

Anna Szmit
Politechnika Łódzka

ANALIZA SEZONOWOŚCI WYBRANYCH PROCESÓW GOSPODARCZYCH

1. Wstęp

Dekompozycja szeregu czasowego, wyodrębniając poszczególne jego składowe, przypisuje określoną wagę i definiuje lub zakłada charakter wpływu trendu, wahań sezonowych, cyklicznych i losowych. Sposób uwzględniania sezonowości zjawiska może być różny i różnice te dotyczą nie tylko zakładanego kształtu przebiegu kolejnych sezonów czy sposobu nakładania się sezonowości na trend, ale także metody zapisu parametrów sezonowych dla konkretnych założeń co do przebiegu zjawiska. Nawet podstawowe narzędzia dostarczają niekiedy interesującego materiału do badań – niekiedy bowiem np. przyjmuje się w różnych metodach takie same założenia w stosunku do kompozycji szeregu, a dekompozycja zachodzi w nich w różny sposób. Celem artykułu jest porównanie wpływu sposobu, w jaki wybrane metody ujmują sezonowość zjawiska, na jakość dopasowania do danych empirycznych na kilku przykładach.

2. Metody uwzględniania sezonowości

Podstawową metodą uwzględniania sezonowości są wskaźniki sezonowości¹, które dodaje się do wygładzonych wartości szeregu czasowego (lub mnoży). Zakłada się w ten sposób, że szereg jest sumą (iloczynem) składowych: trendu (czasami też cyklu) i sezonowej. To samo założenie przyjmuje się w modelach regresji ze zmiennymi sezonowymi w postaci zero-jedynkowej, np. w metodzie Kleina dla

¹ Metody wyznaczania wskaźników sezonowości dla modeli addytywnych i multiplikatywnych zaprezentowano np. w pracy [Cieślak 1999, s. 88-90].

postaci addytywnej lub dla szeregu z sezonowością multiplikatywną, przyjmując postać (przy założeniu trendu liniowego):

$$\hat{y}_{it} = \sum_{i=1}^r \alpha_{0i} \cdot Q_i + \sum_{i=1}^r \alpha_i \cdot Q_i \cdot t = \sum_{i=1}^r \alpha_{0i} \cdot Q_i + \sum_{i=1}^r \alpha_i \cdot Q_i^i, \quad (1)$$

gdzie: $t=1, 2, \dots, n$ obserwacji, $i=1, 2, \dots, r$ sezonów w cyklu, $\tilde{y}_{it}$ jest wartością wygładzoną szeregu (trendem), Q_i są zmiennymi równymi 1 w i -tej fazie cyklu sezonowego i 0 poza nią, $Q_i^i = Q_i \cdot t$ albo, zamiast Q_i^i stosując zmienne², $Q_i^j = Q_i \cdot j$, gdzie $j=1, \dots, s$ jest numerem cyklu.

Warto zauważyć, że metoda wskaźników i model regresji nie odpowiadają sobie dokładnie. W tej pierwszej najpierw wyznacza się wartości wygładzone – trend, potem wskaźniki sezonowości, w tej drugiej obie składowe są wyodrębniane równocześnie. Druga metoda jest bardziej zgodna z założeniem o konstrukcji szeregu, gdyż np. na wstępnie wyodrębniany trend częściowo może wpływać kształt cyklu sezonowego.

Odrębnym problemem jest liczba parametrów modelu. Model addytywny Kleina (lub inne odpowiadające mu metody³) ma liczbę parametrów porównywalną do metody wskaźników. Model (1) dla $r > 2$ ma więcej parametrów przy podobnym sposobie opisu szeregu niż model liniowy z multiplikatywnymi wskaźnikami sezonowości. Uwzględnienie tylu parametrów zwłaszcza dla małych prób może być niekorzystne, a poza tym w niektórych przypadkach niekoniecznie poprawia stopień odzwierciedlenia badanego zjawiska. Niekiedy może wystarczyć uwzględnienie w modelu zmiennych Q_i^i lub $Q_i^{j,i}$, bez zmiennych zero-jedynkowych, tzn. wykorzystanie modeli:

$$\hat{y}_{it} = \alpha'_0 + \sum_{i=1}^r \alpha'_i \cdot Q_i^i \quad (2)$$

lub

$$\hat{y}_{it} = \gamma'_0 + \sum_{i=1}^r \gamma'_i \cdot Q_i^{i,i}. \quad (3)$$

W inny sposób ujmują sezonowość modele autoregresji, wykorzystujące dla takich zjawisk najczęściej (por. [Enders 1995, s. 112 i nast.]) wartości zmiennej objaśnianej opóźnione tylko o pełny cykl, tzn. o $r, 2r$ itd. sezonów⁴. Modele autoregresji nie zakładają żadnej konkretnej postaci trendu i dlatego są dość podatne na jego zmiany czy nieregularny charakter. W związku z tym dobrze opisują zjawiska, dla których kształt trendu jest trudny do zdefiniowania lub zmienia się dynamicznie, mogą też być pomocne, gdy zjawisko poddaje się pewnym cyklom, np. koniunkturalnym. Modele ze zmiennymi opóźnionymi są jednak bardziej zależne od bieżącej wartości składnika losowego, co może w niekorzystny sposób wpływać na opis szeregu.

² Por. [Łapińska-Sobczak 1998, s. 122-124].

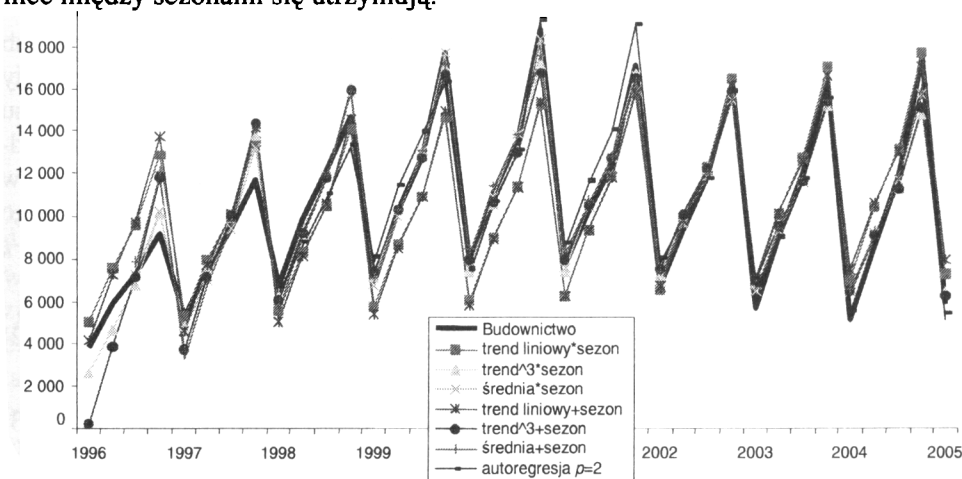
³ Przyjmuje się także [Zieliński 1967], że zmiennych zero-jedynkowych jest r , ale nie ma wyrazu wolnego. W innym podejściu (np. [Witkowska 2002, s. 82-83]) Q_i dla brakującego r -tego podokresu są równe -1 .

⁴ Istnieją oczywiście także inne metody, jak choćby modele ARMA, metoda jednoimiennych trendów, analiza harmoniczna czy metoda Wintersa, jednak w artykule konieczne było ograniczenie się do wybranych.

3. Przykłady sezonowych zjawisk gospodarczych

Do porównań wyników dekompozycji szeregów czasowych wybrano dane charakteryzujące kilka zjawisk gospodarczych, o których wiadomo, że charakteryzują się sezonowością. Wykorzystano dwa szeregi kwartalne: wartość dodana PKB w sektorze budownictwa (zmienna *budownictwo*), przeciętne zatrudnienie w rolnictwie, leśnictwie i łowiectwie (zmienna *rolnictwo*) oraz średnie miesięczne ceny ziemniaków uzyskiwanych przez rolników na targowiskach w kolejnych miesiącach (zmienna *ziemniaki*). Okres objęty badaniem obejmował lata 1995-2005.

Zmienna *budownictwo* (rys. 1) ma trend oraz sezonowość, przy czym amplituda wahań sezonowych zmienia się w czasie. Jednak charakter tych zmian trudno określić wyraźnie jako proporcjonalny do średniej rocznej. W początkowym okresie amplituda wahań rośnie wraz z poziomem zjawiska, potem poziom spada, ale różnice między sezonami się utrzymują.



Rys. 1. Szereg czasowy wartości dodanej PKB w budownictwie oraz wartości teoretyczne wybranych modeli
Źródło: opracowanie własne na podstawie danych statystycznych.

Dla tego szeregu wyznaczono wartości teoretyczne dla modeli z multiplikatywnymi (*) i addytywnymi (+) wskaźnikami sezonowości, wykorzystując trzy metody wygładzania szeregu: trend liniowy oszacowany jako:

$$\hat{y}_t = 102,89t + 8479,9,$$

(60,45) (1317,5) – wartości błędów standardowych ocen parametrów,

(0,098) (0,0000002) – wartość p testu istotności parametrów,

trend wielomianowy stopnia trzeciego (w tabeli oznaczony jako trend^3), o równaniu:

$$\hat{y}_t = 0,5507t^3 - 46,002t^2 + 1141,7t + 3238,1.$$

(0,64) (36,71) (604,62) (2688,72)

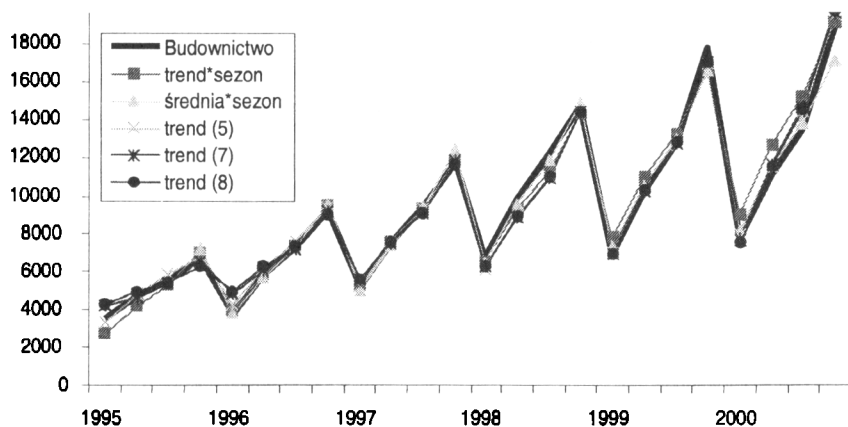
(0,39) (0,22) (0,068) (0,24)

oraz dla porównania średnią ruchomą czterookresową. Wyznaczono także wartości modelu autoregresji dla $p = 2$. Porównanie sum kwadratów reszt SE , wartości średniego modułu błędu MAE , wartości średniej z modułów błędów procentowych $MAPE$, wartości współczynnika R^2 wraz z wynikami testów normalności i istotności autokorelacji składnika losowego zamieszczono w tab. 1. Można zauważyć, że modele addytywne gorzej dopasowują się do danych niż modele ze wskaźnikami sezonowości multiplikatywnymi. Wszystkie modele średnio wiernie odwzorowują szereg empiryczny, jednak dla niektórych obserwacji odbiegają znacznie od danych oryginalnych. Ponadto należy zaznaczyć, że rozkład reszt dla wszystkich modeli można uznać za normalny, jednak w wielu przypadkach dał się zaobserwować zwłaszcza sezonowy charakter reszt. Wynika to zapewne ze wspomnianych właściwości szeregu, których żadna z metod nie odzwierciedla dokładnie. Należałoby być może posłużyć się innymi, być może opisowymi modelami do analizy tego szeregu w całości, a być może rozdzielić dwa okresy i osobno je modelować.

Tabela 1. Oceny wybranych modeli zmiennej *budownictwo* dla lat 1996-2005

Model	SE	MPE	$MAPE$	R^2	Normalność	Istotność ρ_1	Istotność ρ_4
Trend liniowy * sezon	95605788,6	1351,32	14,70%	0,836	$p > 0,1$	$p < 0,05$	$p < 0,05$
Trend ³ * sezon	29081069,2	690,51	7,63%	0,950	$p > 0,1$		$p < 0,05$
Średnia * sezon	15280551,3	514,72	5,29%	0,974	$p > 0,1$		
Trend liniowy + sezon	110540759,2	1418,21	15,63%	0,811	$p > 0,1$	$p < 0,05$	$p < 0,05$
Trend ³ + sezon	55142079,1	901,35	11,23%	0,906	$p > 0,1$		$p < 0,05$
Średnia + sezon	30870717,3	707,21	7,55%	0,947	$p > 0,1$		
Autoregresja $p=2$	26727792,4	832,27	7,76%	0,954	$p > 0,1$	$p < 0,05$	

Źródło: opracowanie własne na podstawie danych statystycznych.



Rys. 2. Szereg czasowy zmiennej *budownictwo* z lat 1995-2000 oraz wartości teoretyczne wybranych modeli

Źródło: opracowanie własne na podstawie danych statystycznych.

Aby zbadać różne możliwości wykorzystania zmiennych zero-jedynkowych, opisane w poprzedniej części referatu, wykorzystano fragment przedstawianego

szeregu, z lat 1995-2000, o liniowym charakterze trendu. Na rysunku 2 i w tabeli 2 można zauważyć, że wartości teoretyczne kilku modeli trendu liniowego (oraz średniej ruchomej) z multiplikatywną sezonowością są znacznie lepiej niż poprzednio dopasowane do rzeczywistych.

Tabela 2. Oceny kilku modeli szeregu zmiennej *budownictwo* dla lat 1995-200

Model	SE	MAE	MAPE	R ²	Skorygowany R ²	Normalność	Istotność ρ_1	Istotność ρ_4
Trend liniowy*sezon	12467360,7	538,71	6,22%	0,969		0,05<p<0,1	p<0,05	
Średnia*sezon	6268177,7	388,98	3,61%	0,984		p<0,01		p<0,05
Trend (1)	5791466,5	362,59	4,07%	0,986	0,979	p>0,1	p<0,05	
Trend (2)	7677586,2	400,24	5,10%	0,981	0,977	p>0,1	p<0,05	
Trend (3)	8302336,6	441,79	5,77%	0,979	0,975	p>0,1	p<0,05	

Źródło: opracowanie własne na podstawie danych statystycznych.

Dla poszczególnych modeli otrzymano następujące parametry:

- t.liniowy*sezon:

$$\hat{y}_{it} = (3526,2 + 461,5 \cdot t) \cdot c_i, \quad c_1 = 0,685, \quad c_2 = 0,929, \quad c_3 = 1,077, \quad c_4 = 1,311,$$

- średnia ruchoma*sezon: $c_1 = 0,659, \quad c_2 = 0,911, \quad c_3 = 1,084, \quad c_4 = 1,354,$

- trend (1):

$$\hat{y}_{it} = 3101 \cdot Q_1 + 4179 \cdot Q_2 + 4563 \cdot Q_3 + 4271 \cdot Q_4 + 233 \cdot Q_1^4 + 335 \cdot Q_2^2 + 429 \cdot Q_3^3 + 636 \cdot Q_4^4,$$

(465,6) (496,5) (528,0) (560,1) (36,0) (36,0) (36,0) (36,0)

(5 · 10⁻⁶) (3 · 10⁻⁷) (2 · 10⁻⁷) (1 · 10⁻⁶) (8 · 10⁻⁶) (7 · 10⁻⁸) (2 · 10⁻⁹) (6 · 10⁻¹²)

$$\hat{y}_{it} = 3965,1 + 176,2 \cdot Q_1^4 + 348,5 \cdot Q_2^2 + 464,9 \cdot Q_3^3 + 653,8 \cdot Q_4^4,$$

- trend (2): (268,9) (27,0) (25,3) (24,0) (22,8)

(7 · 10⁻¹²) (3 · 10⁻⁶) (2 · 10⁻¹¹) (6 · 10⁻¹⁴) (4 · 10⁻¹⁷)

- trend (3):

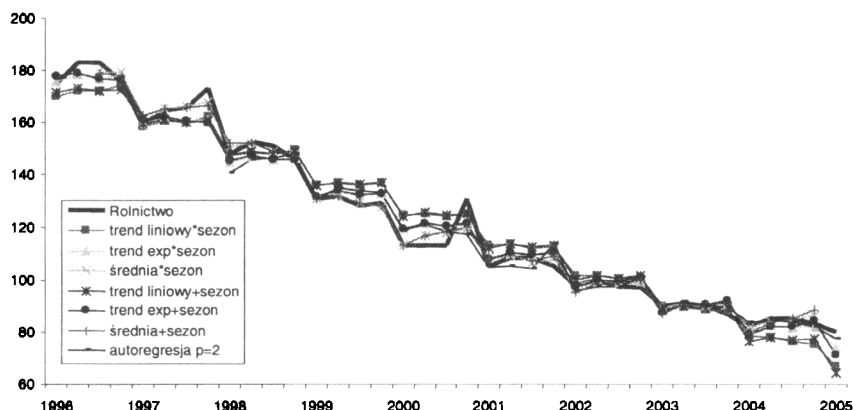
$$\hat{y}_{it} = 3579,1 + 659,9 \cdot Q_1^4 + 1324,0 \cdot Q_2^2 + 1843,7 \cdot Q_3^3 + 2704,2 \cdot Q_4^4$$

(307,7) (99,2) (99,2) (99,2) (99,2)

(4 · 10⁻¹⁰) (2 · 10⁻⁶) (4 · 10⁻¹¹) (1 · 10⁻¹³) (1 · 10⁻¹⁶)

Można zauważyć, że w tym przypadku dokładność odwzorowania danych była istotnie większa niż poprzednio oraz że była większa dla modeli ze zmiennymi sezonowymi niż dla trendu ze wskaźnikami sezonowości, który także uzyskał gorszy wynik w teście normalności. Natomiast pomiędzy ocenami trzech modeli (1), (2) i (3) nie zaznaczają się duże różnice, zwłaszcza dla skorygowanego R², jednak najdokładniejszy jest model (1) z największą liczbą zmiennych. Model (3) jest mniej dokładny niż model (2) z tą samą liczbą zmiennych. Reszty modeli nie mają sezonowości, choć nie są wolne od autokorelacji rzędu 1.

Dane dotyczące zmiennej *rolnictwo* od początku roku 1996 do pierwszego kwartału roku 2005 przedstawiono na rys. 3 wraz z wartościami teoretycznymi wybranych modeli.



Rys. 3. Szereg czasowy zmiennej *rolnictwo* w kolejnych kwartałach oraz wartości teoretyczne modeli
Źródło: opracowanie własne na podstawie danych statystycznych.

Widoczne jest stosunkowo niskie zatrudnienie w pierwszych kwartałach roku. Poza tym wysokość wahań sezonowych wydaje się być proporcjonalna do średniego poziomu zjawiska w danym okresie. Zastosowano dla tego szeregu wskaźniki sezonowości zarówno addytywne, jak i multiplikatywne, dla średniej ruchomej, trendu liniowego oraz trendu wykładniczego (w tabeli „exp”).

Tabela 3. Oceny wybranych modeli zmiennej *rolnictwo* dla okresu 1996-2005

Model	SE	MAE	MAPE	R ²	Normalność	Istotność ρ_1	Istotność ρ_4
Trend liniowy *sezon	1664,1	5,620	4,84%	0,958	$p>0,1$	$p<0,05$	$p<0,05$
Trend exp*sezon	726,6	3,403	2,77%	0,982	$p>0,1$	$p<0,05$	
Średnia*sezon	305,7	2,097	1,75%	0,992	$p<0,01$		
Trend liniowy +sezon	1737,8	5,753	5,01%	0,956	$p>0,1$	$p<0,05$	$p<0,05$
Trend exp+sezon	826,3	3,716	3,11%	0,979	$p>0,1$		
Średnia+sezon	379,4	2,440	2,11%	0,990	$0,02<p<0,05$		
Trend exp 01	709,7	3,414	2,79%	0,982	$p<0,01$	$p<0,05$	

Źródło: opracowanie własne na podstawie danych statystycznych.

W tabeli 3 można zauważyć, że lepsze dopasowanie dają wskaźniki multiplikatywne oraz że trend wykładniczy lepiej dopasowuje się do danych niż trend liniowy. Dlatego uzupełniono zestaw modeli o model trend exp 01 ze zmiennymi zero-jedynkowymi zamiast wskaźników sezonowości o parametrach oszacowanych jako (błędy standardowe oraz prawdopodobieństwa odnoszą się do modelu liniowego względem $\ln y$):

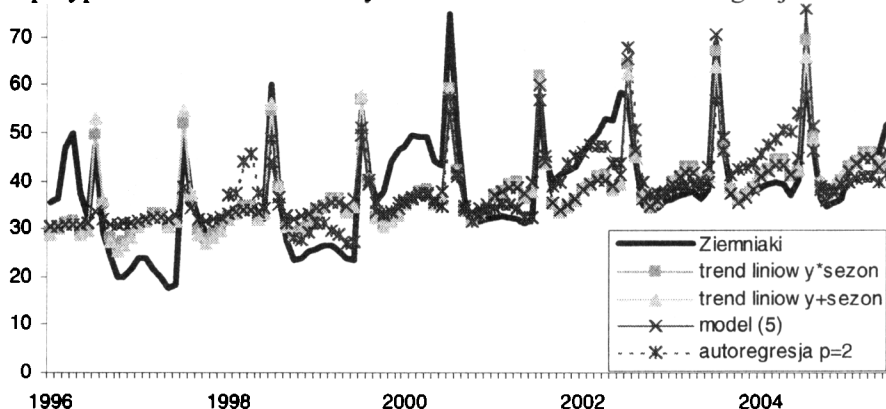
$$\hat{y}_t = \exp(-0,0243 \cdot t + 5,198 \cdot Q_1 + 5,240 \cdot Q_2 + 5,252 \cdot Q_3 + 5,281 \cdot Q_4)$$

$$(0,0006) \quad (0,017) \quad (0,017) \quad (0,017) \quad (0,018)$$

$$(3 \cdot 10^{-29}) \quad (2 \cdot 10^{-57}) \quad (4 \cdot 10^{-57}) \quad (8 \cdot 10^{-57}) \quad (3 \cdot 10^{-57})$$

Dla tego modelu uzyskano lepsze wyniki dopasowania niż dla modelu ze wskaźnikami sezonowości, choć zastrzeżenia budzi jego rozkład reszt.

Następnie analizowano szereg wartości zmiennej *ziemniaki*, dla której można zauważyć pewną dwuletnią cykliczność średniego poziomu cen widoczną przez znaczną część obserwowanego okresu (rys. 4). Można się zatem spodziewać, że w tym przypadku właściwe może być zastosowanie modelu autoregresji.



Rys. 4. Szereg czasowy wielkości przeciętnej ceny ziemniaków oraz wartości teoretyczne wybranych modeli

Źródło: opracowanie własne na podstawie danych statystycznych.

Dla szeregu przedstawionych danych wyznaczono także wartości wskaźników sezonowości (addytywne i multiplikatywne) dla trendu liniowego oraz model ze zmiennymi zero-jedynkowymi. Parametry oceny modeli zamieszczono w tab. 4.

Tabela 4. Porównanie ocen wybranych modeli zmiennej *ziemniaki* dla lat 1996-2005

Model	SE	MAE	MAPE	R ²	Normalność	Istotność ρ_1	Istotność ρ_4
Trend liniowy*sezon	5668,7	5,49	20,16%	0,571	$p < 0,01$	$p < 0,05$	$p < 0,05$
Trend liniowy+sezon	5640,3	5,57	20,17%	0,573	$p < 0,01$	$p < 0,05$	$p < 0,05$
Model (1)	5635,7	5,61	22,37%	0,573	$p < 0,01$	$p < 0,05$	$p < 0,05$
Autoregresja $p = 2$	3799,3	5,09	12,48%	0,712	$p < 0,01$	$p < 0,05$	

Źródło: opracowanie własne na podstawie danych statystycznych.

W tym przypadku do wszystkich modeli można mieć spore zastrzeżenia, w odniesieniu zarówno do stopnia dopasowania, jak i do własności reszt. Wybrane metody nie opisują zjawiska w akceptowalny sposób. Tak jak można było przypuszczać, stosunkowo najlepsze dopasowanie – choć nie jest to także zadawalający poziom – uzyskano w tym przypadku dla modelu autoregresji rzędu 2*12:

$$\hat{y}_t = 10,654 + 0,1485y_{t-12} + 0,6472y_{t-24}$$

$$(3,097) \quad (0,067) \quad (0,066)$$

$$(0,0008) \quad (0,03) \quad (1 \cdot 10^{-15})$$

4. Prognozy

Dla wybranych metod wykonano prognozy wygaście, odcinając ostatni badany rok i wyznaczając na nowo parametry modeli, a następnie porównując je z wartościami rzeczywistymi. Dla zmiennej *budownictwo* dla lat 1996-2005 i dla zmiennej *ziemniaki* nie zdecydowano się wykonać prognoz, uznając wyniki otrzymane w procesie estymacji za niewystarczająco dobre. Na podstawie zachowania zmiennej *budownictwo* z lat 1995-1999 wykonano prognozę na rok 2000 metodami oznaczanymi jako trend liniowy*sezon, trend (1), trend (2) i trend (3). Dla zmiennej *rolnictwo* wykonano prognozę na ostatni rok metodami trend liniowy*sezon, trend exp*sezon, trend exp+sezon i trend exp 01. Wyniki zamieszczono w tab. 5.

Tabela 5. Średnie błędy prognoz dla wybranych metod i zmiennych

Zmienna <i>rolnictwo</i>			Zmienna <i>budownictwo</i> 1995-2000		
Model	MAE	MAPE	Model	MAE	MAPE
Trend liniowy *sezon	14,49	17,48%	trend liniowy*sezon	538,71	18,71%
Trend exp*sezon	5,28	6,39%	trend (1)	362,59	8,77%
Trend exp+sezon	5,30	6,45%	trend (2)	400,24	8,66%
Trend exp 01	5,31	6,42%	trend (3)	441,79	9,05%

Źródło: opracowanie własne na podstawie danych statystycznych.

Należy zauważyć, że wartości błędów prognoz są gorsze niż wielkości błędów MAPE, gdy cały szereg był traktowany jako okres estymacji. To oczywiście niekorzystnie świadczy o zastosowaniu przedstawionych metod do prognozowania tych szeregów. Dla zmiennej *budownictwo* pogłębiła się jeszcze różnica w wynikach uzyskiwanych modelami regresji w stosunku do metody wskaźników, dla zmiennej zaś *rolnictwo* trend liniowy jest na pewno nieodpowiedni, choć wszystkie metody nie doszacowują wartości rzeczywistych.

5. Podsumowanie

Zaprezentowano wyniki obserwacji wpływu sposobu uwzględniania sezonowości w wybranych procesach gospodarczych. Dla szeregu cen ziemniaków nie udało się znaleźć właściwej metody opisu wśród zastosowanych. W przypadku wartości dodanej PKB w budownictwie łatwiej znaleźć taką metodę, gdy szereg jest ujmowany częściowo. Wówczas, podobnie jak dla wielkości kwartalnego zatrudnienia w rolnictwie, uzyskano dobry stopień dopasowania do danych. Jednak wszystkie metody wykazują niekorzystne tendencje w prognozach.

Należy zauważyć, że użycie zmiennych zero-jedynkowych (lub zmiennych od nich pochodnych, w rodzaju Q^i_t) dawało dla rzeczywistych procesów zazwyczaj lepsze wyniki niż wyznaczenie wskaźników sezonowości dla tego samego kształtu trendu. Dobroć dopasowania pogarsza się, gdy stosuje się metody z mniejszą liczbą

parametrów (model (2), a zwłaszcza (3)), choć nie zawsze – co widać w prognozach – jest to wpływ oczywisty. Stosunkowo dobre wyniki otrzymywano dla modeli wygładzanych poprzez średnią ruchomą. Należy jednak pamiętać, że modele takie mają bardzo ograniczone zastosowanie prognostyczne i mogą być raczej używane jako punkt odniesienia dla innych metod opisu.

Literatura

- Cieślak M. (red.) (1999), *Prognozowanie gospodarcze. Metody i zastosowania*, Wydawnictwo Naukowe PWN, Warszawa.
- Box G.E.P., Jenkins G.M. (1983), *Analiza szeregów czasowych. Prognozowanie i sterowanie*, PWN, Warszawa.
- Czerwiński Z., Guzik B. (1980), *Prognozowanie ekonometryczne*, PWE, Warszawa.
- Dittman P. (2000), *Metody prognozowania sprzedaży w przedsiębiorstwie*, AE, Wrocław.
- Enders W. (1995), *Applied Econometric Time Series*, John Wiley & Sons, Inc., USA.
- Łapińska-Sobczak N. (red.) (1998), *Opisowe modele ekonometryczne. Elementy teorii, przykłady i zadania*, Wydawnictwo Uniwersytetu Łódzkiego, Łódź.
- Statistica PL dla Windows* (1997), Wydawnictwo Statsoft, Kraków.
- Witkowska D. (2002), *Sztuczne sieci neuronowe i metody statystyczne*, Wydawnictwo C.H. Beck, Warszawa.
- Welfe A. (1998), *Ekonometria*, PWE, Warszawa.
- Zeliaś A. (1997), *Teoria prognozy*, PWE, Warszawa.

SEASONALITY ANALYSIS OF SOME OF ECONOMIC PROCESS

Summary

A subject of paper is analysis of influence of seasonality method implemented in model over quality of the investigated occurrence description, in which there is trend as well as seasonality. Special intention is paid on time series with multiplicative seasonality. These methods were implemented to decomposition of some economic occurrences with seasonal character.