PRZEGLĄD STATYSTYCZNY
R. LVII – ZESZYT 4 – 2010

CZESŁAW DOMAŃSKI

# UWAGI O TESTACH JARQUE'A-BERA

## 1. MIARY SKOŚNOŚCI I KURTOZY

W literaturze statystycznej prezentuje się wiele miar skośności i spłaszczenia (kurtozy). Występują różne podziały tych miar, często jako kryterium uwzględnia się wielkość wpływu na ich wartość pojedynczych obserwacji nietypowych. Miary te często są stosowane do budowy statystyk testowych weryfikujących hipotezę o skośności rozkładów (por. Domański [4] i [5]).

Przypuśćmy, że mamy $n$ niezależnych obserwacji $x_1, x_2, \ldots, x_n$ zmiennej losowej $X$.

Dla $n$-elementowej próby określamy moment centralny $k$-tego rzędu:

$$m_k = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^k,\ k = 2, 3, \ldots \tag{1}$$

Statystyki $\sqrt{b_1} = \dfrac{m_3}{\sqrt{m_2^3}}$ $i$ $b_2 = \dfrac{m_4}{m_2^2}$ (2)

są nieobciążonymi estymatorami parametrów

$$\sqrt{\beta_1} = \frac{\mu_3}{\sigma^3}\ i\ \beta_2 = \frac{\mu_4}{\sigma^4}. \tag{3}$$

Jeśli zmienna losowa $X$ ma rozkład normalny o parametrach $\mu$ i $\sigma$, to wiadomo, że $\mu_3 = 0$, $\mu_4 = 3\sigma^4$, stąd $\sqrt{\beta_1} = 0$ $i$ $\beta_2 = 3$.

W artykule zostaną przedstawione własności testów weryfikujących hipotezy o skośności rozkładów zaproponowanych przez Jarque'a-Berego [7] oraz ich modyfikacji zaproponowanych przez Godfreya i Orme'a [6].

## 2. TESTY JARQUE'A-BEREGO I MODYFIKACJE

Test Jarque'a-Berego [7] oparty jest na miarach skośności i spłaszczenia (kurtozy) rozkładu, czyli jest testem omnibus, ponieważ uwzględnia jednocześnie odstępstwa od normalności wywołane zarówno przez skośności i spłaszczenia. Statystyka ta najczęściej przedstawiona jest wzorem:

$$JB = n\left[\frac{\left(\sqrt{b_1}\right)^2}{6} + \frac{(b_2 - 3)^2}{24}\right]. \tag{4}$$

Statystyka (4) przy założeniu prawdziwości hipotezy zerowej $H_0$, że rozkład badany jest rozkładem normalnym, ma asymptotyczny rozkład $\chi^2$ o dwóch stopniach swobody. Wydzielając pierwszy składnik, uwzględniający jedynie skośność rozkładu, dostajemy postać:

$$JB_1 = n\frac{\left(\sqrt{b_1}\right)^2}{6}. \tag{5}$$

Statystyka ta przy założeniu prawdziwości hipotezy $H_0: \sqrt{\beta_1} = 0$ (współczynnik skośności równa się zero) ma asymptotyczny rozkład $\chi^2$ o jednym stopniu swobody.

Wyodrębniając drugi składnik ze wzoru (4) otrzymujemy:

$$JB_2 = n\frac{(b_2 - 3)^2}{24}. \tag{6}$$

Statystyka (6) przy założeniu prawdziwości hipotezy zerowej $H_0$: $\beta_2 = 3$ (współczynnik spłaszczenia równa się 3) ma asymptotyczny rozkład $\chi^2$ o jednym stopniu swobody.

W wielu pracach wykazano, że test oparty na statystyce (5) nie rozróżnia w sposób wiarygodny rozkładów symetrycznych i asymetrycznych w przypadku występowania wysokiej kurtozy. Istotną przyczyną występowania nieodpowiedniej własności tego testu jest fakt, że wzór na wariancję statystyki $\sqrt{b_1}$ został wyprowadzony przy założeniu wystąpienia niskiej kurtozy. Bera i John [2] wyraźnie stwierdzili, że statystyki (5) i (6) nie są poprawnymi testami skośności i kurtozy, ponieważ asymptotyczny rozkład tych statystyk jest wyprowadzony przy założonej normalności.

W szczególności wzór określający asymptotyczną wariancję $D^2\left(\sqrt{b_1}\right) = 6/n$ wykorzystany we wzorze (5) będzie niepoprawny. Na przykład, przy dużej kurtozie będziemy mieli do czynienia z proporcjonalnie większą liczbą obserwacji nietypowych, co często występuje w praktyce przy badaniu rozkładów stóp zwrotu instrumentów finansowych (por. Piontek [9]). Pearson [8] wykazał, że wychodząc od rozkładów z krótkimi ogonami i przesuwając się w kierunku rozkładów o długich ogonach, momenty będą rosły coraz szybciej ze względu na ekstremalne wartości pochodzące z ogonów rozkładów prawdopodobieństwa. W konsekwencji wariancja statystyki (5) będzie rosła i dla rozkładów o grubych ogonach formuła $D^2\left(\sqrt{b_1}\right) = 6/n$ będzie powodowała niedoszacowanie wariancji. W przypadku dużej kurtozy, będziemy odrzucać hipotezę zerową zakładając symetrię zbyt często.

Godfrey i Orme [6] zaproponowali odpowiednią modyfikację poszerzając zakres zastosowań tego testu zwłaszcza do badania rozkładów o grubych ogonach (por. także Bera, Premarante 2001).

Znalezienie formuły dla wariancji statystyk $\sqrt{b_1}$ oraz $b_2$ (por. (2)), które będą poprawne w przypadku błędnej specyfikacji modelu, umożliwia wykorzystanie rodziny rozkładów Pearsona oraz zastosowania modyfikacji White'a (1987).

Zmodyfikowana statystyka przyjmuje postać:

$$RS = n\frac{\left(\sqrt{b_1}\right)^2}{9 + m_6 m_2^{-3} - 6m_4 m_2^{-2}} \tag{7}$$

gdzie $m_i$ oznaczają odpowiednie $i$-te momenty zwykłe rozkładu zmiennej losowej $X$.

Statystyka (7) przy założeniu prawdziwości hipotezy $H_0$ ma również asymptotyczny rozkład $\chi^2$ o jednym stopniu swobody.

## 3. PORÓWNANIE TESTÓW OPARTYCH NA STATYSTYCE $\left(\sqrt{b_1}\right)$

W celu zbadania własności rozważanych testów opartych na statystykach (5) i (7) wygenerowaliśmy metodą Monte Carlo dane z rozkładu normalnego (10000 powtórzeń). Uzyskaliśmy wynik świadczący o tym, że rozmiar zmodyfikowanego testu opartego na statystyce RS jest na ogół bliższy założonym poziomom istotności w szczególności dla $\alpha = 1\%$ (por. tabl. 1).

W tabeli 1 przedstawione są wyniki dotyczące rozmiaru testu dla rozkładu Studenta ($t_7$) o 7 stopniach swobody. W tym przypadku standardowy test niedoszacowuje wariancję, prowadząc do błędnego zbyt częstego odrzucania hipotezy $H_0$:$\sqrt{\beta} = 0$.

Wartość wyrażenia $\left(9 + \mu_6\eta_2^{-2} - 6\mu_4\eta_2^{-2}\right)/n = 104/n$ we wzorze (7) jest zdecydowanie większa niż standardowa wielkość. Prawdopodobieństwo odrzucenia hipotezy $H_0$ dla testu zmodyfikowanego jest również w tym przypadku bliskie nominalnym poziomom 5% i 1%. Proces generujący dane z rozkładu beta (2,2), ma charakter platykuryczny $\beta_2 < 3$ oraz symetryczny. W tym przypadku standardowy test cechuje się znacznie bardzo małymi oszacowaniami wartościami rozmiaru. Zmodyfikowany test tym razem również ma rozmiar bliższy nominalnym poziomom 5% i 1%.

Dla rozkładu Laplace'a, który cechuje symetrię oraz leptokurozę ($\beta_2 > 0$) standardowy test odrzuca prawdziwą hipotezę zbyt często, podczas gdy zmodyfikowany test utrzymuje swój rozmiar.

Tabela 2 przedstawia rezultaty symulacji dotyczące mocy rozważanych testów. Dla dwóch pierwszych rozkładów, które generują dane z dodatnią asymetrią ($\chi_4^2$ oraz beta 1,2) zarówno test standardowy, jak i zmodyfikowany cechuje się bardzo wysoką mocą, przy czym dla rozkładu beta (1,2) wyższą moc posiada test zmodyfikowany.

Analiza wyników prezentowanych w tablicy 1 i 2 potwierdza tezę, że standardowy test oparty na statystyce określonej wzorem (5) daje błędne wyniki w przypadku występowania dużej kurtozy. W konsekwencji powoduje to zbyt częste, błędne odrzucenie prawdziwej hipotezy zerowej mówiącej o symetrii. Test zmodyfikowany oparty na statystyce (7) daje znacznie lepsze rezultaty. Ponadto cechuje się bardzo dobrymi właściwościami odnoszącymi się do jego rozmiaru i mocy.

Tabela 1

Oszacowania rozmiarów testów skośności (asymetrii) na podstawie 10000 powtórzeń

| Wielkość próby | Test | | | |
|---|---|---|---|---|
| | standardowy ($JB_1$) | | zmodyfikowany (RS) | |
| | $\alpha$ = 5% | $\alpha$ = 1% | $\alpha$ = 5% | $\alpha$ = 1% |
| Rozkład $N(0,1)$ | | | | |
| 30 | 3,29 | 0,83 | 3,88 | 0,59 |
| 50 | 4,35 | 1,07 | 4,43 | 0,71 |
| 100 | 4,88 | 1,12 | 4,55 | 0,81 |
| 200 | 4,83 | 0,89 | 4,51 | 0,63 |
| 250 | 4,74 | 0,91 | 4,68 | 0,81 |
| 300 | 4,96 | 0,89 | 4,96 | 0,73 |
| 400 | 5,17 | 1,06 | 4,81 | 0,90 |
| 500 | 4,99 | 0,91 | 5,00 | 0,85 |
| 600 | 4,88 | 1,05 | 4,73 | 0,80 |
| 700 | 4,62 | 0,94 | 4,42 | 0,65 |
| 800 | 4,92 | 1,05 | 5,00 | 0,82 |
| 900 | 4,52 | 0,89 | 4,43 | 0,78 |
| 1000 | 5,18 | 0,87 | 4,98 | 0,88 |
| Rozkład Studenta $t_7$ | | | | |
| 30 | 15,36 | 7,71 | 3,08 | 0,39 |
| 50 | 20,93 | 11,99 | 2,66 | 0,35 |
| 100 | 27,85 | 17,45 | 3,02 | 0,27 |
| 200 | 33,48 | 22,27 | 3,44 | 0,31 |
| 250 | 34,72 | 23,29 | 2,81 | 0,20 |
| 300 | 36,79 | 24,98 | 3,28 | 0,39 |
| 400 | 37,72 | 26,16 | 3,45 | 0,34 |
| 500 | 39,91 | 27,27 | 3,50 | 0,36 |
| 600 | 39,65 | 28,12 | 3,38 | 0,37 |
| 700 | 41,44 | 29,09 | 3,67 | 0,43 |
| 800 | 41,72 | 29,94 | 3,50 | 0,30 |
| 900 | 43,05 | 30,91 | 3,65 | 0,33 |
| 1000 | 43,29 | 31,47 | 3,68 | 0,57 |

cd. tabeli 1

| Wielkość próby | Test | | | |
|---|---|---|---|---|
| | standardowy ($JB_1$) | | zmodyfikowany (RS) | |
| | $\alpha = 5\%$ | $\alpha = 1\%$ | $\alpha = 5\%$ | $\alpha = 1\%$ |
| Rozkład Beta (2,2) | | | | |
| 30 | 0,17 | 0,01 | 3,05 | 0,63 |
| 50 | 0,11 | 0,00 | 3,03 | 0,53 |
| 100 | 0,17 | 0,00 | 3,39 | 0,42 |
| 200 | 0,11 | 0,00 | 2,79 | 0,41 |
| 250 | 0,04 | 0,00 | 2,75 | 0,44 |
| 300 | 0,07 | 0,00 | 3,17 | 0,37 |
| 400 | 0,10 | 0,00 | 2,96 | 0,47 |
| 500 | 0,14 | 0,00 | 3,16 | 0,53 |
| 600 | 0,08 | 0,00 | 3,04 | 0,50 |
| 700 | 0,11 | 0,00 | 2,73 | 0,39 |
| 800 | 0,08 | 0,00 | 2,80 | 0,42 |
| 900 | 0,13 | 0,00 | 2,89 | 0,41 |
| 1000 | 0,06 | 0,00 | 2,98 | 0,45 |
| Rozkład Laplace'a ($\lambda = 2$) | | | | |
| 30 | 26,18 | 15,49 | 2,85 | 0,19 |
| 50 | 33,47 | 21,58 | 2,28 | 0,14 |
| 100 | 39,74 | 27,50 | 2,74 | 0,18 |
| 200 | 44,07 | 31,77 | 3,05 | 0,32 |
| 250 | 46,07 | 33,10 | 3,06 | 0,16 |
| 300 | 47,49 | 35,26 | 3,41 | 0,29 |
| 400 | 47,73 | 35,08 | 3,41 | 0,34 |
| 500 | 48,56 | 36,13 | 3,59 | 0,31 |
| 600 | 49,30 | 36,77 | 3,65 | 0,32 |
| 700 | 49,95 | 37,05 | 3,68 | 0,33 |
| 800 | 49,34 | 37,49 | 4,00 | 0,46 |
| 900 | 50,83 | 38,22 | 3,96 | 0,44 |
| 1000 | 50,65 | 38,76 | 4,07 | 0,52 |

Źródło: obliczenia własne.

Tabela 2

Oszacowania rozmiarów testów skośności metodą Monte Carlo oparte na 10000 powtórzeń

| Wielkość próby | Test | | | |
|---|---|---|---|---|
| | standardowy ($JB_1$) | | zmodyfikowany (RS) | |
| | $\alpha$ = 5% | $\alpha$ = 1% | $\alpha$ = 5% | $\alpha$ = 1% |
| Rozkład $\chi^2_4$ | | | | |
| 30 | 60,90 | 39,24 | 85,02 | 58,70 |
| 50 | 87,79 | 70,81 | 97,51 | 91,58 |
| 100 | 99,68 | 97,91 | 99,99 | 99,90 |
| 200 | 100 | 100 | 100 | 100 |
| 250 | 100 | 100 | 100 | 100 |
| 300 | 100 | 100 | 100 | 100 |
| 400 | 100 | 100 | 100 | 100 |
| 500 | 100 | 100 | 100 | 100 |
| 600 | 100 | 100 | 100 | 100 |
| 700 | 100 | 100 | 100 | 100 |
| 800 | 100 | 100 | 100 | 100 |
| 900 | 100 | 100 | 100 | 100 |
| 1000 | 100 | 100 | 100 | 100 |
| Rozkład Beta (1,2) | | | | |
| 30 | 11,97 | 2,36 | 49,93 | 21,04 |
| 50 | 27,33 | 6,70 | 73,39 | 49,87 |
| 100 | 66,90 | 31,92 | 96,21 | 88,18 |
| 200 | 97,98 | 84,17 | 99,96 | 96,97 |
| 250 | 99,57 | 95,09 | 100 | 100 |
| 300 | 99,95 | 98,43 | 100 | 100 |
| 400 | 100 | 99,93 | 100 | 100 |
| 500 | 100 | 100 | 100 | 100 |
| 600 | 100 | 100 | 100 | 100 |
| 700 | 100 | 100 | 100 | 100 |
| 800 | 100 | 100 | 100 | 100 |
| 900 | 100 | 100 | 100 | 100 |
| 1000 | 100 | 100 | 100 | 100 |

cd. tabeli 2

| Wielkość próby | Test | | | |
|---|---|---|---|---|
| | standardowy ($JB_1$) | | zmodyfikowany (RS) | |
| | $\alpha$ = 5% | $\alpha$ = 1% | $\alpha$ = 5% | $\alpha$ = 1% |
| Rozkład Beta (2,1) | | | | |
| 30 | 12,78 | 36,87 | 2,15 | 14,85 |
| 50 | 26,96 | 61,62 | 7,03 | 35,59 |
| 100 | 68,04 | 92,60 | 31,55 | 76,87 |
| 200 | 97,78 | 99,87 | 85,19 | 99,01 |
| 250 | 99,68 | 100 | 94,82 | 99,86 |
| 300 | 99,93 | 100 | 98,64 | 99,99 |
| 400 | 100 | 100 | 99,96 | 100 |
| 500 | 100 | 100 | 100 | 100 |
| 600 | 100 | 100 | 100 | 100 |
| 700 | 100 | 100 | 100 | 100 |
| 800 | 100 | 100 | 100 | 100 |
| 900 | 100 | 100 | 100 | 100 |
| 1000 | 100 | 100 | 100 | 100 |

Źródło: obliczenia własne.

## 4. UWAGI KOŃCOWE

Warto prześledzić własności testu opartego na statystyce określonej wzorem (6) oraz jego modyfikacji postaci:

$$RS^* = n\frac{(b_2 - 3)^2}{24 + 64b_1 + 8b_3}$$

gdzie: $b_1$ i $b_2$ dane są wzorem (2), $b_3 = \frac{m_3 m_5}{m_2^4}$,

przy czym $m_i$ ($i$ = 1, 2, ..., 5) oznaczają momenty zwykłe zmiennej losowej.

*Uniwersytet Łódzki*

## LITERATURA

[1] Asai M., Dashzeveg U., [2006], *Distribution – Free Test for Symmetry with anApplication to S&P Index Returns*, http://home.soka.ac.jpfaculty/economics?DP/004.pdf
[2] Bera A.K., John S., [1983], *Test for multivariate normality with Pearson alternatives*, Communication Statistics – Theory and Methods, 12, s. 103-117.
[3] Bera A., Premarante G., [2001], *Adjusting the Tests for Skewness and Kurtosis for Distributional Misspecifications*, http://www.business.uiuc.edu/Working_Papers/papers/01-0116.pdf
[4] Domański Cz., [1990], *Testy statystyczne*, Państwowe Wydawnictwo Ekonomiczne, Warszawa.
[5] Domański Cz., [2009], *Testy normalności oparte na momentach*, Prace i Materiały Wydziału Zarządzania Uniwersytetu Gdańskiego 4/2, s. 569-577.
[6] Godfrey L., Orme C., [1991], *Testing for Skewness of Regression Distribuances*, Economic Letters 37, s. 31-34.
[7] Jarque C.M., Bera A.K., [1987], *A Tests of Observations and Regression Residuals*, International Statistical Review, vo. 55, s. 163-172.
[8] Pearson E.S., [1963], *Some problems arising in approximating to probability distributions using moments*, Biometrika 50, s. 95-112.
[9] Piontek K., [2007], *Pomiar i testowanie skośności rozkładów stóp zwrotu instrumentów finansowych*, Prace Naukowe Akademii Ekonomicznej we Wrocławiu, Taksonomia 14, s. 122-130.
[10] White H., [1982], *Maximum likelihood estimation of misspecified models*, Econometrica, 50, s. 1-25.

Praca wpłynęła do redakcji w lipcu 2010 r.

## UWAGI O TESTACH JARQUE'A-BERA

### Streszczenie

W opracowaniu tym analizie symulacyjnej poddane zostały rozmiary testów skośności i kurtozy Jargue'a-Bera oraz ich modyfikacji. Eksperymenty symulacyjne dotyczą wybranych czterech rodzin rozkładów prawdopodobieństwa i obejmuję 10000 powtórzeń każdy.

**Słowa kluczowe:** statystyczne testy skośności i kurtozy, rozmiar testu, eksperymenty symylacyjne

## SOME REMARKS ON JARQUE-BERA TESTS

### Summary

The paper involves a description of simulation experiments aimed at determining the size of Jargue-Bera tests for skewness and kurtosis, as well as selected modifications of those tests. All experiments refer to four different families of probability distributions and consist of one thousand runs each.

**Key words:** statistical tests for skewness and kurtosis, the size of a test, simulation experiments.