

Stanisław Jędrusik*, Andrzej Paliński*, Wojciech Chmiel**,
Piotr Kadłuczka**

Testowanie wsteczne modeli wartości narażonej na stratę

Backtesting Value at Risk Models

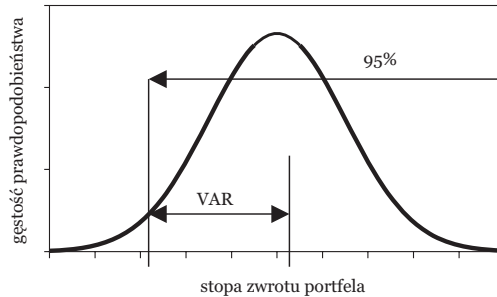
1. Wstęp

Metoda Wartości Narażonej na Stratę VaR (*Value-at-Risk*) znajduje zastosowanie nie tylko w praktyce bankowej, ale również w innych obszarach wymagających oszacowania ryzyka i efektywnego nim zarządzania. Do sukcesu tej metody w dużym stopniu przyczyniło się opublikowanie i upowszechnienie przez jeden z największych banków inwestycyjnych JP Morgan własnego systemu analizy i zarządzania ryzykiem *RiskMetrics*, opartego właśnie na metodologii VaR. W chwili obecnej bank JP Morgan udostępnia kilka wariantów podstawowego systemu, lepiej przystosowanych do konkretnych dziedzin (np. *CreditMetrics*). Powyższe cechy metody VaR znalazły odzwierciedlenie w dyrektywie Bazylejskiego Komitetu do spraw Nadzoru Bankowego rekomendującej metodę VaR jako dopuszczalną metodę pomiaru ryzyka finansowego w działalności bankowej.

Idea metody VaR polega na obliczeniu tzw. potencjalnej straty portfela. W jej wyniku otrzymuje się poziom straty portfela, który może być wyrażony w jednostkach pieniężnych, przy czym straty wyższe od obliczonej mogą się pojawiać z prawdopodobieństwem nie większym niż określona początkowo wartość – zwykle jest to 1% lub 5%, co przedstawia rysunek 1.

* Wydział Zarządzania, Akademia Górniczo-Hutnicza w Krakowie

** Wydział Elektrotechniki, Automatyki, Informatyki i Elektroniki, Akademia Górniczo-Hutnicza w Krakowie



Rys. 1. Ilustracja metody VaR

Dyrektywa, o której mowa wyżej, nie preferuje konkretnego modelu VaR rozumianego jako metoda uzyskiwania rozkładu prawdopodobieństwa zysku/straty portfela. Banki w tej dziedzinie mogą korzystać z modeli opracowanych przez swoich ekspertów, bądź też oprzeć się na bogatej literaturze przedmiotu. Dyrektywa ta precyzuje natomiast wymagania stawiane potencjalnym modelom VaR odnośnie uzyskiwanych wyników.

W poprawnie funkcjonującym modelu VaR w dłuższej perspektywie czasowej liczba przekroczeń poziomu VaR w stosunku do ogólnej liczby notowań powinna oscylować wokół poziomu założonego na wstępie. Znaczne przekroczenie tego poziomu oznacza w konsekwencji, że model VaR zaniża poziom ryzyka, i tym samym należy go odrzucić lub zmodyfikować. Z kolei wartości niższe od założonych oznaczają przeszacowanie ryzyka, co prowadzi do nieuzasadnionego podwyższenia kosztów zabezpieczenia portfela.

VaR w ujęciu statystycznym jest kwantylem rzędu p rozkładu prawdopodobieństwa prognozowanej stopy zwrotu portfela. Jeśli rozkład stóp zwrotu aktywów wchodzących w skład portfela dobrze opisuje rozkład normalny, to poziom VaR może być wyznaczony na podstawie zależności:

$$VaR_t = \begin{cases} 1,65 \cdot \sigma_t & \text{dla } p = 0,05 \\ 2,33 \cdot \sigma_t & \text{dla } p = 0,01 \end{cases} \quad (1)$$

gdzie σ_t oznacza odchylenie standardowe rozkładu prawdopodobieństwa prognozowanej stopy zwrotu portfela w chwili t .

2. Testowanie wsteczne

Metody oceny poprawności modelu VaR są bardzo istotnym elementem każdego systemu zarządzania ryzykiem finansowym. W literaturze przedmiotu tego rodzaju analiza nosi nazwę testowania wstecznego [2, 7].

Do chwili obecnej opracowano wiele testów, które można zaliczyć do kategorii „testowania wstecznego” [1, 3, 5, 6]. W niniejszym artykule skoncentrowano się na dwóch z nich – uznawanych za przełomowe w opinii większości – a mianowicie na teście Kupca [8] i teście Christoffersena [4]. Celem badań było sprawdzenie, który z dwóch modeli (Model błędzenia przypadkowego i Model GARCH) generuje lepsze prognozy VaR. Wyniki przeprowadzonych eksperymentów na bazie portfeli z polskiego rynku finansowego poprzedzone zostały omówieniem matematycznych podstaw obydwu testów.

2.1. Test Kupca

Z definicji wynika, że przekroczenia poziomu VaR powinny pojawiać się z częstotliwością równą p . Łatwo zauważyć, że prawdopodobieństwo uzyskania x przekroczeń w próbie o długości n – przy założeniu, że model jest poprawny – opisane jest rozkładem dwumianowym i wynosi

$$pr(x) = \binom{n}{x} p^x (1-p)^{n-x} \quad (2)$$

Idea testu Kupca opiera się na statystyce ilorazu wiarygodności. Funkcja wiarygodności przy założeniu poprawności modelu (hipoteza zerowa) ma postać:

$$L_0 = p^x (1-p)^{n-x} \quad (3)$$

W wypadku hipotezy alternatywnej funkcja wiarygodności wynosi

$$L_A = \hat{p}^x (1-\hat{p})^{n-x} \quad (4)$$

gdzie $\hat{p} = x/n$.

Statystykę opartą na ilorazie wiarygodności można zatem zapisać w postaci

$$LR_{uc} = 2[\log L_A - \log L_0] \quad (5)$$

Łatwo udowodnić, że zdefiniowana wyżej statystyka ma asymptotycznie rozkład χ^2 z jednym stopniem swobody. Hipotezę zerową odrzuca się, jeśli wartość tej statystyki przekroczy poziom krytyczny rozkładu $\chi^2(1)$.

2.2. Test Christoffersena

Test Kupca pozwala oszacować, czy liczba przekroczeń poziomu VaR znajduje się w dopuszczalnym przedziale. Wyniki tego testu nic jednak nie mówią o rozkładzie w czasie kolejnych przekroczeń. W poprawnie funkcjonującym modelu

przekroczenia poziomu VaR powinny mieć charakter losowy, tj. następować niezależnie od siebie. Brak niezależności, objawiający się często nadmierną koncentracją przekroczeń na pewnym odcinku czasu i jednocześnie nadmiernym rozproszeniem na innym odcinku – oznacza, że model VaR niezbyt dobrze odwzorowuje cechy analizowanego procesu. Warto w tym miejscu zaznaczyć, że brak niezależności nie oznacza automatycznie, iż model nie będzie spełniał kryterium Kupca. Te dwie cechy są do pewnego stopnia niezależne.

Zaproponowany przez Christoffersena test niezależności przekroczeń opiera się również na ilorazie wiarygodności. Jeśli I_t oznacza zmienną losową określoną wzorem:

$$I_{t+1} = \begin{cases} 1 & \text{dla } R_{t+1} < -VaR_t(p) \\ 0 & \text{w p.p.} \end{cases} \quad (6)$$

gdzie R_{t+1} oznacza stopę zwrotu portfela w okresie $t+1$,

to przy założeniu niezależności kolejnych przekroczeń zmienne losowe I_t są również niezależne i mają rozkład dwumianowy o wartości oczekiwanej p .

Hipotezę alternatywną można sformułować na wiele różnych sposobów. P.F. Christoffersen jako alternatywę przyjął proces Markowa pierwszego rzędu. Jeśli $\{I_t\}$ jest procesem Markowa, to prawdopodobieństwa przejścia $pr(I_{t+1}|I_t)$ opisane są następującą macierzą przejść:

$$\begin{bmatrix} 1 - \pi_{01} & \pi_{01} \\ 1 - \pi_{11} & \pi_{11} \end{bmatrix} \quad (7)$$

gdzie $\pi_{ij} = pr(I_{t+1} = j | I_t = i)$.

Funkcja wiarygodności hipotezy alternatywnej przyjmuje zatem postać

$$L_A = (1 - \pi_{01})^{T_{00}} \pi_{01}^{T_{01}} (1 - \pi_{11})^{T_{10}} \pi_{11}^{T_{11}} \quad (8)$$

gdzie T_{ij} oznacza liczbę przejść ze stanu i do stanu j ,

$$\pi_{01} = \frac{T_{01}}{T_{00} + T_{01}} \quad (9)$$

$$\pi_{11} = \frac{T_{11}}{T_{10} + T_{11}} \quad (10)$$

Przy założeniu hipotezy zerowej przekroczenia poziomu VaR powinny mieć stałą warunkową średnią, co implikuje:

$$\pi_{01} = \pi_{11} = p \quad (11)$$

Funkcję wiarygodności w tym przypadku można zapisać

$$L_0 = (1-p)^{T_{00}+T_{10}} p^{T_{01}+T_{11}} \quad (12)$$

a statystyka oparta na ilorazie wiarygodności ma postać

$$LR_{ind} = 2[\log L_A - \log L_0] \quad (13)$$

Podobnie jak w wypadku testu Kupca, rozkład statystyki LR_{ind} jest zbieżny do rozkładu χ^2 z jednym stopniem swobody.

3. Wyniki badań

Poniżej podjęto próbę oceny dwóch modeli VaR. Pierwszym z nich był to klasyczny model błędzenia przypadkowego. Formalnie rzecz biorąc za pomocą modelu błędzenia przypadkowego opisuje się szereg wartości portfela w kolejnych chwilach czasu. Ponieważ szereg stóp zwrotu jest jego pochodną, więc użycie tego terminu wydaje się przynajmniej częściowo uzasadnione

$$y_t = C + \varepsilon_t \quad (14)$$

gdzie:

y_t – logarytmiczna stopa zwrotu portfela w chwili t ,

ε_t – ciąg niezależnych zmiennych losowych o rozkładzie normalnym i odchyleniu standardowym σ_t niezmiennym w czasie.

W drugim modelu dopuszczono zmienność warunkowej wariancji procesu zgodnie z modelem GARCH

$$\sigma_t^2 = \kappa + G \sigma_{t-1}^2 + A \varepsilon_{t-1}^2 \quad (15)$$

Porównanie modelu ze stałą i zmienną warunkową wariancją pozwoliło na zbadanie, czy uwzględnienie zmienności warunkowej wariancji wpływa na jakość prognozy VaR. Badania przeprowadzono na grupie portfeli z polskiego rynku finansowego. Pod uwagę wzięto: portfele walutowe: EUR/PLN i USD/PLN, portfele akcyjne: WIG i WIG 20 oraz portfele funduszy inwestycyjnych: NFI01 i NFO1.

Długość próbkki wahała się od 1893 notowań w wypadku szeregu EUR/PLN do 3404 notowań w wypadku szeregu USD/PLN. Obliczenia przeprowadzono dla dziennych stóp zwrotu każdego portfela. Do estymacji parametrów użyto 250 początkowych stóp zwrotu. Dla modelu błędzenia przypadkowego (RW) oszacowanie parametrów sprowadzało się do obliczenia odchylenia standardowego. Aby

uzyskać lepsze dopasowanie modelu do danych, procedurę estymacji parametrów powtarzano co 25 dni. Następnie na podstawie wzoru (1) (tj. przy założeniu normalności rozkładów) obliczono jednodniowe prognozy VaR dla wszystkich pozostałych dni.

Ostatnim krokiem było obliczenie wartości statystyk Kupca i Christoffersena wszystkich portfeli i wszystkich modeli VaR. Oszacowano wartości statystyk dla najczęściej przyjmowanych wartości $p = 0,05$ i $p = 0,01$. Wyniki obliczeń zawiera tabela 1. Przedstawiono w niej wyniki obydwu testów oraz wynik testu łącznego:

$$LR_{cc} = LR_{uc} + LR_{ind} \quad (16)$$

Statystyka LR_{cc} ma również rozkład χ^2 , ale z dwoma stopniami swobody.

Tabela 1

Wyniki testów Kupca i Christoffersena dla 6 portfeli inwestycyjnych z polskiego rynku finansowego

Portfel	Model	$p = 0,01$			$p = 0,05$		
		LR_{uc}	LR_{ind}	LR_{cc}	LR_{uc}	LR_{ind}	LR_{cc}
WIG	Random Walk	3,91	8,71	12,62	0,77	17,82	18,58
	GARCH(1,1)	3,14	1,86	5,00	0,30	16,25	16,55
WIG20	Random Walk	2,82	0,84	3,66	2,63	1,72	4,35
	GARCH(1,1)	3,20	0,06	3,26	0,13	1,40	1,53
NFI 1	Random Walk	10,75	2,25	13,00	0,26	1,27	1,53
	GARCH(1,1)	11,49	1,03	12,53	1,06	0,37	1,43
NFI 2	Random Walk	12,27	6,58	18,84	0,01	4,48	4,49
	GARCH(1,1)	6,14	1,75	7,89	0,09	0,53	0,62
EUR	Random Walk	0,01	0,15	0,16	2,69	0,02	2,70
	GARCH(1,1)	0,05	0,12	0,17	0,37	0,30	0,66
USD	Random Walk	1,93	0,89	2,82	2,70	4,97	7,67
	GARCH(1,1)	3,64	3,14	6,78	0,63	1,26	1,89

Źródło: opracowanie własne.

Analizując wyniki zawarte w tabeli 1, należy stwierdzić, że w większości wypadków otrzymane wartości statystyk są mniejsze niż wartości krytyczne rozkładów $\chi^2(1)$ i $\chi^2(2)$, które wynoszą odpowiednio: 3,842 i 5,992 przy poziomie istotności 95%. Dla modelu RW i statystyk LR_{uc} i LR_{ind} w 7 przypadkach (na 24 możliwe) nastąpiło przekroczenie wartości krytycznych.

Model GARCH dla tych samych statystyk daje jedynie 3 przypadki przekroczenia wartości krytycznej. Podobne wyniki uzyskano dla testu łącznego LR_{cc} . W przypadku modelu RW odnotowano 5 przekroczeń na ogólną liczbę 12, a dla modelu GARCH – 4 przekroczenia poziomu krytycznego. W praktyce oznacza to, że w odniesieniu do większości portfeli nie ma podstaw do odrzucenia hipotezy zerowej, a tym samym obydwa modele VaR są poprawne. Z powyższej analizy wynika również, że wyraźnie lepsze wyniki uzyskuje się, stosując model GARCH niż RW.

Modele przyjęte do analizy nie były w żaden sposób formalnie dobierane. Nie ulega wątpliwości, że w rodzinie GARCH można znaleźć model znacznie lepiej pasujący do danych niż klasyczny proces GARCH(1,1). Celem badań nie było jednak poszukiwanie najlepszego modelu GARCH dla badanych portfeli, ale weryfikacja tezy, czy uwzględnienie zmienności warunkowej wariancji pozwala w znaczący sposób poprawić jakość prognozy VaR. Z tych powodów warto spojrzeć na wyniki zawarte w tabeli 1 pod innym kątem, tj. istotnie jest nieprzekroczenie poziomu krytycznego, ale dla którego modelu uzyskuje się lepsze (czyli niższe) wartości odpowiednich statystyk. Na łączną liczbę 36 obliczonych wartości statystyk jedynie w 10 przypadkach modele RW dominują nad modelami GARCH. W rozbiciu na konkretne statystyki sytuacja ta nie zmienia się zasadniczo, aczkolwiek pojawiają się pewne charakterystyczne cechy. W wypadku LR_{uc} odnotowano 6 przypadków, kiedy model RW był lepszy od GARCH i tyle samo przypadków, kiedy było odwrotnie. Oznacza to, że obydwa modele jednakowo dobrze prognozują liczbę przekroczeń poziomu VaR. Gdyby takie wyniki potwierdziły się dla większej liczby portfeli, oznaczałoby to, że model GARCH nie daje istotnych korzyści w kategoriach kryterium Kupca. Statystyka LR_{ind} potwierdza już jednak regułę, że modele GARCH są lepsze niż RW. Na 12 możliwych porównań jedynie w 2 przypadkach model RW dawał lepsze wyniki niż GARCH.

Na przykładzie tej statystyki widać przewagę modelu GARCH. Zastosowanie modelu GARCH nie pozwoliło na uzyskanie wyraźnie lepszych wartości statystyki LR_{uc} , ale dla statystyki LR_{ind} sytuacja jest już bardziej czytelna. Ponieważ w modelu RW warunkowa wariancja jest stała, to przewaga modelu GARCH musi wynikać z faktu, że szeregi stóp zwrotu nie są homoskedastyczne, a poziom zmienności warunkowej wariancji jest duży.

Ze względu na różnorodność aktywów finansowych wchodzących w skład portfeli warto spojrzeć na wyniki z tabeli 1 jeszcze z innego punktu widzenia, a mianowicie: dla których portfeli model GARCH jest lepszy niż RW i odwrotnie? Liczba analizowanych portfeli w poszczególnych kategoriach nie upoważnia, co prawda, do wysuwania kategoriycznych wniosków, ale nie sposób nie zauważyć, że dla portfeli giełdowych (WIG, WIG20) i portfeli funduszy inwestycyjnych (NFI01 i NFI02) model GARCH daje znacząco lepsze wyniki niż RW. Sytuacja

portfeli walutowych jest niejednoznaczna – żaden z modeli nie dominuje wyraźnie. Należy podejrzewać, że jest to wynik bądź niedopasowania modelu GARCH(1,1) do danych walutowych, bądź też zbyt niskiego stopnia zmienności warunkowej wariancji w szeregach walutowych.

Reasumując, wyniki przeprowadzonych badań potwierdziły tezę o przewadze modelu GARCH nad modelem RW w prognozowaniu poziomu VaR. Zaobserwowane prawidłowości, mimo iż liczba badanych portfeli nie była zbyt duża, wydają się zgodne z teoretycznymi przesłankami odnośnie badanych modeli. Nie ulega wątpliwości, że przewaga modelu GARCH jest wynikiem efektu zmienności warunkowej wariancji, obserwowanego na większości szeregów finansowych. Klasyczny model RW tego zjawiska nie uwzględnia, stąd prognoza VaR uzyskana na jego podstawie jest nieco gorsza niż na podstawie modelu GARCH.

Literatura

- [1] Berkowitz J., *Testing density forecasts, with applications to risk management*, „Journal of Business and Economic Statistics” 2001, No. 19, s. 465–474.
- [2] Berkowitz, J., O’Brien J., *How Accurate are the Value-at-Risk Models at Commercial Banks?*, „Journal of Finance” 2002, No. 57, s. 1093–1111.
- [3] Campbell S.D., *A Review of Backtesting and Backtesting Procedures* [working paper], Finance and Economics Discussion Series 2005.
- [4] Christoffersen P.F., *Evaluating Interval Forecasts*, „International Economic Review” 1998, No. 39, s. 841–862.
- [5] Christoffersen P.F., Pelletier D., *Backtesting Value-at-Risk. A Duration-Based Approach*, „Journal of Financial Econometrics” 2004, No. 2, s. 84–108.
- [6] Hass M., *New Methods in Backtesting*, Financial Engineering Research Center, Bonn 2001.
- [7] Jędrusik S., Paliński A., *Analiza porównawcza modeli wartości narażonej na stratę (VaR) dla portfela akcji*. [w:] idem, *Zarządzanie firmą – teoria i praktyka. Wybrane zagadnienia*, Kraków 2001, s. 157–169.
- [8] Kupiec, P., *Techniques for Verifying the Accuracy of Risk Management Models*, „Journal of Derivatives” 1995, No. 2, s. 73–84.