Inference for a class of non-ergodic non-Gaussian regression

Hiroki Masuda

Kyushu University

The LiU Seminar Series in Statistics and Mathematical Statistics
February 2, 2021

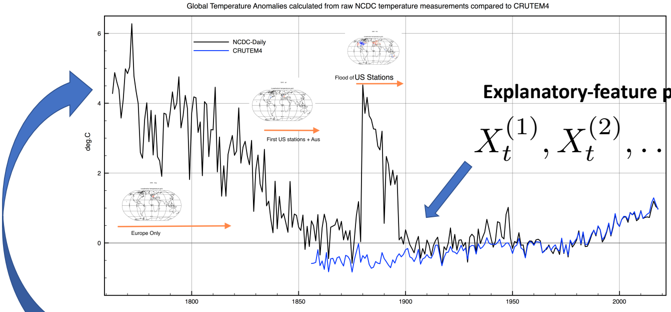
Outline

1. Introduction
2. Locally stable regression: Joint inference
3. Optimal stable Ornstein-Uhlenbeck regression (optional)

1. Introduction
2. Locally stable regression: Joint inference
3. Optimal stable Ornstein-Uhlenbeck regression (optional)

Put simply

- Inference for non-ergodic process regression models:
  - **Trend** and **Scale** structures
  - Activity index of **non-Gaussian noise**



Explanatory-feature process(es)

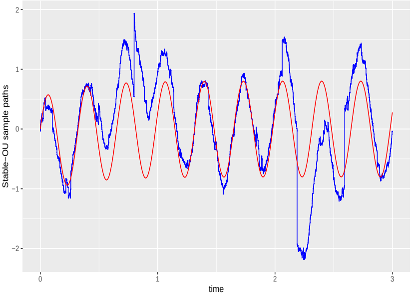
$$X_t^{(1)}, X_t^{(2)}, \ldots, X_t^{(M)}$$

Figure: ©clivebest.com

$$Y_{t+dt} \leftarrow Y_t + (\textbf{Trend})dt + (\textbf{Scale}) \cdot d(\textbf{Noise})_t$$

- Non-ergodic dynamic regression $\{(X_{t_j}, Y_{t_j})\}_{j=0}^{n}$ ($t_j := jh$, $h := T/n$):

$$Y_{t_j} \leftarrow Y_{t_{j-1}} + A(X_{t_{j-1}}, Y_{t_{j-1}})h + C(X_{t_{j-1}}, Y_{t_{j-1}})\Delta_j J$$

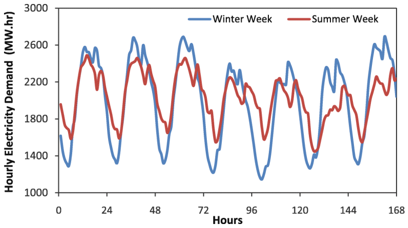


Probabilistic (path-)structure of **noise increments**

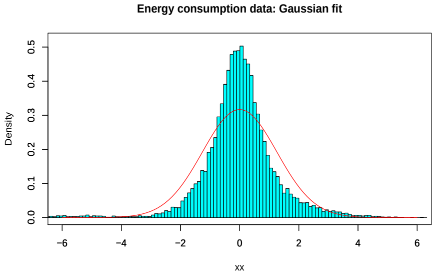
- Quantitatively manages **trend and scale** estimation precisions
- Brings about pros & cons of non-Gaussianity

Potential applications (ongoing)

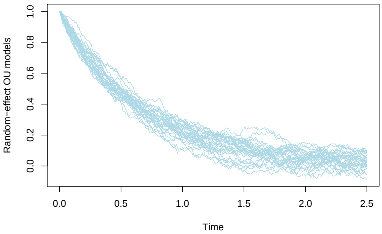
- Electricity-consumption prediction
  - Seasonal/Periodic nature, often time-inhomogeneous and/or non-ergodic
  - (Multivariate) Ornstein-Uhlenbeck regression



©https://doi.org/10.1186/s13705-019-0224-1



- Population-dynamics (mixed-effects) modeling
  - Non-Gaussian stable OU modeling for foraging brain dynamics.
  - PK/PD mixed-effects kinetics [Lavielle, 2015]
  - Extensions to large interacting system, mean-field model, ...



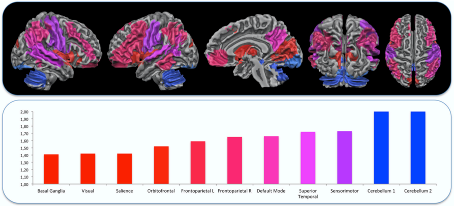


Fig 5. The graph represents the values of α for different networks from the lowest (red) to the highest. On the top of the graphs the corresponding brain areas are shown.

doi:10.1371/journal.pone.0161702.g005

What's known: location-scale inference for ergodic Lévy SDE

$$Y_t = Y_0 + \int_0^t a(Y_s;\alpha,\gamma)ds + \int_0^t c(Y_{s-};\gamma)dJ_s,$$
$$(Y_{t_j})_{j=0}^n,\ \ t_j = jh,\ \ T_n := nh \to \infty,\ \ nh^2 \to 0.$$

- **Gaussian** quasi-likelihood for $\theta = (\alpha,\gamma)$:

$$\mathcal{L}\left(Y_{t_j}\,\middle|\,Y_{t_{j-1}} = y\right) \overset{\theta}{\approx} N\left(y + a(y;\alpha,\gamma)h,\ c^2(y;\gamma)h\right)$$

  leads to $\sqrt{T_n}$-A.N. estimator, with convergence of moments.

- Stepwise-proceeding tools available:
  - **Estimate** first γ, then α. [Uehara and Masuda, 2017];
  - **Select** first $c(y;\gamma)$, then $a(y;\alpha,\gamma)$. [Eguchi and Masuda, 2018], [Eguchi and Masuda, 2019], and [Eguchi and Uehara, 2019].

Case of non-ergodic locally β-stable driven SDE?

1 Introduction

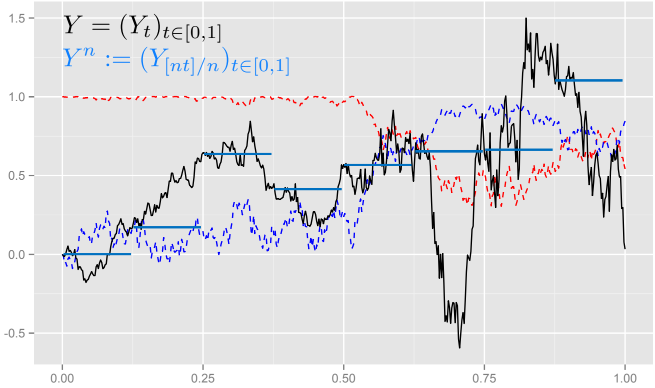
2 Locally stable regression: Joint inference

3 Optimal stable Ornstein-Uhlenbeck regression (optional)

Non-ergodic locally stable regression model (semiparametric)

$$Y_t = Y_0 + \int_0^t a(X_s, Y_s; \alpha) ds + \int_0^t c(X_{s-}, Y_{s-}; \gamma) dJ_s, \quad t \in [0, T],$$

$$\{(X_{t_j}, Y_{t_j})\}_{j=0}^n, \ \ t_j = jh, \ \ h = h_n = T/n \to 0 \ (n \to \infty).$$



$$Y_t = Y_0 + \int_0^t a(X_s, Y_s; \alpha) ds + \int_0^t c(X_{s-}, Y_{s-}; \gamma) dJ_s, \quad t \in [0, T],$$
$$\{(X_{t_j}, Y_{t_j})\}_{j=0}^n, \ t_j = jh, \ h = h_n = T/n \to 0 \ (n \to \infty).$$

1. Smooth, non-degenerate **coefficients** $a(x, y; \alpha)$, $c(x, y; \gamma)$
2. Locally stable **noise** $J = J^\beta$: $\mathcal{L}\big(h^{-1/\beta} J_h\big) \Rightarrow S_\beta$ as $h \to 0$.
   - $\mathbb{E}(e^{iuS_\beta}) = e^{-|u|^\beta} \sim$ pdf $\phi_\beta(y)$
   - (General tempered) Stable, Normal tempered stable, Generalized hyperbolic, Student-t, Meixner, ...
3. **Parameter** $\theta = (\alpha, \beta, \gamma) \in \Theta_\alpha \times \Theta_\beta \times \Theta_\gamma = \Theta$:
   - $\Theta_\alpha \subset \mathbb{R}^{p_\alpha}$, $\overline{\Theta}_\beta \subset [1, 2)$, $\Theta_\gamma \subset \mathbb{R}^{p_\gamma}$
   - $\exists$True $\theta_0 = (\alpha_0, \beta_0, \gamma_0) \in \Theta$ (bounded convex domain)

Regularity conditions

$$Y_t = Y_0 + \int_0^t a(X_s, Y_s; \alpha)ds + \int_0^t c(X_{s-}, Y_{s-}; \gamma)dJ_s, \quad t \in [0, T].$$

1. (Coefficients and covariate process)
   1. (a, c) is smooth in (x, y, θ), and Lipschitz at $\theta = \theta_0$;
   2. $0 < \inf_{(x,y;\gamma)} c(x, y; \gamma) \leq \sup_{(x,y;\gamma)} c(x, y; \gamma) < \infty$;
   3. $t \mapsto X_t \in \mathbb{R}^q$ is a **"good" Itô semimartingale with jumps**, of the form

$$X_t = X_0 + \int_0^t a'_s ds + \int_0^t b'_s dw'_s + \int_0^t c'_{s-} dJ'_s.$$

2. (Identifiability and non-degeneracy)
   1. $t \mapsto (a(X_t, Y_t; \alpha, \gamma), c(X_t, Y_t; \gamma))$ and $t \mapsto (a(X_t, Y_t; \alpha_0, \gamma_0), c(X_t, Y_t; \gamma_0))$ on $[0, T]$ a.s. coincide if and only if $\theta = \theta_0$.
   2. $(x, y) \mapsto \partial_\gamma \log c(x, y; \gamma)$ is not a constant function for each γ.
3. $\mathcal{L}(J_1)$ is symmetric, and $\exists \epsilon' > 0 \,\forall$ m'ble. f s.t. $\exists K \geq 0,\ |f(y)| \lesssim 1 + \log^K(1 + |y|)$,

$$\sqrt{n} \left| \mathbb{E}\{f(h^{-1/\beta} J_h)\} - \int f(y)\phi_\beta(y)dy \right| = o(n^{-\epsilon'})$$

   - cf. [Clément and Gloter, 2020], [Kulik, 2019] for sufficient conditions.

- Euler-Maruyama ($\Delta_j \xi := \xi_{t_j} - \xi_{t_{j-1}}$ and $f_{j-1}(\theta) := f(X_{t_{j-1}}, Y_{t_{j-1}}; \theta)$)

$$Y_{t_j} = Y_{t_{j-1}} + \int_{t_{j-1}}^{t_j} a(X_s, Y_s; \alpha) ds + \int_{t_{j-1}}^{t_j} c(X_{s-}, Y_{s-}; \gamma) dJ_s$$
$$\overset{\mathbb{P}_\theta}{\approx} Y_{t_{j-1}} + a_{j-1}(\alpha)h + c_{j-1}(\gamma) h^{1/\beta} \cdot \underbrace{h^{-1/\beta}\Delta_j J}_{\Rightarrow S_\beta \text{ as } h \to 0.}$$

- **Conditional** stable quasi-likelihood:

$$\mathbb{H}_n(\alpha, \beta, \gamma) := \sum_{j=1}^n \log \left\{ \frac{1}{h_n^{1/\beta} c_{j-1}(\gamma)} \phi_\beta \left(\frac{\Delta_j Y - a_{j-1}(\alpha)h}{h^{1/\beta} c_{j-1}(\gamma)} \right) \right\}$$

Goal: Model descriptive location-scale model selection

- Selection consistency among $\{a_l\}_{l \le L}$ and $\{c_m\}_{m \le M}$ for $n \to \infty$
- Good (feasible) estimator

Marginal stable quasi-likelihood: a quasi model evidence

$$Y_{t_j} \overset{\mathbb{P}_\theta}{=} Y_{t_{j-1}} + \int_{t_{j-1}}^{t_j} a(X_s, Y_s; \alpha)ds + \int_{t_{j-1}}^{t_j} c(X_{s-}, Y_{s-}; \gamma)dJ_s^\beta$$

For $\theta' := (\alpha, \gamma)$ and a **prior distribution** $\mathfrak{p}(\theta')d\theta'$ on $\Theta' := \Theta_\alpha \times \Theta_\gamma$,

$$\mathfrak{L}_n := \log \left(\int_{\Theta'} \exp\{\mathbb{H}_n(\alpha, \beta_0, \gamma)\}\mathfrak{p}(\theta')d\theta' \right).$$

- $\mathfrak{p}(\theta'_0) > 0$, and continuous at θ'_0

- $D_n = D_n(\beta_0) := \mathrm{diag}\left(\sqrt{n}h^{1-1/\beta_0}I_{p_\alpha}, \sqrt{n}I_{p_\gamma}\right)$: Diagonal-rate matrix
- $\Delta_n(\theta_0) := D_n^{-1}\partial_{\theta'}\mathbb{H}_n(\theta_0)$: Quasi score
- $\left(\Delta_n(\theta_0), -D_n^{-1}\partial_{\theta'}^2\mathbb{H}_n(\theta'_0)D_n^{-1}\right) \xrightarrow{\mathcal{L}} \left(\Gamma_0^{1/2}\eta, \Gamma_0\right)$, $\eta \sim N(0, I_{p_\alpha+p_\gamma}) \perp\!\!\!\perp \mathcal{F}$

Stochastic expansion of non-ergodic model evidence

$$Y_{t_j} = Y_{t_{j-1}} + \int_{t_{j-1}}^{t_j} a(X_s, Y_s; \alpha) ds + \int_{t_{j-1}}^{t_j} c(X_{s-}, Y_{s-}; \gamma) dJ_s$$

$$\mathfrak{L}_n = \log \left(\int_{\Theta'} \exp\{\mathbb{H}_n(\alpha, \beta_0, \gamma)\} \mathfrak{p}(\theta') d\theta' \right)$$

Theorem 2.1 (Stochastic expansion of $\mathfrak{L}_n$)

$$\begin{aligned} -2\mathfrak{L}_n &= -2\mathbb{H}_n(\theta_0) - 2\log|D_n^{-1}| - 2\log \mathfrak{p}(\theta_0) - (p_\alpha + p_\gamma)\log(2\pi) \\ &\quad + \log|\Gamma_0| - \Gamma_0^{-1}\left[\Delta_n(\theta_0)^{\otimes 2}\right] + o_p(1) \\ &= -2\mathbb{H}_n(\theta_0) + p_\alpha \log\left(nh^{2(1-1/\beta_0)}\right) + p_\gamma \log n + O_p(1) \end{aligned}$$

- $-2\mathfrak{L}_n = -2\mathbb{H}_n(\hat{\theta}_n) + p_\alpha \log\left(nh^{2(1-1/\hat{\beta}_n)}\right) + p_\gamma \log n + O_p(1)$?
- $\left(\sqrt{n}h^{1-1/\beta_0}(\hat{\alpha}_n - \alpha_0), \sqrt{n}\log(1/h)\beta_0^{-2}(\hat{\beta}_n - \beta_0), \sqrt{n}(\hat{\gamma}_n - \gamma_0)\right) = O_p(1)$

Stable quasi-MLE (SQMLE)

Theorem 2.2 (Asymptotic mixed normality (AMN) of SQMLE)

∃Local maxima $\hat{\theta}_n = (\hat{\alpha}_n, \hat{\beta}_n, \hat{\gamma}_n) \in \Theta$ *of* $\mathbb{H}_n$*, for which*

$$\left(\sqrt{n}h^{1-1/\beta_0}(\hat{\alpha}_n - \alpha_0),\ \sqrt{n}\log(1/h)\beta_0^{-2}(\hat{\beta}_n - \beta_0),\ \sqrt{n}(\hat{\gamma}_n - \gamma_0)\right)$$

$$\xrightarrow{\mathcal{L}} \textit{Mixed Normal}_{p_\alpha+1+p_\gamma}\left(0,\ \mathrm{diag}\left(\mathcal{I}_{T,\alpha}(\theta_0)^{-1},\ \mathcal{I}_{T,(\beta,\gamma)}(\theta_0)^{-1}\right)\right).$$

$$\mathcal{I}_{T,\alpha}(\theta_0) := C_1(\beta_0)\frac{1}{T}\int_0^T \frac{\{\partial_\alpha a(X_t, Y_t; \alpha_0)\}^{\otimes 2}}{c(X_t, Y_t; \gamma_0)^2} dt,$$

$$\mathcal{I}_{T,(\beta,\gamma)}(\theta_0) := C_2(\beta_0)\frac{1}{T}\int_0^T \begin{pmatrix} 1 & \text{sym.} \\ \partial_\gamma \log c(X_t, Y_t; \gamma_0) & \{\partial_\gamma \log c(X_t, Y_t; \gamma_0)\}^{\otimes 2} \end{pmatrix} dt,$$

$$C_1(\beta) := \int \frac{\{\partial \phi_\beta(y)\}^2}{\phi_\beta(y)} dy, \quad C_2(\beta) := \int \frac{\{\phi_\beta(y) + y\partial\phi_\beta(y)\}^2}{\phi_\beta(y)} dy.$$

- AMN does hold even when $\overline{\Theta}_\beta \subset (2/3,\ 2)$.
- One-step improvement is possible for stable OU regression (M. arXiv:20200608)

Remarks on estimation

- An approximate confidence region can be constructed:

$$\left| \left(\mathcal{I}_{c,T}(\hat{\theta}_n)^{1/2} \sqrt{n} h^{1-1/\hat{\beta}_n} (\hat{\alpha}_n - \alpha_0), \right.\right.$$
$$\left.\left. \mathcal{I}_{a,T}(\hat{\theta}_n)^{1/2} \begin{pmatrix} \sqrt{n} \log(1/h) \hat{\beta}_n^{-2} (\hat{\beta}_n - \beta_0) \\ \sqrt{n}(\hat{\gamma}_n - \gamma_0) \end{pmatrix} \right) \right|^2 \xrightarrow{\mathcal{L}} \chi^2(p).$$

- [Clément and Gloter, 2020] for joint estimation with $\beta \in (0, 2)$ via trajectory-fitting in Markovian case.
- *In some cases*, stepwise manner is expected to give no loss of asymptotic efficiency with maintaining asymptotic orthogonality between $(\hat{\beta}_n, \hat{\gamma}_n)$ and $\hat{\alpha}_n$.

Location-scale model selection

$$Y_t = Y_0 + \int_0^t a(X_s, Y_s; \alpha) ds + \int_0^t c(X_{s-}, Y_{s-}; \gamma) dJ_s, \quad t \in [0, T]$$

- Model selection **for location and scale structures**?
  - Objective: most parsimonious correct $a(x, y; \alpha)$ and $c(x, y; \gamma)$
  - Driving locally β-stable J is a *nuisance* element.
    - The index β here is *not* allowed to be time-varying.
- e.g. **Nested** coefficients (with c'_k, c''_l bounded) such as:

$$c(x, y; \gamma) = \exp\left(\sum_{k=1}^{p_\gamma} \gamma'_k c'_k(y) + \sum_{l=1}^{q_\gamma} \gamma''_l c''_l(x) \right)$$

$$a(x, y; \alpha) = \sum_{k=1}^{p_\alpha} \alpha'_k a'_k(y) + \sum_{l=1}^{q_\alpha} \alpha''_l a''_l(x)$$

or, with possible interaction between x and y.

- **Doubly indexed and correctly specified** candidate coefficients:

$$Y_t = Y_0 + \int_0^t a_l(X_s, Y_s; \alpha_l)ds + \int_0^t c_m(X_{s-}, Y_{s-}; \gamma_m)dJ_s$$

for $l = 1, \dots, L$ and $m = 1, \dots, M$, with *common* unknown $\beta_0 \geq 1$.

- Optimal model indices $(k_\star, l_\star)$, assumed to be a.s. unique.

Theorem 2.3 (Selection consistency of LS-QBIC)

$$\mathcal{S}_{l,m;n} := -2\mathbb{H}_{l,m;n}(\hat{\theta}_{l,m;n}) + p_{\alpha_l} \log\left(nh^{2(1-1/\hat{\beta}_{l,m;n})}\right) + p_{\gamma_m} \log n$$

If $l \neq l_\star$ or $m \neq m_\star$, and if "$\mathcal{M}_{l_\star,m_\star}$ is nested in (l,m)th model", then

$$\mathbb{P}\left(\mathcal{S}_{l,m;n} > \mathcal{S}_{l_\star,m_\star;n}\right) \to 1.$$

- Can be proved as in [Eguchi and Masuda, 2018, Sect 5].
- $\sqrt{n}\log(1/h)$-rate of $\hat{\beta}_n$ used, probably $\sqrt{n}$-one insufficient.

Locally stable regression: Summary and agenda

Locally stable regression, without stability assumption

$$Y_t = Y_0 + \int_0^t a(X_s, Y_s; \alpha) ds + \int_0^t c(X_{s-}, Y_{s-}; \gamma) dJ_s, \quad t \in [0, T],$$

$$\{(X_{t_j}, Y_{t_j})\}_{j=0}^n, \ t_j = jh, \ h = h_n = T/n \to 0 \ (n \to \infty).$$

- **Regression** framework, extending [Masuda, 2019] and [Clément and Gloter, 2020]; no unit-root type problem.
- **Joint** SQMLE asymptotics: estimation and selection
- **Stepwise/Adaptive** counterparts, (partially) possible;
  - Second-order increments $\Delta_j^2 Y := \Delta_j Y - \Delta_{j-1} Y$ effective.
- Agenda: YUIMA R package and GUI implementation; Optimization theory/algorithm (SGLD, est.)

References I

Brouste, A. and Masuda, H. (2018).
Efficient estimation of stable Lévy process with symmetric jumps.
Stat. Inference Stoch. Process., 21(2):289–307.

Clément, E. and Gloter, A. (2020).
Joint estimation for SDE driven by locally stable Lévy processes.
Electron. J. Stat., 14(2):2922–2956.

Eguchi, S. and Masuda, H. (2018).
Schwarz type model comparison for LAQ models.
Bernoulli, 24(3):2278–2327.

Eguchi, S. and Masuda, H. (2019).
Data driven time scale in Gaussian quasi-likelihood inference.
Stat. Inference Stoch. Process., 22(3):383–430.

Eguchi, S. and Uehara, Y. (2019).
Schwartz type model selection for ergodic stochastic differential equation models.
arXiv preprint arXiv:1904.12398.

Kulik, A. (2019).
Approximation in law of locally α-stable Lévy-type processes by non-linear regressions.
Electron. J. Probab., 24:Paper No. 83, 45.

References II

Lavielle, M. (2015).
Mixed effects models for the population approach.
Chapman & Hall/CRC Biostatistics Series. CRC Press, Boca Raton, FL.
Models, tasks, methods and tools, With contributions by Kevin Bleakley.

Masuda, H. (2019).
Non-Gaussian quasi-likelihood estimation of SDE driven by locally stable Lévy process.
Stochastic Process. Appl., 129(3):1013–1059.

Uehara, Y. and Masuda, H. (2017).
Stepwise estimation of a Lévy driven stochastic differential equation.
Proc. Inst. Statist. Math. (Japanese), 65(1):21–38.

1 Introduction

2 Locally stable regression: Joint inference

3 Optimal stable Ornstein-Uhlenbeck regression (optional)

- Preprint arXiv:2006.04630

Objective, formally (arXiv:2006.04630)

- **Stable Ornstein-Uhlenbeck (OU) regression model**

$$dY_t = (\boldsymbol{\mu} \cdot \boldsymbol{X}_t - \lambda Y_t)dt + \sigma dJ_t$$

  - β-stable Lévy noise: $\mathbb{E}(e^{iuJ_t}) = \exp(-t|u|^\beta)$; $\quad h^{-1/\beta} J_h \stackrel{d}{=} J_1$
  - Non-random continuous covariate process $\boldsymbol{X}$ in $\mathbb{R}^q$ *
  - Unknown parameter: $\theta := (\lambda, \boldsymbol{\mu}, \beta, \sigma) \in \mathbb{R} \times \mathbb{R}^q \times (0,2) \times (0,\infty)$
  - Available data: $\{(\boldsymbol{X}_t)_{t\in[0,1]}, (Y_{t_j})_{j=0}^n\}$, $t_j = jh$, $h = h_n = 1/n \to 0$

- **Goal** ($^\exists$true $\theta_0 = (\lambda_0, \boldsymbol{\mu}_0, \beta_0, \sigma_0)$):
  - *Asymptotic mixed normality (AMN)* of MLE
  - *Asymptotic efficiency* of MLE

partly extending the previous study [Brouste and Masuda, 2018].

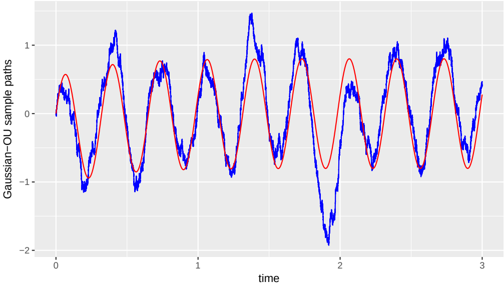
*Could be partially weakened/modified.

Remarks on Wiener driven case ($\beta = 2$)

- Sample $(Y_{t_j})_{j=0}^n$ from $dY_t = (\mu - \lambda Y_t)dt + \sigma dw_t$
  - (λ, σ) jointly estimable *only when* $nh_n \to \infty$
  - *Essentially different* local-likelihood asymptotics w.r.t. λ:

$\lambda > 0$ (Ergodic)	LAN (Local asymptotic normality)
$\lambda < 0$ (Non-ergodic)	LAMN
$\lambda = 0$ (Unit-root)	Locally asymptotically Brownian functional

- Gaussian high variability in small time "masks" trend information.



Explicit log-likelihood function

$$dY_t = (\boldsymbol{\mu} \cdot \boldsymbol{X}_t - \lambda Y_t)dt + \sigma dJ_t$$

- càdlàg solution: for each $t > s$,

$$Y_t \stackrel{\mathbb{P}_\theta}{=} e^{-\lambda(t-s)} Y_s + \boldsymbol{\mu} \cdot \int_s^t e^{-\lambda(t-s)} \boldsymbol{X}_s ds + \sigma \int_s^t e^{-\lambda(t-s)} dJ_s$$

- By the *scaling property* of J,

$$\ell_n(\theta) = \ell_n(\lambda, \boldsymbol{\mu}, \beta, \sigma) = \sum_{j=1}^n \log \left(\frac{1}{\sigma h^{1/\beta} \eta(\lambda\beta h)^{1/\beta}} \phi_\beta \left(\epsilon_j(\theta)\right) \right)$$

  - $\epsilon_j(\theta) := \dfrac{Y_{t_j} - e^{-\lambda h} Y_{t_{j-1}} - \boldsymbol{\mu} h \cdot \zeta_j(\lambda)}{\sigma h^{1/\beta} \eta(\lambda\beta h)^{1/\beta}} \stackrel{\mathbb{P}_\theta}{\sim}$ i.i.d. $\mathcal{L}(J_1)$
  - $\eta(x) := \dfrac{1}{x}(1 - e^{-x})$, $\quad \zeta_j(\lambda) := \dfrac{1}{h} \displaystyle\int_{t_{j-1}}^{t_j} e^{-\lambda(t_j - s)} \boldsymbol{X}_s ds$

Important ingredients

- Non-diagonal and non-symmetric norming matrix

$$\varphi_n = \varphi_n(\theta_0) := \mathrm{diag}\left\{ \frac{1}{\sqrt{n}h^{1-1/\beta}} I_{1+q},\ \frac{1}{\sqrt{n}} \begin{pmatrix} \varphi_{11,n}(\theta_0) & \varphi_{12,n}(\theta_0) \\ \varphi_{21,n}(\theta_0) & \varphi_{22,n}(\theta_0) \end{pmatrix} \right\}$$

- $$\begin{cases} \beta^{-2}\log(1/h)\varphi_{11,n} + \sigma^{-1}\varphi_{21,n} \to \overline{\varphi}_{21}; \\ \beta^{-2}\log(1/h)\varphi_{12,n} + \sigma^{-1}\varphi_{22,n} \to \overline{\varphi}_{22}; \\ \varphi_{11,n} \to \overline{\varphi}_{11}; \qquad \varphi_{12,n} \to \overline{\varphi}_{12}; \qquad |\overline{\varphi}_{11}\overline{\varphi}_{22} - \overline{\varphi}_{12}\overline{\varphi}_{21}| > 0 \end{cases}$$

- Non-degenerate Fisher information matrix ($\lambda \in \mathbb{R}$ implicit here)

$$\begin{aligned} \mathcal{I}(\theta_0) &= \mathcal{I}(\theta_0; \overline{\varphi}) = \mathcal{I}(\lambda_0, \boldsymbol{\mu}_0, \beta_0, \sigma_0; \overline{\varphi}) \\ &:= \mathrm{diag}\left\{ \frac{1}{\sigma_0^2}\mathbb{E}\{g(\epsilon)^2\} \int_0^1 \begin{pmatrix} Y_t^2 & -Y_t \boldsymbol{X}_t^\top \\ -Y_t \boldsymbol{X}_t & \boldsymbol{X}_t^{\otimes 2} \end{pmatrix} dt, \right. \\ &\left. \begin{pmatrix} \overline{\varphi}_{11} & \overline{\varphi}_{12} \\ -\overline{\varphi}_{21} & -\overline{\varphi}_{22} \end{pmatrix}^\top \begin{pmatrix} \mathbb{E}\{f(\epsilon)^2\} & \mathbb{E}\{\epsilon f(\epsilon) g(\epsilon)\} \\ \mathbb{E}\{\epsilon f(\epsilon) g(\epsilon)\} & \mathbb{E}\{(1+\epsilon g(\epsilon))^2\} \end{pmatrix} \begin{pmatrix} \overline{\varphi}_{11} & \overline{\varphi}_{12} \\ -\overline{\varphi}_{21} & -\overline{\varphi}_{22} \end{pmatrix} \right\} \end{aligned}$$

- $\epsilon \sim \mathcal{L}(J_1)$ under $\mathbb{P}_{\theta_0}$, $f(y) := \left.\frac{\partial_\beta \phi_\beta}{\phi_\beta}(y)\right|_{\beta=\beta_0}$, $g(y) := \left.\frac{\partial \phi_\beta}{\phi_\beta}(y)\right|_{\beta=\beta_0}$

Local structure of likelihood

Theorem 3.1

$$\ell_n\left(\theta_0+\varphi_n(\theta_0)u\right)-\ell_n\left(\theta_0\right)-\left(\Delta_n(\theta_0)[u]-\frac{1}{2}\mathcal{I}(\theta_0)[u,u]\right)\xrightarrow{p}0$$

- $\Delta_n(\theta_0):=\varphi_n(\theta_0)^\top\partial_\theta\ell_n(\theta_0)$
- $(\Delta_n(\theta_0),\,\mathcal{I}(\theta_0))\xrightarrow{\mathcal{L}}\left(\mathcal{I}(\theta_0)^{1/2}Z,\,\mathcal{I}(\theta_0)\right)$, $Z\sim N_{q+3}(0,I)\perp\!\!\!\perp(Y_0,J)$
- $\det\mathcal{I}(\theta_0)>0$ *a.s.*

- Followed by the lower bound, e.g.

$$\liminf_{n\to\infty}\mathbb{E}_{\theta_0}\left\{L\left(\varphi_n^{-1}(\theta_0)(\hat{\theta}_n-\theta_0)\right)\right\}\ge\iint L(w^{-1/2}z)(\mathbb{P}^{\mathcal{I}(\theta_0)}\otimes\mathbb{P}^{N_{q+3}(0,I)})(dw,dz),$$

$$\liminf_{\delta\to 0}\liminf_{n\to\infty}\sup_{\theta;|\varphi_n^{-1}(\theta_0)(\theta-\theta_0)|\le\delta}\mathbb{E}_\theta\left\{\left(\frac{\sqrt{n}}{\log(1/h)}(\hat{\sigma}_n-\sigma_0)\right)^2\right\}\ge C.$$

Optimality of the MLE

Theorem 3.2 (Asymptotic optimality of the MLE)

There exists a local maximum $\hat{\theta}_n$ of ℓ_n with prob.$\to 1$, s.t. $(\tilde{\varphi}_n := (\overline{\varphi}_{kl,n}))$

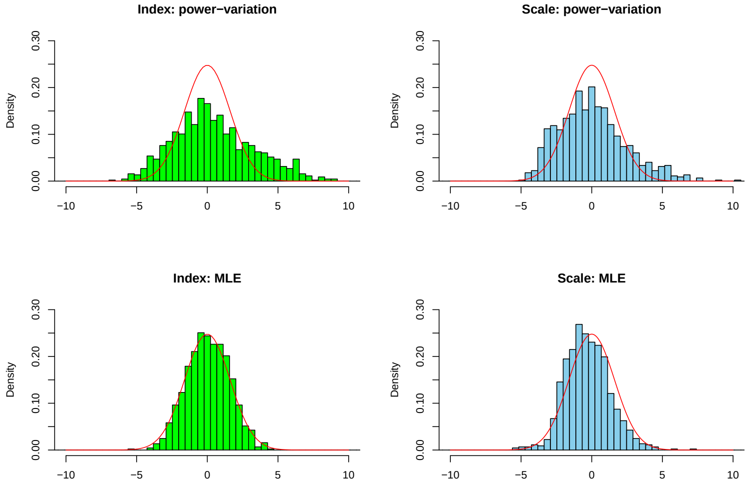
$$\left(\sqrt{n} h^{1-1/\beta_0} \begin{pmatrix} \hat{\lambda}_n - \lambda_0 \\ \hat{\boldsymbol{\mu}}_n - \boldsymbol{\mu}_0 \end{pmatrix}, \ \sqrt{n} \tilde{\varphi}_n(\theta_0)^{-1} \begin{pmatrix} \hat{\beta}_n - \beta_0 \\ \hat{\sigma}_n - \sigma_0 \end{pmatrix} \right) \xrightarrow{\mathcal{L}} \mathcal{L}' \otimes \mathcal{L}'', \quad (\mathbb{P}_{\theta_0})$$

- Straightforward to construct a consistent estimator $\widehat{\mathcal{I}}_n$ of $\mathcal{I}(\theta_0)$.
- Asymptotically linear dependency of MLE:

$$\frac{\sqrt{n}}{\log(1/h)} (\log \hat{\sigma}_n - \log \sigma_0) = {}^{\exists} C_{\theta_0} \sqrt{n} (\hat{\beta}_n - \beta_0) + O_p\left(\frac{1}{\log(1/h)} \right),$$

- **Non-Gaussianity is essential here.**

- **Stable Lévy process**: Histograms of the MLE and the $(q, 2q)$-moment estimator (ME) with $q = 0.1$ for (β, σ) (power-variation type), based on $1,000$ independent simulated paths of sample size $n = 2^9 = 512$: [Brouste and Masuda, 2018]



Remarks

$$dY_t = (\boldsymbol{\mu} \cdot \boldsymbol{X}_t - \lambda Y_t)dt + \sigma dJ_t, \quad \{(\boldsymbol{X}_t)_{t\in[0,1]},\, (Y_{t_j})_{j=0}^n\}$$

- **Time scale**, the terminal time $T \equiv 1$ **w.l.g.**:

$$\begin{aligned} dY'_t &= (T\boldsymbol{\mu} \cdot \boldsymbol{X}'_t - \lambda T Y'_t)dt + T^{1/\beta}\sigma d\overline{J}_t \\ &=: (\boldsymbol{\mu}'_T \cdot \boldsymbol{X}'_t - \lambda'_T Y'_t)dt + \sigma'_T d\overline{J}_t, \qquad t \in [0,1], \end{aligned}$$

with $\xi'_t := \xi_{tT}$ and $\overline{J}_t := T^{-1/\beta} J_{tT}$, i.e. *parameter re-scaling*:

$$\theta = (\lambda, \boldsymbol{\mu}, \beta, \sigma) \quad \leftarrow \quad \theta_T = (T\lambda, T\boldsymbol{\mu}, \beta, T^{1/\beta}\sigma)$$

- (Naive) Model assessment for significance of $X_{k,t}$ $(k = 1, \dots, q)$
  - e.g. If $\mu_{k,0} \neq 0$, the estimated p-value is to be small:

$$\hat{\mathfrak{p}}_n(\mu_k) := 2\Phi\left(- \left| \widehat{(\mathcal{I}_{\mu_k})}_n^{-1/2} \sqrt{n} h^{1-1/\hat{\beta}_n} \hat{\mu}_{k,n} \right| \right)$$

Summary and agenda

$$dY_t = (\boldsymbol{\mu} \cdot \boldsymbol{X}_t - \lambda Y_t)dt + dJ_t, \quad \{(\boldsymbol{X}_t)_{t\in[0,1]}, (Y_{t_j})_{j=0}^n\}$$

Merits of high-frequency scenario in handling dependent data

- LAMN property

"Individual" optimal rates	λ	$\boldsymbol{\mu}$	β	σ
Stable OU	$\sqrt{n}h^{1-1/\beta}$	$\sqrt{n}h^{1-1/\beta}$	$\sqrt{n}$	$\dfrac{\sqrt{n}}{\log(1/h)}$

- "Unified" asymptotic mixed normality of an MLE sequence
- Non-ergodicity and existence of a unit root does not matter here.

- Thinned $(\boldsymbol{X}_{t_j})_{j=0}^n$, random X, etc. could be handled (β restricted).
- Possible refinements:
  - (Quasi-)Likelihood analysis: a closer look at the log-LF;
  - Theoretical derivation of IC and/or model-confidence sets.

References (reprise) I

Brouste, A. and Masuda, H. (2018).
Efficient estimation of stable Lévy process with symmetric jumps.
Stat. Inference Stoch. Process., 21(2):289–307.

Clément, E. and Gloter, A. (2020).
Joint estimation for SDE driven by locally stable Lévy processes.
Electron. J. Stat., 14(2):2922–2956.

Eguchi, S. and Masuda, H. (2018).
Schwarz type model comparison for LAQ models.
Bernoulli, 24(3):2278–2327.

Eguchi, S. and Masuda, H. (2019).
Data driven time scale in Gaussian quasi-likelihood inference.
Stat. Inference Stoch. Process., 22(3):383–430.

Eguchi, S. and Uehara, Y. (2019).
Schwartz type model selection for ergodic stochastic differential equation models.
arXiv preprint arXiv:1904.12398.

Kulik, A. (2019).
Approximation in law of locally α-stable Lévy-type processes by non-linear regressions.
Electron. J. Probab., 24:Paper No. 83, 45.

References (reprise) II

Lavielle, M. (2015).
Mixed effects models for the population approach.
Chapman & Hall/CRC Biostatistics Series. CRC Press, Boca Raton, FL.
Models, tasks, methods and tools, With contributions by Kevin Bleakley.

Masuda, H. (2019).
Non-Gaussian quasi-likelihood estimation of SDE driven by locally stable Lévy process.
Stochastic Process. Appl., 129(3):1013–1059.

Uehara, Y. and Masuda, H. (2017).
Stepwise estimation of a Lévy driven stochastic differential equation.
Proc. Inst. Statist. Math. (Japanese), 65(1):21–38.