HSC Research Report

HSC/04/04

Periodic correlation vs. integration and cointegration

(*Okresowa korelacja a integracja i kointegracja*)

Ewa Broszkiewicz-Suwaj*
Agnieszka Wyłomańska*

* Institute of Mathematics, Wrocław University of Technology, Poland

Hugo Steinhaus Center
Wrocław University of Technology
Wyb. Wyspiańskiego 27, 50-370 Wrocław, Poland
http://www.im.pwr.wroc.pl/~hugo/

Ewa Broszkiewicz-Suwaj, Agnieszka Wyłomańska
Instytut Matematyki , Politechnika Wrocławska

OKRESOWA KORELACJA A INTEGRACJA I KOINTEGRACJA

1. Wstęp

W artykule omówiono nowe podejście do pojęcia integracji i kointegracji. Pokazano, że ażdy szereg czasowy okresowo skorelowany można podzielić na podszeregi, które są zintegrowane, a dodatkowo istnieje duże prawdopodobieństwo ich kointegracji. Zatem udowodnienie, że badany szereg jest okresowo skorelowany jednocześnie pokazuje, że podszeregi utworzone z niego są zintegrowane i nie jest konieczne badanie integracji innymi metodami.

W pierwszej części pracy omówiono własności procesów okresowo skorelowanych oraz przedstawiono jedną z metod opartą na analizie spektralnej służącą do wykrywania okresu funkcji korelacji. Metodę tę wykorzystano do zbadania okresowej korelacji w danych rzeczywistych opisujących średnie dzienne spotowe ceny energii elektrycznej na Giełdzie Nord Pool z okresu od 30 grudnia 1996 do 1 grudnia 1998 roku. W drugim rozdziale zajęto się zagadnieniem integracji oraz zbadano ją dla podszeregów wyznaczonych z szeregu wyjściowego. Podszeregi te opisują spotowe ceny energii w poszczególnych dniach tygodnia. Na zakończenie omówiono zjawisko kointegracji i przetestowano hipotezę o jej istnieniu.

2. Procesy okresowo skorelowane

Analiza szeregów czasowych w znacznej mierze bazuje na stacjonarności danych. Jednak w większości przypadków założenie, że badany szereg jest niestacjonarny jest zbyt proste. Procesy okresowo skorelowane są klasą procesów, które generalnie są niestacjonarne, ale wykazują wiele własności procesów stacjonarnych. Spotykane są miedzy innymi w klimatologii, hydrologii i ekonomii.

Definicja 1. Proces stochastyczny $\{X(t),\ t \in I\}$ o skończonym drugim momencie nazwany jest okresowo skorelowanym (ang. periodically correlated) z okresem T , jeśli T jest najmniejszą liczbą taką, że dla każdych t oraz u ze zbioru I spełnione są następujące warunki [4] (*i*) $m_X(t) = E(X(t)) = m_X(t+T)$

(*ii*) $R_X(t,u) = E((X(t) - m_X(t))(X(u) - m_X(u))) = R_X(t+T, u+T).$

Szereg czasowy okresowo skorelowany jest stacjonarny dla *T=1*.

Poniżej przedstawiono algorytm wykrywania okresowej korelacji dla zadanego szeregu czasowego $\{X(1), X(2) \ ,..., X(N)\}$ o średniej $m_X(n) \equiv 0$. Jest to metoda oparta na analizie spektralnej [2,6].

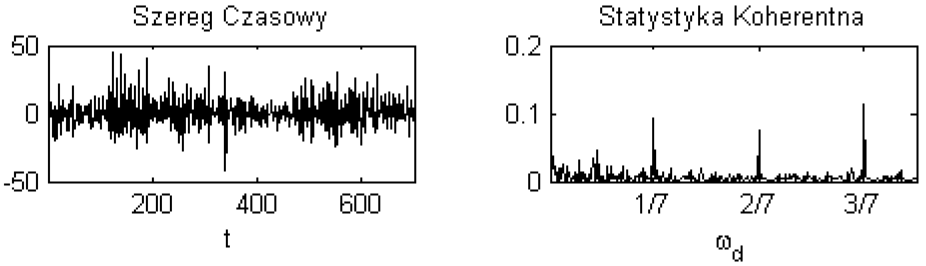
1. Wyznaczamy dyskretną transformatę Fouriera $I_N(\omega) = \sum_{n=1}^{N} X(n)\exp(-i\omega(n-1))$ w punktach $\omega_k = \dfrac{2\pi(k-1)}{N}$, $k = 1,2,\ldots,N$.

2. Obliczamy dwuwymiarową koherencję próbkową (ang. sample coherence)

$$|\gamma(\omega_p, \omega_q, M)|^2 = \frac{\left|\sum_{m=0}^{M-1} I_N(\omega_{p+m})\overline{I_N(\omega_{q+m})}\right|^2}{\sum_{m=0}^{M-1}\left|I_N(\omega_{p+m})\right|^2 \sum_{m=0}^{M-1}\left|I_N(\omega_{q+m})\right|^2} \quad \text{w każdym punkcie } (p,q).$$

3. Wyznaczamy wartości jednowymiarowej statystyki koherentnej (ang. coherent statistic) $|\gamma(0, \omega_d, N)|^2$. Statystyka ta przyjmuje wartości z przedziału $[0,1]$. Jeżeli na wykresie pojawią się wyraźne „piki" w punktach $\omega_d, 2\cdot\omega_d, 3\cdot\omega_d, \ldots,$ to znaczy, iż zadany szereg jest okresowo skorelowany z okresem $T = \dfrac{1}{\omega_d}$.

Istnieją również inne metody wykrywania okresowej korelacji oparte na analizie spektralnej [1,2,6]. Na potrzeby artykułu zbadano okresową korelację danych pochodzących ze skandynawskiej giełdy energii (średnie dzienne spotowe ceny z okresu od 30.12 1996 do 01.12.1998) . Po uprzednim usunięciu średniej poprzez zróżnicowanie danych z siedmiodniowym opóźnieniem przeprowadzono opisany test.



Rys. 1. Szereg czasowy przedstawiający dzienne zróżnicowane tygodniowo ceny z Giełdy Nord Pool oraz wykres statystyki koherentnej dla tego szeregu. Źródło: opracowania własne.

„Piki” występujące w częstotliwościach *1/7, 2/7,...* wskazują na okres funkcji korelacji, który jest równy *7* dni (tygodniowa okresowa korelacja).

3. Integracja i kointegracja

Definicja 2. Szereg niestacjonarny *{X(n), n є Z}*, który można sprowadzić do szeregu stacjonarnego, obliczając przyrosty ($\Delta X(n) = X(n) - X(n-1)$) *d* razy, nazywamy szeregiem zintegrowanym stopnia *d* (*I(d)*) [3,7].

Istnieją testy, które pozwalają na zbadanie stopnia integracji, np. test Dickeya-Fullera (DF) oraz rozszerzony test Dickeya-Fullera (ADF) [3, 5]. Do prostego testowania, czy szereg jest zintegrowany pożytecznym narzędziem jest test oparty na statystyce Durbina-Watsona:

$$IDW = \frac{\sum (X(n) - X(n-1))^2}{\sum (X(n) - \overline{X}(n))^2} \tag{1}$$

Jeśli wartość statystyki *IDW* jest mniejsza niż *0.5* istnieje podejrzenie braku integracji. Natomiast, gdy jej wartość jest bliska 2 wówczas można przyjąć hipotezę o zintegrowaniu szeregu [3].

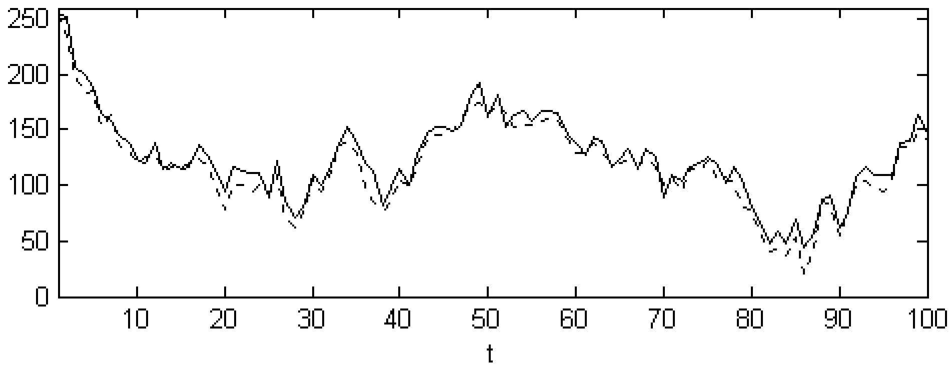
Okresowa korelacja danego szeregu czasowego jest równoważna integracji *I(0)* jego podszeregów, powstałych w następujący sposób

$$Y(v,n) = X(nT + v) \qquad v = 1,2,...T, \tag{2}$$

gdzie *X(n)* jest wyjściowym szeregiem niestacjonarnym, w którym wykryto okresową korelację. Własność ta wynika z następujących zależności, które spełnione są dla dowolnych *n* i *k* ze zbioru liczb naturalnych:

$$m_{Y(v)}(n) = m_X(nT+v) = m_X(kT+v) = m_{Y(v)}(k) = const$$
$$R_{Y(v)}(n,k) = R_X(nT+v, kT+v) = R_X(v, |k-n|T+v) = R_{Y(v)}(0, |k-n|).$$

Faktem jednak jest, że dla rzeczywistych danych okresową korelację uzyskujemy dopiero po zróżnicowaniu zadanego szeregu co może spowodować utratę długookresowych zależności. Interesujące może być zatem zachowanie podszeregów powstałych jak we wzorze (2) dla niezróżnicowanego szeregu pierwotnego.



Rys. 2. Dwa przykładowe (piątek, sobota) podszeregi powstałe z szeregu opisującego ceny energii na Giełdzie Nord Pool. Źródło: opracowania własne.

Poniżej przedstawiono wyniki otrzymane przy badaniu integracji dla podszeregów opisujących ceny energii z Giełdy Nord Pool w poszczególnych dniach tygodnia

(*7* podszeregów). Do badania stopnia integracji wykorzystano szybką metodę badania hipotezy o kointegracji opartą na statystyce Durbina-Watsona oraz wykresy funkcji autokorelacji (ACF).

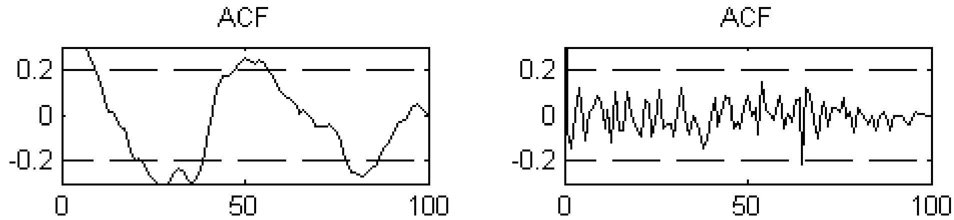
Cena	Poniedziałek	Wtorek	Środa	Czwartek	Piątek	Sobota	Niedziela
IDW	0.2187	0.1819	0.2470	0.2606	0.2034	0.1507	0.1955

Tabela 1. Wartość statystyki *IDW* dla szeregów obejmujących ceny w poszczególnych dniach tygodnia. Żródło: opracowania własne.

Wartość statystyki *IDW* dla danych pierwotnych (Tabela 1) jest istotnie mniejsza niż *0.5* dla każdego z szeregów, zatem możemy odrzucić hipotezę o integracji stopnia *0* dla każdego z nich. Wartości statystyki *IDW* dla przyrostów pierwszego rzędu poszczególnych szeregów (Tabela 2) wskazują na fakt, że szeregi te są stacjonarne. Zatem można wnioskować o integracji *I(1)* poszczególnych podszeregów.

Cena	Poniedziałek	Wtorek	Środa	Czwartek	Piątek	Sobota	Niedziela
IDW	2.1190	1.9728	2.1501	2.3200	2.2453	1.8699	2.1508

Tabela 2. Wartość statystyki *IDW* dla przyrostów szeregów obejmujących ceny w poszczególnych dniach tygodnia. Źródło: opracowania własne.



Rys. 3. Przykładowy wykres funkcji autokorelacji dla podszeregu w wersji niezróżnicowanej (panel lewy) i zróżnicowanej (panel prawy). Przerywana linia oznacza 95% przedziały ufności dla ruchu Browna.

Źródło: opracowania własne.

Dla wszystkich badanych podszeregów wykresy funkcji autokorelacji mają zbliżoną postać. Podobnie jak poprzedni test wskazują one na brak stacjonarności przed różnicowaniem i jej obecność po zróżnicowaniu.

Z pojęciem integracji ściśle związane jest zagadnienie kointegracji.

Definicja 4. Mówimy, że szeregi czasowe Y(1,n), Y(2,n),...Y(m,n) są skointegrowane stopnia d, b, gdzie d ≥ b ≥ 0 i piszemy Y(1,n), Y(2,n),...Y(m,n)~ CI (d, b), jeżeli:

1) są one zintegrowane stopnia *d*,
2) istnieje kombinacja liniowa tych zmiennych, np. $a_1Y(1,n) + a_2Y(2,n)+...a_mY(m,n)$, która jest zintegrowana stopnia $d – b$. Wektor $[a_1, a_2, ...a_m]$ nazywamy kointegrującym.

Test kointegracji przebiega dwuetapowo. W pierwszym etapie badamy stopień zintegrowania poszczególnych szeregów, co daje nam wskazówki do dalszego badania kointegracji. W drugim etapie należy wyestymować wektor kointegrujący postaci

$[1, - a_1, -a_2, \dots , -a_6]$, który pojawia się w następującym modelu regresji

$$Y(1,n) = a_1Y(2,n) + a_2Y(3,n) + ...a_6Y(7,n) + \nu(n) \quad (3)$$

Można go estymować używając metody najmniejszych kwadratów a następnie testować hipotezę o kointegracji przy użyciu testów Dickeya-Fullera. Inną prostą metodą testowania tej hipotezy jest użycie analogonu statystyki Durbina-Watsona. Test ten polega na wyznaczeniu statystyki Durbina-Watsona dla oszacowania odchyleń szeregu od głównej trajektorii, które przy założeniu prawdziwości hipotezy o kointegracji są stacjonarne [3]

$$CIDW = \frac{\sum(\hat{v}(n) - \hat{v}(n-1))^2}{\sum(\hat{v}(n) - \overline{v}(n))^2}, \tag{4}$$

gdzie $\hat{v}(n)$ są resztami wyznaczonymi metodą najmniejszych kwadratów dla równania (3). Jeśli wartość statystyki *CIDW* jest mniejsza od współczynnika determinancji R^2

$$R^2 = 1 - \frac{\sum \hat{v}^2(n)}{\sum(Y(1,n) - \overline{Y}(1,n))^2},$$

postuluje się odrzucenie hipotezy o kointegracji. W przeciwnym wypadku, gdy $CIDW>R^2$, kointegracja może występować [3].

W naszych analizach badamy kointegrację szeregów, które opisują cenę energii elektrycznej w poszczególnych dniach tygodnia. Jako zmienną opisywaną przyjmujemy ceny z poniedziałku, pozostałe szeregi tworzą zmienne opisujące. Dla takiego przypadku współczynnik determinancji wynosi *0.9524*, natomiast wartość statystyki *CIDW* jest równa *2.1523*. Zatem szeregi te są skointegrowane *CI(1,1).* Współczynniki kointegracji dla tego modelu umieszczone są w tabeli 3.

a1	*a2*	*a3*	*a4*	*a5*	*a6*
1,3473	-0,3781	0,0246	-0,0376	0.2434	-0,2087

Tabela 3. Wartości współczynników kointegracji. Źródło: opracowania własne.

4. Podsumowanie

Powyższa analiza wskazuje na fakt, że okresowa korelacja jest ściśle związana z zagadnieniem integracji, a co za tym idzie z kointegracją. Szereg okresowo skorelowany można przetransformować na zbiór podszeregów, które wykazują własność integracji, zatem nie jest konieczne badanie jej innymi metodami. Ponadto podszeregi w swej niezróżnicowanej postaci mogą być ze sobą skointegrowane. Należy podkreślić, że sama okresowość szeregu nie wystarcza do tego, by był on złożeniem podszeregów zintegrowanych. Dlatego do analizy wykorzystano szereg okresowo skorelowany opisujący spotowe ceny energii elektrycznej. Przedstawiona metodologia pozwala na badanie i opisywanie procesów wykazujących okresową korelację bez utraty zależności długookresowych. Służy ona także jako potwierdzenie słuszności metod wykrywania okresowej korelacji opartych na analizie spektralnej.

Bibliografia:

[1] Broszkiewicz-Suwaj E., 2003, *Wykrywanie okresowej korelacji danych z TGE SA w oparciu o analizę spektralną*, Rynek Terminowy 20, 92-95.

[2] Broszkiewicz-Suwaj E., Makagon A., Weron R., Wyłomańska A., 2004, *On detecting and modeling periodic correlation in financial data,* Physica A 336, 196-205.

[3] Charemza W. W., Deadman D.F., 1997, Nowa ekonometria, Polskie Wydawnictwo Ekonomiczne, Warszawa.

[4] Gladyshev E.G., 1961, *Periodically correlated random sequences*, Soviet Math. 2, 385-388.

[5] Hamilton J., D., 1994, *Time Series Analysis,* Princeton University Press, Princeton.

[6] Hurd H., Gerr N.L., 1991, *Graphical methods for detemining the presence of periodic correlation,* J. Time Ser. Anal. 12 (4), 337-350.

[7] Welfe A., 1998, *Ekonometria*, Polskie Wydawnictwo Ekonomiczne.

HSC Research Report Series 2004

For a complete list please visit http://ideas.repec.org/s/wuu/wpaper.html

01 *Finding the optimal exercise time for American warrants on WIG20 futures (Wyznaczanie optymalnego momentu wykonania warrantów amerykańskich na kontrakty futures na indeks WIG20)* by Bartosz Stawiarski

02 *Power markets in Poland and worldwide (Rynki energii elektrycznej w Polsce i na świecie)* by Rafał Weron

03 *Principal Components Analysis in implied volatility modeling (Analiza składowych głównych w modelowaniu implikowanej zmienności)* by Rafał Weron and Sławomir Wójcik

04 *Periodic correlation vs. integration and cointegration (Okresowa korelacja a integracja i kointegracja)* by Ewa Broszkiewicz-Suwaj and Agnieszka Wyłomańska