Inference for Semimartingales in the Presence of Noise

Mark Podolskij (CREATES and University of Aarhus, Denmark)

Joint work with J. Jacod and M. Vetter

Sandbjerg

January 29 – February 01, 2008

Set up

- ✘ We consider a 1-dimensional semimartingale ("true process") of the form

$$X = X_0 + \int_0^\cdot a_u \mathrm{d}u + \int_0^\cdot \sigma_u \mathrm{d}W_u + (x 1_{\{|x|\le 1\}}) * (\mu - \nu) + (x 1_{\{|x|>1\}}) * \mu$$

defined on some filtered probability space $(\Omega^0, \mathcal{F}^0, (\mathcal{F}_t^0)_{t\ge 0}, P^0)$. Here a is locally bounded, σ is cádlág adapted, μ is a jump measure of X and ν is its predictable compensator. The observed process Z, defined on $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\ge 0}, P)$, is given by

$$Z_t = X_t + U_t \ , \qquad t \ge 0.$$

The observation times are $t_i = i\Delta_n$ with $\Delta_n \to 0$. The probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\ge 0}, P)$ is constructed in such a way that

$$E[U_t|\mathcal{F}^0] = 0, \qquad \alpha_t^2 = E[U_t^2|\mathcal{F}^0] \ \text{ is cádlág }, \tag{1}$$

and, conditionally on $\mathcal{F}^0$, $U_t \perp U_s$ for $t \neq s$.

Equivalent representation of the model

- ✘ The process Z can be described by the following equation

$$Z_t = X_t + h(X_t)\varepsilon_t \ , \qquad t \geq 0 \ ,$$

where

$$\alpha_t = h(X_t) \ , \qquad E[\varepsilon_t | \mathcal{F}^0] = 0, \qquad E[\varepsilon_t^2 | \mathcal{F}^0] = 1 \ , \qquad t \geq 0 \ ,$$

and, conditionally on $\mathcal{F}^0$, $\varepsilon_t \perp \varepsilon_s$ for $t \neq s$.

- ✘ The latter representation can be interpreted as a time continuous nonlinear regression model.

Statement of the problem

- We are interested in the estimation of some characteristics of the true process X, i.e.
  - Estimation of the quadratic variation (incl. CLT)

$$[X,X]_t = \int_0^t \sigma_u^2 du + \sum_{u \le t} |\Delta X_u|^2.$$

  - Estimation of jumps (incl. CLT)

$$\sum_{u \le t} |\Delta X_u|^p \ , \qquad p > 2.$$

  - Estimation of the quantities (incl. CLT)

$$\int_0^t |\sigma_u|^p du$$

  for any $p > 0$ when X is continuous.

Some examples

(I) (*Additive i.i.d. noise*) Consider the process

$$Z_t = X_t + U_t$$

with $X \perp U$, U_t is i.i.d. and $EU_t = 0$, $EU_t^2 = \omega^2$. Such a process obviously satisfies the conditions of (1) and $\alpha_t^2 = \omega^2$.

(II) (*Additive i.i.d. noise + rounding*) Consider the process

$$Z_t = \gamma \Big[\frac{X_t + V_t}{\gamma} \Big] \ ,$$

where $X \perp V$, V_t is i.i.d. $V_t \sim U([0,\gamma])$ and $\gamma > 0$. Then Z satisfies the assumptions of (1) and

$$\alpha_t^2 = \gamma^2 \Big(\Big\{ \frac{X_t}{\gamma} \Big\} - \Big\{ \frac{X_t}{\gamma} \Big\}^2 \Big).$$

Some statistics

- We choose a sequence of integers $k_n\sqrt{\Delta_n} = \theta + o(\Delta_n^{1/4})$ for some $\theta > 0$ and consider a function $g : [0,1] \to \mathbb{R}$ (which is continuous, piecewise C^1 with piecewise Lipschitz derivative) with $g(0) = g(1) = 0$. Typical examples of such a function g are given by $g_1(x) = x \wedge (1-x)$ or $g_2(x) = \sin(\pi x)$. Moreover, we introduce the notation

$$G_p = \int_0^1 |g(s)|^p ds \ , \qquad H_p = \int_0^1 |g'(s)|^p ds.$$

- Next, we define the local moving average by

$$\bar{Z}_i^n = \sum_{j=1}^{k_n-1} g(\frac{j}{k_n})\Delta_{i+j}^n Z = -\sum_{j=1}^{k_n-1} \Big(g(\frac{j}{k_n}) - g(\frac{j-1}{k_n})\Big) Z_{\frac{i+j-1}{n}} \ ,$$

where $\Delta_i^n Z = Z_{\frac{i}{n}} - Z_{\frac{i-1}{n}}$. This quantity has been proposed by Podolskij & Vetter (2006) (see also Jacod, Li, Mykland, Podolskij & Vetter (2007)).

Intuition

- ✘ Assume for a moment that $X = W$, $W \perp U$, U_t is i.i.d. and $EU_t = 0$, $EU_t^2 = \omega^2$. A simple calculation shows that

$$\overline{W}_i^n \overset{asy}{\sim} N(0, \Delta_n k_n G_2) \ ,$$

and

$$\overline{U}_i^n \overset{asy}{\sim} N(0, k_n^{-1} H_2 \omega^2).$$

- ✘ This calculation shows that (with our choice of the sequence k_n) the influence of the Brownian part $\overline{W}_i^n$ and of the noise part $\overline{U}_i^n$ are balanced, i.e.

$$\overline{W}_i^n = O_p(\Delta_n^{1/4}) = \overline{U}_i^n \ ,$$

(which leads later to an optimal rate of convergence). This result remains true for general continuous semimartingales X.

Main results

- The core statistic of our method is defined by

$$V(Z,p)_t^n = \sum_{i=0}^{[t/\Delta_n]-k_n+1} |\overline{Z}_i^n|^p \ , \qquad p \geq 0.$$

- For bias corrections we need to define the following statistic

$$\hat{V}(Z)_t^n = \sum_{i=0}^{[t/\Delta_n]} |\Delta_i^n Z|^2.$$

Notice that $\frac{\Delta_n}{2}\hat{V}(Z)_t^n \xrightarrow{P} \int_0^t \alpha_u^2 du$.

- Before we proceed with the asymptotic results we introduce the process

$$Q_t(p) = E[|U_t|^p \ |\mathcal{F}^0], \qquad p > 0.$$

Main results: limit in probability

Theorem 1:

(i) Assume that $p > 2$ and $Q_t(p)$ is locally bounded. Then we have

$$\frac{1}{k_n} V(Z,p)_t^n \xrightarrow{P} G_p \sum_{u \le t} |\Delta X_u|^p.$$

(ii) If $p = 2$ and $Q_t(4)$ is locally bounded we have

$$\frac{1}{k_n} V(Z,2)_t^n - \frac{H_2}{2k_n^2} \hat{V}(Z)_t^n \xrightarrow{P} G_2 [X,X]_t.$$

(iii) Assume that X is continuous and $Q_t(2p)$ is locally bounded. Then we have

$$\Delta_n^{1-\frac{p}{4}} V(Z,p)_t^n \xrightarrow{P} \mu_p \int_0^t \left(\theta G_2 \sigma_u^2 + \frac{H_2}{\theta} \alpha_u^2\right)^{p/2} du \ ,$$

where $\mu_p = E[|v|^p]$ with $v \sim N(0,1)$.

Main results: central limit theorems

Before we present the central limit theorems we need to introduce some additional notations. We set

$$\phi_1(s) = \int_s^1 g'(u)g'(u-s)du \ , \quad \phi_2(s) = \int_s^1 g(u)g(u-s)du \ , \qquad s \in [0,1]$$

$$\Phi_{ij} = \int_0^1 \phi_i(s)\phi_j(s)ds \ , \qquad i,j = 1,2$$

Next, let B a new Brownian motion and $(V_m^-, V_m^+, U_m^-, U_m^+)_{m\geq 1}$ a sequence of random variables with

$$(V_m^-, V_m^+, U_m^-, U_m^+) \ \text{ i.i.d } \sim N(0, \text{diag}(\Psi_1^-, \Psi_1^+, \Psi_2^-, \Psi_2^+)) \ ,$$

where $\Psi_1^-, \Psi_1^+, \Psi_2^-, \Psi_2^+$ depend on the function g. B and $(V_m^-, V_m^+, U_m^-, U_m^+)_{m\geq 1}$ are independent, both defined on the extension of the probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, P)$, and independent of $\mathcal{F}$.

Main results: central limit theorems

✘ Finally, we introduce the processes

$$
\begin{aligned}
L(p)_t &= \frac{p}{G_p} \sum_{T_m \le t} \operatorname{sgn}(\Delta X_{T_m}) |\Delta X_{T_m}|^{p-1} \Big(\sqrt{\theta} \{ \sigma_{T_m -} U_m^- + \sigma_{T_m} U_m^+ \} \\
&+ \frac{1}{\sqrt{\theta}} \{ \alpha_{T_m -} V_m^- + \alpha_{T_m} V_m^+ \} \Big) , \\
\bar{L}_t &= \int_0^t \gamma_u \, dB_u ,
\end{aligned}
$$

where (T_m) are jump times of X and

$$
\gamma_u^2 = 4 \Big(\Phi_{22} \theta \sigma_u^4 + 2 \Phi_{12} \frac{\sigma_u^2 \alpha_u^2}{\theta} + \Phi_{11} \frac{\alpha_u^4}{\theta^3} \Big).
$$

✘ Notice that $L(p)_t$ and $\bar{L}_t$ are both mixed normal processes!

Stable convergence

In the following we will show some *stable* central limit theorems associated with Theorem 1. Let us briefly describe this concept (see Renyi (1963), Aldous & Eagleson (1978) or Jacod & Shiryaev (2003) for more details on stable convergence).

Definition: A sequence of random variables Y_n converges stably in law with limit Y ($Y_n \xrightarrow{\mathcal{D}_{st}} Y$), defined on an appropriate extension $(\Omega', \mathcal{F}', P')$ of the original probability space $(\Omega, \mathcal{F}, P)$, if and only if for any $\mathcal{F}$-measurable random variable V the convergence in distribution

$$(Y_n, V) \xrightarrow{\mathcal{D}} (Y, V)$$

holds.

Main results: central limit theorems

Theorem 2: For any $t > 0$ we obtain the following assertions.

(i) Assume that $p > 3$ and $Q_t(2p)$ is locally bounded. Then we have

$$\Delta_n^{-1/4}\Big(\frac{1}{k_n}V(Z,p)_t^n - G_p\sum_{u\leq t}|\Delta X_u|^p\Big) \xrightarrow{\mathcal{D}_{st}} L(p)_t.$$

(ii) If $p = 2$ and $Q_t(8)$ is locally bounded we have

$$\Delta_n^{-1/4}\Big(\frac{1}{k_n}V(Z,2)_t^n - \frac{H_2}{2k_n^2}\hat{V}(Z)_t^n - G_2[X,X]_t\Big) \xrightarrow{\mathcal{D}_{st}} L(2)_t + \bar{L}_t.$$

In particular, when X is continuous we have

$$\Delta_n^{-1/4}\Big(\frac{1}{k_n}V(Z,2)_t^n - \frac{H_2}{2k_n^2}\hat{V}(Z)_t^n - G_2\int_0^t \sigma_u^2\,du\Big) \xrightarrow{\mathcal{D}_{st}} \bar{L}_t.$$

Remarks

- The conditional variances

$$\Gamma(p)_t = E[|L(p)_t|^2 \ |\mathcal{F}] \ , \qquad \bar{\Gamma}_t = E[|\bar{L}_t|^2 \ |\mathcal{F}]$$

can be estimated, and, by the properties of stable convergence, we can get a feasible version of Theorem 2, i.e. for any estimators $\Gamma(p)_t^n \xrightarrow{P} \Gamma(p)_t$, $\bar{\Gamma}(p)_t^n \xrightarrow{P} \bar{\Gamma}(p)_t$ we have

$$\frac{\Delta_n^{-1/4}\Big(\frac{1}{k_n}V(Z,2)_t^n - \frac{H_2}{2k_n^2}\hat{V}(Z)_t^n - G_2[X,X]_t\Big)}{\sqrt{\Gamma(p)_t^n + \bar{\Gamma}(p)_t^n}} \xrightarrow{\mathcal{D}} N(0,1).$$

- As a by-product we obtain consistent estimates of σ_s^2, σ_{s-}^2, α_s^2, α_{s-}^2 (which are robust to jumps of the process X!).

Remarks

- Assume that X is continuous. As in Theorem 2 we can obtain a central limit theorem for estimators of $\int_0^t |\sigma_u|^p du$ for any even number $p > 0$. Suppose moreover that $X = \sigma W$ and $\alpha_t = \alpha$. In this case the rate $\Delta_n^{1/4}$ is known to be optimal (see Gloter & Jacod (2001)). Finally, we can minimize the conditional variance of the limit $\bar{L}_t$ by choosing an optimal θ.

- In Podolskij & Vetter (2007) we study the asymptotic behaviour of the functionals

$$V(Z,r,l)_t^n = \sum_{i=0}^{[t/\Delta_n]-2k_n+1} |\overline{Z}_i^n|^r |\overline{Z}_{i+k_n}^n|^l , \qquad r,l \geq 0.$$

  We provide consistent estimates for the quantities $\int_0^t |\sigma_u|^p du$ (for even $p > 0$), which are robust to certain specification of jumps (in the WLLN and CLT). This theory can be applied to construct tests for jumps in the presence of noise.