

Marcin Błażejowski

Wyższa Szkoła Bankowa w Toruniu

Symulacyjna ocena rozmiaru testu Dickeya na stacjonarność procesów o wysokiej częstotliwości obserwowania

Streszczenie. Celem artykułu jest ocena rozmiaru testu stacjonarności szeregów czasowych o wysokiej częstotliwości obserwowania, zaproponowanego przez D.A. Dickeya w 2009 r., weryfikującego hipotezę o sezonowej integracji procesu $SI_d(1)$. Test ten rozszerza zastosowanie testu sezonowego pierwiastka jednostkowego DHF (Dickey, Hasza, Fuller)¹ o przypadki częstotliwości cyklu dla $d = 5, 6, 7, 21, 24, 26, 31, 48, 52, 168, 365, \dots$

Słowa kluczowe: procesy sezonowo zintegrowane, test stacjonarności Dickeya, modelowanie cykliczności dla danych o wysokiej częstotliwości obserwowania, Gretl

Wprowadzenie

Modelowanie cykliczności procesów gospodarczych o wysokiej częstotliwości obserwowania wymaga określenia, czy obserwowane cykliczne wahania średniej procesu mają charakter deterministyczny czy stochastyczny. Prawidłowe rozróżnienie charakteru tych wahań ma zasadnicze znaczenie dla wyników

¹ D.A. Dickey, D. Hasza, W. Fuller, *Testing for Unit Roots in Seasonal Time Series*, „Journal of the American Statistical Association” 1984, nr 79.

analizy oraz jakości generowanych prognoz². Dotychczas znane były dwa testy sezonowych pierwiastków jednostkowych:

- 1) test DHF dla danych półrocznych, kwartalnych i miesięcznych³,
- 2) test HEGY dla danych kwartalnych⁴.

D.A. Dickey w 2009 r. opublikował nowy test⁵ weryfikujący hipotezę o występowaniu pierwiastka jednostkowego dla procesów o wysokiej częstotliwości obserwowania. Możliwe jest zatem poszukiwanie sezonowych pierwiastków jednostkowych dla cykli o długości⁶:

- 1) $d = 24$ – cykl o okresie doby dla danych godzinowych,
- 2) $d = 5, 6, 7$ – cykl o okresie tygodnia dla danych dziennych w układach tygodnia 5-, 6- i 7-dniowego,
- 3) $d = 22, 26, 31$ – cykl o okresie miesiąca dla danych dziennych w układach tygodnia 5-, 6- i 7-dniowego,
- 4) $d = 52$ – sezonowy pierwiastek jednostkowy dla danych tygodniowych,
- 5) $d = 261, 313, 365$ – sezonowy pierwiastek jednostkowy dla danych dziennych w układach tygodnia 5-, 6- i 7-dniowego i inne.

W artykule zostaną przedstawione podstawowe informacje o tym teście oraz jego implementacji w programie Gretl. Zaprezentowane zostaną także wyniki symulacji Monte Carlo w celu oceny rozmiaru tego testu dla danych dziennych w układzie tygodnia 5-, 6- i 7-dniowego.

1. Test stacjonarności Dickeya dla danych o wysokiej częstotliwości obserwowania

Zaproponowany przez Dickeya test stacjonarności dla danych o wysokiej częstotliwości obserwowania ma układ hipotez bardzo podobny do wcześniejszych testów pierwiastka jednostkowego, których współautorem był Dickey, tj. ADF i DHF:

² D.A. Dickey, *Stationarity Testing in High-Frequency Seasonal Time Series*, „SAS Global Forum 2009: Statistics and Data Analysis” 2009.

³ D.A. Dickey, D. Hasza, W. Fuller, *Testing for Unit Roots in Seasonal Time Series*, „Journal of the American Statistical Association” 1984, nr 79.

⁴ S. Hylleberg, R.F. Engle, C.W. Granger, B. Yoo, *Seasonal Integration and Cointegration*, „Journal of Econometrics” 1990, nr 44.

⁵ Artykuł, o którym mowa, został opublikowany w materiałach konferencyjnych SAS Global Forum 2009, w którym Dickey przedstawił zarys nowego testu, dowód asymptotycznej zbieżności statystyki testu do rozkładu normalnego, oraz dwa przykłady empiryczne. W artykule: D.A. Dickey, Y. Zhang, *Seasonal unit root tests in long periodicity cases*, „Journal of the Korean Statistical Society” 2010, nr 39, s. 271-279, tekst ten rozszerzono i poddano procesowi recenzyjnemu.

⁶ T. Kufel, *Ekonometryczna analiza cykliczności procesów gospodarczych o wysokiej częstotliwości obserwowania*, Wydawnictwo Naukowe UMK, Toruń 2010.

- $H_0 : \delta_d = 0$ (proces posiada sezonowy pierwiastek jednostkowy dla cyklu d),
- $H_1 : \delta_d < 0$ (proces nie posiada sezonowego pierwiastka sezonowego dla cyklu d).

Statystyka testu τ , którą uzyskuje się w 4-etapowej procedurze testowej, podobnie jak w testach ADF oraz DHF, posiada niestandardowy rozkład. Warto jednak zwrócić uwagę na fakt, że po pewnej korekcji statystyka ta sprowadzana jest do zmiennej losowej o asymptotycznym rozkładzie $N(0,1)$, co ma ogromne znaczenie przy weryfikacji hipotez o pierwiastku jednostkowym dla nietypowych wartości d . Procedura testowa składa się z następujących etapów.

Etap 1 – oczyszczenie testowanego szeregu Y_t ze składników deterministycznych, takich jak trend oraz cykliczność. Ze względu na potencjalnie dużą liczbę zmiennych zero-jedynkowych do opisu cykliczności Dickey sugeruje wykorzystanie jedynie wybranych harmonik. Do oszacowania parametrów wykorzystuje się KMNK i uzyskuje się nowy, oczyszczony proces y_t .

Etap 2 – estymacja modelu autoregresyjnego $ARMA(p,0)$ dla sezonowych różnic procesu y_t , czyli dla szeregu $\Delta_d y_t$. Reszty tego modelu powinny charakteryzować się własnościami białego szumu.

Etap 3 – za pomocą ocen parametrów modelu oszacowanego w etapie 2 przeprowadza się filtrację procesu y_t . Jeżeli H_0 jest prawdziwa, to uzyskany na drodze tej filtracji nowy proces z_t powinien mieć charakter sezonowego błędzenia przypadkowego dla okresu d .

Etap 4 – za pomocą KMNK szacuje się regresję sezonowych różnic procesu z_t jako funkcji opóźnień rzędu d poziomów tego procesu oraz opóźnionych sezonowych różnic procesu y_t dla opóźnień $s = 1, 2, \dots, p$ według formuły:

$$\Delta_d z_t = \alpha z_{t-d} + \sum_{s=1}^p \beta_s \Delta_d y_{t-s} + v_t.$$

Statystykę τ uzyskuje się jako $\tau = \frac{\hat{\alpha}}{s(\hat{\alpha})}$, która ma niestandardowy rozkład.

Sprowadzenie tej statystyki do zmiennej o rozkładzie $N(0,1)$ przeprowadza się według formuły:

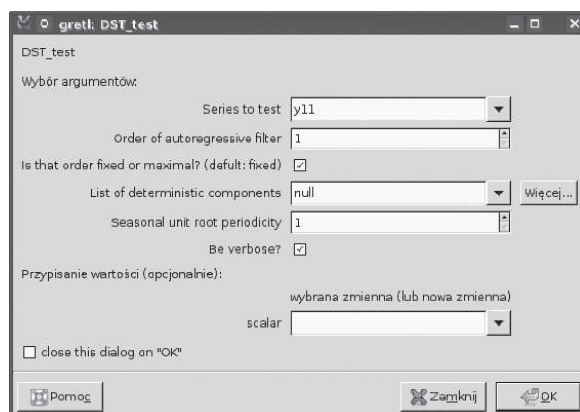
$$u = \tau + \frac{1 + k\sqrt{2}}{2\sqrt{d}},$$

gdzie k oznacza liczbę deterministycznych składników wykorzystanych w etapie 1. Warto zauważyć, że każda uwzględniona zmienna deterministyczna zwiększa wartość korekty, przesuując tym samym skorygowaną wartość statystyki w kierunku obszaru braku podstaw do odrzuceń H_0 , co może sugerować wrażliwość tego testu na obecność składników deterministycznych do opisu cykliczności.

Dokładny opis tego testu, wraz z dowodami oraz przykładami empirycznymi, można znaleźć w artykule Dickeya⁷ oraz Dickeya, Zhanga⁸.

2. Implementacja testu Dickeya w programie Gretl

Implementacja testu stacjonarności Dickeya dla danych o wysokiej częstotliwości obserwowania w Gretlu jest nowym elementem, jego pierwsza wersja ukazała się 20 sierpnia 2010 r. Procedura, roboczo nazwana *DST_test*, została napisana w wewnętrznym języku skryptowym Gretla i udostępniona w formie pakietu funkcji na publicznym serwerze pakietów przez Marcina Błażejowskiego⁹. Rysunek 1 prezentuje okno wyboru parametrów testu.



Rys. 1. Okno wyboru opcji testu stacjonarności Dickeya w programie Gretl

Źródło: opracowanie własne.

W wierszu „Series to test” należy wybrać z listy testowany szereg. W wierszu „Order of autoregressive filter” podaje się rząd p modelu autoregresyjnego $ARMA(p,0)$, który zostanie wykorzystany do przefiltrowania oryginalnego szeregu. Opcja „Is that order fixed or maximal? (default: fixed)” służy do określenia, czy wskazany w poprzednim wierszu rząd p jest dokładną wartością, czy jest to maksymalny możliwy rząd modelu filtrującego. W tym drugim przypadku zostaną oszacowane wszystkie możliwe modele $ARMA(p_i,0)$ dla $p_i = \{1, 2, \dots, p\}$ i do filtracji zostanie automatycznie wybrana specyfikacja minimalizująca kryterium

⁷ D.A. Dickey, wyd. cyt.

⁸ D.A. Dickey, Y. Zhang, wyd. cyt.

⁹ Kod tego pakietu został napisany na podstawie artykułu: D.A. Dickey, wyd. cyt., oraz oryginalnego kodu dla programu SAS napisanego przez Dickeya dla szeregu „Natural Gas”. W tym miejscu chciałbym serdecznie podziękować prof. Dickeyowi za udostępnienie mi swojego programu.

informacyjne Schwarza BIC. W wierszu „List of deterministic components” należy wskazać wszystkie składniki deterministyczne, jakie mają być wykorzystane do filtracji testowanego procesu w etapie 1 (jest to argument opcjonalny, który domyślnie ma wartość *null*). W wierszu „Seasonal unit root periodicity” wskazuje się długość cyklu d , dla którego sprawdzane jest występowanie sezonowego pierwiastka jednostkowego. Ostatnim argumentem procedury jest opcja „Be verbose?”, które określa, czy po realizacji testu powinny być pokazane wyniki wszystkich 4 etapów procedury *DST_test* (ustawienie domyślne), czy jedynie powinna zostać zwrócona wartość p -value wyznaczonej statystyki testu.

4. Symulacyjna ocena rozmiaru testu Dickeya dla danych dziennych

Tabele 1-2 zawierają opisy scenariuszy eksperymentów symulacyjnych Monte Carlo oraz ich wyników do oceny rozmiaru testu¹⁰ stacjonarności Dickeya dla danych dziennych. Każdy wariant był powtarzany 1000 razy.

Tabela 1. Eksperyment symulacyjny dla danych dziennych, proces sezonowego błędzenia przypadkowego z pojedynczym pierwiastkiem jednostkowym $t = 1001, 1002, \dots, 3000$, liczba replikacji $m = 1000$

Model generujący	
$y_t = 1 \cdot y_{t-d} + \varepsilon_t$	$\varepsilon_t \sim IID(0, \sigma^2)$
lub w zapisie operatorowym: $(1 - L^d)y_t = \varepsilon_t$	
	$d = \{dd, dm, dr\}$
	$\sigma^2 = 1, t = 1, 2, \dots, 3000$
Eksperyment A	
$p = 1$, brak zmiennych deterministycznych	
$dd = \{5, 6, 7\}, dm = \{22, 26, 31\}, dr = \{261, 313, 365\}$	

Źródło: opracowanie własne.

W eksperymencie A wygenerowano procesy sezonowego błędzenia przypadkowego o okresie tygodnia, miesiąca oraz roku dla danych dziennych w układzie tygodnia 5-, 6- i 7-dniowego, co dało łącznie 9 scenariuszy symulacyjnych.

¹⁰ W niniejszym artykule rozmiar testu rozumiany jest jako supremum prawdopodobieństwa błędu I rodzaju (por. A. Jakiel-Rokita, R. Magiera, *Modele i metody statystyki matematycznej w zadaniach*, wyd. 3 zm., Oficyna Wydawnicza GIS, Wrocław 2005, s. 92).

Tabela 2. Eksperyment symulacyjny dla danych dziennych, proces sezonowego błędzenia przypadkowego z podwójnym pierwiastkiem jednostkowym $t = 1001, 1002, \dots, 3000$, liczba replikacji $m = 1000$

Model generujący	
$y_t = 1 \cdot y_{t-d_1} + 1 \cdot y_{t-d_2} + 1 \cdot y_{t-(d_1+d_2)} \varepsilon_t$	$\varepsilon_t \sim IID(0, \sigma^2)$
lub w zapisie operatorowym: $(1 - L^{d_1})(1 - L^{d_2})y_t = \varepsilon_t$	
$(d_1, d_2) = \{(dd, dm), (dd, dr), (dm, dr)\}$	
$\sigma^2 = 1, t = 1, 2, \dots, 3000$	
Eksperyment B	
$p = 1$, brak zmiennych deterministycznych	
$(dd, dm, dr) = \{(5, 22, 261), (6, 26, 313), (7, 31, 365)\}$	

Źródło: opracowanie własne.

W eksperymencie B wygenerowano procesy o podwójnych (złożonych) pierwiastkach jednostkowych dla wszystkich par cykli o długości tygodnia, miesiąca i roku w układzie tygodnia 5-, 6- i 7-dniowego, co dało łącznie 18 scenariuszy symulacyjnych.

Każdy wygenerowany szereg miał długość $n = 3000$, jednak podczas weryfikacji hipotezy o występowaniu pierwiastka jednostkowego pominięto pierwszych 1000 obserwacji, co w rezultacie dało efektywną długość testowanych szeregów $n = 2000$. Wyniki zrealizowanych symulacji zawarte są w tabelach 3 i 4.

Tabela 3. Procent odrzuceń prawdziwej hipotezy zerowej dla różnych typów danych dziennych – wyniki eksperymentu symulacyjnego A

Rodzaj tygodnia	Poziom istotności α	Długość cyklu d		
		$dd = 5$	$dm = 22$	$dr = 261$
5-dniowy	0,10	12,1	10,6	10,4
	0,05	6,3	5,1	4,9
	0,01	0,9	0,8	1,1
		$dd = 6$	$dm = 26$	$dr = 313$
6-dniowy	0,10	11,4	10,3	9,6
	0,05	5,3	4,6	4,7
	0,01	1,2	1,2	0,9
		$dd = 7$	$dm = 31$	$dr = 365$
7-dniowy	0,10	11,4	9,5	8,1
	0,05	5,7	4,5	3,6
	0,01	1,1	1,1	0,7

Źródło: opracowanie własne.

Tabela 4. Procent odrzuceń prawdziwej hipotezy zerowej dla różnych typów danych dziennych – wyniki eksperymentu symulacyjnego B

Rodzaj tygodnia	Poziom istotności α	Długości cykli (d_1, d_2)					
		(dd, dm)		(dd, dr)		(dm, dr)	
		$dd = 5$	$dm = 22$	$dd = 5$	$dr = 261$	$dm = 22$	$dr = 261$
5-dniowy	0,10	47,8	58,3	60,5	90,0	43,5	65,9
	0,05	42,0	55,1	44,1	89,1	36,0	63,6
	0,01	32,8	49,8	16,1	88,0	20,8	59,2
		$dd = 6$	$dm = 26$	$dd = 6$	$dr = 313$	$dm = 26$	$dr = 313$
6-dniowy	0,10	10,9	20,0	59,3	86,9	35,3	59,6
	0,05	8,4	18,4	44,9	85,8	26,0	57,0
	0,01	5,4	15,6	19,3	84,1	13,6	52,0
		$dd = 7$	$dm = 31$	$dd = 7$	$dr = 365$	$dm = 31$	$dr = 365$
7-dniowy	0,10	67,5	74,1	55,0	62,2	28,6	51,5
	0,05	61,4	71,7	41,0	60,5	18,9	49,4
	0,01	49,0	68,8	15,3	57,4	8,9	43,0

Źródło: opracowanie własne.

Tabela 3 zawiera podsumowanie uzyskanych wyników w eksperymencie symulacyjnym A, przy czym kursywą zaznaczono przypadki, w których procent odrzuceń prawdziwej hipotezy zerowej jest mniejszy od założonego poziomu istotności. Można zauważyć, że zgodnie z założeniami testu jego rozmiar spada wraz ze wzrostem długości cyklu d . W przypadku weryfikacji hipotezy o pierwiastku jednostkowym o cyklu tygodniowym, we wszystkich wariantach uzyskano niewielkie przekroczenia założonego poziomu istotności α . W przypadku weryfikacji hipotez dla cyklu miesięcznego niewielkie przekroczenia założonego poziomu α uzyskano w 5 wariantach, natomiast przy weryfikacji hipotez o pierwiastku jednostkowym dla cyklu rocznego (sezonowości) niewielkie przekroczenie założonego poziomu istotności zanotowano dwa razy dla cyklu o długości $d = 261$.

Zupełnie inne wyniki uzyskano w eksperymencie symulacyjnym B, tj. dla procesów, w których występowały równoległe dwa pierwiastki jednostkowe (podsumowanie tych wyników znajduje się w tab. 4). We wszystkich przebadanych przypadkach uzyskano znaczne przekroczenia założonego poziomu istotności, przy czym dla cyklu o długości $d = 261$ ten procent dochodził do 90, natomiast dla cyklu o długości $d = 26$ współwystępującego z pierwiastkiem o długości $d = 6$ wyniki wskazują na mniejszy rozmiar testu. Jednak mimo że rozmiar testu spada wraz z długością cyklu d , należy stwierdzić, że test Dickeya nie ma mocy w przypadku szeregów o złożonych pierwiastkach sezonowych (jego rozmiar

mierzony prawdopodobieństwem błędu I rodzaju jest bardzo wysoki i niejednokrotnie osiąga wartość kilkudziesięciu procent).

Podsumowanie

D.A. Dickey w roku 2009 opublikował nowy test do weryfikowania hipotezy o występowaniu sezonowych pierwiastków jednostkowych dla danych o wysokiej częstotliwości obserwowania. Test ten wypełnia lukę, jaka dotąd istniała, pozwalając na testowanie sezonowych pierwiastków jednostkowych dla cykli o różnej długości. Asymptotyczna zbieżność statystyki tego testu do rozkładu $N(0,1)$ stwarza duże możliwości aplikacyjne, jednak jego moc oraz rozmiar wciąż wymagają badań w tym zakresie. Wyniki analiz symulacyjnych zaprezentowane w artykule wskazują, że rozmiar testu co prawda spada wraz ze wzrostem d , jednak ponownie rośnie w przypadku procesów o złożonych pierwiastkach sezonowych.

Warto także wspomnieć, że zaprezentowana tutaj implementacja testu Dickeya w programie Gretl jest prawdopodobnie pierwsza (i jedyna) na świecie wśród programów statystyczno-ekonometrycznych.

Literatura

- Box G.E.P., Jenkins G., *Analiza szeregów czasowych. Prognozowanie i sterowanie*, PWN, Warszawa 1983.
- Dickey D.A., *Stationarity Testing in High-Frequency Seasonal Time Series*, „SAS Global Forum 2009: Statistics and Data Analysis” 2009.
- Dickey D.A., Zhang Y., *Seasonal unit root tests in long periodicity cases*, „Journal of the Korean Statistical Society” 2010, nr 39, s. 271-279.
- Dickey D.A., Hasza D., Fuller W., *Testing for Unit Roots in Seasonal Time Series*, „Journal of the American Statistical Association” 1984, nr 79.
- Hylleberg S., Engle R.F., Granger C.W., Yoo B., *Seasonal Integration and Cointegration*, „Journal of Econometrics” 1990, nr 44.
- Jakiel-Rokita A., Magiera R., *Modele i metody statystyki matematycznej w zadaniach*, wyd. 3 zmienione, Oficyna Wydawnicza GIS, Wrocław 2005.
- Kufel T., *Ekonometryczna analiza cykliczności procesów gospodarczych o wysokiej częstotliwości obserwowania*, Wydawnictwo Naukowe UMK, Toruń 2010.