Cauchy quasi-likelihood in SDE estimation

Hiroki Masuda

Graduate School of Mathematics, Kyushu University, Japan

SAPS VIII
University of Maine, Le Mans
March 21, 2011

Brief summary

- When observing a discrete-time but high-frequency sample

$$X_0, X_{h_n}, X_{2h_n}, \dots, X_{nh_n}, \quad \text{where } h_n \to 0,$$

from the (semi-)parametric Lévy driven SDE

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

how can we estimate $\theta_0 = (\alpha_0, \gamma_0)$, the true value of $\theta := (\alpha, \gamma)$?

- We will provide an estimator $\hat{\theta}_n = (\hat{\alpha}_n, \hat{\gamma}_n)$ s.t.

$$\left\{\left(\sqrt{n}h_n^{1-1/\beta}(\hat{\alpha}_n - \alpha_0),\ \sqrt{n}(\hat{\gamma}_n - \gamma_0)\right)\right\}_{n\in\mathbb{N}} \quad \text{is asymp. normal,}$$

with β denoting the activity index of the Lévy process Z.

Brief summary

- When observing a discrete-time but high-frequency sample

$$X_0, X_{h_n}, X_{2h_n}, \dots, X_{nh_n}, \quad \text{where } h_n \to 0,$$

from the (semi-)parametric Lévy driven SDE

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

how can we estimate $\theta_0 = (\alpha_0, \gamma_0)$, the true value of $\theta := (\alpha, \gamma)$?

- We will provide an estimator $\hat{\theta}_n = (\hat{\alpha}_n, \hat{\gamma}_n)$ s.t.

$$\left\{ \left(\sqrt{n} h_n^{1-1/\beta}(\hat{\alpha}_n - \alpha_0),\ \sqrt{n}(\hat{\gamma}_n - \gamma_0) \right) \right\}_{n\in\mathbb{N}} \quad \text{is asymp. normal,}$$

with β denoting the activity index of the Lévy process Z.

Gaussian Quasi-Likelihood Estimator (GQLE) @ SAPS 7

- **Consists of fitting one-step conditional mean and variances:**
  - **Originally due to Wedderburn (1974).**

  To formulate the estimation procedure, it is enough to have

$$E[X_{t_j}|X_{t_{j-1}}] = m_{j-1}(\theta) \text{ and } Var[X_{t_j}|X_{t_{j-1}}] = v_{j-1}(\theta).$$

  explicitly.
- **The GQLE is formally given by the argmax of**

$$\theta \mapsto \sum_{j=1}^{n} \log\left\{\frac{1}{\sqrt{v_{j-1}(\theta)}}\phi\left(\frac{X_{t_j} - m_{j-1}(\theta)}{\sqrt{v_{j-1}(\theta)}}\right)\right\},$$

  ϕ **denoting the** $\mathcal{N}(0,1)$**-density.**

GQLE for discretely observed Lévy driven SDE *

- **Based on $X_{h_n}, X_{2h_n}, \dots, X_{nh_n}$ stemming from the ergodic**

$$dX_t = a(X_t, \alpha)dt + c(X_t, \gamma)dZ_t,$$

we want to estimate $\theta = (\alpha, \gamma)$, where Z is a Lévy process s.t. $E[Z_t] = 0$ and $E[Z_t^2] = t$.

- **"Aggressive" approximation $\mathcal{L}(Z_{h_n}) \approx \mathcal{N}(0, h_n)$ for small h_n:**

$$\begin{aligned} X_{jh_n} &\approx X_{(j-1)h_n} + a(X_{(j-1)h_n}, \alpha_0)h_n \\ &\qquad + c(X_{(j-1)h_n}, \gamma_0)(Z_{jh_n} - Z_{(j-1)h_n}) \\ &\sim \mathcal{N}\left(X_{(j-1)h_n} + a(X_{(j-1)h_n}, \alpha_0)h_n, c(X_{(j-1)h_n}, \gamma_0)^2 h_n\right), \end{aligned}$$

making the GQLE procedure explicit.

*M (2010, preprint) and the references therein.

GQLE for discretely observed Lévy driven SDE *

- Based on $X_{h_n}, X_{2h_n}, \dots, X_{nh_n}$ stemming from the ergodic

$$dX_t = a(X_t, \alpha)dt + c(X_t, \gamma)dZ_t,$$

we want to estimate $\theta = (\alpha, \gamma)$, where Z is a Lévy process s.t. $E[Z_t] = 0$ and $E[Z_t^2] = t$.

- "Aggressive" approximation $\mathcal{L}(Z_{h_n}) \approx \mathcal{N}(0, h_n)$ for small h_n:

$$\begin{aligned} X_{jh_n} &\approx X_{(j-1)h_n} + a(X_{(j-1)h_n}, \alpha_0)h_n \\ &\qquad + c(X_{(j-1)h_n}, \gamma_0)(Z_{jh_n} - Z_{(j-1)h_n}) \\ &\sim \mathcal{N}\left(X_{(j-1)h_n} + a(X_{(j-1)h_n}, \alpha_0)h_n, c(X_{(j-1)h_n}, \gamma_0)^2 h_n\right), \end{aligned}$$

making the GQLE procedure explicit.

*M (2010, preprint) and the references therein.

Resulting phenomena

$$dX_t = a(X_t, \alpha)dt + c(X_t, \gamma)dZ_t$$

- The GQLE $\hat{\theta}_n = (\hat{\alpha}_n, \hat{\gamma}_n)$ are asymptotically normal:

$$\left(\sqrt{nh_n}(\hat{\alpha}_n - \alpha_0), \sqrt{n}(\hat{\gamma}_n - \gamma_0)\right) \to^d \mathcal{N}(0, V') \quad \text{if } \nu(\mathbb{R}) = 0;$$
$$\left(\sqrt{nh_n}(\hat{\alpha}_n - \alpha_0), \sqrt{nh_n}(\hat{\gamma}_n - \gamma_0)\right) \to^d \mathcal{N}(0, V'') \quad \text{if } \nu(\mathbb{R}) > 0,$$

  where ν is the Lévy measure of Z.
- Existence of “any” jump part in Z slows down the convergence rate.

Our goal of this talk

- **Provide an estimator of the true value of $\theta = (\alpha, \gamma)$ in**

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

  based on $X_0, X_{h_n}, X_{2h_n}, \dots, X_{nh_n}$ ($h_n \to 0$).
- **We want to deal with pure-jump Z with infinite activity; e.g. Generalized hyperbolic, Meixner, tempered stable, etc.**

Non-Gaussian Quasi-Likelihood Estimation (NGQLE)

Target:

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

- Z is a **pure-jump Lévy process of infinite activity.**
- **The parameter** $\theta := (\alpha, \gamma) \in \Theta_\alpha \times \Theta_\gamma = \Theta \subset \mathbb{R}^p$, **a bounded convex domain, the true value** $\theta_0 := (\alpha_0, \gamma_0) \in \Theta$.

Notation:

- $\Delta_j X := X_{jh_n} - X_{(j-1)h_n}$.
- $f_{j-1}(\theta) := f(X_{(j-1)h_n}, \theta)$ **for any function of the form** $f(x, \theta)$.

A1. Regularity of the coefficients

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

1. a and c are smooth in $\mathbb{R} \times \Theta$.
2. $a(\cdot, \alpha_0)$ and $c(\cdot, \gamma_0)$ are globally Lipschitz.
3. $\exists c \in (1, \infty)$ s.t. $\forall (x, \gamma)$: $0 < c^{-1} \leq c(x, \gamma) \leq c$.
4. If X is not a Lévy process, then
   $\exists c', M > 0$ s.t. $\forall |x| \geq M$: $xa(x, \alpha_0) \leq -c'|x|^2$.

* X can be then ergodic under the true image measure P_0, under a good behavior of ν around the origin.
  The unique invariant distribution will be denoted by $\pi_0(dx)$.
  (cf. Alexey Kulik (2009) and M (2007)).

A2. Driving noise

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

1. $\mathcal{L}(Z_1)$ is symmetric around 0, and the Lévy measure ν of Z fulfils

$$\nu(dz) = {}^{\exists}g_0(z)dz \quad \text{s.t.} \quad g_0(z) = \frac{c_0}{|z|^{1+\beta}}\{1 + O(|z|)\}, \quad |z| \to 0.$$

* Then $\mathcal{L}(h^{-1/\beta}Z_h)$ admits a positive density $f_h(y)$ s.t. $\mathcal{L}(h^{-1/\beta}Z_h) \underset{h\to 0}{\Rightarrow}$ β-stable law with the C.F. $u \mapsto \exp(-|u|^\beta)$ for some $\beta \in (0,2)$: ϕ_β denotes the density.

2. There exist constant $\epsilon_n \to 0$ and Lebesgue-integrable λ s.t.

$$\sqrt{n}\int |f_h(y) - \phi_\beta(y)|dy \to 0.$$

* This holds for, e.g., the NIG Z if $nh_n^{2-\kappa} \to 0$ for some $\kappa > 0$.

Construction of our estimator

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

- **Again, the naive Euler type approximation:**

$$\begin{aligned} X_{jh_n} &\approx^{P_0} X_{(j-1)h_n} + a_{j-1}(\alpha_0)h_n + c_{j-1}(\gamma_0)\Delta_j Z \\ &= X_{(j-1)h_n} + a_{j-1}(\alpha_0)h_n + c_{j-1}(\gamma_0)h_n^{1/\beta}\frac{\Delta_j Z}{h_n^{1/\beta}} \end{aligned}$$

$\Rightarrow$ $\epsilon_{nj}(\theta_0) := \dfrac{\Delta_j X - a_{j-1}(\alpha_0)h_n}{h_n^{1/\beta}c_{j-1}(\gamma_0)} \overset{d}{\approx}$ β-stable (density ϕ_β, say).

- We define our estimator $\hat{\theta}_n = (\hat{\alpha}_n, \hat{\gamma}_n)$ through the quasi-likelihood:

$$\hat{\theta}_n \in \underset{\theta\in\Theta}{\text{argmax}} \sum_{j=1}^n W_{j-1} \log\left\{\frac{1}{h_n^{1/\beta}c_{j-1}(\gamma)}\phi_\beta\left(\epsilon_{nj}(\theta)\right)\right\}.$$

Construction of our estimator

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

- **Again, the naive Euler type approximation:**

$$\begin{aligned} X_{jh_n} &\approx^{P_0} X_{(j-1)h_n} + a_{j-1}(\alpha_0)h_n + c_{j-1}(\gamma_0)\Delta_j Z \\ &= X_{(j-1)h_n} + a_{j-1}(\alpha_0)h_n + c_{j-1}(\gamma_0)h_n^{1/\beta}\frac{\Delta_j Z}{h_n^{1/\beta}} \end{aligned}$$

$\Rightarrow$ $\epsilon_{nj}(\theta_0) := \dfrac{\Delta_j X - a_{j-1}(\alpha_0)h_n}{h_n^{1/\beta}c_{j-1}(\gamma_0)} \overset{d}{\approx}$ **β-stable (density ϕ_β, say).**

- **We define our estimator $\hat{\theta}_n = (\hat{\alpha}_n, \hat{\gamma}_n)$ through the quasi-likelihood:**

$$\hat{\theta}_n \in \operatorname*{argmax}_{\theta\in\Theta} \sum_{j=1}^{n} W_{j-1} \log\left\{\frac{1}{h_n^{1/\beta}c_{j-1}(\gamma)}\phi_\beta\left(\epsilon_{nj}(\theta)\right)\right\}.$$

Construction of our estimator

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

- **Again, the naive Euler type approximation:**

$$\begin{aligned} X_{jh_n} &\approx^{P_0} X_{(j-1)h_n} + a_{j-1}(\alpha_0)h_n + c_{j-1}(\gamma_0)\Delta_j Z \\ &= X_{(j-1)h_n} + a_{j-1}(\alpha_0)h_n + c_{j-1}(\gamma_0)h_n^{1/\beta}\frac{\Delta_j Z}{h_n^{1/\beta}} \end{aligned}$$

$\Rightarrow$ $\epsilon_{nj}(\theta_0) := \dfrac{\Delta_j X - a_{j-1}(\alpha_0)h_n}{h_n^{1/\beta}c_{j-1}(\gamma_0)} \overset{d}{\approx} \beta$**-stable (density ϕ_β, say).**

- **We define our estimator $\hat{\theta}_n = (\hat{\alpha}_n, \hat{\gamma}_n)$ through the quasi-likelihood:**

$$\hat{\theta}_n \in \underset{\theta\in\Theta}{\text{argmax}} \sum_{j=1}^{n} W_{j-1} \log\left\{\frac{1}{h_n^{1/\beta}c_{j-1}(\gamma)}\phi_\beta\left(\epsilon_{nj}(\theta)\right)\right\}.$$

A3. Sampling rate

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

1. $\beta \geq 1$ if X is a Lévy process (we then do not need $nh_n \to \infty$).
2. Otherwise, $\beta > 1$, $nh_n \to \infty$, and

$$\exists \epsilon_0 > 0 \quad \text{s.t.} \quad \limsup_{n\to\infty} nh_n^{3-2/\beta-\epsilon_0} < \infty.$$

A4. Weight function; for heavy-tailed cases

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

1. **$W : \mathbb{R} \to \mathbb{R}_+$ is bounded, and $W \equiv 1$ if X is a Lévy process.**
2. **If X is not a Lévy process, then $\exists\ K : \mathbb{R} \to \mathbb{R}_+$ s.t.**
    1. $\sup_{\theta\in\Theta} W(x)\{|\partial_\alpha a(x,\alpha)| + |\partial_\alpha a(x,\alpha)|^2 + |\partial_\alpha^2 a(x,\alpha)| + |\partial_\gamma c(x,\gamma)| + |\partial_\gamma c(x,\gamma)|^2 + |\partial_\gamma^2 c(x,\gamma)|\} \leq K(x)$,
    2. $\sup_{t\in\mathbb{R}_+} E_0[K(X_t)] < \infty$.

A5. Nonsingularity and identifiability

For $g(y) := \frac{\partial \phi_\beta}{\phi_\beta}(y)$,

1. $\det\{ \int W(x) \frac{[\partial_\alpha a(x,\alpha_0)]^{\otimes 2}}{c(x,\gamma_0)^2} \pi_0(dx)\} \cdot \det\{ \int W(x) \frac{[\partial_\gamma c(x,\gamma_0)]^{\otimes 2}}{c(x,\gamma_0)^2} \pi_0(dx)\} \neq 0.$
2. $\iint W(x) \frac{\partial_\alpha a(x,\alpha)}{c(x,\gamma)^2} \{a(x,\alpha_0) - a(x,\alpha)\} \partial g(\frac{c(x,\gamma_0)}{c(x,\gamma)} y) \phi_\beta(y) dy \pi_0(dx) = 0$ iff $\theta = \theta_0$.
3. $\iint W(x) \frac{\partial_\gamma c(x,\gamma)}{c(x,\gamma)} \{1 + \frac{c(x,\gamma_0)}{c(x,\gamma)} y g(\frac{c(x,\gamma_0)}{c(x,\gamma)} y)\} \phi_\beta(y) dy \pi_0(dx) = 0$ iff $\theta = \theta_0$.

Set $\pi_0 = \phi_\beta$ in case where X is a Lévy process.

Main claim: Asymptotic Normality

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

$$\left(\sqrt{n}h_n^{1-1/\beta}(\hat{\alpha}_n - \alpha_0), \sqrt{n}(\hat{\gamma}_n - \gamma_0)\right) \Rightarrow \mathcal{N}\left(0, \mathrm{diag}[U(\theta_0)^{-1}, V(\theta_0)^{-1}]\right),$$

where

$$U(\theta_0) = \int W(x)\frac{\{\partial_\alpha a(x,\alpha_0)\}^{\otimes 2}}{c(x,\gamma_0)^2}\pi_0(dx) \cdot \int \frac{\partial\phi_\beta(y)^2}{\phi_\beta(y)}dy,$$

$$V(\theta_0) = \int W(x)\frac{\{\partial_\gamma c(x,\gamma_0)\}^{\otimes 2}}{c(x,\gamma_0)^2}\pi_0(dx) \cdot \int \frac{\{\phi_\beta(y) + y\partial\phi_\beta(y)\}^2}{\phi_\beta(y)}dy$$

Set $W \equiv 1$ and $\pi_0 = \phi_\beta$ in case where X is a Lévy process.

A comparison with the Gaussian quasi-likelihood case

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

Contrast	Rates	
	α	γ
Gaussian QL	$\sqrt{nh_n}$	$\sqrt{nh_n}$
Non-Gaussian (Stable) QL	$\sqrt{n}h_n^{1-1/\beta}$	$\sqrt{n}$

- GQLE is easier to use, but NGQLE has better performance.
- Both are somewhat robust for the specification of the Lévy measure.
- However, we conjecture that the NGQLE is asymptotically optimal.

A comparison with the Gaussian quasi-likelihood case

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

Contrast	Rates α	Rates γ
Gaussian QL	$\sqrt{nh_n}$	$\sqrt{nh_n}$
Non-Gaussian (Stable) QL	$\sqrt{n}h_n^{1-1/\beta}$	$\sqrt{n}$

- **GQLE is easier to use, but NGQLE has better performance.**
- Both are somewhat robust for the specification of the Lévy measure.
- However, we conjecture that the NGQLE is asymptotically optimal.

A comparison with the Gaussian quasi-likelihood case

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

Contrast	Rates	
	α	γ
Gaussian QL	$\sqrt{nh_n}$	$\sqrt{nh_n}$
Non-Gaussian (Stable) QL	$\sqrt{n}h_n^{1-1/\beta}$	$\sqrt{n}$

- **GQLE is easier to use, but NGQLE has better performance.**
- **Both are somewhat robust for the specification of the Lévy measure.**
- **However, we conjecture that the NGQLE is asymptotically optimal.**

A comparison with the Gaussian quasi-likelihood case

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

Contrast	Rates α	Rates γ
Gaussian QL	$\sqrt{nh_n}$	$\sqrt{nh_n}$
Non-Gaussian (Stable) QL	$\sqrt{n}h_n^{1-1/\beta}$	$\sqrt{n}$

- **GQLE is easier to use, but NGQLE has better performance.**
- **Both are somewhat robust for the specification of the Lévy measure.**
- **However, we conjecture that the NGQLE is asymptotically optimal.**

A small numerical example: NIG Lévy process

- We set $X_t = \alpha t + \gamma Z_t$ with $\mathcal{L}(Z_t) = NIG(a, 0, t, 0)$ for some (unknown) $a > 0$, hence

$$\frac{X_t - \alpha t}{\gamma t} \sim NIG(at, 0, 1, 0) \to^d \text{ standard Cauchy.}$$

- $\theta_0 = (\alpha_0, \gamma_0) \leftarrow (-3, 2)$, $\beta = 1$, and $a = 2$.
- 1000 iterations with $n = 500$ and $h_n = 1/n$.
- Results.

	Sample median	Stable QLE α	Stable QLE γ
Mean	-2.9961	-2.9942	1.9781
S.D.	0.1430	0.1272	0.1237
Max	-2.5186	-2.5852	2.3635
Min	-3.4808	-3.4704	1.6225

A small numerical example: NIG Lévy process

- We set $X_t = \alpha t + \gamma Z_t$ with $\mathcal{L}(Z_t) = NIG(a, 0, t, 0)$ for some (unknown) $a > 0$, hence

$$\frac{X_t - \alpha t}{\gamma t} \sim NIG(at, 0, 1, 0) \to^d \text{standard Cauchy.}$$

- $\theta_0 = (\alpha_0, \gamma_0) \leftarrow (-3, 2)$, $\beta = 1$, and $a = 2$.
- 1000 iterations with $n = 500$ and $h_n = 1/n$.
- Results.

	Sample median	Stable QLE α	Stable QLE γ
Mean	-2.9961	-2.9942	1.9781
S.D.	0.1430	0.1272	0.1237
Max	-2.5186	-2.5852	2.3635
Min	-3.4808	-3.4704	1.6225

Achieving the normality of the NGQLE