PROGNOZOWANIE FINANSOWYCH SZEREGÓW CZASOWYCH

Andrzej Sokołowski
Akademia Ekonomiczna w Krakowie, Zakład Statystyki

W opracowaniu tym przedstawiono pewną grupę podstawowych modeli wykorzystywanych przy prognozowaniu finansowych szeregów czasowych oraz przykład analizy konkretnego szeregu czasowego. Składnik losowy obecny we wszystkich zjawiskach ekonomicznych jest reprezentowany przez proces stochastyczny, czyli ciąg zmiennych losowych o jednakowych rozkładach prawdopodobieństwa, zależnych od nielosowego parametru *t*, który reprezentuje czas. W literaturze przedmiotu i zastosowaniach spotyka się wiele klas modeli wykorzystywanych do opisu i prognozowania zjawisk finansowych. Rozsądne przedstawienie ich w jednej, krótkiej prezentacji jest niemożliwe, dlatego wybrano tylko (i tak dość liczną) pewną grupę modeli określoną przez następujące warunki:

- zmienna czasowa jest zmienną skokową, a więc modelowane dane dotyczą równoodległych momentów lub okresów czasu,
- rozpatrujemy kształtowanie się tylko jednej wielkości, czyli mamy do czynienia z jednowymiarowymi procesami stochastycznymi,
- prezentujemy tu tylko modele liniowe, czyli takie, w których wielkość zjawiska powiązana jest funkcją liniową z impulsami losowymi.

Modele wyjściowe

Biały szum

Biały szum to ciąg niezależnych zmiennych losowych o jednakowych rozkładach prawdopodobieństwa ze skończonymi wartościami przeciętnymi i wariancjami. Jeżeli są to rozkłady normalne z wartością przeciętną zero, to mamy do czynienia z *gaussowskim białym szumem.* Taki proces czysto losowy ma wartości funkcji autokorelacji równe zeru dla każdego opóźnienia. Uzyskane z próby oceny funkcji autokorelacji nie są oczywiście równe zeru, jednak na ogół różnią się od niego nieznacznie.

Liniowy szereg czasowy

Szereg czasowy nazywany jest *liniowym*, jeżeli można go zapisać w postaci

$$r_t = \mu + \sum_{i=0}^{\infty} \psi_i a_{t-i} \ ,$$

gdzie μ jest wartością przeciętną szeregu, a_t białym szumem, a współczynnik ψ_0=0. Struktura dynamiczna szeregu zależy od wartości współczynników ψ_i, które nazywane są też *wagami.*

Modele stacjonarne

Model autoregresji AR(p)

Jeżeli wartość szeregu jest skorelowana ze swoją poprzednią wartością, to rozsądnym modelem jest równanie

$$r_t = \phi_0 + \phi_1 r_{t-1} + a_t \, .$$

Jest to model autoregresji rzędu 1, co zapisujemy jako AR(1). W opisie szeregu możemy sięgać głębiej w przeszłość, powiedzmy p jednostek czasu wstecz. Wówczas mamy do czynienia z modelem autoregresji rzędu p, czyli AR(p).

$$r_t = \phi_0 + \phi_1 r_{t-1} + ... + \phi_p r_{t-p} + a_t \, .$$

Do identyfikacji rzędu modelu przydatna jest *funkcja autokorelacji cząstkowej*, która przyjmuje wartości nieistotnie różne od zera dla opóźnień większych od p. Przy pomocy takiego modelu prognozę buduje się krok po kroku, poprzez rekurencyjne podstawianie wartości. Przy stacjonarnych procesach AR(p) prognoza taka zmierza do przeciętnej wartości procesu, a wariancja błędu prognozy zmierza do wariancji procesu.

Model średniej ruchomej MA(q)

Proces średniej ruchomej jest uogólnieniem białego szumu, powstającym poprzez „wygładzenie” go pewnego rodzaju jednostronną średnią ruchomą o nierównych wagach. Proces MA(q) jest zawsze (słabo) stacjonarny. Jeden z możliwych sposobów zapisu takiego modelu to

$$r_t = c_0 + a_t - \theta_1 a_{t-1} - ... - \theta_q a_{t-q}$$

Wartość przeciętna takiego procesu jest równa c_0. Do identyfikacji rzędu procesu wykorzystuje się funkcję autokorelacji, której ostatnia wartość istotnie większa od zera wskazuje na rząd q. Prognozę przy pomocy modelu MA(q) uzyskuje się na drodze rekurencyjnej, przy czym bardzo szybko zmierza ona do wartości przeciętnej procesu.

Model ARMA(p,q)

Model ten łączy idee modeli średniej ruchomej i autoregresji. Ogólny zapis modelu może być przedstawiony jako

$$r_t = \phi_0 + \sum_{i=1}^{p} \phi_i r_{t-i} + a_t - \sum_{i=1}^{q} \Theta_i a_{t-i}$$

Do identyfikacji składowych procesu można wykorzystać *rozszerzoną funkcję autokorelacji* (EACF). Najpierw wyznaczamy modele AR o coraz większym rzędzie, a dla ich reszt liczymy funkcje autokorelacji. Wyniki przedstawiamy w postaci tabeli dwudzielczej, której wiersze odpowiadają rzędowi autoregresji, a kolumny rzędowi średniej ruchomej. Nieistotne autokorelacje, oznaczane przez zero, powinny w tej tabeli utworzyć trójkąt, którego lewy górny wierzchołek wskazuje właściwe parametry p oraz q. Proces ARMA prognozujemy analogicznie jak procesy omawiane uprzednio, na drodze obliczeń rekurencyjnych.

Modele niestacjonarne

Błądzenie przypadkowe

Szereg czasowy określany jest mianem błądzenia przypadkowego, jeżeli jego przebieg jest generowany następującym modelem

$$p_t = p_{t-1} + a_t ,$$

gdzie p_0 jest wartością startową procesu, zaś a_t białym szumem. Jeżeli biały szum ma rozkład symetryczny z wartością przeciętną zero, to prawdopodobieństwo tego, że szereg w następnej obserwacji „pójdzie” w górę, jest takie samo, że „pójdzie” w dół.

Błądzenie przypadkowe z dryftem

W wielu szeregach finansowych opisujących kształtowanie się logarytmów stóp zwrotu zauważono występowanie dodatniej wartości przeciętnej, zazwyczaj o małej wielkości. Oznacza to, że model odpowiedni dla takiej sytuacji ma postać

$$p_t = \mu + p_{t-1} + a_t .$$

Wielkość μ jest wartością przeciętną różnicy $(p_t - p_{t-1})$ i nazywana jest *dryftem*. Reprezentuje ona na przykład trend, jaki występuje w logarytmie p_t.

Model ARIMA(p,d,q)

Większość ekonomicznych szeregów czasowych to realizacje procesów niestacjonarnych. Typową niestacjonarnością jest obecność trendu. Możemy go wyeliminować przez różnicowanie. Krotność (stopień, rząd) różnicowania określony jest przez stopień wielomianu opisującego trend. W modelu ARIMA ten stopień różnicowania oznaczony jest przez d.

Sezonowy model ARIMA(p,d,q)(P,D,Q)

Wahania okresowe występujące w szeregu też stanowią pewien rodzaj niestacjonarności, a w każdym razie są składnikiem regularnym, który powinien zostać wyeliminowany z szeregu przed próbą oszacowania mieszanego modelu autoregresji średniej ruchomej. Wahania regularne eliminuje się poprzez różnicowanie sezonowe, wyliczając $(y_t - y_{t-s})$, gdzie s jest okresem (długością) wahania regularnego. Rząd różnicowania sezonowego oznaczony jest przez D.

Model ARFIMA(p,d,q)

Model ten jest nazywany modelem „z długą pamięcią”. Jest uogólnieniem procesu ARIMA poprzez dopuszczenie, aby rząd różnicowania d był liczbą niecałkowitą. Zazwyczaj rozważa się $-0,5<d<0,5$.

Warunkowe modele heteroskedastyczne

Tego typu modele są wykorzystywane w ekonometrii do modelowania kształtowania się zmienności (*volatility)* stóp zwrotu. Zmienność jest mierzona warunkową wariancją. Ma znaczenie również w szacowaniu wartości narażonej na ryzyko *(value at risk)*. W modelach jednowymiarowych zakłada się, że logarytmy stopy zwrotu (r_t) nie są niezależne w czasie, choć przyjmuje się występowanie tylko autokorelacji niskich rzędów. Warunkowość obecna w tych modelach oznacza, że wartość przeciętna i wariancja procesu mogą być wyrażone wzorami

$$\begin{aligned} \mu_t &= E(r_t | F_{t-1}) \\ \sigma_t^2 &= V(r_t | F_{t-1}) \end{aligned},$$

czyli są uwarunkowane informacjami dostępnymi w momencie (t-1).

Model ARCH(m)

Model ten zakłada, że odchylenia od wartości przeciętnej stóp zwrotu (a_t) mogą być objaśnione przez funkcję kwadratową ich wartości opóźnionych. Mamy zatem

$$\begin{aligned} a_t &= \sigma_t \varepsilon_t \\ \sigma_t^2 &= \alpha_0 + \alpha_1 a_{t-1}^2 + \ldots + \alpha_m a_{t-m}^2 \end{aligned}$$

Przez $\{\varepsilon_t\}$ oznaczamy ciąg niezależnych zmiennych losowych o jednakowym rozkładzie z wartością zero i wariancją 1. Najczęściej przyjmuje się tu standardowy rozkład normalny lub standaryzowany rozkład Studenta.

Model GARCH(m,s)

W tym podejściu zakłada się, że przy opisie kształtowania się logarytmów stóp zwrotu główne równanie procesu może być zapisane jako proces ARMA. Przyjmując, że a_t to logarytm stóp zwrotu, od którego odjęto wartość średnią, mamy

$$a_t = \sigma_t \varepsilon_t$$
$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{m} \alpha_i a_{t-i}^2 + \sum_{j=1}^{s} \beta_j \sigma_{t-j}^2 .$$

Założenie $\sum_{i=1}^{\max(m,s)} (\alpha_i + \beta_i) < 1$ powoduje, że bezwarunkowa wariancja a_t jest skończona, a warunkowa wariancja σ_t^2 zmienia się w czasie.

Istnieją pewne specjalne wersje omawianego modelu. I tak model IGARCH jest odpowiednikiem modelu ARIMA dla niejednorodnej wariancji procesu. Z kolei model GARCH-M (M to skrót od *in mean*) opisuje sytuacje, w których poziom stóp zwrotu zależy od zmienności, czyli do powyższych równań dochodzi relacja

$$r_t = \mu + c\sigma_t^2 + a_t .$$

Wykładniczy model EGARCH uwzględnia różny wpływ ε_t, w zależności od tego, czy realizacja tego procesu jest dodatnia czy ujemna.

Model CHARMA

To podejście wykorzystuje losowe współczynniki, które kształtują zachowanie się warunkowej wariancji. Model ma postać

$$r_t = \mu + a_t$$
$$a_t = \delta_{1t} a_{t-1} + \delta_{2t} a_{t-2} + \ldots + \delta_{mt} a_{t-m} + \eta_t ,$$

gdzie $\{\eta_t\}$ jest gaussowskim białym szumem, zaś $\{\delta_t\} = \{(\delta_{1t}, \ldots, \delta_{mt})'\}$ jest ciągiem wektorów losowych o jednakowych rozkładach i zerowych wartościach przeciętnych. Podobną konstrukcją jest model RCA, czyli model autoregresji z losowymi parametrami. W CHARMA zmienność zależy od opóźnionych wartości a_t, zaś w RCA, od opóźnionych wartości r_t.

Model SV

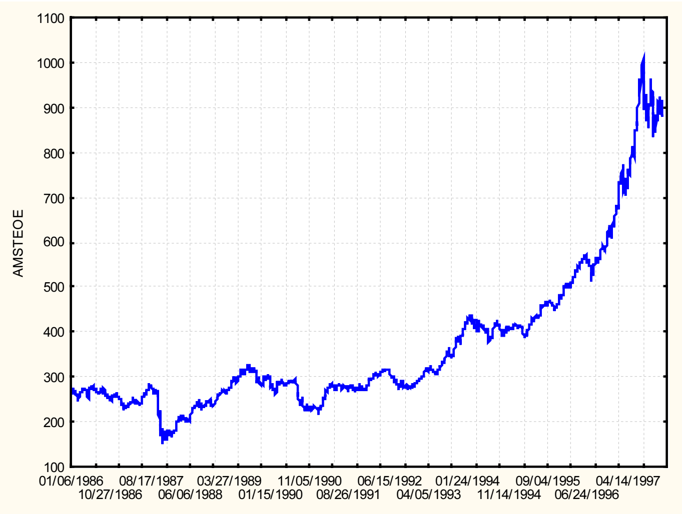
Jest to model zmienności stochastycznej (stochastic voaltility), w którym wykorzystuje się opóźnione logarytmy wariancji. Wygodny zapis uzyskujemy poprzez użycie operatora opóźnienia wstecznego B, którego działanie można opisać jako $(1 - B^m) y_t = y_t - y_{t-m}$. W tej konwencji model SV ma postać

$$a_t = \sigma_t \varepsilon_t$$
$$(1 - \alpha_1 B - \ldots - \alpha_m B^m) \ln(\sigma_t^2) = \alpha_0 + v_t .$$

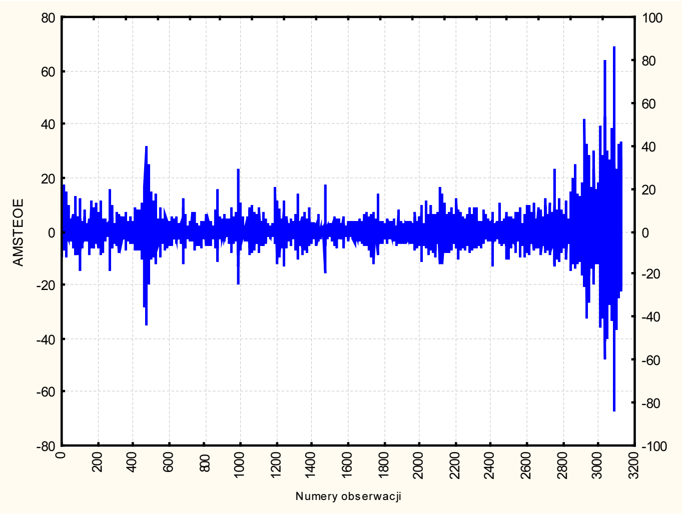
W tym modelu mamy dwa niezależne składniki losowe ε_t oraz v_t, które są procesami gaussowskimi, lecz pierwszy z nich ma wariancję 1, a drugi wariancję stałą, ale niekoniecznie równą jedności.

Analiza indeksu giełdy w Amsterdamie

Szereg czasowy obejmuje dane za okres od 6 stycznia 1986 roku do końca 1987 roku i został udostępniony przez P.H.Fransesa i D. van Dijka, autorów książki „Non-linear time series models in empirical finance" (Cambrigde University Press, 2002).



Rys. 1. Wartości indeksu giełdy w Amsterdamie



Rys. 2. Drugie różnice indeksu giełdy w Amsterdamie

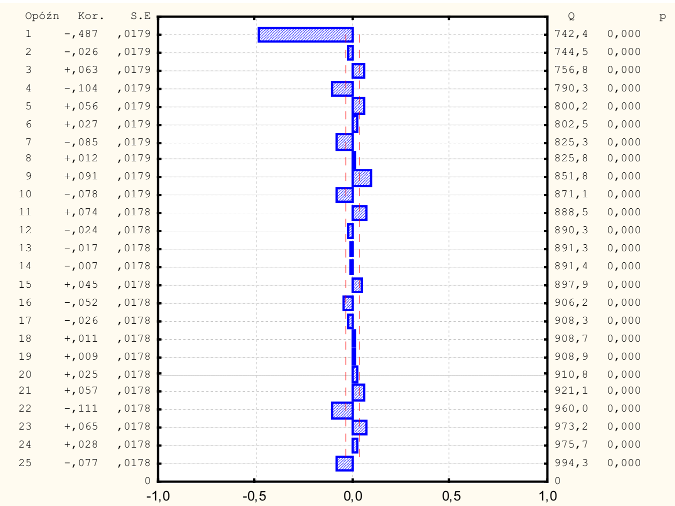
Tydzień w tym szeregu trwa od poniedziałku do piątku. W przypadku świąt, w których giełda nie funkcjonuje, przyjęto wartość z dnia poprzedniego. Dzięki tym zabiegom szereg charakteryzuje się regularnością próbkowania, umożliwiającą analizę wahań okresowych.

Jak widać na rysunku 1, w szeregu wystąpił trend wykładniczy, który jednak załamał się przy końcu analizowanego okresu. W celu wyeliminowania tego trendu zastosowano dwukrotne różnicowanie rzędu pierwszego, otrzymując szereg przedstawiony na rysunku 2.

Strukturę harmoniczną tego szeregu można rozpoznać poprzez analizę funkcji autokorelacji (rysunek 3). Widzimy, że proces charakteryzuje się bardzo mocną ujemną autokorelacją rzędu pierwszego, co oznacza, że bezpośrednio po wzrostach następują spadki i na odwrót. Związki nie ograniczają się do wpływu „dzień po dniu”, gdyż również wiele autokorelacji rzędów wyższych od jedności wykazuje istotność statystyczną. Warto zwrócić uwagę na to, iż nawet bardzo małe współczynniki korelacji okazują się istotne statystycznie, co jest spowodowane znaczną długością szeregu czasowego, którym dysponujemy. Jednym z dopuszczalnych modeli ARIMA, które uzyskano w trakcie analizy tego szeregu jest ARIMA(9,2,0), a więc model, który posiada tylko część autoregresyjną. Jeżeli przez $y_t^{(2)}$ oznaczymy drugie różnice oryginalnego szeregu, to oszacowany model można zapisać jako

$$y_t^{(2)} = -0{,}9318 y_{t-1}^{(2)} - 0{,}8638 y_{t-2}^{(2)} - 0{,}7727 y_{t-3}^{(2)} - 0{,}7786 y_{t-4}^{(2)} - 0{,}6587 y_{t-5}^{(2)}$$
$$- 0{,}5534 y_{t-6}^{(2)} - 0{,}5128 y_{t-7}^{(2)} - 0{,}3690 y_{t-8}^{(2)} - 0{,}1323 y_{t-9}^{(2)}.$$

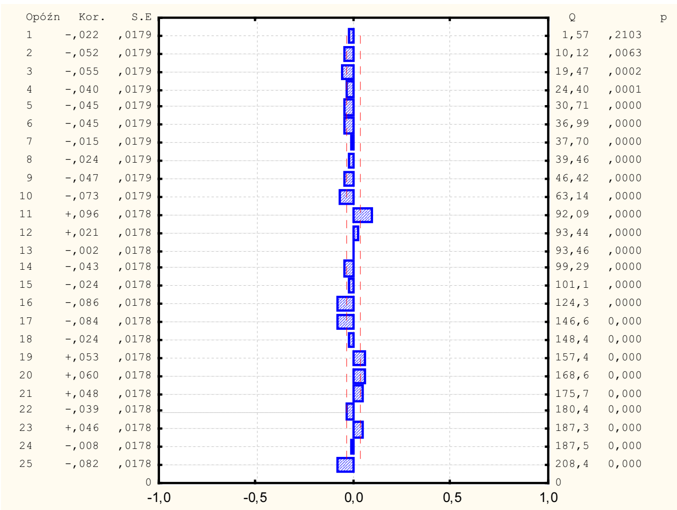
Wszystkie parametry są wysoce istotne statystycznie, gdyż wszystkie *wartości p* są mniejsze od 0,00001.



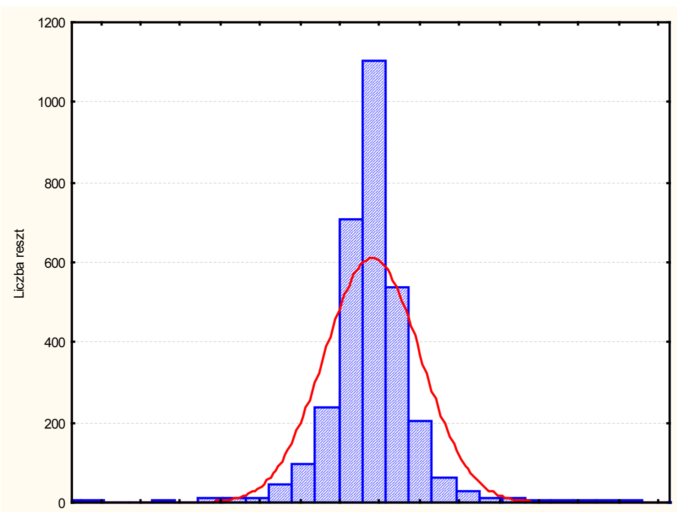
Rys. 3. Funkcja autokorelacji drugich różnic szeregu

Przedstawiony model wystarczająco dobrze opisuje kształtowanie się przeciętnego poziomu indeksu giełdy w Amsterdamie. Analiza reszt pokazuje, że to, co pozostaje, jest losowe.

Funkcja autokorelacji reszt, przedstawiona na rysunku 3, ma bardzo małe wartości dla wszystkich widocznych opóźnień, szczególnie dla ich małych wartości. Rozkład reszt, który widać na Rysunku 4 ma kształt nieco odbiegający od rozkładu normalnego. Jest więcej bardzo małych reszt, niż to wynika z rozkładu normalnego – co jest zjawiskiem korzystnym. Z kolei pojawiają się bardzo duże (co do modułu) reszty. Jest to związane z niejednorodnością wariancji, którą możemy zaobserwować na rysunku 2, gdy zmienność zdecydowanie rośnie przy końcu analizowanego okresu. Należy więc poddać modelowaniu heteroskedastyczny składnik losowy. W tym celu obliczono kwadraty reszt modelu ARIMA dla szeregu oryginalnego i dla tych kwadratów poszukiwano adekwatnego modelu. Okazał się nim model ARIMA(2,0,2). Oceny parametrów tego modelu zamieszczono w tabeli 1.



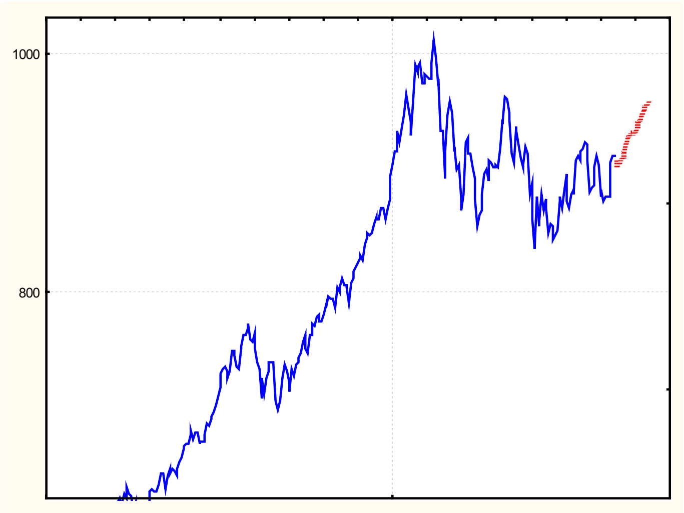
Rys. 4. Funkcja autokorelacji reszt



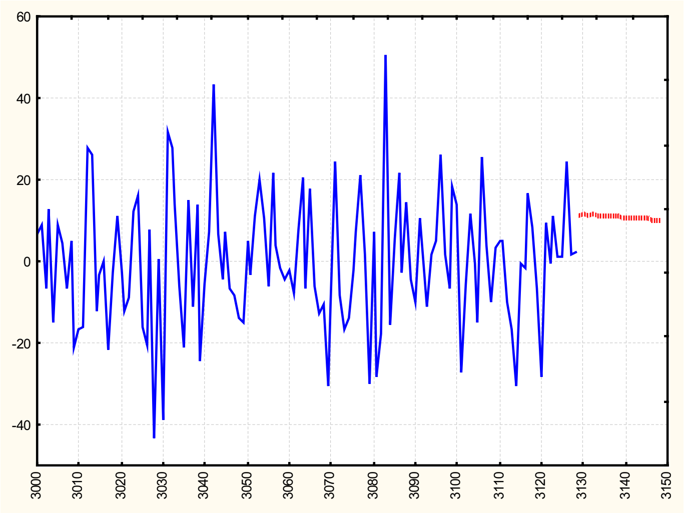
Rys. 5. Rozkład reszt

Tabela 1. Oceny parametrów modelu dla wariancji składnika resztowego

Parametr	Ocena	Wartość p
wyraz wolny	24,48935	0,0178
p(1)	0,32436	0,0015
p(2)	0,64510	0,0000
q(1)	0,30509	0,0056
q(2)	0,48877	0,0000



Rys. 6. Prognozy wyznaczone z modelu ARIMA(9,2,0)



Rys. 7. Prognoza poziomu odchylenia standardowego

Oszacowane modele mogą zostać wykorzystane do prognozowania. Na rysunku 6 przedstawiono prognozę uzyskaną na podstawie modelu ARIMA(9,2,0) dla oryginalnych wartości indeksu giełdy w Amsterdamie. Dla większej przejrzystości rysunek ten zawiera tylko wartości z końca analizowanego okresu. Z szeregu empirycznego widać, że wariancja mimo wszystko wydaje się stabilizować.

Rysunek 7 przedstawia prognozy odchylenia standardowego wyznaczone z modelu ARIMA(2,0,2), który dotyczy kwadratów reszt, nałożone na szereg czasowy wartości reszt. Dlatego prognoza znajduje się tylko po dodatniej stronie wartości szeregu. Nasz model, który opisuje kształtowanie się warunkowej niejednorodnej wariancji, jest modelem GARCH(2,2).