Asymptotics for Student-Lévy regression *

Hiroki Masuda[1,2 †]   Lorenzo Mercuri[3]   Yuma Uehara[4]

[1]University of Tokyo and [2]JST CREST
[3]University of Milan
[4]Kansai University

ICIAM 2023
Waseda University
August 25, 2023

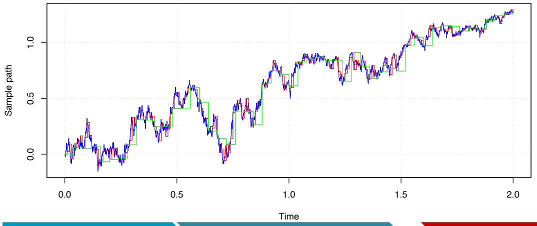
*This version: August 24, 2023

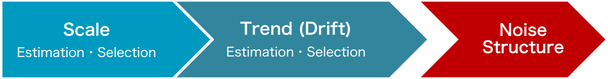
†Partly supported by JSPS KAKENHI and JST CREST

Objective in general

- Inference for **coefficients** and driving **non-Gaussian noise** process

$$X_{t+dt} \leftarrow X_t + (\text{Trend})dt + (\text{Scale}) \cdot d(\text{Noise})$$





Masuda, H., Mercuri, L., and Uehara, Y. (2022).
Noise inference for ergodic Lévy driven SDE.
Electron. J. Statist., 16(1):2432–2474.

Masuda, H., Mercuri, L., and Uehara, Y. (2023).
Quasi-likelihood analysis for Student-Lévy regression.
arXiv:2306.16790.

Objective in this talk: A non-trivial case study with t-noise

- Discrete-time sample $\{(X_{t_j}, Y_{t_j})\}_{j=0}^{[nT_n]}$ with $t_j = jh = j/n$ ($h = 1/n$) from

$$Y_t = X_t \cdot \mu + \sigma J_t, \qquad t \in [0, T_n].$$

  - $X = (X_t)$ is a càdlàg stochastic process in $\mathbb{R}^q$ satisfying some regularity conditions.
  - $J = (J_t)$ is a Lévy process such that the unit-time distribution $\mathcal{L}(J_1) = t_\nu$:

$$t_\nu \sim \text{pdf} \quad f(x;\nu) := \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\pi}\Gamma(\frac{\nu}{2})}(1+x^2)^{-(\nu+1)/2}.$$

- Parameter of interest: $\theta := (\mu, \sigma, \nu) \in \Theta_\mu \times \Theta_\sigma \times \Theta_\nu = \Theta$
- **Our objective** is to estimate the true value $\theta_0 = (\mu_0, \sigma_0, \nu_0) \in \Theta$ from $\{(X_{t_j}, Y_{t_j})\}_{j=0}^{[nT_n]}$ when:
  - $T_n \to \infty$ as $n \to \infty$;
  - Θ is a bounded convex domain in $\mathbb{R}^q + 2$ such that $\overline{\Theta} \subset \mathbb{R}^q \times (0,\infty) \times (0,\infty)$.

- [Massing, 2019] studied the LAN property for Student-t Lévy process **when** $\nu > 1$ **is known**.

Sample $\{(X_{t_j}, Y_{t_j})\}_{j=0}^{[nT_n]}$ from $Y_t = X_t \cdot \mu + \sigma J_t$, where $t_j = jh = j/n$ and $\mathcal{L}(J_1) = t_\nu := t_\nu(0,1)$.

$$\sigma^{-1}\left(Y_{t_j} - Y_{t_{j-1}} - (X_{t_j} - X_{t_{j-1}}) \cdot \mu\right) \overset{\theta}{=} J_{t_j} - J_{t_{j-1}} \sim J_h$$

- The Student-t distribution $t_\nu(0,1)$ is **not closed under convolution**:

$$\varphi_\nu(u) := E[e^{iuJ_1}] = \frac{2^{1-\nu/2}}{\Gamma(\nu/2)}|u|^{\nu/2}K_{\nu/2}(|u|), \qquad u \in \mathbb{R},$$

$$\text{where} \qquad K_\nu(t) = \frac{1}{2}\int_0^\infty s^{\nu-1}\exp\left\{-\frac{t}{2}\left(s + \frac{1}{s}\right)\right\} ds.$$

- Even if X is non-random, the exact log-likelihood is **highly intractable** ($\Delta_j \xi := \xi_{t_j} - \xi_{t_{j-1}}$):

$$\ell_n(\theta) = \sum_{j=1}^n \log\left(\sigma^{-1}\left(\frac{2^{1-\nu/2}}{\Gamma(\nu/2)}\right)^h \frac{1}{\pi}\int_0^\infty \cos\left(\frac{\Delta_j Y - \Delta_j X \cdot \mu}{\sigma}u\right) u^{\nu h/2}\left(K_{\nu/2}(u)\right)^h du\right)$$

  - The numerical integration $\int_0^\infty (\dots)du$ can be unstable, particularly for very small h.

$$Y_t = X_t \cdot \mu + \sigma J_t$$

$$\sigma^{-1}\left(Y_{t_j} - Y_{t_{j-1}} - (X_{t_j} - X_{t_{j-1}}) \cdot \mu\right) \overset{\theta}{=} J_{t_j} - J_{t_{j-1}} \sim J_h$$

Proposal for $\hat{\theta}_n = (\hat{\mu}_n, \hat{\sigma}_n, \hat{\nu}_n)$: Adaptive (two-stage) quasi-likelihood inference

1. Estimate (μ_0, σ_0) based on the **Cauchy quasi-likelihood**
2. Estimate ν_0 based on the **Student-t quasi-likelihood** through the "unit-time" residual sequence

$$\hat{\epsilon}_i := \hat{\sigma}_n^{-1}\left(Y_i - Y_{i-1} - \hat{\mu}_n \cdot (X_i - X_{i-1})\right), \qquad i = 1, \dots, [T_n],$$

Points in our study

- We can **asymptotically efficiently** estimate (μ_0, σ_0), with leaving ν_0 unknown.
- The two-stage procedure enables us to estimate ν_0 **as if we directly observe** $(J_i)_{i=1}^{[T_n]}$.
  - We will **thin the full data in the first stage** to obtain a "clean" result.

$$\{(X_{t_j}, Y_{t_j})\}_{j=0}^{[nT_n]} \quad \text{from} \quad \begin{cases} Y_t = X_t \cdot \mu + \sigma J_t \\ J_1 \sim t_\nu \end{cases} \quad \Rightarrow \quad \sqrt{??}\begin{pmatrix} \hat{\mu}_n - \mu_0 \\ \hat{\sigma}_n - \sigma_0 \end{pmatrix} = O_p(1)$$

- We have $h^{-1}J_h \xrightarrow{\mathcal{L}} t_1$ as $h \to 0$, whatever the degrees of freedom $\nu_0 > 0$ is:

$$\left(E_\theta\left[\exp(iuh^{-1}J_1)\right]\right)^h = (1+o(1))\left|u\nu^{-1/2}h^{-1}\right|^{\nu h/2} K_{\nu/2}\left(\left|uh^{-1}\right|\right)^h \to e^{-|u|}.$$

- We look at the **Cauchy quasi-(log-)likelihood** (QLF) in such a way:

$$Y_{t_j} \big| \{Y_{t_{j-1}} = y,\, X_{t_{j-1}},\, X_{t_j}\} \overset{\theta}{\approx} \text{Cauchy}\left(y + \mu\cdot(X_{t_j} - X_{t_{j-1}}),\, h\sigma\right)$$

  - We can make an inference for (μ, σ) without knowing the value ν_0;

Using the Cauchy QLF is not free:

- The information of ν_0 gets disappeared in the small-time limit.
- The Cauchy-approximation errors accumulate for $T_n \to \infty$.

- We use data on $[0, B_n]$, where

$$B_n \le T_n, \qquad B_n \to \infty, \qquad \exists \epsilon' \in (0,1),\ B_n \lesssim n^{1-\epsilon'}. \tag{2.1}$$

- Technically (Essentially?) we needed in the proof that $(\phi_1(y) := \pi^{-1}(1+y^2)^{-1})$

$$\sqrt{nB_n} \int |\zeta(y)|\, |f_h(y) - \phi_1(y)|\, dy \lesssim n^{r-1}\sqrt{nB_n} = \sqrt{n^{2r-1}B_n} \to 0,$$

for any $r > 0$ small enough and logarithmic-growth-order function ζ.

  - (2.1) restricts the speed of $T_n \to \infty$ as $T_n = O(n^{1-\epsilon'})$, when using the whole sample (i.e. $B_n = T_n$).

$$Y_t = X_t \cdot \mu + \sigma J_t$$

Assumption 2.1 (Covariate process X)

We can associate an $(\mathcal{F}_t)$-adapted process $t \mapsto X'_t$ having càdlàg sample paths s.t.:

1. *Some integrability conditions (omitted: [Masuda et al., 2023, Aasumption 2.1])*
2. *There exists a probability measure $\pi_X(dx)$ on the Borel space $(\mathbb{R}^q, \mathcal{B}^q)$ such that*

$$\forall f \in L^1(\pi_0), \quad \frac{1}{T}\int_0^T f(X'_t)dt \xrightarrow{p} \int f(x')\pi_0(dx'), \qquad T \to \infty.$$

3. *The integral $S_0 := \int x'^{\otimes 2}\pi_0(dx')$ is finite and positive-definite.*

- One may think of $X_t = \int_0^t X'_s ds$, where sample paths of X' may have certain discontinuity.

- Let $a := (\mu, \sigma)$, $a_0 := (\mu_0, \sigma_0)$, $N_n := [nB_n]$ with $T_n/N_n \to 0$, and $\Gamma_{a,0} := (2\sigma_0^2)^{-1}\mathrm{diag}(S_0, 1)$.
- **Cauchy quasi-MLE**: $\hat{a}_n := (\hat{\mu}_n, \hat{\sigma}_n) \in \mathrm{argmax}_{a \in \overline{\Theta_\mu \times \Theta_\sigma}} \mathbb{H}_{1,n}(a)$, where (apart from a constant)

$$\mathbb{H}_{1,n}(a) = -\sum_{j=1}^{N_n} \left\{ \log \sigma + \log \left(1 + \left(\frac{\Delta_j Y - \mu \cdot \Delta_j X}{h\sigma} \right)^2 \right) \right\}.$$

Theorem 2.2 (Cauchy quasi-likelihood estimation of (μ, σ))

1. *Asymptotic normality:* $\hat{u}_{a,n} := \sqrt{N_n}(\hat{a}_n - a_0) = \Gamma_{a,0}^{-1}\Delta_{a,n} + o_p(1) \xrightarrow{\mathcal{L}} N\left(0, \Gamma_{a,0}^{-1}\right)$
2. *Tail-probability estimate:* $\forall L > 0, \quad \sup_{r>0} \sup_n P\left[|\hat{u}_{a,n}| > r\right] r^L < \infty$

- Thanks to the LAN property [Ivanenko et al., 2015], we can conclude the **asymptotic efficiency**.

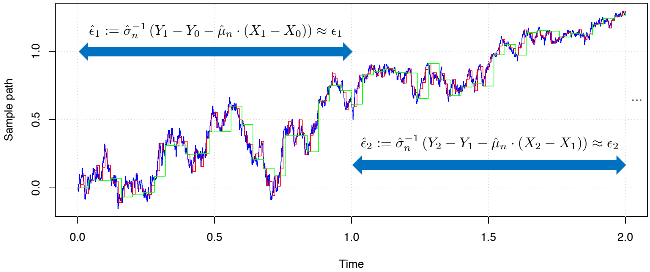
Estimation of the degrees of freedom ν from $\{(X_{t_j}, Y_{t_j})\}_{j=0}^{[nT_n]}$ under Assumption 2.1

$$\begin{cases} Y_t = X_t \cdot \mu + \sigma J_t \\ J_1 \sim t_\nu \end{cases} \Rightarrow \sqrt{N_n}\begin{pmatrix} \hat{\mu}_n - \mu_0 \\ \hat{\sigma}_n - \sigma_0 \end{pmatrix} \xrightarrow{\mathcal{L}} N_{q+1}(0, \Gamma_a^{-1}) \quad \Rightarrow \quad \sqrt{?}(\hat{\nu}_n - \nu_0) = O_p(1)$$

- Define the **unit-time** residual sequence ($i = 1, \dots, [T_n]$):

$$\hat{\epsilon}_i := \hat{\sigma}_n^{-1}\left(Y_i - Y_{i-1} - \hat{\mu}_n \cdot (X_i - X_{i-1})\right) \overset{\theta}{\approx} J_i - J_{i-1} =: \epsilon_i \sim \text{i.i.d. } t_\nu$$

- We make use $\hat{\epsilon}_i \approx \epsilon_i$ to estimate ν as if $\hat{\epsilon}_1, \dots, \hat{\epsilon}_{[T_n]}$ are observed t_ν-i.i.d. samples.



- We consider the quasi-likelihood function

$$\widehat{M}_n(\nu) = \sum_{i=1}^{[T_n]} \rho(\hat{\epsilon}_i; \nu),$$

where ($f(\cdot;\nu)$ is the noise t_ν density)

$$\rho(\epsilon;\nu) := \log f(\epsilon;\nu) = \log\left(\frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\pi}\Gamma(\frac{\nu}{2})} (1+x^2)^{-(\nu+1)/2} \right)$$

- We define the **Student quasi-MLE** of ν by any element $\hat{\nu}_n \in \mathrm{argmax}_{\nu \in \overline{\Theta}_\nu} \widehat{M}_n(\nu)$.
  - $\widehat{M}_n(\cdot)$ is a.s. convex.
  - Any other reasonable contrast function can be considered as well.

- We're left to investigate the asymptotic behavior of the random function $\widehat{M}_n(\cdot)$.

- Let $\Delta_{\nu,n} := \frac{1}{\sqrt{T_n}} \sum_{i=1}^{[T_n]} \partial_\nu \rho(\epsilon_i; \nu_0)$ and $\Gamma_\nu := \frac{1}{4}\left(\psi_1\left(\frac{\nu_0}{2}\right) - \psi_1\left(\frac{\nu_0+1}{2}\right)\right)$.
- Finer (High-frequency) sampling for estimating (μ, σ) makes the plugging-in effect ignorable:

Theorem 3.1 (Student-t quasi-likelihood estimation of ν)

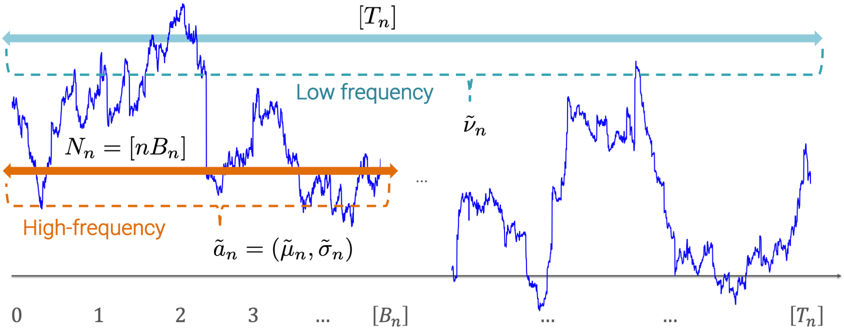
1. *Asymptotic normality:* $\hat{u}_{\nu,n} := \sqrt{T_n}(\hat{\nu}_n - \nu_0) = \Gamma_{\nu,0}^{-1}\Delta_{\nu,n} + o_p(1) \xrightarrow{\mathcal{L}} N\left(0, \Gamma_{\nu,0}^{-1}\right)$
2. *Tail-probability estimate:* $\forall L > 0, \quad \sup_{r>0} \sup_n P\left[|\hat{u}_{\nu,n}| > r\right] r^L < \infty$

- Question: **Joint** asymptotic distribution of $(\hat{u}_{a,n},\ \hat{u}_{\nu,n}) = \left(\sqrt{N_n}(\hat{a}_n - a_0),\ \sqrt{T_n}(\hat{\nu}_n - \nu_0)\right)$?

Joint asymptotic normality?

- We now assume $\frac{B_n}{T_n} \to 0$: we thin the data for the 1st-stage inference of $a = (\mu, \sigma)$, then

$$\left(\sqrt{N_n}(\hat{a}_n - a_0),\ \sqrt{T_n}(\hat{\nu}_n - \nu_0)\right) \xrightarrow{\mathcal{L}} N\left(0,\ \mathrm{diag}\{\Gamma_{a,0}^{-1},\ \Gamma_{\nu,0}^{-1}\}\right).$$



A modified setting: Student-Lévy driven Markovian SDE

- Discrete-time sample $(Y_{t_j})_{j=0}^{[nT_n]}$ for $t_j = jh = j/n$ ($h := 1/n$) and $T_n \to \infty$ from

$$dY_t = \mu \cdot b(Y_t)dt + \sigma dJ_t, \qquad Y_0 = 0.$$

  - $J = (J_t)$ is a Lévy process s.t. $J_1 \sim$ pdf $f(x;\nu) := \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\pi}\Gamma(\frac{\nu}{2})}(1+x^2)^{-(\nu+1)/2}$.
  - $\sup_{y\in\mathbb{R}}(|b(y)| \vee |\partial_y b(y)|) < \infty$ and $\limsup_{|y|\to\infty} \frac{\mu_0 \cdot b(y)}{y} < 0$.
- $\theta := (\mu, \sigma, \nu) =: (a, \nu) \in \Theta$, bounded convex domain s.t. $\overline{\Theta} \subset \mathbb{R}^q \times (0,\infty) \times (1,\infty)$.

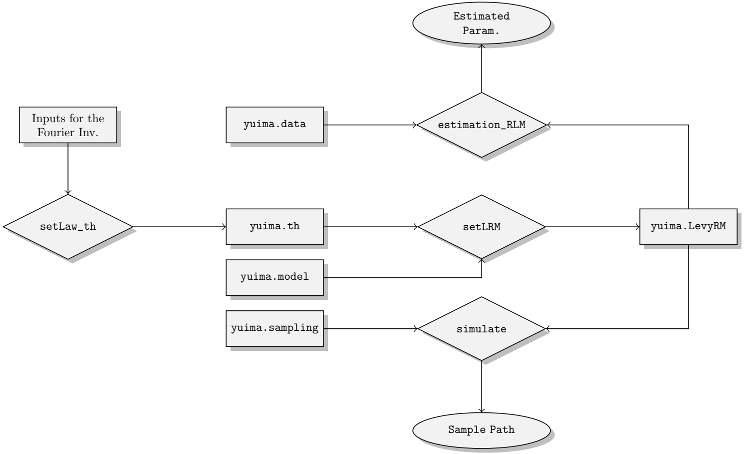
Theorem 3.2 (Joint asymptotic normality with orthogonality)

Under the above setup (including $B_n/T_n \to 0$), we have, with $\tilde{S}_n := N_n^{-1}\sum_{j=1}^{N_n} b_{j-1}^{\otimes 2}$,

$$\left(\mathrm{diag}\left(\frac{1}{2\tilde{\sigma}_n^2}\tilde{S}_n,\ \frac{1}{2\tilde{\sigma}_n^2}\right)^{1/2}\sqrt{N_n}(\tilde{a}_n - a_0),\ \left\{\frac{1}{4}\left(\psi_1\left(\frac{\tilde{\nu}_n}{2}\right) - \psi_1\left(\frac{\tilde{\nu}_n+1}{2}\right)\right)\right\}^{1/2}\sqrt{T_n}(\tilde{\nu}_n - \nu_0)\right) \xrightarrow{\mathcal{L}} N_{q+2}(0, I_{q+2}).$$

New classes/methods in yuima package [Brouste et al., 2014]

- New classes/methods in yuima package [Brouste et al., 2014] have been composed.



Case 1: Periodic deterministic regressors

- $Y_t = (\mu_1 \cos 5t + \mu_2 \sin t) + \sigma J_t$, $J_1 \sim t_3$; $(\mu_{1,0}, \mu_{2,0}, \sigma_0, \nu_0) = (5, -1, 3, 3)$
- $T_n = 50$, $B_n = 15$ $(T_n/N_n = 1/15)$, $h = 0.02$; $N_n = B_n/h = 750$; #MC $= 1000$

	Laguerre Method			COS Method			FFT Method		
h	0.01	0.005	1/365	0.01	0.005	1/365	0.01	0.005	1/365
$\hat{\mu}_1$	4.968	5.019	4.976	4.939	5.065	4.875	4.934	5.069	4.876
	(0.029)	(0.020)	(0.014)	(0.041)	(0.049)	(0.054)	(0.042)	(0.049)	(0.054)
$\hat{\mu}_2$	-1.065	-1.044	-0.914	-1.192	-1.195	-0.555	-1.179	-1.178	-0.528
	(0.146)	(0.104)	(0.074)	(0.210)	(0.247)	(0.277)	(0.213)	(0.248)	(0.277)
$\hat{\sigma}_0$	2.770	2.794	2.657	3.993	6.654	10.010	4.055	6.674	10.010
	(0.101)	(0.072)	(0.051)	(0.145)	(0.172)	(0.192)	(0.148)	(0.172)	(0.193)
$\hat{\nu}$	3.462	3.281	3.067	2.827	2.223	2.227	5.749	8.310	15.175
	(0.615)	(0.580)	(0.538)	(0.492)	(0.377)	(0.378)	(1.063)	(1.571)	(2.940)
sec.	2.07	2.02	2.07	1.10	1.24	1.34	0.15	0.24	0.39

Table: Estimated parameters for $h = 0.01, 0.005, 1/365$ and different integration methods. The number of points for the inversion of the characteristic function is $N = 180$. The parenthesis shows the asymptotic standard error. The last row reports the seconds necessary for the simulation of a sample path.

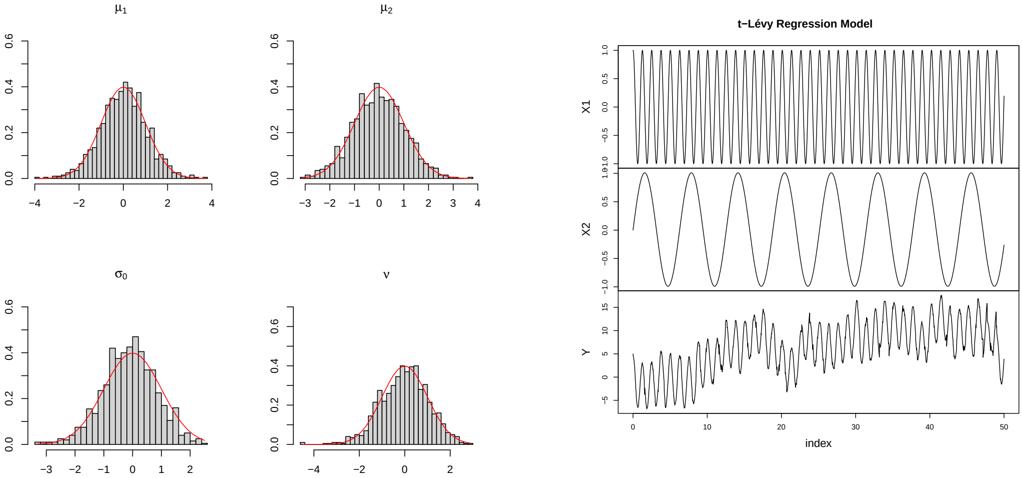


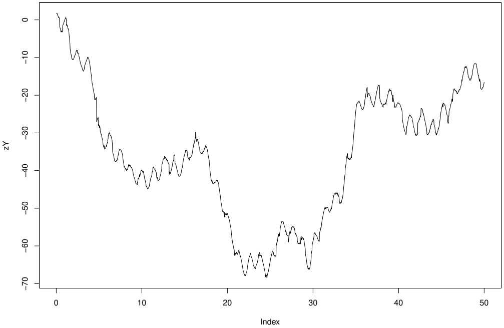
Figure: Studentized histograms ($\mu_1 = 5$, $\mu_2 = -1$, $\sigma_0 = 3$, $\nu = 2$) and path plots.

Case 2: Integrated Gaussian-OU regressor

$$Y_t = (\mu_1 \cos 5t + \mu_2 X_t) + \sigma J_t, \quad J_1 \sim t_3,$$

$$X_t = \int_0^t \int_0^s \exp\left(-\frac{1}{2}(s-u)\right) dw_u ds$$

- $T_n = 50, \quad B_n = 10, \quad h = 0.02,$
  $N_n = B_n/h = 500; \quad \#\mathrm{MC} = 1000$
    - $T_n/N_n = 1/10$
- $(\mu_{1,0}, \mu_{2,0}, \sigma_0, \nu_0) = (2, 5, 3, 3);$
  $\Theta_{\mu,\sigma} = [0.01, 10] \times [-10, 10] \times [0.1, 5]$



	$\hat{\mu}_{1,n}$	$\hat{\mu}_{2,n}$	$\hat{\sigma}_n$	$\hat{\eta}_n$
mean	2.00	5.00	2.67	2.73
sd	0.04	0.23	0.15	0.48

Summary and agenda

$$\{(X_{t_j}, Y_{t_j})\}_{j=0}^{[nT_n]} \text{ from } \begin{cases} Y_t = X_t \cdot \mu + \sigma J_t \\ J_1 \sim t_\nu \end{cases} \overset{\text{Cauchy QMLE}}{\Longrightarrow} \sqrt{N_n}\begin{pmatrix} \hat{\mu}_n - \mu_0 \\ \hat{\sigma}_n - \sigma_0 \end{pmatrix} \xrightarrow{\mathcal{L}} N_{q+1}(0, \Gamma_a^{-1})$$

$$\overset{\text{Student-}t\text{ QMLE}}{\Longrightarrow} \sqrt{T_n}(\hat{\nu}_n - \nu_0) \xrightarrow{\mathcal{L}} N_1(0, \Gamma_\nu^{-1})$$

- $N_n := nB_n$ with $B_n \le T_n$, $B_n \to \infty$, $\exists\epsilon' \in (0,1)$, $B_n \lesssim n^{1-\epsilon'}$, and $T_n \to \infty$, with $T_n/N_n \to 0$.

1. We do **not** know the asymptotic phenomenon in estimating ν from full data $\{(X_{t_j}, Y_{t_j})\}_{j=0}^{[nT_n]}$.
2. Implemented in the `yuima` package in CRAN, managing the **numerical** issues.

More general and systematic noise inference scheme through the β-**stable quasi-likelihood**:

$$dX_t = a(X_t, \alpha)dt + c(X_{t-}, \gamma)dJ_t^\beta, \qquad h^{-1/\beta} J_h^\beta \xrightarrow{\mathcal{L}} S_\beta \quad (h \to 0)$$

$$(X_{t_0}, X_{t_1}, \ldots, X_{t_{[nT_n]}}): \ t_j = jh; \ \ h = 1/n \to 0; \ \ T_n \to \infty.$$

References

Brouste, A., Fukasawa, M., Hino, H., Iacus, S. M., Kamatani, K., Koike, Y., Masuda, H., Nomura, R., Ogihara, T., Shimizu, Y., Uchida, M., and Yoshida, N. (2014).
The yuima project: A computational framework for simulation and inference of stochastic differential equations.
Journal of Statistical Software, 57(4):1–51.

Ivanenko, D., Kulik, A. M., and Masuda, H. (2015).
Uniform LAN property of locally stable Lévy process observed at high frequency.
ALEA Lat. Am. J. Probab. Math. Stat., 12(2):835–862.

Massing, T. P. G. (2019).
Stochastic Properties of Student-Lévy Processes with Applications.
PhD thesis, Universität Duisburg-Essen.

Masuda, H., Mercuri, L., and Uehara, Y. (2023).
Quasi-likelihood analysis for Student-Lévy regression.
arXiv:2306.16790.