Właściwości testu Jarque-Bera gdy w danych występuje obserwacja nietypowa.

Paweł Strawiński

Uniwersytet Warszawski
Wydział Nauk Ekonomicznych

16 stycznia 2006

Streszczenie

W artykule analizowane są właściwości statystyczne testu normalności rozkładu Jarque-Bera (1980). Analiza przeprowadzona jest techniką symulacji. Pokazano, że przy zachowanej normalności rozkładu zmiennej losowej test Jarque-Bera na dobre właściwości nawet w dużych próbach. Z kolei występowanie obserwacji nietypowych sprawia, że moc testu jest bardzo niska.

Streszczenie

Article analyses statistical properties of Jarque-Bera (1980) normality test. The analysis is made with use of simulation technique. It is shown that under random variable normal distribution assumption the Jarque-Bera test has good statistical properties even in large samples. However, in the presence of outliers, the power of the test is low.

Dla zapewnienia poprawności oszacowań modelu regresji liniowej wymagane jest by składniki losowe miały rozkład normalny lub przynajmniej ich rozkład dążył asymptotycznie do rozkładu normalnego. Przy założeniu, że czynnik stochastyczny równania regresji jest zmienną losową o skończonych dwóch pierwszych momentach, na podstawie Centralnego Twierdzenia Granicznego wiadomo, że jeżeli próba będzie dostatecznie duża, to rozkłady analizowanych estymatorów będą dążyły według rozkładu do rozkładu normalnego.

Istnieje szereg testów sprawdzających normalność rozkładu zmiennej losowej. W ekonometrii najczęściej używanym testem, z uwagi na swoją prostotę i znaną nieskomplikowaną postać rozkładu asymptotycznego, jest test Jarque-Bera (Bera i Jarque 1980). Konstrukcja statystyki testowej bazuje na wartościach momentów rozkładu zmiennej losowej obliczonych na podstawie próby empirycznej i porównaniu ich z momentami teoretycznymi rozkładu normalnego.

Test weryfikuje hipotezę o jednowymiarowej normalności zmiennej losowej przeciwko innemu dowolnemu rozkładowi

$$H_0 : \varepsilon \sim N(0, \sigma^2)$$

$$H_1 : \varepsilon \quad \text{ma inny rozkład}$$

za pomocą następującej statystyki

$$JB = n\Big[\frac{w}{6} + \frac{(k-3)^2}{24}\Big] \tag{1}$$

gdzie n jest liczebnością próby, w jest współczynnikiem skośności

$$w = \frac{\sum_{i=1}^{n} e_i^3}{(\sum_{i=1}^{n} e_i^2)^{\frac{3}{2}}} \tag{2}$$

a k współczynnikiem kurtozy

$$k = \frac{\sum_{i=1}^{n} e_i^2}{(\sum_{i=1}^{n} e_i^4)^2} \tag{3}$$

Rozkład normalny jest rozkładem symetrycznym, więc jego skośność wynosi $w = 0$. Kurtoza rozkładu normalnego wynosi $k = 3$. Jarque i Bera (1980) pokazali, że zaproponowana przez nich statystyka JB przy spełnionej hipotezie zerowej ma rozkład chi-kwadrat z dwoma stopniami swobody.

Z nierówności Czebyszewa i tzw. "reguły 3 sigm" wynika, że nietypowa obserwacja znacznie różniąca się od wartości przeciętnej powinna pojawiać się w próbie nie częściej niż raz na 300 losowań. Gdy taka obserwacja występuje, to obliczone na podstawie próby momenty mogą wskazywać na istotną różnicę między rozkładem empirycznym a rozkładem normalnym, pomimo tego że o "nienormalności" decyduje wyłącznie jedna obserwacja. Uzależnienie od wartości momentów próbkowych jest pewnym mankamentem statystyki JB. Jest to powodem jej niskiej mocy w dużych próbach i braku odporności na występowanie obserwacji nietypowych, czy niewielkie odstępstwa od normalności.

W celu sprawdzenia zachowania się powszechnie stosowanego test na normalność zmiennej Jarque-Bera (Bera i Jarque 1980) i zbadania rzeczywistych rozmiarów testów przeprowadziłem symulację. W pierwszym kroku procedury symulacyjnej wylosowano po 100.000 prób o różnych wielkościach. Zdecydowano się analizować zarówno małe próby liczące sto, dwieście i pięćset elementów, jak również duże próby liczące tysiąc i pięć tysięcy elementów. Wybór został podyktowany najczęściej spotykanymi rzędami wielkości prób w badaniach empirycznych. Wszystkie próby wylosowano ze standaryzowanego rozkładu normalnego o średniej 0 i wariancji 1. Następnie powtórzono losowanie za każdym razem wprowadzając m razy większą obserwację do próby. Element poddany transformacji był wybierany w sposób losowy. Tak wprowadzona obserwacja była odpowiednikiem obserwacji nietypowej. Na podstawie każdej próby obliczono statystykę testu JB. Rezultaty eksperymentu przedstawiono w tabeli

Kolumna $p - value$ zawiera nominalny poziom istotności testu, czyli prawdopodobieństwo popełnienia błędu pierwszego rodzaju. Kolejna kolumna zawiera wartości statystyk odczytane z tablic. JB jest odpowiednim kwantylem

Tabela 1: Symulacja poprawności testu JB

rozmiar próby	p-value	$\chi^2(2)$	JB	m=2	m=5	m=10
n=100	0.05	5.99	5.443	12.597	2308	15836
	0.25	7.38	7.952	26.981	3853	19270
	0.01	9.21	12.716	69.151	6199	22799
n=200	0.05	5.99	5.715	10.391	3479	60950
	0.25	7.38	8.009	20.397	7182	85832
	0.01	9.21	12.014	52.576	13975	114901
n=500	0.05	5.99	5.850	8.184	3308	184097
	0.25	7.38	7.747	13.436	8091	325262
	0.01	9.21	10.882	28.051	19221	542424
n=1000	0.05	5.99	5.934	7.216	2217	238230
	0.25	7.38	7.542	10.362	5964	505319
	0.01	9.21	10.059	18.017	15667	1015309
n=5000	0.05	5.99	6.068	6.442	590	128700
	0.25	7.38	7.529	8.221	1718	347836
	0.01	9.21	9.516	11.000	5061	931517

Źródło: Obliczenia własne.

rozkładu statystyki testowej Jarque-Bera uzyskanym w wyniku przeprowadzonej symulacji, bez wprowadzania obserwacji o nietypowej wielkości. Została ona umieszczona w celu sprawdzenia dokładności procedury generującej liczby z rozkładu normalnego. Odchylenia wartości JB od wartości wynikającej z tablicy rozkładu $\chi^2(2)$ będą traktowane jako wzorzec. Kolejne kolumny przedstawiają kwantyle rozkładu statystyki otrzymane z symulacji zawierających obserwację nietypową wielkości m.

Rezultaty testu wskazują, że gdy losowana próba pochodzi z rozkładu normalnego to odchylenie wartości statystyki testowej od wartości teoretycznej jest niewielkie, nie przekracza 10 % wartości uzyskanej statystyki. Wraz ze wzrostem rozmiaru próby wartości empiryczne statystyki JB zbiegają do wartości teoretycznych odczytanych z tablic rozkładu.

W przypadku gdy wylosowana próba zastaje zakłócona przez wprowadzenie obserwacji nietypowej uzyskiwane wartości statystyki testowej zdecydowanie rosną. W próbach o przeciętnej wielkości do 500 elementów zbieżność do rozkładu granicznego nie jest obserwowana. Wyłącznie dla bardzo dużych prób zachodzi zbieżność, jednak jej tempo jest niskie.

Kolejna tabela przedstawia rzeczywiste poziomy istotności testu.

Tabela 2: Rzeczywisty poziom istotności testu JB

rozmiar próby	p-value	$\chi^2(2)$	JB	m=2	m=5	m=10
n=100	0.05	5.99	0.042	0.101	0.469	0.716
	0.25	7.38	0.029	0.082	0.451	0.704
	0.01	9.21	0.019	0.067	0.433	0.693
n=200	0.05	5.99	0.045	0.093	0.456	0.706
	0.25	7.38	0.030	0.073	0.436	0.694
	0.01	9.21	0.015	0.057	0.419	0.683
n=500	0.05	5.99	0.047	0.082	0.423	0.684
	0.25	7.38	0.029	0.059	0.400	0.670
	0.01	9.21	0.016	0.042	0.378	0.658
n=1000	0.05	5.99	0.049	0.073	0.393	0.666
	0.25	7.38	0.027	0.048	0.367	0.651
	0.01	9.21	0.013	0.031	0.345	0.636
n=5000	0.05	5.99	0.052	0.061	0.315	0.605
	0.25	7.38	0.027	0.034	0.285	0.585
	0.01	9.21	0.011	0.017	0.260	0.567

Źródło: Obliczenia własne.

Należy podkreślić, że wartości z tabeli 2 nie powinny stanowić podstawy dla informacji o rzeczywistych wartościach testu. Rzeczywisty poziom istotności testu jest uzależniony od liczby obserwacji nietypowych i specyficznego charakteru poszczególnych obserwacji. Wartości pokazane w tabeli odzwierciedlają uśredniony efekt występowania obserwacji nietypowej, posiadając jedynie charakter ilustrujący problem związany z zastosowaniem procedury testowej zaproponowanej przez Jarque i Bera.

Wyniki przeprowadzonego eksperymentu potwierdzają, że prawdziwy rozmiar testu jest wyższy od nominalnego, oraz wskazują na niską moc testu. Rzeczywisty rozkład statystyki testowej bardzo wolno zbiega do rozkładu asymptotycznego $\chi^2(2)$. Dla przeciętnych rozmiarów prób używanych w badaniach ekonometrycznych dużo częściej niż to jest zasadne test odrzuca hipotezę zerową. Gdy rozkład jest zaburzony przez obserwację nietypową wartość statystyki testowej staje się ogromna w stosunku do wartości pozwalającej na stwierdzenie braku podstaw do odrzucenia hipotezy zerowej. Test sygnalizuje, że rozkład znacznie odbiega od normalnego, mimo że wpływ na podjęie decyzji o nienormalności rozkładu ma pojedyncza obserwacja.

Za pomocą przedstawionej symulacji wykazano słabość procedury testowej opartej o momenty rozkładu w obliczu występowania obserwacji znacznie odbiegających od wartości przeciętnych. Procedura testowa jest asymptotycznie poprawna, jednak z uwagi na bardzo niskie tempo zbieżności do rozkładu asymptotycznego występują poważne trudności w jej stosowaniu.

Ekonometrykom skupiającym swoją uwagę na zastosowaniach zależy by rozkłady empiryczne posiadały zbliżony rozkład do zakładanego. Testy oparte na momentach nie weryfikują takich hipotez. Najistotniejszym elementem dla poprawności wnioskowania statystycznego jest zachowanie się ogonów rozkładu. Jeśli rozkłady na swoich końcach są zbliżone do rozkładu zakładanego to rzeczywiste rozmiary testów statystycznych będą równe zakładanym. Jak rozkład się zachowuje w pozostałej części nie jest aż tak istotne w kontekście decyzji podejmowanych na podstawie uzyskiwanych wyników.

Literatura

[1] Bera Anil Kumar, Jarque Carols (1980) "Efficient Test for Normality, Heteroscedasticity and Serial Independence of Regression Residuals",

Economic Letters vol. 6. pp. 255-259.