Bootstrap method for misspecified stochastic differential equation models

Yuma Uehara

The Institute of Statistical Mathematics

Risk and Statistics - 2nd ISM-UUlm Joint Workshop
October 8 - October 10, 2019

Outline

Setting

$$\text{Data-generating model}: \boldsymbol{dX_t = A(X_t)dt + C(X_{t-})dZ_t},$$
$$\text{Statistical model}: \boldsymbol{dX_t = a(X_t, \alpha)dt + c(X_{t-}, \gamma)dZ_t}$$

- Our estimation target: $\boldsymbol{\theta := (\gamma, \alpha) \in \Theta_\gamma \times \Theta_\alpha := \Theta}$, and the parameter space $\boldsymbol{\Theta}$ is a bounded convex space.
- The parameter spaces $\boldsymbol{\Theta_\gamma}$ and $\boldsymbol{\Theta_\alpha}$ are subsets of $\mathbb{R}^{p_\gamma}$ and $\mathbb{R}^{p_\alpha}$, respectively.
- The drift coefficients $\boldsymbol{A}$ and $\boldsymbol{a}$, and the scale coefficients $\boldsymbol{C}$ and $\boldsymbol{c}$ are Lipschitz continuous, and smooth enough, and they and their derivatives are of at most polynomial growth. We further suppose that $\boldsymbol{1/c}$ and $\boldsymbol{1/C}$ are bounded away from 0.

Setting (cont'd)

- The driving noise Z is a standard Wiener process (hereafter it is sometimes written as w), or a pure-jump Lévy process with $E[Z_t] = 0, E[Z_t^2] = t$, $E[|Z_t|^q] < \infty$, and $E[|X_t|^q] < \infty$ for any $q > 0$. Furthermore, we assume that the Blumenthal-Getoor index (BG-index) of Z is smaller than 2, that is, for the Lévy measure ν_0 of Z,

$$\beta := \inf_{\gamma} \left\{ \gamma \geq 0 : \int_{|z| \leq 1} |z|^{\gamma} \nu_0(dz) < \infty \right\} < 2.$$

- There exists a probability measure π_0 such that for every $q > 0$, we can find constants $a > 0$ and $C_q > 0$ for which

$$\sup_{t \in \mathbb{R}_+} \exp(at) ||P_t(x, \cdot) - \pi_0(\cdot)||_{h_q} \leq C_q h_q(x),$$

for any $x \in \mathbb{R}$ where $h_q(x) := 1 + |x|^q$.

- Then we have the ergodic theorem: as $T \to \infty$, for any polynomial growth function f, $\frac{1}{T} \int_0^T f(X_t) dt \xrightarrow{p} \int f(x) \pi_0(dx)$.

Setting (cont'd)

- Observation: From the solution process $\boldsymbol{X}$, we suppose that we observe discrete but high-frequency samples $(\boldsymbol{X}_{t_j})_{j=0}^{n}$ under the so-called "rapidly increasing design":
  $t_j := t_j^n = jh_n, T_n := nh_n \to \infty, nh_n^2 \to 0.$
- Model misspecification: Our statistical model is possibly misspecified, i.e., for all $\theta \in \Theta$, $A(x) \neq a(x, \alpha)$ and $C(x) \neq c(x, \gamma)$ on the set S, and $\pi_0(S) > 0$.
- In general, we cannot avoid the model misspecification. The theory of misspecified models is considered in many papers. For example, Berk (1966), Huber (1967), and White (1984), to mention few. Especially, for stochastic differential equation models, see McKeague (1984), Uchida and Yoshida (2011), Kutoyants (2017), Uehara (2019), and so on.
- In this talk, we consider the four cases: the correctly specified diffusion case, the misspecified diffusion case, the correctly specified pure-jump Lévy driven case, and the misspecified pure-jump Lévy driven case.

Model setting

Gaussian quasi-likelihood estimation

- Diffusion case
- Pure-jump Lévy driven case

Bootstrap method for block sum

- Pure-jump Lévy driven case
- Diffusion case

Modified bootstrap method

Summary

Gaussian quasi-likelihood estimation

- $\Delta_j X := X_{t_j} - X_{t_{j-1}}$, $\Delta_j Z := Z_{t_j} - Z_{t_{j-1}}$, $f_s(\theta) := f(X_{s-}, \theta)$, $f_j(\theta) := f(X_{t_j}, \theta)$, $\phi(x; \mu, \Sigma)$: the density function of the normal distribution whose mean and variance are μ and Σ, respectively.
- We define the Gaussian quasi-likelihood estimator $\hat{\theta}_n := (\hat{\gamma}_n, \hat{\alpha}_n)$ by

$$\hat{\gamma}_n = \underset{\gamma \in \bar{\Theta}_\gamma}{\operatorname{argmax}} \sum_{j=1}^{n} \log \phi(\Delta_j X; 0, h_n c_{j-1}^2(\gamma)),$$

$$\hat{\alpha}_n = \underset{\alpha \in \bar{\Theta}_\alpha}{\operatorname{argmax}} \sum_{j=1}^{n} \log \phi(\Delta_j X; h_n a_{j-1}(\alpha), h_n c_{j-1}^2(\hat{\gamma}_n)).$$

- The asymptotic behavior of $\hat{\theta}_n$ is studied in all cases, for instance, see Kessler (1997), Uchida and Yoshida (2011), Masuda (2013), and Uehara (2019).

Optimal value

- We define the optimal value $\theta^\star := (\gamma^\star, \alpha^\star)$ by

$$\gamma^\star := \operatorname*{argmax}_{\gamma \in \bar{\Theta}_\gamma} \int_{\mathbb{R}} -\left(\log c^2(x,\gamma) + \frac{C^2(x)}{c^2(x,\gamma)} \right) \pi_0(dx) (=: \mathbb{G}_1(\gamma)),$$

$$\alpha^\star := \operatorname*{argmax}_{\alpha \in \bar{\Theta}_\alpha} \int_{\mathbb{R}} -\frac{(A(x) - a(x,\alpha))^2}{c^2(x,\gamma^\star)} \pi_0(dx) (=: \mathbb{G}_2(\alpha)).$$

- In the correctly specified case, the optimal value $\theta^\star$ corresponds to the true value.
- We assume the model separability

$$\mathbb{G}_1(\gamma) - \mathbb{G}_1(\gamma^\star) \le -\chi_\gamma |\gamma - \gamma^\star|^2,$$
$$\mathbb{G}_2(\alpha) - \mathbb{G}_2(\alpha^\star) \le -\chi_\alpha |\alpha - \alpha^\star|^2.$$

Misspecification bias

We illustrate how the misspecification effect appears (since the drift part is almost the same, we look at the scale part).

- In the misspecified diffusion case,

$$\text{scaled (quasi-)score function} = \frac{1}{\sqrt{T_n}}\int_0^{T_n} b(X_s,\theta^\star)ds + o_p(1)$$

- In the misspecified pure-jump Lévy driven case,

$$\text{scaled (quasi-)score function} = \underbrace{\frac{1}{\sqrt{T_n}}\int_0^{T_n}\int_{\mathbb{R}} \bar{m}(X_{s-},\theta^\star)\tilde{N}(ds,dz)}_{\text{CLT term}} + \frac{1}{\sqrt{T_n}}\int_0^{T_n} b(X_s,\theta^\star)ds + o_p(1).$$

- ν_0 and $\tilde{N}(ds,dz) = N(ds,dz) - ds\nu_0(dz)$ are the corresponding compensated Poisson random measure, and Lévy measure, respectively.

Misspecification bias

- b is the misspecification bias, and $\int b(x, \theta^\star)\pi_0(dx) = 0$. Especially, $b \equiv 0$ in the correctly specified case.
- Although the limit theory for integrals of functional of Markov process is developed in some literature (e.g. Bhattacharya (1982), Komorowski and Walczuk (2012)), its sufficient conditions are difficult to check or the joint asymptotic distribution with the main term is not trivial.
- How to correct the misspecification bias?

Diffusion case (Uchida and Yoshida (2011))

- Let $\mathcal{A}$ be the infinitesimal generator of X.
- Itô's formula: for a smooth enough f, $f(X_t) = f(X_0) + \int_0^t \mathcal{A}f(X_s)ds + \int \partial_x f(X_s)C(X_s)dw_s$.
- If there exists a function f solving $\mathcal{A}f = b$, we can transform the bias term:

$$\int_0^{T_n} b(X_s, \theta^\star)ds = f(X_{T_n}) - f(X_0) - \int_0^{T_n} \partial_x f(X_s)C(X_s)dw_s.$$

- Since the equation $\mathcal{A}f = b$ (so-called Poisson equation) is the second order differential equation, the existence of the solution and its regularity is ensured (cf. Pardoux and Veretennikov (2001)).

Pure-jump Lévy driven case (Uehara (2019))

- We cannot apply the same approach as the diffusion case since the equation $\mathcal{A}f = b$ has the integral operator with respect to the Lévy measure.
- Instead of $\mathcal{A}$, we consider the extended infinitesimal generator $\tilde{\mathcal{A}}$ of X, and the corresponding extended Poisson equations (cf. Kulik and Veretennikov (2011)).

Definition (Kulik and Veretennikov (2011))

We say that a measurable function $g : \mathbb{R} \to \mathbb{R}$ belongs to the domain of the extended generator $\tilde{\mathcal{A}}$ of a càdlàg homogeneous Feller Markov process Y taking values in $\mathbb{R}$ if there exists a measurable function $b : \mathbb{R} \to \mathbb{R}$ such that the process

$$g(Y_t) - \int_0^t b(Y_s)ds, \qquad t \in \mathbb{R}^+,$$

is well defined and is a local martingale with respect to the natural filtration of Y and every measure $P_x(\cdot) := P(\cdot|Y_0 = x),\ x \in \mathbb{R}$. For such a pair (g, b), we write $g \in \mathrm{Dom}(\tilde{\mathcal{A}})$ and $\tilde{\mathcal{A}}g \overset{EPE}{=} b$.

Pure-jump Lévy driven case (cont'd)

Uehara (2019), Proposition 3.5

The potential function

$$g(x) := \int_0^\infty E^x[b(X_t, \theta^\star)]dt$$

is the unique solution of $\tilde{A}g = b$, and it satisfies that for all $p \in (1, \infty)$ and $q = \frac{p}{p-1}$,

$$\sup_{x,y\in\mathbb{R}, x\neq y} \frac{|g(x) - g(y)|}{|x-y|^{1/p}(1 + |x|^q + |y|^q)} < \infty.$$

- Combined with the martingale representation theorem, we have a similar transformation to the diffusion case.

Asymptotic distribution

- Let $A_n := \mathrm{diag}\{a_n I_{p_\gamma}, \sqrt{T_n} I_{p_\alpha}\}$ where $a_n = \sqrt{n}$ in the correctly specified diffusion case, and otherwise, $a_n = \sqrt{T_n}$.

Theorem

- Tail probability estimates: for any $L > 0$ and $r > 0$, there exists a positive constant C_L such that

$$\sup_{n\in\mathbb{N}} P\left(\left|A_n(\hat{\theta}_n - \theta^\star)\right| > r\right) \le \frac{C_L}{r^L}. \qquad (1)$$

- Asymptotic normality:

$$A_n(\hat{\theta}_n - \theta^\star) \xrightarrow{\mathcal{L}} N(0, \mathcal{I}^{-1}\Sigma(\mathcal{I}^{-1})^\top).$$

The form of $\mathcal{I}$

- The matrix $\mathcal{I} = \begin{pmatrix} \mathcal{I}_\gamma & O \\ \mathcal{I}_{\alpha\gamma} & \mathcal{I}_\alpha \end{pmatrix}$ is common and defined as

$$\begin{aligned}
\mathcal{I}_\gamma &= 4\int_{\mathbb{R}} \frac{(\partial_\gamma c(x,\gamma^\star))^{\otimes 2}}{c^4(x,\gamma^\star)} C^2(x)\pi_0(dx) \\
&\quad - 2\int_{\mathbb{R}} \frac{\partial_\gamma(c(x,\gamma^\star)\partial_\gamma c(x,\gamma^\star))}{c^4(x,\gamma^\star)}(C^2(x) - c^2(x,\gamma^\star))\pi_0(dx), \\
\mathcal{I}_\alpha &= 2\int_{\mathbb{R}} \frac{(\partial_\alpha a(x,\alpha^\star))^{\otimes 2}}{c^2(x,\gamma^\star)}\pi_0(dx) \\
&\quad - 2\int_{\mathbb{R}} \frac{\partial_\alpha^{\otimes 2} a(x,\alpha^\star)}{c^2(x,\gamma^\star)}(A(x) - a(x,\alpha^\star))\pi_0(dx), \\
\mathcal{I}_{\alpha\gamma} &= 2\int_{\mathbb{R}} \partial_\alpha a(x,\alpha^\star)\partial_\gamma^\top c^{-2}(x,\gamma^\star)(a(x,\alpha^\star) - A(x))\pi_0(dx).
\end{aligned}$$

- It is easy to construct a consistency estimator $\hat{\mathcal{I}}_n$ of $\mathcal{I}$.

The form of Σ (correctly specified diffusion case)

$$\boldsymbol{\Sigma} = 2\mathcal{I} = 2\,\mathrm{diag}\{\mathcal{I}_\gamma, \mathcal{I}_\alpha\}$$

$$= \begin{pmatrix} 8\int_{\mathbb{R}} \frac{(\partial_\gamma c(x,\gamma^\star))^{\otimes 2}}{c^2(x,\gamma^\star)}\pi_0(dx) & O \\ O & 4\int_{\mathbb{R}} \frac{(\partial_\alpha a(x,\alpha^\star))^{\otimes 2}}{c^2(x,\gamma^\star)}\pi_0(dx) \end{pmatrix}.$$

The form of Σ (misspecified diffusion case)

$$\Sigma_\gamma = 4\int (\partial_x f_1(x) C(x))^{\otimes 2}\pi_0(dx),$$

$$\Sigma_{\alpha\gamma} = 4\int \left(\frac{\partial_\alpha a(x,\alpha^\star)}{c^2(x,\gamma^\star)} - \partial_x f_2(x)\right) C^2(x)(\partial_x f_1(x))^\top \pi_0(dx),$$

$$\Sigma_\alpha = 4\int \left[\left(\frac{\partial_\alpha a(x,\alpha^\star)}{c^2(x,\gamma^\star)} - \partial_x f_2(x)\right) C(x)\right]^{\otimes 2}\pi_0(dx),$$

where the functions f_1 and f_2 are the solution of the following Poisson equations:

$$\mathcal{A}f_1^{(j_1)}(x) = \frac{\partial_{\gamma^{(j_1)}} c(x,\gamma^\star)}{c^3(x,\gamma^\star)}(c^2(x,\gamma^\star) - C^2(x)),$$

$$\mathcal{A}f_2^{(j_2)}(x) = \frac{\partial_{\alpha^{(j_2)}} a(x,\alpha^\star)}{c^2(x,\gamma^\star)}(A(x) - a(x,\alpha^\star)),$$

for $j_1 \in \{1,\dots,p_\gamma\}$ and $j_2 \in \{1,\dots,p_\alpha\}$.

The form of Σ (pure-jump Lévy driven case)

$$\Sigma_\gamma = 4\int_{\mathbb{R}}\int_{\mathbb{R}}\left(\frac{\partial_\gamma c(x,\gamma^\star)}{c^3(x,\gamma^\star)}C^2(x)z^2 + g_1(x+C(x)z) - g_1(x)\right)^{\otimes 2}\pi_0(dx)\nu_0(dz),$$

$$\begin{aligned}\Sigma_{\alpha\gamma} = -4\int_{\mathbb{R}}\int_{\mathbb{R}}&\left(\frac{\partial_\gamma c(x,\gamma^\star)}{c^3(x,\gamma^\star)}C^2(x)z^2 + g_1(x+C(x)z) - g_1(x)\right)\\ &\left(\frac{\partial_\alpha a(x,\alpha^\star)}{c^2(x,\gamma^\star)}C(x)z + g_2(x+C(x)z) - g_2(x)\right)^\top \pi_0(dx)\nu_0(dz),\end{aligned}$$

$$\Sigma_\alpha = 4\int_{\mathbb{R}}\int_{\mathbb{R}}\left(\frac{\partial_\alpha a(x,\alpha^\star)}{c^2(x,\gamma^\star)}C(x)z + g_2(x+C(x)z) - g_2(x)\right)^{\otimes 2}\pi_0(dx)\nu_0(dz),$$

where the functions g_1 and g_2 are the solution of the following extended Poisson equations:

$$\tilde{\mathcal{A}}g_1^{(j_1)}(x) \overset{EPE}{=} -\frac{\partial_{\gamma^{(j_1)}} c(x,\gamma^\star)}{c^3(x,\gamma^\star)}(c^2(x,\gamma^\star) - C^2(x)),$$

$$\tilde{\mathcal{A}}g_2^{(j_2)}(x) \overset{EPE}{=} -\frac{\partial_{\alpha^{(j_2)}} a(x,\alpha^\star)}{c^2(x,\gamma^\star)}(A(x) - a(x,\alpha^\star)),$$

for $j_1 \in \{1,\dots,p_\gamma\}$ and $j_2 \in \{1,\dots,p_\alpha\}$ (In the correctly specified case, g_1 and g_2 are identically 0).

Numerical experiment

We suppose that the data-generating model is the following Lévy driven Ornstein-Uhlenbeck model:

$$dX_t = -\frac{1}{2}X_t dt + dZ_t, \quad X_0 = 0,$$

and that the parametric model is described as:

$$dX_t = \alpha(1 - X_t)dt + \frac{\gamma}{\sqrt{1 + X_t^2}} dZ_t, \quad \alpha, \gamma > 0.$$

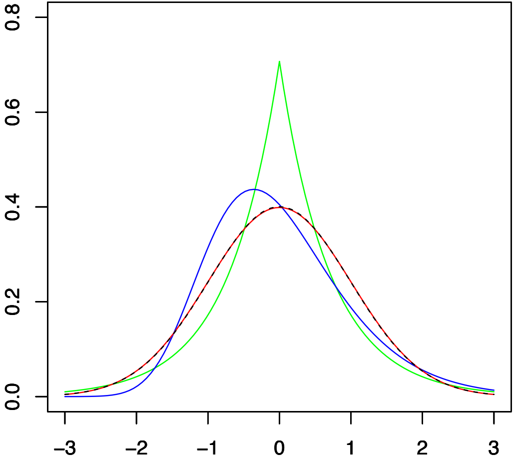
We conduct numerical experiments in the four situations:

1. $\mathcal{L}(Z_t) = NIG(10, 0, 10t, 0)$,
2. $\mathcal{L}(Z_t) = bGamma(t, \sqrt{2}, t, \sqrt{2})$,
3. $\mathcal{L}(Z_t) = NIG(25/3, 20/3, 9/5t, -12/5t)$,
4. $\mathcal{L}(Z_t) = N(0, t)$.

We generated 10000 paths of each SDE based on Euler-Maruyama scheme and constructed the estimators.

Density plot at $t = 1$

Figure: (i) $NIG(10, 0, 10, 0)$ (black dotted line), (ii) $bGamma(1, \sqrt{2}, 1, \sqrt{2})$ (green line), (iii) $NIG(25/3, 20/3, 9/5, -12/5)$ (blue line), and $N(0, 1)$ (red line).



Estimators

Solving the corresponding estimating equations, the GQMLE are calculated as:

$$\hat{\alpha}_n = -\frac{\sum_{j=1}^{n}(X_{j-1}-1)(X_j - X_{j-1})(X_{j-1}^2+1)}{h_n\sum_{j=1}^{n}(X_j-1)^2(X_{j-1}^2+1)},$$

$$\hat{\gamma}_n = \sqrt{\frac{1}{T_n}\sum_{j=1}^{n}(X_j - X_{j-1})^2(X_{j-1}^2+1)}.$$

Optimal values

Since

$$\mathbb{G}_1(\gamma) = -2\log\gamma - \frac{2}{\gamma^2} + \int_{\mathbb{R}} \log(1+x^2)\pi_0(dx),$$

$$\mathbb{G}_2(\alpha) = -\frac{1}{\gamma^\star}\left[\frac{1}{4}\int_{\mathbb{R}} x^3\pi_0(dx) + \alpha\left(1 - \int_{\mathbb{R}} x^3\pi_0(dx) + \int_{\mathbb{R}} x^4\pi_0(dx)\right) + \alpha^2\left(3 - 2\int_{\mathbb{R}} x^3\pi_0(dx) + \int_{\mathbb{R}} x^4\pi_0(dx)\right)\right],$$

the optimal values $\gamma^\star$ and $\alpha^\star$ are calculated as

$$\gamma^\star = \sqrt{2},$$

$$\alpha^\star = \frac{1 - \int_{\mathbb{R}} x^3\pi_0(dx) + \int_{\mathbb{R}} x^4\pi_0(dx)}{2(3 - 2\int_{\mathbb{R}} x^3\pi_0(dx) + \int_{\mathbb{R}} x^4\pi_0(dx))}.$$

Results

Table: The performance of our estimators; the mean is given with the standard deviation in parenthesis. The target optimal values are given in the first line of each items.

T_n	n	h_n	(i) (0.33,1.41)		(ii) (0.37, 1.41)		(iii) (0.37, 1.41)		diffusion (0.33, 1.41)	
			$\hat{\alpha}_n$	$\hat{\gamma}_n$	$\hat{\alpha}_n$	$\hat{\gamma}_n$	$\hat{\alpha}_n$	$\hat{\gamma}_n$	$\hat{\alpha}_n$	$\hat{\gamma}_n$
50	1000	0.05	0.38	1.41	0.40	1.39	0.40	1.39	0.38	1.41
			(0.12)	(0.11)	(0.16)	(0.29)	(0.15)	(0.19)	(0.13)	(0.10)
100	5000	0.02	0.37	1.41	0.39	1.39	0.38	1.39	0.36	1.41
			(0.09)	(0.08)	(0.11)	(0.23)	(0.11)	(0.15)	(0.09)	(0.08)
100	10000	0.01	0.36	1.41	0.37	1.39	0.38	1.40	0.36	1.41
			(0.08)	(0.07)	(0.09)	(0.22)	(0.10)	(0.15)	(0.08)	(0.07)

Summary of Gaussian quasi-likelihood estimator

- Even though the model is misspecified, the Gaussian quasi-likelihood estimator has the consistency and asymptotic normality.
- However, it is hard to construct a consistent estimator of its asymptotic variance due to the solution of the (extended) Poisson equations which is essential to correct the misspecification bias.
- To conduct fundamental statistical methods, we need to approximate $A_n \mathcal{I}_n^{1/2}(\hat{\theta}_n - \theta^\star)$.

Model setting

Gaussian quasi-likelihood estimation

- Diffusion case
- Pure-jump Lévy driven case

Bootstrap method for block sum

- Pure-jump Lévy driven case
- Diffusion case

Modified bootstrap method

Summary

Estimating equations

- To approximate the distribution of $A_n \mathcal{I}_n^{1/2}(\hat{\theta}_n - \theta^\star)$, we consider a bootstrap method.
- Regard the Gaussian quasi-likelihood estimator as a root of the estimating equations:

$$\sum_{j=1}^{n} \frac{\partial_\gamma c_{j-1}(\gamma)}{c_{j-1}^3(\gamma)} \left\{ h_n c_{j-1}^2(\gamma) - (\Delta_j X)^2 \right\} = 0,$$

$$\sum_{j=1}^{n} \frac{\partial_\alpha a_{j-1}(\alpha)}{c_{j-1}^2(\hat{\gamma}_n)} \left(\Delta_j X - h_n a_{j-1}(\alpha) \right) = 0.$$

Weighted bootstrap method for estimating equations (Chatterjee and Bose (2005))

- We now consider the situation where samples $\{Y_j\}_{j=1}^n$ are in hand.
- We define Z-estimator $\hat{\beta}_n$ as a root of the estimating equation:

$$\sum_{j=1}^{n} \psi(Y_j, \beta) = 0,$$

where ψ is an appropriate function, $\beta^\star$ is the optimal value, and $(\psi(Y_j, \beta^\star))_{j=1}^n$ is a martingale difference.

- The consistency and asymptotic normality of $\hat{\beta}_n$ can be shown under sufficient regularity conditions.

Weighted bootstrap estimator

- We define the weighted bootstrap estimator $\hat{\beta}_n^{\mathbf{B}}$ by a root of

$$\sum_{j=1}^{n} w_j \psi(Y_j, \beta) = 0,$$

  where the bootstrap weights $(w_j)_{j=1}^n$ is i.i.d. random variables being independent of $(Y_j)_{j=1}^n$ and satisfies

$$E[w_1] = 1, Var[w_1] = 1, E[w_1^4] < \infty.$$

- We write $P_{\mathbf{B}}$ as the bootstrap probability measure conditioned by the observed data.

$$F(x) := P\left(\left[\sum_{j=1}^{n} \partial_\beta \psi(Y_j, \hat{\beta}_n)\right]^{1/2} (\hat{\beta}_n - \beta^\star) \le x\right),$$

$$F^{\mathsf{B}}(x) := P_{\mathsf{B}}\left(\left[\sum_{j=1}^{n} \partial_\beta \psi(Y_j, \hat{\beta}_n)\right]^{1/2} (\hat{\beta}_n^{\mathsf{B}} - \hat{\beta}_n) \le x\right).$$

Under sufficient regularity conditions, we have

Chatterjee and Bose (2005), Theorem 3.2

$$\left[\sum_{j=1}^{n} \partial_\beta \psi(Y_j, \hat{\beta}_n)\right]^{1/2} (\hat{\beta}_n^{\mathsf{B}} - \hat{\beta}_n) = -a_n^{-1} \sum_{j=1}^{n} (w_j - 1)\psi(Y_j, \hat{\beta}_n) + r_{n,\mathsf{B}}, \quad (2)$$

where a_n is the convergence rate of $\hat{\beta}_n$, and $r_{n,\mathsf{B}}$ is a random variable such that for any $\epsilon > 0$, $P_{\mathsf{B}}(|r_{nB}| > \epsilon) = o_p(1)$. Furthermore, it follows that

$$\sup_{x\in\mathbb{R}} \left|F(x) - F^{\mathsf{B}}(x)\right| \xrightarrow{p} 0.$$

Bootstrap estimator

- With the bootstrap weights $\{w_j\}_{j=1}^n$, consider the bootstrap estimator $\tilde{\theta}_n^{\mathsf{B}} := (\tilde{\gamma}_n^{\mathsf{B}}, \tilde{\alpha}_n^{\mathsf{B}})$ defined by a root of

$$\sum_{j=1}^n w_j \frac{\partial_\gamma c_{j-1}(\gamma)}{c_{j-1}^3(\gamma)} \left[h_n c_{j-1}^2(\gamma) - (\Delta_j X)^2\right] = 0,$$

$$\sum_{j=1}^n w_j \frac{\partial_\alpha a_{j-1}(\alpha)}{c_{j-1}^2(\hat{\gamma}_n)} \left(\Delta_j X - h_n a_{j-1}(\alpha)\right) = 0.$$

- $\left\{\frac{\partial_\gamma c_{j-1}(\gamma^\star)}{c_{j-1}^3(\gamma^\star)} \left[h_n c_{j-1}^2(\gamma^\star) - (\Delta_j X)^2\right]\right\}_{j=1}^n$ and $\left\{\frac{\partial_\alpha a_{j-1}(\alpha^\star)}{c_{j-1}^2(\hat{\gamma}_n)} \left(\Delta_j X - h_n a_{j-1}(\alpha^\star)\right)\right\}_{j=1}^n$ are asymptotically martingale diffenrece.

Inconsistency

- We consider the misspecified pure-jump Lévy driven case.
- $\Phi(x,\mu,\Sigma)$: the cumulative distribution function of the normal distribution whose mean and variance are μ and Σ, respectively.

$$\sup_{x\in\mathbb{R}}\left|P^{\mathsf{B}}\left(\sqrt{T_n}\hat{\mathcal{I}}_n^{1/2}(\tilde{\theta}_n^{\mathsf{B}}-\hat{\theta}_n)\le x\right)-\Phi(x,0,\Sigma^{\mathsf{spec}})\right|\xrightarrow{p}0,$$

where $\Sigma^{\mathsf{spec}}:=\begin{pmatrix}\Sigma'_{\gamma} & \Sigma'_{\alpha\gamma}\\ \Sigma'^{\top}_{\alpha\gamma} & \Sigma'_{\alpha}\end{pmatrix}\neq\Sigma^{\mathsf{miss}}$ is defined by

$$\Sigma'_{\gamma}=\int_{\mathbb{R}}\int_{\mathbb{R}}\left(\frac{\partial_\gamma c(x,\gamma^\star)}{c^3(x,\gamma^\star)}C^2(x)z^2\right)^{\otimes 2}\pi_0(dx)\nu_0(dz),$$

$$\Sigma'_{\alpha\gamma}=-\int_{\mathbb{R}}\int_{\mathbb{R}}\left(\frac{\partial_\gamma c(x,\gamma^\star)}{c^3(x,\gamma^\star)}C^2(x)z^2\right)\left(\frac{\partial_\alpha a(x,\alpha^\star)}{c^2(x,\gamma^\star)}C(x)z\right)^{\top}\pi_0(dx)\nu_0(dz),$$

$$\Sigma'_{\alpha}=\int_{\mathbb{R}}\int_{\mathbb{R}}\left(\frac{\partial_\alpha a(x,\alpha^\star)}{c^2(x,\gamma^\star)}C(x)z\right)^{\otimes 2}\pi_0(dx)\nu_0(dz).$$

- Index(j)-wise weighted bootstrap method does not reflect the effect of the model misspecification.

Weighted bootstrap method for block sum

$$\text{Data-generating model}: dX_t = A(X_t)dt + C(X_{t-})dZ_t,$$
$$\text{Statistical model}: dX_t = a(X_t,\alpha)dt + c(X_{t-},\gamma)dZ_t$$

- We now consider weighted bootstrap method for block sum to reflect the model misspecification.
- We divide $\{1,\dots,n\}$ into k_n-blocks $(B_i)_{i=1}^{k_n}$ defined by:

$$B_i := \{j \in \{1,\dots,n\} : (i-1)c_n + 1 \le j \le ic_n\},$$

where $c_n = \frac{n}{k_n}$, and here c_n is supposed to be a positive integer for simplicity.

Block weighted bootstrap estimator

- Let the bootstrap weights $(w_j)_{j=1}$ be i.i.d. random variables being independent of $X = (X_t)_{t \geq t}$ and satisfies
$$E[w_1] = 1, Var[w_1] = 1, E[w_1^4] < \infty.$$
- With the bootstrap weights $\{w_i\}_{i=1}^{k_n}$, we define weighted block bootstrap estimator $\hat{\theta}_n^{\mathsf{B}} := (\hat{\gamma}_n^{\mathsf{B}}, \hat{\alpha}_n^{\mathsf{B}})$ as a root of

$$\sum_{i=1}^{k_n} w_i \sum_{j \in B_{k_i}} \frac{\partial_\gamma c_{j-1}(\gamma)}{c_{j-1}^3(\gamma)} \left[h_n c_{j-1}^2(\gamma) - (\Delta_j X)^2 \right] = 0,$$
$$\sum_{i=1}^{k_n} w_i \sum_{j \in B_{k_i}} \frac{\partial_\alpha a_{j-1}(\alpha)}{c_{j-1}^2(\hat{\gamma}_n)} (\Delta_j X - h_n a_{j-1}(\alpha)) = 0.$$

Asymptotic result (pure-jump Lévy driven case)

- $\xi(s,z) = (\xi_1(s,z), \xi_2(s,z))$, where

$$\xi_1(s,z) = \frac{\partial_\gamma c_{s-}(\gamma^\star)}{c_{s-}^3(\gamma^\star)} C_{s-}^2 z^2 + g_1(X_{s-} + C_{s-}z) - g_1(X_{s-}),$$

$$\xi_2(s,z) = \frac{\partial_\alpha a_{s-}(\alpha^\star)}{c_{s-}^2(\gamma^\star)} C_{s-} z + g_2(X_{s-} + C_{s-}z) - g_2(X_{s-}).$$

Theorem

1. Stochastic expansion:

$$A_n \hat{\mathcal{I}}_n^{1/2} (\hat{\theta}_n^{\mathsf{B}} - \hat{\theta}_n) = A_n^{-1} \sum_{i=1}^{k_n} (w_i - 1) \int_{(i-1)c_n h_n}^{i c_n h_n} \int_{\mathbb{R}} 2\xi(s,z) \tilde{N}(ds,dz) + r_{nB}.$$

2. Approximation:

$$\sup_{x\in\mathbb{R}} \left| P^{\mathsf{B}} \left(A_n \hat{\mathcal{I}}_n^{1/2} \left(\hat{\theta}_n^{\mathsf{B}} - \hat{\theta}_n \right) \le x \right) - P \left(A_n \hat{\mathcal{I}}_n^{1/2} \left(\hat{\theta}_n - \theta^\star \right) \le x \right) \right| \xrightarrow{p} 0.$$

Asymptotic result (misspecified diffusion case)

- $\xi(s) = (\xi_1(s), \xi_2(s))$, where

$$\xi_1(s) = \partial_x f_1(X_s) C_s, \quad \xi_2(s) = \frac{\partial_\alpha a_s(\alpha^\star) - \partial_x f_2(X_s)}{c_s^2(\gamma^\star)} C_s.$$

Theorem

1. Stochastic expansion:

$$\begin{aligned} &A_n \hat{\mathcal{I}}_n^{1/2}(\hat{\theta}_n^{\mathsf{B}} - \hat{\theta}_n) \\ &= A_n^{-1} \sum_{i=1}^{k_n} (w_i - 1) \int_{(i-1)c_n h_n}^{i c_n h_n} 2\xi(s) dw_s + r_{nB}. \end{aligned}$$

2. Approximation:

$$\sup_{x \in \mathbb{R}} \left| P^{\mathsf{B}} \left(A_n \hat{\mathcal{I}}_n^{1/2} \left(\hat{\theta}_n^{\mathsf{B}} - \hat{\theta}_n \right) \le x \right) - P \left(A_n \hat{\mathcal{I}}_n^{1/2} \left(\hat{\theta}_n - \theta^\star \right) \le x \right) \right| \xrightarrow{p} 0.$$

Asymptotic result (correctly specified diffusion case)

- Recall that A_n becomes $\mathbf{diag}\{\sqrt{n}I_{p_\gamma}, \sqrt{T_n}I_{p_\alpha}\}$ only in this case.
- Let B_n be $\mathbf{diag}\{\sqrt{n}h_n I_{p_\gamma}, \sqrt{T_n}I_{p_\alpha}\}$.
- $\xi(s) = (\xi_1(s), \xi_2(s))$, where

$$\xi_1(s) = 2\frac{\partial_\gamma c_s(\gamma^\star)}{C_s}w_s, \quad \xi_2(s) = \frac{\partial_\alpha a_s(\alpha^\star)}{C_s}.$$

Theorem

1. Stochastic expansion:

$$A_n\hat{\mathcal{I}}_n^{1/2}(\hat{\theta}_n^{\mathsf{B}} - \hat{\theta}_n) = B_n^{-1}\sum_{i=1}^{k_n}(w_i - 1)\int_{(i-1)c_n h_n}^{ic_n h_n} 2\xi(s)dw_s + r_{nB}.$$

2. Approximation:

$$\sup_{x\in\mathbb{R}}\left|P^{\mathsf{B}}\left(A_n\hat{\mathcal{I}}_n^{1/2}\left(\hat{\theta}_n^{\mathsf{B}} - \hat{\theta}_n\right) \le x\right) - P\left(A_n\hat{\mathcal{I}}_n^{1/2}\left(\hat{\theta}_n - \theta^\star\right) \le x\right)\right| \xrightarrow{p} 0.$$

- However, this bootstrap method is not a unified one since we cannot identify A_n in practice.

Balancing term

$$\tilde{b}_{2,n} := \frac{1}{n}\sum_{j=1}^{n}\left[\frac{(\Delta_j X)^4}{3h_n^2} - \frac{2(\Delta_j X)^2 c_{j-1}^2(\hat{\gamma}_n)}{h_n} + c_{j-1}^4(\hat{\gamma}_n)\right]$$

- To solve the problem, we introduce the balancing term $b_n := b_{1,n} + b_{2,n}$ defined by

$$b_{1,n} := \frac{\sum_{j=1}^{n}(\Delta_j X)^4}{\sum_{j=1}^{n}(\Delta_j X)^2},$$

$$b_{2,n} := \exp\left[-\left(\left|\tilde{b}_{2,n}\right| + \left|\tilde{b}_{2,n}\right|^{-1}\right)\right].$$

The role of $b_{1,n}$

- In the diffusion case,

$$\frac{b_{1,n}}{3h_n} \xrightarrow{p} \frac{\int C^4(x)\pi_0(dx)}{\int C^2(x)\pi_0(dx)}.$$

- In the pure-jump Lévy driven case,

$$b_{1,n} \xrightarrow{p} \frac{\int C^4(x)\pi_0(dx) \int z^4\nu_0(dz)}{\int C^2(x)\pi_0(dx)}.$$

⇒The term $b_{1.n}$ distinguishes whether the driving noise is a standard Wiener process or pure-jump Lévy process.

The role of $b_{2,n}$

- In the pure-jump Lévy driven case, $\tilde{b}_{2,n} \to \infty$.
- In the diffusion case,

$$\tilde{b}_{2,n} \xrightarrow{p} \int (C^2(x) - c^2(x, \gamma^\star))^2 \pi_0(dx) =: b_2.$$

  Hence, in the correctly specified case, $b_2 = 0$, and in the misspecified case, $b_2 \neq 0$.

- When $x \to 0$ and $x \to \infty$, the function $h(x) := \exp[-(|x| + |x|^{-1})]$ to 0.

$\Rightarrow b_{2,n}$ distinguishes whether the misspecified diffusion case or not.

Asymptotic behavior of b_n

Proposition

1. In the correctly specified diffusion case,

$$\frac{b_n}{3h_n} \xrightarrow{p} \frac{\int c^4(x,\gamma^\star)\pi_0(dx)}{\int c^2(x,\gamma^\star)\pi_0(dx)}.$$

2. In the misspecified diffusion case,

$$b_n \xrightarrow{p} \exp\left[-\left(b_2 + b_2^{-1}\right)\right] \neq 0.$$

3. In the pure-jump Lévy driven case,

$$b_n \xrightarrow{p} \frac{\int c^4(x,\gamma^\star)\pi_0(dx) \int z^4 \nu_0(dz)}{\int c^2(x,\gamma^\star)\pi_0(dx)}.$$

⇒Only in the correctly specified case, the convergence rate of b_n is h_n.

Modified bootstrap method

Let $\hat{A}_n := \mathrm{diag}\left\{\sqrt{\frac{T_n}{b_n}} I_{p_\gamma}, \sqrt{T_n} I_{p_\alpha}\right\}$, and $\hat{B}_n := \mathrm{diag}\left\{\sqrt{T_n b_n} I_{p_\gamma}, \sqrt{T_n} I_{p_\alpha}\right\}$.

- We consider the approximation of $\hat{A}_n \hat{\mathcal{I}}_n^{1/2}(\hat{\theta}_n - \theta^\star)$ instead of $A_n \hat{\mathcal{I}}_n^{1/2}(\hat{\theta}_n - \theta^\star)$.

Theorem

For $\delta \in (\frac{1}{2}, 1)$, suppose that $k_n = O(T_n^\delta)$.

$$
\begin{aligned}
&\hat{A}_n \hat{\mathcal{I}}_n^{1/2}(\hat{\theta}_n^{\mathsf{B}} - \hat{\theta}_n) \\
&= \hat{B}_n^{-1} \sum_{i=1}^{k_n} (w_i - 1) \sum_{j \in B_{k_i}} \begin{pmatrix} \frac{\partial_\gamma c_{j-1}(\hat{\gamma}_n)}{c_{j-1}^3(\hat{\gamma}_n)} \left\{ h_n c_{j-1}^2(\hat{\gamma}_n) - (\Delta_j X)^2 \right\} \\ \frac{\partial_\alpha a_{j-1}(\hat{\alpha}_n)}{c_{j-1}^2(\hat{\gamma}_n)} \left(\Delta_j X - h_n a_{j-1}(\hat{\alpha}_n) \right) \end{pmatrix} + r_{nB}.
\end{aligned}
$$

Approximation of the distribution

Theorem

In all cases, we have

$$\sup_{x \in \mathbb{R}} \left| P^{\mathsf{B}} \left(\hat{A}_n \hat{\mathcal{I}}_n^{1/2} \left(\hat{\theta}_n^{\mathsf{B}} - \hat{\theta}_n \right) \leq x \right) - P \left(\hat{A}_n \hat{\mathcal{I}}_n^{1/2} \left(\hat{\theta}_n - \theta^{\star} \right) \leq x \right) \right| \xrightarrow{p} 0.$$

- Thanks to this theorem, we can construct confidence intervals and hypothesis testing based on the bootstrap quantile $c_{n,q'}^{\mathsf{B}}$, and it has theoretical validity.

Remark

- To calculate $c^{\mathsf{B}}_{n,q}$, we need to generate L weighted bootstrap estimators $\left\{\hat{\theta}^{\mathsf{B}}_{n,l}\right\}_{l=1}^{L}$ for large $L \in \mathbb{N}$.
- The stochastic expansion suggests that in order to obtain $c^{\mathsf{B}}_{n,q}$, it suffices to generate the bootstrapped quasi-score function

$$\hat{B}_n^{-1}\sum_{i=1}^{k_n}(w_{i,l}-1)\sum_{j\in B_{k_i}}\begin{pmatrix}\frac{\partial_\gamma c_{j-1}(\hat{\gamma}_n)}{c^3_{j-1}(\hat{\gamma}_n)}\left\{h_n c^2_{j-1}(\hat{\gamma}_n)-(\Delta_j X)^2\right\}\\ \frac{\partial_\alpha a_{j-1}(\hat{\alpha}_n)}{c^2_{j-1}(\hat{\gamma}_n)}\left(\Delta_j X - h_n a_{j-1}(\hat{\alpha}_n)\right)\end{pmatrix},$$

instead of $\hat{A}_n\hat{\Gamma}_n\left(\hat{\theta}^{\mathsf{B}}_{n,l}-\hat{\theta}_n\right)$. Importantly, its generation only require the optimization to get $\hat{\theta}_n$ while calculating $\sqrt{T_n}\hat{\Gamma}_n\left(\hat{\theta}^{\mathsf{B}}_{n,l}-\hat{\theta}_n\right)$ entails some optimization method such as quasi-Newton method for each l, thus resulting much smaller computational effort.

Summary

$$\text{Data-generating model}: dX_t = A(X_t)dt + C(X_{t-})dZ_t,$$
$$\text{Statistical model}: dX_t = a(X_t, \alpha)dt + c(X_{t-}, \gamma)dZ_t$$

- We present a constructible random vector which approximates the distribution of the Gaussian quasi-likelihood estimator by the weighted block bootstrap method.
- By introducing a balancing term, our method can be applied to all cases without the specification of the case.
- Problem: How to choose the block size?