Outlier-resistant inference without jump-detection filter *

Shoichi Eguchi[1] Hiroki Masuda[2]

[1]Faculty of Information Science and Technology, Osaka Institute of Technology

[2]Graduate School of Mathematical Sciences, University of Tokyo

Fall School "Time Series, Random Fields and beyond"

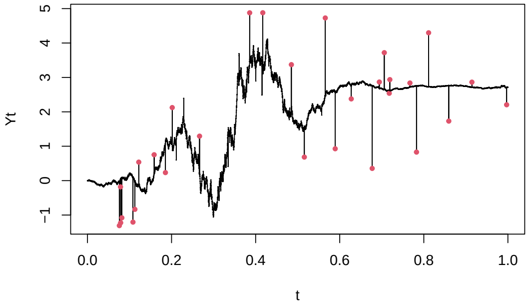
Ulm University

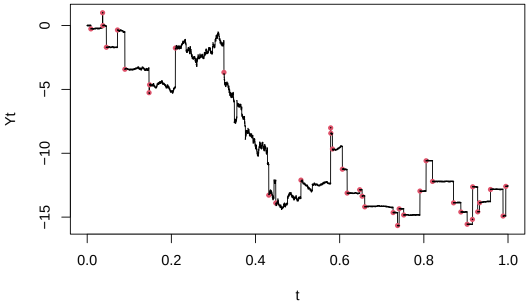
September 26, 2024 †

*Supported by JSPS KAKENHI (22H01139) and by JST CREST (JPMJCR2115).

†This version: September 26, 2024

Continuous-time volatility modeling





- Want to focus on the black line.
- Reagarding the red points as external enemy.

Objective

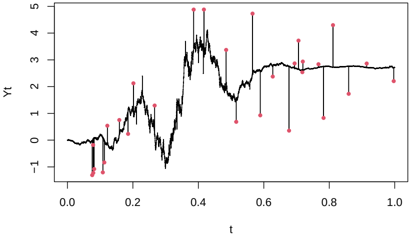
Robust parametric inference for the volatility, ignoring **discontinuous contaminations**

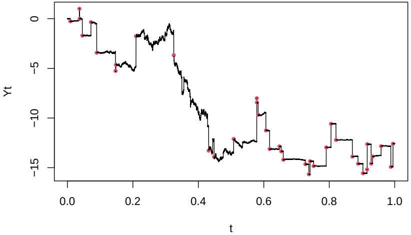
$$\begin{cases} Y_t^\star = Y_0^\star + \int_0^t \mu_{s-} ds + \int_0^t \sigma(X_{s-}^\star, \theta) dw_s + J_t \\ X_t^\star = X_0^\star + \int_0^t \mu'_{s-} ds + \int_0^t \sigma'_{s-} dw'_s + J'_t \end{cases}$$

$$\begin{cases} Y_t = Y_t^\star + \sum_{j=1}^n \Upsilon_j I(t = t_j) \\ X_t = X_t^\star + \sum_{j=1}^n \Upsilon'_j I(t = t_j) \end{cases}$$

$$\overset{\text{Obs.}}{\Longrightarrow} \quad \{(X_{t_j}, Y_{t_j})\}_{j=0}^n, \quad t_j = jT/n =: jh$$

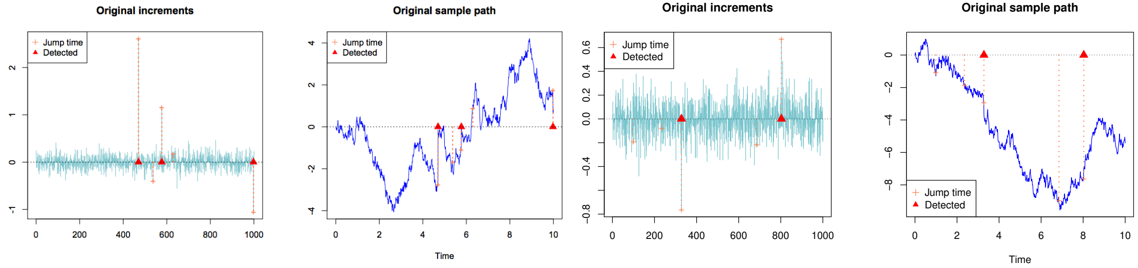
- Asymptotic inference for θ
- Parametric diffusion coeff. $\sigma(x, \theta)$





Outline

How to handle jumps



- **Pruning (Selection) type** for both ergodic and non-ergodic cases:

$$\text{Ignore } Y_{t_j} - Y_{t_{j-1}} \quad \text{if} \quad |Y_{t_j} - Y_{t_{j-1}}| \geq Ch^{\kappa}$$

  - **Local jump detection filter: [Shimizu and Yoshida, 2006], [Ogihara and Yoshida, 2011], ...**
  - Self-normalized statistics bases test [Masuda and Uehara, 2021]: Remove large increments
  - Global jump detection filter [Inatsugu and Yoshida, 2021]: Order-statistics based "deformation"

Our interest: simple and (relatively) tuning-resistant way?

$$Y_t^\star = Y_0^\star + \int_0^t \mu_{s-} ds + \int_0^t \sigma(X_{s-}^\star, \theta) dw_s + J_t, \qquad X_t^\star = X_0^\star + \int_0^t \mu'_{s-} ds + \int_0^t \sigma'_{s-} dw'_s + J'_t,$$

$$Y_t = Y_t^\star + \sum_{j=1}^n \Upsilon_j I(t = t_j), \qquad X_t = X_t^\star + \sum_{j=1}^n \Upsilon'_j I(t = t_j) \overset{\text{Obs.}}{\Longrightarrow} \{(X_{t_j}, Y_{t_j})\}_{j=0}^n, \quad t_j = jT/n =: jh$$

- Application of **the density-power divergence** [Basu et al., 1998]? (like the Box-Cox trans.)

$$(f; g) \mapsto \int \left(f^{1+\lambda} - \left(1 + \frac{1}{\lambda}\right) f^\lambda g + \frac{1}{\lambda} g^{1+\lambda} \right) d\mu \geq 0$$

  - Converges to the Kullback-Leibler divergence (from g to f) for $\lambda \downarrow 0$.

- Asymptotic inference theory with effectively ignoring contaminations?
- No heavy computational loading, and no sensitive fine-tuning.

Previous studies for ergodic diffusion ($T = nh \to \infty$)

- [Lee and Song, 2013] considered $dY_t = \mu(Y_t, \theta)dt + \sigma dw_t$ and proved

$$\left(\sqrt{nh}(\hat{\theta}_n(\lambda) - \theta_0),\ \sqrt{n}(\sigma_n(\lambda) - \sigma_0)\right) \xrightarrow{\mathcal{L}} N(0,\ V_0(\lambda))$$

through minimizing the the (explicit) Euler-scheme based objective function:

$$(\theta, \sigma) \mapsto \int \phi(y; Y_{t_{j-1}} + \mu_{j-1}(\theta)h, \sigma^2 h)^{1+\lambda} dy - \left(1 + \frac{1}{\lambda}\right) \sum_{j=1}^{n} \phi(Y_{t_j}; Y_{t_{j-1}} + \mu_{j-1}(\theta)h, \sigma^2 h)^{\lambda}$$

- [Song, 2017] similarly considered $dY_t = \mu(Y_t, \theta)dt + \sigma(Y_t, \gamma)dw_t$ and proved the consistency

$$(\hat{\theta}_n(\lambda),\ \sigma_n(\lambda)) \xrightarrow{p} (\theta_0, \sigma_0).$$

- Yet, they/he did no theoretical consideration about **what happens under contaminations**.

- We are given a complete filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in T}, P)$.

Setup: Stochastic dynamic regression with jumps and spikes

$$Y_t^\star = Y_0^\star + \int_0^t \mu_{s-} ds + \int_0^t \sigma(X_{s-}^\star, \theta) dw_s + J_t, \qquad X_t^\star = X_0^\star + \int_0^t \mu'_{s-} ds + \int_0^t \sigma'_{s-} dw'_s + J'_t,$$

$$Y_t = Y_t^\star + \sum_{j=1}^n \Upsilon_j I(t = t_j), \qquad X_t = X_t^\star + \sum_{j=1}^n \Upsilon'_j I(t = t_j) \overset{\text{Obs.}}{\Longrightarrow} \{(X_{t_j}, Y_{t_j})\}_{j=0}^n, \quad t_j = jT/n =: jh$$

- Jumps $J_t = \sum_{0<s\le t} \Delta Y_s^\star$ ($\Delta Y_s^\star := Y_s^\star - Y_{s-}^\star$) and $J'_t = \sum_{0<s\le t} \Delta X_s^\star$
- Spikes $\Upsilon_j = \Upsilon_{n,j}$ and $\Upsilon'_j = \Upsilon'_{n,j}$ are $\mathcal{F}_{t_j}$-m'ble r.v's.
- Fake **Gaussian quasi-likelihood** (QL) based on the Euler approx. $Y_{t_j} \overset{P_\theta}{\approx} Y_{t_{j-1}} + h\sigma_{j-1}(\theta)\Delta_j w$:

$$\mathbb{H}_n(\theta) := \sum_{j=1}^n \log \phi_d \left(Y_{t_j};\ Y_{t_{j-1}},\ hS_{j-1}(\theta) \right) =: \sum_{j=1}^n \log \phi_j(\theta)$$

  - $S := \sigma^{\otimes 2}$, $f_{j-1}(\theta) := f(X_{t_{j-1}}, Y_{t_{j-1}}; \theta)$.

Density-power weighting for Gaussian quasi-(log-)likelihood

$$(f;g)\mapsto \int \left(f^{1+\lambda} - \left(1+\frac{1}{\lambda}\right) f^{\lambda} g + \frac{1}{\lambda} g^{1+\lambda} \right) d\mu =: \text{Const.} + \int f^{1+\lambda} d\mu - \left(1+\frac{1}{\lambda}\right) \int f^{\lambda} g d\mu \ge 0$$

- Applying the above density-power form $\left(\phi_{j-1}(y;\theta) := \phi\left(y;\, Y_{t_{j-1}},\, hS_{j-1}(\theta)\right)\right)$,

$$\theta \mapsto \sum_{j=1}^{n} \left(\frac{1}{\lambda}\phi_j(\theta)^{\lambda} - \frac{1}{\lambda+1} \int \phi_{j-1}(y;\theta)^{\lambda+1} dy \right)$$

$$= \sum_{j=1}^{n} \left(\frac{1}{\lambda}\phi_j(\theta)^{\lambda} - h^{-d\lambda/2} \frac{(2\pi)^{-d\lambda/2}}{(\lambda+1)^{1+d/2}} \det(S_{j-1}(\theta))^{-\lambda/2} \right)$$

Definition 2.1 (Density-power Gaussian QL ($\phi(\cdot); = \phi_d(\cdot; 0, I_d)$; $0 < \lambda \le 1$))

$$\mathbb{H}_n(\theta;\lambda) = \sum_{j=1}^{n} \det(S_{j-1}(\theta))^{-\lambda/2} \left(\frac{1}{\lambda}\phi\left(S_{j-1}(\theta)^{-1/2} h^{-1/2} \Delta_j Y\right)^{\lambda} - \frac{(2\pi)^{-d\lambda/2}}{(\lambda+1)^{1+d/2}} \right)$$

- The best possible phenomenon for any regular estimator $\widetilde{\theta}_n$ is well-known [Gobet, 2001]:

$$\sqrt{n}(\widetilde{\theta}_n - \theta_0) \xrightarrow{\mathcal{L}} MN\left(0, \left(\frac{1}{2T}\int_0^T \operatorname{trace}\left((S^{-1}(\partial_\theta S)S^{-1}(\partial_\theta S))_t\right) dt\right)^{-1}\right)$$

- Our $\hat{\theta}_n(\lambda)$ has an asymptotic mixed normality **for each $\boldsymbol{\lambda} > 0$**.

Regularity conditions in brief

$$Y_t^\star = Y_0^\star + \int_0^t \mu_{s-} ds + \int_0^t \sigma(X_{s-}^\star, \theta) dw_s + J_t, \qquad X_t^\star = X_0^\star + \int_0^t \mu'_{s-} ds + \int_0^t \sigma'_{s-} dw'_s + J'_t,$$

$$Y_t = Y_t^\star + \sum_{j=1}^n \Upsilon_j I(t = t_j), \qquad X_t = X_t^\star + \sum_{j=1}^n \Upsilon'_j I(t = t_j) \overset{\text{Obs.}}{\Longrightarrow} \{(X_{t_j}, Y_{t_j})\}_{j=0}^n, \quad t_j = jT/n =: jh$$

Assumption 3.1

1. Smoothness and non-degeneracy of $(x, \theta) \mapsto S(x, \theta)$
2. Integrability of (μ, X, J)
3. **Probabilistic structures of finite-activity jumps**
4. **Probabilistic structures of spike noise**
5. Identifiability, and positive definiteness of the asymptotic covariance

Asymptotic mixed normality for fixed $\lambda > 0$

Theorem 3.2

$$\sqrt{n}(\hat{\theta}_n(\lambda) - \theta_0) \xrightarrow{\mathcal{L}} MN_p\left(0, \Gamma_0(\lambda)^{-1}\Sigma_0(\lambda)\Gamma_0(\lambda)^{-1}\right)$$

$$\Sigma_0^{(k,l)}(\lambda) := \frac{(2\pi)^{-d\lambda}}{4T}\int_0^T \det(S_t)^{-\lambda/2} \times \Big\{ \left(2\lambda(2\lambda+1)^{-(1+d/2)} - \lambda^2(\lambda+1)^{-(2+d)}\right) \times trace\left((S^{-1}\partial_{\theta_k}S)_t\right) trace\left((S^{-1}\partial_{\theta_l}S)_t\right) + 2(2\lambda+1)^{-(1+d/2)} trace\left((S^{-1}(\partial_{\theta_k}S)S^{-1}(\partial_{\theta_l}S))_t\right) \Big\} dt,$$

$$\Gamma_0^{(k,l)}(\lambda) := \frac{(2\pi)^{-d\lambda/2}}{(\lambda+1)^{1+d/2}} \times \frac{1}{2T}\int_0^T \det(S_t)^{-\lambda/2}\Big\{(1-\lambda) trace\left((S^{-1}(\partial_{\theta_k}S)S^{-1}(\partial_{\theta_l}S))_t\right) - \lambda^2 trace\left((S^{-1}(\partial_{\theta_k}S))_t\right) trace\left((S^{-1}(\partial_{\theta_l}S))_t\right)\Big\} dt$$

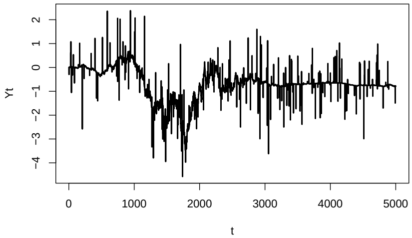
(We wrote $f_t := f(X_t^\star, \theta_0)$).

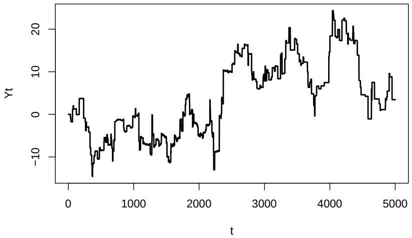
- $\lim_{\lambda\downarrow 0}\Sigma_0(\lambda) = \lim_{\lambda\downarrow 0}\Gamma_0(\lambda) = \frac{1}{2T}\int_0^T \mathrm{trace}\left((S^{-1}(\partial_{\theta_k}S)S^{-1}(\partial_{\theta_l}S))_t\right) dt$ a.s. (Fisher-info. matrix)

Simulation designs: Gaussian additive processes with contamination

- Model: $Y_t = \int_0^t \exp\left\{\frac{1}{2}(\theta_1 X_{1,s} + \theta_2 X_{2,s})\right\} dw_s$, $(X_{1,t_j}, X_{2,t_j}) = \big(\cos(2j\pi/n), \sin(2j\pi/n)\big)$

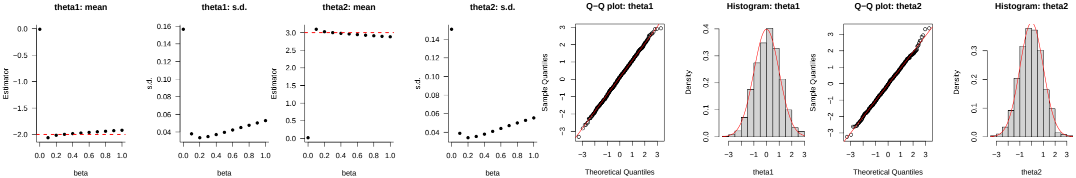
1. **Spike noise**: $Y_{t_j} = Y^i_{o,t_j} + p_j Y_{c,t_j}$, $t_j = j/n \in [0,1]$
   - $Y_{c,t_1}, \dots, Y_{c,t_n} \sim$ i.i.d. $N(0,1)$; $p_0, p_1, \dots, p_n \sim$ i.i.d. Bernoulli(0.05) [5% contamination], $(\theta_{1,0}, \theta_{0,2}) = (-2, 3)$, $n = 5000$, $\sharp$MC $= 1000$.
2. **Compound Poisson jumps**: $Y_t = \int_0^t \exp\left\{\frac{1}{2}(-2X_{1,s} + 3X_{2,s})\right\} dw_s + J_t$, $Y_0 = 0$
   - $\mathcal{L}(J_t) = CP(\lambda t, N(0,5))$ with $\lambda = 0.05n$; $n = 5000$, $\sharp$MC $= 1000$ [250 jumps in average]



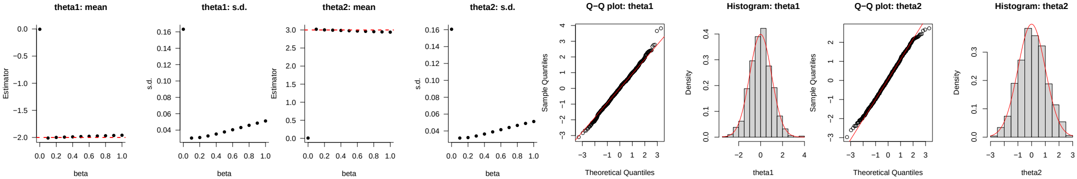


- Checking the robustness

1. Spike case

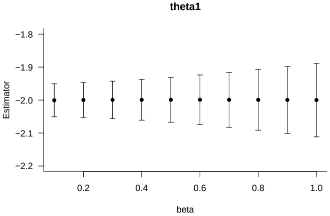


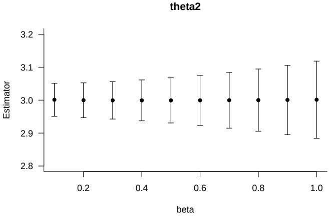
2. Jump case



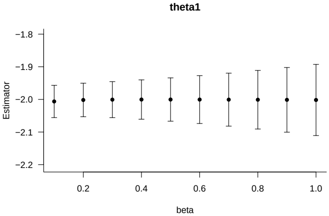
- Plots of averaged estimates with confidence intervals (Theorem 3.2); $(\theta_{1,0}, \theta_{0,2}) = (-2, 3)$.

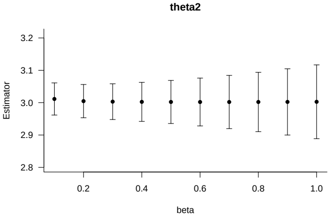
1. Spike case





2. Jump case





Summary: Robust volatility inference with single fine tuning

$$Y_t^\star = Y_0^\star + \int_0^t \mu_{s-} ds + \int_0^t \sigma(X_{s-}^\star, \theta) dw_s + J_t, \qquad X_t^\star = X_0^\star + \int_0^t \mu'_{s-} ds + \int_0^t \sigma'_{s-} dw'_s + J'_t,$$

$$Y_t = Y_t^\star + \sum_{j=1}^n \Upsilon_j I(t = t_j), \qquad X_t = X_t^\star + \sum_{j=1}^n \Upsilon'_j I(t = t_j) \overset{\text{Obs.}}{\Longrightarrow} \{(X_{t_j}, Y_{t_j})\}_{j=0}^n, \quad t_j = jT/n =: jh$$

Explicit density-power Gaussian QLF

$$\mathbb{H}_n(\theta;\lambda) := \sum_{j=1}^n \det(S_{j-1}(\theta))^{-\lambda/2} \left(\frac{1}{\lambda} \phi\big(S_{j-1}(\theta)^{-1/2} h^{-1/2} \Delta_j Y\big)^\lambda - \frac{(2\pi)^{-d\lambda/2}}{(\lambda+1)^{1+d/2}} \right)$$

$$\sqrt{n}(\hat{\theta}_n(\lambda) - \theta_0) \xrightarrow{\mathcal{L}} MN_p\left(0, \Gamma_0(\lambda)^{-1}\Sigma_0(\lambda)\Gamma_0(\lambda)^{-1}\right)$$

Concluding remarks I

- **Controlling the tuning parameter** as $\lambda = \lambda_n \downarrow 0$ at appropriate rate:

$$\sqrt{n}(\hat{\theta}_n(\lambda_n) - \theta_0) \xrightarrow{\mathcal{L}} MN_p\left(0, \left(\frac{1}{2T}\int_0^t \mathrm{trace}\left((S^{-1}(\partial_\theta S)S^{-1}(\partial_\theta S)(X_t,\theta_0)\right) dt\right)^{-1}\right)$$

  - For construction of BIC type model selection criterion based on $\mathbb{H}_n(\theta;\lambda)$ (Ongoing).

- Applicable to **Hölder based divergence**, also known as γ-divergence, as well.
  - [Windham, 1995] and [Jones et al., 2001]; also [Fujisawa and Eguchi, 2008].
  - Approx. martingale-estimating-function version can effectively handle heteroskedasticity.

$$\mathbb{H}_n(\theta;\lambda) := \sum_{j=1}^{n} \det(S_{j-1}(\theta))^{-\lambda/(2(\lambda+1))}\phi\left(S_{j-1}(\theta)^{-1/2}h^{-1/2}\Delta_j Y\right)^{\lambda}.$$

Concluding remarks II

- The approximation technique will apply to **various other "Gaussian type" QL inferences**:

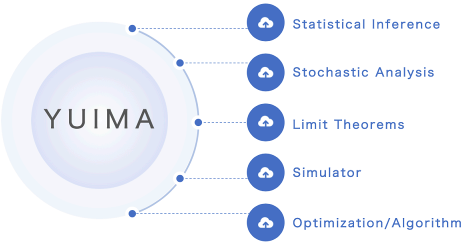
$$\mathbb{H}_n(\theta) \leftarrow \sum_{j=1}^{n} \log \phi_d \left(Y_{t_j};\, \mu_h(X_{t_{j-1}}, \theta),\, \Sigma_h(X_{t_{j-1}}, \theta; h) \right)$$

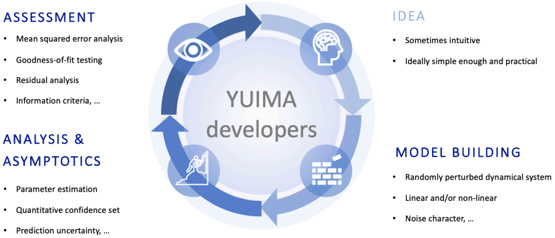
$$\mu_h(X_{t_{j-1}}, \theta) \approx E_\theta^{j-1}[Y_{t_j}], \qquad \Sigma_h(X_{t_{j-1}}, \theta; h) \approx \mathrm{Cov}_\theta^{j-1}[Y_{t_j}],$$

such as location-scale time series model, ergodic diffusion, **ergodic Lévy driven SDE**[1], small diffusion, small-Lévy driven SDE, etc.

Concluding remarks III

- Implementation of "`qmlerobust`" in yuima R package [Brouste et al., 2014] (by S. Eguchi);
  User input: T, n, $\sigma(x,\theta)$, and $\lambda \in (0,1]$ (with the rule of thumb "use $\lambda = 0.2 \sim 0.1$").





[1] Essentially requires different consideration! Ongoing.

Related references I

Basu, A., Harris, I. R., Hjort, N. L., and Jones, M. C. (1998).
Robust and efficient estimation by minimising a density power divergence.
Biometrika, 85(3):549–559.

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.

Fujisawa, H. and Eguchi, S. (2008).
Robust parameter estimation with a small bias against heavy contamination.
J. Multivariate Anal., 99(9):2053–2081.

Gobet, E. (2001).
Local asymptotic mixed normality property for elliptic diffusion: a Malliavin calculus approach.
Bernoulli, 7(6):899–912.

Inatsugu, H. and Yoshida, N. (2021).
Global jump filters and quasi-likelihood analysis for volatility.
Ann. Inst. Statist. Math., 73(3):555–598.

Jones, M. C., Hjort, N. L., Harris, I. R., and Basu, A. (2001).
A comparison of related density-based minimum divergence estimators.
Biometrika, 88(3):865–873.

Related references II

Lee, S. and Song, J. (2013).
Minimum density power divergence estimator for diffusion processes.
Ann. Inst. Statist. Math., 65(2):213–236.

Masuda, H. and Uehara, Y. (2021).
Estimating diffusion with compound Poisson jumps based on self-normalized residuals.
J. Statist. Plann. Inference, 215:158–183.

Ogihara, T. and Yoshida, N. (2011).
Quasi-likelihood analysis for the stochastic differential equation with jumps.
Stat. Inference Stoch. Process., 14(3):189–229.

Shimizu, Y. and Yoshida, N. (2006).
Estimation of parameters for diffusion processes with jumps from discrete observations.
Stat. Inference Stoch. Process., 9(3):227–277.

Song, J. (2017).
Robust estimation of dispersion parameter in discretely observed diffusion processes.
Statist. Sinica, 27(1):373–388.

Windham, M. P. (1995).
Robustifying model fitting.
J. Roy. Statist. Soc. Ser. B, 57(3):599–609.