Annals of the Institute of Statistical Mathematics (2023) 75:533–565
https://doi.org/10.1007/s10463-022-00854-2

Bootstrap method for misspecified ergodic Lévy driven stochastic differential equation models

Yuma Uehara[1]

Received: 31 March 2022 / Revised: 2 August 2022 / Accepted: 21 September 2022 /
Published online: 10 November 2022

Abstract
In this paper, we consider possibly misspecified stochastic differential equation models driven by Lévy processes. Regardless of whether the driving noise is Gaussian or not, Gaussian quasi-likelihood estimator can estimate unknown parameters in the drift and scale coefficients. However, in the misspecified case, the asymptotic distribution of the estimator varies by the correction of the misspecification bias, and consistent estimators for the asymptotic variance proposed in the correctly specified case may lose theoretical validity. As one of its solutions, we propose a bootstrap method for approximating the asymptotic distribution. We show that our bootstrap method theoretically works in both correctly specified case and misspecified case without assuming the precise distribution of the driving noise.

Keywords Misspecified model · Lévy driven stochastic differential equation · Bootstrap method

1 Introduction

We suppose that the data-generating structure is the following one-dimensional stochastic differential equation

$$dX_t = A(X_t)dt + C(X_{t-})dZ_t, \tag{1}$$

defined on a stochastic basis $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, P)$ where

This work was supported by JST CREST Grant Number JPMJCR14D7, Japan, and JSPS KAKENHI Grant Number JP19K20230.

✉ Yuma Uehara
y-uehara@kansai-u.ac.jp

[1] Department of Mathematics, Faculty of Engineering Science, Kansai University, 3-3-35 Yamate-cho, Suita-city, Osaka 564-8680, Japan

- the driving noise Z can either be a standard Wiener process or a pure-jump Lévy process;
- $\mathcal{F}_t = \sigma(X_0) \vee \sigma(Z_s; s \leq t)$;
- The initial variable X_0 is $\mathcal{F}_0$-measurable.

In this paper, we consider the situation where high-frequency samples $\mathbb{X} := (X_{t_j})_{j=0}^n$ from the solution path X are obtained in the so-called "rapidly increasing design": $t_j = t_j^n := jh_n$, $T_n := nh_n \to \infty$, and $nh_n^2 \to 0$. To deal with the effect of model misspecification being inevitable in statistical modeling, we consider the following parametric model on $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, P)$:

$$dX_t = a(X_t, \alpha)dt + c(X_{t-}, \gamma)dZ_t. \tag{2}$$

Here, the functional forms of drift coefficient $a : \mathbb{R} \times \Theta_\alpha \mapsto \mathbb{R}$, and scale coefficient $c : \mathbb{R} \times \Theta_\gamma \mapsto \mathbb{R}$ are supposed to be known except for the drift parameter α and scale parameter γ. We also suppose that α and γ belong to bounded convex domains $\Theta_\alpha \subset \mathbb{R}^{p_\alpha}$ and $\Theta_\gamma \subset \mathbb{R}^{p_\gamma}$, respectively. We note that the coefficients are possibly misspecified, that is, the parametric family $\{(a(\cdot, \alpha), c(\cdot, \gamma)); \alpha \in \Theta_\alpha, \gamma \in \Theta_\gamma\}$ does not include the true coefficients $(A(\cdot), C(\cdot))$. From now on, the terminologies "misspecification," "misspecified," and "correctly specified" will be used for the above meaning unless otherwise mentioned. In this framework, there are four possible cases:

1. Correctly specified diffusion case: the driving noise Z is a standard Wiener process and the coefficients are correctly specified;
2. Misspecified diffusion case: the driving noise Z is a standard Wiener process and the coefficients are missspecified;
3. Correctly specified pure-jump Lévy driven case: the driving noise Z is a pure-jump Lévy process and the coefficients are correctly specified;
4. Misspecified pure-jump Lévy driven case: the driving noise Z is a pure-jump Lévy process and the coefficients are missspecified.

For the estimation of θ, we consider Gaussian quasi-likelihood estimation. It is a tractable and powerful tool for estimating mean and variance structure in the sense that we need not to assume the precise distribution of the error variable. For various statistical models including (2), its theoretical property has been analyzed. In particular, for correctly specified diffusion models, the asymptotic behavior of the Gaussian quasi-maximum likelihood estimator (GQMLE) is verified for example by Yoshida (1992); Genon-Catalot and Jacod (1993), and Kessler (1997). As for correctly specified non-Gaussian Lévy driven SDE models, Masuda (2013) clarified the theoretical property of the GQMLE. Although its convergence rate is slower than the correctly specified diffusion case, it still has the consistency and asymptotic normality. The papers Uchida and Yoshida (2011) and Uehara (2019) extended the results to the case where the drift and (or) scale coefficient are (is) misspecifed. In these papers, the misspecification bias is handled by the theory of (extended) Poisson equation, and inevitably the asymptotic distribution of the GQMLE contains the solution. Hence, the estimators of the asymptotic variance which have been proposed for the correctly specified case does not work in the misspecified case. As a result, the confidence intervals and hypothesis testing based on the estimators no longer have theoretical validity in the misspecified case. This is a serious problem since in practice, we cannot avoid the risk of model misspecification. The primary object of this paper is to overcome this issue.

When it is tough to evaluate the asymptotic distribution of some statistic directly, bootstrap methods originally introduced by Efron (1979) often serve as a good prescription. As for high-frequently observed settings, bootstrap methods also do and indeed for various purposes such as estimating realized volatility distribution (Gonçalves and Meddahi 2009), making statistical inference in jump regressions (Li et al. 2017), executing jump tests (Dovonon et al. 2019), kinds of the methods have been proposed. In this paper, we follow this direction. We construct a block bootstrap Gaussian quasi-score function which can uniformly approximate the asymptotic distribution of the GQMLE both in the correctly specified and misspecified case. More specifically, we divide $\{1,\dots,n\}$ (n denotes the sample size) into k_n blocks, and for each block, we generate a bootstrap weight. Based on the weights, we construct the bootstrap score function and estimator. Furthermore, by introducing a adjustment term, our method can uniformly approximate the asymptotic distribution without specifying the distribution of the driving noise although the convergence rate of the scale parameter is different in the correctly specified diffusion case.

Here, we introduce some notations and conventions used throughout this paper. We largely abbreviate "n" from the notation like $t_j = t_j^n$ and $h = h_n$. For any vector variable $x = \left(x^{(i)}\right)$, $\frac{\partial}{\partial x^{(i)}}$ stands for the partial derivative with respect to the i-th component of x, and we write $\partial_x = \left(\frac{\partial}{\partial x^{(i)}}\right)_i$. I_d and O denote the d-dimensional identity matrix and zero matrix, respectively. $\top$ stands for the transpose operator, and $v^{\otimes 2} := vv^\top$ for any matrix v. The convergences in probability and in distribution are denoted by $\xrightarrow{p}$ and $\xrightarrow{\mathcal{L}}$, respectively. All limits appearing below are taken for $n \to \infty$ unless otherwise mentioned. For two nonnegative real sequences (a_n) and (b_n), we write $a_n \lesssim b_n$ if there exists a positive constant C and $N \in \mathbb{N}$ such that $a_n \leq Cb_n$ for any $n \geq N$. For any process Y, $\Delta_j Y$ denotes the j-th increment $Y_{t_j} - Y_{t_{j-1}}$. For any matrix-valued function f on $\mathbb{R} \times \Theta$, we write $f_s(\theta) = f(X_s, \theta)$. The Lévy measure of Z is written as $\nu_0(dz)$, and the associated compensated Poisson random measure is represented by $\tilde{N}(ds, dz)$. $\mathcal{A}$ and $\tilde{\mathcal{A}}$ stand for the infinitesimal generator and extended generator of X, respectively.

The rest of this paper is organized as follows. In Sect. 2, we provide a brief overview of the Gaussian quasi-likelihood estimation. We also introduce assumptions used throughout of this paper. Section 3 is the main body of this paper: First, we construct an adjustment term for uniformly dealing with the difference of the convergence rate of the scale parameter γ, and after that we propose our bootstrap method, and show its theoretical property. Section 4 presents the finite sample performance of our method. All of their proofs are given in Sect. 5.

2 Gaussian quasi-likelihood estimation and assumptions

Since the explicit form of the transition probability of X cannot be obtained in general, the estimation based on the genuine likelihood function is impractical. In this section, we briefly explain the Gaussian quasi-likelihood estimation for our model, and introduce assumptions for its asymptotic results and our main results. Building on the discrete-time approximation of (2):

$$X_{t_j} \approx X_{t_{j-1}} + h_n a_{t_{j-1}}(\alpha) + c_{t_{j-1}}(\gamma)\Delta_j Z,$$

we consider the stepwise Gaussian quasi-likelihood (GQL) function defined as follows:

$$\mathbb{H}_{1,n}(\gamma) = -\frac{1}{2h_n}\sum_{j=1}^{n}\left\{ h_n \log c_{t_{j-1}}^2(\gamma) + \frac{(\Delta_j X)^2}{c_{t_{j-1}}^2(\gamma)} \right\},$$

$$\mathbb{H}_{2,n}(\alpha,\gamma) = -\frac{1}{2h_n}\sum_{j=1}^{n}\frac{(\Delta_j X - h_n a_{t_{j-1}}(\alpha))^2}{c_{t_{j-1}}^2(\gamma)}.$$

Based on this GQL function, we define Gaussian quasi maximum likelihood estimator (GQMLE) $\hat{\theta}_n := (\hat{\gamma}_n, \hat{\alpha}_n)$ by

$$\hat{\gamma}_n \in \underset{\gamma\in\bar{\Theta}_\gamma}{\operatorname{argmax}}\, \mathbb{H}_{1,n}(\gamma),\ \ \hat{\alpha}_n \in \underset{\alpha\in\bar{\Theta}_\alpha}{\operatorname{argmax}}\, \mathbb{H}_{2,n}(\alpha,\hat{\gamma}_n).$$

We define an optimal parameter $\theta^\star := (\gamma^\star, \alpha^\star)$ of θ by

$$\gamma^\star \in \underset{\gamma\in\bar{\Theta}_\gamma}{\operatorname{argmax}}\, \mathbb{H}_1(\gamma), \quad \alpha^\star \in \underset{\alpha\in\bar{\Theta}_\alpha}{\operatorname{argmax}}\, \mathbb{H}_2(\alpha),$$

where $\mathbb{H}_1 : \Theta_\gamma \mapsto \mathbb{R}$ and $\mathbb{H}_2 : \Theta_\alpha \mapsto \mathbb{R}$ are defined as follows:

$$\mathbb{H}_1(\gamma) = -\frac{1}{2}\int_{\mathbb{R}}\left(\log c^2(x,\gamma) + \frac{C^2(x)}{c^2(x,\gamma)}\right)\pi_0(dx), \tag{3}$$

$$\mathbb{H}_2(\alpha) = -\frac{1}{2}\int_{\mathbb{R}} c^{-2}(x,\gamma^\star)(A(x) - a(x,\alpha))^2\pi_0(dx). \tag{4}$$

Now, we introduce the technical assumptions for our main results. Some comments on each assumption will be given after the all assumptions are mentioned. Recall that the parameter space Θ is supposed to be a bounded convex domain. We assume the following identifiability condition for $\mathbb{H}_1(\gamma)$ and $\mathbb{H}_2(\alpha)$:

Assumption 1 $\theta^\star \in \Theta$, and there exist positive constants χ_γ and χ_α such that for all $(\gamma,\alpha) \in \Theta$,

$$\mathbb{H}_1(\gamma) - \mathbb{H}_1(\gamma^\star) \le -\chi_\gamma|\gamma - \gamma^\star|^2, \tag{5}$$

$$\mathbb{H}_2(\alpha) - \mathbb{H}_2(\alpha^\star) \leq -\chi_\alpha |\alpha - \alpha^\star|^2. \tag{6}$$

In the rest of this paper, we sometimes omit the optimal value $\theta^\star$, for instance, the abbreviated symbols f_s and $f_{t_{j-1}}$ are used instead of $f_s(\theta^\star)$, and $f_{t_{j-1}}(\theta^\star)$, respectively.

Assumption 2

1. The coefficients A and C are Lipschitz continuous and twice differentiable, and their first and second derivatives are of at most polynomial growth.
2. The drift coefficient $a(\cdot, \alpha^\star)$ and scale coefficient $c(\cdot, \gamma^\star)$ are Lipschitz continuous, and $c(x, \gamma) \neq 0$ for every (x, γ).
3. For each $i \in \{0, 1\}$ and $k \in \{0, \ldots, 5\}$, the following conditions hold:
   - The coefficients a and c admit extension in $\mathcal{C}(\mathbb{R} \times \bar{\Theta})$ and have the partial derivatives $(\partial_x^i \partial_\alpha^k a, \partial_x^i \partial_\gamma^k c)$ possessing extension in $\mathcal{C}(\mathbb{R} \times \bar{\Theta})$.
   - There exists nonnegative constant $C_{(i,k)}$ satisfying

$$\sup_{(x,\alpha,\gamma)\in\mathbb{R}\times\bar{\Theta}_\alpha\times\bar{\Theta}_\gamma} \frac{1}{1+|x|^{C_{(i,k)}}} \Big\{ |\partial_x^i \partial_\alpha^k a(x, \alpha) | + |\partial_x^i \partial_\gamma^k c(x, \gamma)| + |c^{-1}(x, \gamma)| \Big\} < \infty. \tag{7}$$

Note that since we impose the extension condition in Assumption 2, $\mathbb{H}(\theta) := (\mathbb{H}_1(\gamma), \mathbb{H}_2(\alpha))$ admit extension in $\mathcal{C}(\bar{\Theta})$ as well.

For a function $\rho : \mathbb{R} \to \mathbb{R}_+$ and a signed measure m on a one-dimensional Borel space, we define

$$||m||_\rho = \sup\{|m(\rho)| : f \text{is} \mathbb{R} - \text{valued}, m - \text{measurable, and satisfies} |f| \leq \rho\}.$$

Assumption 3

1. 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

$$\sup_{t\in\mathbb{R}_+} \exp(at)||P_t(x, \cdot) - \pi_0(\cdot)||_{h_q} \leq C_q h_q(x), \tag{8}$$

   for any $x \in \mathbb{R}$ where $h_q(x) := 1 + |x|^q$.
2. For any $q > 0$, we have

$$\sup_{t\geq 0} E[|X_t|^q] < \infty. \tag{9}$$

We introduce a $p \times p$-matrix $\Gamma := \begin{pmatrix} \Gamma_\gamma & O \\ \Gamma_{\alpha\gamma} & \Gamma_\alpha \end{pmatrix}$ whose components are defined by:

$$\Gamma_\gamma := \int_{\mathbb{R}} \frac{\partial_\gamma^{\otimes 2} c(x,\gamma^\star)c(x,\gamma^\star) - (\partial_\gamma c(x,\gamma^\star))^{\otimes 2}}{c^4(x,\gamma^\star)}(C^2(x) - c^2(x,\gamma^\star))\pi_0(dx)$$
$$- 2\int_{\mathbb{R}} \frac{(\partial_\gamma c(x,\gamma^\star))^{\otimes 2}}{c^4(x,\gamma^\star)} C^2(x)\pi_0(dx),$$
$$\Gamma_{\alpha\gamma} := \int_{\mathbb{R}} \partial_\alpha a(x,\alpha^\star)\partial_\gamma^\top c^{-2}(x,\gamma^\star)(a(x,\alpha^\star) - A(x))\pi_0(dx),$$
$$\Gamma_\alpha := -\int_{\mathbb{R}} \frac{\partial_\alpha^{\otimes 2} a(x,\alpha^\star)}{c^2(x,\gamma^\star)}(a(x,\alpha^\star) - A(x))\pi_0(dx) - \int_{\mathbb{R}} \frac{(\partial_\alpha a(x,\alpha^\star))^{\otimes 2}}{c^2(x,\gamma^\star)}\pi_0(dx).$$

The matrix Γ is the probability limit of the Hessian matrix of the GQL function.

Assumption 4 $-\Gamma_\gamma$, $-\Gamma_\alpha$, and $-\Gamma$ are positive definite.

Next, we introduce the assumption on the driving noise. The diffusion case corresponds with (1), and the pure-jump Lévy driven case does with (2).

Assumption 5 Either (1) or (2) is satisfied.

1. **Diffusion case**: Z is a standard Wiener process (in this case, Z is often written as w). Furthermore, there exist the solutions of the following Poisson equations:

$$\mathcal{A}f_1(x) = \frac{\partial_\gamma c(x,\gamma^\star)}{c^3(x,\gamma^\star)}\left(C^2(x) - c^2(x,\gamma^\star)\right), \tag{10}$$

$$\mathcal{A}f_2(x) = \frac{\partial_\alpha a(x,\alpha^\star)}{c^2(x,\gamma^\star)}\left(A(x) - a(x,\alpha^\star)\right), \tag{11}$$

 and the solutions f_1 and f_2 are differentiable, and they and their derivatives are of at most polynomial growth.
2. **Pure-jump Lévy driven case**: Z is a pure-jump Lévy process satisfying
   - $E[Z_1] = 0$, $Var[Z_1] = 1$, and $E[|Z_1|^q] < \infty$ for all $q > 0$.
   - The Blumenthal–Getoor index (BG-index) of Z is smaller than 2, that is,

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

We make comments on our assumptions below.

- Under Assumption 2 and Assumption 5, the existence and uniqueness of the strong solution of (2) and its Markov and time-homogeneous property are guaranteed (cf. Applebaum 2009, Sect. 6).
- The Poisson equations (10) and (11) will play important role in dealing with the misspecification bias (cf. Uchida and Yoshida 2011). A sufficient condition for

the existence and regularity of f_1 and f_2 is given for example in Pardoux and Veretennikov (2001). Moreover, in one-dimensional case, the explicit forms of f_1 and f_2 are presented in (Uchida and Yoshida 2011, Remark 2.2).

- In the pure-jump Lévy driven case, a similar misspecification bias also exists. However, we cannot follow the same way as the diffusion case: the generator $\mathcal{A}$ of X is given by

$$\mathcal{A}f(x) = A(x)\partial_x f(x) + \int_{\mathbb{R}} \big(f(x + C(x)z) - f(x) - \partial_x f(x)C(x)z\big)\nu_0(dz),$$

for a suitable function f. The integral with respect to ν_0 makes it difficult to ensure the existence and regularity of the solutions of equations like $\mathcal{A}f = g$ with some functions f and g. Here, g corresponds with the misspecification bias term. Alternatively, as in Uehara (2019), we invoke the theory of extended Poisson equations introduced by Kulik and Veretennikov (2011) to deal with the misspecification bias. Its definition is as follows:

Definition 6 (Kulik and Veretennikov 2011, Definition 2.1) We say that a measurable function $f : \mathbb{R} \to \mathbb{R}$ 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 $g : \mathbb{R} \to \mathbb{R}$ such that the process

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

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 (f, g), we write $f \in \operatorname{Dom}(\tilde{\mathcal{A}})$ and $\tilde{\mathcal{A}}f \overset{EPE}{=} g$.

As for the Lévy driven SDE case, the Feller property holds under Assumption 2 (cf. Masuda 2007, 3.1.1 (ii)), and we consider the following extended Poisson equations:

$$\tilde{\mathcal{A}}g_1(x) = -\frac{\partial_\gamma c(x, \gamma^\star)}{c^3(x, \gamma^\star)}\big(c^2(x, \gamma^\star) - C^2(x)\big), \tag{12}$$

$$\tilde{\mathcal{A}}g_2(x) = -\frac{\partial_\alpha a(x, \alpha^\star)}{c^2(x, \gamma^\star)}\big(A(x) - a(x, \alpha^\star)\big). \tag{13}$$

(Uehara 2019, Proposition 3.5) shows the existence and uniqueness of g_1 and g_2 under Assumption 2. Assumption 3 and Assumption 5-(2). Although the regularity of g_1 and g_2 is not obtained except for the limited case, its weighted Hölder continuity is also ensured under the same assumptions, and it is enough for our asymptotic result. For more details, see the discussion in (Uehara 2019, Sect. 3).

Building on these assumptions, we can derive the asymptotic normality of $\hat{\theta}_n$. For its technical details, we refer to the references presented in the next theorem.

Theorem 7 *Under Assumptions 1–5, we can deduce the consistency and asymptotic normality of the GQMLE* $\hat{\theta}_n$: $\hat{\theta}_n \xrightarrow{p} \theta^\star$, *and*

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

where A_n *denotes the rate matrix of* $\hat{\theta}_n$ (*cf. Table* 1). *In each case,*

$$\Sigma := \begin{pmatrix} \Sigma_\gamma & \Sigma_{\alpha\gamma} \\ \Sigma_{\alpha\gamma}^\top & \Sigma_\alpha \end{pmatrix}$$

is explicitly given below:

- Correctly specified diffusion case (Kessler 1997; Uchida and Yoshida 2012):

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

- Misspecified diffusion case (Uchida and Yoshida 2011):

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

- Correctly specified pure-jump Lévy driven case (Masuda 2013; Masuda and Uehara 2017)

Table 1 GQL approach for ergodic diffusion models and ergodic Lévy driven SDE models

Model	Rates of convergence		References
	drift	Scale	
Correctly specified diffusion	$\sqrt{T_n}$	$\sqrt{n}$	Kessler (1997), Uchida and Yoshida (2012)
Misspecified diffusion	$\sqrt{T_n}$	$\sqrt{T_n}$	Uchida and Yoshida (2011)
Correctly specified Lévy driven SDE	$\sqrt{T_n}$	$\sqrt{T_n}$	Masuda (2013), Masuda and Uehara (2017)
Cisspecified Lévy driven SDE	$\sqrt{T_n}$	$\sqrt{T_n}$	Uehara (2019)

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

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

$$\Sigma_\alpha = \int_{\mathbb{R}} \left(\frac{\partial_\alpha a(x,\alpha^\star)}{c(x,\gamma^\star)} \right)^{\otimes 2} \pi_0(dx).$$

- Misspecified pure-jump Lévy driven case (Uehara 2019)

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

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

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

where the functions g_1 and g_2 are the solution of (12) and (13).

Remark 8 In the case where either of the coefficients is correctly specified, we can also derive a similar asymptotic result to the above. It is worth noticing that when Z is a standard Wiener process and the drift coefficient is misspecified, the convergence rate of $\hat{\gamma}_n$ is still $\sqrt{n}$ as in the correctly specified case (cf. Uchida and Yoshida 2011) since the fluctuation of the drift part is dominated by the diffusion part in L_p-sense ($p \geq 2$). By using this estimate, we consider the stepwise estimation procedure.

Remark 9 From Assumption 1 and Assumption 2, we have

$$\partial_\alpha \mathbb{H}_2(\alpha^\star) = \int_{\mathbb{R}} \partial_\alpha a(x,\alpha^\star) c^{-2}(x,\gamma^\star)(A(x) - a(x,\alpha))\pi_0(dx) = 0.$$

Hence, the off-diagonal part $\Gamma_{\alpha\gamma}$ of Γ becomes zero matrix for a scale coefficient $c(x,\gamma)$ being linear with respect to γ, and it also does when the drift coefficient is correctly specified.

Remark 10 Regardless of whether the model is correctly specified or not, it is easy to construct a consistent estimator of Γ. The $p \times p$ matrix $\hat{\Gamma}_n$ defined by

$$\hat{\Gamma}_n = \begin{pmatrix} \frac{1}{n}\partial_\gamma^{\otimes 2}\mathbb{H}_{1,n}(\hat{\gamma}_n) & O \\ \frac{1}{nh_n}\partial_\gamma\partial_\alpha\mathbb{H}_{2,n}(\hat{\alpha}_n,\hat{\gamma}_n) & \frac{1}{nh_n}\partial_\alpha^{\otimes 2}\mathbb{H}_{2,n}(\hat{\alpha}_n,\hat{\gamma}_n) \end{pmatrix},$$

is one example and this matrix works both in the correctly specified and misspecified case. However, the solutions of the (extended) Poisson equations are hard to estimate to the best of the author's knowledge, and thus we will rely on the bootstrap approach in the next section.

3 Main results

3.1 Adjustment of convergence rate

From Table 1, the difference of the convergence rate can be seen with respect to the scale estimator $\hat{\gamma}_n$; more specifically, its convergence rate is $\sqrt{n}$ in the correctly specified case, and otherwise it is $\sqrt{T_n}$. However, since no one can distinguish whether the statistical model is correctly specified or not, as a matter of course, we cannot identify A_n in advance. Therefore, we need a constructible alternative of A_n for uniformly dealing with the all cases below. To satisfy the demand, we introduce the following adjustment term

$$b_n := b_{1,n} + b_{2,n},$$

where $b_{1,n}$ and $b_{2,n}$ are defined as

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

$$b_{2,n} = \exp\left(-\left\{ \left| \frac{1}{n}\sum_{j=1}^n \left[\frac{(\Delta_j X)^4}{3h_n^2} - \frac{2(\Delta_j X)^2 c_{t_{j-1}}^2(\hat{\gamma}_n)}{h_n} + c_{t_{j-1}}^4(\hat{\gamma}_n) \right] \right| + \left| \frac{1}{n}\sum_{j=1}^n \left[\frac{(\Delta_j X)^4}{3h_n^2} - \frac{2(\Delta_j X)^2 c_{t_{j-1}}^2(\hat{\gamma}_n)}{h_n} + c_{t_{j-1}}^4(\hat{\gamma}_n) \right] \right|^{-1} \right\} \right).$$

Obviously, it can be constructed only by the observed data. The next proposition provides the asymptotic behavior of b_n.

Proposition 11 *Suppose that Assumption 2, Assumption 3, and Assumption 5 hold. Then, the adjustment term b_n behaves as follows:*

- In the correctly specified diffusion case,

$$\frac{b_n}{3h_n} \xrightarrow{p} \frac{\int_{\mathbb{R}} c^4(x,\gamma^\star)\pi_0(dx)}{\int_{\mathbb{R}} c^2(x,\gamma^\star)\pi_0(dx)}. \tag{14}$$

- In the misspecified diffusion case,

$$b_n \xrightarrow{p} \exp\Bigg(-\Bigg\{\int_{\mathbb{R}} (C^2(x)-c^2(x,\gamma^\star))^2\pi_0(dx) + \left[\int_{\mathbb{R}} (C^2(x)-c^2(x,\gamma^\star))^2\pi_0(dx)\right]^{-1}\Bigg\}\Bigg) \neq 0. \tag{15}$$

- In the pure-jump Lévy driven case,

$$b_n \xrightarrow{p} \frac{\int_{\mathbb{R}} c^4(x,\gamma^\star)\pi_0(dx)\int_{\mathbb{R}} z^4\nu_0(dz)}{\int_{\mathbb{R}} c^2(x,\gamma^\star)\pi_0(dx)}. \tag{16}$$

Hereafter, we write $b^\star$ as the (scaled) limit of b_n given in Proposition 11. The importance of Proposition 11 is that the convergence rate of b_n is h_n only in the correctly specified diffusion case, that is, the convergence rate of the scale estimator is equivalent to $\sqrt{\frac{T_n}{b_n}}$ up to constant. Thus, the new matrix

$$\hat{A}_n := \begin{pmatrix} \sqrt{\frac{T_n}{b_n}} I_{p_\gamma} & O \\ O & \sqrt{T_n} I_{p_\alpha} \end{pmatrix},$$

is constructed by the observed data, and serves as a good alternative of A_n. Let

$$B^\star = \begin{pmatrix} \frac{1}{\sqrt{b^\star}} I_{p_\gamma} & O \\ O & I_{p_\alpha} \end{pmatrix}.$$

A simple application of Slutky's lemma with the asymptotic normality of $\hat{\theta}_n$ gives the following corollary.

Corollary 12 *Under Assumptions 1–5, we have*

$$\hat{A}_n \hat{\Gamma}_n(\hat{\theta}_n - \theta^\star) \xrightarrow{\mathcal{L}} N\big(0, B^\star \Sigma B^\star\big). \tag{17}$$

3.2 Bootstrap Gaussian quasi-maximum likelihood estimator

From now on, we consider the approximation of the distribution of $\hat{A}_n\hat{\Gamma}_n(\hat{\theta}_n - \theta^\star)$ instead of $A_n\hat{\Gamma}_n(\hat{\theta}_n - \theta^\star)$ since we can avoid checking whether the model is misspecified or not and the driving noise is Wiener or not.

We divide the set $\{1, \dots, n\}$ into k_n-blocks $(B_{k_i})_{i=1}^{k_n}$ defined by:

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

where $c_n = \frac{n}{k_n}$, and here k_n and c_n are supposed to be a positive integer for simplicity. With bootstrap weights $\{w_i\}_{i=1}^{k_n}$, we define the bootstrap Gaussian quasi-score function $\mathbb{G}_n^{\mathbf{B}}(\theta) := \left(\mathbb{G}_{1,n}^{\mathbf{B}}(\gamma), \mathbb{G}_{2,n}^{\mathbf{B}}(\alpha) \right)^\top$ as:

$$\mathbb{G}_{1,n}^{\mathbf{B}}(\gamma) = \sum_{i=1}^{k_n} w_i \sum_{j \in B_{k_i}} \frac{\partial_\gamma c_{t_{j-1}}(\gamma)}{c_{t_{j-1}}^3(\gamma)} \left[h_n c_{t_{j-1}}^2(\gamma) - (\Delta_j X)^2 \right],$$

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

We define our bootstrap estimator $\hat{\theta}_n^{\mathbf{B}} := (\hat{\gamma}_n^{\mathbf{B}}, \hat{\alpha}_n^{\mathbf{B}})$ as the solution of

$$\left| \mathbb{G}_{1,n}^{\mathbf{B}}(\hat{\gamma}_n^{\mathbf{B}}) \right| + \left| \mathbb{G}_{2,n}^{\mathbf{B}}(\hat{\alpha}_n^{\mathbf{B}}) \right| = 0.$$

For the bootstrap weights and block size, we assume:

Assumption 13 There exists a positive $\delta \in \left(\frac{1}{2}, 1\right)$ such that $k_n = O\left(T_n^\delta\right)$, and the bootstrap weights $\{w_i\}_{i=1}^{k_n}$ are i.i.d. random variables and independent of $X = (X_t)_{t \ge 0}$ with $E[w_i] = 1$, $E[w_i^2] = 1$, and $E[|w_i|^{2+\delta'}] < \infty$, for some $\delta' > 0$.

In the rest of this paper, $P^{\mathbf{B}}$ stands for the probability of bootstrap random variables, conditional on $\mathcal{F}$. Analogously, $E^{\mathbf{B}}$ represents the expectation with respect to $P^{\mathbf{B}}$. More specifically, for any bootstrap quantity $U_n(\cdot, \omega)$ and measurable set A,

$$P^{\mathbf{B}}\left(U_n \in A\right) = P^{\mathbf{B}}\left(U_n(\cdot, \omega) \in A | \mathbb{X}\right),$$

where $\omega \in \Omega$. Regarding $P^{\mathbf{B}}$, r_{nB} denotes a generic random vector fulfilling

$$P^{\mathbf{B}}(|r_{nB}| > M) = o_p(1),$$

for any $M > 0$. Its explicit form depends on each context.

Remark 14 For such a weighted bootstrap procedure, several papers often assume the additional condition $E[w_i^3] = 1$ in order to fit the first three moments of the bootstrap distribution. A popular candidate of the distribution of w_i is proposed by Mammen (1993):

$$w_i = \begin{cases} \frac{1-\sqrt{5}}{2} & \text{with probability } p = \frac{\sqrt{5}+1}{2\sqrt{5}} \\ \frac{1+\sqrt{5}}{2} & \text{with probability } 1-p \end{cases}$$

We have the following other choice: let $\frac{w_i}{4}$ be the beta distribution whose density function $f_{\frac{w_i}{4}}$ is given by

$$f_{\frac{w_i}{4}}(x) = \begin{cases} \frac{1}{B(\frac{1}{2},\frac{3}{2})} x^{\frac{1}{2}-1}(1-x)^{\frac{3}{2}-1} & 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

where B denotes the beta function. Then the distribution of w_i is continuous and satisfies Assumption 13 and $E[w_i^3] = 1$. In our case, the latter one often gives numerically good results. This may be because k_n is not so large and the limit distribution of the normalized bootstrap estimator is continuous.

Remark 15 Chatterjee and Bose (2005) deals with a similar bootstrap estimating equation for martingale difference arrays. In the paper, the block size is equivalent to the sample size n as well as i.i.d. case. In contrast, our block size is much smaller. This is for making the misspecification bias in the bootstrap distribution asymptotically negligible. More specifically, the bias can be written as

$$\frac{1}{\sqrt{T_n}} \sum_{i=1}^{k_n} (w_i - 1) \sum_{j \in B_{k_i}} (f_{t_j} - f_{t_{j-1}}) = \frac{1}{\sqrt{T_n}} \sum_{i=1}^{k_n} (w_i - 1)(f_{ic_n h_n} - f_{[(i-1)c_n+1]h_n}),$$

where f denotes a solution of the (extended) Poisson equations introduced in the previous section. Under our assumptions, the bias term is evaluated as $O_p(\sqrt{k_n}/\sqrt{T_n})$ (cf. the proof of Theorem 17), and thus we need a stringent upper bound for the block size; $\delta < 1$ in Assumption 13 is essential.

Example 16 Suppose that $p_\alpha = p_\gamma = 1$ and that the coefficients are written as $a(x,\alpha) = \alpha a(x)$ and $c(x,\gamma) = \gamma c(x)$ for some $\mathbb{R}$-valued smooth functions $a(x)$ and $c(x)$ with $\gamma > 0$. Then, given $\mathbb{X}$, the bootstrap estimator is calculated as

$$\hat{\gamma}_n^{\mathbf{B}} = \sqrt{\frac{k_n \sum_{i=1}^{k_n} w_i \sum_{j \in B_{k_i}} \frac{(\Delta_j X)^2}{c_{t_{j-1}}^2}}{T_n \sum_{i=1}^{k_n} w_i}}, \quad \hat{\alpha}_n^{\mathbf{B}} = \frac{\sum_{i=1}^{k_n} w_i \sum_{j \in B_{k_i}} \frac{\Delta_j X}{c_{t_{j-1}}^2} a_{t_{j-1}}}{h_n \sum_{i=1}^{k_n} w_i \sum_{j \in B_{k_i}} \frac{a_{t_{j-1}}^2}{c_{t_{j-1}}^2}}.$$

Let

$$\hat{B}_n = \begin{pmatrix} \frac{1}{\sqrt{T_n b_n}} I_{p_\gamma} & O \\ O & \frac{1}{\sqrt{T_n}} I_{p_\alpha} \end{pmatrix}, \quad \bar{\Gamma}_n = \begin{pmatrix} \frac{1}{n} \partial_\gamma^{\otimes 2} \mathbb{H}_{1,n}(\hat{\gamma}_n) & O \\ O & \frac{1}{T_n} \partial_\alpha^{\otimes 2} \mathbb{H}_{2,n}(\hat{\alpha}_n, \hat{\gamma}_n) \end{pmatrix}.$$

For each $j \in \{1, \ldots, n\}$, define the indicator function $\chi_j(s)$ by

$$\chi_j(s) = \begin{cases} 1, & s \in [t_{j-1}, t_j), \\ 0, & \text{otherwise.} \end{cases}$$

The following theorem ensures the existence of the bootstrap estimator and the bootstrap consistency of our method.

Theorem 17 *Under Assumptions 1–5 and Assumption 13, we have*

$$P^{\mathbf{B}}\left(\hat{\theta}_n^{\mathbf{B}} \in \Theta\right) = 1 - o_p(1), \tag{18}$$

and $\hat{\theta}_n^{\mathbf{B}}$ admits the following stochastic expansion:

$$\hat{A}_n \bar{\Gamma}_n (\hat{\theta}_n^{\mathbf{B}} - \hat{\theta}_n) = \hat{B}_n \sum_{i=1}^{k_n} (w_i - 1) \sum_{j \in B_{k_i}} \begin{pmatrix} \dfrac{\partial_\gamma c_{t_{j-1}}(\hat{\gamma}_n)}{c^3_{t_{j-1}}(\hat{\gamma}_n)} \left[h_n c^2_{t_{j-1}}(\hat{\gamma}_n) - (\Delta_j X)^2 \right] \\ \dfrac{\partial_\alpha a_{t_{j-1}}(\hat{\alpha}_n)}{c^2_{t_{j-1}}(\hat{\gamma}_n)} \left(\Delta_j X - h_n a_{t_{j-1}}(\hat{\alpha}_n) \right) \end{pmatrix} + r_{nB}. \tag{19}$$

Furthermore, the first term of the right-hand-side of (19) can be expressed as:

- In the correctly specified diffusion case,

$$\begin{aligned} &\hat{B}_n \sum_{i=1}^{k_n} (w_i - 1) \sum_{j \in B_{k_i}} \begin{pmatrix} \dfrac{\partial_\gamma c_{t_{j-1}}(\hat{\gamma}_n)}{c^3_{t_{j-1}}(\hat{\gamma}_n)} \left\{ h_n c^2_{t_{j-1}}(\hat{\gamma}_n) - (\Delta_j X)^2 \right\} \\ \dfrac{\partial_\alpha a_{t_{j-1}}(\hat{\alpha}_n)}{c^2_{t_{j-1}}(\hat{\gamma}_n)} \left(\Delta_j X - h_n a_{t_{j-1}}(\hat{\alpha}_n) \right) \end{pmatrix} \\ &= \hat{B}_n \sum_{i=1}^{k_n} (w_i - 1) \sum_{j \in B_{k_i}} \int_{(i-1)c_n h_n}^{i c_n h_n} \begin{pmatrix} 2 \dfrac{\partial_\gamma c_{t_{j-1}}}{c_{t_{j-1}}} (w_s - w_{t_{j-1}}) \\ \dfrac{\partial_\alpha a_{t_{j-1}}}{c_{t_{j-1}}} \end{pmatrix} \chi_j(s) dw_s + r_{nB}. \end{aligned} \tag{20}$$

- In the misspecified diffusion case,

$$\begin{aligned} &\hat{B}_n \sum_{i=1}^{k_n} (w_i - 1) \sum_{j \in B_{k_i}} \begin{pmatrix} \dfrac{\partial_\gamma c_{t_{j-1}}(\hat{\gamma}_n)}{c^3_{t_{j-1}}(\hat{\gamma}_n)} \left\{ h_n c^2_{t_{j-1}}(\hat{\gamma}_n) - (\Delta_j X)^2 \right\} \\ \dfrac{\partial_\alpha a_{t_{j-1}}(\hat{\alpha}_n)}{c^2_{t_{j-1}}(\hat{\gamma}_n)} \left(\Delta_j X - h_n a_{t_{j-1}}(\hat{\alpha}_n) \right) \end{pmatrix} \\ &= \hat{B}_n \sum_{i=1}^{k_n} (w_i - 1) \sum_{j \in B_{k_i}} \int_{(i-1)c_n h_n}^{i c_n h_n} \begin{pmatrix} \partial_x f_1(X_s) \\ \dfrac{\partial_\alpha a_s}{c_s^2} - \partial_x f_2(X_s) \end{pmatrix} C_s \chi_j(s) dw_s + r_{nB}. \end{aligned} \tag{21}$$

- In the pure-jump Lévy driven case,

$$\hat{B}_n \sum_{i=1}^{k_n}(w_i - 1) \sum_{j\in B_{k_i}} \begin{pmatrix} \frac{\partial_\gamma c_{t_{j-1}}(\hat{\gamma}_n)}{c^3_{t_{j-1}}(\hat{\gamma}_n)} \left\{ h_n c^2_{t_{j-1}}(\hat{\gamma}_n) - (\Delta_j X)^2 \right\} \\ \frac{\partial_\alpha a_{t_{j-1}}(\hat{\alpha}_n)}{c^2_{t_{j-1}}(\hat{\gamma}_n)} \left(\Delta_j X - h_n a_{t_{j-1}}(\hat{\alpha}_n) \right) \end{pmatrix}$$
$$= \hat{B}_n \sum_{i=1}^{k_n}(w_i - 1) \int_{(i-1)c_n h_n}^{ic_n h_n} \int_{\mathbb{R}} \begin{pmatrix} \frac{\partial_\gamma c_{s-}}{c_{s-}} z^2 \\ \frac{\partial_\alpha a_s}{c_{s-}} z \end{pmatrix} \tilde{N}(ds, dz) + r_{nB}. \tag{22}$$

- In the misspecified pure-jump Lévy driven case,

$$\hat{B}_n \sum_{i=1}^{k_n}(w_i - 1) \sum_{j\in B_{k_i}} \begin{pmatrix} \frac{\partial_\gamma c_{t_{j-1}}(\hat{\gamma}_n)}{c^3_{t_{j-1}}(\hat{\gamma}_n)} \left\{ h_n c^2_{t_{j-1}}(\hat{\gamma}_n) - (\Delta_j X)^2 \right\} \\ \frac{\partial_\alpha a_{t_{j-1}}(\hat{\alpha}_n)}{c^2_{t_{j-1}}(\hat{\gamma}_n)} \left(\Delta_j X - h_n a_{t_{j-1}}(\hat{\alpha}_n) \right) \end{pmatrix}$$
$$= \hat{B}_n \sum_{i=1}^{k_n}(w_i - 1) \int_{(i-1)c_n h_n}^{ic_n h_n} \int_{\mathbb{R}} \begin{pmatrix} \frac{\partial_\gamma c_{s-}}{c^3_{s-}} C^2_{s-} z^2 + g_1(X_{s-} + C_{s-}z) - g_1(X_{s-}) \\ \frac{\partial_\alpha a_s}{c^2_{s-}} C_{s-} z + g_2(X_{s-} + C_{s-}z) - g_2(X_{s-}) \end{pmatrix}$$
$$\tilde{N}(ds, dz) + r_{nB}. \tag{23}$$

Moreover, we get the following convergence for all cases:

$$\sup_{x\in\mathbb{R}^{p_\alpha+p_\gamma}} \left| P^{\mathbf{B}}\left(\hat{A}_n \bar{\Gamma}_n \left(\hat{\theta}^{\mathbf{B}}_n - \hat{\theta}_n\right) \le x\right) - P\left(\hat{A}_n \hat{\Gamma}_n \left(\hat{\theta}_n - \theta^\star\right) \le x\right) \right| \xrightarrow{p} 0. \tag{24}$$

Remark 18 In order to obtain the bootstrap percentile and confidence intervals, we need to generate the bootstrap samples $\{\hat{A}_n \bar{\Gamma}_n (\hat{\theta}^{\mathbf{B}}_{n,l} - \hat{\theta}_n)\}_{l=1}^L$, for large $L \in \mathbb{N}$ in practice. However, the calculation of each $\hat{\theta}^{\mathbf{B}}_{n,l}$ often entails some optimization method such as the quasi-Newton method, resulting in high computational complexity. For such a problem, the stochastic expansion (19) shown in Theorem 17 suggests that we can use the following quasi-score function based bootstrap quantity:

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

instead of $\hat{A}_n \bar{\Gamma}_n (\hat{\theta}^{\mathbf{B}}_{n,l} - \hat{\theta}_n)$. Importantly, once $\hat{\theta}_n$ is obtained, we can generate $\{\hat{\eta}^{\mathbf{B}}_{n,l}\}_{l=1}^L$ without any optimization, and thus drastically relieving the computational load.

4 Numerical experiment

We consider the following data-generating model and statistical model:

$$dX_t = -\frac{1}{2}X_t dt + dZ_t, \quad X_0 = 0,$$
$$dX_t = -\frac{1}{2}X_t dt + \frac{\gamma}{\sqrt{1+X_t^2}}dZ_t, \quad \gamma > 0. \tag{25}$$

As for the distribution of Z_1, we consider the two cases $(i)\mathcal{L}(Z_1) = N(0,1)$ and $(ii)\mathcal{L}(Z_1) = bgamma(1, \sqrt{2}, 1, \sqrt{2})$, where $bgamma(\delta_1, \gamma_1, \delta_2, \gamma_2)$ is defined as the law of $\tau_1 - \tau_2$ where for each $i \in \{1,2\}$, τ_i stands for a gamma random variable whose Lévy density is

$$f_{\tau^i}(z) = \frac{\delta_i}{z}e^{-\gamma_i z}, \quad z > 0.$$

For two pairs of the sample size and Terminal time $(n, T_n) = (5 \times 10^4, 200), (10^5, 500)$, we independently generate 1000 paths of (25) based on Euler–Maruyama scheme. Concerning the bootstrap weights, we choose the beta distribution based random variables given in Remark 14. We set k_n as 25 and 50 for each pair. Then, we have $\log_{500} 25 \approx 0.518$, $\log_{200} 25 \approx 0.608$, $\log_{500} 50 \approx 0.629$ and $\log_{200} 50 \approx 0.738$. Hence, in this setting, all of Assumptions 1–5 and Assumption 13 hold with the optimal value $\gamma^\star = \sqrt{2}$ (cf. Uchida and Yoshida 2011 and Uehara 2019). Table 2 shows the actual coverage rate of all the 99% bootstrap interval constructed by 1000 bootstrap replication. Each coverage rate is approaching as n and T_n increases. Compared with case (i), the coverage rate of case (ii) is slightly worse; this is probably because of the high kurtosis of the bilateral gamma distribution which makes the asymptotic variance of $\hat{\gamma}_n$ large.

5 Proofs

Hereafter, $R(x)$ denotes a generic function being of at most polynomial growth. Its form may vary depending on context.

Table 2 Coverage rate of 99% bootstrap confidence interval

n	T_n	k_n	Coverage rate (i)	Coverage rate (ii)
10^5	500	25	0.962	0.952
10^5	500	50	0.969	0.944
5×10^4	200	25	0.935	0.924
5×10^4	200	50	0.939	0.907

5.1 Auxiliary results for blocked sums

We first prepare some lemmas repeatedly used in the proof of our main results. All of their proofs are presented in Supplementary material.

Lemma 19 *Suppose that Assumptions 2–3 and Assumption 5 hold. Then, we have*

$$\sum_{i=1}^{k_n}\left|\sum_{j\in B_{k_i}} R_{t_{j-1}}\int_{t_{j-1}}^{t_j}(A_s-A_{t_{j-1}})ds\right|^2=O_p\left(\frac{n^2h_n^3}{k_n}\right), \tag{26}$$

$$\sum_{i=1}^{k_n}\left|\sum_{j\in B_{k_i}} R_{t_{j-1}}\left(\int_{t_{j-1}}^{t_j}A_s ds\right)^2\right|^2=O_p\left(\frac{n^2h_n^4}{k_n}\right), \tag{27}$$

$$\sum_{i=1}^{k_n}\left|\sum_{j\in B_{k_i}} R_{t_{j-1}}\int_{t_{j-1}}^{t_j}C_{s-}dZ_s\right|^2=O_p(T_n). \tag{28}$$

Lemma 20 *Suppose that Assumptions 2-3 hold. Then, under Assumption 5-(1), we have*

$$\sum_{i=1}^{k_n}\left|\sum_{j\in B_{k_i}} R_{t_{j-1}}\left[\left(\int_{t_{j-1}}^{t_j}C_s dw_s\right)^2-h_nC_{t_{j-1}}^2\right]\right|^2=O_p\left(nh_n^2\right), \tag{29}$$

and under Assumption 5-(2),

$$\sum_{i=1}^{k_n}\left|\sum_{j\in B_{k_i}} R_{t_{j-1}}\left[\left(\int_{t_{j-1}}^{t_j}C_{s-} dZ_s\right)^2-h_nC_{t_{j-1}}^2\right]\right|^2=O_p\left(T_n\right). \tag{30}$$

We will say that a matrix-valued function f on $\mathbb{R}$ is centered if $\pi_0(f)=0$ in the rest of this section.

Lemma 21 *Suppose that a centered matrix-valued function f is differentiable, and that it and its derivative are of at most polynomial growth. Then, under Assumption 2, Assumption 3, and Assumption 5, we have*

$$\sum_{i=1}^{k_n}\left|\sum_{j\in B_{k_i}} f_{t_{j-1}}\right|^2=O_p\left(\sqrt{nk_n}\right). \tag{31}$$

The following lemma verifies the probability limit of the sum of squared Wiener integrals.

Lemma 22 *Suppose that Assumption 2, Assumption 3, and Assumption 5-(1) hold. We further assume that a differentiable function f on $\mathbb{R}$ and its derivative are of at most polynomial growth. Then, we have*

$$\begin{aligned}&\frac{1}{nh_n^2}\sum_{i=1}^{k_n}\left\{\int_{(i-1)c_nh_n}^{ic_nh_n}\left[\sum_{j=(i-1)c_n+1}^{ic_n} f_{t_{j-1}}(w_s-w_{t_{j-1}})\chi_j(s)\right]dw_s\right\}^2\\&\xrightarrow{p}\frac{1}{2}\int_{\mathbb{R}}(f(x))^2\pi_0(dx),\end{aligned} \tag{32}$$

$$\frac{1}{T_n}\sum_{i=1}^{k_n}\left(\int_{(i-1)c_nh_n}^{ic_nh_n}\sum_{j=(i-1)c_n+1}^{ic_n} f_s\chi_j(s)dw_s\right)^2\xrightarrow{p}\int_{\mathbb{R}}(f(x))^2\pi_0(dx). \tag{33}$$

We next show a similar convergence result to Lemma 22 when the driving noise is a pure-jump Lévy process.

Lemma 23 *Suppose that Assumptions 2–3 and Assumption 5-(2) hold. Moreover, for two functions f_1 and f_2 on $\mathbb{R}$, we assume the following conditions:*

1. *f_1 is differentiable, and it and its derivative is of at most polynomial growth.*
2. *There exists a positive constant K such that for any $p \in (1,\infty)$ and $q=\frac{p}{p-1}$,*

$$\sup_{x,y\in\mathbb{R},x\neq y}\frac{|f_2(x)-f_2(y)|}{|x-y|^{1/p}(1+|x|^{qK}+|y|^{qK})}<\infty.$$

Let $f(x,z)=f_1(x)z^\delta+f_2(x+z)-f_2(x)$ for a fixed $\delta\geq 1$. Then, we have

$$\frac{1}{T_n}\sum_{i=1}^{k_n}\left(\int_{(i-1)c_nh_n}^{ic_nh_n}\int f(X_{s-},z)\tilde{N}(ds,dz)\right)^2\xrightarrow{p}\int_{\mathbb{R}}\int_{\mathbb{R}}(f(x,z))^2\pi_0(dx)\nu_0(dz).$$

5.2 Proof of Proposition 11

For simplicity, we write

$$\xi_j=\Delta_jX-C_{t_{j-1}}\Delta_jZ=\int_{t_{j-1}}^{t_j}A_sds+\int_{t_{j-1}}^{t_j}(C_{s-}-C_{t_{j-1}})dZ_s,$$

and we divide the proof into the diffusion case and pure-jump Lévy driven case.

5.2.1 Diffusion case

For any $q\geq 2$, from Burkholder's inequality and Assumption 2, we have

$$E[\left|\xi_j\right|^q] \lesssim h_n^q, \tag{34}$$

$$E^{j-1}[\left|\xi_j\right|^q] \lesssim h_n^q R_{t_{j-1}}. \tag{35}$$

Combined with Hölder's inequality and the ergodic theorem, it follows from (Genon-Catalot and Jacod 1993, Lemma 9) that

$$\frac{1}{T_n}\sum_{j=1}^{n}(\Delta_j X)^2 = \frac{1}{T_n}\sum_{j=1}^{n} C_{t_{j-1}}^2 (\Delta_j Z)^2 + O_p\left(\sqrt{h_n}\right) \xrightarrow{p} \int C^2(x)\pi_0(dx), \tag{36}$$

$$\frac{1}{nh_n^2}\sum_{j=1}^{n}(\Delta_j X)^4 = \frac{1}{n}\sum_{j=1}^{n} \frac{(\Delta_j Z)^4}{h_n^2} C_{t_{j-1}}^4 + O_p\left(\sqrt{h_n}\right) \xrightarrow{p} 3\int C^4(x)\pi_0(dx). \tag{37}$$

Hence, Slutsky's theorem leads to

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

Next, we look at $b_{2,n}$. By applying the Taylor's expansion, we have

$$\begin{aligned}
&\frac{1}{n}\sum_{j=1}^{n}\left[-2\frac{(\Delta_j X)^2}{h_n} c_{t_{j-1}}^2(\hat{\gamma}_n) + c_{t_{j-1}}^4(\hat{\gamma}_n)\right] \\
&= \frac{1}{n}\sum_{j=1}^{n}\left[-2\frac{(\Delta_j X)^2}{h_n} c_{t_{j-1}}^2 + c_{t_{j-1}}^4\right] \\
&\quad + \frac{1}{n}\int_0^1 \sum_{j=1}^{n}\left[-2\frac{(\Delta_j X)^2}{h_n}\partial_\gamma c_{t_{j-1}}^2(\hat{\gamma}_n + u(\gamma^\star - \hat{\gamma}_n))\right. \\
&\quad \left. + \partial_\gamma c_{t_{j-1}}^4(\hat{\gamma}_n + u(\gamma^\star - \hat{\gamma}_n))\right] du[\hat{\gamma}_n - \gamma^\star].
\end{aligned}$$

Let a_n be the convergence rate of $\hat{\gamma}_n$. Since $\partial_\gamma c^2$ and $\partial_\gamma c^4$ are of at most polynomial growth and $a_n(\hat{\gamma}_n - \gamma^\star) = O_p(1)$, we can deduce that the second term of the right-hand side is $O_p(\frac{1}{a_n})$. Hence, (34) and (35) lead to

$$\begin{aligned}
&\frac{1}{n}\sum_{j=1}^{n}\left[\frac{(\Delta_j X)^4}{3h_n^2} - 2\frac{(\Delta_j X)^2}{h_n} c_{t_{j-1}}^2(\hat{\gamma}_n) + c_{t_{j-1}}^4(\hat{\gamma}_n)\right] \\
&= \frac{1}{n}\sum_{j=1}^{n}\left[\frac{(\Delta_j Z)^4}{3h_n^2} C_{t_{j-1}}^4 - 2\frac{(\Delta_j Z)^2}{h_n} c_{t_{j-1}}^2 + c_{t_{j-1}}^4\right] + O_p\left(\sqrt{h_n} \vee \frac{1}{a_n}\right).
\end{aligned}$$

In the misspecified case, we can similarly observe that the first term of the right-hand side converges to

$$\int (C^2(x) - c^2(x))^2 \pi_0(dx),$$

in probability. Then, (15) follows from the continuous mapping theorem. In the correctly specified case, we obtain

$$E^{j-1}\left[\frac{(\Delta_j Z)^4}{3h_n^2} C^4_{t_{j-1}} - 2\frac{(\Delta_j Z)^2}{h_n} c^2_{t_{j-1}} + c^4_{t_{j-1}}\right] = 0,$$
$$E^{j-1}\left[\left(\frac{(\Delta_j Z)^4}{3h_n^2} C^4_{t_{j-1}} - 2\frac{(\Delta_j Z)^2}{h_n} c^2_{t_{j-1}} + c^4_{t_{j-1}}\right)^2\right] \lesssim R_{t_{j-1}}.$$

(Genon-Catalot and Jacod 1993, Lemma 9) yields that for a positive constant $\delta \in (0, 1/2)$, we have

$$\frac{1}{n}\sum_{j=1}^{n}\left[\frac{(\Delta_j Z)^4}{3h_n^2} C^4_{t_{j-1}} - 2\frac{(\Delta_j Z)^2}{h_n} c^2_{t_{j-1}} + c^4_{t_{j-1}}\right] = o_p\left(n^{-\frac{1}{2}+\delta}\right).$$

Since $a_n = \sqrt{n}$ and $\sqrt{h_n} \vee \frac{1}{\sqrt{n}} = \sqrt{h_n}$, we arrive at

$$\frac{1}{n}\sum_{j=1}^{n}\left[\frac{(\Delta_j X)^4}{3h_n^2} - 2\frac{(\Delta_j X)^2}{h_n} c^2_{t_{j-1}}(\hat{\gamma}_n) + c^4_{t_{j-1}}(\hat{\gamma}_n)\right] = o_p\left(n^{-\frac{1}{2}+\delta} \wedge h_n^{\frac{1}{2}-\delta}\right) = o_p\left(h_n^{\frac{1}{2}-\delta}\right), \tag{38}$$

and consequently,

$$\frac{b_{2,n}}{h_n} \lesssim \frac{1}{h_n}\exp\left(-\frac{h_n^{-\frac{1}{2}+\delta}}{|o_p(1)|}\right) = o_p(1),$$

and this concludes (14).

5.2.2 Pure-jump Lévy driven case

Since the route is quite similar to the diffusion case, we omit some details. For any $q \geq 2$, Burkholder's inequality and Assumption 2 imply that

$$E[\left|\xi_j\right|^q] \lesssim h_n^2, \tag{39}$$

$$E^{j-1}[\left|\xi_j\right|^q] \lesssim h_n^2 R_{t_{j-1}}. \tag{40}$$

Hence by making use of (Genon-Catalot and Jacod 1993, Lemma 9), we get

$$\frac{1}{T_n}\sum_{j=1}^{n}(\Delta_j X)^2 = \frac{1}{T_n}\sum_{j=1}^{n} C_{t_{j-1}}^2(\Delta_j Z)^2 + O_p\left(\sqrt{h_n}\right) \xrightarrow{p} \int C^2(x)\pi_0(dx), \tag{41}$$

$$\frac{1}{nh_n}\sum_{j=1}^{n}(\Delta_j X)^4 = \frac{1}{n}\sum_{j=1}^{n} C_{t_{j-1}}^4 \frac{(\Delta_j Z)^4}{h_n} + O_p\left(\sqrt{h_n}\right) \xrightarrow{p} \int C^4(x)\pi_0(dx)\int z^4\nu_0(dz), \tag{42}$$

and it is immediate from Slutsky's theorem that

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

From (42) and a similar estimates to the diffusion case, we have

$$\frac{1}{n}\sum_{j=1}^{n}\left[\frac{(\Delta_j X)^4}{3h_n^2} - 2\frac{(\Delta_j X)^2}{h_n}c_{t_{j-1}}^2(\hat{\gamma}_n) + c_{t_{j-1}}^4(\hat{\gamma}_n)\right] \to \infty.$$

Since the function $h(x) = \exp[-(|x| + 1/|x|)]$ tends to 0 as $x \to \infty$, we obtain (16). □

5.3 Proof of Theorem 17

The essence of this proof stems from Chatterjee and Bose (2005). For simplicity, we hereafter write

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

$$\eta_j(\alpha,\gamma) = \frac{\partial_\alpha a_{t_{j-1}}(\alpha)}{c_{t_{j-1}}^2(\gamma)}\left[\Delta_j X - h_n a_{t_{j-1}}(\alpha)\right].$$

Since the matrix $\bar{\Gamma}_n$ is block diagonal, we divide the proof of (19) into the scale part:

$$\sqrt{\frac{T_n}{b_n}}\frac{1}{n}\partial_\gamma^{\otimes 2}\mathbb{H}_{1,n}(\hat{\gamma}_n)(\hat{\gamma}_n^{\mathbf{B}} - \hat{\gamma}_n) = \frac{1}{\sqrt{T_n b_n}}\sum_{i=1}^{k_n}(w_i - 1)\sum_{j\in B_{k_i}}\zeta_j(\hat{\gamma}_n) + r_{nB} \tag{43}$$

$$= \frac{1}{\sqrt{T_n b_n}}\sum_{i=1}^{k_n}(w_i - 1)\sum_{j\in B_{k_i}}\zeta_j(\gamma^\star) + r_{nB}, \tag{44}$$

and drift part:

$$\sqrt{T_n}\frac{1}{T_n}\partial_\gamma^{\otimes 2}\mathbb{H}_{2,n}(\hat{\alpha}_n)(\hat{\alpha}_n^{\mathbf{B}}-\hat{\alpha}_n)=\frac{1}{\sqrt{T_n}}\sum_{i=1}^{k_n}(w_i-1)\sum_{j\in B_{k_i}}\eta_j(\hat{\alpha}_n,\gamma^\star)+r_{nB} \tag{45}$$

$$=\frac{1}{\sqrt{T_n}}\sum_{i=1}^{k_n}(w_i-1)\sum_{j\in B_{k_i}}\eta_j(\alpha^\star,\gamma^\star)+r_{nB}. \tag{46}$$

5.3.1 Scale part

Define the function $\mathbb{H}_n^{\mathbf{B}}$ on $\mathbb{R}^{p_\gamma}$ by

$$\mathbb{H}_n^{\mathbf{B}}(t)=\frac{1}{\sqrt{T_n b_n}}\sum_{i=1}^{k_n}w_i\sum_{j\in B_j}\left\{\zeta_j\left(\hat{\gamma}_n+\sqrt{\frac{b_n}{T_n}}t\right)-\zeta_j(\hat{\gamma}_n)\right\}-\frac{1}{n}\partial_\gamma^{\otimes 2}\mathbb{H}_{1,n}(\hat{\gamma}_n)t.$$

First, we show that

$$E^{\mathbf{B}}\left[\sup_{|t|\le K_n}\left|\mathbb{H}_n^{\mathbf{B}}(t)\right|^2\right]=o_p(1), \tag{47}$$

for any positive sequence (K_n) fulfilling $K_n=o_p(\sqrt{k_n})$. For instance, $K_n=T_n^{1/4}$ satisfies the above condition. Combined with Proposition 11 and the consistency of $\hat{\gamma}_n$, we have $\sqrt{\frac{b_n}{T_n}}K_n=o_p(1)$ and we may and do assume $\hat{\gamma}_n+\sqrt{\frac{b_n}{T_n}}t\in\Theta_\gamma$ as long as $|t|\le K_n$. By applying Taylor's formula twice, we obtain

$$\mathbb{H}_n^{\mathbf{B}}(t)=\mathbb{H}_{1,n}^{\mathbf{B}}(t)+\mathbb{H}_{2,n}^{\mathbf{B}}(t)+\mathbb{H}_{3,n}^{\mathbf{B}}(t),$$

where for notational simplicity, we write

$$\mathbb{H}_{1,n}^{\mathbf{B}}(t)=\frac{1}{T_n}\sum_{i=1}^{k_n}(w_i-1)\sum_{j\in B_{k_i}}t^\top\partial_\gamma\zeta_j(\gamma^\star),$$

$$\mathbb{H}_{2,n}^{\mathbf{B}}(t)=\frac{1}{T_n}\sum_{i=1}^{k_n}(w_i-1)\sum_{j\in B_{k_i}}\int_0^1\partial_\gamma^{\otimes 2}\zeta_j(\hat{\gamma}_n+s(\gamma^\star-\hat{\gamma}_n))ds[t,\hat{\gamma}_n-\gamma^\star],$$

$$\mathbb{H}_{3,n}^{\mathbf{B}}(t)=\sqrt{\frac{b_n}{T_n^3}}\sum_{i=1}^{k_n}w_i\sum_{j\in B_{k_i}}\int_0^1\partial_\gamma^{\otimes 2}\zeta_j\left(\hat{\gamma}_n+\sqrt{\frac{b_n}{T_n}}ut\right)du[t,t].$$

From now on, we separately look at these three ingredients. Assumption 13 yields that

$$E^{\mathbf{B}}\left[\sup_{|t|\le K_n}\left|\mathbb{H}^{\mathbf{B}}_{1,n}(t)\right|^2\right] \le \frac{K_n^2}{T_n^2}\sum_{i=1}^{k_n}\left|\sum_{j\in B_{k_i}}\partial_\gamma\zeta_j(\gamma^\star)\right|^2.$$

Decompose $\partial_\gamma\zeta_j(\gamma^\star)$ as $\partial_\gamma\zeta(\gamma^\star)=\zeta_{1,j}+\zeta_{2,j}$ where

$$\zeta_{1,j}=h_n\frac{c_{t_{j-1}}\partial_\gamma^{\otimes 2}c_{t_{j-1}}-(\partial_\gamma c_{t_{j-1}})^{\otimes 2}}{c_{t_{j-1}}^4}\left(c_{t_{j-1}}^2-C_{t_{j-1}}^2\right)+2h_n\frac{(\partial_\gamma c_{t_{j-1}})^{\otimes 2}}{c_{t_{j-1}}^4}C_{t_{j-1}}^2,$$

$$\zeta_{2,j}=\frac{c_{t_{j-1}}\partial_\gamma^{\otimes 2}c_{t_{j-1}}-3(\partial_\gamma c_{t_{j-1}})^{\otimes 2}}{c_{t_{j-1}}^4}\left[h_nC_{t_{j-1}}^2-(\Delta_jX)^2\right].$$

Jensen's inequality and the ergodic theorem yield that

$$\frac{1}{T_n^2}\sum_{i=1}^{k_n}\left|\sum_{j\in B_{k_i}}\zeta_{1,j}\right|^2\le\frac{1}{nk_n}\sum_{i=1}^{k_n}\sum_{j\in B_{k_i}}\zeta_{1,j}^2=O_p\left(k_n^{-1}\right). \tag{48}$$

It follows from Lemma 19 and Lemma 20 that

$$\frac{1}{T_n^2}\sum_{i=1}^{k_n}\left|\sum_{j\in B_{k_i}}\zeta_{2,j}\right|^2=\begin{cases}O_p\left(n^{-1}\right), & \text{in the diffusion case,}\\ O_p\left(T_n^{-1}\right), & \text{in the pure-jump Lévy driven case.}\end{cases}$$

From these estimates, we get

$$E^{\mathbf{B}}\left[\sup_{|t|\le K_n}\left|\mathbb{H}^{\mathbf{B}}_{1,n}(t)\right|^2\right]=O_p(K_n^2k_n^{-1})=o_p(1). \tag{49}$$

Assumption 13 leads to

$$E^{\mathbf{B}}\left[\sup_{|t|\le K_n}\left|\mathbb{H}^{\mathbf{B}}_{2,n}(t)\right|^2\right]\le\frac{K_n^2}{T_n^2}|\hat\gamma_n-\gamma^\star|^2\sum_{i=1}^{k_n}\left|\sum_{j\in B_{k_i}}\int_0^1\partial_\gamma^{\otimes 2}\zeta_j(\hat\gamma_n+s(\gamma^\star-\hat\gamma_n))ds\right|^2.$$

From Assumption 2, there exists a positive constants M_1 and M_2 such that

$$\begin{aligned}&\left|\int_0^1\partial_\gamma^{\otimes 2}\zeta_j(\hat\gamma_n+s(\gamma^\star-\hat\gamma_n))ds\right|\\&\lesssim\left(1+|X_{t_{j-1}}|^{M_1}\right)\left\{h_n(1+|X_{t_{j-1}}|^{M_2})+\left[(\Delta_jX)^2-h_nC_{t_{j-1}}\right]\right\}.\end{aligned}$$

Combined with Jensen's inequality, Lemma 19, and Lemma 20, we obtain

$$E\left[\left|\sum_{j\in B_{k_i}}\int_0^1\partial_\gamma^{\otimes 2}\zeta_j(\hat\gamma_n+s(\gamma^\star-\hat\gamma_n))ds\right|^2\right]\lesssim\frac{T_n^2}{k_n^2}. \tag{50}$$

Hence, the tightness of $\sqrt{T_n}(\hat{\gamma}_n - \gamma^\star)$ implies that

$$E^{\mathbf{B}}\left[\sup_{|t|\le K_n}\left|\mathbb{H}^{\mathbf{B}}_{2,n}(t)\right|^2\right] = O_p\left(K_n^2 k_n^{-1} T_n^{-1}\right) = o_p(1). \tag{51}$$

We rewrite $\mathbb{H}^{\mathbf{B}}_{3,n}(t)$ as

$$\begin{aligned}\mathbb{H}^{\mathbf{B}}_{3,n}(t) &= \sqrt{\frac{b_n}{T_n^3}}\sum_{i=1}^{k_n}(w_i-1)\sum_{j\in B_{k_i}}\int_0^1 \partial_\gamma^{\otimes 2}\zeta_j\left(\hat{\gamma}_n + \sqrt{\frac{b_n}{T_n}}ut\right)du[t,t]\\ &+ \sqrt{\frac{b_n}{T_n^3}}\sum_{j=1}^{n}\int_0^1 \partial_\gamma^{\otimes 2}\zeta_j\left(\hat{\gamma}_n + \sqrt{\frac{b_n}{T_n}}ut\right)du[t,t].\end{aligned}$$

Recall that $b_n = O_p(1)$. By utilizing Cauchy–Schwartz inequality and the estimates in (50), we have

$$\begin{aligned}&E^{\mathbf{B}}\left[\sup_{|t|\le K_n}\left|\sqrt{\frac{b_n}{T_n^3}}\sum_{i=1}^{k_n}(w_i-1)\sum_{j\in B_{k_i}}\int_0^1 \partial_\gamma^{\otimes 2}\zeta_j\left(\hat{\gamma}_n + \sqrt{\frac{b_n}{T_n}}ut\right)du[t,t]\right|^2\right]\\ &\lesssim \frac{K_n^2 b_n}{T_n^3}\sum_{i=1}^{k_n}\sup_{|t|\le K_n}\left|\sum_{j\in B_{k_i}}\int_0^1 \partial_\gamma^{\otimes 2}\zeta_j\left(\hat{\gamma}_n + \sqrt{\frac{b_n}{T_n}}ut\right)du\right|^2\\ &= O_p(K_n^2 k_n^{-1} T_n^{-1}).\end{aligned}$$

Similarly, for any $C > 0$, we have

$$\sup_{|t|\le K_n}\left|\sqrt{\frac{b_n}{T_n^3}}\sum_{j=1}^{n}\int_0^1 \partial_\gamma^{\otimes 2}\zeta_j\left(\hat{\gamma}_n + \sqrt{\frac{b_n}{T_n}}ut\right)du[t,t]\right|^2 = O_p\left(K_n^2 T_n^{-1}\right),$$

so that

$$E^{\mathbf{B}}\left[\sup_{|t|\le K_n}\left|\mathbb{H}^{\mathbf{B}}_{3,n}(t)\right|^2\right] = O_p\left(K_n^2 T_n^{-1}\right) = o_p(1). \tag{52}$$

Putting (49), (51), and (52) together, we arrive at (47).

Next, we observe that

$$P^{\mathbf{B}}\left(\inf_{|t|=K_n}\left[-\frac{1}{\sqrt{T_n b_n}}\sum_{i=1}^{k_n} w_i \sum_{j\in B_{k_i}} t^\top \zeta_j\left(\hat{\gamma}_n + \sqrt{\frac{b_n}{T_n}}t\right)\right] > 0\right) = 1 - o_p(1). \tag{53}$$

By using the estimates in Kessler (1997); Uchida and Yoshida (2011); Masuda (2013), and Uehara (2019), it is easy to observe that

$$\frac{1}{n}\partial_\gamma^{\otimes 2}\mathbb{H}_{1,n}(\hat{\gamma}_n)\xrightarrow{p} -\Gamma_\gamma > 0. \tag{54}$$

We hereafter write the smallest eigenvalue of $-\Gamma_\gamma$ and $\frac{1}{n}\partial_\gamma^{\otimes 2}\mathbb{H}_{1,n}(\hat{\gamma}_n)$ as λ_γ and $\lambda_{\gamma,n}$, respectively. From (54), we may and do assume that for a fixed $\delta \in (0, \lambda_\gamma)$ and enough large n, $\frac{1}{n}\partial_\gamma^{\otimes 2}\mathbb{H}_{1,n}(\hat{\gamma}_n)$ is positive definite and $|\lambda_\gamma - \lambda_{\gamma,n}| < \delta$ without loss of generality. For such n, and the same sequence (K_n) as the estimates of $\mathbb{H}_n^{\mathbf{B}}$, it follows that

$$
\begin{aligned}
&P^{\mathbf{B}}\left(\inf_{|t|=K_n}\left[-\frac{1}{\sqrt{T_n b_n}}\sum_{i=1}^{k_n} w_i \sum_{j\in B_{k_i}} t^\top \zeta_j\left(\hat{\gamma}_n + \sqrt{\frac{b_n}{T_n}}t\right)\right] > 0\right)\\
&\geq P^{\mathbf{B}}\left(\inf_{|t|=K_n}\left\{-\frac{1}{\sqrt{T_n b_n}}\sum_{i=1}^{k_n} w_i \sum_{j\in B_{k_i}} t^\top \zeta_j\big(\hat{\gamma}_n\big) + t^\top \mathbb{H}_n^{\mathbf{B}}(t) + \frac{1}{n}\partial_\gamma^{\otimes 2}\mathbb{H}_{1,n}(\hat{\gamma}_n)[t,t]\right\} > 0\right)\\
&\geq 22P^{\mathbf{B}}\left(-\sup_{|t|=K_n}\frac{1}{\sqrt{T_n b_n}}\sum_{i=1}^{k_n} w_i \sum_{j\in B_{k_i}} t^\top \zeta_j\big(\hat{\gamma}_n\big) - \sup_{|t|=K_n}\left|t^\top \mathbb{H}_n^{\mathbf{B}}(t)\right| >\right.\\
&\left. -\inf_{|t|=K_n}\frac{1}{n}\partial_\gamma^{\otimes 2}\mathbb{H}_{1,n}(\hat{\gamma}_n)[t,t]\right)\\
&\geq 1 - P^{\mathbf{B}}\left(\left|\frac{1}{\sqrt{T_n b_n}}\sum_{i=1}^{k_n} w_i \sum_{j\in B_{k_i}} \zeta_j\big(\hat{\gamma}_n\big)\right| + \sup_{|t|=K_n}\left|\mathbb{H}_n^{\mathbf{B}}(t)\right| \geq (\lambda_\gamma - \delta)K_n\right)\\
&\geq 1 - P^{\mathbf{B}}\left(\left|\frac{1}{\sqrt{T_n b_n}}\sum_{i=1}^{k_n} w_i \sum_{j\in B_{k_i}} \zeta_j\big(\hat{\gamma}_n\big)\right| \geq \frac{(\lambda_\gamma - \delta)K_n}{2}\right)\\
&- P^{\mathbf{B}}\left(\sup_{|t|=K_n}\left|\mathbb{H}_n^{\mathbf{B}}(t)\right| \geq \frac{(\lambda_\gamma - \delta)K_n}{2}\right).
\end{aligned}
$$

For abbreviation, let

$$M_n = M(K_n) = \frac{(\lambda_\gamma - \delta)K_n}{2}.$$

Then, Taylor's expansion and Chebychev's inequality gives

$$
\begin{aligned}
&E\left[P^{\mathbf{B}}\left(\left|\frac{1}{\sqrt{T_n b_n}}\sum_{i=1}^{k_n} w_i \sum_{j\in B_{k_i}} \zeta_j(\hat{\gamma}_n)\right| \geq M_n\right)\right] \\
&\leq \frac{4}{M_n^2 T_n b_n}\sum_{i=1}^{k_n}\left|\sum_{j\in B_{k_i}}\zeta_j(\gamma^\star)\right|^2 + \frac{4|\hat{\gamma}_n-\gamma^\star|^2}{M_n^2 T_n b_n}\sum_{i=1}^{k_n}\left|\sum_{j\in B_{k_i}}\partial_\gamma\zeta_j(\gamma^\star)\right|^2 \\
&+ \frac{4|\hat{\gamma}_n-\gamma^\star|^4}{M_n^2 T_n b_n}\sum_{i=1}^{k_n}\left|\sum_{j\in B_{k_i}}\int_0^1 \partial_\gamma^{\otimes 2}\zeta_j(\hat{\gamma}_n + s(\gamma^\star-\hat{\gamma}_n))ds\right|^2.
\end{aligned}
$$

$\zeta_j(\gamma^\star)$ can be decomposed as:

$$
\zeta_j(\gamma^\star) = h_n\frac{\partial_\gamma c_{t_{j-1}}}{c_{t_{j-1}}^3}(c_{t_{j-1}}^2 - C_{t_{j-1}}^2) + \frac{\partial_\gamma c_{t_{j-1}}}{c_{t_{j-1}}^3}\left[h_n C_{t_{j-1}}^2 - (\Delta_j X)^2\right].
$$

Notice that from Proposition 11, $T_n b_n = O_p(nh_n^2)$ in the correctly specified diffusion case, and $T_n b_n = O_p(T_n)$ in the other cases, and that the function

$$
\frac{\partial_\gamma c(x,\gamma^\star)}{c^3(x,\gamma^\star)}(c^2(x,\gamma^\star) - C^2(x))
$$

is centered. Hence,

$$
\frac{1}{T_n b_n}\sum_{i=1}^{k_n}\left|\sum_{j\in B_{k_i}}\zeta_j(\gamma^\star)\right|^2 = O_p(1), \tag{55}
$$

is straightforward from Lemma 19, Lemma 20, and Lemma 21. From Proposition 11, we have

$$
\sqrt{\frac{T_n}{b_n}}(\hat{\gamma}_n - \gamma^\star) = O_p(1),
$$

and the estimates of $\mathbb{H}_{1,n}^{\mathbf{B}}$ and $\mathbb{H}_{2,n}^{\mathbf{B}}$ imply that

$$
\begin{aligned}
&\frac{4|\hat{\gamma}_n-\gamma^\star|^2}{M_n^2 T_n b_n}\sum_{i=1}^{k_n}\left|\sum_{j\in B_{k_i}}\partial_\gamma\zeta_j(\gamma^\star)\right|^2 + \frac{4|\hat{\gamma}_n-\gamma^\star|^4}{M_n^2 T_n b_n}\sum_{i=1}^{k_n}\left|\sum_{j\in B_{k_i}}\int_0^1 \partial_\gamma^{\otimes 2}\zeta_j(\hat{\gamma}_n + s(\gamma^\star-\hat{\gamma}_n))ds\right|^2 \\
&= o_p(1).
\end{aligned}
$$

Hence, we obtain (53).

Let $t = t_n$ be a root of the equation

$$
\mathbb{G}_{1,n}^{\mathbf{B}}\left(\hat{\gamma}_n + \sqrt{\frac{b_n}{T_n}}t\right) = 0, \tag{56}
$$

if it exists, and otherwise, t be an arbitrary element of $\mathbb{R}^{p_\gamma}$. For such t, we define

$$\hat{\gamma}_n^{\mathbf{B}} = \hat{\gamma}_n + \sqrt{\frac{b_n}{T_n}}t.$$

On the set

$$\inf_{|t|=K_n}\left[-\frac{1}{\sqrt{T_n b_n}}\sum_{i=1}^{k_n} w_i \sum_{j\in B_{k_i}} t^\top \zeta_j\left(\hat{\gamma}_n + \sqrt{\frac{b_n}{T_n}}t\right)\right] > 0,$$

the continuity of $\zeta(\gamma)$ and (Ortega and Rheinboldt 1970, Theorem 6.3.4) ensure that a root t of (56) does exist within $|t| \leq K_n$. Since $\sqrt{b_n/T_n}K_n = o_p(1)$, (53) and the consistency of $\hat{\gamma}_n$ lead to

$$P^{\mathbf{B}}(\hat{\gamma}_n^{\mathbf{B}} \in \Theta_\gamma) = 1 - o_p(1). \tag{57}$$

Finally, Chebyshev's inequality, (47) and (53) yield that for any $M > 0$,

$$\begin{aligned}
&P^{\mathbf{B}}\left(\left|\sqrt{\frac{T_n}{b_n}}\frac{1}{n}\partial_\gamma^{\otimes 2}\mathbb{H}_{1,n}(\hat{\gamma}_n)(\hat{\gamma}_n^{\mathbf{B}} - \hat{\gamma}_n) - \frac{1}{\sqrt{T_n b_n}}\sum_{i=1}^{k_n}(w_i - 1)\sum_{j\in B_{k_i}}\zeta_j(\hat{\gamma}_n)\right| > M\right)\\
&\leq P^{\mathbf{B}}\left(\inf_{|t|=K_n}\left[-\frac{1}{\sqrt{T_n b_n}}\sum_{i=1}^{k_n} w_i \sum_{j\in B_{k_i}} t^\top \zeta_j\left(\hat{\gamma}_n + \sqrt{\frac{b_n}{T_n}}t\right)\right] \leq 0\right)\\
&+ P^{\mathbf{B}}\left(\left\{\inf_{|t|=K_n}\left[-\frac{1}{\sqrt{T_n b_n}}\sum_{i=1}^{k_n} w_i \sum_{j\in B_{k_i}} t^\top \zeta_j\left(\hat{\gamma}_n + \sqrt{\frac{b_n}{T_n}}t\right)\right] > 0\right\}\right.\\
&\left.\cap\left\{\sup_{|t|\leq K_n}|\mathbb{H}_n^{\mathbf{B}}(t)| > M\right\}\right)\\
&\leq P^{\mathbf{B}}\left(\inf_{|t|=K_n}\left[-\frac{1}{\sqrt{T_n b_n}}\sum_{i=1}^{k_n} w_i \sum_{j\in B_{k_i}} t^\top \zeta_j\left(\hat{\gamma}_n + \sqrt{\frac{b_n}{T_n}}t\right)\right] \leq 0\right)\\
&+ \frac{1}{M^2}E^{\mathbf{B}}\left[\sup_{|t|\leq K_n}|\mathbb{H}_n^{\mathbf{B}}(t)|^2\right]\\
&= o_p(1).
\end{aligned} \tag{58}$$

Combined with the estimates of $\mathbb{H}_{1,n}^{\mathbf{B}}$ and $\mathbb{H}_{2,n}^{\mathbf{B}}$, we obtain (43) and (44).

5.3.2 Drift part

Since the route is a similar to (43) and (44), we sometimes omit the details below. Introduce the function $\mathbb{U}_n^{\mathbf{B}}$ on $\mathbb{R}^{p_\alpha}$ by

$$\mathbb{U}_n^{\mathbf{B}}(t) = \frac{1}{\sqrt{T_n}} \sum_{i=1}^{k_n} w_i \sum_{j \in B_j} \left\{ \eta_j\left(\hat{\alpha}_n + \frac{t}{\sqrt{T_n}}, \hat{\gamma}_n \right) - \eta_j\left(\hat{\alpha}_n, \hat{\gamma}_n\right) \right\} - \frac{1}{T_n} \partial_\alpha^{\otimes 2} \mathbb{H}_{2,n}(\hat{\alpha}_n) t,$$

and Taylor's expansion gives

$$\mathbb{U}_n^{\mathbf{B}}(t) = \mathbb{U}_{1,n}^{\mathbf{B}}(t) + \mathbb{U}_{2,n}^{\mathbf{B}}(t) + \mathbb{U}_{3,n}^{\mathbf{B}}(t),$$

where

$$\mathbb{U}_{1,n}^{\mathbf{B}}(t) = \frac{1}{T_n} \sum_{i=1}^{k_n} (w_i - 1) \sum_{j \in B_{k_i}} t^\top \partial_\alpha \eta_j(\alpha^\star, \hat{\gamma}_n),$$

$$\mathbb{U}_{2,n}^{\mathbf{B}}(t) = \frac{1}{T_n} \sum_{i=1}^{k_n} (w_i - 1) \sum_{j \in B_{k_i}} \int_0^1 \partial_\alpha^{\otimes 2} \eta_j(\hat{\alpha}_n + s(\alpha^\star - \hat{\alpha}_n), \hat{\gamma}_n) ds[t, \hat{\alpha}_n - \alpha^\star],$$

$$\mathbb{U}_{3,n}^{\mathbf{B}}(t) = \frac{1}{T_n^{3/2}} \sum_{i=1}^{k_n} w_i \sum_{j \in B_{k_i}} \int_0^1 \partial_\alpha^{\otimes 2} \eta_j\left(\hat{\alpha}_n + \frac{1}{\sqrt{T_n}} ut, \hat{\gamma}_n \right) du[t, t].$$

As can be seen in the proof of (43), it is sufficient for (45) to show that

$$E^{\mathbf{B}}\left[\sup_{|t| \le K_n} \left| \mathbb{U}_n^{\mathbf{B}}(t) \right|^2 \right] = o_p(1), \tag{59}$$

$$\frac{1}{T_n} \sum_{i=1}^{k_n} \left| \sum_{j \in B_{k_i}} \eta_j\left(\alpha^\star, \hat{\gamma}_n\right) \right|^2 = O_p(1), \tag{60}$$

where (K_n) denotes the same positive sequence as the previous part. We first show (59). Decompose $\mathbb{U}_{1,n}^{\mathbf{B}}(t)$ as

$$\begin{aligned} \mathbb{U}_{1,n}^{\mathbf{B}}(t) &= \frac{1}{T_n} \sum_{i=1}^{k_n} (w_i - 1) \sum_{j \in B_{k_i}} t^\top \partial_\alpha \eta_j(\alpha^\star, \gamma^\star) \\ &\quad + \frac{1}{T_n} \sum_{i=1}^{k_n} (w_i - 1) \sum_{j \in B_{k_i}} \int_0^1 \partial_\gamma \partial_\alpha \eta_j(\alpha^\star, \hat{\gamma}_n + u(\gamma^\star - \hat{\gamma}_n)) du[t, \hat{\gamma}_n - \gamma^\star]. \end{aligned}$$

Since $\partial_\alpha \eta_j(\alpha^\star, \gamma^\star)$ can be rewritten as:

$$\partial_\alpha \eta_j(\alpha^\star, \gamma^\star) = \frac{\partial_\alpha^{\otimes 2} a_{t_{j-1}}}{c_{t_{j-1}}^2} \int_{t_{j-1}}^{t_j} (A_s - a_{t_{j-1}}) ds + \frac{\partial_\alpha^{\otimes 2} a_{t_{j-1}}}{c_{t_{j-1}}^2} \int_{t_{j-1}}^{t_j} C_{s-} dZ_s$$
$$+ h_n \left[\frac{\partial_\alpha^{\otimes 2} a_{t_{j-1}}}{c_{t_{j-1}}^2} a_{t_{j-1}} + \frac{\partial_\alpha^{\otimes 2} a_{t_{j-1}} - \left(\partial_\alpha a_{t_{j-1}} \right)^{\otimes 2}}{c_{t_{j-1}}^2} \right],$$

it follows from Jensen's inequality and Lemma 19 that

$$E^{\mathbf{B}} \left[\sup_{|t| \le K_n} \left| \frac{1}{T_n} \sum_{i=1}^{k_n} (w_i - 1) \sum_{j \in B_{k_i}} t^\top \partial_\alpha \eta_j(\alpha^\star, \gamma^\star) \right|^2 \right] = O_p(K_n^2 k_n^{-1}) = o_p(1).$$

Again applying Jensen's inequality, we obtain

$$E^{\mathbf{B}} \left[\sup_{|t| \le K_n} \left| \frac{1}{T_n} \sum_{i=1}^{k_n} (w_i - 1) \sum_{j \in B_{k_i}} \int_0^1 \partial_\gamma \partial_\alpha \eta_j(\alpha^\star, \hat{\gamma}_n + u(\gamma^\star - \hat{\gamma}_n)) du [t, \hat{\gamma}_n - \gamma^\star] \right|^2 \right]$$
$$\lesssim \frac{K_n^2 |\hat{\gamma}_n - \gamma^\star|^2}{T_n^2} \sum_{i=1}^{k_n} \left| \sum_{j \in B_{k_i}} \int_0^1 \partial_\gamma \partial_\alpha \eta_j(\alpha^\star, \hat{\gamma}_n + u(\gamma^\star - \hat{\gamma}_n)) du \right|^2$$
$$\le \frac{K_n^2 |\hat{\gamma}_n - \gamma^\star|^2}{T_n^2} \sum_{i=1}^{k_n} \sup_{\gamma \in \Theta} \left| \sum_{j \in B_{k_i}} \partial_\gamma \partial_\alpha \eta_j(\alpha^\star, \gamma) \right|^2.$$

For any $q > p_\gamma$, Sobolev's inequality (cf. Adams 1973) gives

$$E \left[\sup_{\gamma \in \Theta} \left| \sum_{j \in B_{k_i}} \partial_\gamma \partial_\alpha \eta_j(\alpha^\star, \gamma) \right|^q \right]$$
$$\lesssim \sup_{\gamma \in \Theta} E \left[\left| \sum_{j \in B_{k_i}} \partial_\gamma \partial_\alpha \eta_j(\alpha^\star, \gamma) \right|^q \right] + \sup_{\gamma \in \Theta} E \left[\left| \sum_{j \in B_{k_i}} \partial_\gamma^{\otimes 2} \partial_\alpha \eta_j(\alpha^\star, \gamma) \right|^q \right].$$

Now, we focus on the first term of the right-hand side. $\partial_\gamma \partial_\alpha \eta_j(\alpha^\star, \gamma)$ can be decomposed as:

$$\partial_\gamma \partial_\alpha \eta_j(\alpha^\star, \gamma) = \partial_\gamma c_{t_{j-1}}^{-2}(\gamma) \partial_\alpha^{\otimes 2} a_{t_{j-1}} \left(\int_{t_{j-1}}^{t_j} A_s ds + \int_{t_{j-1}}^{t_j} C_{s-} dZ_s \right)$$
$$- h_n \partial_\gamma c_{t_{j-1}}^{-2}(\gamma) \left(\partial_\alpha a_{t_{j-1}} \right)^{\otimes 2}.$$

Jensen's inequality gives

$$\sup_{\gamma\in\Theta_\gamma} E\left[\left|\sum_{j\in B_{k_i}}\left[\partial_\gamma c_{t_{j-1}}^{-2}(\gamma)\partial_\alpha^{\otimes 2}a_{t_{j-1}}\int_{t_{j-1}}^{t_j}A_s ds - h_n\partial_\gamma c_{t_{j-1}}^{-2}(\gamma)\left(\partial_\alpha a_{t_{j-1}}\right)^{\otimes 2}\right]\right|^q\right] \lesssim \frac{T_n^q}{k_n^q}.$$

Burkholder's inequality leads to

$$\begin{aligned}
&\sup_{\gamma\in\Theta_\gamma} E\left[\left|\sum_{j\in B_{k_i}}\partial_\gamma c_{t_{j-1}}^{-2}(\gamma)\partial_\alpha^{\otimes 2}a_{t_{j-1}}\int_{t_{j-1}}^{t_j}C_{s-}dZ_s\right|^q\right]\\
&= \sup_{\gamma\in\Theta_\gamma} E\left[\left|\int_{(i-1)c_nh_n}^{ic_nh_n}\sum_{j\in B_{k_i}}\partial_\gamma c_{t_{j-1}}^{-2}(\gamma)\partial_\alpha^{\otimes 2}a_{t_{j-1}}C_{s-}\chi_j(s)dZ_s\right|^q\right]\\
&\lesssim \frac{T_n^{q/2-1}}{k_n^{q/2-1}}\sup_{\gamma\in\Theta}\sum_{j\in B_{k_i}}\int_{t_{j-1}}^{t_j}E\left[\left|\partial_\gamma c_{t_{j-1}}^{-2}(\gamma)\partial_\alpha^{\otimes 2}a_{t_{j-1}}C_s\right|^q\right]ds\\
&\lesssim \frac{T_n^{q/2}}{k_n^{q/2}},
\end{aligned}$$

so that for any $q > p_\gamma$,

$$E\left[\sup_{\gamma\in\Theta}\left|\sum_{j\in B_{k_i}}\partial_\gamma\partial_\alpha\eta_j(\alpha^\star,\gamma)\right|^2\right] \leq E\left[\sup_{\gamma\in\Theta}\left|\sum_{j\in B_{k_i}}\partial_\gamma\partial_\alpha\eta_j(\alpha^\star,\gamma)\right|^q\right]^{2/q} \lesssim \frac{T_n^2}{k_n^2}.$$

Analogously, we can evaluate $\sup_{\gamma\in\Theta} E\left[\left|\sum_{j\in B_{k_i}}\partial_\gamma^{\otimes 2}\partial_\alpha\eta_j(\alpha^\star,\gamma)\right|^q\right]$, and the tightness of $\sqrt{T_n}(\hat{\gamma}_n - \gamma^\star)$ leads to

$$\begin{aligned}
&E^{\mathbf{B}}\left[\sup_{|t|\leq K_n}\left|\frac{1}{T_n}\sum_{i=1}^{k_n}(w_i-1)\sum_{j\in B_{k_i}}\int_0^1\partial_\gamma\partial_\alpha\eta_j(\alpha^\star,\hat{\gamma}_n+u(\gamma^\star-\hat{\gamma}_n))du[t,\hat{\gamma}_n-\gamma^\star]\right|^2\right]\\
&= O_p\left(K_n^2k_n^{-1}T_n^{-1}\right).
\end{aligned}$$

Hence, we obtain

$$E^{\mathbf{B}}\left[\sup_{|t|\leq K_n}\left|\mathbb{U}_{1,n}^{\mathbf{B}}(t)\right|^2\right] = o_p(1).$$

By taking a similar route, it is easy to see

$$E^{\mathbf{B}}\left[\sup_{|t|\leq K_n}\left|\mathbb{U}_{2,n}^{\mathbf{B}}(t)\right|^2\right] + E^{\mathbf{B}}\left[\sup_{|t|\leq K_n}\left|\mathbb{U}_{3,n}^{\mathbf{B}}(t)\right|^2\right] = o_p(1),$$

and in turn we get (59). We next show (60). Taylor's formula leads to

$$\frac{1}{T_n}\sum_{i=1}^{k_n}\left|\sum_{j\in B_{k_i}}\eta_j\big(\alpha^\star,\hat{\gamma}_n\big)\right|^2$$
$$\lesssim \frac{1}{T_n}\sum_{i=1}^{k_n}\left|\sum_{j\in B_{k_i}}\eta_j\big(\alpha^\star,\gamma^\star\big)\right|^2+\frac{|\hat{\gamma}_n-\gamma^\star|^2}{T_n}\sum_{i=1}^{k_n}\left|\sum_{j\in B_{k_i}}\int_0^1\partial_\gamma\eta_j\big(\alpha^\star,\hat{\gamma}_n+u(\gamma^\star-\hat{\gamma}_n)\big)du\right|^2, \tag{61}$$

and the first term of the right-hand side is $O_p(1)$ from an easy application of Lemma 19. As for the second term of the right-hand side, it is $O_p(k_n^{-1})$ by making use of the same argument based on Sobolev's inequality presented above. Hence, (60) follows, and by mimicking the proof of (53), we get

$$P^{\mathbf{B}}\left(\inf_{|t|=K_n}\left[-\frac{1}{\sqrt{T_n}}\sum_{i=1}^{k_n}w_i\sum_{j\in B_{k_i}}t^\top\eta_j\left(\hat{\alpha}_n+\frac{t}{\sqrt{T_n}},\hat{\gamma}_n\right)\right]>0\right)=1-o_p(1).$$

Again by using (Ortega and Rheinboldt 1970, Theorem 6.3.4), it turns out that the equation

$$\mathbb{G}^{\mathbf{B}}_{2,n}\left(\hat{\alpha}_n+\frac{t}{\sqrt{T_n}}\right)=0, \tag{62}$$

has a root t within $|t|\leq K_n$ on the set

$$\inf_{|t|=K_n}\left[-\frac{1}{\sqrt{T_n}}\sum_{i=1}^{k_n}w_i\sum_{j\in B_{k_i}}t^\top\eta_j\left(\hat{\alpha}_n+\frac{t}{\sqrt{T_n}},\hat{\gamma}_n\right)\right]>0.$$

Hence, for $\hat{\alpha}^{\mathbf{B}}_n := \hat{\alpha}_n+t/\sqrt{T_n}$, we get

$$P^{\mathbf{B}}(\hat{\alpha}^{\mathbf{B}}_n\in\Theta_\alpha)=1-o_p(1).$$

Mimicking (58), we get (45). It remains to prove (46), but it automatically follows from the estimates of (61). Combined with (43) and (44), we obtain (19). Moreover, Taylor's expansion and the calculation up to here lead to

$$\hat{A}_n\bar{\Gamma}_n^{1/2}(\hat{\theta}^{\mathbf{B}}_n-\hat{\theta}_n)=\hat{B}_n\sum_{i=1}^{k_n}(w_i-1)\sum_{j\in B_{k_i}}\begin{pmatrix}\dfrac{\partial_\gamma c_{t_{j-1}}}{c^3_{t_{j-1}}}\left[h_nc^2_{t_{j-1}}-(\Delta_jX)^2\right]\\ \dfrac{\partial_\alpha a_{t_{j-1}}}{c^2_{t_{j-1}}}\left(\Delta_jX-h_na_{t_{j-1}}\right)\end{pmatrix}+r_{nB}. \tag{63}$$

Now, we move to the proof of (20), (21), (22), and (23). In the correctly specified case, by taking Lemma 19 and Lemma 20 into consideration, (20) and (22) are trivial from (63). Concerning the misspecified case, it is enough for (21) and (23) to show that for each $l\in\{1,2\}$,

$$\frac{1}{\sqrt{T_n}}\sum_{i=1}^{k_n}(w_i-1)\big(f_{l,ic_nh_n}-f_{l,[(i-1)c_n+1]h_n}\big)=r_{nB}, \tag{64}$$

$$\frac{1}{\sqrt{T_n}}\sum_{i=1}^{k_n}(w_i-1)\big(g_{l,ic_nh_n}-g_{l,[(i-1)c_n+1]h_n}\big)=r_{nB}. \tag{65}$$

Since f_1 and f_2 is at most polynomial growth, it follows that for each $l\in\{1,2\}$, there exist positive constants C_1 and C_2 satisfying

$$\max_{i\in\{1,\dots,k_n\}}E\left[\left|f_{l,ic_nh_n}-f_{l,[(i-1)c_n+1]h_n}\right|^2\right]\le C_1+\sup_{t\ge 0}E[|X_t|^{C_2}]<\infty.$$

Hence, we have

$$\begin{aligned}&E^{\mathbf{B}}\left[\frac{1}{T_n}\sum_{i=1}^{k_n}(w_i-1)^2\left|f_{1,ic_nh_n}-f_{1,[(i-1)c_n+1]h_n}\right|^2\right]\\&=\frac{1}{T_n}\sum_{i=1}^{k_n}\left|f_{1,ic_nh_n}-f_{1,[(i-1)c_n+1]h_n}\right|^2\\&=O_p\left(\frac{k_n}{T_n}\right),\end{aligned}$$

and (64). Since g_1 and g_2 are also polynomial growth from their weighted Hölder continuity, (65) can be deduced in the same way.

It remains to prove (24). However, (24) follows from $\mathcal{F}$-conditional Lindeberg-Feller central limit theorem by making use Lemma 22 and Lemma 23. Hence, the proof is complete. □

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10463-022-00854-2.

Acknowledgements The author would like to thank Professor Y.Koike for his constructive advice on bootstrap weights. He is also grateful to the anonymous referees for their valuable comments.

References

Adams, R. A. (1973). Some integral inequalities with applications to the imbedding of Sobolev spaces defined over irregular domains. *Transactions of the American Mathematical Society, 178,* 401–429.

Applebaum, D. (2009). *Lévy processes and stochastic calculus*, Cambridge studies in advanced mathematics, vol 116, 2nd ed. Cambridge University Press, Cambridge, https://doi.org/10.1017/CBO9780511809781,

Chatterjee, S., Bose, A. (2005). Generalized bootstrap for estimating equations. *The Annals of Statistics, 33*(1), 414–436. https://doi.org/10.1214/009053604000000904.

Dovonon, P., Gonçalves, S., Hounyo, U., Meddahi, N. (2019). Bootstrapping high-frequency jump tests. *Journal of the American Statistical Association, 114*(526), 793–803. https://doi.org/10.1080/01621459.2018.1447485.

Efron, B. (1979). Bootstrap methods: another look at the jackknife. *The Annals of Statistics, 7*(1), 1–26. http://links.jstor.org/sici?sici=0090-5364(197901)7:1%3C1:BMALAT%3E2.0.CO;2-6&origin=MSN.

Genon-Catalot, V., Jacod, J. (1993). On the estimation of the diffusion coefficient for multi-dimensional diffusion processes. *Annales de l'Institut Henri Poincaré Probabilités et Statistiques, 29*(1), 119–151. http://www.numdam.org/item?id=AIHPB_1993__29_1_119_0.

Gonçalves, S., Meddahi, N. (2009). Bootstrapping realized volatility. *Econometrica Journal of the Econometric Society, 77*(1), 283–306. https://doi.org/10.3982/ECTA5971.

Kessler, M. (1997). Estimation of an ergodic diffusion from discrete observations. *Scandinavian Journal of Statistics Theory and Applications, 24*(2), 211–229. https://doi.org/10.1111/1467-9469.00059.

Kulik, A. M., Veretennikov, A. Y. (2011). Extended Poisson equation for weakly ergodic Markov processes. *Theory of Probability and Mathematical Statistics, 85,* 22–38.

Li, J., Todorov, V., Tauchen, G., Chen, R. (2017). Mixed-scale jump regressions with bootstrap inference. *Journal of Econometrics, 201*(2), 417–432. https://doi.org/10.1016/j.jeconom.2017.08.017.

Mammen, E. (1993). Bootstrap and wild bootstrap for high-dimensional linear models. *The Annals of Statistics, 21*(1), 255–285. https://doi.org/10.1214/aos/1176349025.

Masuda, H. (2007). Ergodicity and exponential β-mixing bounds for multidimensional diffusions with jumps. *Stochastic Processes and their Applications, 117*(1), 35–56. https://doi.org/10.1016/j.spa.2006.04.010.

Masuda, H. (2013). Convergence of gaussian quasi-likelihood random fields for ergodic lévy driven sde observed at high frequency. *The Annals of Statistics, 41*(3), 1593–1641.

Masuda, H., Uehara, Y. (2017). On stepwise estimation of lévy driven stochastic differential equation (japanese). *Proceedings of the institute of statistical mathematics* 65(1).

Ortega, J. M., Rheinboldt, W. C. (1970). *Iterative solution of nonlinear equations in several variables*. New York-London: Academic Press.

Pardoux, E., Veretennikov, A. Y. (2001). On the Poisson equation and diffusion approximation. *I. The Annals of Probability, 29*(3), 1061–1085. https://doi.org/10.1214/aop/1015345596.

Uchida, M., Yoshida, N. (2011). Estimation for misspecified ergodic diffusion processes from discrete observations. *ESAIM Probability and Statistics, 15,* 270–290. https://doi.org/10.1051/ps/2010001.

Uchida, M., Yoshida, N. (2012). Adaptive estimation of an ergodic diffusion process based on sampled data. *Stochastic Processes and their Applications, 122*(8), 2885–2924. https://doi.org/10.1016/j.spa.2012.04.001.

Uehara, Y. (2019). Statistical inference for misspecified ergodic Lévy driven stochastic differential equation models. *Stochastic Processes and their Applications, 129*(10), 4051–4081. https://doi.org/10.1016/j.spa.2018.11.007.

Yoshida, N. (1992). Estimation for diffusion processes from discrete observation. *Journal of Multivariate Analysis, 41*(2), 220–242. https://doi.org/10.1016/0047-259X(92)90068-Q.