zastosowanie pakietu statystycznego R do prognozowania inter- i ekstrapolacyjnego w szeregach czasowych z wahaniami sezonowymi dla luk niesystematycznych. Pakiet ten, będący zintegrowanym środowiskiem do obliczeń statystycznych, cechuje się szybkością wykonywania procedur, elastycznością oraz modułowością. W obliczeniach zostały wykorzystane m.in. procedury: `lm()`, `stepAIC()`. Zastosowanie pakietu zostanie zilustrowane na przykładzie kształtowania się miesięcznych wielkości skupu mleka. Prognozy inter- i ekstrapolacyjne będą budowane na podstawie predyktorów opartych na klasycznych oraz hierarchicznych modelach szeregu czasowego z liniowym trendem i periodycznym składnikiem sezonowym.

Słowa kluczowe: R, luki niesystematyczne, brakujące dane, szeregi czasowe.

1. Wstęp

Problem kompletności danych jest jednym z trudniejszych problemów stojących przed badaczem analizującym dane w postaci szeregów czasowych. Występowanie luk może niekiedy uniemożliwić przeprowadzenie modelowania lub też zmniejszyć jakość uzyskiwanych wyników. Do prognozowania brakujących danych wykorzystywane są różne metody oraz narzędzia. Mają one oczywiście swoje wady i zalety. Przykładowo metody numeryczne nie mogą być wykorzystywane bezpośrednio w przypadku danych ze składnikiem sezonowym. Dodatkowo niektóre z nich wymagają występowania z góry określonej liczby obserwacji w szeregu, najczęściej odnoszących się do obserwacji początkowych.

W niniejszej pracy przedstawiono wykorzystanie pakietu R do prognozowania inter- oraz ekstrapolacyjnego dla danych w postaci szeregów czasowych, w których występują luki o charakterze niesystematycznym. Występowanie tego rodzaju

luk oznacza, że dostępne są przynajmniej pojedyncze informacje o kształtowaniu się badanego zjawiska w każdym z podokresów. Badanie zostało wykonane dla 4 wariantów luk i 9 typów modeli.

Wszystkie obliczenia przedstawione poniżej wykonano w bezpłatnym pakiecie statystycznym R, dostępnym pod adresem <http://cran.r-project.org/>. Został on wykorzystany przede wszystkim ze względu na ilość dostępnych funkcji oraz elastyczność i łatwość doboru procedur. W analizie wykorzystano moduły Stats oraz MASS. W tym celu zbudowano algorytm do automatycznego obliczania prognoz, zarówno inter-, jak i ekstrapolacyjnych w szeregach czasowych o minimalnej długości 24 obserwacji, z wykorzystaniem modeli z periodycznym składnikiem sezonowym.

1.1. Modele tendencji rozwojowej z periodycznym składnikiem sezonowym

Do obliczeń luk interpolacyjnych zostały wykorzystane klasyczne modele szeregu czasowego z sezonowością periodyczną opisane za pomocą:

- zmiennych zero-jedynkowych (TM12),
- wielomianu trygonometrycznego (W.tryg),

a także regularne modele hierarchiczne dwu- oraz trzypostopniowe dla danych miesięcznych.

Ogólny zapis modelu tendencji rozwojowej z periodycznym składnikiem sezonowym w postaci zmiennej 0/1 jest następujący [Zawadzki 2003]:

$$Y_t = \beta_1 t + \beta_0 + \sum_{k=1}^m d_{0k} Q_{kt} + U \quad (1)$$

z założeniem:

$$\sum_{k=1}^m d_{0k} = 0, \quad (2)$$

gdzie: β_1, β_0 – parametry trendu liniowego,

d_{0k} – wartość parametru w podokresie k ,

Q_{kt} – zmienna 0/1 przyjmująca wartość 1 w podokresie k i 0 w pozostałych podokresach.

Ogólny zapis modelu tendencji rozwojowej z periodycznym składnikiem sezonowym w postaci wahań harmonicznnych jest następujący:

$$Y_t = \beta_1 t + \beta_0 + \sum_{i=0}^{\frac{m}{2}} (a_{ji} \cos \omega_i t + b_{ji} \sin \omega_i t), \quad j = 0, 1, \dots, r, \quad (3)$$

gdzie:

$$\omega_i = \frac{2\pi \cdot i}{m}, i = 0, 1, \dots, \frac{m}{2}, \quad (4)$$

- β_1, β_0 – parametry trendu liniowego,
- I – numer harmoniki,
- M – długość cyklu,
- a_{ji}, b_{ji} – parametry opisujące wahania sezonowe.

W przypadku prognozowania brakujących danych za pomocą modelu tendencji rozwojowej ze składnikiem sezonowym postaci wielomianu trygonometrycznego mamy do czynienia z wyborem wyłącznie statystycznie istotnych wahań harmonicznych. Wyboru tego w tym przypadku dokonano za pomocą regresji krokowej. W efekcie stworzono modele oszczędne, w których ilość oraz oceny parametrów zależą w głównej mierze od ilości oraz rozmieszczenia luk.

W modelowaniu i prognozowaniu zmiennych ekonomicznych mogą być wykorzystywane także regularne modele hierarchiczne. Ogólny zapis dwustopniowego modelu hierarchicznego jest następujący [Zawadzki 2003]:

$$Y_{srt} = \alpha_1 t + \alpha_0 + \sum_{s=1}^{\frac{m}{p_1}} b_{0s} Q_{st} + \sum_{r=1}^{\frac{m}{p_2}} b_{0sr} Q_{str} + U_{srt}, \quad (5)$$

z założeniem:

$$\sum_s b_{0s} = \sum_r b_{0sr} = 0. \quad (6)$$

Ogólny zapis trzystopniowego modelu hierarchicznego jest następujący [Zawadzki 2003]:

$$Y_{srlt} = \alpha_1 t + \alpha_0 + \sum_{s=1}^{\frac{m}{p_1}} b_{0s} Q_{st} + \sum_{r=1}^{\frac{m}{p_2}} b_{0sr} Q_{str} + U_{2t} + \sum_{l=1}^{\frac{m}{p_3}} b_{0srl} Q_{srlt} + U_{srlt} \quad (7)$$

z założeniem:

$$\sum_{s=1}^{\frac{m}{p_1}} b_{0s} = \sum_{r=1}^{\frac{m}{p_1 p_2}} b_{0sr} = \sum_{l=1}^{\frac{m}{p_1 p_2 p_3}} b_{0srl} = 0, \quad (8)$$

$Q_{st}, Q_{str}, Q_{srlt}$ – zmienna 0/1,

p_1, p_2, p_3 – podzielniki odpowiadające kolejnym stopniom hierarchii.

W przypadku, gdy analizowane dane mają charakter miesięczny, liczba modeli hierarchicznych wynosi siedem, w tym cztery dwustopniowe oraz trzy trójstopniowe. Dokładne informacje przedstawiono w tabeli 1.

Tabela 1. Specyfikacja regularnych modeli hierarchicznych

Model	Czynnik pierwszy	Czynnik drugi	Czynnik trzeci	Liczba szacowanych parametrów
	rodzaj zmienności			
TH26	półrocze w roku	miesiąc w półroczu		6
TH34	kwartał w roku	miesiąc w kwartale		5
TH43	okres 4 miesięcy w roku	miesiąc w okresie czteromiesięcznym		5
TH62	okres 2 miesięcy w roku	miesiąc w okresie 2 miesięcy		6
TH232	półrocze w roku	2 miesiące w półroczu	miesiąc w okresie 2 miesięcy	4
TH223	półrocze w roku	kwartał w półroczu	miesiąc w kwartale	4
TH322	okres 4 miesięcy w roku	2 miesiące w okresie 4 miesięcy	miesiąc w okresie 2 miesięcy	4

Źródło: [Zawadzki 2003].

2. Pakiet statystyczny R w prognozowaniu brakujących danych

Pakiet R jest środowiskiem, które dzięki dostępności olbrzymiej ilości modułów jest na tyle elastyczne, że można je wykorzystywać nie tylko w ściśle pojętej statystyce, ale również w innych sferach nauki. Pierwsza wersja R została napisana przez Roberta Gentlemana i Ross Ihake pracujących na wydziale statystyki Uniwersytetu w Auckland. Obecnie jest rozwijany przez ludzi z całego świata, skupionych wokół fundacji The R Foundation for Statistical Computing. Dostępne są jego wersje nie tylko na platformę Windows, ale również na Linux oraz MacOS. Zarówno program, jak i pakiety w przeważającej większości udostępnione zostały za darmo, na licencji GNU GPL [Biecek 2008].

Do budowy algorytmu wykorzystano podstawowe funkcje pakietu R zawarte w pakiecie Stats, a także `stepAIC()` z pakietu MASS. Zamiennie można również użyć funkcji `step()` dostępnej w pakiecie Stats

Do budowy poszczególnych 0/1 zmiennych objaśniających oraz harmonik wykorzystano polecenie `matrix()`. Następnie łączono je w jeden zbiór za pomocą polecenia `data.frame()`. Na jego podstawie budowano funkcje regresji z użyciem polecenia `lm()`. W przypadku modelu z wielomianem trygonometrycznym dodatkowo użyto regresji krokowej za pomocą funkcji `stepAIC()`, aby wyodrębnić jedynie istotne statystyczne parametry (tabela 2).

Tabela 2. Procedura budowy modeli W.tryg oraz TH223 w języku R

W.tryg	TH223
<pre>#zmienne objaśniające s <- 1:12 s1 <- matrix(sin((2*pi*s)/12), length(t), 1) s2 <- matrix(sin((4*pi*s)/12), length(t), 1) s3 <- matrix(sin((6*pi*s)/12), length(t), 1) s4 <- matrix(sin((8*pi*s)/12), length(t), 1) s5 <- matrix(sin((10*pi*s)/12), length(t), 1) c1 <- matrix(cos((2*pi*s)/12), length(t), 1) c2 <- matrix(cos((4*pi*s)/12), length(t), 1) c3 <- matrix(cos((6*pi*s)/12), length(t), 1) c4 <- matrix(cos((8*pi*s)/12), length(t), 1) c5 <- matrix(cos((10*pi*s)/12), length(t), 1) c6 <- matrix(cos((12*pi*s)/12), length(t), 1) data.x <- data.frame(t, s1, s2, s3, s4, s5, c1, c2, c3, c4, c5, c6) #funkcja regresji model.lm <- lm(Yt ~., data = data.x) model.lm <- stepAIC(model.lm, trace=F)</pre>	<pre>#zmienne objaśniające in6.12 <- matrix(c(1,1,1,1,1,1,-1,-1,-1,-1,-1), length(t), 1) in3.6.1 <- matrix(c(1,1,1,-1,-1), length(t), 1) in1.3 <- matrix(c(1,0,-1), length(t), 1) in2.3 <- matrix(c(0,1,-1), length(t), 1) data.x.th223 <- data.frame(t, in6.12, in3.6.1, in1.3, in2.3) #funkcja regresji model.lm.th223 <- lm(Yt ~., data = data.x.TH223)</pre>

Źródło: opracowanie własne na podstawie <http://cran.r-project.org/>.

Na podstawie oszacowanych modeli wyznaczono prognozy interpolacyjne dla miesięcy, w których występują luki w danych, oraz ekstrapolacyjne dla 12 kolejnych okresów. Wykorzystano do tego celu funkcje predict() pakietu Stats (tabela 3).

Tabela 3. Procedura obliczania prognoz dla modeli W.tryg oraz TH223 w języku R

W.tryg	TH223
<pre>data.x.fc <- data.frame(t=(length(t)+1):(length(t)+12), s1, s2, s3, s4, s5, c1, c2, c3, c4, c5, c6) model.lm.fc <- predict(model.lm, int="p", newdata=data.x.fc[1:12,])</pre>	<pre>data.x.fc <- data.frame(t=(length(t)+1):(length(t)+12), in6.12, in3.6.1, in1.3, in2.3) model.lm.th223.fc <- predict(model.lm.th223, int="p", newdata=data.x.fc[1:12,])</pre>

Źródło: opracowanie własne na podstawie <http://cran.r-project.org/>.

Kolejnym etapem było zliczenie luk interpolacyjnych, weryfikacja prognoz, obliczenie średnich względnych błędów prognoz. Liczbę luk w danym szeregu obliczono, wykorzystując prostą pętlę for(). W celu obliczenia średnich względnych błędów prognoz zbudowano funkcje error(), która umożliwia wyznaczenie błędów zarówno dla prognoz inter-, jak i ekstrapolacyjnych o zadanym horyzoncie (tab. 4).

Tabela 4. Procedura zliczania luk w szeregu, funkcja error oraz średnie względne błędy inter- oraz ekstrapolacyjne dla modeli W.tryg oraz TH223

Procedura zliczania luk w szeregu	
<pre>na.c <- 0 for(a in 1:length(t)) { if(is.na(Yt[[a]]) == TRUE) { na.c <- na.c + 1}}</pre>	
Procedura obliczania średnich względnych błędów prognoz inter- i ekstrapolacyjnych	
<pre>error <- function(real, teo, okr) { if(okr == 0) { b.i <- sum((abs(real[1:length(t)] - teo))/real[1:length(t)])/(na.c) b.i <- round(b.i, digits = 5) return(b.i)} if(okr > 0) { b.e <- sum((abs(real[(length(t)+1):(length(t)+okr)] - teo[1:okr]))/real[(length(t)+1):(length(t)+okr)])/(okr) b.e <- round(b.e, digits = 5) return(b.e)}}</pre>	
W.tryg	TH223
<pre># błędy interpolacyjne bi.lm <- error(Y, model.wl[,1], 0) # błędy ekstrapolacyjne h=3 be.lm.h3 <- error(Y, model.lm.fc[,1], 3) # błędy ekstrapolacyjne h=6 be.lm.h6 <- error(Y, model.lm.fc[,1], 6) # błędy ekstrapolacyjne h=9 be.lm.h9 <- error(Y, model.lm.fc[,1], 9) # błędy ekstrapolacyjne h=12 be.lm.h12 <- error(Y, model.lm.fc[,1], 12)</pre>	<pre># błędy interpolacyjne bi.th223 <- error(Y, model.wl[,8], 0) # błędy ekstrapolacyjne h=3 be.lm.th223.h3 <- error(Y, model.lm.th223.fc[,1], 3) # błędy ekstrapolacyjne h=6 be.lm.th223.h6 <- error(Y, model.lm.th223.fc[,1], 6) # błędy ekstrapolacyjne h=9 be.lm.th223.h9 <- error(Y, model.lm.th223.fc[,1], 9) # błędy ekstrapolacyjne h=12 be.lm.th223.h12 <- error(Y, model.lm.th223.fc[,1], 12)</pre>

Źródło: opracowanie własne na podstawie <http://cran.r-project.org/>.

Kończącym etapem działania programu jest wydruk, na którym zawarto średnie względne błędy prognoz dla poszczególnych modeli (tab. 5).

Tabela 5. Średnie względne błędy prognoz inter- oraz ekstrapolacyjnej dla wariantu I

	B.Int.	B.Eks.h=3	B.Eks.h=6	B.Eks.h=9	B.Eks.h=12
W.tryg	4.328	11.223	9.969	7.193	6.086
TM12	2.593	8.787	7.718	5.640	4.607
TH26	8.532	4.233	8.305	8.142	8.133
TH34	6.194	12.560	10.941	8.962	7.325
TH43	4.213	5.275	6.258	5.135	4.862
TH62	3.210	7.466	6.716	5.051	4.378
TH232	8.566	3.695	7.910	7.962	8.082
TH223	8.471	3.488	8.081	8.093	7.993
TH322	4.186	5.478	6.651	5.506	5.185

Źródło: opracowanie własne.

W tabeli 6 zestawione wszystkie funkcje, których użyto do zbudowania programu.

Tabela 6. Predefiniowane procedury wykorzystane do budowy algorytmu

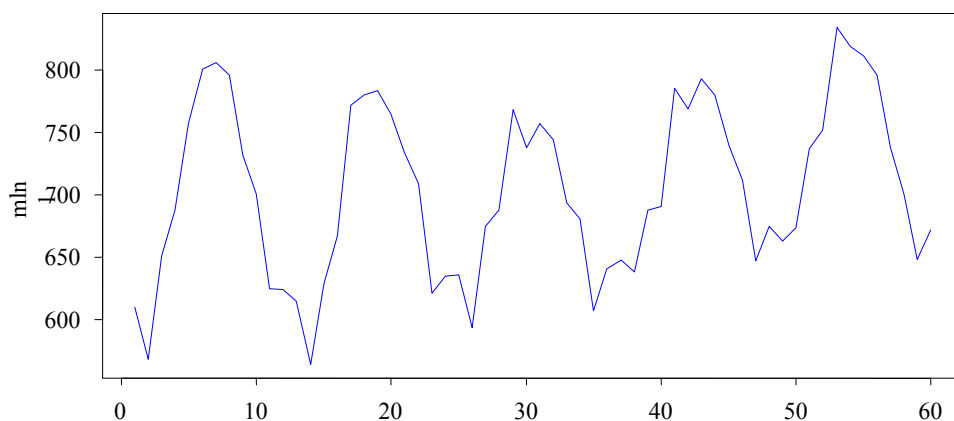
Polecenie	Pakiet	Opis
read.table()	utils	wczytuje dane w formie tabeli do określonego zbioru
require()	base	wczytuje dodatkowy pakiet, który będzie używany do analizy
matrix()	base	tworzy macierz o zadanych parametrach
data.frame()	base	tworzy jeden zbiór z kilku macierzy
length()	base	zwraca długość [ilość elementów] zmiennej
for()	base	tworzy pętlę o zadanej ilości powtórzeń
if)...Ele	base	polecenie warunkowe
lm()	stats	oblicza parametry modelu liniowego na podstawie zadanych zmiennych
stepAIC() / step()	MASS / stats	regresja krokowa oparta na kryterium informacyjnym akaikie
predict()	stats	oblicza wartości teoretyczne na podstawie obiektu zawierającego model
function()	base	tworzy funkcję zdefiniowaną przez użytkownika
is.na()	base	Sprawdza, czy dany element jest zbiorem pustym (NA)

Źródło: opracowanie własne na podstawie <http://cran.r-project.org/>.

Działanie programu zostało zaprezentowane na poniższym przykładzie.

3. Przykład empiryczny

Analiza została wykonana na przykładzie kształtowania się skupu mleka krowiego w Polsce. Użyto w tym celu danych pochodzących z Biuletynów Statystycznych Głównego Urzędu Statystycznego z lat 2005-2009. Dane z lat 2005-2008 zostały wykorzystane w prognozowaniu luk interpolacyjnych. Rok 2009 został wykorzystany do empirycznej weryfikacji prognoz ekstrapolacyjnych. Kształtowanie się badanej zmiennej przedstawiono na rysunku 1.



Rys. 1. Skup mleka w Polsce w latach 2005-2009

Źródło: opracowanie własne na podstawie Biuletynów Statystycznych GUS.

Cechą charakterystyczną badanej zmiennej są jej wyraźne wahania sezonowe. Maksimum natężenia w każdym roku przypada na okresy letnie, natomiast minima na okresy zimowe. Ma to swoje odzwierciedlenie w charakterze i cyklu chowu bydła mlecznego. Zauważalna jest także tendencja wzrostowa badanej zmiennej.

3.1. Warianty luk w danych

Badanie wykonano dla następujących rozmieszczeń luk niesystematycznych (X oznacza lukę):

Wariant I

Lata	1	2	3	4	5	6	7	8	9	10	11	12
2005							×		×		×	
2006	×		×		×		×		×		×	
2007	×		×		×		×		×		×	
2008	×		×		×							

Źródło: opracowanie własne.

Wariant II

Lata	1	2	3	4	5	6	7	8	9	10	11	12
2005		×		×		×	×		×		×	
2006	×		×		×		×		×		×	
2007	×		×		×		×		×		×	
2008	×		×		×			×		×		×

Źródło: opracowanie własne.

Wariant III

Lata	1	2	3	4	5	6	7	8	9	10	11	12
2005	×		×		×			×		×		×
2006	×		×		×			×		×		×
2007		×		×		×	×		×		×	
2008		×		×		×	×		×		×	

Źródło: opracowanie własne.

Wariant IV

Lata	1	2	3	4	5	6	7	8	9	10	11	12
2005	×						×					
2006		×				×		×				×
2007			×		×				×		×	
2008				×						×		

Źródło: opracowanie własne.

3.2. Wyniki analizy – prognozy interpolacyjne

W tabeli 7 zestawiono średnie względne błędy prognoz dla czterech wariantów luk interpolacyjnych.

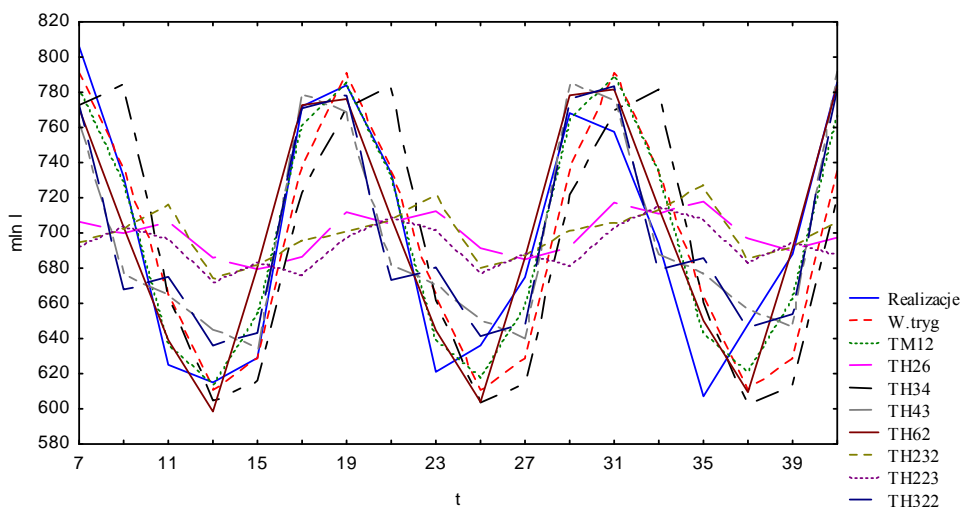
Tabela 7. Średnie względne błędy prognoz interpolacyjnych skupu mleka dla czterech wariantów luk (w %)

Wariant	W.tryg	TM12	TH26	TH34	TH43	TH62	TH232	TH223	TH322
I	4,328	2,593	8,532	6,194	4,213	3,210	8,566	8,471	4,186
II	4,263	2,870	8,551	5,919	4,492	2,962	8,584	8,462	4,449
III	3,628	3,428	8,389	5,717	4,585	2,653	8,329	8,367	4,513
IV	2,664	2,589	9,101	5,992	5,529	2,605	9,122	9,291	5,148

Źródło: opracowanie własne.

Średnie względne błędy prognoz interpolacyjnych zawierają się w przedziale od 2,589% (TM12, wariant I) do 9,291% (TH223, wariant IV). Najlepsze wyniki w prognozowaniu interpolacyjnym w wariantach I, II i IV osiągnięto po zastosowaniu modelu TM12. W wariantach II i III lepszy okazał się model hierarchiczny dwustopniowy TH62. Najgorsze prognozy, niezależnie od wariantu, generowały modele hierarchiczne TH26, TH232 oraz 223. Dawały one błędy wyższe o ok. 6 punktów procentowych od najlepszych modeli w danym wariantach.

Model zawierający wielomian trygonometryczny był we wszystkich wariantach gorszy od modelu niehierarchicznego zawierającego zmienną 0/1 oraz TH62. Kształtowanie się absolutnych odchyień prognoz interpolacyjnych od wartości rzeczywistych dla poszczególnych luk przedstawiono na rysunku 2.



Rys. 2. Prognozy interpolacyjne dla wariantu I

Źródło: opracowanie własne.

3.3. Wyniki analizy – prognozy ekstrapolacyjne

W tabeli 8 zestawiono średnie względne błędy prognoz ekstrapolacyjnych zbudowanych na podstawie szeregów, w których luki uzupełniono za pomocą modeli z wahaniami harmonicznymi.

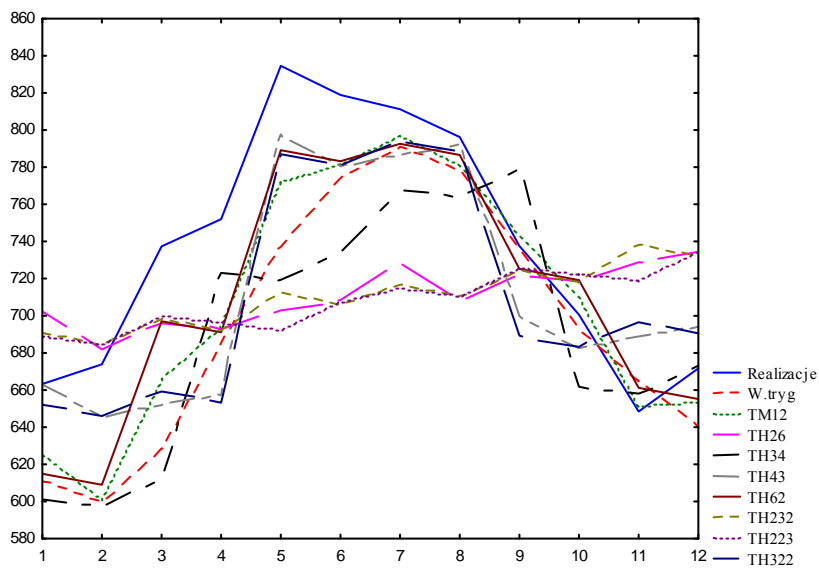
Tabela 8. Średnie względne błędy prognoz ekstrapolacyjnych skupu mleka dla czterech wariantów luk (w %)

Wariant	Horyzont prognozy	Model								
		W.tryg	TM12	TH26	TH34	TH43	TH62	TH232	TH223	TH322
I	h = 3	11,223	8,787	4,233	12,560	5,275	7,466	3,695	3,488	5,478
	h = 6	9,969	7,718	8,305	10,941	6,258	6,716	7,910	8,081	6,651
	h = 9	7,193	5,640	8,142	8,962	5,135	5,051	7,962	8,093	5,506
	h = 12	6,086	4,607	8,133	7,325	4,862	4,378	8,082	7,993	5,185
II	h = 3	10,312	9,035	4,431	13,221	5,192	6,684	4,092	4,064	4,844
	h = 6	9,804	8,335	8,575	11,941	6,143	6,678	8,376	8,539	6,335
	h = 9	7,362	6,110	8,163	9,646	5,009	4,855	8,048	8,246	5,316
	h = 12	6,255	5,030	8,102	8,157	4,799	4,294	8,038	8,060	5,032
III	h = 3	5,342	6,991	4,745	3,762	6,454	7,465	3,569	3,026	5,373
	h = 6	4,688	5,775	8,608	4,094	7,318	6,744	7,989	7,708	6,624
	h = 9	4,160	4,524	8,482	4,439	6,035	5,054	8,091	7,872	5,472
	h = 12	3,523	3,562	8,228	4,406	5,455	4,457	7,964	7,754	5,013
IV	h = 3	6,681	6,954	4,509	6,550	6,188	6,076	4,525	4,373	5,770
	h = 6	6,170	6,371	8,410	6,155	7,159	6,219	8,356	8,175	7,148
	h = 9	4,632	4,867	8,111	5,544	6,147	4,570	8,021	7,861	6,279
	h = 12	3,969	3,987	8,098	4,860	5,426	4,145	8,071	7,992	5,420

Źródło: opracowanie własne.

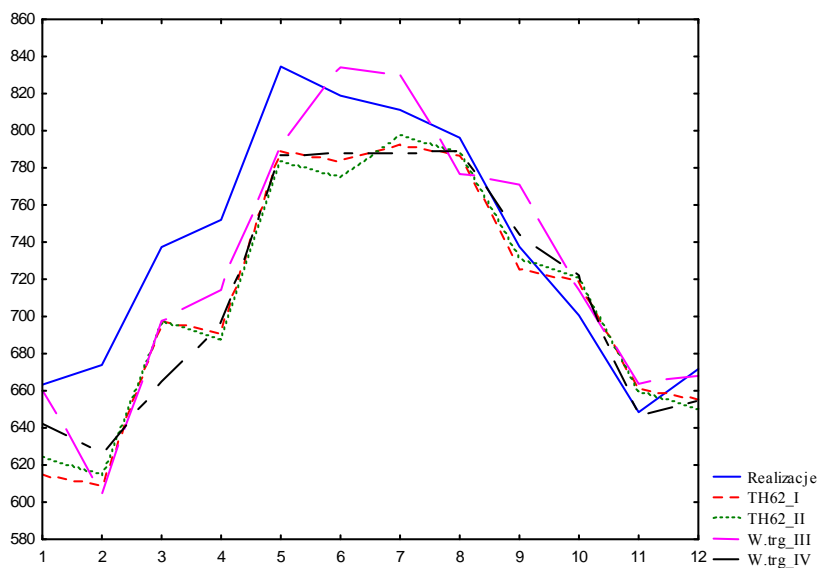
Średnie względne błędy prognoz ekstrapolacyjnych zawierają się w przedziale od 3,026% (TH223, wariant III) do 13,221% (TH34, wariant II). Najlepsze prognozy trzymiesięczne we wszystkich wariantach generował model TH223. W przypadku prognoz sześciomiesięcznych w dwóch pierwszych wariantach najlepszy okazał się model TH43, a w dwóch kolejnych TH34. Dla dziewięciomiesięcznego horyzontu jakościowo najlepsze były prognozy generowane przez model TH62 (wariant I, II, IV) oraz model zawierający wielomian trygonometryczny (wariant III). Również te same modele dały najlepsze prognozy na okres 1 roku – model TH62 w wariantach I i II, a model z wielomianem trygonometrycznym w wariantach III i IV.

Należy zauważyć, że w odróżnieniu od prognoz interpolacyjnych użycie modeli hierarchicznych wpłynęło pozytywnie na jakość prognoz ekstrapolacyjnych. Charakteryzowały się one czasem nawet 3-krotnie niższymi błędami niż modele niehierarchiczne. Nie stwierdzono natomiast znaczących różnic między jakością prognoz generowanych między modelami hierarchicznymi z dwoma i trzema zmiennymi. Kształtowanie się prognoz ekstrapolacyjnych w wariantach I przedstawiono na rysunku 3. Na rysunku 4 zestawiono najlepsze prognozy wygenerowane dla poszczególnych wariantów luk.



Rys. 3. Prognozy ekstrapolacyjne skupu mleka dla wariantu I

Źródło: opracowanie własne.



Rys. 4. Najlepsze prognozy ekstrapolacyjne dla poszczególnych wariantów

Źródło: opracowanie własne.

4. Podsumowanie

W artykule wskazano możliwość wykorzystania pakietu R w prognozowaniu brakujących danych o charakterze niesystematycznym w szeregach czasowych z sezonowością. Niewątpliwą zaletą przedstawionej procedury jest szybkość wieloetażowej analizy, obejmującej wybór składowych szeregu czasowego, estymacje parametrów, obliczanie prognoz inter- i ekstrapolacyjnych oraz błędów prognoz. Pewną niegodnością związaną z jego stosowaniem stanowi konieczność tworzenia od podstaw algorytmu na przykład w przypadku rozszerzenia procesu na luki systematyczne.

Duża liczba metod, które mogą być wykorzystane do budowy prognoz sprzyja umacnianiu się tendencji do „symbiozy” różnych narzędzi zawartych w różnych pakietach statystycznych. Przykładem jest moduł pakietu Statistica umożliwiający korzystanie z procedur środowiska R. Dzięki temu narzędzia takie jak R mogą być wykorzystywane przez coraz szerszą grupę użytkowników.

Literatura

- Biecek P., *Przewodnik po pakiecie R.*, <http://cran.r-project.org/doc/contrib/Biecek-R-basics.pdf> (6.07.2010).
Zawadzki J., *Zastosowanie hierarchicznych modeli szeregów czasowych w prognozowaniu zmiennych ekonomicznych z wahaniami sezonowymi*, Wydawnictwo Akademii Rolniczej w Szczecinie, Szczecin 2003.

THE R APPLICATION IN FORECASTING UNSYSTEMATIC LACKS IN SEASONAL TIME SERIES

Summary: This study presents the R application in forecasting unsystematic lacks in the seasonal time series. This integrated environment for statistical computing is characterized by speed of performance of procedures, flexibility and modularity. In calculations the following procedures are used: `lm()`, `stepAIC()` and others. The package application is illustrated on the example of the monthly evolution of the volume of buying milk. Forecasts will be constructed on classical and hierarchical models of time series with linear trend and periodic seasonal component.

Informacje dodatkowe

Skrypt na podstawie, którego powstał artykuł, można otrzymać drogą mailową (moesterreich@zut.edu.pl).