Metody analizy długozasięgowej

Andrzej Zacharewicz

Warsztat analizy zależności długoterminowej jest wciąż rozwijany i udoskonalany. Od czasów Hursta (1951) i jego analizy R/S powstało wiele nowych metod. Wiele z nich jest na tyle nowych, że posiadają dopiero status sposobów pomocniczych, przy pomocy których „patrzymy" na dane w nowy sposób. Niektóre z nich nie są jeszcze do końca zbadane i sformalizowane, ale stanowią doskonałe narzędzia obróbki szeregów czasowych. Oto metody zaimplementowane w programie Long Memory Analysis.

1. Analiza R/S.

Budowniczy tam rzecznych, hydrolog H.E. Hurst pracując nad projektem tamy na rzece Nil dysponował 847-letnim zapisem poziomu Nilu zostawionym przez Egipcjan. Wielu hydrologów przypuszczało, że poziom rzeki jest procesem całkowicie losowym, niezależnym od przeszłości. Jednak Hurst, analizując zapisy Egipcjan odkrył, że dane wcale nie reprezentują takiej losowości, chociaż standardowe metody analizy statystycznej na nic takiego nie wskazywały. Dlatego Hurst stworzył całkiem nową metodę analizy danych.

W roku 1908 Einstein opublikował pracę na temat ruchu Browna. Udowodnił w niej, że cząsteczka poruszająca się ruchem Browna pokonuje w czasie t odległość $R = \sqrt{t}$ Hurst przy pomocy przeskalowanego zasięgu (*rescaled range, R/S*) rozszerzył ten model na następujący:

$$\left(\frac{R}{S}\right)_n = cn^H \tag{1.1}$$

gdzie S jest odchyleniem standardowym przyrostów w czasie n, c jest pewną dodatnią stałą, n oznacza ilość obserwacji (u nas będzie to ilość elementów szeregu czasowego), zaś H jest **wykładnikiem Hursta.**

Aby wyznaczyć H należy najpierw dla różnych n obliczyć wartości średnie R/S, a następnie, przy pomocy regresji liniowej, rozwiązać równanie

$$\ln \mathbb{E}(R/S)_n = \ln c + H \ln n.$$

Wystarczy zatem wykreślić w skali podwójnie logarytmicznej $\mathbb{E}(R/S)_n$ względem n. Nachylenie krzywej będzie wtedy estymować H.

Hurst rozszerzył model Einsteina z $t^{0.5}$ do t^H. Jeśli dane pochodzą z ciągu niezależnych zmiennych losowych o jednakowym rozkładzie, to otrzymujemy asymptotyczny wzór Hursta:

$$E(R/S)_n = \sqrt{n\frac{\pi}{2}}$$

Sedno analizy R/S tkwi w przeskalowaniu zasięgu. Zwrot ten jest często używany w analizie procesów samopodobnych. Polega to na analizowaniu zasięgu R jaki ma szereg czasowy w różnych

odstępach czasu *n*. Porównując go z podobnym zasięgiem w przypadku niezależnych zmiennych losowych (i.i.d.) możemy wysunąć wnioski, co do niezależności i ewentualnej długości pamięci naszego procesu.

Istnieją trzy klasy wartości wykładnika Hursta. Jeśli *H=0,5* to proces zachowuje się tak jak błądzenie losowe, czyli i.i.d. Jeśli nie, szereg nie jest niezależny:

- ***H=0,5*** - biały szum (*white noise*) - szereg i.i.d.,
- ***0<H<0.5*** - proces pokonuje mniejszą drogę niż błądzenie losowe. Dlatego też ma tendencję do częstego „zawracania". Gdy jest wzrastający, to zaczyna szybko maleć i odwrotnie, po spadku szybko następuje wzrost. Niewiele empirycznych danych posiada tę własność, taką cechę ma np. zmienność procesów rynkowych,
- ***0.5<H<1*** - takie szeregi czasowe pokonują większy dystans niż błądzenie losowe. Gdy proces ma tendencję wzrostową, to z dużym prawdopodobieństwem tą tendencję utrzyma i podobnie w przypadku spadku. Nazywa się to zwykle „efektem Josepha", przypominającym nam, że po siedmiu latach szczęścia przyjdzie następne siedem lat niedoli. Podobne procesy rozpatruje się w ubezpieczeniach, gdzie nagłe katastrofy wywołują lawinę zgłoszeń.

Istnieje dokładny algorytm opisujący sposób liczenia średniej wartości $(R/S)_n$, będącej estymatorem przeskalowanego zasięgu $E(R/S)_n$. Z małymi modyfikacjami zastosowano go w pakiecie LMA. Ze względu na język programowania, w którym powstał LMA (C++) niektóre fragmenty zostały zoptymalizowane pod kątem szybkości wykonania. Analiza R/S jest jedną z bardziej czasochłonnych metod analizy długozasięgowej.

Podczas badania zwrotów giełdowych metodą analizy *R/S* pojawia się problem zbyt małej liczby danych. Hurst znał historię zmian poziomu Nilu na przestrzeni kilkuset lat, więc mógł z dużą dokładnością obliczyć *R/S* dla dużych *n*. Często jednak, tak jak w przypadku danych KGHM, dla dużych *n* liczba podciągów jest zbyt mała, aby uśrednione wartości dawały dobre przybliżenie. Trzeba więc ograniczyć się do krótkich podciągów. Poza tym, że taka analiza nie będzie już długozasięgowa, okazuje się, że dla małych *n* nie można sprawdzić, czy badany ciąg składa się z niezależnych zmiennych losowych o jednakowym rozkładzie, korzystając z asymptotycznego wzoru Hursta

$$E(R/S)_n = \sqrt{n\frac{\pi}{2}} \tag{1.2}$$

Innymi słowy, jeśli do wyznaczenia *H* zostały użyte zbyt małe wartości *n*, to fakt, że wartość *H* jest różna od 0,5 wcale nie implikuje niezależności badanego procesu. W roku 1976 Anis i Lloyd zaproponowali inny wzór, który modyfikował wzór Hursta dla małych *n*:

$$E(R/S)_n = \begin{cases} \dfrac{\Gamma(\frac{n-1}{2})}{\sqrt{\pi}\Gamma(\frac{n}{2})} \displaystyle\sum_{i=1}^{n-1}\sqrt{\frac{n-i}{i}} & \text{dla } n \le 340, \\ \sqrt{\dfrac{2n}{\pi(n-1)^2}} \displaystyle\sum_{i=1}^{n-1}\sqrt{\frac{n-i}{i}} & \text{dla } n > 340. \end{cases} \tag{1.3}$$

Dla *n*>340 wzór ten jest przybliżony wzorem Stirlinga, w celu skrócenia obliczeń. Ta propozycja daje znacznie lepsze przybliżenie średnich wartości $(R/S)_n$ niż wzór Hursta. Jednak dla *n*<40

istnieje nadal znaczna różnica pomiędzy wzorem (1.3), a wynikiem symulacji komputerowej. Dopiero w roku 1994 Peters wprowadził poprawkę, która zapewniła dostateczną zgodność przybliżenia z obserwowanymi w symulacjach wartościami:

$$E(R/S)_n = \begin{cases} \dfrac{n-0{,}5}{n} \dfrac{\Gamma(\frac{n-1}{2})}{\sqrt{\pi}\Gamma(\frac{n}{2})} \displaystyle\sum_{i=1}^{n-1} \sqrt{\frac{n-i}{i}} & \text{dla } n \le 340, \\ \dfrac{n-0{,}5}{n} \sqrt{\dfrac{2n}{\pi(n-1)^2}} \displaystyle\sum_{i=1}^{n-1} \sqrt{\frac{n-i}{i}} & \text{dla } n > 340. \end{cases} \tag{1.4}$$

Badając dzienne zwroty giełdowe, przyjmujemy, że efekt długiej pamięci występuje wtedy, gdy wartość H wyznaczona empirycznie (przy pomocy algorytmu przedstawionego poniżej) różni się od wartości teoretycznej przynajmniej o $\sqrt{1/N}$ $1/N$ jest wariancją H, gdy ciąg składa się z N zmiennych i.i.d. o rozkładzie $N(0,1)$.

Algorytm wyznaczania $\mathbb{E}(R/S)_n$

1. Ciąg zwrotów o długości N podziel na d podciągów o długości n, przy czym można tego dokonać tylko dla tych n, dla których $d \cdot n = N$.

2. Dla każdego podciągu $m = 1, ..., d$:

a) wyznacz średnie (arytmetyczne) wartości zwrotów (E_m) oraz empiryczne odchylenia standardowe (S_m);

b) przeskaluj wartości $Z_{i,m}$ przez odjęcie średniej wartości zwrotów w tym podciągu:
$X_{i,m} = Z_{i,m} - E_m$, dla $i=1,...,n$;

c) skonstruuj skumulowany ciąg przeskalowanych zwrotów:
$Y_{i,m} = \sum_{j=1}^{i} X_{j,m}$ dla $i=1,...,n$;

d) oblicz zasięg: $R_m = \max\{Y_{1,m},..., Y_{n,m}\} - \min\{Y_{1,m},..., Y_{n,m}\}$;

e) przeskaluj zasięg: R_m/S_m.

3. Średnia wartość przeskalowanego zasięgu dla podciągów o długości n wynosi więc

$$(R/S)_n = \frac{1}{d} \sum_{m=1}^{d} \frac{R_m}{S_m}$$

Jako przykład niech posłuży analiza indeksu miedziowego z giełdy w Johannesburgu (JOHCOP). W tabeli przedstawiono estymowane parametry wraz z odpowiadającymi im przedziałami. $\sqrt{1/N}$ = 0,019. Tak więc dane wykazują zależność gdy różnica pomiędzy *H* teoretycznym i *H* empirycznym będzie większa (co do modułu) od 0,019.

n	H empiryczne	H teoretyczne	różnica H
10 – 176	**0,6381**	**0,5792**	**0,0589**
176 – 440	0,5173	0,5279	-0,0105
440 – 1320	**0,4538**	**0,5169**	**-0,0631**

Na tym przykładzie możemy zaobserwować wszystkie trzy typy szumu. Choć powyżej 440 dni wykładnik Hursta nie jest zbyt dobrze zdefiniowany, to na potrzeby przykładu możemy przyjąć, że istnieje.

- Na przedziale krótkoterminowym, **do 176 dni** obserwujemy długozasięgową zależność danych. Proces zwrotów indeksu ma więc pamięć sięgającą ponad pół roku (najczęściej przyjmuje się 250 dni roboczych w roku). Na tym odcinku *H*=0,63. Jest to wynik znacznie mniejszy od otrzymanego przez Hursta podczas badania poziomu Nilu, ale jest znaczący.
- Analizując zakres **176 – 440 dni** nie stwierdzamy żadnej zależności. Obliczony wykładnik Hursta jest nieco mniejszy niż teoretyczny, ale różnica mieści się w dopuszczalnym przedziale.
- W zakresie **powyżej 440 dni** wykładnik Hursta jest mniejszy od 0,5. Oznacza to sytuację, w której szereg ten oglądany z perspektywy ponad 440 dni (interwał czasowy szeregu = 440 dni) będzie pokonywał mniejszą drogę niż szereg niezależnych danych. W przypadku wielkich wzrostów następować będzie dążenie do spadków i odwrotnie. Następuje swoista „auto-stabilizacja" procesu.

2. Analiza V.

Analiza V powstała jako uzupełnienie i modyfikacja analizy R/S. Jest ona szczególnie przydatna w szukaniu cykli nieokresowych. **Cykle nieokresowe** (*nonperiodic cycles*) są naturalnym rozszerzeniem pojęcia cyklu okresowego. Sinusoida ma ustalony okres: 2π. W przypadku procesów losowych okresy jednak nie są na ogół stałe, są losowe. Również wahanie takiego okresu wokół jakiegoś **okresu głównego** nie jest zerowe i może zmieniać się zgodnie z jakimś rozkładem prawdopodobieństwa. W niektórych przypadkach rozkład ten ma dobrze zdefiniowaną średnią i wariancję. Takie procesy nazywamy procesami z cyklami nieokresowymi.

Używając analizy R/S łatwo można pokazać, że procesy rynkowe, w zakresie krótkoterminowym, mają dobrze zdefiniowany wykładnik Hursta, stąd trudno jest określić ich „główny okres". Ale patrząc na strukturę długoterminową rynek zachowuje się inaczej. Aby zaobserwować tę różnicę stworzono statystykę V:

$$V = \frac{(R/S)}{\sqrt{n}}$$

Postępowanie jest podobne jak w przypadku analizy R/S. Różnice polegają na rysowaniu na osi y statystyki V, a nie jej logarytmu, oraz na dzieleniu statystyki przez pierwiastek z *n*. W praktyce analiza V pozwala na uwydatnienie na wykresie tych miejsc (zakresów), w których pojawiają się cykle zależności. Na przykładzie (rys. 2), który jest odpowiednikiem analizy z rys. 1, dokładnie to powiększenie widać. Ponieważ próbka ma długość tylko 2640 dni roboczych, więc z dwóch wyróżniających się okresów odrzucamy ten drugi (660 dni). Dla *n*=660 ilość podciągów, na podstawie których liczymy zasięg wynosi 4, więc jest zbyt mała, aby z zadowalającą dokładnością wyliczyć średni zasięg. Pomijając ten punkt otrzymujemy główny okres procesu równy 240 dni. Jest to niecały rok biznesowy.

Na wykresie przerywaną linią zaznaczona jest statystyka V dla ciągu niezależnych zmiennych losowych. Jak widać nie mamy tu żadnych okresów, ani punktów charakterystycznych. Jest to typowa cecha błądzenia losowego – brak głównego okresu.

3. Korelacja opóźniona (*partial correlation*).

Analizując długość pamięci danego procesu możemy skorzystać ze standardowej metody na badanie liniowej zależności. Jest nią korelacja pomiędzy dwoma procesami, z tym, że jeden z szeregów zostaje przesunięty w czasie o pewną stałą (*lag*). Jeśli porównujemy w ten sposób dwa różne procesy, to mówimy o **korelacji opóźnionej** lub **częściowej** (*lagged correlation* bądź *partial correlation*). Gdy zaś takiej analizie poddajemy jeden proces, porównując dwa identyczne, ale wzajemnie przesunięte szeregi, mówimy o **autokorelacji** (*autocorrelation*). Oto przykład estymatora korelacji opóźnionej dla dwóch szeregów czasowych o długości *n*:

$$\rho(\tau) = \frac{\sum_{t=\tau}^{n} [X_{t-\tau} - \bar{x}][Y_t - \bar{y}]}{\sqrt{\sum_{t=1}^{n} [X_t - \bar{x}]^2 \sum_{t=1}^{n} [Y_t - \bar{y}]^2}} \tag{1.5}$$

Wykresy autokorelacji niosą ze sobą wiele informacji. Dokładnie widzimy prędkość „utraty pamięci”. Często jednak wyniki otrzymane przy pomocy autokorelacji nie pokrywają się z wynikami analizy R/S czy analizy V. Jest tak dlatego, że metoda korelacji opóźnionej jest bardzo czuła na krótkie odległości, więc łatwiej jest nią badać krótkoterminowe zależności na rynku. W miarę wzrastania opóźnienia (*lag*) przedmiotem analizy są coraz mniejsze fragmenty procesu, więc dokładność obliczeń się zmniejsza. Jako rozsądny przedział opóźnienia przyjmuje się więc *N*/5. Chcąc więc analizować korelację w perspektywie roku musimy dysponować aż pięcioletnim zbiorem danych.

Na rys. 3 przedstawiony jest przykład zastosowania autokorelacji w praktyce. Znany już indeks miedziowy wykazuje zależność pomiędzy ostatnimi dwoma, czterema i sześcioma miesiącami. Piki malejące w odległościach dwóch miesięcy świadczą o tym, że proces kwadratów zwrotów indeksu JOHCOP jest skorelowany z danymi sprzed dwóch miesięcy. Same zwroty giełdowe nie wykazują autokorelacji. Już dla przesunięcia kilku dni korelacja wpada w przedział ufności ruchu Browna. Inaczej dzieje się z wartościami bezwzględnymi oraz z kwadratami zwrotów. Oto trzy wykresy ilustrujące autokorelację odpowiednio zwrotów, ich modułów oraz ich kwadratów (rys. 4).

Na wykresie autokorelacji samych zwrotów nie widać żadnej zależności długoterminowej. Pojawia się ona dopiero przy analizie wartości bezwzględnej zwrotów. Krzywa wolno opada, co oznacza, że proces dla coraz większych odległości traci pamięć. Analiza kwadratów zwrotów daje jeszcze inny obraz autokorelacji. Wyraźniej zaznaczają się konkretne wartości, dla których proces

wykazuje zależność długoterminową. Szum losowy, czyli korelacja mieszcząca się w przedziale ufności błądzenia losowego, jest też tutaj odpowiednio „zduszony” do zera. Dlatego też w analizie zależności liniowej najczęściej wybiera się autokorelację kwadratów zwrotów.

4. Wykres wariancji (*variance plot*).

Główną cechą procesów z długą pamięcią jest to, że **wariancja średniej z próbki** zbiega do zera wolniej niż $1/n$. Wyraża się to wzorem:

$$Var(\overline{X}_n) \approx cn^{2H-2}, \tag{1.6}$$

gdzie $c>0$, a H jest znanym już wykładnikiem Hursta. Oto prosty algorytm wykreślania „variance plot” oraz estymacji parametru H.

1. Ciąg zwrotów $\{Z_t\}$ o długości n podziel na jednakowe podciągi o długości $k0(2;n/2)$. Otrzymamy w ten sposób m_k podciągów.

2. Dla każdego podciągu $m=1, ..., m_k$ oblicz jego empiryczną średnią

$$\overline{Z}_m(k) = \frac{1}{k}\sum_{j=1}^{k} Z_j$$

oraz średnią z całości

$$\overline{Z}(k) = \frac{1}{m_k}\sum_{j=1}^{m_k} \overline{Z}_j(k)$$

3. Oblicz empiryczną wariancję średnich $\overline{Z}_m(k)$

$$s^2(k) = \frac{1}{m_k - 1}\sum_{j=1}^{m_k} (\overline{Z}_j(k) - \overline{Z}(k))^2$$

4. Powtórz punkty 1-3 dla $k0(2;n/2)$.

5. Wykreśl (jako punkty) $\ln(s^2(k))$ względem $\ln(k)$.

Dla dużych wartości k, punkty na wykresie powinny układać się w linię prostą, o ujemnym współczynniku nachylenia równym $2H$-2. W przypadku zależności krótkoterminowej, lub danych niezależnych prostą graniczną będzie prosta o nachyleniu –1. Natomiast w przypadku zależności długoterminowej dane ułożą się w prostą o mniejszym (co do wartości bezwzględnej) współczynniku nachylenia. Tak więc variance plot daje nam zgrubną informację o istnieniu zależności długozasięgowej w danym procesie, pod warunkiem, że zależność ta jest wystarczająco silna. Małe odchyłki od H=0,5 są trudne do zweryfikowania nawet dla długich próbek. Mimo tych wad metoda ta jest prosta w użyciu, a variance plot czytelny i łatwy w interpretacji.

Oto przykłady dwóch procesów: modelu ARCH(1) (dobranego tak, aby nie wykazywał zależności długoterminowej) oraz indeksu S&P500 na przestrzeni lat 1961–1991. Variance plot pierwszego procesu (rys. 5) obrazuje dokładnie to, co jest główną cechą modelu ARCH: brak zależności długoterminowej. Punkty układają się dokładnie wokół linii H=0,5. W takim przypadku nie ma wątpliwości co do braku pamięci takiego szeregu.

Na rys. 6 sytuacja wygląda ciekawiej. Po pierwsze, dane nie tworzą prostej. Możemy jednak odrzucić pierwsze punkty obserwacji (do około 10) i otrzymamy już prostą odpowiadającą H=0,8. Jest to co prawda zgrubna aproksymacja, ale daje ogólny wgląd w zależności długoterminowe tego procesu.

5. Variogram.

Variogramy są często używane w geostatystyce. Na przykład w procesach przestrzennych są wykorzystywane razem z metodologią nazywaną *„kriging”*. Variogram o opóźnieniu k (*lag*) jest zdefiniowany następująco:

$$V(k) = \frac{1}{2} E\left[(X_t - X_{t-k})^2\right] \tag{1.7}$$

Jeśli X_t jest procesem stacjonarnym z kowariancją $\gamma(k)$ i korelacją $\rho(k)$, to $V(k)$ zbiega do granicy:

$$V(k) = \gamma(0)(1-\rho(k)) = V(\infty)(1-\rho(k)).$$

Istnieje alternatywną metoda wyznaczania variogramów. Wystarczy po prosty wykreślić średnie wartości $(X_i\text{-}X_j)^2$ względem $|i\text{-}j|$ dla $i<j$ $(i,j=1,...,n)$. Otrzymamy wtedy rozrzucone punkty. Taki algorytm został zastosowany w LMA. Oś y została jednak odpowiednio przeskalowana, tzn. wartości $(X_i\text{-}X_j)^2$ zostały jeszcze poddane standaryzacji, czyli podzielone przez empiryczną wariancję z próby. W ten sposób otrzymujemy linię referencyjną na poziomie 1. Procesom bez pamięci będą odpowiadać variogramy złożone z punktów „oscylujących” wokół tej prostej.

Odczyt variogramu nie jest tak intuicyjny jak w przypadku variance plot. Ma on jednak tę zaletę, że jest dobrze zdefiniowany dla procesów niestacjonarnych. Na przykład dla ARIMA(0,1,0) $X_t=X_{t-1}+0_t$ z szumem losowym 0_t, $V(k)$ jest równe $\frac{1}{2}k\sigma_\varepsilon^2$, natomiast dla procesu niestacjonarnego z trendem liniowym $X_t=at+0_t$ mamy:

$$V(k) = \frac{1}{2}a^2k^2 + \sigma_\varepsilon^2 \approx a^2k^2.$$

Jako przykład niech posłużą dwa variogramy, będące wynikiem analizy szeregu ARCH(1) oraz indeksu giełdowego DIJA na przełomie prawie całego XX wieku. Rys. 7 obrazuje proces krótkoterminowy. Tutaj wariancja oscyluje wokół wariancji całego szeregu co widać w równomiernym jej rozłożeniu. Świadczy to o braku pamięci, gdyż w takim przypadku, wariancja modelu w miejscach, gdzie zależności długoterminowe mają duży wpływ, byłaby odpowiednio zaburzona. Tutaj jednak już dla lag=3 proces zdaje się być pozbawiony takich cech.

Na rys. 8 przedstawiono variogram z bardzo długiej próbki indeksu Dow Jonesa. Tutaj proces wykazuje wspomniane wcześniej zaburzenia wariancji. Dla małych opóźnień jest ona niewielka, co może świadczyć o długoterminowej zależności wartości bez-względnej zwrotów tego indeksu. Zależność ta sięga aż 1100 dni biznesowych, czyli ponad 4 lata. Wynik ten został zaprezentowany przez Petersa, jednak przy pomocy całkiem innej metody, analizy R/S. Tam zależność obejmowała 1044 dni.

6. Zmienność krótko- i długoterminowa (*fine & coarse volatilities*).

Inną własnością szeregów czasowych jest zależność między **zmiennością krótkoterminową** (*fine*) i **długoterminową** (*coarse*). Dla danego procesu zmienność krótkoterminowa została zdefiniowana jako średnia wartości bezwzględnych pięciu ostatnich zwrotów (dla danych dziennych odpowiada to jednemu tygodniowi):

$$\sigma_i = \frac{1}{5}\sum_{t=m}^{m+4}|Z_t| \tag{1.8}$$

natomiast zmienność długoterminowa – jako wartość bezwzględna z sumy pięciu ostatnich zwrotów:

$$\varsigma_i = \left|\sum_{t=m}^{m+4} Z_t\right| \tag{1.9}$$

Oczywiście podział na przedziały o długości 5 jest umowny i zależy od typu danych. Mogą to być odcinki zawierające 6, 10 bądź 20 sąsiednich zwrotów. Np. dla badanego co 30 minut kursu wymiany USD/DEM (Olsen), zmienność krótkoterminowa została zdefiniowana jako średnia z wartości bezwzględnych sześciu ostatnich zwrotów, co odpowiada 3-godzinnemu przedziałowi czasowemu.

Dzięki takiej definicji otrzymujemy różne „punkty obserwacyjne" rynku. Zmienność krótkoterminowa będzie odpowiadała horyzontowi graczy krótkoterminowych, zwracających uwagę na małe i krótkie wahania na rynku oraz czerpiących z tego korzyści. Natomiast zmienność długoterminowa będzie odzwierciedleniem handlarzy długoterminowych, stosujących inną strategię polegającą na obserwacji rynku w szerszej skali czasowej.

Oto przykład zastosowania metody fine&coarse volatilities. Na rysunku 9 przedstawiono wykres korelacji między zmiennością krótko- i długoterminową dla modelu GARCH(4,4). Parametry zostały tak dobrane, aby proces miał charakter długoterminowy.

GARCH(4,4)		
i	b_i	c_i
0		$2\cdot10^{-6}$
1	0,1	0,1
2	0,1	0,1
3	0,1	0,1
4	0,2	0,2

Objawia się to wolnym opadaniem krzywej auto-korelacji. Mimo tego różnica pomiędzy lewą i prawą stroną jest znikoma (mieści się w 95% przedziale ufności ruchu Browna). Nie można więc mówić o wpływie jednej zmienności na drugą. Rys. 10 jest odpowiednikiem rys. 9 z tym, że tutaj analizowany jest szereg HARCH(4096). Oto wartości jego współczynników:

HARCH(4096)	
i	c_i
0	10^{-7}
1 – 100	0,001
100 – 500	0,001
500 – 1000	0,01
1000 - 4096	0,02

Parametry modelu zostały tak dobrane, aby przedstawić właśnie tę cechę danych finansowych – wpływ zmienności długoterminowej na krótkoterminową – zmienność długoterminowa pozwala lepiej przewidzieć zmienność krótkoterminową niż odwrotnie. Wykres różnicy między korelacją lewej i prawej strony przedstawia ten wpływ sięgający aż ośmiu ostatnich tygodni (zakładając, że dane są danymi dziennymi). Dla danych finansowych, zwłaszcza kursów wymiany walutowej, istnieje makrorynkowe uzasadnienie tego faktu. Zmienność długoterminowa ma wpływ na krótkoterminową ponieważ poziom zmienności długoterminowej odzwierciedla oczekiwane rozmiary trendów na rynku. Z jednej strony ma to istotny wpływ na zachowanie się handlarzy krótkoterminowych, którzy w reakcji na zmianę poziomu tej zmienności zmieniają sposób handlu, powodując zmianę zmienności krótkoterminowej. Z drugiej strony, wahania poziomu zmienności krótkoterminowej nie maja wpływu na strategię handlarzy długoterminowych, którzy kierują się raczej wiadomościami fundamentalnymi. Z rysunków 9-10 widać, że modele HARCH przejawiają tę własność, natomiast modele GARCH już nie.

7. Gęstość empiryczna (*frequency distribution*).

Wyznaczanie empirycznej gęstości rozkładu szeregu czasowego jest rzeczą prostą. Zbiór wartości procesu należy podzielić na *n* jednakowych odcinków. Następnie zlicza się ile punktów z szeregu wpada do poszczególnych przedziałów. Istotny jest przy tym odpowiedni dobór stałej *n*. Przy zbyt małej ilości przedziałów będą one odpowiednio szerokie, co da niezbyt precyzyjne wyniki. Za duże *n* może przy zbyt małej długości procesu powodować efekt „dziur", gdyż nie starczy danych do poprawnego zapełnienia wszystkich przedziałów.

Z wykresu gęstości empirycznej (*frequency distribution*) zwanego czasem histogramem można wyczytać wiele interesujących własności. W prosty sposób możemy porównać nasz proces z procesem gaussowskim. Możemy zobaczyć grubość ogonów, a przede wszystkim skośność danych (*skewness*).

Oto dwa rysunki obrazujące to zagadnienie. Na rys. 11 przedstawiono histogram zwrotów błądzenia losowego (realizacja zmiennej losowej o rozkładzie normalnym N(0,1)). Skośność dla rozkładu normalnego jest zerowa, więc wykres gęstości jest symetryczny względem średniej (badanie skośności jest jednym ze sposobów testowania zgodności próbki z rozkładem normalnym).

Rys. 12 jest histogramem trajektorii modelu TARCH(2,2,2). Oto parametry tego modelu:

i	a_i	a_i`	c_i
0	0,0007	-0,0001	0,0001
1	0,2	0,2	0,01
2	0,1	-0,1	0,02

Zostały one odpowiednio dobrane do potrzeb tego przykładu. Trajektoria procesu, po obliczeniu jej z szeregu zwrotów, ma prawie liniowy wzrost i byłaby kiepskim estymatorem realnego procesu, ale właśnie dzięki temu, że większość zwrotów ma znak dodatni uzyskujemy prawostronną skośność.

Ta niesymetryczność odzwierciedla zachowanie się inwestorów w zależności od tego, czy poprzedniego dnia ceny spadły, czy wzrosły. Współczynniki a_i' odpowiedzialne za zachowanie się inwestorów w czasie, gdy ceny spadają, przyjmują wartości ujemne co oznacza, że reakcją na spadek cen jest sprzedaż papierów i odwrotnie, w przypadku hossy inwestorzy chętnie kupują obiecujące papiery. Z drugiej strony taka sytuacja prowadzi do efektu umocnienia się trendów. Występują wtedy w modelu grupy zbyt mocno skorelowanych ze sobą wzrostów i spadków.