Przemysław Jeziorski
Uniwersytet Ekonomiczny w Katowicach
Wydział Informatyki i Komunikacji
Zakład Demografii i Statystyki Ekonomicznej
przemyslaw.jeziorski@ue.katowice.pl

WYBRANE TESTY NIEOBCIĄŻONOŚCI MIAR RYZYKA NA PRZYKŁADZIE VALUE AT RISK

Streszczenie: Celem pracy jest analiza porównawcza testów nieobciążoności służących do oceny poprawności szacowania ryzyka metodą Value at Risk. Przedstawiono wybrane testy, które weryfikują liczbę przekroczeń oraz ich niezależność.

Słowa kluczowe: nieobciążoność, ryzyko, Value at Risk.

Wprowadzenie

Artykuł podejmuje tematykę oceny nieobciążoności Value at Risk (VaR). Nieobciążoność jest rozumiana jako zgodność liczby przekroczeń VaR z założonym poziomem istotności oraz niezależność przekroczeń w czasie. Liczba przekroczeń powinna odpowiadać poziomowi istotności – w szczególności, gdy poziom istotności wynosi 0,05, w przypadku 100 obserwacji liczba przekroczeń powinna wynosić 5. Istniejące rozwiązania do oceny poprawności estymatorów VaR, m.in. test Kupca, uwzględniają w ocenie nieobciążoności wyłącznie binarną zmienną mówiącą o tym, czy rzeczywista wartość rynkowej stopy zwrotu przekroczyła szacowaną miarę zagrożenia.

Kluczową kwestią staję się ocena efektywności oszacowań modeli VaR. Według Giacomini i Komunjer [2005] kiedy dostępne są prognozy miar ryzyka, pożądane jest posiadanie formalnej procedury testowej, dla której nie byłoby niezbędne posiadanie wiedzy o założonym modelu oraz skupianie się na procedurze estymacji modelu. W literaturze istnieje wiele testów (*backtests*), takich jak np. Kupca [1995], Christoffersena [1998] oraz Engle'a i Manganelliego [2004].

Przedstawione w pracy testy nieobciążoności uwzględniają mechanizm, dzięki któremu możliwa jest ocena efektywności modelu VaR oraz udzielenie odpowiedzi na pytanie, dlaczego pomiar ryzyka jest obciążony. Ponadto dobrze znany test Kupca [1995] charakteryzuje się niską mocą i często nie odrzuca modelu, który powinien zostać odrzucony.

W niniejszej pracy zostaną przedstawione trzy metody szacowania VaR: metoda wariancji-kowariancji, metoda historyczna oraz metoda regresji kwantylowej. Value at Risk zostanie wyznaczony dla indeksu WIG20 dla szerokości okna obserwacji 20, 60, 120, 250 oraz 500 sesji. Celem pracy jest porównanie jakości szacowanych VaR za pomocą testu Kupca, Christoffersena oraz Engle'a i Manganelliego.

1. Wybrane testy nieobciążoności

Value at Risk pozwala oszacować maksymalną stratę, jakiej można się spodziewać dla założonego poziomu istotności w zadanym horyzoncie czasu. Jeśli R_t oznacza stopę zwrotu portfela w okresie t oraz $\tau^* \in (0,1)$ oznacza założony poziom istotności, to VaR (V_t) jest zdefiniowany za pomocą wyrażenia:

$$\Pr\left[R_t < V_t \middle| F_{t-1}\right] = \tau^* \tag{1}$$

gdzie F_{t-1} są to informacje (rozkład stóp zwrotu) dostępne w okresie $t - 1$. Z powyższej definicji wynika, że V_t jest warunkowym kwantylem rzędu τ^* z R_t.

Procedura testowa poprawności modelu VaR polega na sprawdzeniu, czy jest spełnione wyrażenie $\Pr\left[R_t < V_t \middle| F_{t-1}\right] = \tau^*$, co oznacza, że V_t jest poprawnie oszacowanym kwantylem rzędu τ^* stóp zwrotu R_t.

Poprawnie skonstruowany model VaR dla poziomu istotności τ^* zwraca warunkowy kwantyl rzędu τ^* ze stóp zwrotu R_t. Celem jest zweryfikowanie hipotezy zerowej, głoszącej, że VaR poprawnie szacuje warunkowe kwantyle dla założonego poziomu istotności τ^*. W artykule zostanie przedstawiony test zawierający więcej informacji (zmiennych) w porównaniu z istniejącymi testami, posiadający tym samym większą moc dla skończonej liczebności próby.

W pierwszej kolejności zostanie zdefiniowana funkcja wskaźnikowa określająca sekwencję przekroczeń VaR:

$$H_t = \begin{cases} 1; gdy\ R_t < V_t \\ 0; gdy\ R_t \geq V \end{cases} \tag{2}$$

Z definicji prawdopodobieństwo występujących przekroczeń VaR powinno spełniać warunek: $\Pr(H_t = 1 | F_{t-1}) = \tau^*$.

1.1. Test Kupca [1995]

Nieparametryczny test opierający się na proporcji przekroczeń. Zakładając wielkość próby T oraz liczbę przekroczeń $N = \sum_{t=1}^{T} H_t$, celem testu jest określenie, czy $\hat{p} \equiv N/T$ jest statystycznie równe $\tau^* : H_0 : p = E(H_t) = \tau^*$.

Prawdopodobieństwo wystąpienia N przekroczeń w próbie T jest opisywane rozkładem dwumianowym. Hipoteza zerowa $H_0 : p = \tau^*$ może być zweryfikowana testem LR (nazywanym również bezwarunkowym testem pokrycia). Test odrzuca hipotezę zerową mówiącą o poprawności modelu VaR, jeżeli frakcja przekroczeń VaR w próbie jest statystycznie różna od τ^*. Kupiec zwracał jednak uwagę, że test ma niską moc dla skończonych prób, a większa moc dla testu pojawia się w chwili występowania bardzo dużej liczby obserwacji. Statystyka testowa dla testu posiada postać [Kupiec, 1995]:

$$LR = -2\ln\left[(1-\alpha)^{T-N}\alpha^{N}\right] + 2\ln\left[\left(1-\frac{N}{T}\right)^{T-N}\left(\frac{N}{T}\right)^{N}\right] \tag{3}$$

1.2. Test Christoffersena [1998]

Test Kupca [1995] weryfikuje hipotezę o poprawnej liczbie przekroczeń, jednak test nie reaguje na występowanie skupień w przekroczeniach VaR. Istotne jest zidentyfikowanie przekroczeń, które nie spełniają warunku niezależności. Model VaR, który ma niewłaściwą liczbę przekroczeń lub przekroczenia nie posiadają własności niezależności powinien być odrzucony. Niezależność jest ważną cechą, gdyż skupianie się przekroczeń VaR powoduje wzrost ryzyka i przyczynia się do kumulacji strat.

Sekwencja przekroczeń posiada własność efektywności, jeżeli dla zasobu informacji F_{t-1} zachodzi warunek: $E[H_t | F_{t-1}] = \tau^*$.

Tak sformułowany test nie wymaga zakładania rozkładu teoretycznego dla procesu, który jest prognozowany. Jest to istotna zaleta, gdyż jakiekolwiek założenie rozkładu do opisu zjawiska w zastosowaniach ekonomicznych jest bardzo dyskusyjne. Standardowa metoda oceny poprawności oszacowań VaR skupia się wyłącznie na sprawdzeniu czy $\hat{p} \equiv N/T$ zmierza do prawdziwej wartości poziomu istotności τ^*. W prezentowanym podejściu będzie również testowana warunkowa efektywność.

W pierwszej kolejności zostanie zbudowane kryterium określające obserwacje poza próbą (*out-of-sample*). Założono zasób informacji w okresie F_{t-1} mówiący o sekwencji przekroczeń $F_{t-1} = \{H_{t-1}, H_{t-2}, H_{t-3}, \ldots, H_1\}$.

Celem testu nie jest badanie istotności modelu, lecz skupienie się na tym, aby frakcja przekroczeń była zbieżna z założoną oraz aby sekwencja przekroczeń była niezależna.

Test Christoffersena [1998] posiada trzy wersje. Pierwsza wersja testuje wyłącznie poprawność bezwarunkowej zbieżności do założonego poziomu istotności. Drugi test weryfikuje wyłącznie hipotezę o niezależności, natomiast trzeci test jest połączeniem dwóch poprzednich i pozwala weryfikować warunkową zbieżność do założonego poziomu istotności [Christoffersen, 1998].

1.2.1. Bezwarunkowa zbieżność do poziomu istotności

Zakładając sekwencję przekroczeń $\{H_t\}$, celem jest zweryfikowanie hipotezy o poprawnej wartości frakcji odpowiadającej założonemu poziomowi istotności $E[H_t] = \tau^*$ wobec alternatywy $E[H_t] \neq \tau^*$. Prawdopodobieństwo dla pierwszej hipotezy jest określone wyrażeniem:

$$L(p; H_1, H_2, \ldots, H_t) = (1-p)^{n_0} p^{n_1} \tag{4}$$

oraz prawdopodobieństwo dla alternatywy:

$$L(\pi; H_1, H_2, \ldots, H_t) = (1-\pi)^{n_0} \pi^{n_1} \tag{5}$$

Testowanie bezwarunkowej zbieżności do poziomu istotności może zostać sprowadzone do standardowego testu ilorazu wiarygodności:

$$LR_{uc} = -2\log\left[\frac{L(p; H_1, H_2, \ldots, H_t)}{L(\pi; H_1, H_2, \ldots, H_t)}\right] \overset{asy}{\sim} \chi^2(s-1) = \chi^2(1) \tag{6}$$

gdzie $\hat{\pi} = n_1/(n_0 + n_1)$ jest szacunkiem π oraz s = 2 jest liczbą możliwych wartości, jaka może wystąpić w zbiorze sekwencji $\{H_t\}$.

Test weryfikuje czy występująca frakcja przekroczeń w próbie odpowiada założonemu poziomowi *p*.

1.2.2. Niezależność przekroczeń

Test pozwala weryfikować hipotezę o niezależności w sekwencji przekroczeń $\{H_t\}$. Założono binarny łańcuch Markova pierwszego rzędu. Przybliżona funkcja prawdopodobieństwa dla tego procesu ma postać:

$$L(\Pi_1; H_1, H_2, \ldots, H_t) = (1-\pi_{01})^{n_{00}} \pi_{01}^{n_{01}} (1-\pi_{11})^{n_{10}} \pi_{11}^{n_{11}} \tag{7}$$

gdzie n_{ij} jest liczbą obserwacji z wartością j poprzedzoną wartością i.

Zakłada się, że sekwencją przekroczeń $\{H_t\}$ można oszacować łańcuch Markova pierwszego rzędu dla modelu, co w konsekwencji pozwoli zweryfikować hipotezę o niezależności sekwencji przekroczeń. Prawdopodobieństwo dla procesu może zostać oszacowane jako:

$$L(\Pi_2; H_1, H_2, \ldots, H_t) = (1-\pi_2)^{(n_{00}+n_{10})} \pi_2^{(n_{01}+n_{11})} \tag{8}$$

gdzie macierz Π_2 może być oszacowana przy użyciu wyrażenia:

$$\hat{\pi}_2 = (n_{01} + n_{11}) / (n_{00} + n_{10} + n_{01} + n_{11}) \tag{9}$$

Zgodnie z Hoelem [1954] iloraz wiarygodności LR dla testu niezależności posiada asymptotycznie rozkład chi-kwadrat z $(s-1)^2$ stopniami swobody:

$$LR_{ind} = -2\log\left[\frac{L(\hat{\Pi}_2; H_1, H_2, \ldots, H_t)}{L(\hat{\Pi}_1; H_1, H_2, \ldots, H_t)}\right] \overset{asy}{\sim} \chi^2\left((s-1)^2\right) = \chi^2(1) \tag{10}$$

W dalszym ciągu sekwencja przekroczeń $\{H_t\}$ jest opisana binarną zmienną, dlatego s = 2. Istotny jest fakt, że wartość statystyki testowej jest niezależna od założonego poziomu istotności p i tym samym test weryfikuje hipotezę wyłącznie o niezależności sekwencji przekroczeń.

1.2.3. Zbieżności do założonego poziomu istotności oraz niezależności przekroczeń

Powyższe dwa testy (bezwarunkowej zbieżności oraz niezależności) mogą zostać połączone w jeden test, który weryfikuje warunkową zbieżność do założonego poziomu istotności. Statystyka testowa LR dla testu warunkowej zbieżności do poziomu istotności posiada asymptotyczny rozkład chi-kwadrat z $s(s-1)$ stopniami swobody.

$$LR_{cc} = -2\log\left[\frac{L(p; H_1, H_2, \ldots, H_t)}{L(\hat{\Pi}_1; H_1, H_2, \ldots, H_t)}\right] \overset{asy}{\sim} \chi^2(s(s-1)) = \chi^2(2) \qquad (11)$$

Test pozwala weryfikować niezależność przekroczeń oraz bierze pod uwagę wyłącznie autokorelacje rzędu pierwszego w sekwencji przekroczeń.

1.3. Podejście Engle'a i Manganelliego [2004]

Engle i Manganelli zaproponowali test, który posiada wiele wersji. Używając poprzedniego zapisu, zmienna losowa $Hit_t = H_t - \tau^*$ została wykorzystana przez autorów do konstrukcji dynamicznego warunkowego testu kwantyla (DQ). Zmienna Hit_t może być opisana za pomocą szeregu zmiennych objaśniających:

$$\begin{aligned} Hit_t &= \delta_0 + \delta_1 Hit_{t-1} + \ldots + \delta_p Hit_{t-p} + \\ &\delta_{p+1} VaR_t + \delta_{p+2} H_{year1,t} + \ldots + \delta_{p+2+n} H_{yearn,t} + u_t \end{aligned} \qquad (12)$$

gdzie wektor X_t może zawierać opóźnione Hit_t, V_t i jego opóźnienia.

Statystyka testowa (DQ) posiada postać:

$$DQ = \frac{\hat{\delta}_{OLS}^T X^T X \hat{\delta}_{OLS}}{\theta(1-\theta)} \overset{asy}{\sim} \chi^2(p+n+2) \qquad (13)$$

Zaproponowany test posiada statystykę chi-kwadrat z liczbą stopni $p + n + 2$, gdzie p jest rzędem macierzy X_t. Warto zauważyć, że test DQ może być używany do oceny poprawności każdego modelu VaR szacowanego podejściami innymi niż zaproponowane przez autorów modele CaViaR [Engle, Manganelli, 2004].

2. Badania empiryczne

Weryfikacji zostaną poddane modele VaR oszacowane metodą wariancji-kowariancji, metodą historyczną oraz metodą regresji kwantylowej dla WIG20. Modele zostały oszacowane na podstawie danych z okresu od 31.10.2006 r. do 28.10.2010 r. Szerokość okna obserwacji wynosiła 20, 60, 120, 250 oraz 500 sesji. Przyjęto poziom istotności dla modeli 1%, 2% oraz 5%. Okres od 29.10.2010 r. do 25.10.2012 r. posłużył do weryfikacji modeli i obejmował łącznie 500 sesji. Porównane ze sobą zostały test Kupca (LR), test przekroczeń Christoffersena (LR_{uc}), test niezależności Christoffersena (LR_{ind}), test przekroczeń i niezależności

Christoffersena oraz test Engle'a i Manganelliego (DQ). Kolumna N/T przedstawia obserwowaną liczbę przekroczeń – szarym wypełnieniem oznaczono odchylenia o ±30% od założonego poziomu istotności. W kolumnach *p-value* zostały oznaczone szarym wypełnieniem wartości mniejsze od 0,05 (założony poziom istotności dla testów), który świadczy o odrzuceniu hipotezy o poprawnym oszacowaniu modelu VaR.

Tabela 1. Wyniki nieobciążoności dla WIG20 – metoda wariancji-kowariancji

WIG20 – metoda wariancji-kowariancji (prognozy jednosesyjne)									
T	N	N/T	alfa	L. sesji	p-value (LR)	p-value (LRuc)	p-value (LRind)	p-value (LRcc)	p-value (DQ)
500	15	**3,0%**	1%	20	**0,000**	**0,017**	0,629	0,050	**0,000**
500	20	**4,0%**	2%		**0,005**	0,063	0,433	0,129	**0,000**
500	32	6,4%	5%		0,168	0,363	0,209	0,292	**0,006**
500	14	**2,8%**	1%	60	**0,001**	**0,029**	0,205	**0,041**	**0,000**
500	18	**3,6%**	2%		**0,021**	0,130	0,134	0,102	**0,000**
500	31	6,2%	5%		0,235	0,433	0,611	0,629	0,067
500	14	**2,8%**	1%	120	**0,001**	**0,029**	0,061	**0,016**	**0,000**
500	16	**3,2%**	2%		0,078	0,245	0,093	0,123	**0,000**
500	27	5,4%	5%		0,685	0,789	**0,015**	**0,050**	**0,000**
500	12	**2,4%**	1%	250	**0,008**	0,079	**0,008**	**0,006**	**0,000**
500	13	2,6%	2%		0,359	0,546	**0,011**	**0,033**	**0,000**
500	23	4,6%	5%		0,678	0,784	**0,005**	**0,018**	**0,000**
500	10	**2,0%**	1%	500	**0,048**	0,192	0,098	0,108	**0,000**
500	15	**3,0%**	2%		0,137	0,327	0,076	0,127	**0,000**
500	24	4,8%	5%		0,836	0,892	**0,021**	0,0663	**0,000**

Źródło: Opracowanie własne.

Tabela 2. Wyniki nieobciążoności dla WIG20 – metoda historyczna

WIG20 – metoda symulacji historycznej (prognozy jednosesyjne)									
T	N	N/T	alfa	L. sesji	p-value (LR)	p-value (LRuc)	p-value (LRind)	p-value (LRcc)	p-value (DQ)
500	26	**5,2%**	1%	20	**0,000**	**0,000**	0,825	**0,000**	**0,000**
500	29	**5,8%**	2%		**0,000**	**0,001**	0,692	**0,004**	**0,000**
500	44	**8,8%**	5%		**0,000**	**0,020**	0,738	0,060	**0,000**
500	20	**4,0%**	1%	60	**0,000**	**0,001**	0,883	**0,003**	**0,000**
500	23	**4,6%**	2%		**0,000**	**0,019**	0,570	0,053	**0,000**
500	34	**6,8%**	5%		0,079	0,247	0,760	0,474	0,059
500	15	**3,0%**	1%	120	**0,000**	**0,017**	0,238	**0,028**	**0,000**
500	20	**4,0%**	2%		**0,005**	0,063	0,184	0,073	**0,000**
500	27	5,4%	5%		0,685	0,789	0,099	0,241	**0,001**
500	11	**2,2%**	1%	250	**0,020**	0,125	0,121	0,092	**0,000**
500	12	2,4%	2%		0,536	0,683	**0,037**	0,104	**0,000**
500	25	5,0%	5%		1,000	1,000	**0,003**	**0,011**	**0,000**
500	6	1,2%	1%	500	0,663	0,774	0,204	0,426	**0,000**
500	11	2,2%	2%		0,753	0,836	0,121	0,291	**0,000**
500	22	4,4%	5%		0,530	0,679	**0,037**	0,102	**0,000**

Źródło: Opracowanie własne.

Tabela 3. Wyniki nieobicążoności dla WIG20 – metoda regresji kwantylowej

WIG20 – metoda regresji kwantylowej (prognozy jednosesyjne)									
T	N	N/T	alfa	L. sesji	p-value (LR)	p-value (LRuc)	p-value (LRind)	p-value (LRcc)	p-value (DQ)
500	18	**3,6%**	1%	20	**0,000**	**0,003**	0,444	**0,009**	**0,000**
500	18	**3,6%**	2%		**0,021**	0,130	0,444	0,233	**0,001**
500	18	3,6%	5%		0,131	0,320	0,444	0,448	0,097
500	8	**1,6%**	1%	60	0,215	0,414	0,737	0,672	0,320
500	9	1,8%	2%		0,745	0,830	0,705	0,902	0,816
500	15	**3,0%**	5%		**0,027**	0,145	0,629	0,304	0,086
500	5	1,0%	1%	120	1,000	1,000	0,834	0,974	0,949
500	11	2,2%	2%		0,753	0,836	0,431	0,711	0,197
500	18	3,6%	5%		0,131	0,320	0,782	0,578	0,316
500	6	1,2%	1%	250	0,663	0,774	0,204	0,426	**0,000**
500	7	1,4%	2%		0,311	0,505	0,247	0,407	**0,027**
500	15	**3,0%**	5%		**0,027**	0,145	0,076	0,071	**0,001**
500	4	0,8%	1%	500	0,641	0,759	0,867	0,937	0,883
500	4	**0,8%**	2%		**0,029**	0,151	0,867	0,351	0,160
500	14	**2,8%**	5%		**0,014**	0,106	0,205	0,120	**0,008**

Źródło: Opracowanie własne.

Podsumowanie

Zaprezentowane testy wyraźnie różnią się mocą – test Kupca wielokrotnie uznaje model za poprawnie oszacowany, podczas gdy test DQ odrzuca hipotezę głoszącą nieobciążoność modelu. Cztery dodatkowo przedstawione testy dają bardziej obszerną informację w porównaniu z testem Kupca. Pozwalają zidentyfikować powód odrzucenia modelu VaR – brak zgodności liczby przekroczeń z zakładanym poziomem istotności lub brak niezależności liczby przekroczeń. Ponadto test DQ posiada zauważalnie wyższą moc i reaguje zarówno na brak niezależności liczby przekroczeń, jak i na niezgodność liczby przekroczeń z założonym poziomem istotności dla modelu VaR. W pracy zaprezentowano także metodę regresji kwantylowej do wyznaczania prognoz VaR, która pozwoliła znacznie częściej osiągnąć nieobiążoność (test *DQ*) w porównaniu z metodą wariancji i kowariancji oraz metodą historyczną.

Literatura

Christoffersen P.F. (1998), *Evaluating Interval Forecasts*, „International Economic Review", Vol. 39, Iss. 4, s. 841-862.

Engle R.F., Manganelli S. (2004), *CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles*, „Journal of Business and Economic Statistics", Vol. 22, Iss. 4, s. 367-381.

Giacomini R., Komunjer I. (2005), *Evaluation and Combination of Conditional Quantile Forecasts*, „Journal of Business and Economic Statistics", Vol. 23, Iss. 4, s. 416-431.

Hoel P.G. (1954), *A Test for Markov Chains*, „Biometrika", Vol. 41, s. 430-433.

Kupiec P. (1995), *Techniques for Verifying the Accuracy of risk Measurement Models*, „Journal of Derivatives", Vol. 3, No. 2, s. 73-84.

SOME UNBIASEDNESS TEST ON THE EXAMPLE OF VALUE AT RISK

Summary: Article contain comparative analysis of unbiasedness test in calculation of Value at Risk. Author presented selected tests which verify number and the independence of exceedances.

Keywords: unbiasedness, risk, Value at Risk.