

Ryszard Doman*

**WYKORZYSTANIE DANYCH WYSOKIEJ CZĘSTOTLIWOŚCI
W PROGNOZOWANIU ZMIENNOŚCI
POLSKICH INDEKSÓW GIEŁDOWYCH
ZA POMOCĄ MODELI ZMIENNOŚCI STOCHASTYCZNEJ**

Streszczenie. Modele zmienności stochastycznej (SV) stanowią drugą, obok bardziej znanych modeli typu GARCH, klasę modeli wykorzystywanych do prognozowania zmienności instrumentów finansowych. W przeciwieństwie do modeli rodziny GARCH, w modelach SV ewolucja zmienności w czasie jest opisywana za pomocą dwóch nieskorelowanych procesów stochastycznych. W niniejszym artykule modele SV są stosowane do prognozowania dziennej zmienności indeksów Giełdy Papierów Wartościowych w Warszawie. Otrzymywane prognozy odnoszone są do dziennej zmienności zrealizowanej, rozumianej jako suma kwadratów zwrotów śróddziennych. Ponadto badany jest wpływ, jaki na jakość prognoz ma wprowadzenie do modelu SV dziennej zmienności zrealizowanej jako dodatkowej zmiennej objaśniającej.

Słowa kluczowe: prognozowanie, zmienność stochastyczna, zmienność zrealizowana, dane wysokiej częstotliwości.

1. WSTĘP

Problematyka związana z pojęciem zmienności instrumentów finansowych należy do najaktywniej rozwijających się w ciągu ostatnich dziesięciu lat obszarów badawczych w dziedzinie ekonometrii finansowych szeregów czasowych. Definiowana na ogół jako wariancja warunkowa stopy zwrotu z instrumentu finansowego lub pierwiastek z tej wariancji, zmienność jest jednym z najważniejszych parametrów w tak ważnych dziedzinach finansów, jak wycena pochodnych instrumentów finansowych, alokacja portfela inwestycji czy zarządzanie ryzykiem. Okoliczności te wymuszają potrzebę konstruowania dobrych modeli zmienności. Na podstawie prowadzonych od dłuższego czasu badań empirycznych ukształtował się zbiór powszechnie akceptowanych

*Dr, Wydział Matematyki i Informatyki, Uniwersytet im. Adama Mickiewicza w Poznaniu.

opinii na temat zmienności instrumentów finansowych. Najistotniejsze z nich stwierdzają, że zmienność sama fluktuuje w czasie, wykazuje trwałość w tym sensie, że dzisiejsze zwroty mają duży wpływ na przyszłą zmienność oraz charakteryzuje się powracaniem do średniej, co oznacza, że dzisiejsza informacja nie ma wpływu na prognozy zmienności, gdy horyzont prognozy dąży do nieskończoności. Dobry model zmienności powinien oczywiście uwzględniać wymienione fakty. Najistotniejszym jednak żądaniem, któremu model zmienności pretendujący do miana dobrego powinien sprostać, są jego zadowalające zdolności prognostyczne. Chronologicznie pierwszymi modelami zmienności uwzględniającymi heteroskedastyczność były modele ARCH, zaproponowane w roku 1982 przez Engle'a (1982). W roku 1986 w pracy Bollersleva (1986) pojawiły się ich uogólnienia w postaci modeli GARCH. W modelach tych zmienność instrumentu finansowego w danym dniu jest funkcją deterministyczną kwadratów jego opóźnionych zwrotów oraz swoich opóźnionych wartości. W kolejnych latach rodzina modeli GARCH rozrosła się do pokaźnych rozmiarów, obejmując modele uwzględniające w coraz większym stopniu odnotowywane empirycznie cechy zmienności. Równoległe wyrosła konkurencyjna klasa modeli zmienności stochastycznej (SV), które charakteryzują się tym, że modelowana przez nie zmienność zależy nie tylko od swoich opóźnionych wartości, ale również od niezależnych zaburzeń losowych. Przegląd prac obrazujących rozwój badań związanych z modelami zmienności stochastycznej zawiera m. in. artykuł, którego autorami są Ghysels, Harvey i Renault (1996). Standardowy model SV zawiera jedynie trzy parametry wymagające estymacji. Jest to niewiele w porównaniu z rozbudowanymi modelami klasy GARCH, gdzie do niełatwych problemów należy adekwatna specyfikacja modelu. W przypadku modeli zmienności stochastycznej istotnym wyzwaniem jest natomiast estymacja modelu, co wynika z niemożności bezpośredniego stosowania metody największej wiarygodności.

Jak było wspomniane, *experimentum crucis* dla modelu zmienności powinny być jego zdolności prognostyczne. Do czasu ukazania się pracy Andersena i Bollersleva (1998) panowało powszechne przekonanie, że prognozy zmienności instrumentu finansowego dostarczane przez modele parametryczne charakteryzują się słabą jakością. Prognozy te były jednak odnoszone do, co prawda nieobciążonego, ale zanieczyszczonego szumem, estymatora zmienności dziennej, jakim jest kwadrat zwrotu dziennego. Andersen i Bollerslev zaproponowali i uzasadnili teoretycznie alternatywne podejście do mierzenia zmienności, w którym rolę kwadratu zwrotu dziennego przejęła suma kwadratów zwrotów śróddziennych, odpowiadających obserwacjom o większej częstotliwości. Przy tym nowym rozumieniu zmienności, nazwanej zmiennością zrealizowaną, okazało się, że prognozy dostarczane przez parametryczne modele zmienności są radykalnie lepsze. Zwiększająca

się dostępność danych finansowych o częstotliwości śróddziennej zainspirowała również badania nad ich potencjalnym wykorzystaniem jako źródła informacji do prognozowania na okres poza próbą. W pracach Blaira, Poon i Taylora (2001) oraz Martensa (2002) wykazano zdecydowaną poprawę w prognozowaniu zmienności za pomocą modeli GARCH jako konsekwencję użycia zmienności zrealizowanej w charakterze dodatkowej zmiennej objaśniającej w równaniu wariancji warunkowej. Wyniki Hol i Koopmana (2002) potwierdziły poprawę jakości prognoz również w przypadku modeli SV, po rozszerzeniu ich o informacje o zmienności śróddziennej. W niniejszym opracowaniu badamy wpływ, jaki na jakość prognoz zmienności dziennej dwóch polskich indeksów giełdowych WIG i WIG20, otrzymywanych za pomocą modelu SV, ma włączenie do równań modelu opóźnionych wartości zmienności zrealizowanej. Prognozy zmienności odnoszone są do zmienności zrealizowanej szacowanej za pomocą zwrotów 20-minutowych oraz jej zmodyfikowanej wersji nie uwzględniającej bezpośrednio tzw. zwrotu nocnego.

2. MODELE ZMIENNOŚCI STOCHASTYCZNEJ

Niech $r_t = 100(\ln(P_t) - \ln(P_{t-1}))$ oznacza procentowy zwrot logarytmiczny z badanego instrumentu finansowego, gdzie P_t jest ceną instrumentu w dniu t . Jeśli przez Ω_{t-1} oznaczymy zbiór informacji na temat r_t dostępnych do dnia $t-1$ włącznie, to zmienność instrumentu finansowego w dniu t definiuje się zazwyczaj jako wariancję warunkową

$$\sigma_t^2 = \text{Var}(r_t | \Omega_{t-1}) = E((r_t - \mu_t)^2 | \Omega_{t-1}) \quad (1)$$

gdzie $\mu_t = E(r_t | \Omega_{t-1})$ jest średnią warunkową. W związku z tym przy modelowaniu zmienności będziemy rozważać szereg $y_t = r_t - \mu_t$. W przypadku danych empirycznych średnia warunkowa jest na ogół pomijana jako nieistotna lub modeluje się ją jako stacjonarny proces ARMA(p, q). W uogólnionym modelu autoregresyjnym heteroskedastyczności warunkowej, GARCH(1, 1), postuluje się, by

$$y_t = \sigma_t \varepsilon_t \quad (2)$$

$$\sigma_t^2 = \omega + \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2$$

gdzie ε_t jest szeregiem niezależnych standaryzowanych zmiennych losowych o jednakowym rozkładzie. W alternatywnym modelu zmienności stochastycznej (SV), wprowadzonym przez Melino i Turnbulla (1990), zmienność σ_t^2 nie

jest funkcją jedynie elementów zbioru informacji Ω_{t-1} , ale zależy od dodatkowych zaburzeń. Podstawowy model SV jest określony równaniami

$$\begin{aligned} y_t &= \sigma_t \varepsilon_t \\ \sigma_t^2 &= \sigma^{*2} \exp(h_t) \end{aligned} \quad (3)$$

$$h_t = \varphi h_{t-1} + \sigma_\eta \eta_t$$

gdzie ε_t i η_t mają standardowy rozkład normalny oraz są szeregowo i wzajemnie niezależne. Opisany model (3) jest odpowiednikiem dla procesów z czasem dyskretnym modelu z czasem ciągłym, wykorzystywanego w pracy Hulla i White'a (1987), dotyczącej wyceny opcji. Heurystyczna interpretacja modelu SV jest taka, że zaburzenia η_t reprezentują intensywność napływu nowych informacji, mierzoną przez σ_t , natomiast ε_t reprezentują zaburzenia związane z ilością nowych informacji. Jeśli $|\varphi| < 1$, to h_t jest stacjonarnym procesem AR(1), przy czym $h_t \sim N(0, \sigma_\eta^2 / (1 - \varphi^2))$. Zatem σ_t^2 ma rozkład log-normalny ze średnią

$$E(\sigma_t^2) = \sigma^{*2} \exp\left(0,5 \frac{\sigma_\eta^2}{1 - \varphi^2}\right) \quad (4)$$

Obecność innowacji $\sigma_\eta \eta_t$ powoduje, że model SV jest bardziej elastyczny w porównaniu z modelami klasy GARCH, ale z drugiej strony fakt ten znacząco utrudnia estymację parametrów. Ponieważ σ_t^2 nie daje się zrekonstruować na podstawie historii procesu zwrotów, nie można stosować standardowych technik metody największej wiarygodności. Obecnie znanych i stosowanych jest kilka alternatywnych procedur estymacji. Stosuje się tutaj zarówno techniki bayesowskie (głównie metody Monte Carlo typu łańcuchów Markowa), jak i niebayesowskie (metoda quasi-największej wiarygodności, estymacja za pomocą filtrów Kalmana, uogólniona metoda momentów).

W naszych badaniach empirycznych przyjęto podejście, które zastosowali Koopman i Hol-Uspensky (2002). W największym skrócie, polega ono na użyciu metod Monte Carlo z funkcją ważności (por. Osiewalski, 2001) do konstrukcji funkcji wiarygodności dla modelu SV. Nieliniowa zależność pomiędzy logarytmiczną zmiennością h_t i zwrotem (z usuniętą średnią warunkową) y_t uniemożliwia wyliczenie funkcji wiarygodności za pomocą metod liniowych, takich jak filtr Kalmana. Funkcję wiarygodności dla modelu SV wyrażamy więc w postaci:

$$L(\psi) = p(y|\psi) = \int p(y, \theta|\psi) d\theta = \int p(y|\theta, \psi) p(\theta|\psi) d\theta \quad (5)$$

gdzie $\psi = (\varphi, \sigma_\eta, \sigma^{*2})'$, $\theta = (h_1, \dots, h_T)'$.

Jako funkcję ważności wybiera się warunkową gęstość gaussowską $g(\theta|y, \psi)$ pokrywającą się z rozwinięciem Taylora rzędu drugiego prawdziwego modelu wokół θ . Następnie za pomocą filtru Kalmana wylicza się funkcję wiarygodności dla przybliżonego liniowego modelu gaussowskiego. Okazuje się, że funkcja wiarygodności dla modelu SV jest równoważna funkcji wiarygodności modelu gaussowskiego, pomnożonej przez czynnik korygujący, który można estymować przy użyciu symulacji, losując z gaussowskiej gęstości $g(\theta|y, \psi)$.

Za pomocą przedstawionych metod prognozy 1-dniowe z modelu SV wylicza się jako:

$$E(\sigma_{T+1|T}^2) = \hat{\delta}^{*2} \exp(\hat{h}_{T+1|T} + 0,5p_{T+1|T}) \quad (6)$$

gdzie $\hat{\delta}^{*2}$ jest estymatorem największej wiarygodności dla σ^{*2} , $\hat{h}_{T+1|T}$ jest estymatorem dla h_{T+1} przy danych T obserwacjach, a $p_{T+1|T}$ jest jego błędem średniokwadratowym. Jeśli horyzont prognozy jest równy j dniom, to prognoza jest zdefiniowana wzorem:

$$E(\sigma_{T+j|T}^2) = \hat{\delta}^{*2} \exp(\hat{h}_{T+j|T} + 0,5p_{T+j|T}) \quad (7)$$

Estymatory $\hat{h}_{T+j|T}$ i $p_{T+j|T}$ wylicza się za pomocą metod symulacyjnych, wprowadzonych przez Durбина i Koopmana (2000), a dla $j \geq 2$, wartości $\hat{h}_{T+j|T}$ oraz $p_{T+j|T}$ wylicza się jako:

$$\begin{aligned} \hat{h}_{T+j|T} &= \hat{\varphi}^{j-1} h_{T+1|T}, \\ p_{T+j|T} &= \hat{\varphi}^{2(j-1)} p_{T+1|T} + \sum_{i=0}^{j-2} \hat{\varphi}^{2i} \sigma_{\eta}^2. \end{aligned} \quad (8)$$

Jeśli $N \rightarrow \infty$, to $E(\sigma_{T+N|T}^2)$ dąży do $\hat{\sigma}^{*2} \exp\left(0,5 \frac{\hat{\delta}_{\eta}^2}{1 - \hat{\varphi}^2}\right)$, czyli do estymatora wariancji bezwarunkowej.

Dzienny model SV można rozszerzyć, wprowadzając do równania logarytmicznej zmienności dodatkowe zmienne objaśniające. Jedną z możliwych specyfikacji jest model SVX_k , w którym równanie to przyjmuje postać:

$$h_t = \varphi h_{t-1} + \gamma(1 - \varphi L)x_{t-k} + \sigma_{\eta} \eta_t \quad (9)$$

gdzie L oznacza operator opóźnienia. Początkowa specyfikacja dla procesu logarytmicznej zmienności modelu SVX_k zakłada, że x_{1-k} jest stałe i znane, tak że $h_1 \sim N(\gamma x_{1-k}, \sigma_{\eta}^2/(1 - \varphi^2))$. Można również rozważać ograniczony model SVX_k nakładając warunek $\varphi = 0$. Model ten, w którym równanie logarytmicznej zmienności przybiera postać

$$h_t = \gamma x_{t-k} + \sigma_\eta \eta_t \quad (10)$$

będziemy oznaczać przez $SVR X_k$. Wprowadzenie zmiennej objaśniającej x_t w równaniu dla h_t nie wpływa na zależności nieliniowe między y_t i h_t . Zatem przedstawione wyżej metody estymacji i prognozowania dla modelu SV mogą być stosowane również dla modeli rozszerzonych. Szczegóły postępowania w tym przypadku opisane są w pracy Hol i Koopmana (2000).

3. ZMIENNOŚĆ ZREALIZOWANA

Jest znanym faktem, że prognozy zmienności dostarczane przez parametryczne modele zmienności typu GARCH czy SV, słabo aproksymują kwadraty faktycznych zwrotów dziennych. Sytuacja poprawia się jednak radykalnie, jeśli prognozy te odnosi się do tzw. dziennej zmienności zrealizowanej. Pojęcie to wraz z teoretycznym uzasadnieniem zostało wprowadzone przez Andersena i Bollersleva (1998). Załóżmy, że proces P_t ceny instrumentu finansowego jest procesem z czasem ciągłym, przy czym przedział długości 1 odpowiada jednemu dniu. Definiujemy szereg czasowy zwrotów logarytmicznych ceny P_t odpowiadający m obserwacjom w ciągu dnia wzorem:

$$r_{(m),t} = 100(\ln(P_t) - \ln(P_{t-1/m})) \quad (11)$$

gdzie $t = 1/m, 2/m, \dots$. Jeśli $m = 1$, to zwrot dzienny $r_{(1),t}$ oznaczamy przez r_t . W dalszym ciągu będziemy też oznaczać $p_t = 100 \ln(P_t)$. Jeśli proces p_t spełnia stochastyczne równanie różniczkowe

$$dp_t = \mu_t dt + \sigma_t dW_t \quad (12)$$

gdzie μ_t jest całkowalne, prognozowalne i o skończonej wariacji na $[t, t+1]$, σ_t jest ściśle dodatnie i całkowalne z kwadratem, a W_t jest standardowym procesem Wienera, to przy pewnych dodatkowych założeniach o σ_t z podstawowych własności całki Itô wynika, że

$$\text{Var}_t(r_{t+1}) = E_t \left(\int_0^1 \sigma_{t+\tau}^2 d\tau \right) \quad (13)$$

gdzie $\text{Var}_t(\cdot)$ i $E_t(\cdot)$ oznaczają warunkową wariancję i wartość oczekiwaną. Całka $\int_0^1 \sigma_{t+\tau}^2 d\tau$ określa tzw. zmienność scałkowaną, która jest odpowiednikiem zmienności w dniu $t+1$ w kontekście czasu ciągłego. Z drugiej strony, można wykazać, że

$$\int_0^1 \sigma_{t+\tau}^2 d\tau = p \lim_{m \rightarrow \infty} \left(\sum_{j=1 \dots m} r_{(m), t+j/m}^2 \right) \quad (14)$$

Równość (14) uzasadnia wprowadzenie pojęcia dziennej zmienności zrealizowanej, odpowiadającej m obserwacjom w ciągu dnia:

$$Real_t = \sum_{i=0}^{m-1} r_{(m), t-i/m}^2 \quad (15)$$

Z przedstawionej teorii wynika, że nieobserwowalną zmienność można aproksymować za pomocą sumy kwadratów zwrotów śróddziennych. W praktyce, w związku z efektami mikrostruktury rynkowej, przejścia granicznego nie można traktować dosłownie. W badaniach empirycznych stosuje się zazwyczaj dane o odstępach nie krótszym niż 5 minut. Liczne badania empiryczne, za którymi stoją również uzasadnienia teoretyczne, potwierdzają, że jakość prognoz zmienności dostarczanych przez modele parametryczne poprawia się radykalnie, gdy zamiast do kwadratów zwrotów dziennych, odnosi się je do zmienności zrealizowanej. W niniejszym artykule ze względu na stosunkowo niską płynność, występującą na GPW w Warszawie, badamy zwroty śróddzienne, odpowiadające obserwacjom o odstępach 20-minutowym. W związku ze specyfiką notowań giełdowych pierwszy zwrot śróddzienny, tzw. zwrot nocny, obejmujący okres od zamknięcia giełdy w dniu poprzednim do otwarcia w danym dniu, ma szczególny charakter. Sensowne jest zatem wyodrębnienie tego zwrotu. W dalszym ciągu zmienność zrealizowaną w dniu t uwzględniającą zwrot nocny, będziemy oznaczać $Real1_t$. Natomiast sumę kwadratów śróddziennych bez kwadratu zwrotu nocnego oznaczamy przez $Real2_t$. W niektórych badaniach (np. Martens, 2002) twierdzi się, że optymalnym estymatorem dziennej zmienności w przypadku danych giełdowych jest wielkość

$$Real3_t = (1 + c)Real2_t \quad (16)$$

Uwzględnia ona zwrot nocny dzięki przeskalowaniu, ale nie wprowadza do miary zmienności szumu związanego z tym zwrotem. Wśród ekspertów nie ma jednoznacznych opinii co do metody wyznaczania stałej c . W niniejszym opracowaniu za Hol i Koopmanem (2002) przyjmujemy $c = \frac{\sigma_{co}^2}{\sigma_{oc}^2}$, gdzie σ_{co}^2 oznacza wariancję zwrotu między kursem zamknięcia a kursem otwarcia następnego dnia, a σ_{oc}^2 jest wariancją zwrotu dziennego.

4. KRYTERIA OCENY JAKOŚCI PROGNOZ

Opisane w 2 paragrafie prognozy zmienności dziennej będziemy odnosić do dziennej zmienności zrealizowanej *Real1* oraz *Real3*. Do oceny jakości prognoz stosujemy następujące miary:

(a) współczynnik R^2 regresji Minzera-Zarnowitza:

$$RealS_{T+h+j} = a + bE(\sigma_{T+h+j|T+j}^2) + u_{T+h+j}, \quad j = 0, \dots, N-1 \quad (17)$$

(b) średni błąd bezwzględny (*MAE*):

$$MAE(h) = \frac{1}{N} \sum_{j=1}^{N-1} |E(\sigma_{T+h+j|T+j}^2) - RealS_{T+h+j}| \quad (18)$$

(c) błąd średni (*ME*):

$$ME(h) = \frac{1}{N} \sum_{j=1}^{N-1} |E(\sigma_{T+h+j|T+j}^2) - RealS_{T+h+j}| \quad (19)$$

(d) skorygowany błąd bezwzględny (*AMAE*):

$$AMAE(h) = \frac{1}{N} \sum_{j=0}^{N-1} \left| \frac{E(\sigma_{T+h+j|T+j}^2) - RealS_{T+h+j}}{E(\sigma_{T+h+j|T+j}^2) + RealS_{T+h+j}} \right| \quad (20)$$

(e) współczynnik rozbieżności Theila (*TIC*):

$$TIC(h) = \frac{\sqrt{\frac{1}{N} \sum_{j=0}^{N-1} (E(\sigma_{T+h+j|T+j}^2) - RealS_{T+h+j})^2}}{\sqrt{\frac{1}{N} \sum_{j=0}^{N-1} (E(\sigma_{T+h+j|T+j}^2))^2 + \frac{1}{N} \sum_{j=0}^{N-1} (RealS_{T+h+j})^2}} \quad (21)$$

We wzorach (17)–(21) T oznacza liczbę obserwacji w próbie, N jest liczbą prognoz na okres poza próbą, a h jest długością horyzontu prognozy. Ostatnie dwie miary są znormalizowane, tzn. przyjmują wartości pomiędzy zero a jeden.

5. OPIS DANYCH

Zbiór danych empirycznych badanych w tym artykule składa się z dziennych i śróddziennych obserwacji notowań indeksów WIG i WIG20 na Giełdzie Papierów Wartościowych w Warszawie. Dane dzienne pochodzą

z okresu od 29.12.1994 r. do 14.03.2003 r. Notowania śróddzienne z 20-minutowym odstępem pomiędzy obserwacjami dotyczą okresu od 17.11.2000 r. do 14.03.2003 r. Pierwsze notowanie śróddzienne pochodzi z godziny 10.05, a ostatnie z godziny 16.00. Jeśli notowania były okresowo zawieszane, to zostały uzupełnione ostatnią wartością z okresu przed zawieszeniem. Zwroty śróddzienne w dniu t różne od zwrotu nocnego liczone są według wzoru:

$$r^{t,d} = 100(\ln(P_{t,d}) - \ln(P_{t,d-1})) \quad (22)$$

gdzie $d = 1, \dots, 18$. Zwrot nocny określamy przez $r_{t,N} = 100(\ln(P_{t,0}) - \ln(P_{t-1,18}))$. Zwroty dzienne wyliczane są jako $r_t = 100(\ln(P_t) - \ln(P_{t-1}))$, gdzie P_t jest kursem zamknięcia.

6. WYNIKI EMPIRYCZNE

Zwroty z badanych indeksów do dnia 31.12.2002 r. zostały wykorzystane do estymacji modeli. Okres od 2.01. do 14.03.2003 r., obejmujący 52 dni sesyjne, był okresem prognozy poza próbą. Zestawienie szeregów zwrotów będących podstawą estymowanych modeli zawiera tabela 1. W tabeli 2 przedstawione są statystyki opisowe tych szeregów.

Tabela 1. Badane szeregi zwrotów i ich nazwy

Indeks	Nazwa szeregu zwrotów	Okres	Liczba obserwacji
WIG	WIGK	17.11.2000–14.03.2003	528 + 52 = 580
WIG	WIGD	02.01.1995–14.03.2003	1996 + 52 = 2048
WIG20	WIG20K	17.11.2000–14.03.2003	528 + 52 = 580
WIG20	WIG20D	02.01.1995–14.03.2003	1996 + 52 = 2048

Tabela 2. Statystyki opisowe badanych szeregów zwrotów

Szereg	Średnia	Min	Max	Wariancja	Skośność	Kurtoza
WIGK	-0,0271	-4,5160	4,4220	1,6228	0,2856	3,5717
WIGD	0,0302	-10,286	7,8933	3,0916	-0,1090	5,8562
WIG20K	-0,0670	-5,0446	5,4830	2,7619	0,3405	3,5880
WIG20D	0,0201	-14,161	13,709	4,3811	-0,0751	6,4367

Rozważane szeregi zwrotów dziennych, poza szeregiem WIGD, nie wykazywały obecności istotnie nietrywialnej średniej warunkowej. W przypadku szeregu WIGD usunięta średnia warunkowa miała postać $\mu_t = 0,1361y_{t-1}$. Do opisanych szeregów zwrotów dziennych były dopasowywane modele:

$$y_t = \sigma_t \varepsilon_t$$

$$\sigma_t^2 = \sigma^{*2} \exp(h_t)$$

$$\text{SV: } h_t = \varphi h_{t-1} + \sigma_\eta \eta_t \quad (23)$$

$$\text{SVX}_k: h_t = \varphi h_{t-1} + \gamma(1 - \varphi L)x_{t-k} + \sigma_\eta \eta_t$$

$$\text{SVRX}_k: h_t = \gamma x_{t-k} + \sigma_\eta \eta_t$$

przy czym jako dodatkową zmienną objaśniającą x_t próbowaliśmy zastosować $\ln(\text{Real1}_t)$, czyli logarytm sumy kwadratów wszystkich śróddziennych zwrotów 20-minutowych w dniu t oraz $\ln(\text{Real2}_t)$, gdzie Real2_t różni się od Real1_t brakiem kwadratu zwrotu nocnego. Okazało się, że wśród rozszerzonych modeli zmienności stochastycznej SVX_k i SVRX_k istotne parametry φ lub γ wystąpiły jedynie w SVRX_3 . Jest to zjawisko raczej nietypowe. W pracy Hol i Koopmana (2002) w ogóle nie bierze się pod uwagę faktu, że zmienność dzienna może być istotnie objaśniana przez zmienność zrealizowaną w inny sposób niż poprzez zależność od jej wielkości z dnia poprzedniego.

Estymacja modeli i wyliczenia prognoz zostały wykonane za pomocą programów autorstwa S.J. Koopmana, napisanych w języku Ox i korzystających z pakietu *SsfPack* 2.2 Koopmana, Shepharda i Doornika (1999). Ponieważ pewne parametry są estymowane w postaci transformowanej, zamiast błędów standardowych pod oszacowaniami podane są końce asymptotycznych (asymetrycznych) 95-procentowych przedziałów ufności (tabela 3 i 4).

Tabela 3. Wyestymowane parametry modeli SV dopasowanych do szeregu zwrotów indeksu WIG

Szereg	Model	φ	σ_η^2	σ^{*2}	γ
WIGD	SV	0,9551	0,0517	2,1588	
		0,9239–0,9739	0,0305–0,0877	1,7185–2,7120	
WIGK	SV	0,9526	0,0134	1,5094	
		0,7796–0,9913	0,0017–0,1043	1,1999–1,8989	
	SVRX_3 $x_t = \ln(\text{Real2}_t)$		0,2764	1,5220	0,21214 (0,0809)
			0,1167–0,6546	1,2796–1,8103	

Tabela 4. Wyestymowane parametry modeli SV dopasowanych do szeregu zwrotów indeksu WIG20

Szereg	Model	φ	σ_{η}^2	σ^{*2}	γ
WIGD	SV	0,9528	0,0517	3,1576	
		0,9218–0,9718	0,0311–0,0859	2,5401–3,9252	
WIGK	SV	0,9750	0,0059	2,5593	
		0,8737–0,9954	0,0009–0,0408	1,9570–3,3468	
	SVRX ₃ $x_t = \ln(Real1_t)$		0,2863	0,6493	0,1843
			0,1262–0,6493	1,7118–2,4957	(0,0663)

W każdym z wyestymowanych modeli SV logarytmiczna zmienność h_t jest procesem kowariancyjnie stacjonarnym, a bezwarunkowe wariancje wyliczone dla szeregów WIGD, WIGK, WIG20D, WIG20K ze wzoru

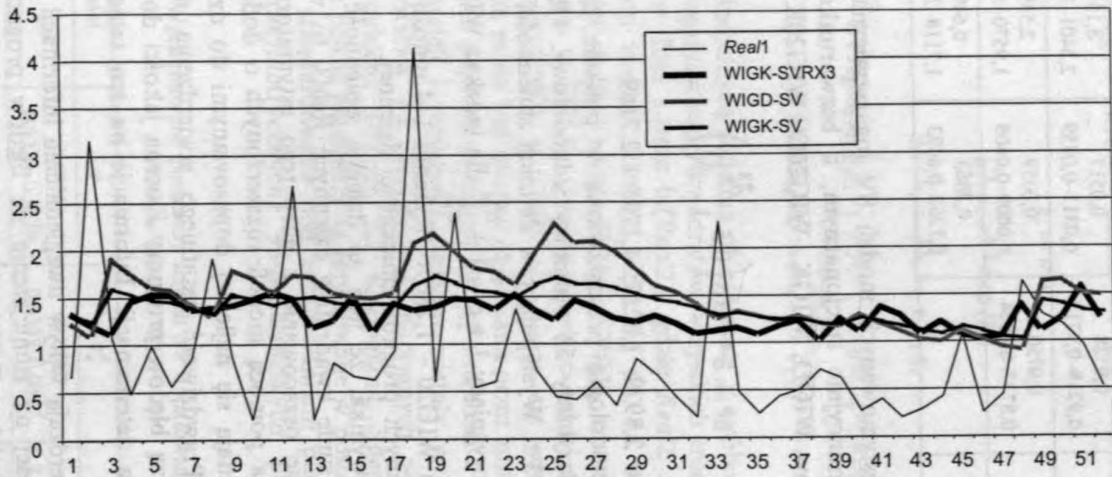
$$\sigma^2 = \sigma^{*2} \exp\left(0,5 \frac{\delta_{\eta}^2}{1 - \hat{\varphi}^2}\right) \quad (24)$$

są równe odpowiednio 2,8979; 1,6227; 4,1795 i 2,7169.

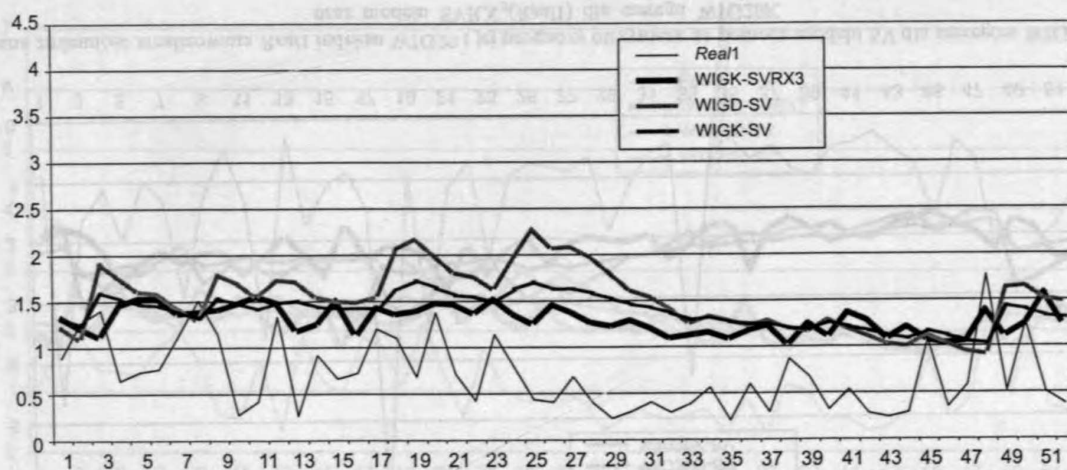
Posługując się metodologią przedstawioną w punkcie 2, dla każdego szeregu zwrotów wyliczyliśmy 52 prognozy jednodniowe. Objęło to okres od 2.01. do 14.03.2003 r. Występujący w definicji zmienności zrealizowanej *Real2* współczynnik korygujący $1 + c = 1 + \frac{\sigma_{co}^2}{\sigma_{oc}^2}$ dla indeksu WIG ma wartość 1,36814, a dla indeksu WIG20 – 1,28676. Rysunki 1–4 umożliwiają wstępną ocenę jakości otrzymanych prognoz zmienności dziennej.

Z rysunków 1–4 wynika, że jeśli za dzienną zmienność zrealizowaną przyjąć skorygowaną sumę kwadratów dziennych *Real3*, to wszystkie prognozy dają w zasadzie oszacowanie z góry tego estymatora zmienności. Prognozy wyliczone za pomocą modeli rozszerzonych o dodatkowe informacje śróddzienne wydają się najlepiej dopasowanymi do dziennej zmienności zrealizowanej. Prawdziwość ostatniego stwierdzenia należy jednak skonfrontować z analizą błędów prognozy i oceną jakości dopasowania za pomocą współczynnika determinacji. Informacje na ten temat są zawarte w tabelach 5 i 6.

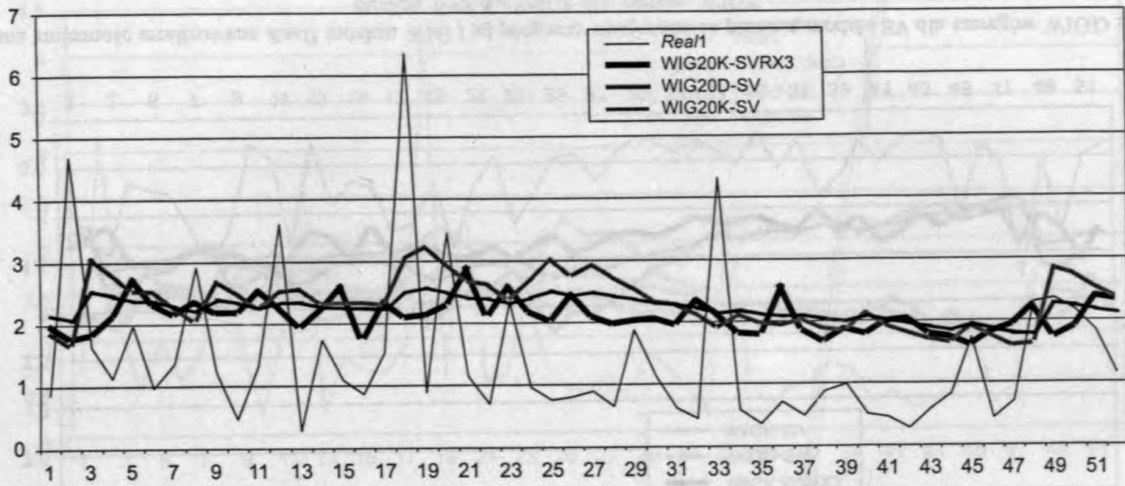
Tabele 5 i 6 potwierdzają wpływ uwzględnienia informacji o opóźnionej zmienności zrealizowanej na zmniejszenie się błędów prognozy. Niemniej jednak w przypadku indeksu WIG20 lepiej dopasowane, jeśli uwzględnić wielkość współczynnika R^2 , wydają się prognozy oparte na wieloletnich danych dziennych.



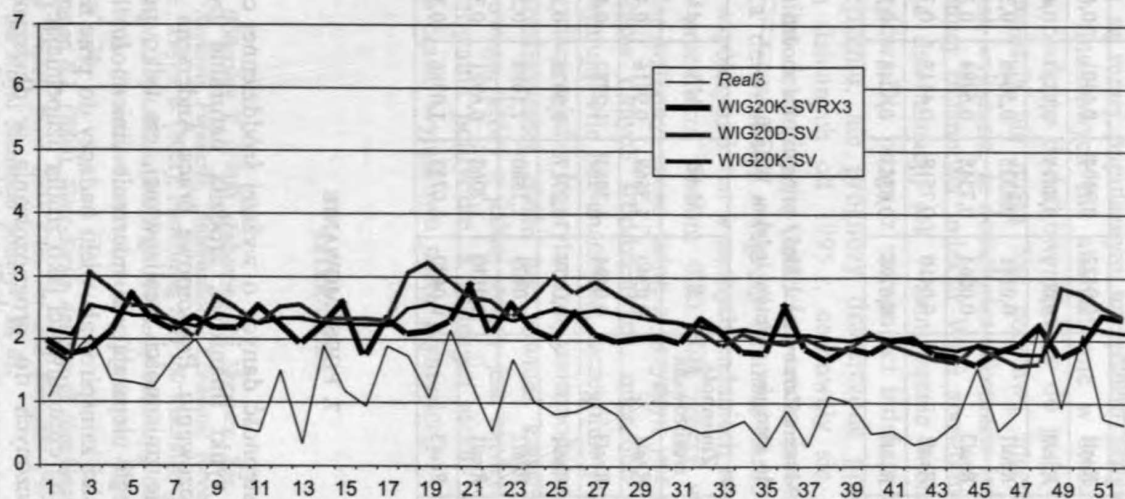
Rys. 1. Dzienna zmienność zrealizowana *Real1* indeksu WIG i jej prognozy otrzymane za pomocą modelu SV dla szeregów WIGD i WIGK oraz modelu $SVRX_3(Real2)$ dla szeregu WIGK



Rys. 2. Dzienna zmienność zrealizowana $Real3$ indeksu WIG i jej prognozy otrzymane za pomocą modelu SV dla szeregów WIGD i WIGK oraz modelu $SVRX_3(Real2)$ dla szeregu WIGK



Rys. 3. Dzienna zmienność zrealizowana *Real1* indeksu WIG20 i jej prognozy otrzymane za pomocą modelu SV dla szeregów WIG20D i WIG20K oraz modelu $SVRX_3(Real1)$ dla szeregu WIG20K.



Rys. 4. Dzienna zmienność zrealizowana $Real3$ indeksu WIG20 i jej prognozy otrzymane za pomocą modelu SV dla szeregów WIG20D i WIG20K oraz modelu $SVRX_3(Real1)$ dla szeregu WIG20K

Tabela 5. Współczynnik Mincera-Zarnowitza i błędy prognozy dla modeli dopasowanych do szeregów zwrotów indeksu WIG

Szereg Model	Zmienność zrealizowana	R^2	MAE	ME	AMAE	TIC
WIGD	Real1	0,0122	0,9404	0,6401	0,4192	0,3782
SV	Real3	0,0019	0,8815	0,8337	0,4231	0,4134
WIGK	Real1	0,0181	0,8325	0,5048	0,3952	0,3590
SV	Real3	0,0064	0,7348	0,6984	0,3899	0,3698
WIGK SVRX ₃	Real1	0,0410	0,7518	0,4112	0,3718	0,3494
$x_t = \ln(Real2_t)$	Real3	0,0796	0,6363	0,6048	0,3615	0,3394

Tabela 6. Współczynnik Mincera-Zarnowitza i błędy prognozy dla modeli dopasowanych do szeregów zwrotów indeksu WIG20

Szereg Model	Zmienność zrealizowana	R^2	MAE	ME	AMAE	TIC
WIG20D	Real1	0,0240	1,3668	0,9116	0,4001	0,3589
SV	Real3	0,0494	1,3053	1,2782	0,4147	0,3987
WIG20K	Real1	0,0278	1,2717	0,8022	0,3852	0,3503
SV	Real3	0,0409	1,1880	1,1688	0,3981	0,3771
WIG20K SVRX ₃	Real1	0,0130	1,2068	0,6921	0,3734	0,3484
$x_t = \ln(Real2_t)$	Real3	0,0490	1,0753	1,0586	0,3755	0,3609

7. PODSUMOWANIE

Coraz większa dostępność danych o wyższej śróddziennej częstotliwości, generowanych przez rynki finansowe, zrodziła naturalne pytanie o ich przydatność w prognozowaniu. Przełomowa praca Andersena i Bollersleva, wprowadzająca pojęcie zmienności zrealizowanej, nie tylko przyczyniła się do rozwoju metodologii ulepszających mierzenie zmienności instrumentów finansowych, ale również zainspirowała wielu badaczy do prac nad wykorzystaniem danych o wyższej częstotliwości jako źródła uzupełniających informacji w modelach przeznaczonych do prognozowania w dłuższym horyzoncie czasowym. Efektem tych prac są modyfikacje klasycznych parametrycznych modeli zmienności dziennej, takich jak modele z rodziny GARCH czy SV, rozszerzające je o dodatkowe zmienne objaśniające, będące funkcjami szeregów danych śróddziennych. Celem niniejszego opracowania była weryfika-

cja przydatności tych modeli w kontekście polskiego rynku finansowego, który od niedawna w coraz szerszym zakresie kreuje i udostępnia finansowe dane śróddzienne. Modele zmienności stochastycznej z powodu trudności metodologicznych związanych z ich estymacją i większej czasochłonności obliczeniowej są mniej popularnym narzędziem prognozowania zmienności instrumentów finansowych niż alternatywne w stosunku do nich modele GARCH. Stanowiło to dodatkowy bodziec do podjęcia badań. Otrzymane wyniki z pewnością nie mogą być podstawą zdecydowanych wniosków. Niemniej jednak wykazano, że dodatkowe informacje śróddzienne w przypadku prognoz dziennej zmienności indeksu WIG spowodowały zarówno wzrost współczynnika determinacji, jak i zmniejszenie wielkości wszystkich rozważanych błędów. Dla prognoz zmienności indeksu WIG20 błędy również uległy zmniejszeniu, ale prognozy dostarczone przez model oparty na obserwacjach dziennych od 1995 r. cechowały się nieznacznie wyższym współczynnikiem determinacji. Oddzielną rzeczą jest ogólnie niska wartość współczynnika determinacji z regresji Minzera-Zarnowitza. Należy jednak zauważyć, że współczynnik ten w znanych badaniach empirycznych podobnego typu rzadko przekracza poziom 0,20. Siłą rzeczy wyniki naszych badań mają charakter wstępny, czego jedną z przyczyn jest stosunkowo niewielka długość szeregów danych śróddziennych, mierzona liczbą dni sesyjnych, które one obejmują. Nie bez znaczenia, szczególnie jeśli chodzi o nietypowe opóźnienie w relacji między logarytmiczną zmiennością a informacją o przeszłej zmienności zrealizowanej, może być również kwestia specyfiki polskiego rynku finansowego, który z jednej strony jest bardzo wrażliwy na różnego rodzaju zawirowania polityczne, a z drugiej strony charakteryzuje się, od dłuższego już czasu, wyjątkowo niską płynnością.

LITERATURA

- Andersen T.G., Bollerslev T. (1998), *Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts*, „International Economic Review”, 39.
- Blair B., Poon S., Taylor S. (2001), *Forecasting S&P100 Volatility: The Incremental Information Content of Implied Volatilities and High Frequency Returns*, „Journal of Econometrics”, 105.
- Bollerslev T. (1986), *Generalized Autoregressive Conditional Heteroscedasticity*, „Journal of Econometrics”, 31.
- Durbin J., Koopman S.J. (2000), *Time Series Analysis of Non-Gaussian Observations Based on State Space Models from Both Classical and Bayesian Perspectives* (with discussion), „Journal of the Royal Statistical Society”, B, 62.
- Engle R.F. (1982), *Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation*, „Econometrica”, 50.
- Ghysels E., Harvey A., Renault E. (1996), *Stochastics Volatility*, [w:] Maddala G., Rao C. (eds), *Handbook of Statistics, 14; Statistical Methods in Finance*, North-Holland, Amsterdam.

- Hol E., Koopman S.J. (2000), *Forecasting the Variability of Stock Index Returns with Stochastic Volatility Models and Implied Volatility*, Tinbergen Institute, „Discussion Paper”, 104/4.
- Hol E., Koopman S.J. (2002), *Stock Index Volatility Forecasting with High Frequency Data*, Tinbergen Institute, „Discussion Paper”, 068/4.
- Hull J., White A. (1987), *The Pricing of Options on Assets with Stochastic Volatilities*, „Journal of Finance”, 42.
- Koopman S.J., Hol Uspensky E. (2002), *The Stochastic Volatility in Mean Model: Empirical Evidence from International Stock Markets*, „Journal of Applied Econometrics”, 17(6).
- Koopman S.J., Shephard N., Doornik J. (1999), *Statistical Algorithms for Models in State Space Using SsfPack2.2.*, „Econometrics Journal”, 2.
- Martens M. (2002), *Measuring and Forecasting S&P500 Index-Futures Volatility Using High Frequency Data*, „Journal of Futures Markets”, 22.
- Melino A., Turnbull S. (1990), *Pricing Foreign Currency Options with Stochastic Volatility*, „Journal of Econometrics”, 45.
- Osiewalski J., (2001), *Ekonometria bayesowska w zastosowaniach*, Wydawnictwo Akademii Ekonomicznej w Krakowie, Kraków.

Ryszard Doman

APPLICATION OF HIGH-FREQUENCY DATA IN FORECASTING POLISH STOCK INDICES BY MEANS OF STOCHASTIC VOLATILITY MODELS

Summary

Stochastic volatility (SV) models form a class of models applied to financial instrument volatility forecasting that is alternative to the one consisting of better known GARCH models. In contrast to GARCH models, the time-varying volatility in SV models is described by means of two uncorrelated stochastic processes. In this paper we apply stochastic volatility models to forecasting the daily volatility of the Warsaw Stock Exchange indices. The obtained forecasts are evaluated against the daily realized volatility understood as a sum of squared intraday returns. We also investigate the impact of entering the realized volatility as an additional explanatory variable on the quality of the forecasts.