Estimation of Integrated Covariances in the Simultaneous Presence of Nonsynchronicity, Noise and Jumps

Yuta Koike

Graduate School of Mathematical Sciences

University of Tokyo

JAPAN

September 3, 2013

Outline

- Introduction
- Model
- Construction of the estimator
- Main result
- Simulation
- Conclusions

Introduction

- The aim of this talk: Estimating integrated covariances separately from jumps using high-frequency financial data
  - Important for identifying the sources of risks (systematic or idiosyncratic, normal or non-normal, etc.)
- We try to deal with the following problems:
  - Nonsynchronous observation times
  - Microstructure noise
  - Time endogeneity

Model: Latent log-prices

- $\underline{Z^1, Z^2}$ Latent log-price processes of two assets

$$\mathrm{d}Z_t^k = \underbrace{a_t^k \mathrm{d}t + \sigma_{t-}^k \mathrm{d}W_t^k}_{\text{continuous part}} + \underbrace{c_{t-}^k \mathrm{d}L_t^k}_{\text{jump}}, \quad \mathrm{d}[W^1, W^2]_t = \eta_t \mathrm{d}t.$$

  - $\mathcal{B}^{(0)} = (\Omega^{(0)}, \mathcal{F}^{(0)}, (\mathcal{F}_t^{(0)}), P^{(0)})$: Stochastic basis
  - W^k : Standard Wiener process on $\mathcal{B}^{(0)}$
  - L^k: Pure jump Lévy process on $\mathcal{B}^{(0)}$
  - a^k, σ^k, c^k, η: Càdlàg $(\mathcal{F}_t^{(0)})$-adapted processes

- Objective Integrated covariance: $IC_t = \int_0^t \sigma_s^1 \sigma_s^2 \eta_s \mathrm{d}s$

Model: Observation times

- $\underline{\mathcal{I} = (S^i)_{i=0}^\infty, \mathcal{J} = (T^j)_{j=0}^\infty}$ Sequences of $(\mathcal{F}_t^{(0)})$-stopping times satisfying $S^i \uparrow \infty$, $T^i \uparrow \infty$ as $i \to \infty$.
- $n \in \mathbb{N}$: Parameter representing the observation frequency
- $\mathcal{I}$ and $\mathcal{J}$ depend on n and assume that

$$n^{1-\varepsilon}\left[\sup_{i:S^i \le t}(S^i - S^{i-1}) \vee \sup_{j:T^j \le t}(T^j - T^{j-1})\right] \to^p 0$$

as $n \to \infty$ for any $\varepsilon, t > 0$ ($S^{-1} = T^{-1} := 0$).

Model: Microstructure noise

- $\underline{\mathsf{Z}^1_{S^i}, \mathsf{Z}^2_{T^j}}$ Noisy observation data of Z^1 and Z^2 observed at each times in $\mathcal{I}$ and $\mathcal{J}$ respectively:

$$\mathsf{Z}^1_{S^i} = Z^1_{S^i} + U^1_{S^i}, \qquad \mathsf{Z}^2_{T^j} = Z^2_{T^j} + U^2_{T^j}.$$

- $\underline{(U^1_{S^i})_{i=0}^{\infty},\ (U^2_{T^j})_{j=0}^{\infty}}$ Centered independent random variables, conditionally on $\mathcal{F}^{(0)}$
  - $Q_t(\omega^{(0)}, \mathrm{d}u)$: Conditional law of (U^1_t, U^2_t) (a transition probability from $(\Omega^{(0)}, \mathcal{F}^{(0)}_t)$ into $\mathbb{R}^2$ with $\int uQ_t(\mathrm{d}u) = 0$)
  - An appropriate stochastic basis $(\Omega, \mathcal{F}, (\mathcal{F}_t), P)$ is constructed in the same way as Jacod *et al.* (2009)

Construction of the estimator

- The aim of this talk: Estimating IC_t from the observation data $(\mathsf{Z}^1_{S^i}, \mathsf{Z}^2_{T^j})_{i,j:S^i,T^j \leq t}$ as $n \to \infty$ for every $t \geq 0$.
- If both jumps and noise are absent, we can use the Hayashi-Yoshida estimator (Hayashi & Yoshida, 2005):

$$\sum_{i,j:S^i \vee T^j \leq t} (Z^1_{S^i} - Z^1_{S^{i-1}})(Z^2_{T^j} - Z^2_{T^{j-1}}) 1_{\{[S^{i-1}, S^i) \cap [T^{j-1}, T^j) \neq \emptyset\}}$$

- Our approach
  - Reconstructing the returns of the continuous parts from the observed returns
  - Constructing a Hayashi-Yoshida type estimator based on the reconstructed returns

Removing the noise: Pre-averaging

- Choose a positive integer k_n satisfying $k_n = \theta\sqrt{n} + o(n^{1/4})$ for some $\theta > 0$ (e.g., $k_n = \lceil \theta\sqrt{n} \rceil$)

- Choose a weight function g on $[0,1]$. Here $g(x) = x \wedge (1-x)$ is used for simplicity.

- Pre-averaging (in tick time) (cf. Podolskij & Vetter, 2009):

$$\overline{\mathsf{Z}}^1(\mathcal{I})^i = \sum_{p=1}^{k_n-1} g\left(\frac{p}{k_n}\right)(\mathsf{Z}^1_{S^{i+p}} - \mathsf{Z}^1_{S^{i+p-1}}).$$

- $(\overline{\mathsf{Z}}^1(\mathcal{I})^i)_i$ seem to be "noise-free" returns

Removing the noise: Pre-averaging

- Naive approach (in the absence of jumps) Summing the cross-products $\overline{Z}^1(\mathcal{I})^i \overline{Z}^2(\mathcal{J})^j 1_{\{[S^i, S^{i+k_n}) \cap [T^j, T^{j+k_n}) \neq \emptyset\}}$ over i, j and scaling appropriately (cf. Christensen *et al.*, 2010)

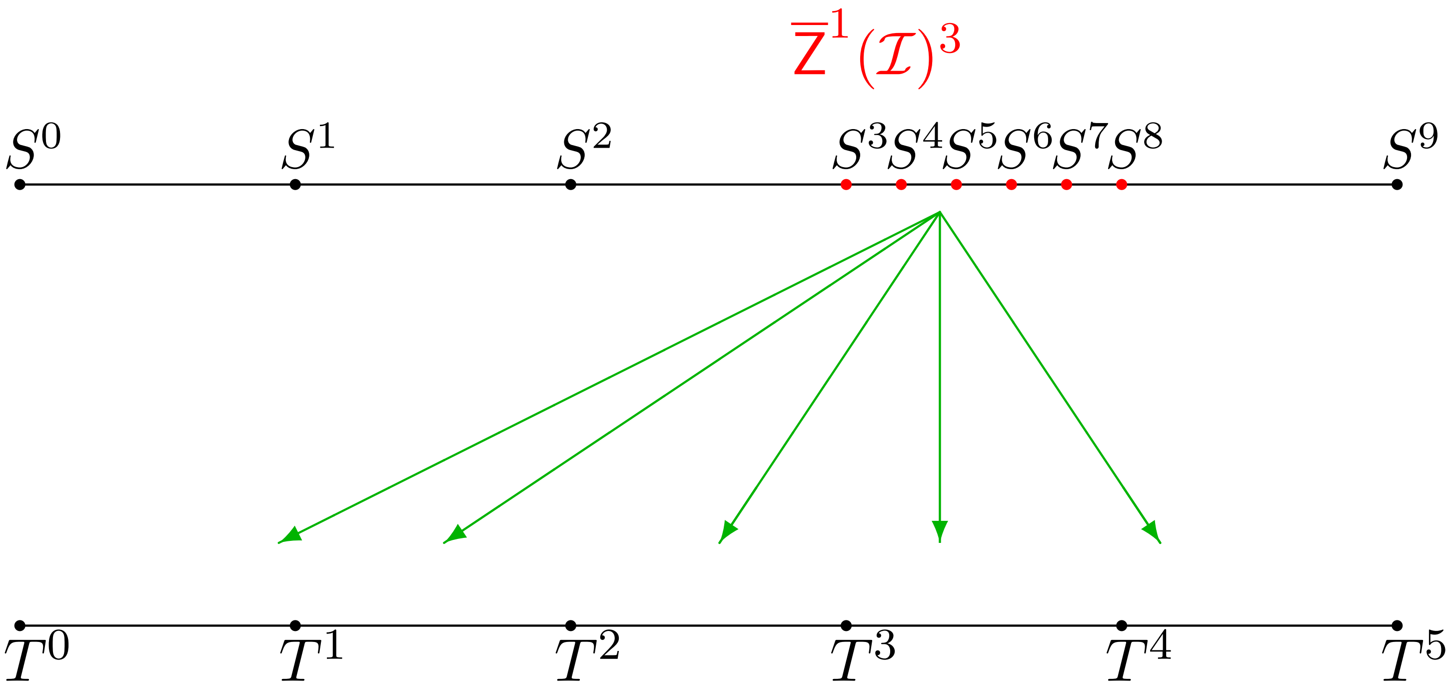
  $\Longrightarrow$ But this approach is possibly ineffective and intractable (due to using the common pre-averaging window k_n)

- For example, if $[S^i, S^{i+m}) \subset [T^j, T^{j+1})$ for some i, j, m,

$$\overline{Z}^1(\mathcal{I})^i \overline{Z}^2(\mathcal{J})^{j'} 1_{\{[S^i, S^{i+k_n}) \cap [T^{j'}, T^{j'+k_n}) \neq \emptyset\}} \qquad j' = 0, 1, \ldots$$

  will involve many non-overlap cross-products

  $\Longrightarrow$ We partially pre-synchronize the data



$\overline{\mathsf{Z}}^1(\mathcal{I})^3$
S^0
S^1
S^2
S^3
S^4
S^5
S^6
S^7
S^8
S^9
T^0
T^1
T^2
T^3
T^4
T^5

Pre-synchronization: Refresh time

- Refresh time sampling Define $(\widehat{S}^k, \widehat{T}^k, R^k)_{k=0}^{\infty}$ sequentially by $\widehat{S}^0 := S^0$, $\widehat{T}^0 := T^0$, $R^0 := S^0 \vee T^0$ and

$$\widehat{S}^k := \min\{S^i | S^i > R^{k-1}\}, \quad \widehat{T}^k := \min\{T^j | T^j > R^{k-1}\}$$

$$R^k := \widehat{S}^k \vee \widehat{T}^k$$

- Pre-averaging in refresh time

$$\overline{\mathsf{Z}}_i^1 := \overline{\mathsf{Z}}^1(\widehat{\mathcal{I}})^i = \sum_{p=1}^{k_n-1} g\left(\frac{p}{k_n}\right)(\mathsf{Z}^1_{\widehat{S}^{i+p}} - \mathsf{Z}^1_{\widehat{S}^{i+p-1}}),$$

$$\overline{\mathsf{Z}}_j^2 := \overline{\mathsf{Z}}^2(\widehat{\mathcal{J}})^j = \sum_{q=1}^{k_n-1} g\left(\frac{q}{k_n}\right)(\mathsf{Z}^2_{\widehat{T}^{j+q}} - \mathsf{Z}^2_{\widehat{T}^{j+q-1}})$$

Removing the jumps: Thresholding

- Large pre-averaging data will involve jumps
- Thresholding Remove the pre-averaging data exceeding predetermined threshold values (cf. Aït-Sahalia *et al.*, 2012; Podolskij & Ziggel, 2010):

$$\widetilde{\mathsf{Z}}_i^1 = \overline{\mathsf{Z}}_i^1 1_{\{|\overline{\mathsf{Z}}_i^1|^2 \leq \varrho_n^1(\widehat{S}^i)\}}, \quad \widetilde{\mathsf{Z}}_j^2 = \overline{\mathsf{Z}}_j^2 1_{\{|\overline{\mathsf{Z}}_j^2|^2 \leq \varrho_n^2(\widehat{T}^i)\}}$$

where $\varrho_n^k(t) = \alpha_n^k(t)\rho_n$ with

- $\alpha_n^k(t)$: Sequence of positive-valued stochastic processes
- ρ_n: Sequence of positive numbers

Construction of the estimator

□ Our estimator $PTHY_t^n$ is defined by

$$PTHY_t^n = \frac{1}{(\psi_{HY} k_n)^2} \sum_{i,j:\widehat{S}^{i+k_n} \vee \widehat{T}^{j+k_n} \leq t} \widetilde{\mathsf{Z}}_i^1 \widetilde{\mathsf{Z}}_j^2 \bar{K}^{ij},$$

- $\psi_{HY} = \int_0^1 g(x)\mathrm{d}x = \frac{1}{4}$ (Normalizing factor),
- $\bar{K}^{ij} = 1_{\{[\widehat{S}^i, \widehat{S}^{i+k_n}) \cap [\widehat{T}^j, \widehat{T}^{j+k_n}) \neq \emptyset\}}$ (Hayashi-Yoshida type kernel)

Main result: Notation

- $\mathbb{D}(\mathbb{R}_+)$: Skorohod space
- $\Upsilon_t(\omega^{(0)})$: Covariance matrix of $Q_t(\omega^{(0)}, \mathrm{d}u)$
- $\Gamma^k = [R^{k-1}, R^k)$ for each k
- $(\mathcal{H}^n_t)$: Filtration generated by W^k, a^k, σ^k, c^k $(k = 1, 2)$, η, Υ, $\sum_i 1_{\{S^i \leq \cdot\}}$ and $\sum_j 1_{\{T^j \leq \cdot\}}$
- For each $\rho > 0$, define the processes $G(\rho)^n$ and χ^n by

$$G(\rho)^n_s = E\left[\left(n|\Gamma^k|\right)^\rho \left|\mathcal{H}^n_{R^{k-1}}\right.\right], \quad \chi^n_s = P(\widehat{S}^k = \widehat{T}^k \big| \mathcal{H}^n_{R^{k-1}})$$

when $s \in \Gamma^k$ ($| \cdot |$ denotes the Lebesgue measure).

Main result: Conditions

- Condition on the duration (standard and necessary for computing the asymptotic variance explicitly)

[A1] (i) There exists a càdlàg $\mathbf{F}^{(0)}$-adapted process G such that G and G_- do not vanish and $G(1)^n \to^p G$ in $\mathbb{D}(\mathbb{R}_+)$ as $n \to \infty$.

(ii) There exists a càdlàg $\mathbf{F}^{(0)}$-adapted process χ such that $\chi^n \to^p \chi$ in $\mathbb{D}(\mathbb{R}_+)$ as $n \to \infty$.

(iii) $\sup_{0 \le s \le t} G(2)^n_s$ is tight as $n \to \infty$ for every t.

Main result: Conditions

- Condition to deal with the time endogeneity
  - General case: Strong predictability type condition (cf. Hayashi & Yoshida, 2011)

[A2] There exists a constant $\xi \in (0, \frac{1}{2})$ such that S^i and T^i are $(\mathcal{F}^{(0)}_{(t-n^{-\xi})_+})$-stopping times for every $n, i \in \mathbb{N}$.

  - Finite activity case: Continuity condition on the conditionally expected duration

[A2′] (i) S^i and T^i are $(\mathcal{F}^{(0)}_t)$-predictable times for every i
(ii) G and χ are Itô-type semimartingales
(iii) $n^\delta(G(1)^n - G) \to^p 0$ and $n^\delta(\chi^n - \chi) \to^p 0$ in $\mathbb{D}(\mathbb{R}_+)$ for some $\delta > 0$

Main result: Conditions

- Condition on the activity of the jump processes (standard); for each $\beta \in [0,2]$:

  [K$_\beta$] It holds that $\int_{\mathbb{R}} |x|^\beta \wedge 1 F^k(\mathrm{d}x) < \infty$ for $k = 1, 2$, where F^k is the Lévy measure of L^k.

- Set

$$\kappa = \frac{7585}{1161216}, \qquad \overline{\kappa} = \frac{151}{20160}, \qquad \widetilde{\kappa} = \frac{1}{24}$$

  (quantities appearing in the asymptotic variance and determined by the choice of g)

Theorem

Under standard regularity conditions on the coefficient processes, the moment of the noise process and the threshold processes,

(i) if [A1], [A2] and [K$_\beta$] are satisfied for some $\beta \in [0,1)$, then

$$n^{1/4}(PTHY^n - IC) \xrightarrow{d_s} \int_0^\cdot w_s \mathrm{d}\widetilde{W}_s \quad \text{in} \quad \mathbb{D}(\mathbb{R}_+) \tag{1}$$

as $n \to \infty$, where $\widetilde{W}$ is a standard Wiener process (defined on an extension of $\mathcal{B}$) independent of $\mathcal{F}$ and w is given by

$$w_s^2 = \psi_{HY}^{-4}\Bigg[\theta\kappa(\sigma_s^1\sigma_s^2)^2(1+\eta_s^2)G_s + \theta^{-3}\widetilde{\kappa}\left\{\Upsilon_s^{11}\Upsilon_s^{22} + \left(\Upsilon_s^{12}\chi_s\right)^2\right\}\frac{1}{G_s} + \theta^{-1}\overline{\kappa}\left\{(\sigma_s^1)^2\Upsilon_s^{22} + (\sigma_s^2)^2\Upsilon_s^{11} + 2\sigma_s^1\sigma_s^2\eta_s\Upsilon_s^{12}\chi_s\right\}\Bigg].$$

(ii) if [A1], [A2′] and [K$_0$] are satisfied, then (1) holds true.

Simulation

□ Latent $Z^k = X^k + J^k$, $k = 1, 2$

- X^1, X^2 (continuous part): Bivariate SV1F model (same parametrization as Barndorff-Nielsen *et al.* (2011))
- $R = 0.91$ (Correlation between X^1 and X^2)
- J^1, J^2 (Jump): 3 Scenarios
  1. $J^1 = J^2 = 0$ (No jump)
  2. $J^1 = J^2 = L^0$, where L^0 is a stratified NIG-CP process with a single jump per unit time (FA jump)
  3. $J^1 = L^1$, $J^2 = RL^1 + \sqrt{1 - R^2}L^2$, where L^1 and L^2 are mutually independent VG processes (IA jump)

  (L^l has the same parametrization as Veraart (2010), especially the QV of the jumps is the 10% of the IV in their mean values)

- Sampling (S^i) (resp. (T^j)): Poisson arrival times with intensity n/λ^1 (resp. n/λ^2) in $[0, 1]$
    - $n = 23, 400$, $(\lambda^1, \lambda^2) = (3, 6), (10, 20), (30, 60)$
- Noise (cf. Barndorff-Nielsen *et al.*, 2011)

$$U_t^k \sim N\left(0, 0.001\sqrt{\frac{1}{n}\sum_{i=1}^{n}\sigma_{i/n}^4}\right), \quad \mathrm{Corr}(U_t^1, U_t^2) = R$$

- Tuning parameter (cf. Christensen *et al.*, 2013)
    - $k_n = \lceil 0.15\sqrt{N}\rceil$ (N: Number of the refresh times − 1)
    - Threshold (ad-hoc, but easy and fairly effective)

$$\varrho_n^1(\widehat{S}^i) = 2\log(N)^{1.2}\frac{\pi/2}{K - 2k_n + 1}\sum_{p=i-K}^{i-2k_n}|\overline{\mathsf{Z}}^1(\widehat{\mathcal{I}})^p||\overline{\mathsf{Z}}^1(\widehat{\mathcal{I}})^{p+k_n}|$$

with $K = \lceil N^{3/4}\rceil$ (put $\varrho_n^1(\widehat{S}^i) = \varrho_n^1(\widehat{S}^K)$ for $i < K$)

Scenario	1	2	3
PTHY			
$\lambda = (3, 6)$	$-$.002 (.104)	.003 (.104)	.005 (.103)
$\lambda = (10, 20)$	$-$.011 (.135)	$-$.006 (.137)	$-$.003 (.137)
$\lambda = (30, 60)$	$-$.037 (.200)	$-$.033 (.203)	$-$.030 (.200)
BPV			
$\lambda = (3, 6)$	$-$.020 (.136)	.012 (.148)	.017 (.150)
$\lambda = (10, 20)$	$-$.058 (.165)	$-$.028 (.166)	$-$.022 (.165)
$\lambda = (30, 60)$	$-$.142 (.246)	$-$.116 (.236)	$-$.111 (.235)

Note. The bias and rmse (in parenthesis) are reported. Number of repetition= $1,000$. Upper panel: Our estimator, Lower panel: Subsampled bipower covariation based on 5-min returns (benchmark)

Conclusions

- Construction of the estimator
  - Pre-averaging $\Longrightarrow$ Removing the noise
  - Refresh time sampling + Hayashi-Yoshida method $\Longrightarrow$ Handling the nosynchronicity
  - Thresholding $\Longrightarrow$ Separating the jumps
- Asymptotic mixed normality was shown in the cases
  - Finite variation jumps + Strong Predictability
  - Finite activity jumps + Continuity of conditionally expected duration
- More information arXiv: 1302.5202, 1305.1229

References

Aït-Sahalia, Y., Jacod, J. & Li, J. (2012). Testing for jumps in noisy high frequency data. *J. Econometrics* **168**, 207–222.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A. & Shephard, N. (2011). Multivariate realised kernels: Consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading. *J. Econometrics* **162**, 149–169.

Bibinger, M. (2012). An estimator for the quadratic covariation of asynchronously observed Itô processes with noise: Asymptotic distribution theory. *Stochastic Process. Appl.* **122**, 2411–2453.

Christensen, K., Kinnebrock, S. & Podolskij, M. (2010). Pre-averaging estimators of the ex-post covariance matrix in noisy diffusion models with non-synchronous data. *J. Econometrics* **159**, 116–133.

Christensen, K., Podolskij, M. & Vetter, M. (2013). On covariation estimation for multivariate continuous Itô semimartingales with noise in non-synchronous observation schemes. *J. Multivariate Anal.* **120**, 59–84.

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

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

Jacod, J., Li, Y., Mykland, P. A., Podolskij, M. & Vetter, M. (2009). Microstructure noise in the continuous case: The pre-averaging approach. *Stochastic Process. Appl.* **119**, 2249–2276.

Podolskij, M. & Vetter, M. (2009). Estimation of volatility functionals in the simultaneous presence of microstructure noise and jumps. *Bernoulli* **15**, 634–658.

Podolskij, M. & Ziggel, D. (2010). New tests for jumps in semimartingale models. *Stat. Inference Stoch. Process* **13**, 15–41.

Veraart, A. E. (2010). Inference for the jump part of quadratic variation of Itô semimartingales. *Econometric Theory* **26**, 331–368.

Appendix A: Regularity conditions

- Condition on the coefficient processes (standard)

[A3] a^k, σ^k, c^k $(k = 1, 2)$, η and Υ^{ij} $(i, j = 1, 2)$ are Itô-type semimartingales

- Condition on the noise processes (standard)

[A4] $(\int |u|^r Q_t(\cdot, \mathrm{d}u))_{t \geq 0}$ is a locally bounded process for every $r > 0$

Appendix A: Regularity conditions

- Condition on the threshold processes (standard)

[T] (i) We have

$$\rho_n \to 0 \qquad \text{and} \qquad \frac{n^{-\frac{1}{2}+\gamma}\log n}{\rho_n} \to 0$$

as $n \to \infty$ for some $\gamma \in (0, \frac{1}{2})$.

(ii) For each $k \in \{1, 2\}$, there exists a sequence of stopping times $(\tau_m^k)_{m\in\mathbb{N}}$ such that $\tau_m^k \uparrow \infty$ and both $\sup_{0\leq t<\tau_m^k} \alpha_n^k(t)$ and $\sup_{0\leq t<\tau_m^k}[1/\alpha_n^k(t)]$ are tight as $n \to \infty$ for all m.

Appendix A: Precise statement of the main theorem

Theorem

(i) Suppose [A1]–[A4] and [K$_\beta$] hold for some $\beta \in [0,1)$. Suppose also [T] holds with $\rho_n = O(n^{-1/2(2-\beta)})$. Then (1) holds true.
(ii) Suppose [A1], [A2′], [A3]–[A4′], [K$_0$] and [T] hold. Then (1) holds true.