Quasi-Likelihood Estimation of a Lévy-driven SDE and Its Application

Hiroki Masuda

Kyushu University, Japan

hiroki@math.kyushu-u.ac.jp
http://www2.math.kyushu-u.ac.jp/~hiroki/

SAPS VII, March 16, 2009

Underlying parametric model

We observe a discrete-time sample $X_{t_0^n}, X_{t_1^n}, \ldots, X_{t_n^n}$ from

$$dX_t = a(X_t, \alpha)dt + b(X_{t-}, \beta)dZ_t.$$

- $t_i^n = ih_n$ for some positive h_n fulfilling

$$h_n \to 0, \quad T_n := nh_n \to \infty, \quad nh_n^2 \to 0.$$

- $\theta := (\alpha, \beta) \in \Theta_\alpha \times \Theta_\beta = \Theta \subset \mathbb{R}^{p_\alpha} \times \mathbb{R}^{p_\beta} = \mathbb{R}^p$:
  - Θ is a bounded convex domain;
  - $\theta_0 \in \Theta$ denotes the true value of θ.
- Z is a nontrivial centered Lévy process:

$$Z_t = \sigma w_t + \int_0^t \int z\{\mu(ds, dz) - ds\nu(dz)\},$$

characterized by $\sigma \geq 0$ and Lévy measure ν

Objective

Our goal is to provide

I. An asymptotically normal $\hat{\theta}_n = (\hat{\alpha}_n, \hat{\beta}_n)$, without specifying $\mathcal{L}(Z)$.

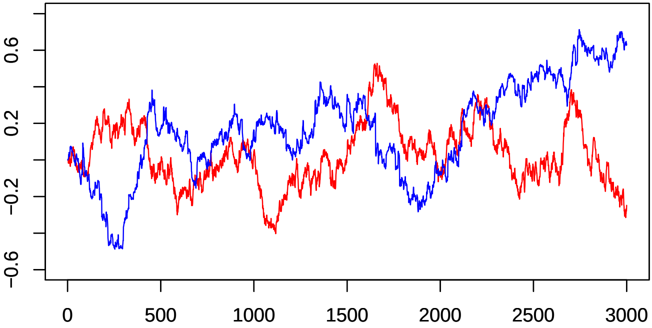
II. Statistics for testing the normality of Z:

$$\mathcal{H}_0\colon \nu(\mathbb{R}) = 0 \quad \text{v.s.} \quad \mathcal{H}_1\colon \nu(\mathbb{R}) \in (0, \infty].$$

We are going to consider

- Approximate quadratic martingale estimating function for **I**.
- Self-normalized partial sums of residuals for **II**.

Focus in II: “diffusion” versus “infinitely many small jumps”



Assumption 1: regularity of the coefficients

Either one of the following holds true.

1. **(Bounded smooth coefficients plus uniformly elliptic $b(x,\beta)$)**
    - $a \in \mathcal{C}^{\infty,2}(\mathbb{R}\times\Theta_\alpha)$ and $b \in \mathcal{C}^{\infty,2}(\mathbb{R}\times\Theta_\beta)$.
    - $\sup_{x,\theta}\{|\partial_x^j\partial_\alpha^k a(x,\alpha)| \vee |\partial_x^j\partial_\beta^k b(x,\beta)|\} < \infty$ for each $j \in \mathbb{Z}_+$ and $k \in \{0,1,2\}$.
    - $\inf_{x,\beta}|b(x,\beta)| > 0$.
2. **(Globally Lipschitz smooth coefficients plus nondegenerate $b(x,\beta)$)**
    - $a \in \mathcal{C}^{\infty,2}(\mathbb{R}\times\Theta_\alpha)$ and $b \in \mathcal{C}^{\infty,2}(\mathbb{R}\times\Theta_\beta)$.
    - $\sup_{x,\theta}\{|\partial_x a(x,\alpha)| \vee |\partial_x b(x,\beta)|\} < \infty$.
    - $\sup_{x,\theta}\{|\partial_x^j\partial_\alpha^k a(x,\alpha)| \vee |\partial_x^j\partial_\beta^k b(x,\beta)|\} \le C(1+|x|)^C$ for each $j \ge 2$ and $k \in \{0,1,2\}$.
    - $\sup_{x,\beta}|b(x,\beta)|^{-1} \le C(1+|x|)^C$.

- **The bounded case 1 enables us to bypass some moment conditions.**

Assumption 2: law of large numbers

There exists a unique invariant distribution π_0 (possibly depending on θ_0) such that $T_n^{-1} \int_0^{T_n} f(X_t)dt \to^p \pi_0(f)$ for any $f \in L^1(\pi_0)$.

- Typically, the ergodicity follows from:
  - The irreducibility and/or the local Doeblin condition;
  - Foster-Lyapunov drift criteria.

  cf. Kulik (2009) and M (2007,2008), among others.
- for the former, e.g., it suffices to have (even for $\sigma \neq 0$):
  - positive Lévy density in $\{z \in \mathbb{R} \backslash \{0\} : |z| < \epsilon\}$;
  - never vanishing b.

I. Approximate quadratic martingale estimating function

- **The SDE in question:** $dX_t = a(X_t, \alpha)dt + b(X_{t-}, \beta)dZ_t$.
- $f_{i-1}(\theta) := f(X_{t^n_{i-1}}, \theta)$ **for a function** f **defined on** $\mathbb{R} \times \Theta$.
- $\Delta_i X := X_{t^n_i} - X_{t^n_{i-1}}$.

We will consider estimators $\hat{\theta}_n = (\hat{\alpha}_n, \hat{\beta}_n)$ **of** θ **such that** $Q_n(\hat{\theta}_n) = 0$, **where** $Q_n : \Theta \to \mathbb{R}^p$ **are random functions given by**

$$Q_n(\theta) = \sum_{i=1}^{n} \begin{pmatrix} A(X_{t^n_{i-1}}, \theta)(\Delta_i X - h_n a_{i-1}(\alpha)) \\ B(X_{t^n_{i-1}}, \theta)\{(\Delta_i X - h_n a_{i-1}(\alpha))^2 - h_n b^2_{i-1}(\beta)\} \end{pmatrix},$$

with some $A(x, \theta) \in \mathbb{R}^{p_\alpha}$ **and** $B(x, \theta) \in \mathbb{R}^{p_\beta}$.

Assumption E1: finite moments of Z

There exists an integer $q > (p \vee 8)$ such that $E[|Z_t|^q] < \infty$, and in particular $E[Z_t] = 0$ and $E[Z_t^2] = t$.

- We require the index order "$q > (p \vee 8)$" to provide consistent estimators of the asymptotic variances of $\hat{\theta}_n$.
- We can always to rescale the centered Z to have variance 1.

Assumption E2: regularity

Either one of the following holds true.

1. In addition to Assumptions 1.1 and 2,
   - $A(x,\cdot) \in \mathcal{C}^2(\Theta)$ and $B(x,\cdot) \in \mathcal{C}^2(\Theta)$ for every $x \in \mathbb{R}$,
   - $\sup_{x,\theta}\{|\partial_\theta^k A(x,\theta)| \vee |\partial_\theta^k B(x,\theta)|\} < \infty$ for $k \in \{0,1,2\}$.
2. In addition to Assumptions 1.2 and 2,
   - $A(x,\cdot) \in \mathcal{C}^2(\Theta)$ and $B(x,\cdot) \in \mathcal{C}^2(\Theta)$ for every $x \in \mathbb{R}$,
   - $\sup_{\theta}\{|\partial_\theta^k A(x,\theta)| \vee |\partial_\theta^k B(x,\theta)|\} \leq C(1+|x|)^C$ for each $k \in \{0,1,2\}$,
   - $\sup_{t\in\mathbb{R}_+} E_0[|X_t|^q] < \infty$ for every $q > 0$.

Assumption E3: identifiability

Put $\pi_0(f(\cdot,\theta)) = \int f(x,\theta)\pi_0(dx)$ for a function on $\mathbb{R}\times\Theta$, and abbreviate $\pi_0(f(\cdot,\theta_0))$ to $\pi_0(f)$.

- $\pi_0(A\partial_\alpha^\top a) \in GL(p_\alpha,\mathbb{R})$ and $\pi_0(B\partial_\beta^\top b^2) \in GL(p_\beta,\mathbb{R})$.
- We have

$$
\begin{aligned}
&|\pi_0(A(\cdot,\theta)\{a(\cdot,\alpha_0) - a(\cdot,\alpha)\})| \\
&\qquad + |\pi_0(B(\cdot,\theta)\{b^2(\cdot,\beta_0) - b^2(\cdot,\beta)\})| = 0
\end{aligned}
$$

iff $\theta = \theta_0$.

Notation

Put $\nu_k = \int z^k \nu(dz)$ for $k \in \mathbb{N}$.

Under the assumptions, the quantities

$$V_0' = \mathrm{diag}(V_{11}, 2V_{22}), \quad V_0'' = \begin{pmatrix} V_{11} & \nu_3 V_{12} \\ \text{Sym.} & \nu_4 V_{22} \end{pmatrix}$$

are well-defined, where

$$V_{11} = \pi_0(\partial_\alpha a A^\top)^{-1} \pi_0(A^{\otimes 2} b^2) \pi_0(A \partial_\alpha^\top a)^{-1},$$
$$V_{12} = \pi_0(\partial_\alpha a A^\top)^{-1} \pi_0(A B^\top b^3) \pi_0(B \partial_\beta^\top b^2)^{-1},$$
$$V_{22} = \pi_0(\partial_\beta b^2 B)^{-1} \pi_0(B^{\otimes 2} b^4) \pi_0(B \partial_\beta^\top b^2)^{-1}.$$

Theorem 1

Suppose Assumptions E1 to E3. Then, there exists a measurable $\mathbb{R}^p$-valued sequence $\hat{\theta}_n$ such that $P_0[Q_n(\hat{\theta}_n) = 0] \to 1$, and any such sequence fulfils $\hat{\theta}_n \to^p \theta_0$. Moreover:

- **if $\nu(\mathbb{R}) = 0$, then $\left(\sqrt{T_n}(\hat{\alpha}_n - \alpha_0), \sqrt{n}(\hat{\beta}_n - \beta_0)\right) \to^d \mathcal{N}_p(0, V_0')$;**
- **if $\nu(\mathbb{R}) > 0$, then $\sqrt{T_n}(\hat{\theta}_n - \theta_0) \to^d \mathcal{N}_p(0, V_0'')$.**

- **The diffusion case is formally well known, but cases where π_0 is possibly "heavy-tailed" do not seem to have been explicitly mentioned as yet; in this case we never have the condition $\sup_{t\in\mathbb{R}_+} E_0[|X_t|^q] < \infty$ even for small $q > 0$.**
- **The convergence rate of $\hat{\beta}_n$ under $\nu(\mathbb{R}) > 0$ is not optimal.**
- **It is easy to give consistent estimators of the asymptotic variances.**
- **The optimal choice when $\nu_3 = 0$ is:**
  $A(x,\theta) \leftarrow \frac{\partial_\alpha a}{b^2}(x,\theta)$ and $B(x,\theta) \leftarrow \frac{\partial_\beta b^2}{b^4}(x,\beta)$.

II. Normality test for the driving process Z

Define

$$\hat{\epsilon}_{ni} = \frac{\Delta_i X - a_{i-1}(\hat{\alpha}_n) h_n}{\sqrt{h_n} b_{i-1}(\hat{\beta}_n)},$$

Writing $\bar{\hat{\epsilon}}_n = \frac{1}{n}\sum_{i=1}^{n} \hat{\epsilon}_{ni}$, we introduce the kth self-normalized residual sums ($k \geq 2$):

$$\hat{\Phi}_n^{(k)} = \frac{\hat{\Psi}_n^{(k)}}{(\hat{\Psi}_n^{(2)})^{k/2}}, \quad \text{where} \quad \hat{\Psi}_n^{(k)} := \frac{1}{n}\sum_{i=1}^{n} (\hat{\epsilon}_{ni} - \bar{\hat{\epsilon}}_n)^k.$$

Our test statistics $\mathcal{T}_n$, $n \in \mathbb{N}$, are defined to be

$$\mathcal{T}_n = \frac{n}{6}\left\{\hat{\Phi}_n^{(3)} - \frac{3\sqrt{h_n}}{n}\sum_{i=1}^{n} \partial_x b_{i-1}(\hat{\beta}_n)\right\}^2 + \frac{n}{24}(\hat{\Phi}_n^{(4)} - 3)^2.$$

Theorem 2

$$\mathcal{T}_n = \frac{n}{6}\left\{\hat{\Phi}_n^{(3)} - \frac{3\sqrt{h_n}}{n}\sum_{i=1}^{n}\partial_x b_{i-1}(\hat{\beta}_n)\right\}^2 + \frac{n}{24}(\hat{\Phi}_n^{(4)} - 3)^2.$$

Under the conditions of Theorem 1, we have:

- $\mathcal{T}_n \to^d \chi^2(2)$ under $\mathcal{H}_0$;
- $P_0[\mathcal{T}_n > K] \to 1$ for every $K > 0$ under $\mathcal{H}_1$.

- The latter claim implies that our test is consistent: i.e., under $\mathcal{H}_1$, we can reject $\mathcal{H}_0$ with probability tending to 1.
- Given sampling points t_i^n, Theorem 2 enables us to perform a consistent test for the normality of the “unobserved” driving Lévy process Z, without any fine-tuning parameter.

Corollary to Theorem 2

Set

$$\mathcal{T}_n' = \frac{n}{6}\left\{\hat{\Phi}_n^{(3)} - \frac{3\sqrt{h_n}}{n}\sum_{i=1}^{n}\partial_x b_{i-1}(\hat{\beta}_n)\right\}^2.$$

Under the conditions of Theorem 1, we have:

- $\mathcal{T}_n' \to^d \chi^2(1)$ **under** $\mathcal{H}_0$;
- $P_0[\mathcal{T}_n' > K] \to 1$ **for every** $K > 0$ **under** $\mathcal{H}_1$.

- **Differently from the i.i.d. sample case, the consistency follows only on using the sample skewness type statistics: but, if $\mathcal{L}(Z_1)$ is symmetric, $\mathcal{T}_n'$ leads to poorer finite-sample performance than $\mathcal{T}_n$.**

Scaled Student diffusion and its NIG-driven counterpart

Consider the parametric model ($\theta = (\alpha', \alpha'', \beta)$)

$$dX_t = \left\{ \alpha' \sigma(X_t) \partial_x \sigma(X_t) - \alpha'' \frac{X_t}{1 + X_t^2} \sigma^2(X_t) \right\} dt + \beta \sigma(X_t) dZ_t.$$

- $\sigma : \mathbb{R} \to \mathbb{R}$ is a known smooth function such that:
  - $\sup_x |\partial_x^k \sigma(x)| < \infty$ for each $k \in \mathbb{Z}_+$;
  - $0 < \underline{\sigma} \leq \sigma(x) \leq \overline{\sigma} < \infty$;
  - $|x \partial_x \sigma(x)| \to 0$ for $|x| \to \infty$.
- $2\alpha_0'' > \beta_0^2$ and $\beta_0 > 0$: then X is ergodic.

If Z is a standard Wiener process, then:

- The tail of π_0 behaves like $C|x|^{-2\alpha_0''/\beta^2}$, and especially, π_0 is a scaled Student-t if $\sigma \equiv 1$;
- X is "polynomially" strong-mixing: use Veretennikov (2006) with a minor modification.

Simulation design

In our setup, $\mathcal{H}_0 : Z$ is a standard Winer process.

To observe empirical powers, we take $\mathcal{L}(Z_t) = NIG(\delta, 0, \delta t, 0)$ for $\delta > 0$ under $\mathcal{H}_1$: Z is the the normal inverse Gaussian Lévy process.

- We set $\sigma \equiv 1$ and $\theta = (\alpha', \beta)$ with $\theta_0 = (1, 1)$:

$$dX_t = -\frac{\alpha' X_t}{1 + X_t^2} dt + \beta dZ_t.$$

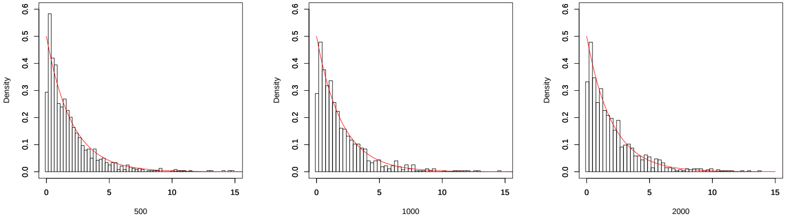
- $A(x, \theta) \leftarrow \frac{\partial_\alpha a}{b^2}(x, \theta)$ and $B(x, \theta) \leftarrow \frac{\partial_\beta b^2}{b^4}(x, \beta)$.
- Two cases for $\mathcal{H}_1$: $\delta = 5$ and 10. The latter is "closer" to $\mathcal{H}_0$.
- We simulate $L = 1000$ independent paths of X under $\mathcal{H}_0$ and $\mathcal{H}_1$ with different Z as above.

Under $\mathcal{H}_0$: empirical sizes

n	$h_n \approx$	$T_n \approx$	5%	1%	Mean	S.D.
500	0.12	60	0.050	0.018	1.9471	2.5204
1000	0.08	79	0.061	0.014	1.9977	2.2567
2000	0.05	105	0.053	0.015	2.0512	2.0917

- **From Theorem 2, both of "Mean" and "S.D." should be close to 2.**

Relative-frequency histogram



- **In each panel, the straight line indicates the target $\chi^2(2)$-density.**

Under $\mathcal{H}_1$: empirical powers

n	$h_n \approx$	$T_n \approx$	$(\alpha, \delta) = (5, 5)$		$(\alpha, \delta) = (10, 10)$	
			5%	1%	5%	1%
500	0.12	60	0.830	0.720	0.213	0.119
1000	0.08	79	0.999	0.996	0.532	0.365
2000	0.05	105	1	1	0.951	0.900

Concluding remarks: further topics

Our estimation procedure and normality test are very easy to implement.

- **A more general class of diffusions with jumps can be treated:**

  $$dX_t = a(X_t, \alpha)dt + b(X_t, \beta)dw_t + c(X_{t-}, \gamma)dZ_t,$$

  with Z being a centered pure-jump Lévy process. Once again the estimator $\hat{\theta}_n$ constructed here may be a good candidate to perform a consistent test for whether or not $\nu(\mathbb{R}) = 0$.
- **Multivariate extension:**
  - **The construction of $\hat{\theta}_n$ remain valid;**
  - **Many possibility for testing the normality.**