Tadeusz Kufel
Nicolaus Copernicus University, Toruń

MODELLING SEASONALLY INTEGRATED PROCESSES AND PROCESSES WITH SEASONAL VOLATILITY FOR DAILY DATA

Abstract: Cyclical, high frequency economic processes contain complicated internal structure with possible time-varying average and variance and cyclical varying average and variance. We can ask the following questions:

Have high frequency economic processes got deterministic or stochastic cyclicity?

Can cyclical variability in high frequency economic processes be described by GARCH model with deterministic cyclicity?

The paper contains discussion of the issues of testing seasonal unit root by new stationarity test for high frequency processes proposed by D.A. Dickey in 2009.

All discussed problems will be illustrated by examples for daily data with analysis of heteroscedasticity by GARCH (q, p) model.

Key words: high frequency economic data, testing seasonal unit root for high frequency processes, GARCH model.

1. Stationarity test for seasonal high frequency processes

In 2009 D.A. Dickey proposed new test for seasonal stationarity for high frequency processes and described it in *Stationarity testing in high-frequency seasonal time series* [Dickey 2009]. Proposed method for testing stationarity for cyclical high frequency processes is a generalization for all possible cycles:

d = 2 – half year data (year cycle),
d = 4 – quarterly data (year cycle),
d = 5 – daily data (week cycle, 5-day week),
d = 6 – daily data (week cycle, 6-day week),
d = 7 – daily data (week cycle, 7-day week),
d = 12 – monthly data (year cycle),
d = 22 – daily data (month cycle, 5 day week),
d = 24 – hourly data (day cycle),
d = 26 – daily data (month cycle, 6-day week),
d = 31 – daily data (month cycle, 7-day week),
d = 36 – decade data (year cycle),

$d = 48$ – half hour data (day cycle),
$d = 52$ – weekly data (year cycle),
$d = 168$ – hourly data (week cycle 7·24),
$d = 261$ – daily data (year cycle, 5-day week),
$d = 313$ – daily data (year cycle, 6-day week),
$d = 365$ – daily data (year cycle, 7-day week).

Testing the stationarity of processes with combined cyclicity boils down to checking if the process has seasonal unit roots. The null hypothesis assumes that the process in question is SI(0) and alternative hypothesis assumes the process in question is nonstationary in variance SI(1)). Testing the above hypotheses is realized according to the following procedure. In the first step analyzed process Y_t should be cleaned from deterministic components which are deterministic trend and combined deterministic cyclicity.

$$\eta_t = Y_t - f(t) = Y_t - P_t - S_t, \tag{1}$$

where: f(t) – deterministic component containing deterministic trend and/or seasonality, $t = 1, 2, 3, \ldots, n = m \cdot d$.

In the second step following autoregressive model for η_t and η_{t-d} has to be estimated:

$$\eta_t = \alpha_d \eta_{t-d} + \varepsilon_t. \tag{2}$$

Table 1. Critical values from corrected normal distribution for Dickey stationarity test for high frequency seasonal time-series

d	$-1/(2\sqrt{d})$	$\alpha = 0.01$	$\alpha = 0.02$	$\alpha = 0.05$	$\alpha = 0.10$
2	–0.35355	–2.67990	–2.40730	–1.99841	–1.63510
4	–0.25000	–2.57635	–2.30375	–1.89485	–1.53155
5	–0.22361	–2.54995	–2.27736	–1.86846	–1.50516
6	–0.20412	–2.53047	–2.25787	–1.84898	–1.48568
7	–0.18898	–2.51533	–2.24273	–1.83384	–1.47053
12	–0.14434	–2.47069	–2.19809	–1.78919	–1.42589
22	–0.10660	–2.43295	–2.16035	–1.75145	–1.38815
24	–0.10206	–2.42841	–2.15581	–1.74692	–1.38361
26	–0.09806	–2.42441	–2.15181	–1.74291	–1.37961
31	–0.08980	–2.41615	–2.14355	–1.73466	–1.37135
36	0.08333	2.40968	2.13708	1.72819	1.36488
48	–0.07217	–2.39852	–2.12592	–1.71702	–1.35372
52	–0.06934	–2.39569	–2.12309	–1.71419	–1.35089
168	–0.03858	–2.36492	–2.09232	–1.68343	–1.32013
261	–0.03095	–2.35730	–2.08470	–1.67580	–1.31250
313	–0.02826	–2.35461	–2.08201	–1.67312	–1.30981
365	–0.02617	–2.35252	–2.07992	–1.67102	–1.30772
∞	0	–2.32635	–2.05375	–1.64485	–1.28155

Source: based on [Dickey 2009].

The null hypothesis $H_0{:}\alpha_d = 0$ means that the process in question has properties of stationary process SI(0) against alternative $H_1{:}\alpha_d < 0$ which means, that analyzed process is seasonally integrated SI(1). Test statistics is given by formula $u_{obl} = \alpha_d / S(\alpha_d)$ and has critical values based on left tailed normal distribution but corrected by formula $1/2\sqrt{d}$, where $m * d = n$. That means, that critical values come from $\left(-\infty; -u_\alpha - 1/\left(2\sqrt{d}\right)\right)$, where d = 2, 4, 5, 6, 7, 12, 22, 24, 26, 31, 36, 48, 52, 168, 261, 313, 365. Examples of critical values are given in Table 1.

Values in Table 1 are generalization of common *DHF* test which tests seasonal unit roots only for d = 2, 4 and 12. In the next section example of usage described test for empirical economic time series will be presented.

2. Empirical example of modelling seasonally integrated process of retail sales

Economic time-series varying due to several factors. Main factors for high frequency time-series are calendar reasons. For example daily data varying due to week cycle (work and free days), month cycle (salary payments) and year cycle (changes in chopping and production structure, different intensity in economic activity, for example vacation time).

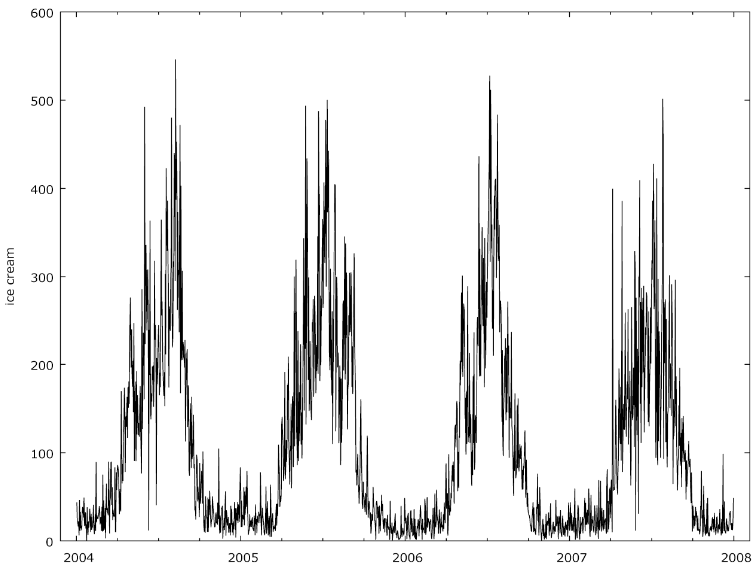

Figure 1. Daily retail ice cream sales process in chosen supermarket in the period from 2004-01-01 to 2007-12-31 (n = 1460)

Source: own research.

In current section example of modelling and predicting the daily retail ice cream sales process recorded in supermarket between 2004-01-01 and 2007-12-31 ($n = 1460$) will be presented (Figure 1).

Sale process has some specific properties, such as huge increase in sales in summer. In the same time variability of that process increases as well.

Analyzed retail daily sales process (7-day week) includes complicated cyclicity of average which consists in:

year cycle ($m = 12$ months) – harmonic approach,

month cycle ($m = 31$ days in month) – harmonic approach,

week cycle ($m = 7$ days in week) – dummy variables approach.

Econometric model of combined cyclicity with use of two approaches: harmonics (for 2-week cycle) and 6 dummy variables (for 1-week cycle) has following specification:

$$Y_t = f(t) + \eta_t = a_0 + \sum_{j=1}^{26} \left(a_j \cos\omega_j t + b_j \sin\omega_j t \right) + \sum_{i=1}^{6} c_i T_{it} + \eta_t. \qquad (3)$$

Describing the year cycle needs harmonic component, because simpler approach based on dummy variables gives imprecise model of varying average of analyzed process. Including in model 26 harmonics with periodicity from 365 days to 365/26=14.04 days (2 weeks) correctly describes variability of average. Set of 6 dummies describes 1 week cycle (see Figure 2).

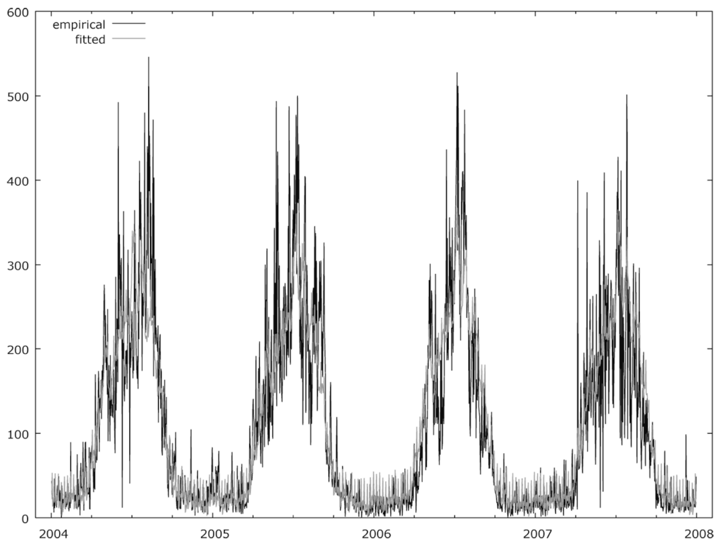

Figure 2. Fitted (from model (3)) and empirical values of daily retail ice cream sales in chosen supermarket between 2004-01-01 and 2007-12-31 ($n = 1460$)

Source: own research.

Residuals characterize specific variability of year cycle with clearly observable several times bigger variability in summer months comparing to winter time. This characteristics points to possible seasonal integration in analyzed process for year cycle ($d = 365$), but presence of seasonal unit roots for one month cycle ($d = 31$) and one week cycle ($d = 7$) should be also checked. Therefore Dickey's stationarity test [Dickey 2009] for high frequency time-series was used to verify different seasonal unit roots in residuals from model (3) (Figure 3).

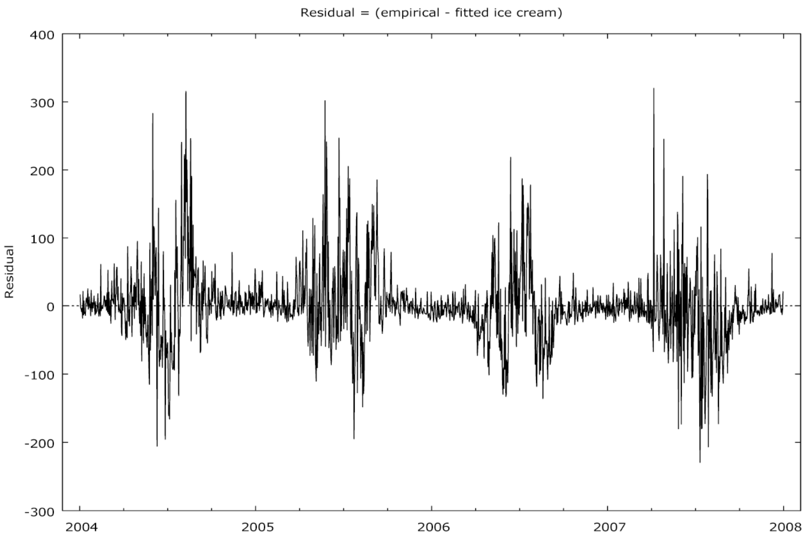

Figure 3. Residuals from model (3) of daily retail ice cream sales in chosen supermarket between 2004-01-01 and 2007-12-31 (n = 1460).

Source: own research.

Stationarity test for seasonal time-series has the following hypotheses:
Test statistics, estimated by model (4):

$$\eta_t = \alpha\eta_{t-d} + \varepsilon_t \tag{4}$$

for residuals from model (3) for $d = 365,31$ and 7, was computed according to the following formula:

$$u_{obl} = a / S(a). \tag{5}$$

Empirical result for above example were as follows:
$d = 365$, $a = -0.101743$, $u_{obl.} = -3.452$, $u_{d=365*} = -1.67102$,
$d = 31$, $a = -0.055457$, $u_{obl.} = -2.098$, $u_{d=31*} = -1.73466$,
$d = 7$, $a = 0.229201$, $u_{obl.} = 8.971$, $u_{d=7*} = -1.83384$.

Critical values u_{d*} were computed according to the following formula:

$$u_{\mathrm{d}*} = -u_{\alpha=0.05} - 1/\left(2\sqrt{d}\right) \tag{6}$$

and are presented in Table 1.

Daily retail ice cream sales process in supermarket has two unit roots for:

– year cycle ($d = 365$), $u_{obl.} = -3.452 < u_{d=365*} = -1.67102$,
– month cycle ($d = 31$), $u_{obl.} = -2.098 < u_{d=31*} = -1.73466$.

For week cycle ($d = 7$), $u_{obl.} = 8.971 > u_{d=7*} = -1.83384$, there is no seasonal integration, which means that the process in question is stationary.

Lack of seasonal unit root for $d = 7$ does not mean that residuals from model (3) did not have ARCH effect. Results for an ARCH test for lag equal to 7 confirmed statistically relevance of autoregressive heteroscedasticity.

The concept of congruent dynamic modelling worked out by Professor Zygmunt Zieliński [Zieliński 1995] assumes that in case of integrated processes model should include lagged variables Y_{t-d}, which means that in analyzed example model should include additional processes Y_{t-365}, Y_{t-31} and deterministic component GARCH(q, p) [Clements, Hendry 1998, pp. 100] for residuals.

$$\begin{aligned} Y_t &= a_0 + \sum_{j=1}^{26}\left(a_j \cos\omega_j t + b_j \sin\omega_j t\right) + \sum_{i=1}^{6} c_i T_{it} + \varepsilon_t, \\ h_t &= \alpha_0 + \sum_{s=1}^{r} \gamma_s Y_{t-s} + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i h_{t-i}. \end{aligned} \tag{7}$$

Parameters of model (7) were estimated by the maximum likelihood method.

Table 2. Model 1: Model GARCH(1,1) with component of deterministic cyclicity and autoregression for data of retail ice cream sales from 2005-01-02 to 2007-12-31 (n = 1094)

Variable	Coefficient	Standard error	*Z* statistics	*P* value	
1	2	3	4	5	
Const	66.7948	6.90451	9.6741	< 0.00001	***
cos_1	–77.5914	8.09611	–9.5838	< 0.00001	***
. . .	. . .	. . .	. . .	. . .	. . .
cos_26	0.11865	1.19529	0.0993	0.92093	
sin_26	4.93581	1.12388	4.3918	0.00001	***
Ddzien_tyg_1	–4.27	2.17569	–1.9626	0.04969	**
Ddzien_tyg_2	–4.85353	2.28997	–2.1195	0.03405	**
Ddzien_tyg_3	–2.45554	2.21805	–1.1071	0.26826	
Ddzien_tyg_4	–8.06458	2.19239	–3.6784	0.00023	***
Ddzien_tyg_5	–3.21787	2.18881	–1.4701	0.14152	
Ddzien_tyg_6	10.9912	2.23061	4.9274	< 0.00001	***

Table 2, cd.

1	2	3	4	5
Lody_1	0.226058	0.0349966	6.4594	< 0.00001 ***
Lody_5	0.0587213	0.0287235	2.0444	0.04092 **
Lody_364	0.129766	0.0281065	4.6169	< 0.00001 ***
Lody_365	–0.138334	0.0281173	–4.9199	< 0.00001 ***
Lody_366	–0.0534532	0.0262173	–2.0389	0.04146 **
alpha(0)	22.2316	5.64526	3.9381	0.00008 ***
alpha(1)	0.329626	0.0466479	7.0663	< 0.00001 ***
beta(1)	0.670374	0.0443272	15.1233	< 0.00001 ***

Mean dependent var	104.7723	S.D. dependent var	110.7182
R-squared	0.784407	S.E. of regression	51.591
Log-likelihood	–5306.758	Akaike criterion	10749.52
Schwarz criterion	11089.35	Hannan–Quinn criterion	10878.11

Source: own research.

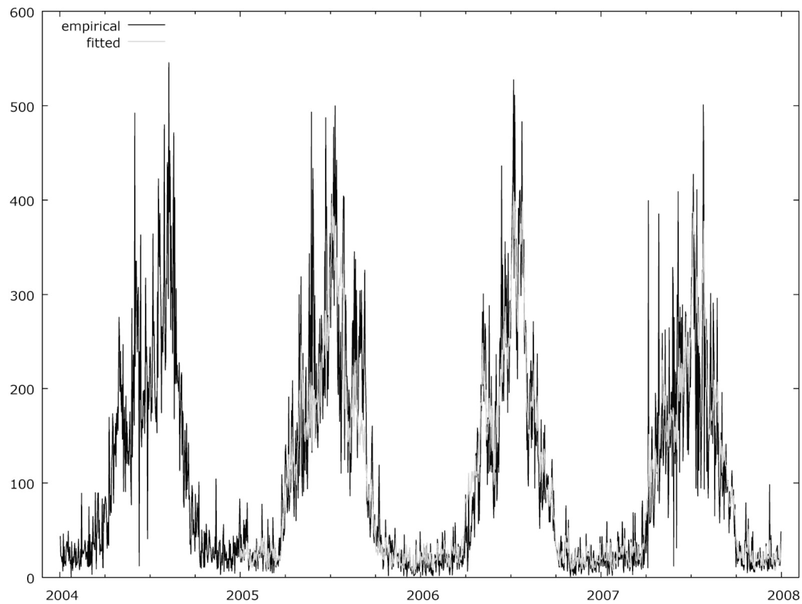

Figure 4. Fitted (from model GARCH(1,1) with seasonal and autoregressive component) and empirical values of daily retail ice cream sales in chosen supermarket between 2005-01-02 and 2007-12-31 (n = 1094)

Source: own research.

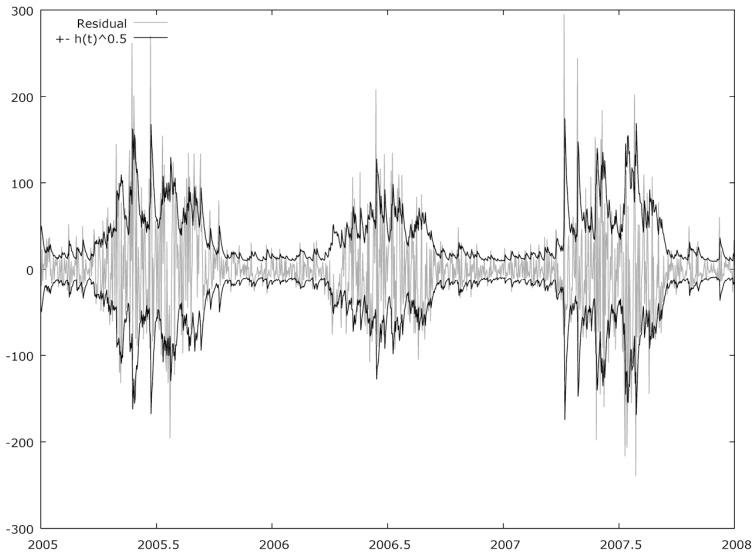

Figure 5. Residuals from model (7) and values of conditional standard deviation of residuals for process of daily retail ice cream sales in chosen supermarket between 2005-01-02 and 2007-12-31 (n = 1094)

Source: own research.

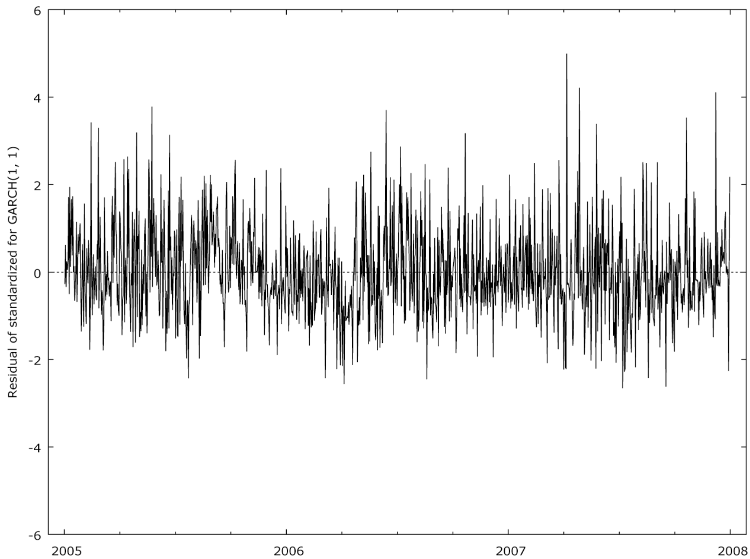

Figure 6. Values of standardized residuals from model (7) for process of daily retail ice cream sales in chosen supermarket between 2005-01-02 and 2007-12-31 (n = 1094)

Source: own research.

Process of standardized residuals has white noise properties. It means that model (7) correctly described variability of average and variance.

Estimated empirical model was the basis for generating forecasts of ice cream sales in chosen supermarket in the period from 2008-01-01 to 2008-06-30 (Figure 7).

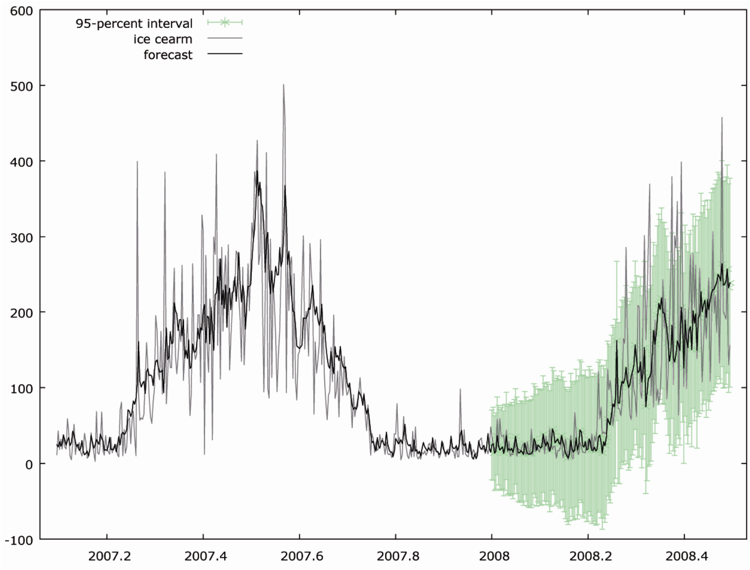

Figure 7. Forecast of ice cream sales in chosen supermarket in period from 2008-01-01 to 2008-06-30 with confidence intervals

Source: own research.

Table 3. Forecast evaluation statistics for retail ice cream sales process generated from model 1 for period from 2008-01-01 to 2008-06-30 (*h* = 182 days)

Forecast evaluation statistics	Statistics	Value
Mean Error	ME =	0.75835
Mean Squared Error	MSE =	4087.6
Root Mean Squared Error	RMSE =	63.935
Mean Absolute Error	MAE =	41.697
Mean Percentage Error	MPE =	–24.192
Mean Absolute Percentage Error	MAPE =	53.411
Theil's U	I =	0.80873
Bias proportion, UM	I1^2 =	0.00014
Regression proportion, UR	I2^2 =	0.03822
Disturbance proportion, UD	I3^2 =	0.96164

Source: own research.

Forecast evaluation statistics for retail ice cream sales process are presented in Table 3.

Mean errors of prediction point to quite big errors in forecasts. Root mean squared error had value of 63.93 PLN, while standard deviation of residuals had value of 51.59 PLN. Theil's factor points to inconsistency in direction as a cause of forecast inconsistency.

3. Summary

In 2009 D.A. Dicey published the article about new stationarity test for cyclical high frequency processes (for example $d = 2, 4, 5, 6, 7, 24, 31, 36, 48, 52, 168, 365$) based on modified critical values from standard normal distribution with critical region given by the following formula:

$$\left(-\infty, -u_\alpha - \frac{1}{2\sqrt{d}}\right].$$

Models for seasonal integrated time-series should be extended by lagged endogenous of d periods (Y_{t-d}).

Stochastic volatility of analyzed process should be described by GARCH(q, p) class model with component for modelling average value of given time-series.

Prediction of high frequency time-series is a difficult task because of very complicated internal process structure, which means combination of deterministic and stochastic cyclicity for periods of one year, month, week and hour.

Furthermore, high frequency time-series characterize outliers and non-typical observations which are difficult to model econometrically.

References

Clements M.P., Hendry D.F., *Forecasting Economic Time Series*, Cambridge University Press, Cambridge 1998.

Dickey D.A., *Stationarity testing in high-frequency seasonal time series*, SAS Global Forum 2009, Statistics and Data Analysis 2009, Paper 235-2009, pp. 1-9.

Dickey D.A., Hasza D.P., Fuller W.A. (1984), *Testing for unit roots in seasonal time series*, Journal of the American Statistical Association, Vol. 79, pp. 355-367.

Engle R., *The econometrics of ultra-high frequency data*, Econometrica, Vol. 68, No 1, pp. 1-22.

Hylleberg S., *Modelling Seasonality*, Oxford University Press, New York 1992.

Hylleberg S., Engle R.F., Granger C.W.J., Yoo B.S., *Seasonal integration and cointegration*, Journal of Econometrics, Vol. 44, pp. 215-238.

Kufel T., *Ekonometria. Rozwiązywanie problemów z wykorzystaniem programu GRETL*, Wydawnictwo Naukowe PWN, Warszawa 2007.

Kufel T., *Ekonometryczne modelowanie cykliczności w procesach sprzedaży na podstawie danych dziennych*, [in:] P. Dittmann, J. Krupowicz (eds.), *Prognozowanie w zarządzaniu firmą*, Prace Naukowe Akademii Ekonomicznej nr 1112, Wydawnictwo Akademii Ekonomicznej, Wrocław 2006, pp. 236-247.

Zieliński Z., *Analiza ekonomicznych procesów stochastycznych. Pisma wybrane*, Wydawnictwo Uniwersytetu Mikołaja Kopernika, Toruń 2002.

Zieliński Z., *Liniowe modele zgodne, opisujące zależności sumacyjnych (zintegrowanych) procesów ekonomicznych*, [in:] A. Zeliaś (ed.), *Przestrzenno-czasowe modelowanie i prognozowanie zjawisk gospodarczych*, Wydawnictwo Akademii Ekonomicznej, Kraków 1995, pp. 77-87.

MODELOWANIE PROCESÓW SEZONOWO ZINTEGROWANYCH ORAZ O SEZONOWEJ ZMIENNOŚCI DLA DANYCH DZIENNYCH

Streszczenie: cykliczne procesy gospodarcze o wysokiej częstotliwości obserwowania mają złożoną strukturę, na którą składają się zmienność wartości średniej oraz cykliczna zmienność wartości średniej, a ponadto prosta zmienność wariancji i cykliczna zmienność wariancji. Zadano dwa pytania badawcze:

Czy w procesach gospodarczych występuje cykliczność deterministyczna czy stochastyczna dla danych o wysokiej częstotliwości?

Czy cykliczną stochastyczną zmienność procesów gospodarczych można opisać modelami GARCH z deterministyczną cyklicznością?

W artykule tym omówiono zagadnienia związane z testowaniem sezonowych pierwiastków jednostkowych za pomocą nowego testu badania stacjonarności procesu dla danych o wysokiej częstotliwości obserwowania zaproponowanego przez D.A. Dickeya w 2009 r. Zagadnienia te zilustrowano przykładami dla danych dziennych wraz z analizą badania heteroskedastyczności reszt z wykorzystaniem modeli GARCH(q, p).