Estimation of the volatility and/or the degree of jump activity, in presence of noise and irregular sampling

Jean Jacod – Viktor Todorov

UPMC (Paris 6) – Northwestern University

The aim

Estimate the **integrated volatility** $C_t = \int_0^t \sigma_s^2\, ds$ of the 1-dimensional process

$$X_t = X_0 + \int_0^t b_s ds + \int_0^t \sigma_s dW_s + \text{jumps}$$

observed at discrete times within the *fixed* time interval $[0,t]$, with a mesh going to 0 (*high frequency setting*) and also **some characteristics of the jumps.**

Three features:

- X has jumps, possibly with *high-activity* (= BG index ≥ 1)
- the sampling times $0 = T(n,0) < T(n,1) < \cdots < T(n,i) < \cdots$ may be irregular, possibly random (*regular sampling* $\Leftrightarrow$ $T(n,i+1) - T(n,i) = \Delta_n$)
- there is a microstructure noise: instead of $X_{T(n,i)}$ we observe

$$Y_i^n = X_{T(n,i)} + \chi_i^n.$$

Notation

$$\Delta(n,i) = T(n,i) - T(n,i-1)$$

$$\Delta_i^n V = V_{T(n,i)} - V_{T(n,i-1)} \qquad V: \text{ any process}$$

Regular sampling means $\Delta(n,i) = \Delta_n$.

Spot Lévy measures of X: the compensator ν of the jump measure of X is assumed to have the factorization

$$\nu(\omega, dt, dx) = dt\, F_{\omega,t}(dx)$$

(this is the "Itô semimartingale property" for the jumps). The measures $F_t = F_{\omega,t}$ are the spot Lévy measure, and $\int (x^2 \wedge 1)\, F_t(dx) < \infty$.

Problems for estimating $C_t = \int_0^t \sigma_s^2 \, ds$ when jumps are present

(no-noise and regular sampling cases)

- *X continuous:* the "optimal" estimator for C_t (in the sense of LAN or LAMN properties)

$$\widehat{C}_t^n = \sum_{i=1}^{[t/\Delta_n]} (\Delta_i^n X)^2$$

the rate is $\frac{1}{\sqrt{\Delta_n}}$, the (conditional) asymptotic variance is $2 \int_0^t \sigma_s^4 \, ds$.

- *X discontinuous:* $\widehat{C}_t^n$ no longer consistent, and one has 2 main methods:

$$\textit{truncation:} \qquad \widehat{C}_t^{\prime n} = \sum_{i=1}^{[t/\Delta_n]} (\Delta_i^n X)^2 \, 1_{\{|\Delta_i^n X| \le u_n\}}$$

$$\textit{multipowers:} \qquad \widehat{C}_t^{\prime\prime n} = \alpha_p \sum_{i=1}^{[t/\Delta_n]-k+1} \prod_{j=0}^{k-1} |\Delta_{i+j}^n X|^{2/k}$$

Both these are consistent, and have a CLT with rate $\frac{1}{\sqrt{\Delta_n}}$ and the optimal variance for truncations and a (slightly) bigger variance for multipowers, under mild assumptions on b_t, σ_t (stronger for multipowers, though, for example $1/\sigma_t^2$ should be locally bounded), PLUS the (strong) additional assumption

$$\int (|x|^r \wedge 1)\, F_t(dx) \quad \text{is locally bounded} \tag{1}$$

for some $r < 1$ ("almost" equivalent to having $\sum_{s \le t} |\Delta X_s|^r < \infty$ for all $t < \infty$).

More: if a sequence S_n is such that $v_n(S_n - C_t)$ is tight for some sequence $v_n \to \infty$, uniformly for *all* X for which b_t, σ_t and (1) are *uniformly bounded*, then the minimax rate v_n satisfies

$$r \le 1 \quad \Rightarrow \quad v_n \preceq \frac{1}{\sqrt{\Delta_n}}, \qquad r > 1 \quad \Rightarrow \quad v_n \preceq \Big(\frac{\log(1/\Delta_n)}{\Delta_n}\Big)^{(2-r)/2}$$

However: when $r > 1$ and the jumps of X are "close enough" to those of a stable process, it becomes possible to get the $\frac{1}{\sqrt{\Delta_n}}$ rate (no contradiction: we have switched from a non-parametric situation to a semi-parametric one).

Assumptions on X (typically a log-price)

We have a filtered space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ on which:

$$X_t = X_0 + \int_0^t b_s\,ds + \int_0^t \sigma_s\,dW_s + \int_0^t \int_E \delta(s,z)(\underline{p} - \underline{q})(ds,dz) + \int_0^t \int_E \delta'(s,z)\underline{p}(ds,dz)$$

$$\sigma_t = \sigma_0 + \int_0^t b_s^\sigma\,ds + \int_0^t H_s^\sigma\,dW_s' + \int_0^t \int_E \delta^\sigma(s,z)(\underline{p} - \underline{q})(ds,dz) + \int_0^t \int_E \delta'^\sigma(s,z)\underline{p}(ds,dz)$$

- W and W' are two correlated Brownian motions.
- $\underline{p}$ is a Poisson measure on $\mathbb{R}_+ \times E$ with (deterministic) compensator $\underline{q}(dt,dz) = dt \otimes \eta(dz)$ (η is a σ-finite measure on the Polish space E).
- "standard" assumptions on the coefficients $b_t, b_t^\sigma, H_t^\sigma$ are locally bounded, and (up to localization). b_t, H_t^σ satisfy for all finite stopping times $T \leq S$:

$$\mathbb{E}(\sup_{s\in[T,S]} |V_s - V_T|^2) \leq K\mathbb{E}(S-T) \tag{2}$$

- $|\delta(t,z)|^{r'}$, $1_{\{\delta'(t,z)\neq 0\}} \leq J(z)$ for some non-random η-integrable function J and some $r' \in [0,2)$, and the same for δ^σ, δ'^σ (up to localization).

MOREOVER, we need a structural assumption on the high-activity jumps of X, expressed in terms of the BG (Blumenthal-Getoor) index, or successive BG indices:

There is an integer $M \geq 0$, numbers $2 > \beta_1 > \cdots \beta_M > 0$, and nonnegative càdlàg processes $a_t^1, \ldots, a_t^M$, such that each $(a_t^m)^{1/\beta_m}$ satisfies (2), and the symmetrized Lévy measures $\breve{F}_t(A) = F_t(A) + F_t(-A)$ are such that the (signed) measure

$$F'_t(dx) = \breve{F}_t(dx) - \sum_{m=1}^{M} \frac{\beta_m a_t^m}{|x|^{1+\beta_m}} 1_{\{0<|x|\leq 1\}}\, dx$$

satisfies ($|F'_t|$ being the "absolute value" of F'_t):

$$|F'_t|([-x,x]^c) \leq \frac{\Gamma}{x^r} \qquad \forall x >\in (0,1].$$

Example:

$$X_t = X_0 + \int_0^t b_s\, ds + \int_0^t \sigma_s\, dW_s + \sum_{m=1}^{M} \int_0^t \gamma_{s-}^m\, dY_s^m + \int_0^t \int_E \delta'(s,z) \underline{p}(ds,dz)$$

with Y^m independent stable or tempered stable processes (with arbitrary dependencies with $\underline{p}$) and γ^m's are Itô semimartingales.

We then have $a_t^m = |\gamma_t^m|^{\beta_m}$.

REGULAR OBSERVATIONS – NON NOISE

Choose a sequence k_n of integer ($\to \infty$) and set

$$L(y)_j^n = \frac{1}{k_n} \sum_{l=0}^{k_n-1} \cos(y(\Delta_{i+1+2l}^n X - \Delta_{i+2+2l}^n X)/\sqrt{\Delta_n}\,)$$

We take differences of two successive returns to “symmetrize” the jump measure around zero and kill the drift.

We have approximately, with $\chi(\beta) = \int_0^\infty \frac{\sin y}{y^\beta}\, dy$:

$$\mathbb{E}(L(y)_i^n \mid \mathcal{F}_i) \approx \exp\left(-y^2 \sigma_{i\Delta_n}^2 - 2 \sum_{m=1}^{M} \chi(\beta_m) y^{\beta_m} \Delta_n^{1-\beta_m/2} a_{i\Delta_n}^m\right),$$

hence the following is a estimator of the spot (squared) volatility:

$$\widehat{c}(y)_j^n = -\frac{1}{y^2}\, \log\left(L(y)_j^n \bigvee \frac{1}{\log(1/\Delta_n)}\right)$$

First estimators of integrated volatility: Introducing a de-biasing term, we set

$$\widehat{C}(y)_t^n = 2k_n\Delta_n \sum_{j=0}^{[t/v_n]-1} \Big(\widehat{c}(y)_j^n - \frac{1}{y^2 k_n}\left(\sinh(y^2\widehat{c}(y)_j^n)\right)^2\Big),$$

where $\sinh(x) = \frac{1}{2}\left(e^x - e^{-x}\right)$ is the hyperbolic sine.

This "estimates" $C_t + \sum_{m=1}^M A^{m,n}(y)_t$, where

$$A^{m,n}(y)_t = y^{\beta_m - 2}\Delta_n^{1-\beta_m/2} A_t^m, \qquad A_t^m = 2\chi(\beta_m)\int_0^t a_s^m\, ds$$

and the normalized error term is

$$Z(y)_t^n = \frac{1}{\sqrt{\Delta_n}}\Big(\widehat{C}(y)_t^n - C_t - \sum_{m=1}^M A^{m,n}(y)_t\Big).$$

The CLT: Let $\mathcal{Y}$ be a finite subset of $(0,\infty)$. Choose k_n and u_n such that

$$k_n\sqrt{\Delta_n} \to 0, \qquad k_n\Delta_n^{1/2-\varepsilon} \to \infty \ \ \forall\varepsilon>0, \qquad u_n \to 0, \qquad \frac{k_n\sqrt{\Delta_n}}{u_n^2} \to 0.$$

Then, we have the (functional) stable convergence in law:

$$\Big(Z(u_n)^n, \, \big(\frac{1}{u_n^2}(Z(yu_n)^n - Z(u_n)^n)\big)_{y\in\mathcal{Y}}\Big) \stackrel{\mathcal{L}-s}{\Longrightarrow} \big(Z, ((y^2-1)\overline{Z})_{y\in\mathcal{Y}}\big),$$

where the limit is defined on an extension $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, (\widetilde{\mathcal{F}}_t)_{t\geq 0}, \widetilde{\mathbb{P}})$ of the original space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ and can be written as

$$Z_t = 2\int_0^t c_s \, dW_s^{(1)}, \qquad \overline{Z}_t = \frac{2}{\sqrt{3}}\int_0^t c_s^2 \, dW_s^{(2)}.$$

where $W^{(1)}$ and $W^{(2)}$ are two independent Brownian motions, independent of the σ-field $\mathcal{F}$.

Estimation of C_t when $\beta_1 < 1$: Recall $A^{m,n}(u_n y)_t \sim (yu_n)^{\beta_m - 2}\Delta_n^{1-\beta_m/2} \int_0^t a_s^m\, ds$, negligible in front of $\sqrt{\Delta_n}$ when $\beta_m < 1$. The previous CLT yields

$$\frac{1}{\sqrt{\Delta_n}}\left(\widehat{C}(u_n)_t^n - C_t\right) \stackrel{\mathcal{L}-\mathrm{s}}{\Longrightarrow} 2\int_0^t c_s\, dW_s^{(1)}$$

So $\widehat{C}(u_n)_t^n$ is an alternative to the realized volatility (or empirical quadratic variation) if there are no jumps, and to the truncated realized volatility or to multipower variations when there are jumps with activity < 1, under exactly the same assumptions on jumps.

Our estimator gives a variance equal to twice the “efficient” variance. This is due to the differencing of increments. If we do not do that we get the optimal variance (and, differencing is not needed in the present case).

Estimation of C_t when $M = 1$: Observe that

$$\overline{C}(y)^n_t = \widehat{C}(u_n y)^n_t - \widehat{C}(u_n)_t = A^{m,n}(u_n y) - A^{m,n}(u_n)_t + \sqrt{\Delta_n}\,(Z(yu_n)^n_t - Z(u_n)_t))$$

Thus we can de-bias $\widehat{C}(y)^n_t$ as follows: take any $\zeta > 1$ and put

$$\widehat{C}'(\zeta)^n_t = \widehat{C}(u_n)^n_t - \frac{(\widehat{C}(\zeta u_n)^n_t - \widehat{C}(u_n)^n_t)^2}{\widehat{C}(\zeta^2 u_n)^n_t - 2\widehat{C}(\zeta u_n)^n_t + \widehat{C}(u_n)^n_t}, \tag{3}$$

with the convention that the ratio above is 0 when its denominator vanishes. Then

Theorem: *When $M = 1$, for any t we have the stable convergence in law*

$$\frac{1}{\sqrt{\Delta_n}}\,(\widehat{C}'(\zeta)^n_t - C_t) \;\stackrel{\mathcal{L}-s}{\longrightarrow}\; 2\int_0^t c_s\,dW^{(1)}_s$$

When $M \geq 2$, one can "iterate" (3), and there are are stopping rules for this sequence of iterations.

Estimation of β_1: For simplicity suppose $M = 1$. Recall

$$\begin{aligned}\overline{C}(y)^n_t &= \widehat{C}(u_n y)^n_t - \widehat{C}(u_n)_t \\ &= u_n^{\beta_1 - 2}(y^{\beta_1 - 2} - 1)\Delta_n^{1-\beta_1/2} A^1_t + \sqrt{\Delta_n}\,(Z(yu_n)^n_t - Z(u_n)_t))\end{aligned}$$

The function $f(x) = \frac{4^x - 1}{2^x - 1}$ is C^∞ on $(1, 2)$, with a C^∞ reciprocal function f^{-1}. Then a natural estimator for β_1 is, for example,

$$\widehat{\beta}^n_t = f^{-1}\Big(\frac{\overline{C}(4)^n_t}{\overline{C}(2)^n_t}\Big). \tag{4}$$

and we have

$$\frac{u_n^{\beta_1 - 2}}{\Delta_n^{(\beta_1 - 1)/2}}\,\big(\widehat{\beta}^n_t - \beta_1\big) \;\stackrel{\mathcal{L}-\mathrm{s}}{\longrightarrow}\; Y$$

for some mixed centered normal y with a known (conditional) variance. The rate is thus almost $1/\Delta_n^{(\beta_1 - 1)/2}$, better than rates obtained by other methods when $\beta_1 > 4/3$.

The choice $k_n \approx 1/\sqrt{\Delta_n}$ and u_n going to 0 very slowly is in fact not “optimal” for this problem, so better (but complicated...) rates can actually be achieved.

NOISE AND IRREGULAR SAMPLING

Assumptions on the observation scheme

Assumption: The inter-observations lags are

$$\Delta(n,i+1) = \Delta_n \lambda_{T(n,i)} \Phi^n_{i+1}$$

- λ_t positive càdlàg adapted, satisfying

$$\mathbb{E}(|\lambda_T - \lambda_S|) \leq \mathbb{E}(T-S)$$

for all stopping times $S \leq T$ and $1/K \leq \lambda_t \leq K$ (up to localization).

- The variables Φ^n_i are positive, $\mathbb{E}(\Phi^n_i) = 1$, and $\sup_{n,i} \mathbb{E}(\Phi^n_i)^p) < \infty$ for all $p > 0$.
- The variables $(\Phi^n_i : i \geq 1)$ are mutually independent and independent of $\mathcal{F}_\infty$

  This implies in particular that $N^n_t = \sum_{i \geq 1} 1_{\{T(n,i) \leq t\}}$ satisfies

$$\Delta_n N^n_t \overset{\text{u.c.p.}}{\Longrightarrow} \Lambda_t := \int_0^t \frac{1}{\lambda_s}\, ds.$$

We denote by $\mathcal{H}^n_\infty$ the σ-field generated by $\mathcal{F}_\infty$ and all $\Phi^n_i : i \geq 1)$.

Two assumptions on the noise

We observe $Y_i^n = X_{T(n,i)} + \gamma'_{T(n,i)} \varepsilon_i^n$, where

(N1): The process γ' is an Itô semimartingale, and the variables $(\varepsilon_i^n : i \geq 1)$ are i.i.d., independent of $\mathcal{H}_\infty^n$, with

$$\mathbb{E}(\varepsilon_i^n) = 0, \qquad \mathbb{E}((\varepsilon_i^n)^2) = 1, \qquad \mathbb{E}(|\varepsilon_i^n|^p) < \infty.$$

(N2): The variables $(\varepsilon_i^n : i \geq 1)$ are independent, conditionally on $\mathcal{H}_\infty^n$ with

$$\mathbb{E}((\varepsilon_i^n)^p \mid \mathcal{H}_\infty^n) = \gamma_{T(n,i)}^{(p)}, \qquad \mathbb{P}(\varepsilon_i^n \in B \mid \mathcal{H}_\infty^n) = \mathbb{P}(\varepsilon_i^n \in B \mid \mathcal{H}_{T(n,i)}^n).$$

(where $(\mathcal{H}_t^n)$ is the smallest filtration containing $(\mathcal{F}_t)$ and for which all $T(n,i)$ are stopping times. Moreover, $\gamma_t^{(1)} = 0$ and $\gamma_t^{(2)} = 1$; moreover γ'_t and all $\gamma_t^{(p)}$ satisfy (2) and are $(\mathcal{F}_t)$-adapted.

(N1) is much stronger than (N2), and not very realistic. Later on, $\gamma_t = (\gamma'_t)^2$ (this is the "variance" of the noise).

An important example satisfying (N2) but not (N1).

Let $\rho_t^j \geq 0$ be càdlàg adapted nonnegative with $\sum_{j\in\mathbb{Z}} \rho_t^j = 1$ and $\rho_t^j = \rho_t^{-j}$ and $\sup_t \sum_{j\in\mathbb{Z}} \rho_t^j |j|^p < \infty$. For each n let $(Z_i^n : i \geq 1)$ be i.i.d. conditionally on $\mathcal{H}_\infty^n$, with density $x \mapsto \sum_{j\in\mathbb{Z}} \rho_{T(n,j)}^j 1_{[j,j+1)}(x)$. The observation at time $T(n,i)$ is

$$Y_i^n = [X_{T(n,i)} + Z_i^n],$$

so we have a additive (modulated) white noise *plus rounding.*

Remark: If we have "pure rounding", i.e. if we observe $[X_{T(n,i)}]$ (or $[X_{T(n,i)}] + \frac{1}{2}$ to "center" the noise), then no consistent estimator for C_t exists.

Pre-averaging

We choose 3 tuning parameters u_n, h_n, k_n all going to ∞: here $u_n > 0$ will be the argument of the empirical characteristic function below, and h_n, k_n (two integers) are window sizes.

The de-noising method is pre-averaging, but other methods could probably be used as well. Take a weight (or, kernel) function g on $\mathbb{R}$ with

$$\begin{array}{l} g \text{ is continuous, piecewise } C^1 \text{ with a piecewise Lipschitz derivative } g', \\ s \notin (0,1) \;\Rightarrow\; g(s) = 0, \qquad \int_0^1 g(s)^2 ds \;>\; 0, \end{array}$$

for example $g(x) = (x \wedge (1-x))\, 1_{[0,1]}(x)$. With a sequence $h_n \to \infty$ of integers, set

$$\begin{array}{lll} g_i^n = g(i/h_n), & \overline{g}_i^n \;=\; g_{i+1}^n - g_i^n & \\ \phi_n = \frac{1}{h_n} \sum_{i\in\mathbb{Z}} (g_i^n)^2, & \overline{\phi}_n = h_n \sum_{i\in\mathbb{Z}} (\overline{g}_i^n)^2, & \widetilde{\phi}_n^{(\beta)} = \frac{1}{h_n} \sum_{i\in\mathbb{Z}} |g_i^n|^\beta, \\ \phi = \int g(u)^2\, du, & \overline{\phi} = \int g'(u)^2\, du, & \widetilde{\phi}^{(\beta)} = \int |g(u)|^\beta\, du \end{array}$$

The *pre-averaged returns* of the observed values Y_i^n are

$$\widetilde{Y}_i^n = \sum_{j=1}^{h_n-1} g_j^n\, (Y_{i+j}^n - Y_{i+j-1}^n) = -\sum_{j=0}^{h_n-1} \overline{g}_j^n\, Y_{i+j}^n.$$

Initial estimators

For any $y > 0$, set

$$L(y)_i^n = \frac{1}{k_n} \sum_{l=0}^{k_n-1} \cos\left(y u_n (\widetilde{Y}^n_{i+2lh_n} - \widetilde{Y}^n_{i+(2l+1)h_n})\right)$$

(a proxy for the real part of the empirical characteristic function of the returns, over a window of $2h_n k_n$ successive observations). Taking a difference above allows us to "symmetrize" the problem.

Then, for any $y > 0$, a natural estimator for the "integrated volatility" over the time interval $[T(n,i), T(n, i+2h_n k_n)]$ is

$$\frac{2k_n}{y^2 u_n^2 \phi_n} v(y)_i^n, \qquad v(y)_i^n = -\log\left(L(y)_i^n \bigvee \frac{1}{h_n}\right).$$

We need to de-bias these estimators, to account for the noise, and also for some intrinsic distortion present even when there is no noise. With $f(x,y) = \frac{1}{2}\big(e^{2x-y} + e^{2x} - 2\big)$, set

$$\widehat{C}(y)_t^n = \tfrac{k_n}{y^2 u_n^2 \phi_n} \sum_{j=0}^{[N_t^n/2h_n k_n]-1} \Big(2v(y)^n_{2jh_nk_n} - \tfrac{1}{k_n}\, f(v(y)^n_{2jh_nk_n}, v(2y)^n_{2jh_nk_n}) \\ -\overline{\phi}_n y^2 u_n^2 \, \textstyle\sum_{l=1} (\Delta^n_{2jh_nk_n+l})^2\Big).$$

Finally, we set

$$Z(y)_t^n = \widehat{C}(y)_t^n - C_t - \frac{2}{\phi_n} \sum_{m=1}^{M} |y|^{\beta_m - 2} u_n^{\beta_m - 2} \widetilde{\phi}_n^{\beta_m} A_t^m$$

The basic CLT

Below, $\mathcal{Y}$ is any finite subset of $(0,\infty)$.

Theorem *under appropriate conditions on u_n, h_n, k_n, plus*

$$\frac{u_n^{\beta_1} h_n^3 \Delta_n}{u_n^{\beta_1} h_n^3 \Delta_n + u_n^4 (1+h_n^2\Delta_n)^2} \to \eta, \qquad \frac{h^2 \Delta_n}{1+h_n^2\Delta_n} \to \eta', \qquad u_n^{8-\beta_1}\Big(\frac{1}{h_n^5\Delta_n} + (h_n\Delta_n)^3\Big) \to 0$$

and with

$$v_n = \sqrt{\frac{h_n^3\Delta_n}{(1+h_n^2\Delta_n)^2 + u_n^{\beta_1-4} h_n^3 \Delta_n}}, \qquad v'_n = \frac{1}{u_n^{(\beta_1-4)/2}}$$

for any $t>0$ the variables $\big(v_n Z(1)_t^n, \big(v'_n (Z(y)_t^n - Z(1)_t^n)\big)_{y\in\mathcal{Y}}\big)$ converge $\mathcal{F}_\infty$-stably in law to $\big(Z(1)_t, (Z(y)'_t)_{y\in\mathcal{Y}}\big)$, which is defined on an extension $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{\mathbb{P}})$ and is, conditionally on $\mathcal{F}$, centered Gaussian with variance-covariance given by

$$\widetilde{\mathbb{E}}(Z(1)_t^2 \mid \mathcal{F}) = \tfrac{1}{\phi} \int_0^t \big(\eta \overline{\psi}_{\beta_1}(1,1) a_s^1 \lambda_s + (1-\eta)\big(\eta' \phi \sigma_s^2 \lambda_s + (1-\eta')\overline{\phi}\gamma_s\big)^2 \tfrac{1}{\lambda_s}\, ds$$
$$\widetilde{\mathbb{E}}(Z(1)_t Z'(y)_t \mid \mathcal{F}) = 0$$
$$\widetilde{\mathbb{E}}(Z'(y)_t\, Z'(y')_t \mid \mathcal{F}) = \frac{\overline{\psi}_{\beta_1}(y,y')}{\phi}\, A_t^1.$$

Estimation of C_t

For simplicity, assume $M = 1$. When $\beta_1 < 1$ we have

$$v_n(\widehat{C}(1)^n_t - C_t) \stackrel{\mathcal{L}-s}{\longrightarrow} Z(1)_t.$$

When $\beta_1 \geq 1$ we have to rely on the same trick as previous, and use the estimators (for some given $\zeta > 1$):

$$\widehat{C}'(\zeta)^n_t = \widehat{C}(u_n)^n_t - \frac{(\widehat{C}(\zeta u_n)^n_t - \widehat{C}(u_n)^n_t)^2}{\widehat{C}(\zeta^2 u_n)^n_t - 2\widehat{C}(\zeta u_n)^n_t + \widehat{C}(u_n)^n_t}.$$

We then get a CLT for these estimators, with the rate v_n above and some limit which is centered Gaussian with an explicit variance, conditionally on $\mathcal{F}$.

Now, under which appropriate conditions does this hold, with v_n as “large” as possible ?

The best rate v_n is always of the form $v_n \asymp \sqrt{h_n}$ (that is, $\eta = 0$ in the previous CLT), and it depends on the assumptions on noise!

Under (N1): (basically, a modulated additive white noise). We can take

$$h_n \asymp \frac{1}{\sqrt{\Delta_n}}, \qquad \begin{cases} u_n \asymp \Delta_n^{-\frac{3(1-\varepsilon)}{16-2\beta_1}}, & k_n \asymp \Delta_n^{-\frac{1-2\varepsilon}{6}} \quad \text{if } \beta_1 \geq \frac{16}{11} \\ u_n \asymp \Delta_n^{-\frac{2(1-\varepsilon)}{16-5\beta_1}}, & k_n \asymp \Delta_n^{-\frac{\beta_1(1-2\varepsilon)}{16-5\beta_1}} \quad \text{if } \beta_1 \leq \frac{16}{11} \end{cases}$$

(unfortunately, the unknown β_1 appears here). Then the rate of convergence is $v_n = \sqrt{h_n} \asymp 1/\Delta_n^{1/4}$, which is known to be the optimal rate in the presence of noise.

Under (N2): (basically "rounding" noise). In this case the rate is slower (although only very slightly so), and we just report the rate v_n below:

$$v_n \asymp \begin{cases} 1/\Delta_n^{\frac{68-13\beta_1}{232-28\beta_1}} & \text{if } \beta_1 \geq \frac{16}{11} \\ 1/\Delta_n^{\frac{32-7\beta_1}{128-34\beta_1}} & \text{if } \beta_1 \leq \frac{16}{11} \end{cases}$$

For regular sampling, even under (N2) and when $\beta_1 < 1$, the truncated realized variance applied with the preaveraged return (with $h_n \asymp 1/\sqrt{\Delta_n}$) enjoys a CLT with rate $1/\Delta_n^{1/4}$. So, perhaps, a modification of the present method could give us the efficient rate $1/\Delta_n^{1/4}$ also when $\beta_1 \geq 1$?

Also, when $M \geq 2$, iterating the formula giving the estimators enough times would give us the same asymptotic results (with, probably, a lot of numerical instability).

Estimation of β_1

Assume again $M = 1$. We use estimator analogous to the estimators in the regular no-noise case, that is

$$\widehat{\beta}_t^n = f^{-1}\Big(\frac{\overline{C}(4)_t^n}{\overline{C}(2)_t^n}\Big).$$

where $f(x) = \frac{4^x - 1}{2^x - 1}$ and

$$\begin{aligned}\overline{C}(y)_t^n &= \widehat{C}(y)_t^n - \widehat{C}(1)_t \\ &= \tfrac{2}{\phi_n}\big(u_n^{\beta_1 - 2}(y^{\beta_1 - 2} - 1)\Delta_n^{1-\beta_1/2} A_t^1 + (Z(y)_t^n - Z(1)_t)\big)\end{aligned}$$

and we have (when the basic CLT holds):

$$u_n^{\beta_1/2}\big(\widehat{\beta}_t^n - \beta_1\big) \stackrel{\mathcal{L}-\mathrm{s}}{\longrightarrow} Y$$

for some mixed centered normal y with a known (conditional) variance. So here again we have to choose $u_n, h_n, k_n)$ with the appropriate conditions, plus u_n as large as possible.

Reporting only the rate for β_1, we find the following optimal rates (depending on the value of β_1)

- If $\sigma_t \equiv 0$ and (N1):
$$\begin{array}{ll} 1/\Delta_n^{\frac{\beta_1(1-\varepsilon)}{10-4\beta_1}} & \text{if } \beta_1 \leq \frac{3}{4} \\ 1/\Delta_n^{\frac{\beta_1(1-\varepsilon)}{7}} & \text{if } \frac{3}{4} \leq \beta_1 \leq \frac{3}{2} \\ 1/\Delta_n^{\frac{\beta_1(1-\varepsilon)}{4+2\beta_1}} & \text{if } \beta_1 \geq \frac{3}{2} \end{array}$$

- If $\sigma_t \equiv 0$ and (N2):
$$\begin{array}{ll} 1/\Delta_n^{\frac{3\beta_1(1-\varepsilon)}{30-11\beta_1}} & \text{if } \beta_1 \leq \frac{3}{4} \\ 1/\Delta_n^{\frac{3\beta_1(1-\varepsilon)}{21+\beta_1}} & \text{if } \frac{3}{4} \leq \beta_1 \leq \frac{3}{2} \\ 1/\Delta_n^{\frac{3\beta_1(1-\varepsilon)}{12+7\beta_1}} & \text{if } \beta_1 \geq \frac{3}{2} \end{array}$$

- If σ_t not 0 and (N1): $\begin{array}{ll} 1/\Delta_n^{\frac{\beta_1(1-\varepsilon)}{16-5\beta_1}} & \text{if } \beta_1 \leq \frac{16}{11} \\ 1/\Delta_n^{\frac{3\beta_1(1-\varepsilon)}{32-4\beta_1}} & \text{if } \beta_1 \geq \frac{16}{11} \end{array}$

- If σ_t not 0 and (N2): $\begin{array}{ll} 1/\Delta_n^{\frac{4\beta_1(1-\varepsilon)}{64-17\beta_1}} & \text{if } \beta_1 \leq \frac{16}{11} \\ 1/\Delta_n^{\frac{9\beta_1(1-\varepsilon)}{116-19\beta_1}} & \text{if } \beta_1 \geq \frac{16}{11} \end{array}$

Some problems:

1 - Optimality concerning β_1

2 - When $M \geq 2$, what does the rate for β_1 become ?, and can we estimate $\beta_2, \beta_, \ldots$?

2 - The multi-dimensional case (for the integrated co-volatilities)

3 - How to choose in practice $u_n h_n, k_n$) (so far, only "mathematical" results are known.