Efektywność estymatorów zmienności opartych na zakresie zmian

Paweł Sakowski

Korefereat do wystąpienia T. Skoczylasa „Modelowanie wariancji finansowych szeregów czasowych przy użyciu Range-based ARCH uwzględniającego heterogeniczność wariancji"

Konferencja WNE UW
Kazimierz nad Wisłą
27-28/9/2013

Uniwersytet Warszawski
Wydział Nauk Ekonomicznych

Plan prezentacji

1. Efektywność estymatorów zmienności
   - klasyczny estymator zmienności
   - estymator Parkinsona (1980)
   - estymator Osbanda (2007)
   - estymator Garmana-Klassa (1980)
   - estymator Rogersa-Satchella (1991)
2. Uwagi do referatu Tomka Skoczylasa

Założenia ogólne

- logarytm ceny akcji podlega jednowymiarowemu ciągłemu błądzeniu losowemu ze stałą dyfuzji D

Założenia ogólne

- logarytm ceny akcji podlega jednowymiarowemu ciągłemu błądzeniu losowemu ze stałą dyfuzji D
- a zatem, prawdopodobieństwo tego, że logarytm ceny x znajdzie się w przedziale $(x, x + dx)$ w momencie t, zakładając jego początkową wartość x_0 w momencie $t = 0$, wynosi:

$$\left(dx/\sqrt{2\pi Dt}\right) \exp[-(x - x_0)^2/2Dt]$$

Założenia ogólne

- logarytm ceny akcji podlega jednowymiarowemu ciągłemu błądzeniu losowemu ze stałą dyfuzji D
- a zatem, prawdopodobieństwo tego, że logarytm ceny x znajdzie się w przedziale $(x, x + dx)$ w momencie t, zakładając jego początkową wartość x_0 w momencie $t = 0$, wynosi:

$$\left(dx/\sqrt{2\pi Dt}\right)\exp[-(x - x_0)^2/2Dt]$$

- porównując to z rozkładem normalnym, zauważamy, że D jest wariancją zmiany logarytmu ceny $(x - x_0)$ w interwale jednostowym

Klasyczny estymator zmienności

Tradycyjny sposób estymacji wariancji D:

Klasyczny estymator zmienności

Tradycyjny sposób estymacji wariancji D:

- obserwujemy $x(t)$ w momentach $t = 0, 1, 2, 3, \ldots, n$

Klasyczny estymator zmienności

Tradycyjny sposób estymacji wariancji D:

- obserwujemy $x(t)$ w momentach $t = 0, 1, 2, 3, \ldots, n$
- definiujemy d_t jako zmiany w t-tym interwale:
  $d_t = x(t) - x(t-1)$, $t = 1, 2, 3, \ldots, n$

Klasyczny estymator zmienności

Tradycyjny sposób estymacji wariancji D:

- obserwujemy $x(t)$ w momentach $t = 0, 1, 2, 3, \ldots, n$
- definiujemy d_t jako zmiany w t-tym interwale:
  $d_t = x(t) - x(t-1)$, $t = 1, 2, 3, \ldots, n$
- obliczamy wyrażenie:

$$D_x = \frac{1}{n-1} \sum_{t=1}^{n} (d_t - \bar{d})^2$$

gdzie

$$\bar{d} = \frac{1}{n} \sum_{m=1}^{n} d_m$$

Klasyczny estymator zmienności

Tradycyjny sposób estymacji wariancji D:

- obserwujemy $x(t)$ w momentach $t = 0, 1, 2, 3, \ldots, n$
- definiujemy d_t jako zmiany w t-tym interwale:
  $d_t = x(t) - x(t-1)$, $t = 1, 2, 3, \ldots, n$
- obliczamy wyrażenie:

$$D_x = \frac{1}{n-1} \sum_{t=1}^{n} (d_t - \bar{d})^2$$

  gdzie

$$\bar{d} = \frac{1}{n} \sum_{m=1}^{n} d_m$$

- dla niskiej liczby obserwacji estymator ten ma małą precyzję!

Klasyczny estymator zmienności

Tradycyjny sposób estymacji wariancji D:

- obserwujemy $x(t)$ w momentach $t = 0, 1, 2, 3, \ldots, n$
- definiujemy d_t jako zmiany w t-tym interwale:
  $d_t = x(t) - x(t-1)$, $t = 1, 2, 3, \ldots, n$
- obliczamy wyrażenie:

$$D_x = \frac{1}{n-1} \sum_{t=1}^{n} (d_t - \bar{d})^2$$

gdzie

$$\bar{d} = \frac{1}{n} \sum_{m=1}^{n} d_m$$

- dla niskiej liczby obserwacji estymator ten ma małą precyzję!
  - zmienność zrealizowana (Andersen *et al.* 2001),
  - modele z rodziny GARCH (Engle 1982, Bollerslev 1986 i następni)

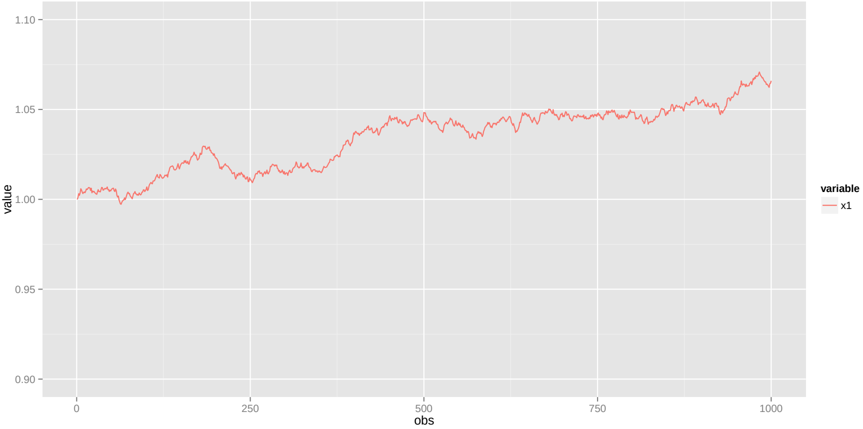
Symulacja Monte Carlo

- ścieżka cen $x(t)$ składa się z $n = 10000$ interwałów, $t = 1, 2, \ldots, 10000$
- ciągłe stopy zwrotu mają rozkład normalny:

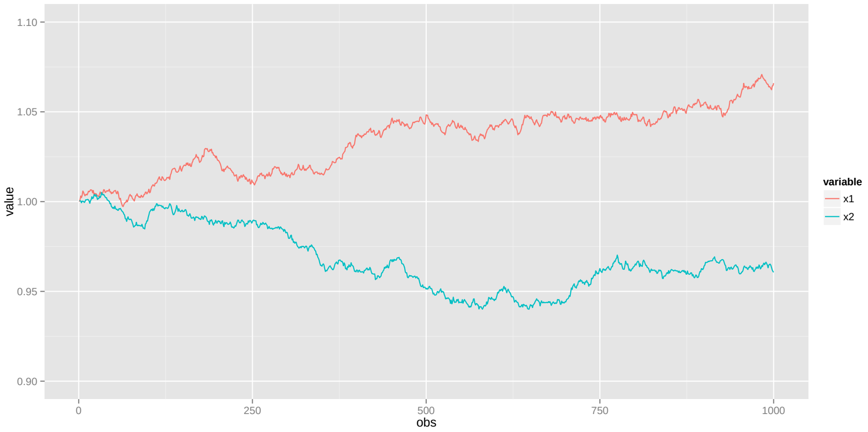
$$\log\left[x(t)\right] - \log\left[x(t-1)\right] \sim N(0, 0.001^2)$$

- liczba powtórzeń symulacji $=$ 10000

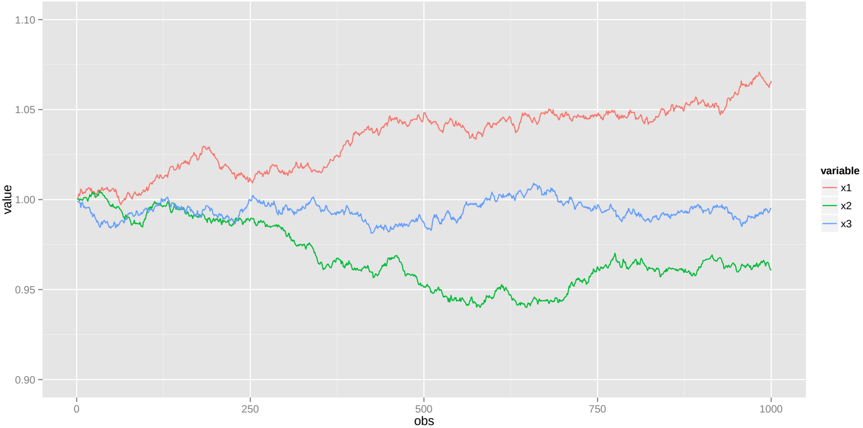
Błądzenie losowe



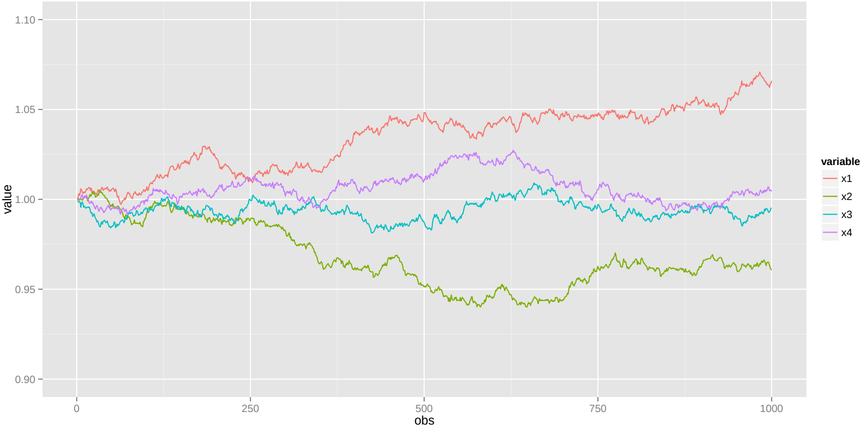
Błądzenie losowe



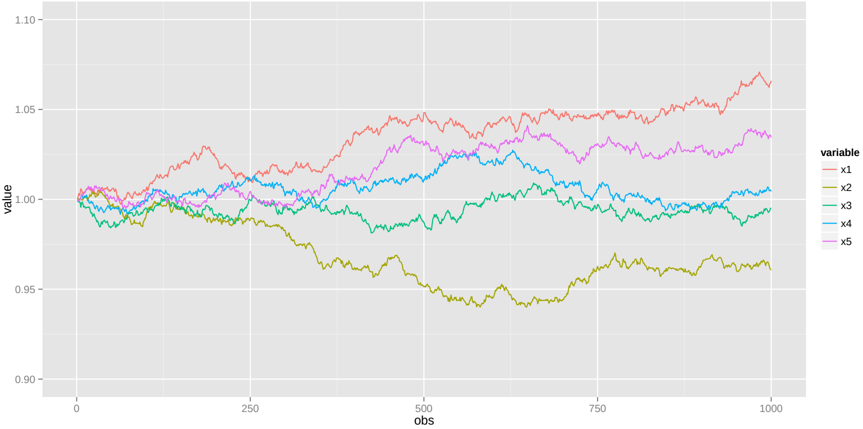
Błądzenie losowe



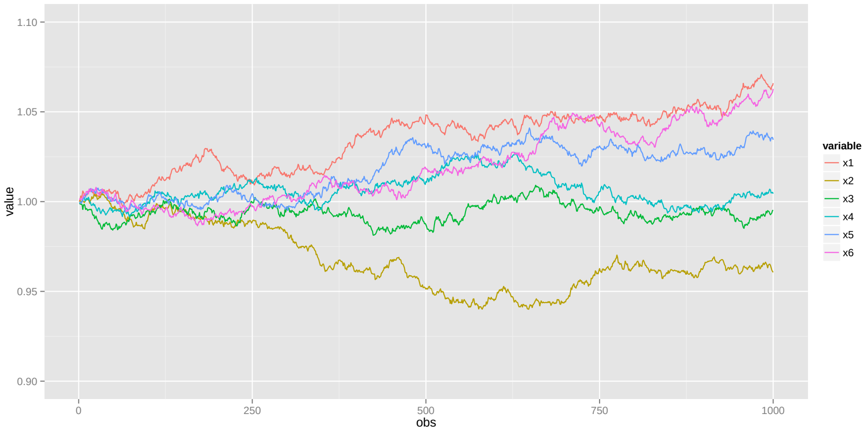
Błądzenie losowe



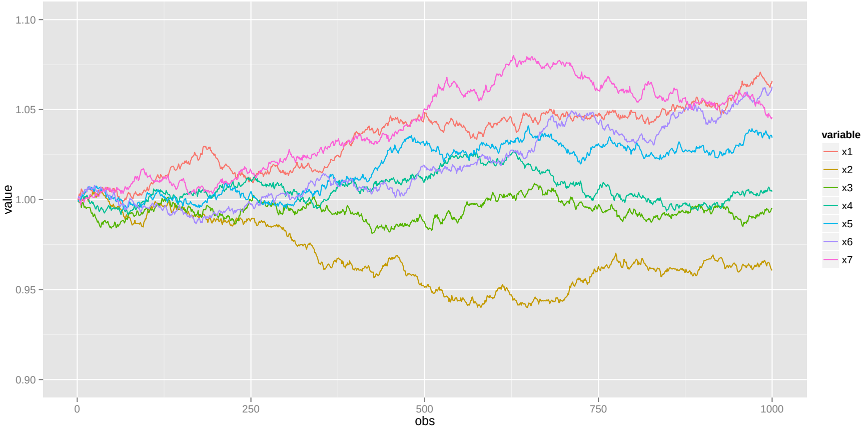
Błądzenie losowe



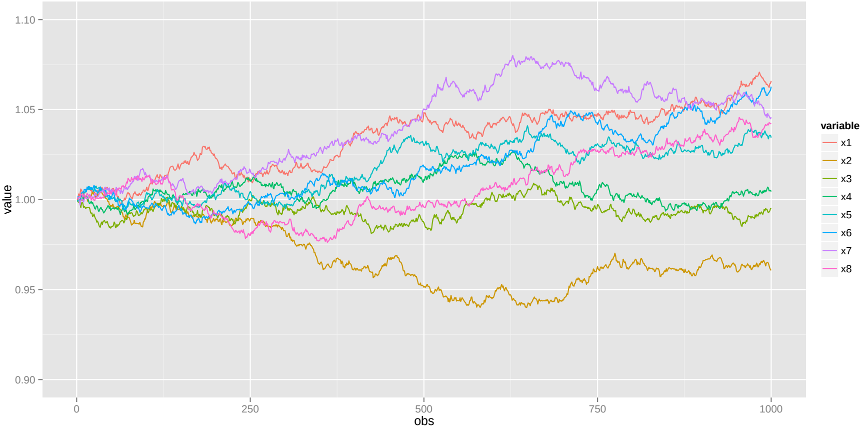
Błądzenie losowe



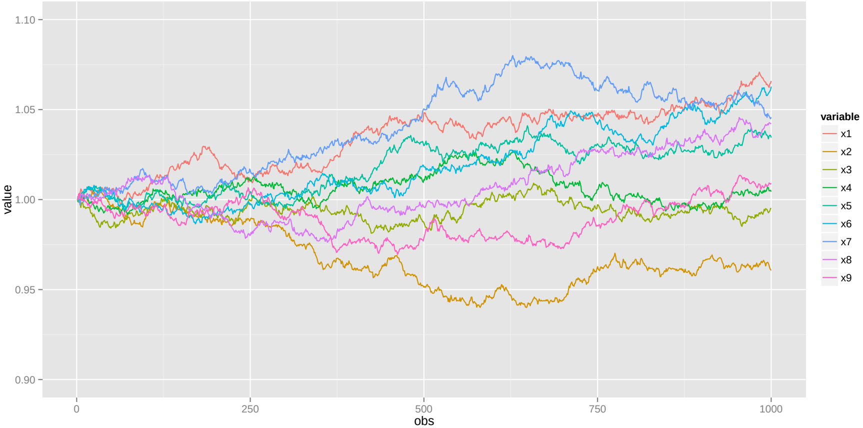
Błądzenie losowe



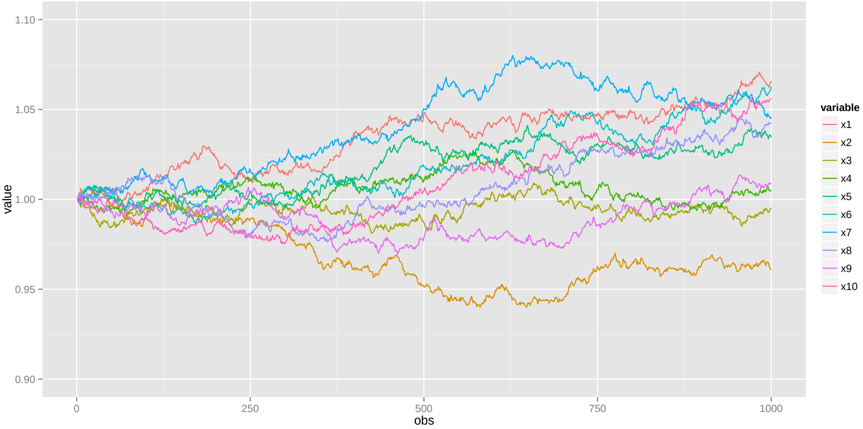
Błądzenie losowe



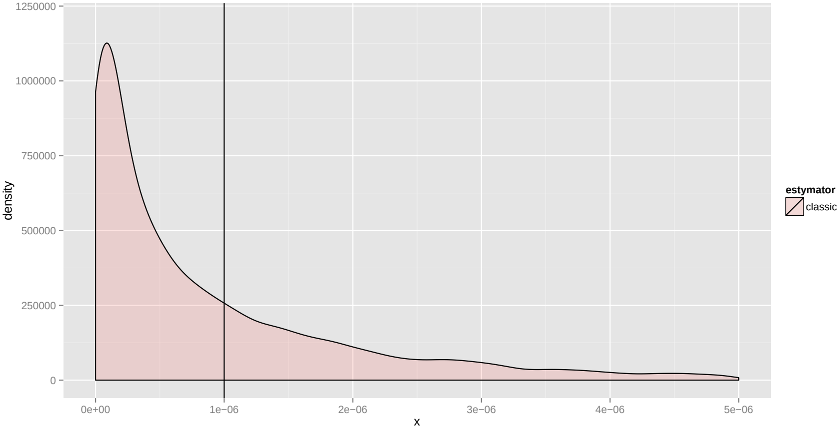
Błądzenie losowe



Błądzenie losowe



Symulowany rozkład klasycznego estymatora zmienności



Estymator D_l Parkinsona (1980)

- zamiast mierzyć $x(n)$ dla $n = 0, 1, 2, 3, \ldots$ obserwujemy różnicę l między logarytmami maksymalnej i minimalnej ceny wewnątrz danego interwału

$$\hat{\sigma}_P^2 = \frac{1}{4 \log 2} (\log H_t - \log L_t)^2$$

gdzie H_t i L_t to odpowiednio ceny maksymalne i minimalne.

Estymator D_l Parkinsona (1980)

- zamiast mierzyć $x(n)$ dla $n = 0, 1, 2, 3, \ldots$ obserwujemy różnicę l między logarytmami maksymalnej i minimalnej ceny wewnątrz danego interwału

$$\hat{\sigma}_P^2 = \frac{1}{4 \log 2} (\log H_t - \log L_t)^2$$

  gdzie H_t i L_t to odpowiednio ceny maksymalne i minimalne.
- intuicja: taki zakres powinien być bardziej efektywnym pomiarem wariancji, w porównaniu ze zmianą ceny między dwoma arbitralnie wybranymi miejscami w czasie (tj. początkiem i końcem danego interwału).

Estymator D_l Parkinsona (1980)

- Rozkład i własności zakresu zmian l są znane. Zachodzi m. in.:

$$E[l^p] = \frac{4}{\sqrt{\pi}}\Gamma\left(\frac{p+1}{2}\right)\left(1-\frac{4}{2^p}\right)\zeta(p-1)(2Dt)^{p/2}$$

gdzie $p \geq$, a $\zeta(x)$ jest funkcją zeta Riemanna.

- zaś w szczególności mamy:

$$E[l] = \sqrt{8Dt/\pi}$$

oraz

$$E[l^2] = (4\log 2)Dt$$

Estymator D_l Parkinsona (1980)

- dla interwału jednostkowego zachodzi także:

$$D = 0.393(E[l])^2 = 0.361E[l^2]$$

- mamy zatem relację:

$$E[l^2] = 1.09(E[l])^2$$

  która jest wygodna przy testowaniu tego, czy obserwowane zakresy zmian l pochodzą z procesu błądzenia losowego

- w rezultacie, mając zbiór zakresów zmian $(l_1, l_2, \cdots, l_n)$ obserwowanych w n jednostkowych interwałach, estymatorem wariancji D jest:

$$D_l = \frac{0.361}{n} \sum_{i=1}^{n} l_i^2$$

porównanie estymatorów D_x i D_l

- wariancja estymatora D_x

$$E[(D_x - D)^2] = \left[\frac{E(x^4)}{E(x^2)} - 1\right]\frac{D^2}{N_x} = \frac{2D^2}{N_x}$$

- wariancja estymatora D_l

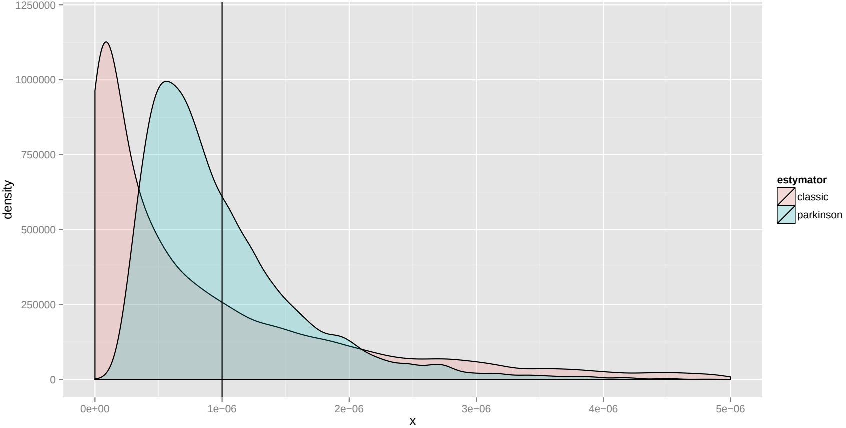
$$E[(D_l - D)^2] = \left[\frac{E(l^4)}{E(l^2)} - 1\right]\frac{D^2}{N_l} = \frac{0.41D^2}{N_l}$$

- A zatem, wariancje estymatorów zbliżone jeśli zachodzi:

$$N_x \approx 5N_l$$

co oznacza, że stosując D_l zamiast D_x możemy zredukować liczbę obserwacji o ok. 80% zachowując tę samą precyzję estymatora!

Symulowany rozkład estymatora Parkinsona (1980)



Inne estymatory oparte na zakresie zmian

- Estymator Osbanda (2007)

$$\begin{aligned}\hat{\sigma}^2_{\mathrm{Osb}} &= 0.84(\log H_t - \log L_t) - \\ &\quad -0.39|\log C_t - \log O_t|\end{aligned}$$

gdzie O_t i C_t to odpowiednio ceny otwarcia i zamknięcia w interwale t.

Inne estymatory oparte na zakresie zmian

- Estymator Osbanda (2007)

$$\begin{aligned}\hat{\sigma}^2_{\mathrm{Osb}} &= 0.84(\log H_t - \log L_t) - \\ &\quad -0.39|\log C_t - \log O_t|\end{aligned}$$

gdzie O_t i C_t to odpowiednio ceny otwarcia i zamknięcia w interwale t.

- Estymator Garmana-Klassa (1980)

$$\begin{aligned}\hat{\sigma}^2_{\mathrm{GK}} &= 0.5[\log H_t - \log L_t]^2 - \\ &\quad -(2\log 2 - 1)[\log C_t - \log O_t]^2\end{aligned}$$

Inne estymatory oparte na zakresie zmian

- Estymator Osbanda (2007)

$$\begin{aligned}\hat{\sigma}^2_{\mathrm{Osb}} &= 0.84(\log H_t - \log L_t) - \\ &\quad -0.39|\log C_t - \log O_t|\end{aligned}$$

  gdzie O_t i C_t to odpowiednio ceny otwarcia i zamknięcia w interwale t.

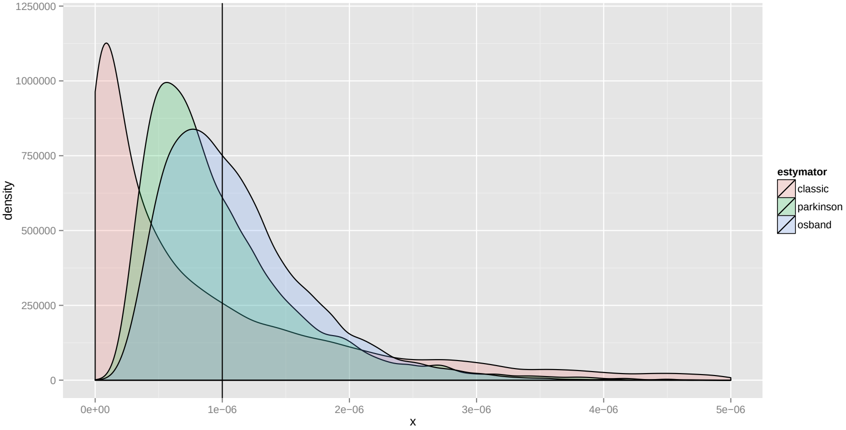
- Estymator Garmana-Klassa (1980)

$$\begin{aligned}\hat{\sigma}^2_{\mathrm{GK}} &= 0.5[\log H_t - \log L_t]^2 - \\ &\quad -(2\log 2 - 1)[\log C_t - \log O_t]^2\end{aligned}$$

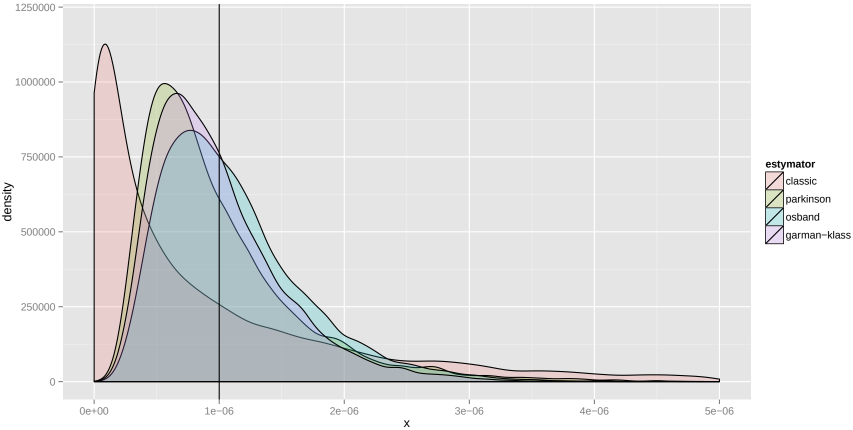
- Estymator Rogersa & Satchella (1991)

$$\begin{aligned}\hat{\sigma}^2_{\mathrm{RS}} &= \log(H_t/O_t)[\log(H_t/O_t) - \log(C_t/O_t)] + \\ &\quad + \log(L_t/O_t)[\log(L_t/O_t) - \log(C_t/O_t)]\end{aligned}$$

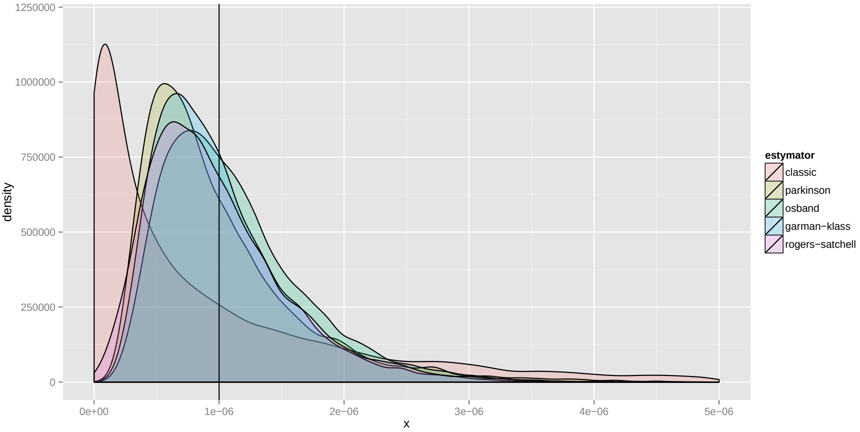
Symulowany rozkład estymatora Osbanda (2007)



Symulowany rozkład estymatora Garmana-Klassa (1980)



Symulowany rozkład estymatora Rogersa-Satchella (1991)



Statystyki próbkowe estymatorów zmienności

prawdziwa wartość $\sigma^2 = 1.0 \times 10^{-6}$

	średnia $\times 10^6$	odch. std. $\times 10^6$	odch. std. do klasycznego	redukcja obs.
klasyczny	1.02	1.44	-	-
Parkinson	0.99	0.65	45%	79.8%
Osband	1.11	0.57	40%	84.2%
Garman-Klass	0.99	0.52	37%	87.9%
Rogers-Satchell	0.99	0.57	40%	84.2%

Uwagi do referatu Tomka Skoczylasa

- Czy model RHARCH może być szacowany za pomocą powszechnie dostępnych pakietów statystycznych?
- Czy MSE prognoz różnią się istotnie od siebie?
  Warto rozważyć raportowanie odch. std. miar MSE.
- Jak zdefiniowana jest miara QLIKE?
- Błędy w oznaczeniach w równaniach?
  Model RGARCH(1,1)

$$\begin{aligned} r_t &= \mu + \varepsilon_t \\ \varepsilon_t &\sim N\Big(0, \frac{1}{4\log 2}\lambda_t^2\Big) \\ \lambda_t &= \omega + \alpha R_{t-1} + \beta \lambda_{t-1} \\ R_{t-1} &= \log(H_{t-1}/L_{t-1}) \overset{?}{=} \lambda_{t-1} \end{aligned}$$

$$\sigma_{\mathrm{P}}^2 = \frac{1}{4\log 2}\Big(\log(H_t/L_t)\Big)^2$$

Dziękuję!

Paweł Sakowski
sakowski@wne.uw.edu.pl