Variance-covariance estimation

Yuta Koike

Mathematics and Informatics Center, The University of Tokyo
CREST JST

June 28, 2019

YSS2019

1. Realized volatility
   - Asymptotic distribution
   - Statistical inference
   - Implementation in YUIMA

2. Covariance estimation
   - Epps effect
   - Hayashi-Yoshida estimator
   - Statistical inference

Realized volatility

- Let $X = (X_t)_{t\in[0,T]}$ be a stochastic process given by the following equation:

$$X_t = X_0 + \int_0^t b_s ds + \int_0^t \sigma_s dB_s,$$

  where
  - $b = (b_s)_{s\in[0,T]}$: càdlàg adapted process (**drift process**)
  - $\sigma = (\sigma_s)_{s\in[0,T]}$: càdlàg adapted process (**volatility process**)
  - $B = (B_s)_{s\in[0,T]}$: Brownian motion

  (càdlàg: right continuous with left limits)

- The **quadratic variation** of X:

$$[X, X]_T = \int_0^T \sigma_s^2 ds$$

Realized volatility

- In high-frequency financial econometrics,
  - X is used as a model for the intraday log-price process of an asset ($[0, T]$ corresponds to one day)
  - Then, $[X, X]_T$ is regarded as a proxy of the variance of X and called the **integrated volatility** or **integrated variance** (IV)
- We observe X at discrete sampling times $0 \leq t_0 < t_1 < \cdots < t_N \leq T$ in $[0, T]$
- The statistic

$$\widehat{[X, X]}_T := \sum_{i=1}^{N} \left(X_{t_i} - X_{t_{i-1}} \right)^2$$

is called the **realized volatility** or **realized variance** (RV)

Realized volatility

- Suppose that $t_0, t_1, \dots, t_N$ are stopping times, i.e. the point process $(\sum_{i=1}^N 1_{\{t_i \leq t\}})_{t\in[0,T]}$ is adapted
- Then, a general theory of stochastic calculus implies that

$$\widehat{[X,X]}_T \to^p [X,X]_T$$

as

$$r_N := t_0 \vee \max_{1\leq i\leq N}(t_i - t_{i-1}) \vee (T - t_N) \to^p 0$$

("$\to^p$" denotes convergence in probability)

  - See e.g. Thm. 4.47 in Chap. I of Jacod and Shiryaev (2003)

$\Rightarrow$ $\widehat{[X,X]}_T$ is a consistent estimator for $[X,X]_T$

Realized volatility: Asymptotic distribution

- Now we assume that t_i's are equi-spaced: $t_i = i\Delta_N$ $(i = 0, 1, \dots, N)$ with $\Delta_N = T/N$
- We are interested in the asymptotic distribution of $\widehat{[X,X]}_T - [X,X]_T$ as $N \to \infty$

Definition 1 (Stable convergence)

A sequence of random variables $U_1, U_2, \dots$ are said to **converge stably in law** to a random variable U_∞ if

$$(U_N, V) \to^d (U_\infty, V)$$

as $N \to \infty$ for any random variable V ("$\to^d$" denotes convergence in law). In this case, we write $U_N \to^{d_s} U_\infty$.

- See Section 2.1 in Podolskij and Vetter (2010) for a concise description of this concept

Realized volatility: Asymptotic distribution

Theorem 1 (Asymptotic mixed normality of realized volatility)

$$\frac{1}{\sqrt{\Delta_N}}\left(\widehat{[X,X]}_T - [X,X]_T\right) \to^{d_s} \sqrt{2IQ_T}\zeta$$

as $N \to \infty$, where

$$IQ_T := \int_0^T \sigma_t^4 dt$$

and ζ is a standard normal variable independent of X

- IQ_T is called the **integrated quarticity**
- Note that IQ_T is NOT independent of X
- See Theorem 5.4.2 in Jacod and Protter (2012) for a proof; see also pp. 341–343 of Podolskij and Vetter (2010) for a sketch of the proof

Realized volatility: Statistical inference

- Assume $IQ_T > 0$ almost surely
- By Theorem 1, we have

$$\left(\frac{1}{\sqrt{\Delta_N}} \left(\widehat{[X,X]}_T - [X,X]_T \right), IQ_T \right) \to^d \left(\sqrt{2IQ_T}\zeta, IQ_T \right)$$

- Thus, the continuous mapping theorem yields

$$\frac{\widehat{[X,X]}_T - [X,X]_T}{\sqrt{\Delta_N \cdot 2IQ_T}} \to^d \zeta \tag{1}$$

- <u>Caution</u>
  - The convergence $\frac{1}{\sqrt{\Delta_N}} \left(\widehat{[X,X]}_T - [X,X]_T \right) \to^d \sqrt{2IQ_T}\zeta$ alone does NOT imply (1) in general because IQ_T is NOT independent of X; this is why we need stable convergence in Theorem 1

- We can construct e.g. confidence intervals for $\widehat{[X,X]}_T$ using (1) once we have a consistent estimator for IQ_T
- To construct a consistent estimator for IQ_T, we introduce the **realized p-variation** ($p > 0$):

$$RPV(X;p)^N_T := N^{p/2-1} \sum_{i=1}^{N} \left| X_{t_i} - X_{t_{i-1}} \right|^p .$$

Theorem 2 (LLN for the realized p-variation)

$$RPV(X;p)^N_T \to^p T^{p/2-1} \cdot m_p \int_0^T |\sigma_s|^p ds$$

as $N \to \infty$. Here, m_p denotes the p-th absolute moment of the standard normal distribution:

$$m_p = \frac{2^{p/2}}{\sqrt{\pi}} \Gamma\left(\frac{p+1}{2}\right).$$

- This is a special case of Theorem 3.4.1 in Jacod and Protter (2012)
- Since $m_4 = 3$, $RPV(X;4)^N_T \to^p T \cdot 3IQ_T$ as $N \to \infty$
- Thus we obtain the following feasible CLT:

$$\frac{\widehat{[X,X]}_T - [X,X]_T}{\sqrt{\frac{2}{3N} RPV(X;4)^N_T}} \to^d \zeta \qquad (N \to \infty) \tag{2}$$

- A $100(1-\alpha)\%$ confidence interval for $[X,X]_T$:

$$\left[\widehat{[X,X]}_T - z_{\alpha/2}\sqrt{\frac{2}{3N} RPV(X;4)^N_T},\ \widehat{[X,X]}_T + z_{\alpha/2}\sqrt{\frac{2}{3N} RPV(X;4)^N_T} \right], \tag{3}$$

where $z_{\alpha/2}$ denotes the $(1-\alpha/2)$-quantile of $N(0,1)$

Realized volatility: Statistical inference

- The CI (3) has the problem that its lower limit is not necessarily positive (we always have $[X, X]_T > 0$ a.s. under the present assumption)
- We can resolve this issue by considering CIs for $\log([X, X]_T)$ rather than $[X, X]_T$
- CIs for $\log([X, X]_T)$ can be derived by the delta method:

Theorem 3 (Delta method for stable convergence)

Let F_N $(N = 1, 2, \dots)$, F_∞ and U_∞ be random variables and $g : \mathbb{R} \to \mathbb{R}$ be a C^1 function. Suppose that $r_N^{-1}(F_N - F_\infty) \to^{d_s} U_\infty$ as $N \to \infty$ for some $r_N > 0$ with $r_N \to 0$. Then

$$r_N^{-1}(g(F_N) - g(F_\infty)) \to^{d_s} g'(F_\infty)U_\infty.$$

Realized volatility: Statistical inference

- A proof of Thm. 3 can be found in e.g. Proposition 2 in Podolskij and Vetter (2010) (they consider the case $r_N = 1/\sqrt{N}$, but this is not essential)
- Applying Thm. 3 with $g(x) = \log x$, we obtain

$$\frac{1}{\sqrt{\Delta_N}}\left\{\log\left(\widehat{[X,X]}_T\right) - \log\left([X,X]_T\right)\right\} \to^{d_s} \frac{\sqrt{2IQ_T}}{[X,X]_T}\zeta \quad (N \to \infty)$$

- Thus we get

$$\frac{\log\left(\widehat{[X,X]}_T\right) - \log\left([X,X]_T\right)}{\sqrt{\frac{2}{3N}RPV(X;4)^N_T} \Big/ \widehat{[X,X]}_T} \to^{d} \zeta \quad (N \to \infty) \tag{4}$$

Realized volatility: Statistical inference

- Hence,

$$\left[\log\left(\widehat{[X,X]}_T\right) - z_{\alpha/2}\frac{\sqrt{\frac{2}{3N}RPV(X;4)^N_T}}{\widehat{[X,X]}_T}, \log\left(\widehat{[X,X]}_T\right) + z_{\alpha/2}\frac{\sqrt{\frac{2}{3N}RPV(X;4)^N_T}}{\widehat{[X,X]}_T}\right]$$

  gives a $100(1-\alpha)\%$ confidence interval for $\log([X,X]_T)$

- By exponential transform, we obtain another $100(1-\alpha)\%$ confidence interval for $[X,X]_T$:

$$\left[e^{\log\left(\widehat{[X,X]}_T\right)-z_{\alpha/2}\frac{\sqrt{\frac{2}{3N}RPV(X;4)^N_T}}{\widehat{[X,X]}_T}}, e^{\log\left(\widehat{[X,X]}_T\right)+z_{\alpha/2}\frac{\sqrt{\frac{2}{3N}RPV(X;4)^N_T}}{\widehat{[X,X]}_T}}\right] \quad (5)$$

Realized volatility: Remarks on irregular sampling schemes

- The convergence (2) is still valid even if the sampling times t_i are NOT equi-spaced under some mild regularity assumptions, as long as t_i's are independent of X
  - See Theorem 3.10 in Hayashi et al. (2011) for details
- Thus, the CIs (3) and (5) are still valid in such a situation
- If t_i's are NOT independent of X, the convergence (2) does NOT hold in general
  - See Fukasawa (2010) and Li et al. (2014) for details

Realized volatility in YUIMA

- The package `yuima` has the function `mpv()` to compute realized *p*-variations
- For a yuima object `x`,
  - Realized volatility: `mpv(x,r=2)`
  - $RPV(X;4)^N_T/3$: `mpv(x,r=4)`
  - The lower and upper limits of the CI (3):

```
mpv(x,r=2)-sqrt(2*mpv(x,r=4)/N)*qnorm(1-alpha/2),
mpv(x,r=2)+sqrt(2*mpv(x,r=4)/N)*qnorm(1-alpha/2)
```

  - The lower and upper limits of the CI (5):

```
exp(log(mpv(x,r=2))-qnorm(1-alpha/2)*sqrt(2*mpv(x,r=4)/N
    )/mpv(x,r=2)),
exp(log(mpv(x,r=2))+qnorm(1-alpha/2)*sqrt(2*mpv(x,r=4)/N
    )/mpv(x,r=2))
```

Realized volatility in YUIMA: SImulation

- Let us estimate the IV of the process $X_t = \log S_t,\ t \in [0,1]$, where S_t is given by the Heston model:

$$\begin{cases} dS_t = \mu S_t dt + \sqrt{V_t} S_t dB_{1,t}, \\ dV_t = -\theta(V_t - v)dt + \gamma\sqrt{V_t}(\rho dB_{1,t} + \sqrt{1-\rho^2} dB_{2,t}), \end{cases}$$

  where $(B_{1,t})_{t\in[0,T]}$ and $(B_{2,t})_{t\in[0,T]}$ are two independent Brownian motions

- In this case we have

$$[X,X]_1 = \int_0^1 V_t dt$$

- **R examples**: `rv.r`, `rv-spx.r`

```r
> ### Realized volatility in Yuima
> ## Simulated data example
> ## We simulate the Heston model
> drift <- c("mu*S", "-theta*(V-v)") # drift coefficient
> diffusion <- matrix(c("sqrt(max(V,0))*S", "gamma*sqrt(max(V,0))*rho",
+                       0, "gamma*sqrt(max(V,0))*sqrt(1-rho^2)"),
+                     2,2) # diffusion coefficient
> mod <- setModel(drift = drift, diffusion = diffusion,
+                 state.variable = c("S", "V"))
> samp <- setSampling(n = 10000)
> heston <- setYuima(model = mod, sampling = samp)
> set.seed(123)
> x0 <- c(1, 0.1) # initial value
> param <- list(mu = 0.03, theta = 3, v = 0.09,
+               gamma = 0.3, rho = -0.6) # true parameters
> result <- simulate(heston, xinit = x0,
+                    true.parameter = param) # simulation
> ## construct the yuima object correspoding to the log-price process
> zdata <- get.zoo.data(result) # extract the zoo data
> x <- zdata[[1]] # extract the first component (price process)
> x <- log(x) # convert to the log-rpice process
```

```
> x <- setData(x) # convert to the yuima object
> # we subsample data to construct observation data
> N <- 100 # number of observations - 1
> x <- subsampling(x, sampling = setSampling(n = N))
> ## Estimation of IV
> (rv <- mpv(x)) # computing the realized volatility
[1] 0.07039031
> (iv <- mean(zdata[[2]][-1])) # "true" integrated volatility
[1] 0.08454659
> ## Construction of CIs for IV
> alpha <- 0.05 # significance level
> # CI based on Eq.(3)
> c(mpv(x,r=2)-sqrt(2*mpv(x,r=4)/N)*qnorm(1-alpha/2),
+   mpv(x,r=2)+sqrt(2*mpv(x,r=4)/N)*qnorm(1-alpha/2))
[1] 0.05239262 0.08838799
> # CI based on Eq.(5)
> c(exp(log(mpv(x,r=2))-qnorm(1-alpha/2)*sqrt(2*mpv(x,r=4)/N )/mpv(x,r
    =2)),
+ exp(log(mpv(x,r=2))+qnorm(1-alpha/2)*sqrt(2*mpv(x,r=4)/N)/mpv(x,r=2)))
[1] 0.05450930 0.09089815
```

Covariance estimation

- Let $X = (X_t)_{t\in[0,T]}$ be a d-dimensional stochastic process given by the following equation:

$$X_t = X_0 + \int_0^t b_s ds + \int_0^t \sigma_s dB_s,$$

  where
  - $b = (b_s)_{s\in[0,T]}$: d-dimensional càdlàg adapted process
  - $\sigma = (\sigma_s)_{s\in[0,T]}$: $d \times r$ matrix-valued càdlàg adapted process
  - $B = (B_s)_{s\in[0,T]}$: r-dimensional Brownian motion
- We denote by $X^{(j)} = (X_t^{(j)})_{t\in[0,T]}$ the j-th component process of X

Covariance estimation

- The quadratic covariation matrix of X:

$$[X,X]_T := \int_0^T \sigma_s \sigma_s^\top ds$$

($\top$ stands for transposition of a matrix)

  - A $d \times d$ (random) matrix, the multi-dimensional extension of the quadratic variation

- For every i, we observe $X^{(j)}$ at discrete sampling times $0 \leq t_0^{(j)} < t_1^{(j)} < \cdots < t_{N_j}^{(j)} \leq T$ in $[0, T]$
- We are interested in estimating $[X,X]_T$ based on the observation data $(X_{t_i}^{(j)})_{i=0}^{N_j}$, $j = 1, \dots, d$

Covariance estimation

- The diagonal entries $[X^{(j)}, X^{(j)}]_T$ can be estimated by the realized volatilities
  $\Longrightarrow$ We focus on the off-diagonal entries, say $[X^{(1)}, X^{(2)}]_T$
- If $N_1 = N_2 =: N$ and $t_i^{(1)} = t_i^{(2)} =: t_i$ for all i (synchronous case), we can use the following natural extension of the realized volatility:

$$\widehat{[X^{(1)}, X^{(2)}]}_T := \sum_{i=1}^{N} \left(X_{t_i}^{(1)} - X_{t_{i-1}}^{(1)} \right) \left(X_{t_i}^{(2)} - X_{t_{i-1}}^{(2)} \right),$$

  which is known as the **realized covariance**
- This is generally not the case. What should we do then?

Epps effect

- <u>A naïve idea</u> Interpolating the data onto a equi-spaced sampling grid to construct synchronized data:
  - Choose a step size $h > 0$ and set

$$\tau_k^{(1)} := \max\{t_i^{(1)} : t_i^{(1)} \leq kh\}, \qquad \tau_k^{(2)} := \max\{t_i^{(2)} : t_i^{(2)} \leq kh\}$$

    for $k = 0, 1, \dots, Th^{-1} =: n_h$ (assume n_h is an integer)
  - Compute the realized covariance based on $(X^{(1)}_{\tau_k^{(1)}})_{k=0}^{n_h}$ and $(X^{(2)}_{\tau_k^{(2)}})_{k=0}^{n_h}$:

$$[\widehat{X^{(1)}, X^{(2)}}]_{T,h} := \sum_{k=1}^{n_h} \left(X^{(1)}_{\tau_k^{(1)}} - X^{(1)}_{\tau_{k-1}^{(1)}} \right) \left(X^{(2)}_{\tau_k^{(2)}} - X^{(2)}_{\tau_{k-1}^{(2)}} \right)$$

- $\tau_k^{(j)} \approx kh$ because the data are observed at a high-frequency, so we expect that $[\widehat{X^{(1)}, X^{(2)}}]_{T,h} \approx [X, X]_T$ when h is sufficiently small

Epps effect

- This is generally NOT true!
  - $\widehat{[X^{(1)}, X^{(2)}]}_{T,h}$ is often strongly downward biased in the absolute value as $h \to 0$ (Epps, 1979)
  - One can show that $\widehat{[X^{(1)}, X^{(2)}]}_{T,h}$ is ALWAYS downward biased in the absolute value when original observation times are non-synchronous under mild regularity condition (Hayashi and Yoshida, 2005, Proposition 2.1)
- This phenomenon is known as the **Epps effect**
- Let us check the Epps effect by simulation

Epps effect: Simulation

- We simulate the following two-dimensional stochastic processes:

$$X_t^{(1)} = B_t^{(1)}, \quad X_t^{(2)} = \rho B_t^{(1)} + \sqrt{1-\rho^2} B_t^{(2)}, \qquad t \in [0,1]$$

  where $B^{(1)}$ and $B^{(2)}$ are two independent Brownian motions and ρ is the correlation parameter; we set $\rho = 0.5$ here

- For each $j = 1, 2$, we generate the sampling times $t_0^{(j)}, t_1^{(j)}, \dots$ so that

$$t_1^{(j)} - t_0^{(j)}, t_2^{(j)} - t_1^{(j)}, \dots$$

  are i.i.d. variables following the exponential distribution with rate λ_j; this is called the **Poisson random sampling** with intensity λ_j
  - $(t_i^{(1)})_{i=1}^{N_1}$ and $(t_i^{(2)})_{i=1}^{N_2}$ are independently generated
  - We set $\lambda_1 = \lambda_2 = n/5$ with $n = 10,000$

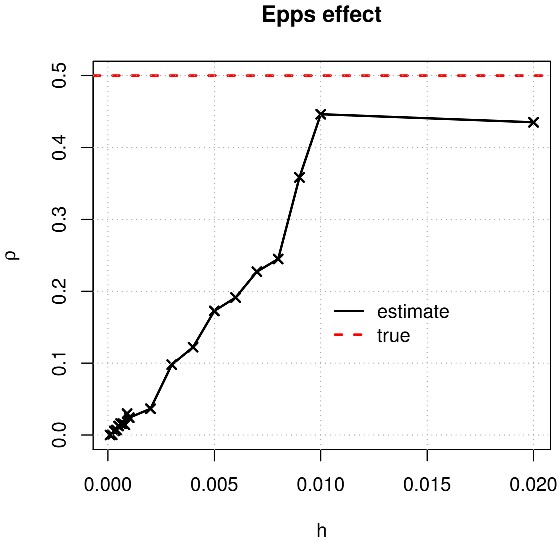
Epps effect: Simulation

- We compute $\widehat{[X^{(1)}, X^{(2)}]}_{T,h}$ with varying h as

$$h = 1/n, 2/n, \dots, 10/n, 20/n, \dots, 100/n, 200/n$$

- The true value of $[X^{(1)}, X^{(2)}]_1$ is $\rho = 0.5$, so they should be close to 0.5 if they correctly estimate $[X^{(1)}, X^{(2)}]_1$
- **R example**: `epps.r`

Epps effect: Simulation



Hayashi-Yoshida estimator

- Hayashi and Yoshida (2005) resolved this problem by proposing the following novel covariance estimator:

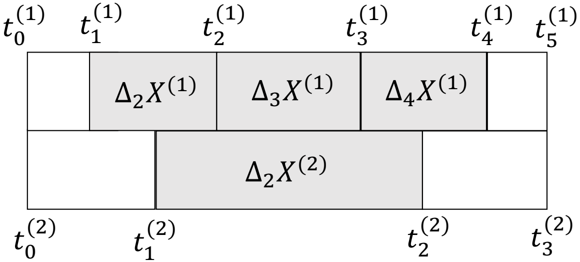
$$\widehat{[X^{(1)}, X^{(2)}]}_T^{HY} := \sum_{i,j} \Delta_i X^{(1)} \Delta_j X^{(2)} 1_{\{(t_{i-1}^{(1)}, t_i^{(1)}] \cap (t_{j-1}^{(2)}, t_j^{(2)}] \neq \emptyset\}}$$

  where $\Delta_i X^{(p)} := X^{(p)}_{t_i^{(p)}} - X^{(p)}_{t_{i-1}^{(p)}}$ for $i = 1, \dots, N_p$ and $p = 1, 2$

- Note that $\widehat{[X^{(1)}, X^{(2)}]}_T^{HY}$ is reduced to the realized covariance when the observation times are synchronous
- In this sense, the Hayashi-Yoshida estimator is a natural extension of the realized covariance

Hayashi-Yoshida estimator

Fig. 1: Hayashi-Yoshida estimator. We sum up cross-products of returns with overlapping observation intervals



Hayashi-Yoshida estimator

- Under mild regularity assumptions, we have

$$\widehat{[X^{(1)}, X^{(2)}]}^{HY}_T \to^p [X^{(1)}, X^{(2)}]_T$$

as

$$r_N := \max_{p=1,2} \left[t_0^{(p)} \vee \max_{1 \le i \le N_p} (t_i^{(p)} - t_{i-1}^{(p)}) \vee (T - t_{N_p}^{(p)}) \right] \to^p 0$$

$\Longrightarrow$ The Hayashi-Yoshida estimator is a consistent estimator for $[X^{(1)}, X^{(2)}]_T$

- See Hayashi and Yoshida (2005) and Hayashi and Kusuoka (2008) for detailed regularity conditions

Hayashi-Yoshida estimator in YUIMA

- The package `yuima` has the function `cce()` to compute the Hayashi-Yoshida estimator from a yuima object
- More precisely, it computes the estimate of the quadratic covariation matrix

$$\left(\widehat{[X^{(i)}, X^{(j)}]}_T^{HY}\right)_{1\leq i,j\leq d}$$

and its correlation matrix counterpart:

$$\left(\frac{\widehat{[X^{(i)}, X^{(j)}]}_T^{HY}}{\sqrt{\widehat{[X^{(i)}, X^{(i)}]}_T^{HY}\widehat{[X^{(j)}, X^{(j)}]}_T^{HY}}}\right)_{1\leq i,j\leq d} \tag{6}$$

Hayashi-Yoshida estimator in YUIMA

- If you implement `cce(psample)` for the simulated data of `epps.r`, you will obtain the following output

```
> cce(psample)
$covmat
          Series 1  Series 2
Series 1 0.9530663 0.4853256
Series 2 0.4853256 1.0424305

$cormat
          Series 1  Series 2
Series 1 1.0000000 0.4869093
Series 2 0.4869093 1.0000000
```

Hayashi-Yoshida estimator: Statistical inference

- The Hayashi-Yoshida estimator also enjoys the asymptotic mixed normality under some regularity assumptions:

$$\frac{1}{\sqrt{\Delta_n^*}}\left(\widehat{[X^{(i)},X^{(j)}]}^{HY}_T - [X^{(i)},X^{(j)}]_T\right) \to^{d_s} \sqrt{\mathbf{AVAR}^{HY}_{T,ij}} \times \zeta \qquad (7)$$

  - $\Delta_n^* \asymp \sum_{i=1}^{N_1}(t_i^{(1)} - t_{i-1}^{(1)})^2 + \sum_{j=1}^{N_2}(t_j^{(2)} - t_{j-1}^{(2)})^2$
  - $\mathbf{AVAR}^{HY}_{T,ij}$: Asymptotic variance (independent of ζ)
- $\mathbf{AVAR}^{HY}_{T,ij}$ has the explicit but somewhat complicated expression (cf. Eq. (4.1) in Hayashi and Yoshida, 2011)

Hayashi-Yoshida estimator: Statistical inference

- The package `yuima` has the function `hyavar()` to construct estimators for $\sqrt{\Delta_n^*} \cdot \mathbf{AVAR}_{T,ij}^{HY}$ $(i,j = 1,\dots,d)$ following Section 8.2 of Hayashi and Yoshida (2011)
- It also computes asymptotic variance estimates for entries of the correlation matrix (6)
- If you implement `hyavar(psample)` for the simulated data of `epps.r`, you will obtain the output shown in the next slide

```
> hyavar(psample)
$covmat
          Series 1  Series 2
Series 1 0.9530663 0.4853256
Series 2 0.4853256 1.0424305

$cormat
          Series 1  Series 2
Series 1 1.0000000 0.4869093
Series 2 0.4869093 1.0000000

$avar.cov
            [,1]        [,2]
[1,] 0.001481935 0.002300558
[2,] 0.002300558 0.002823336

$avar.cor
            [,1]        [,2]
[1,] 0.000000000 0.001614687
[2,] 0.001614687 0.000000000
```

Hayashi-Yoshida estimator: Statistical inference

- The convergence (7) can be used to construct CIs for $[X^{(i)}, X^{(j)}]_T$ as usual
- $100(1-\alpha)\%$ CI for $[X^{(i)}, X^{(j)}]_T$:

$$\Bigg[\widehat{[X^{(i)}, X^{(j)}]}^{HY}_T - z_{\alpha/2}\sqrt{\Delta_n^*}\cdot \mathbf{AVAR}^{HY}_{T,ij}, \\ \widehat{[X^{(i)}, X^{(j)}]}^{HY}_T + z_{\alpha/2}\sqrt{\Delta_n^*}\cdot \mathbf{AVAR}^{HY}_{T,ij}\Bigg]$$

(Of course, we need to replace $\sqrt{\Delta_n^*}\cdot \mathbf{AVAR}^{HY}_{T,ij}$ by its estimate in practice)

Hayashi-Yoshida estimator: Statistical inference

- To construct CIs for a correlation parameter $\rho \in (-1, 1)$, it is often useful to consider **Fisher's z-transformation**
- Suppose that we have an estimator $\hat{\rho}_N$ for ρ such that $r_N^{-1}(\hat{\rho}_N - \rho) \to Z$ as $N \to \infty$, where Z is a centered normal variable with variance v and $r_N > 0$ satisfying $r_N \to 0$
- Let

$$F_N := \tanh^{-1}(\hat{\rho}_N) = \frac{1}{2} \log \left(\frac{1 - \hat{\rho}_N}{1 + \hat{\rho}_N} \right)$$

- The delta method implies that

$$r_N^{-1}(F_N - \tanh^{-1}(\rho)) \to (1 - \rho^2)^{-1} Z$$

as $N \to \infty$

Hayashi-Yoshida estimator: Statistical inference

- Thus we obtain the following $100(1-\alpha)\%$ CI for ρ

$$\left[\tanh\left(F_N - z_\alpha \frac{r_N\sqrt{v}}{1-\hat{\rho}_N^2}\right), \tanh\left(F_N + z_\alpha \frac{r_N\sqrt{v}}{1-\hat{\rho}_N^2}\right)\right]$$

- This CI is always contained in $(-1,1)$
- Application to the simulated data of `epps.r` is shown in the next slide

```
> v <- hyavar(psample)
> alpha <- 0.05 # significance level
> ## CI for covariance
> c(v$covmat[1,2]-qnorm(1-alpha/2)*sqrt(v$avar.cov[1,2]),
    v$covmat[1,2]+qnorm(1-alpha/2)*sqrt(v$avar.cov[1,2]))
[1] 0.3913176 0.5793336
> ## CI for correlation
> c(v$cormat[1,2]-qnorm(1-alpha/2)*sqrt(v$avar.cor[1,2]),
    v$cormat[1,2]+qnorm(1-alpha/2)*sqrt(v$avar.cor[1,2]))
[1] 0.4081517 0.5656669
> ## CI for correlation based on Fisher's z-transformation
> Fn <- atanh(v$cormat[1,2])
> c(tanh(Fn-qnorm(1-alpha/2)*sqrt(v$avar.cor[1,2])/(1 - v$cormat
    [1,2]^2)),
    tanh(Fn+qnorm(1-alpha/2)*sqrt(v$avar.cor[1,2])/(1 - v$cormat
    [1,2]^2)))
[1] 0.4042923 0.5616450
```

References I

T. W. Epps. Comovements in stock prices in the very short run. *J. Amer. Statist. Assoc.*, 74(366):291–298, 1979.

M. Fukasawa. Realized volatility with stochastic sampling. *Stochastic Process. Appl.*, 120:829–852, 2010.

T. Hayashi and S. Kusuoka. Consistent estimation of covariation under nonsynchronicity. *Stat. Inference Stoch. Process.*, 11:93–106, 2008.

T. Hayashi and N. Yoshida. On covariance estimation of non-synchronously observed diffusion processes. *Bernoulli*, 11(2):359–379, 2005.

T. Hayashi and N. Yoshida. Nonsynchronous covariation process and limit theorems. *Stochastic Process. Appl.*, 121:2416–2454, 2011.

T. Hayashi, J. Jacod, and N. Yoshida. Irregular sampling and central limit theorems for power variations: The continuous case. *Ann. Inst. Henri Poincaré Probab. Stat.*, 47(4):1197–1218, 2011.

J. Jacod and P. Protter. *Discretization of Processes*. Springer, 2012.

J. Jacod and A. N. Shiryaev. *Limit theorems for stochastic processes*. Springer, second edition, 2003.

References II

Y. Li, P. A. Mykland, E. Renault, L. Zhang, and X. Zheng. Realized volatility when sampling times are possibly endogenous. *Econometric Theory*, 30: 580–605, 2014.

M. Podolskij and M. Vetter. Understanding limit theorems for semimartingales: a short survey. *Stat. Neerl.*, 64(3):329–351, 2010.