Statistical inference for ergodic non-Gaussian stochastic differential equation models

上原, 悠槙

出版情報：Kyushu University, 2018, 博士（数理学）, 論文博士
バージョン：
権利関係：

Faculty of Mathematics
Kyushu University

Doctoral Dissertation

Statistical inference for ergodic non-Gaussian stochastic differential equation models

by

Yuma Uehara

January 22, 2019

Abstract

This doctoral dissertation is based on the three paper: [44], [43], and [62]. Throughout this dissertation, we look on Lévy driven stochastic differential equation (SDE) models based on high-frequency samples. The models are used for modeling various time varying phenomena. However, due to the complexity of small-time activity of driving Lévy noises, even their estimation theory is still under development, which motivates us to work on this dissertation.

In this dissertation, we propose two estimation method for Lévy driven SDE models. The first one is the stepwise Gaussian quasi-likelihood method which enables us to deal with various kind of driving noises, and to reduce the computational load of calculating compared with the ordinary Gaussian quasi-likelihood method. In deriving the asymptotic behavior of the stepwise Gaussian quasi-likelihood estimator, we consider model misspecification. Model misspecification is essentially inevitable in statistical modeling, but it has not been cared in the estimation theory of Lévy driven stochastic differential equation models. By utilizing the concept of the extended generator of Feller Markov process, we correct the misspecification bias and give the consistency and asymptotic normality of the estimator.

The second one is based on the iterative Jarque-Bera normality test for the estimation of the continuous component of jump diffusion models being included in Lévy driven SDE models. Compared with the existing estimation methods, the method has an advantage in that there is not any sensitive tuning parameter. We will show that our estimator has the same asymptotic behavior as the estimator constructed by non-observed continuous part fluctuation.

In addition to the above two theoretical result, we will give the specification of the functions `qmleLevy` implemented in the YUIMA package on R and `snr` under development. They execute our proposed estimation methods, and some example codes will be exhibited.

Acknowledgement

The author would like to express his deep gratitude to Professor Hiroki Masuda for his passionate guidance and encouragement through the learning precess in his doctoral course. He is very grateful to Professor Ryuei Nishii, Professor Yoshihiko Maesono, and Associate professor Kei Hirose for their valuable comments. He also thanks the past and present member of Masuda laboratory for daily profitable discussion and warm encouragement. At last but not least, my heartfelt appreciation goes to my family who have supported and raised me up to the present.

Yuma Uehara
January, 2019

Contents

Chapter 1

Introduction

Recent development of observation techniques and computers enables us to obtain huge amounts of time series data. To extract beneficial information from such data, a number of statistical models and methods have been proposed up to present. Among them, stochastic differential equation (SDE) models are regarded as good candidates in the sense that they can incorporate the frequency of observed data into modeling by considering (virtual) continuous time axis behind the observations. In this dissertation, we especially focus on SDE models driven by Lévy processes. Lévy processes are defined as stochastically continuous time processes which have stationary and independent increments, and they can be regarded as continuous time random walk. It is known that for a Lévy process Y, there exists an infinitely divisible distribution μ satisfying $Y_1 \sim \mu$. The importance of infinitely divisible distributions in statistical analysis is mentioned by [58], and the comprehensive theoretical review of them is sumarized for instance in [52] and [59]. The class of infinitely divisible distributions contain many widely used distributions, for example, normal distribution, stable distribution, Poisson distribution, generalized hyperbolic distribution, to mention few. Hence changing the corresponding infinitely divisible distribution, we can get a Lévy process well describing non-Gaussian fluctuation (jumps) often seen in time series data. For the theoretical accounts of Lévy processes, we refer to [52], [3], [8], and so on.

By the aforementioned reasons, the estimation theory of Lévy driven SDE models especially based on high-frequency samples has been studied so far, for instance, the threshold based estimation for jump diffusion models by [48] and [57], the least absolute deviation (LAD)-type estimation for Lévy driven Ornstein-Uhlenbeck models by [38], the non-Gaussian stable quasi-likelihood estimation for locally stable driven SDE models by [42], the least square estimation for small Lévy driven SDE models by [33], the Gaussian quasi-likelihood (GQL) for ergodic Lévy driven SDE models by [39] and [43], and so on. These are on parametric methods, and concurrently, nonparametric methods have been investigated, for example, the functional estimation and adaptive estimation for jump diffusion models by [4] and [54], Nadaraya-Watson estimation for stable driven SDE models by [54], and the Fourier based method for Lévy process and Lévy type model by [9].

The primary objective of this dissertation is to further develop the estimation theory of Lévy driven SDE models. The rest of this dissertation is organized as follows. In Chapter 2, the estimation theory under model misspecification is considered. Model misspecification cannot be avoided in statistical modeling, but in Lévy driven SDE models, it has been ignored. In the chapter, we especially reflect on the Gaussian quasi-likelihood estimation which is robust against the distribution of the driving noise. We will give the asymptotic behavior of the corresponding estimator by using the theory of extended Poisson equation proposed by [64]. In Chapter 3, we consider the estimation problem of the continuous part of jump diffusion models being one of the most important subclasses of the Lévy driven SDE models. Although the estimation of the models is often done by threshold based approach ([34], [57]), the approach has an annoying tuning parameter. Our method is the iterative Jarque-Bera normality test based on the self-normalized residual, and we do not have to choose a sensitive quantity. We will show that our estimator has the same distribution as the estimator which is constructed by non-observed continuous part fluctuation. Chapter 4 gives the specification of the R function `qmleLevy` conducting the Gaussian quasi-likelihood estimation of Lévy driven SDE models and `snr` under development for Some example codes will also be presented.

In the end, we table some notations used through this dissertation.

- For any vector v, we describe $v^{(l)}$ as its l-th element.
- For any matrix S, $|S|$ denotes its Frobenius norm.
- I_p represents the p-dimensional identity matrix.
- C denotes a universal positive constant which may vary in each context.
- $x_n \lesssim y_n$ implies that there exists a positive constant C being independent of n satisfying $x_n \leq Cy_n$ for all large enough n.
- $\bar{S}$ denotes the closure of any set S.
- $\top$ denotes the transpose operator, and we write $x^{\otimes 2} = x^\top x$ for any vector x.
- For any vector variable $x = (x^{(i)})$, we write $\partial_x = \left(\frac{\partial}{\partial x^{(i)}}\right)_i$. Here, $\partial_x^{(i)}$ is referred to as a differential operator for any variable $x^{(i)}$.
- The convergence 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.

Chapter 2

Statistical inference for misspecified ergodic Lévy driven SDE models

2.1 Introduction

In statistical modeling, we always face the risk of model misspecification. The statistical theory under model misspecification tells us how close an estimated model is to the data-generating model, and such interpretation is important, for example, in ensuring the reliability of estimation methods, and comparing candidate description models by information criterions. Historically, following pioneering works by [10], [23], and [68], the theory has been investigated up to the present for such reasons. Especially about SDE models, for instance, [45], [60] and [32, Section 3] focus on misspecified diffusion models; [32, Section 4] deal with the misspecification with respect to the intensity function of Poisson processes; [35] considers the situation where the given model is diffusion but the data-generating model has jumps. However, the theory does not seem to be well developed in the context of Lévy driven SDE models whose coefficients take various non-linear form, and indeed the parametric methods introduced above are not discussed under model misspecification.

In this chapter, the data-generating process X which is defined on the complete filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in\mathbb{R}_+}, P)$ is supposed to be the solution of the following Lévy driven SDE:

$$dX_t = A(X_t)dt + C(X_{t-})dZ_t, \tag{2.1.1}$$

where:

- Z is a one-dimensional càdlàg Lévy process without Wiener part. It is independent of the initial variable X_0 and satisfies

$$E[Z_1] = 0, Var[Z_1] = 1, E[|Z_1|^q] < \infty,$$

for all $q > 0$;

- The coefficients $A : \mathbb{R} \mapsto \mathbb{R}$ and $C : \mathbb{R} \mapsto \mathbb{R}$ are Lipschitz continuous;
- $\mathcal{F}_t := \sigma(X_0) \vee \sigma(Z_s; s \le t)$.

We suppose that the discrete but high-frequency observations $(X_{t_0}, \dots, X_{t_n})$ are obtained from X in the so-called "rapidly increasing experimental design", that is,

$$t_j \equiv t_j^n := jh_n, T_n := nh_n \to \infty, nh_n^2 \to 0.$$

For the observations $(X_{t_0}, \dots, X_{t_n})$, we suppose that the following parametric one-dimensional SDE model is allocated:

$$dX_t = a(X_t, \alpha)dt + c(X_{t-}, \gamma)dZ_t, \tag{2.1.2}$$

where the functional forms of the coefficients $a : \mathbb{R} \times \Theta_\alpha \mapsto \mathbb{R}$ and $c : \mathbb{R} \times \Theta_\gamma \mapsto \mathbb{R}$ are supposed to be known except for a finite-dimensional unknown parameter $\theta := (\gamma, \alpha)$ being an element of the bounded convex domain $\Theta := \Theta_\gamma \times \Theta_\alpha \subset \mathbb{R}^p$. We note that the true coefficients $(C, A)(\cdot)$ may not belong to the parametric family $\{(c, a)(\cdot, \theta) : \theta \in \Theta\}$, namely, the misspecification of the coefficient possibly occurs. Hereinafter, the terminologies "misspecified" and "misspecification" will be used for the misspecification with respect to the coefficients unless otherwise mentioned.

To estimate an optimal parameter $\theta^\star$ of θ, we utilize the GQL procedure used in [43]. Concerning misspecified ergodic diffusion models, it is shown in [60] that although the misspecification with respect to their diffusion term deteriorates the convergence rate of the scale (diffusion) parameter, the Gaussian quasi-maximum likelihood estimator (GQMLE) still has asymptotic normality. We will show that asymptotic normality of the GQMLE holds in the misspecified ergodic Lévy driven SDE models as well. To handle the misspecification effect, we will invoke the theory of the extended Poisson equation (EPE) for homogeneous Feller Markov processes established in [64]. Applying the result of [64] for (2.1.1), the existence and weighted Hölder regularity of the solution of EPEs will be shown under a mighty mixing condition on X. Building on the result and martingale representation theorem, we will be able to get the asymptotic normality of our estimator and its tail probability estimates under sufficient regularity and moment conditions on the ingredients of (2.1.1) and (2.1.2). We note that the absence of Wiener part in (2.1.1) is essential while it is not in the correctly specified case, for more details, see Remark 2.3.10.

It will turn out that the convergence rate of the scale parameters is $\sqrt{T_n}$, and it is the same as the correctly specified case. This is different from the diffusion case (cf. Table 2.1). Such difference may be caused from applying the GQL to non-Gaussian driving noises, that is, the efficiency loss of the GQMLE may occur even in the correctly

Table 2.1: GQL approach for ergodic diffusion models and ergodic Lévy driven SDE models

Model	Rates of convergence drift	Rates of convergence scale	Ref.
correctly specified diffusion	$\sqrt{T_n}$	$\sqrt{n}$	[27], [61]
misspecified diffusion	$\sqrt{T_n}$	$\sqrt{T_n}$	[60]
correctly specified Lévy driven SDE	$\sqrt{T_n}$	$\sqrt{T_n}$	[40], [43]
misspecified Lévy driven SDE	$\sqrt{T_n}$	$\sqrt{T_n}$	this chapter

specified case. Indeed, the non-Gaussian stable quasi-likelihood is known to estimate the drift and scale parameters faster than the GQMLE in correctly specified locally β-stable driven SDE models (cf. [42]); each of their convergence rates are $\sqrt{n}h_n^{1-1/\beta}$ and $\sqrt{n}$, respectively. Further, for correctly specified locally β-stable driven Ornstein-Uhlenbeck models, the LAD-type estimators of [38] tend to the true value at the speed of $\sqrt{n}h_n^{1-1/\beta}$ and it is also faster than that of the GQMLE. However, in exchange for its efficiency, the GQL approach is worth considering by the following reasons:

- It does not include any special functions (e.g. Bessel function, Whittaker function, and so on), infinite expansion series and analytically unsolvable integrals, thus computation based on it is not relatively time-consuming.

- It focuses only on the (conditional) mean and covariance structure, thus it does not need so much restriction on the driving noise and is robust against the noise structure. In other words, we can construct reasonable estimators of the drift and scale coefficients in the unified way if only the driving Lévy noise has moments of any order.

Our result ensures that even if the true coefficients are misspecified and take non-linear forms, the staged GQL estimation still works for Lévy driven SDE models and completely inherits its merit written in above.

The rest of this chapter is organized as follows: In Section 2.2, we introduce assumptions and our estimation procedure. Section 2.3 provides our main results in the following turn:

1. the tail probability estimates of the GQMLE (Theorem 2.3.1);

2. the existence and weighted Hölder regularity of the solution of EPEs for Lévy driven SDEs (Proposition 2.3.5);

3. the asymptotic normality of the GQMLE at $\sqrt{T_n}$-rate (Theorem 2.3.7).

A simple numerical experiment is presented in Section 2.4. We give all proofs of our results in Section 2.5.

2.2 Assumptions and Estimation scheme

Before we introduce technical assumptions, we additionally introduce some notations. The part of them will be shared with Chapter 3.

- $P_t(x,\cdot)$ denotes the transition probability of X.
- We write $Y_j = Y_{t_j}$ and $\Delta_j Y = Y_j - Y_{j-1}$ for any stochastic process Y.
- ν_0 represents the Lévy measure of Z.
- $E^j[\cdot]$ denotes the conditional expectation with respect to $\mathcal{F}_{t_j}$.
- η stands for the law of X_0.
- For any matrix valued function f on $\mathbb{R}\times\Theta$, we write $f_s(\theta) = f(X_s,\theta)$; especially we write $f_j(\theta) = f(X_j,\theta)$. We sometimes write f_s and f_{j-1} instead of $f_s(\theta_0)$ or $f_s(\theta^\star)$, and $f_{j-1}(\theta_0)$ or $f_{j-1}(\theta^\star)$ just for simplicity where θ_0 and $\theta^\star$ are the true value and the optimal value of θ, respectively (the definition of $\theta^\star$ will be introduced later).

To derive our asymptotic results, we introduce some assumptions with some technical comments. Most of them are almost the same as in [40], [43], and [44], except for Assumption 2.2.1-(2).

Assumption 2.2.1. *1. $E[Z_1] = 0$, $Var[Z_1] = 1$, and $E[|Z_1|^q] < \infty$ for all $q > 0$.*

2. The Blumenthal-Getoor index (BG-index) of Z is smaller than 2, that is,

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

From [52, Theorem 25.3], it is easy to observe that Assumption 2.2.1 holds if the Lévy measure ν_0 admits a density g with respect to Lebesgue measure satisfying that $g(z) = O(|z|^{-2-\delta})$ as $|z|\to 0$ for some $\delta\in(0,1)$, and that there exist positive constants K_0, K_1 and K_2 such that

$$g(z) \leq K_0(1+|z|^{K_1})e^{-|z|^{K_2}},$$

for all large enough $|z|$. Via standardization, various Lévy processes fulfill them, for example, bilateral gamma process, normal tempered stable process, normal inverse Gaussian process, and variance gamma process.

In the derivation of the asymptotic normality of our estimator, we will evaluate the small time $L_{2-\epsilon}$-moment of X for some $\epsilon > 0$ (cf. Lemma 2.5.3) to handle the solution of extended Poisson equations which are essential to deal with the misspecification effect; thus the additional condition Assumption 2.2.1-(2) is imposed.

Assumption 2.2.2. *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 $k \in \{0, \dots, 5\}$, 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*

$$\sup_{(x,\alpha,\gamma) \in \mathbb{R} \times \bar{\Theta}_\alpha \times \bar{\Theta}_\gamma} \frac{1}{1 + |x|^{C_{(i,k)}}} \left\{ |\partial_x^i \partial_\alpha^k a(x, \alpha)| + |\partial_x^i \partial_\gamma^k c(x, \gamma)| + |c^{-1}(x, \gamma)| \right\} < \infty. \tag{2.2.1}$$

We note that the first part of Assumption 2.2.1 and Assumption 2.2.2 ensures the existence of a unique càdlàg adapted strong solution of SDE (2.1.1) (cf. [3, Theorem 6.2.3 and Theorem 6.2.9]), that is, there exists a measurable function g such that $X = g(X_0, Z)$.

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

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

Assumption 2.2.3. *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*

$$\sup_{t \in \mathbb{R}_+} \exp(at) ||P_t(x, \cdot) - \pi_0(\cdot)||_{h_q} \leq C_q h_q(x), \tag{2.2.2}$$

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

2. For any $q > 0$, we have

$$\sup_{t \in \mathbb{R}_+} E[|X_t|^q] < \infty. \tag{2.2.3}$$

The former property of this assumption is so-called "f-exponentially ergodic" property (cf. [46]), and putting together with the latter condition and the argument in [27, Lemma 8] and [40, Lemma 4.3], it ensures the ergodic theorem, and its moment bound: for any f being differentiable with derivatives of polynomial growth, we have

$$\frac{1}{n}\sum_{j=1}^{n} f_{j-1} \xrightarrow{p} \int_{\mathbb{R}} f(x)\pi_0(dx), \tag{2.2.4}$$

and for any positive constant $K > 0$,

$$E\left[\left|\sqrt{T_n}\left(\frac{1}{n}\sum_{j=1}^{n} f_{j-1} - \int_{\mathbb{R}} f(x)\pi_0(dx)\right)\right|^K\right] < \infty. \tag{2.2.5}$$

The first convergence in probability (2.2.4) is a standard condition assumed in the statistical theory of the ergodic processes, while the second moment bound (2.2.5) is not and is relatively strong. It will be utilized for evaluating the tail probability of the staged GQL random field introduced later. Such evaluation gives the tail probability estimates of our estimator (Theorem 2.3.1), and in turn, the convergence of moments of any order for it (Remark 2.3.9).

The sufficient conditions of the "f-exponentially ergodic" property for (2.1.1) are investigated by many papers such as [31], [36], and [40]. Among them, we introduce a handy one given in [40, Section 5] in the following:

Condition 1 The coefficients A and C are of class $\mathcal{C}^1$, and globally Lipschitz, and the scale coefficient C is bounded.

Condition 2 The drift coefficient A satisfies

$$\limsup_{|x|\to\infty} \mathrm{sgn}(x)A(x) < 0, \tag{2.2.6}$$

and the scale coefficient $C(x) \neq 0$, for every $x \in \mathbb{R}$.

Condition 3 The Lévy measure ν_0 of Z can be decomposed as: $\nu_0 = \nu_0^\star + \nu_0^\sharp$ for the two Lévy measure, where the restriction of $\nu_0^\star$ to some open set of the form $(-\epsilon, 0)\cup(0,\epsilon)$ with some $\epsilon > 0$ admits a continuously differentiable positive density $g^\star$.

Condition 4 $E[Z_t] = 0$ and $\int \exp(q|z|)\nu_0(dz) < \infty$ for some $q > 0$.

Under Condition 1-Condition 4, Assumption 2.2.3 holds true and for its proof, see [40, Proposition 5.4]. We here note that this sufficient condition still allows the nonlinearity of the coefficients. For example, given a Lévy process Z fulfilling Condition 3 and Condition 4, the following SDEs satisfy Condition 1, Condition 2, and Assumption 2.2.2-(1):

1. $dX_t = -X_t dt + \frac{1}{\sqrt{1+X_{t-}^2}} dZ_t$;
2. $dX_t = -\dfrac{X_t}{\sqrt{1+X_t^2}} dt + dZ_t$;
3. $dX_t = -(X_t + 2\sin X_t)dt + \dfrac{3+X_{t-}^2}{1+X_{t-}^2} dZ_t$.

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:

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

Assumption 2.2.4. Γ *is invertible.*

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

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

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

$$\mathbb{G}_1(\gamma) = -\int_{\mathbb{R}} \left(\log c^2(x,\gamma) + \frac{C^2(x)}{c^2(x,\gamma)} \right) \pi_0(dx), \tag{2.2.7}$$

$$\mathbb{G}_2(\alpha) = -\int_{\mathbb{R}} c(x,\gamma^\star)^{-2}(A(x) - a(x,\alpha))^2 \pi_0(dx). \tag{2.2.8}$$

Note that since we impose the extension condition in Assumption 2.2.2, $\mathbb{Y}(\theta) := (\mathbb{Y}_1(\gamma), \mathbb{Y}_2(\alpha))$ admit extension in $\mathcal{C}(\bar{\Theta})$ as well. Recall that the parameter space Θ is supposed to be a bounded convex domain. We assume the following identifiability condition for $\mathbb{G}_1(\gamma)$ and $\mathbb{G}_2(\alpha)$:

Assumption 2.2.5. $\theta^\star \in \Theta$, *and there exist positive constants* χ_γ *and* χ_α *such that for all* $(\gamma, \alpha) \in \Theta$,

$$\mathbb{Y}_1(\gamma) := \mathbb{G}_1(\gamma) - \mathbb{G}_1(\gamma^\star) \le -\chi_\gamma |\gamma - \gamma^\star|^2, \tag{2.2.9}$$

$$\mathbb{Y}_2(\alpha) := \mathbb{G}_2(\alpha) - \mathbb{G}_2(\alpha^\star) \le -\chi_\alpha |\alpha - \alpha^\star|^2. \tag{2.2.10}$$

(2.2.9) and (2.2.10) ensure the separability of the models which will also be used for the tail probability estimates of the staged GQL random fields, and the next remark provides a sufficient and non-stringent condition for them.

Remark 2.2.6. *If the optimal parameter $\theta^\star \in \Theta$ is unique, and $-\Gamma_\gamma$ and $-\Gamma_\alpha$ are positive definite,* (2.2.9) *and* (2.2.10) *hold true for all $(\gamma, \alpha) \in \Theta$ under Assumption 2.2.2-2.2.3. Let $\mathcal{I}_1 : \Theta_\gamma \mapsto \mathbb{R}^{p_\gamma} \otimes \mathbb{R}^{p_\gamma}$ and $\mathcal{I}_2 : \Theta_\alpha \mapsto \mathbb{R}^{p_\alpha} \otimes \mathbb{R}^{p_\alpha}$ be*

$$
\begin{aligned}
\mathcal{I}_1(\gamma) &= 2\int_{\mathbb{R}} \frac{\partial_\gamma^{\otimes 2} c(x,\gamma) c(x,\gamma) - (\partial_\gamma c(x,\gamma))^{\otimes 2}}{c^4(x,\gamma)} (C^2(x) - c^2(x,\gamma))\pi_0(dx) \\
&\quad - 4\int_{\mathbb{R}} \frac{(\partial_\gamma c(x,\gamma))^{\otimes 2}}{c^4(x,\gamma)} C^2(x)\pi_0(dx), \\
\mathcal{I}_2(\alpha) &= -2\int_{\mathbb{R}} \frac{\partial_\alpha^{\otimes 2} a(x,\alpha)}{c^2(x,\gamma^\star)} (a(x,\alpha) - A(x))\pi_0(dx) - 2\int_{\mathbb{R}} \frac{(\partial_\alpha a(x,\alpha))^{\otimes 2}}{c^2(x,\gamma^\star)} \pi_0(dx).
\end{aligned}
$$

From Assumption 2.2.2 and 2.2.3, the Lebesgue dominated convergence theorem implies that these functions are continuous. Thus, for sufficiently small $\epsilon > 0$, we can pick a positive constant δ satisfying $U_\delta(\gamma^\star) \subset \Theta_\gamma$ and

$$
\inf_{\gamma \in U_\delta(\gamma^\star)} \lambda_{min}(-\mathcal{I}_1(\gamma)) > \epsilon
$$

where $U_\delta(\gamma^\star)$ denotes the open ball of radius δ centered at $\gamma^\star$, and $\lambda_{min}(-\mathcal{I}_1(\gamma))$ is a minimum eigenvalue of $-\mathcal{I}_1(\gamma)$. Then, for every $\gamma \in U_\delta(\gamma^\star)$, we have

$$
\mathbb{Y}_1(\gamma) < -\epsilon|\gamma - \gamma^\star|^2,
$$

by Taylor's formula. Concerning $\gamma \in \Theta_\gamma \setminus U_\delta(\gamma^\star)$, it follows that

$$
\mathbb{Y}_1(\gamma) < -\frac{\mathbb{G}_1(\gamma^\star) - \sup_{\gamma \in \Theta_\gamma \setminus U_\delta(\gamma^\star)} \mathbb{G}_1(\gamma)}{\sup_{\gamma_1, \gamma_2 \in \Theta_\gamma \setminus U_\delta(\gamma^\star)} |\gamma_1 - \gamma_2|^2} |\gamma - \gamma^\star|^2.
$$

Hence (2.2.9) *holds true for all $\gamma \in \Theta_\gamma$ with*

$$
\chi_\gamma = \epsilon \vee \frac{\mathbb{G}_1(\gamma^\star) - \sup_{\gamma \in \Theta_\gamma \setminus U_\delta(\gamma^\star)} \mathbb{G}_1(\gamma)}{\sup_{\gamma_1, \gamma_2 \in \Theta_\gamma \setminus U_\delta(\gamma^\star)} |\gamma_1 - \gamma_2|^2}.
$$

(2.2.10) *can be shown as well.*

From now on, we mention our estimation scheme. Recall that we assume that the observation $(X_{t_0}, \dots, X_{t_n})$ is obtained from X with $t_j \equiv t_j^n := jh_n$, $T_n := nh_n \to \infty$, and $nh_n^2 \to 0$. We define our staged GQMLE $\hat{\theta}_n := (\hat{\gamma}_n, \hat{\alpha}_n)$ in the following manner:

1. *Drift-free estimation of γ.* Define the Maximizing-type estimator (so-called M-estimator) $\hat{\gamma}_n$ by
$$\hat{\gamma}_n \in \underset{\gamma\in\bar{\Theta}_\gamma}{\operatorname{argmax}}\, \mathbb{G}_{1,n}(\gamma),$$
for the $\mathbb{R}$-valued random function
$$\mathbb{G}_{1,n}(\gamma) := -\frac{1}{T_n}\sum_{j=1}^{n}\left\{h_n \log c_{j-1}^2(\gamma) + \frac{(\Delta_j X)^2}{c_{j-1}^2(\gamma)}\right\}.$$

2. *Weighted least square estimation of α.* Define the least square type estimator $\hat{\alpha}_n$ by
$$\hat{\alpha}_n \in \underset{\alpha\in\bar{\Theta}_\alpha}{\operatorname{argmax}}\, \mathbb{G}_{2,n}(\alpha),$$
for the $\mathbb{R}$-valued random function
$$\mathbb{G}_{2,n}(\alpha) := -\frac{1}{T_n}\sum_{j=1}^{n}\frac{(\Delta_j X - h_n a_{j-1}(\alpha))^2}{h_n c_{j-1}^2(\hat{\gamma}_n)}.$$

Remark 2.2.7. *Although our estimation method ignores the drift term in the first stage, the effect of it asymptotically vanishes. This is because the scale term dominates the small time behavior of X in L_2-sense. Specifically, we can derive*
$$E_{j-1}\left[\left(\int_{t_{j-1}}^{t_j} f_{s-}dZ_s\right)^2\right] \lesssim h_n f_{j-1}^2, \quad E_{j-1}\left[\left(\int_{t_{j-1}}^{t_j} g_s ds\right)^2\right] \lesssim h_n^2 g_{j-1}^2,$$
for suitable functions f and g. Indeed, it has already been shown that the asymptotic behavior of the scale estimator constructed by our manner is the same as the conventional GQL estimator in the case of correctly specified ergodic diffusion models (cf. [61]) and ergodic Lévy driven SDE models (cf. [43]). Such ignorance should be helpful in reducing the number of simultaneous optimization parameters, thus our estimator is expected to numerically be more stabilized and their calculation should be less time-consuming. Moreover, by choosing appropriate functional forms, each estimation stage is reduced to a convex optimization problem. For example, if $a(\cdot,\cdot)$ and $c(\cdot,\cdot)$ are linear and log-linear with respect to parameters, respectively, then the above argument holds. As for other candidates of their functional form and details, see [43, Example 3.8].

Remark 2.2.8. *We defined the optimal parameter of θ as the argmax point of $\mathbb{G}_1(\gamma)$ and $\mathbb{G}_2(\alpha)$ and, the two functions are the probability limit of the Gaussian quasi-likelihoods $\mathbb{G}_{1,n}(\gamma)$ and $\mathbb{G}_{2,n}(\alpha)$, respectively. Thus, $-\mathbb{G}_1(\gamma)$ and $-\mathbb{G}_2(\alpha)$ can be regarded as Kullback-Leibler (KL) divergence like quantities between the data-generating model and the parametric model*
$$dX_t = a(X_t,\alpha)dt + c(X_{t-},\gamma)dZ_t.$$

Here we first consider the correctly specified case, that is, there exists an element $\theta_0 := (\gamma_0, \alpha_0) \in \Theta$ *such that* $C(x) = c(x, \gamma_0)$ *and* $A(x) = a(x, \alpha_0)$ *for* π_0 *a.s.* x*. Fix a positive constant* $b > 0$*. Then, it can readily be checked that for all* $x > 0$*,*

$$\log x + \frac{b}{x} \geq \log b + 1,$$

and that both sides are equivalent when $x = b$*. Hence, by Assumption 2.2.5,* $\mathrm{argmax}_{\gamma \in \bar{\Theta}_\gamma} \mathbb{G}_1(\gamma)$ *and* $\mathrm{argmax}_{\alpha \in \bar{\Theta}_\alpha} \mathbb{G}_2(\alpha)$ *coincide with* γ_0 *and* α_0*, respectively. In other words, this asserts that the data-generating model certainly attain the minimization of* $-\mathbb{G}_1(\gamma)$ *and* $-\mathbb{G}_2(\alpha)$*. By taking these insight into consideration, we can intuitively interpret the optimality of* $\theta^\star$ *as the parameter value which yields the closest model to the data-generating model measured by the Kullback-Leibler (KL) divergence like quantities* $-\mathbb{G}_1(\gamma)$ *and* $-\mathbb{G}_2(\alpha)$*.*

2.3 Main results

In this section, we state our main results only for the fully misspecified case, that is, both of the true coefficients C and A do not belong to the parametric family $\{(c, a)(\cdot, \theta) : \theta \in \Theta\}$. Concerning the partly misspecified case (i.e. either of C and A is correctly specified), similar results can be derived just as the corollaries (see, Remark 2.3.8). All of their proofs will be given in Appendix.

The first result provides the tail probability estimates of the normalized $\hat{\theta}_n$ which is theoretically essential such as in the deviation of an information criterion, residual analysis, and the measurement of L_q-prediction error.

Theorem 2.3.1. *Suppose that Assumptions 2.2.1-2.2.4 hold. Then, for any* $L > 0$ *and* $r > 0$*, there exists a positive constant* C_L *such that*

$$\sup_{n \in \mathbb{N}} P\left(\left| \sqrt{T_n}(\hat{\theta}_n - \theta^\star) \right| > r \right) \leq \frac{C_L}{r^L}. \tag{2.3.1}$$

In the correctly specified case, such estimates are already shown in [43] under a sufficient moment and regularity conditions, and strong identifiability conditions, and this theorem extends the results to the misspecified case.

Before we state the asymptotic normality of $\hat{\theta}_n$, we roughly explain how the misspecification effect arises in its derivation process, and introduce the useful tool to deal with it. Except for $o_p(1)$ term, each scaled quasi-score function can be decomposed as:

$$\text{(scaled quasi-score function)} = \text{(stochastic integral)} + \text{(misspecification effect term)}, \tag{2.3.2}$$

where the misspecification effect term is expressed as:

$$\sqrt{\frac{h_n}{n}}\sum_{j=1}^{n} g_{j-1}(\theta^\star) = \frac{1}{\sqrt{T_n}}\int_0^{T_n} g_s(\theta^\star)ds + o_p(1), \tag{2.3.3}$$

with a specific measurable function g satisfying $\pi_0(g) = 0$. The celebrated CLT-type theorems for such single functional integration of Markov processes have been reported in many literatures, for example, [11, Theorem 2.1], [24, Theorem VIII 3.65], [29, Theorem 2.1], [65, Corollary 4.1], and the references therein. However the combination with the stochastic integral makes it difficult to clarify the asymptotic behavior of the left-hand-side. To handle this difficulty, we invoke the concept of the extended Poisson equation (EPE) introduced in [64]:

Definition 2.3.2. *[64, 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 $f \in Dom(\tilde{\mathcal{A}})$ and $\tilde{\mathcal{A}}f \overset{EPE}{=} g$.

Remark 2.3.3. *In the previous definition, the terminology "Feller" means that the corresponding transition semigroup T_t is a mapping $C_b(\mathbb{R})$ into $C(\mathbb{R})$. When it comes to X, its homogeneous, Feller and (strong) Markov properties are guaranteed by the argument in [3, Theorem 6.4.6] and [36, 3.1.1 (ii)].*

Remark 2.3.4. *When we consider the misspecified ergodic diffusion models, we also encounter the annoying integral term like (2.3.3). In that case, [60] utilized the theory of the second order differential equations endowed with their infinitesimal generator (cf. [49]) and Itô's formula to derive the asymptotic normality of the GQMLE. However, in our case, the same method cannot be applied since the infinitesimal generator of X contains the integro-operator with respect to the Lévy measure of Z and it is difficult to verify the existence and regularity of the corresponding equation.*

Hereinafter $y^{(i)}$ is referred to as the i-th component of any vector y. We consider the following EPEs:

$$\tilde{\mathcal{A}}f_1^{(j_1)}(x) \overset{EPE}{=} -\frac{\partial_{\gamma^{(j_1)}} c(x,\gamma^\star)}{c^3(x,\gamma^\star)}(c^2(x,\gamma^\star) - C^2(x)), \tag{2.3.4}$$

$$\tilde{\mathcal{A}} f_2^{(j_2)}(x) \overset{EPE}{=} -\frac{\partial_{\alpha^{(j_2)}} a(x,\alpha^\star)}{c^2(x,\gamma^\star)}(A(x) - a(x,\alpha^\star)), \tag{2.3.5}$$

for the extended generator $\tilde{\mathcal{A}}$ of X, $j_1 \in \{1,\dots,p_\gamma\}$ and $j_2 \in \{1,\dots,p_\alpha\}$. The right-hand-side of each EPE corresponds to g in (2.3.2), and it is trivial that they identically 0 when the coefficients are correctly specified.

From now on, E^x is referred to as the expectation operator with the initial condition $X_0 = x$, that is,

$$E^x[g(X_t)] = \int_{\mathbb{R}} g(y) P_t(x, dy),$$

for any measurable function g. The next proposition ensures the existence of the solutions of (2.3.4) and (2.3.5) and verifies their weighted Hölder continuity:

Proposition 2.3.5. *Under Assumption 2.2.1-2.2.3, there exist unique solutions of (2.3.4) and (2.3.5), and the solution vectors* $f_1 := \left(f_1^{(j_1)}\right)_{j_1\in\{1,\dots,p_\gamma\}}$ *and* $f_2 := \left(f_2^{(j_2)}\right)_{j_2\in\{1,\dots,p_\alpha\}}$ *satisfy*

$$\sup_{x,y\in\mathbb{R}, x\neq y} \frac{|f_i(x) - f_i(y)|}{|x-y|^{1/p_i}(1 + |x|^{q_i K_i} + |y|^{q_i K_i})} < \infty, \quad \text{for } i \in \{1,2\},$$

where any $p_i \in (1,\infty)$, $q_i = p_i/(p_i - 1)$, *and some positive constants* K_1 *and* K_2. *Furthermore,*

$$f_1(X_t) + \int_0^t \frac{\partial_\gamma c(X_s,\gamma^\star)}{c^3(X_s,\gamma^\star)}(c^2(X_s,\gamma^\star) - C^2(X_s))ds,$$

and

$$f_2(X_t) + \int_0^t \frac{\partial_\alpha a(X_s,\alpha^\star)}{c^2(X_s,\gamma^\star)}(A(X_s) - a(X_s,\alpha^\star))ds$$

are L_2*-martingale with respect to* $(\mathcal{F}_t, P_x)$ *for every* $x \in \mathbb{R}$, *and their explicit forms are given as follows:*

$$f_1(x) = \int_0^\infty E^x\left[\frac{\partial_\gamma c(X_t,\gamma^\star)}{c^3(X_t,\gamma^\star)}(c^2(X_t,\gamma^\star) - C^2(X_t))\right]dt,$$
$$f_2(x) = \int_0^\infty E^x\left[\frac{\partial_\alpha a(X_t,\alpha^\star)}{c^2(X_t,\gamma^\star)}(A(X_t) - a(X_t,\alpha^\star))\right]dt.$$

Remark 2.3.6. *Thanks to the result of the previous theorem and assumptions on the coefficients,*

$$f_1(X_t) + \int_0^t \frac{\partial_\gamma c(X_s,\gamma^\star)}{c^3(X_s,\gamma^\star)}(c^2(X_s,\gamma^\star) - C^2(X_s))ds,$$

and

$$f_2(X_t) + \int_0^t \frac{\partial_\alpha a(X_s, \alpha^\star)}{c^2(X_s, \gamma^\star)}(A(X_s) - a(X_s, \alpha^\star))ds$$

have finite second-order moments. Thus, slightly refining the argument in [51, the proof of Proposition VII 1.6] with the monotone convergence theorem, the L_2-martingale property of them with respect to $(\mathcal{F}_t, P_x)$ can be replaced by the L_2-martingale property with respect to $(\mathcal{F}_t, P)$ in the previous proposition.

Building on the previous proposition, now we can obtain the asymptotic normality of $\sqrt{T_n}(\hat{\theta}_n - \theta^\star)$:

Theorem 2.3.7. *Under Assumptions 2.2.1-2.2.4, there exists a nonnegative definite matrix $\Sigma \in \mathbb{R}^p \otimes \mathbb{R}^p$ such that*

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

and the form of $\Sigma := \begin{pmatrix} \Sigma_\gamma & \Sigma_{\alpha\gamma} \\ \Sigma_{\alpha\gamma}^\top & \Sigma_\alpha \end{pmatrix}$ is given by:

$$\Sigma_\gamma = 4\int_{\mathbb{R}}\int_{\mathbb{R}} \left(\frac{\partial_\gamma c(x, \gamma^\star)}{c^3(x, \gamma^\star)}C^2(x)z^2 + f_1(x + C(x)z) - f_1(x)\right)^{\otimes 2} \pi_0(dx)\nu_0(dz),$$

$$\Sigma_{\alpha\gamma} = -4\int_{\mathbb{R}}\int_{\mathbb{R}} \left(\frac{\partial_\gamma c(x, \gamma^\star)}{c^3(x, \gamma^\star)}C^2(x)z^2 + f_1(x + C(x)z) - f_1(x)\right) \left(\frac{\partial_\alpha a(x, \alpha^\star)}{c^2(x, \gamma^\star)}C(x)z + f_2(x + C(x)z) - f_2(x)\right)^\top \pi_0(dx)\nu_0(dz),$$

$$\Sigma_\alpha = 4\int_{\mathbb{R}}\int_{\mathbb{R}} \left(\frac{\partial_\alpha a(x, \alpha^\star)}{c^2(x, \gamma^\star)}C(x)z + f_2(x + C(x)z) - f_2(x)\right)^{\otimes 2} \pi_0(dx)\nu_0(dz).$$

Remark 2.3.8. *If either of the coefficients is correctly specified, the right-hand side of the associated EPE (2.3.4) or (2.3.5) is identically 0. Let γ_0 and α_0 be the elements of Θ_γ and Θ_α whose definitions are introduced in Rem 2.2.8. Then we have*

$$\Sigma_\gamma = 4\int_{\mathbb{R}} \left(\frac{\partial_\gamma c(x, \gamma_0)}{c(x, \gamma_0)}\right)^{\otimes 2} \pi_0(dx)\int_{\mathbb{R}} z^4 \nu_0(dz),$$

$$\Sigma_{\alpha\gamma} = -4\int_{\mathbb{R}}\int_{\mathbb{R}} \left(\frac{\partial_\gamma c(x, \gamma_0)}{c(x, \gamma_0)}z^2\right)\left(\frac{\partial_\alpha a(x, \alpha^\star)}{c(x, \gamma_0)}z + f_2(x + c(x, \gamma_0)z) - f_2(x)\right)^\top \pi_0(dx)\nu_0(dz),$$

$$\Sigma_\alpha = 4\int_{\mathbb{R}}\int_{\mathbb{R}} \left(\frac{\partial_\alpha a(x, \alpha^\star)}{c(x, \gamma_0)}z + f_2(x + c(x, \gamma_0)z) - f_2(x)\right)^{\otimes 2} \pi_0(dx)\nu_0(dz),$$

in the case that the scale coefficient is correctly specified and

$$\Sigma_\gamma = 4\int_{\mathbb{R}}\int_{\mathbb{R}}\left(\frac{\partial_\gamma c(x,\gamma^\star)}{c^3(x,\gamma^\star)}C^2(x)z^2 + f_1(x+C(x)z) - f_1(x)\right)^{\otimes 2}\pi_0(dx)\nu_0(dz),$$
$$\Sigma_{\alpha\gamma} = -4\int_{\mathbb{R}}\int_{\mathbb{R}}\left(\frac{\partial_\gamma c(x,\gamma^\star)}{c^3(x,\gamma^\star)}C^2(x)z^2 + f_1(x+C(x)z) - f_1(x)\right)$$
$$\left(\frac{\partial_\alpha a(x,\alpha_0)}{c^2(x,\gamma^\star)}C(x)z\right)^\top \pi_0(dx)\nu_0(dz),$$
$$\Sigma_\alpha = 4\int_{\mathbb{R}}\left(\frac{\partial_\alpha a(x,\alpha_0)}{c^2(x,\gamma^\star)}C(x)\right)^{\otimes 2}\pi_0(dx),$$

in the case that the drift coefficient is correctly specified.

Remark 2.3.9. *Let Y be a random variable which obeys $N(0,\Gamma^{-1}\Sigma(\Gamma^{-1})^\top)$. As a consequence of Theorem 2.3.1 and Theorem 2.3.7, we have*

$$E\left[f\left(\sqrt{T_n}(\hat\theta_n - \theta^\star)\right)\right] \to E[f(Y)], \tag{2.3.6}$$

for any polynomial growth function f. It can be shown in the following way: For any $q>1$, it follows from [15, Lemma 2.2.8] and Theorem 2.3.1 that

$$\begin{aligned} E\left[\left|\sqrt{T_n}(\hat\theta_n - \theta^\star)\right|^q\right] &= \int_0^\infty q x^{q-1} P\left(\left|\sqrt{T_n}(\hat\theta_n - \theta^\star)\right| > x\right)dx \\ &\lesssim \int_0^1 x^{q-1}dx + \int_1^\infty x^{-q}dx < \infty. \end{aligned}$$

Hence $\left|\sqrt{T_n}(\hat\theta_n - \theta^\star)\right|^q$ is asymptotically uniformly integrable from Markov's inequality, and [63, Theorem 2.20] implies (2.3.6).

Remark 2.3.10. *In this remark, we suppose that the data-generating model defined on the probability space $(\Omega,\mathcal{F},(\mathcal{F}_t)_{t\in\mathbb{R}_+},P)$ is supposed to be*

$$dY_t = A(Y_t)dt + B(Y_t)dW_t + C(Y_{t-})dZ_t, \tag{2.3.7}$$

where W is a standard Wiener process independent of (Y_0,Z),

$$\mathcal{F}_t := \sigma(Y_0)\vee\sigma((W_s,Z_s); s\le t),$$

and $B:\mathbb{R}\to\mathbb{R}$ is a measurable function. We look at the following parametric model:

$$dY_t = a(Y_t,\alpha)dt + b(Y_t,\gamma)dW_t + c(Y_{t-},\gamma)dZ_t,$$

where $b : \mathbb{R} \times \Theta_\gamma \to \mathbb{R}$ is a measurable function. Here other ingredients are similarly defined as above and we use the same notations for its transition probability, invariant measure, and so on. When the true coefficients (A, B, C) are correctly specified, the GQMLE still has asymptotic normality and the sufficient conditions for it are easy to check (cf. [39]). However, we note that it is difficult to give such conditions when they are misspecified. This is because our methodology using the martingale representation theorem becomes insufficient due to the presence of Wiener component in the deviation of the asymptotic variance (see, the proof of Theorem 2.3.7). To formally derive a similar result to Theorem 2.3.7, we may additionally have to impose the following condition:

***Condition A**: There exists a unique C^2-solution f on $\mathbb{R}$ of*

$$\begin{aligned}\mathcal{A}f(x) &= A(x)\partial_x f(x) + \frac{1}{2}B(x)\partial_x^2 f(x) + \int_{\mathbb{R}}(f(x+C(x)z) - f(x) - \partial_x f(x)C(x)z)\nu_0(dz)\\ &= g(A(x), B(x), C(x)), \end{aligned} \tag{2.3.8}$$

where $g(A(x), B(x), C(x))$ is a specific function satisfying

$$\int_{\mathbb{R}} g(A(x), B(x), C(x))\pi_0(dx) = 0.$$

Furthermore, the first and second derivatives of f are of at most polynomial growth.

*Under **Condition A**, the limit distribution of the GQMLE can be derived by combining the proof of [61] and Theorem 2.3.7. It is known that the theory of viscosity solutions for integro-differential equations ensures the existence of f in limited situation, for instance, see [5], [6], [21] and [22]. However, it is not so for the regularity of f. As another attempt to confirm **Condition A**, the associated EPE $\tilde{\mathcal{A}}\tilde{f} \overset{EPE}{=} g$ may possibly be helpful. This is because the existence and uniqueness of the solution $\tilde{f}$ of the EPE can be verified in an analogous way to Theorem 2.3.5, and if $\tilde{f}$ admits C^2-property and growth conditions in **Condition A**, then $\tilde{f}$ satisfies (2.3.8). The latter argument can formally be shown as follows:*

It is enough to check $\mathcal{A}\tilde{f} = g$. Since $\tilde{f}(Y_t) - \int_0^t g(A(Y_s), B(Y_s), C(Y_s))ds$ is a martingale with respect to $(\mathcal{F}_t, P_x)$ for all $x \in \mathbb{R}$, we have

$$E^x\left[\tilde{f}(Y_t) - \int_0^t g(A(Y_s), B(Y_s), C(Y_s))ds\right] = \tilde{f}(x).$$

Hence it follows from Itô's formula that as $t \to 0$,

$$\left|\frac{E^x[\tilde{f}(Y_t)] - \tilde{f}(x)}{t} - g(A(x), B(x), C(x))\right|$$

$$
\begin{aligned}
&= \left| \frac{1}{t} \int_0^t \left(E^x[g(A(Y_s), B(Y_s), C(Y_s))] - g(A(x), B(x), C(x)) \right) ds \right| \\
&= \left| \frac{1}{t} \int_0^t \int_0^s E^x[\mathcal{A} g(A(Y_u), B(Y_u), C(Y_u))] du ds \right| \\
&\lesssim t \to 0.
\end{aligned}
$$

In this sketch, we implicitly assume suitable regularity and moment conditions on each ingredient, but they are reduced to be conditions on the true coefficients (A, B, C)*. Thus, verifying the behavior of*

$$\tilde{f}(x) = \int_0^\infty E^x[g(A(Y_t), B(Y_t), C(Y_t))]dt = \int_0^\infty \int_{\mathbb{R}} g(A(y), B(y), C(y)) P_t(x, dy)$$

leads to ***Condition A****. Just for Lévy driven Ornstein-Uhlenbeck models, we can observe the property of* $\tilde{f}(x) = \int_0^\infty E^x[g(A(Y_t), B(Y_t), C(Y_t))]dt$ *based on the explicit form of the solution (cf. Example 2.3.11). Although, for general Lévy driven SDEs, the gradient estimates of their transition probability making use of Malliavin calculus have been investigated lately (cf. [66], [67], and the references therein), the property of* $\tilde{f}(x) = \int_0^\infty E^x[g(A(Y_t), B(Y_t), C(Y_t))]dt$ *is still difficult to be checked as far as the author knows. Since these are out of range of this chapter, we will not treat them later.*

Example 2.3.11. *Here we consider the following Ornstein-Uhlenbeck model:*

$$dX_t = -\alpha X_t dt + dZ_t,$$

for a Lévy process Z *not necessarily being pure-jump type and a positive constant* α*. Applying Itô's formula to* $\exp(\alpha t) X_t$*, we have*

$$X_t = X_0 \exp(-\alpha t) + \int_0^t \exp(\alpha(s-t)) dZ_s$$

and

$$E^x[f(X_t)] = \int_{\mathbb{R}} f(x \exp(-\alpha t) + y) p_t(dy)$$

for a suitable function f*. Here* p_t *is the probability distribution function of* $\int_0^t \exp(\alpha(s-t))dZ_s$ *whose characteristic function* $\hat{p}_t(\cdot)$ *is given by:*

$$\hat{p}_t(u) = \exp\left\{ \int_0^t \psi(\exp(\alpha(s-t))u) ds \right\}, \tag{2.3.9}$$

for $\psi(u) := \log E[\exp(iuZ_1)]$ *(cf. [53, Theorem 3.1]). In this case,* X *fulfills Assumption 2.2.3 provided that Assumption 2.2.1-(1) holds, and that the Lévy measure* ν_0 *of* Z *has a continuously differentiable positive density* g *on an open neighborhood around the* *origin (for more details, see [39, Section 5]). Then, the characteristic function $\hat{p}(\cdot)$ of the invariant measure π_0 is given by*

$$\hat{p}(u) = \exp\left\{\int_0^\infty \psi(\exp(-\alpha s)u)ds\right\}.$$

Under such condition, if f is differentiable and itself and its derivative are of at most polynomial growth, we have

$$\begin{aligned}
&\left|\partial_x\left(\int_0^\infty E^x[f(X_t)]dt\right)\right| \\
&= \left|\int_0^\infty \left(\int_{\mathbb{R}} \partial_x f(x\exp(-\alpha t) + y)p_t(dy)\right)\exp(-\alpha t)dt\right| \\
&\lesssim \int_0^\infty \left\{1 + |x|^K + (1+|x|^{2K})\exp(-at)\right\}\exp(-\alpha t)dt \\
&\lesssim 1 + |x|^{2K},
\end{aligned}$$

for a positive constant K. We can derive similar estimates with respect to its higher-order derivatives in the same way.

Let J be a Lévy process such that its moments of any-order exists and its triplet is $(0, b, \nu^J)$ (cf. [3]). Here b is allowed to be 0. Mimicking the previous example, we write p_t^J as the probability distribution function of $\int_0^t \exp(\alpha(s-t))dJ_s$ for a positive constant $\alpha > 0$ and $\psi^J(u)$ stands for $\log E[\exp(iuJ_1)]$ below. Combining the argument in Remark 2.3.10 and Example 2.3.11, we obtain the following corollary:

Corollary 2.3.12. *For a natural number $k \geq 2$, let f be a polynomial growth C^k-function whose derivatives are of at most polynomial growth. Suppose that the integral of f with respect to the Borel probability measure π_0 whose characteristic function is $\exp\left\{\int_0^\infty \psi^J(\exp(-\alpha s)u)ds\right\}$ is 0, and that ν^J has a continuously differentiable positive density on an open neighborhood around the origin. Then, the function*

$$\begin{aligned}
g(x) :&= \int_0^\infty E^x\left[f\left(x\exp(-\alpha t) + \int_0^t \exp(\alpha(s-t))dJ_s\right)\right]dt \\
&= \int_0^\infty \int_{\mathbb{R}} f(x\exp(-\alpha t) + y)p_t^J(dy)dt
\end{aligned}$$

on $\mathbb{R}$ is the unique solution of the following (first or second order) integro-differential equation

$$-\alpha x\partial_x g(x) + \frac{1}{2}b\partial_x^2 g(x) + \int_{\mathbb{R}} (g(x+z) - g(x) - \partial g(x)z)\,\nu^J(dz) = f(x), \tag{2.3.10}$$

and moreover, g is also a polynomial growth C^k-function.

Remark 2.3.13. *If the Lévy measure ν^J is symmetric (i.e. the imaginary part of ψ^J is 0), the equation (2.3.10) is solvable for many odd functions f as a matter of course. More specifically, for $k \in \mathbb{N}$ and $f(x) = x^{2k+1}$, the solution g is*

$$\begin{aligned} g(x) &= \int_0^\infty \int_{\mathbb{R}} (x\exp(-\alpha t) + y)^{2k+1} p_t^J(dy)dt \\ &= \int_0^\infty \int_{\mathbb{R}} \sum_{i=0}^{2k+1} \frac{(2k+1)!}{i!(2k+1-i)!} (x\exp(-\alpha t))^i y^{2k+1-i} p_t^J(dy)dt. \end{aligned}$$

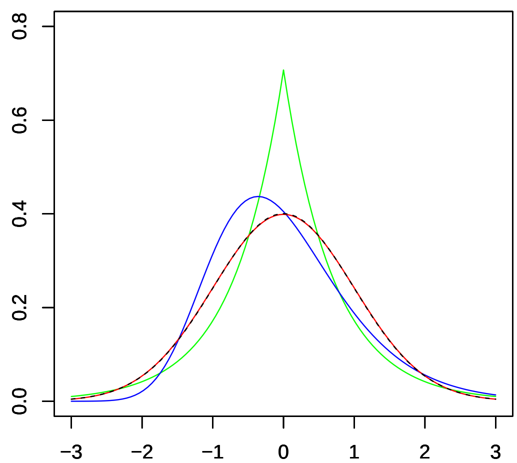
By observing the derivatives of the characteristic function, $\int_{\mathbb{R}} y^{2k+1-i} p_t^J(dy)$ can be expressed by the moments of J, hence the explicit expression of g is available.

Remark 2.3.14. *Beside the estimation of θ, what is of special interest is the inference for ν_0 which may often be an infinite dimensional parameter. Even for (A, C) being constant and specified (i.e. X is a Lévy process with drift), it may be interest in its own right and enormous papers have addressed this problem so far. We refer to [41] for comprehensive accounts under Z being assumed to have a certain parametric structure. As for the situation where just a few information on Z is available, one of plausible attempts is the method of moments proposed in [17], [18], and [47], for example. Especially [47] established a Donsker-type functional limit theorem for empirical processes arising from high-frequently observed Lévy processes. When the coefficients A and C are nonlinear functions but specified, the residual based method of moments for ν_0 by [44] is effective: using the GQMLE $\hat{\theta}_n := (\hat{\gamma}_n, \hat{\alpha}_n)$, we have*

$$\begin{aligned} &\frac{1}{T_n} \sum_{j=1}^n \varphi\left(\frac{\Delta_j X - h_n a_{j-1}(\hat{\alpha}_n)}{c_{j-1}(\hat{\gamma}_n)}\right) \xrightarrow{P} \int_{\mathbb{R}} \varphi(z)\nu_0(dz), \\ &\hat{D}_n \sqrt{T_n} \begin{pmatrix} \hat{\theta}_n - \theta_0 \\ \frac{1}{T_n} \sum_{j=1}^n \varphi\left(\frac{\Delta_j X - h_n a_{j-1}(\hat{\alpha}_n)}{c_{j-1}(\hat{\gamma}_n)}\right) - \int \varphi(z)\nu_0(dz) \end{pmatrix} \xrightarrow{\mathcal{L}} N(0, I_{p+q}), \end{aligned}$$

for an appropriate $\mathbb{R}^q$-valued function φ and a $(p+q) \times (p+q)$ matrix $\hat{D}_n$ which can be constructed only by the observations. For instance, we can choose $\varphi(z) = z^r$ and $\varphi(z) = \exp(iuz) - 1 - iuz$ (to estimate the r-th cumulant of Z and the cumulant function of Z, respectively) as φ; see [44, Assumption 2.7] for the precise conditions on φ. As for misspecified case, if the misspecification is confined within the drift coefficient, then this scheme is still valid thanks to the faster diminishment of the mean activity in small time (cf. Remark 2.2.7).

Figure 2.1: The plot of the density functions of (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).



2.4 Numerical experiments

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:

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

The functional form of the coefficients is the same in [60, Example 3.1]. We conduct numerical experiments in three situations:

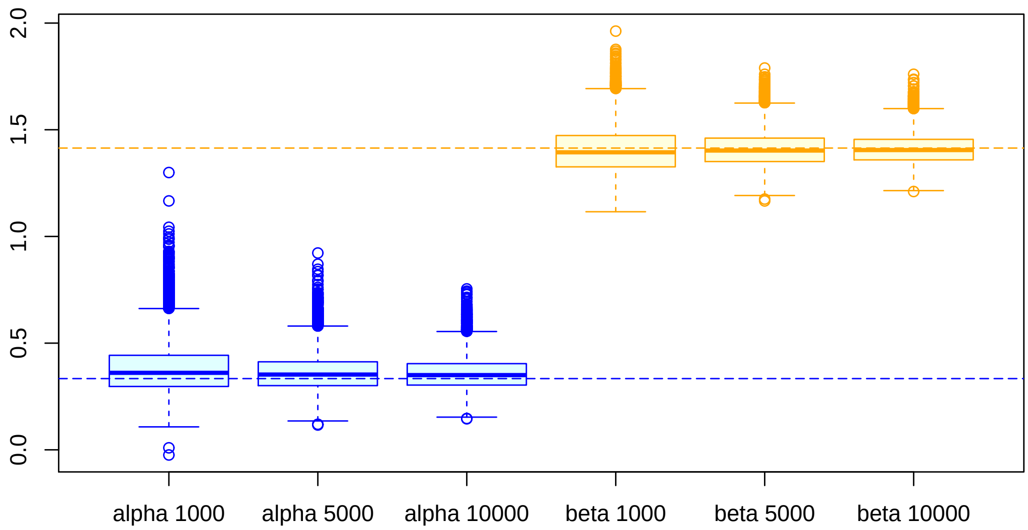
(i) $\mathcal{L}(Z_t) = NIG(10, 0, 10t, 0)$;

(ii) $\mathcal{L}(Z_t) = bGamma(t, \sqrt{2}, t, \sqrt{2})$;

(iii) $\mathcal{L}(Z_t) = NIG(25/3, 20/3, 9/5t, -12/5t)$.

NIG (normal inverse Gaussian) random variable is defined by the normal mean-variance mixture of inverse Gaussian random variable, and *bGamma* (bilateral Gamma) random

Figure 2.2: The boxplot of case (i); the target optimal values are described by dotted lines.



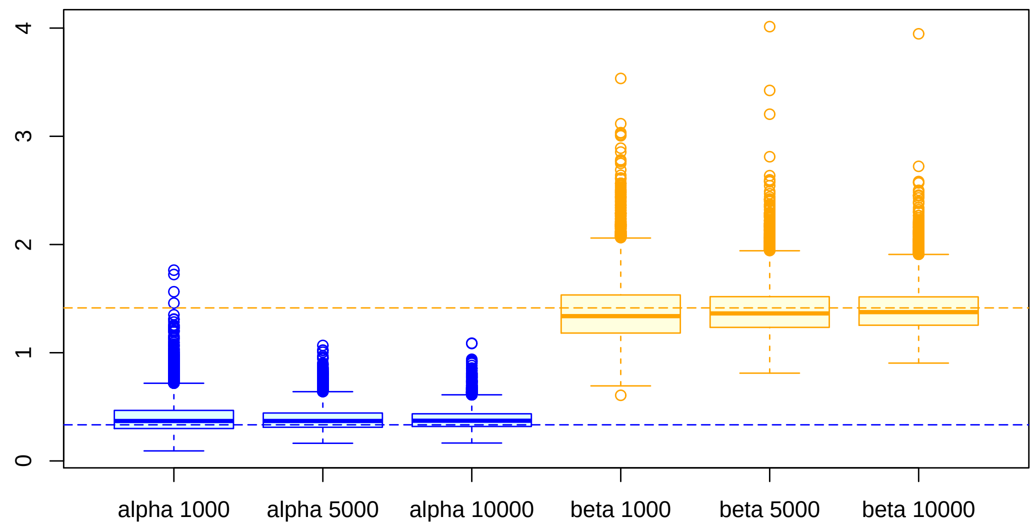
variable is defined by the difference of two independent Gamma random variables. For their technical accounts, we refer to [7] and [30]. To visually observe their non-Gaussianity, each density function at $t = 1$ is plotted with the density of $N(0,1)$ in Figure 2.1 altogether. By taking the limit of (2.3.9), the characteristic function $\hat{p}(\cdot)$ of the invariant measure π_0 is given by

$$\hat{p}(u) = \exp\left\{\int_0^\infty \psi\left(\exp\left(-\frac{s}{2}\right)u\right)ds\right\}, \tag{2.4.1}$$

where $\psi(u) := \log E[\exp(iuZ_1)]$. Differentiating $\hat{p}$, we have $\tilde{\kappa}_j = 2\kappa_j/j$ for the j-th cumulant $\tilde{\kappa}_j$ (resp. κ_j) of $Y \sim \pi_0$ (resp. Z_1). Hence we obtain

$$\begin{aligned}
\mathbb{G}_1(\gamma) &= -2\log\gamma - \frac{2}{\gamma^2} + \int_{\mathbb{R}} \log(1+x^2)\pi_0(dx), \\
\mathbb{G}_2(\alpha) &= -\frac{1}{\gamma^\star}\left\{\frac{1}{4}\int_{\mathbb{R}} x^3\pi_0(dx) + \alpha\left(1 - \int_{\mathbb{R}} x^3\pi_0(dx) + \int_{\mathbb{R}} x^4\pi_0(dx)\right)\right. \\
&\qquad \left. + \alpha^2\left(3 - 2\int_{\mathbb{R}} x^3\pi_0(dx) + \int_{\mathbb{R}} x^4\pi_0(dx)\right)\right\}.
\end{aligned}$$

Figure 2.3: The boxplot of case (ii); the target optimal values are described by dotted lines.



By solving the estimating equations, the target optimal values are given by

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

In the calculation, we used $\int_{\mathbb{R}} x\pi_0(dx) = 0$ and $\int_{\mathbb{R}} x^2 \pi_0(dx) = 1$. Thus, in each case, the optimal parameter $\theta^\star := (\alpha^\star, \gamma^\star)$ is given as follows:

(i) $\theta^\star = \left(803/2406, \sqrt{2}\right) \approx (0.3337, 1.4142)$;

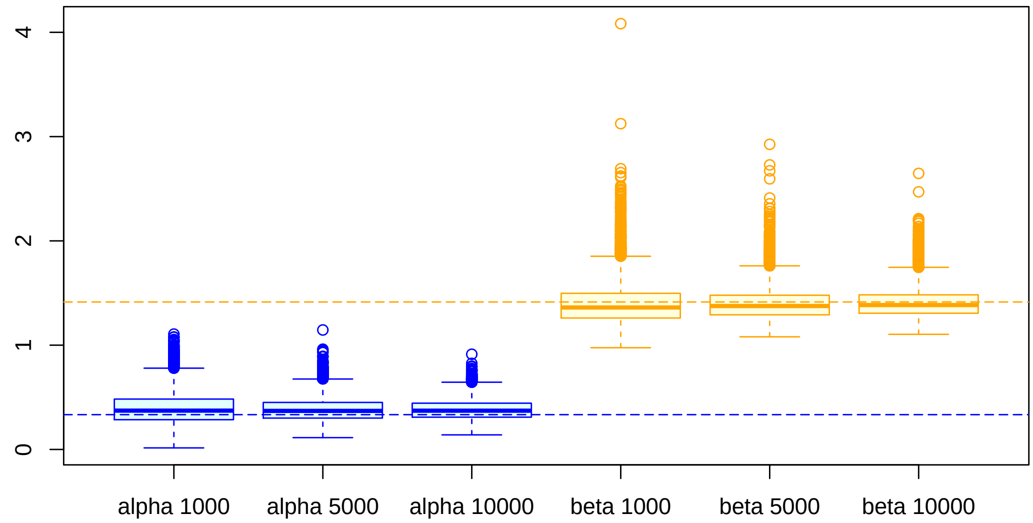
(ii) $\theta^\star = \left(11/30, \sqrt{2}\right) \approx (0.3667, 1.4142)$;

(iii) $\theta^\star = \left(609/1658, \sqrt{2}\right) \approx (0.3673, 1.4142)$.

Here, we write approximated values obtained by rounding off $\theta^\star$ to four decimal places. Solving the corresponding estimating equations, our staged GQMLE are calculated as:

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

Figure 2.4: The boxplot of case (iii); the target optimal values are described by dotted lines.



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

We generated 10000 paths of each SDE based on Euler-Maruyama scheme and constructed the estimators along with the above expressions, independently. In generating the small time increments of the driving noises, we used the function `rng` equipped to YUIMA package in R [12]. Together with the diffusion case (the optimal parameter is$\theta^\star = \left(1/3, \sqrt{2}\right) \approx (0.3333, 1.4142)$), the mean and standard deviation of each estimator is shown in Table 2.2 where n and $h_n = 5n^{-2/3}$ denote the sample size and observation interval, respectively. We also present their boxplots to enhance the visibility. We can observe the followings from the table and boxplots:

- Overall, the estimation accuracy of $\hat{\theta}_n$ improves as T_n and n increase and h_n decrease, and this tendency reflects our main result.
- The result of case (i) is almost the same as the diffusion case. This is thought to be based on the well-known fact that $NIG(\delta, 0, \delta t, 0)$ tends to $N(0, t)$ in total

Table 2.2: 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)

variation norm as $\delta \to \infty$ for any $t > 0$. Indeed, Figure 2.1 shows that the density functions of $NIG(10, 0, 10, 0)$ and $N(0, 1)$ are virtually the same.

- Concerning case (ii), the standard deviation of $\hat{\gamma}_n$ is relatively worse than the other cases. This is natural because the asymptotic variance of $\hat{\gamma}_n$ includes the forth-order-moment of Z, and $bGamma(1, \sqrt{2}, 1, \sqrt{2})$ has the highest kurtosis value as can be seen from Figure 2.1.

- In case (iii), the performance of $\hat{\alpha}_n$ is the worst in this experiment. This may cause from the fact that only $NIG(25/3, 20/3, 9/5, -12/5)$ is not symmetric.

2.5 Appendix

Proof of Theorem 2.3.1 In light of our situation, it is sufficient to check the conditions [A1”], [A4’] and [A6] in [69] for $\mathbb{G}_{1,n}$ and $\mathbb{G}_{2,n}$, respectively. For the sake of convenience, we simply write

$$\mathbb{Y}_{1,n}(\gamma) = \mathbb{G}_{1,n}(\gamma) - \mathbb{G}_{1,n}(\gamma^\star),$$
$$\mathbb{Y}_{2,n}(\alpha) = \mathbb{G}_{2,n}(\alpha) - \mathbb{G}_{2,n}(\alpha^\star).$$

Without loss of generality, we can assume $p_\gamma = p_\alpha = 1$. First we treat $\mathbb{G}_{1,n}(\cdot)$. The conditions hold if we show

$$\sup_{n\in\mathbb{N}} E\left[|\sqrt{T_n}\partial_\gamma \mathbb{G}_{1,n}(\gamma^\star)|^K\right] < \infty, \tag{2.5.1}$$

$$\sup_{n\in\mathbb{N}} E\left[|\sqrt{T_n}(\partial_\gamma^2 \mathbb{G}_{1,n}(\gamma^\star) - \Gamma_\gamma)|^K\right] < \infty, \tag{2.5.2}$$

$$\sup_{n\in\mathbb{N}} E\left[\sup_{\gamma\in\Theta_\gamma} |\partial_\gamma^3 \mathbb{G}_{1,n}(\gamma)|^K\right] < \infty, \tag{2.5.3}$$

$$\sup_{n\in\mathbb{N}} E\left[\sup_{\gamma\in\Theta_\gamma} |\sqrt{T_n}\left(\mathbb{Y}_{1,n}(\gamma) - \mathbb{Y}_1(\gamma)\right)|^K\right] < \infty, \tag{2.5.4}$$

for any $K > 0$. The first two derivatives of $\mathbb{G}_{1,n}$ are given by

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

$$\partial_\gamma^2 \mathbb{G}_{1,n}(\gamma) = -\frac{2}{T_n}\sum_{j=1}^n \left\{\frac{\partial_\gamma^2 c_{j-1}(\gamma)c_{j-1}(\gamma) - (\partial_\gamma c_{j-1})^2}{c_{j-1}^2(\gamma)}h_n \right.$$
$$\left. - \frac{\partial_\gamma^2 c_{j-1}(\gamma)c_{j-1}(\gamma) - 3(\partial_\gamma c_{j-1}(\gamma))^2}{c_{j-1}^4(\gamma)}(\Delta_j X)^2\right\}.$$

We further decompose $\partial_\gamma \mathbb{G}_{1,n}(\gamma^\star)$ as

$$\partial_\gamma \mathbb{G}_{1,n}(\gamma^\star) = -\frac{2}{n}\sum_{j=1}^n \frac{\partial_\gamma c_{j-1}}{c_{j-1}^3}\left(c_{j-1}^2 - C_{j-1}^2\right) + \frac{2}{T_n}\sum_{j=1}^n \frac{\partial_\gamma c_{j-1}}{c_{j-1}^3}\left\{(\Delta_j X)^2 - h_n C_{j-1}^2\right\}.$$

Since the optimal parameter $\theta^\star$ is in Θ, the interchange of the derivative and the integral implies that the function $\partial_\gamma c(x,\gamma^\star)(c^2(x,\gamma^\star) - C^2(x))/c^3(x,\gamma^\star)$ is centered in the sense that its integral with respect to π_0 is 0. Thus [40, Lemma 4.3] and [44, Lemma 5.3] lead to (2.5.1) and (2.5.4). We also have

$$\begin{aligned}
&\partial_\gamma^2 \mathbb{G}_{1,n}(\gamma)\\
&= -\frac{2}{n}\sum_{j=1}^n \left\{\frac{\partial_\gamma^2 c_{j-1}(\gamma)c_{j-1}(\gamma) - (\partial_\gamma c_{j-1})^2}{c_{j-1}^2(\gamma)} - \frac{\partial_\gamma^2 c_{j-1}(\gamma)c_{j-1}(\gamma) - 3(\partial_\gamma c_{j-1}(\gamma))^2}{c_{j-1}^4(\gamma)}C_{j-1}^2\right\}\\
&+ \frac{2}{T_n}\sum_{j=1}^n \frac{\partial_\gamma^2 c_{j-1}(\gamma)c_{j-1}(\gamma) - 3(\partial_\gamma c_{j-1}(\gamma))^2}{c_{j-1}^4(\gamma)}\left\{(\Delta_j X)^2 - h_n C_{j-1}^2\right\}.
\end{aligned}$$

Again applying [40, Lemma 4.3] and [44, Lemma 5.3], we obtain (2.5.2). Via simple calculation, the third and fourth-order derivatives of $\mathbb{G}_{1,n}$ can be represented as

$$\partial_\gamma^i \mathbb{G}_{1,n}(\gamma) = \frac{1}{n}\sum_{j=1}^n g_{j-1}^i(\gamma) + \frac{1}{T_n}\sum_{j=1}^n \tilde{g}_{j-1}^i(\gamma)\left\{(\Delta_j X)^2 - h_n C_{j-1}^2\right\}, \quad \text{for } i \in \{3,4\},$$

with the matrix-valued functions $g^i(\cdot,\cdot)$ and $\tilde{g}^i(\cdot,\cdot)$ defined on $\mathbb{R}\times\Theta_\gamma$, and these are of at polynomial growth with respect to $x \in \mathbb{R}$ uniformly in γ. Hence (2.5.3) follows from

Sobolev's inequality (cf. [1, Theorem 1.4.2]). Thus [69, Theorem 3-(c)] leads to the tail probability estimates of $\hat{\gamma}_n$. We write

$$\upsilon_j = 2\Delta_j X(a_{j-1}(\alpha) - a_{j-1}(\alpha^\star)) + h_n(a_{j-1}^2(\alpha^\star) - a_{j-1}^2(\alpha)).$$

From Taylor's expansion, we get

$$\begin{aligned}
&\mathbb{Y}_{2,n}(\alpha) \\
&= \frac{1}{T_n}\sum_{j=1}^n \frac{\upsilon_j}{c_{j-1}^2(\gamma^\star)} + \left(\int_0^1 \frac{1}{(T_n)^{3/2}}\sum_{j=1}^n \upsilon_j \partial_\gamma c_{j-1}^{-2}(\gamma^\star + u(\hat{\gamma}_n - \gamma^\star))du\right)\left(\sqrt{T_n}(\hat{\gamma}_n - \gamma^\star)\right) \\
&:= \tilde{\mathbb{Y}}_{2,n}(\alpha) + \bar{\mathbb{Y}}_{2,n}(\alpha)\left(\sqrt{T_n}(\hat{\gamma}_n - \gamma^\star)\right).
\end{aligned}$$

Sobolev's inequality leads to

$$\begin{aligned}
&E\left[\left|\sqrt{T_n}\bar{\mathbb{Y}}_{2,n}(\alpha)\right|^K\right] \\
&\leq E\left[\sup_{\gamma\in\Theta_\gamma}\left|\frac{1}{T_n}\sum_{j=1}^n \upsilon_j \partial_\gamma c_{j-1}^{-2}(\gamma)\right|^K\right] \\
&\lesssim \sup_{\gamma\in\Theta_\gamma}\left\{E\left[\left|\frac{1}{T_n}\sum_{j=1}^n \upsilon_j \partial_\gamma c_{j-1}^{-2}(\gamma)\right|^K\right] + E\left[\left|\frac{1}{T_n}\sum_{j=1}^n \upsilon_j \partial_\gamma^2 c_{j-1}^{-2}(\gamma)\right|^K\right]\right\},
\end{aligned}$$

for $K > 1$. The last two terms of the right-hand-side are finite from [44, Lemma 5.3], and the moment bounds of the three functions $\sqrt{T_n}\partial_\alpha^i \bar{\mathbb{Y}}_{2,n}(\alpha)$ ($i \in \{1,2,3\}$) can analogously be obtained. Thus combined with the tail probability estimates of $\hat{\gamma}_n$ and Schwartz's inequality, it suffices to show the conditions for

$$\begin{aligned}
\tilde{\mathbb{Y}}_{2,n}(\alpha) &:= \frac{1}{T_n}\sum_{j=1}^n \frac{\upsilon_j}{c_{j-1}^2(\gamma^\star)}, \\
\tilde{\mathbb{G}}_{2,n}(\alpha) &:= -\frac{1}{T_n}\sum_{j=1}^n \frac{(\Delta_j X - h_n a_{j-1}(\alpha))^2}{h_n c_{j-1}^2(\gamma^\star)},
\end{aligned}$$

instead of $\mathbb{G}_{2,n}(\alpha)$ and $\mathbb{Y}_{2,n}(\alpha)$, respectively. Since their estimates can be proved in a similar way to the first half, we omit the details. □

To derive Proposition 2.3.5, we prepare the next lemma. For L_1 metric $d(\cdot,\cdot)$ on $\mathbb{R}$, we define the coupling distance $W(\cdot,\cdot)$ between any two probability measures P and Q by

$$W(P,Q) : = \inf\left\{\int_{\mathbb{R}^2} d(x,y)d\mu(x,y) : \mu \in M(P,Q)\right\}$$

$$= \inf \left\{ \int_{\mathbb{R}^2} |x - y| d\mu(x, y) : \mu \in M(P, Q) \right\},$$

where $M(P, Q)$ denotes the set of all probability measures on $\mathbb{R}^2$ with marginals P and Q. $W(\cdot, \cdot)$ is called the probabilistic Kantrovich-Rubinstein metric (or the first Wasserstein metric). The following assertion gives the exponential estimates of $W(P_t(\cdot, \cdot), \pi_0)$:

Lemma 2.5.1. *If Assumption 2.2.3 holds, then for any $q > 1$, there exists a positive constant C_q such that for all $x \in \mathbb{R}$,*

$$W(P_t(x, \cdot), \pi_0) \leq C_q \exp(-at)(1 + |x|^q).$$

Proof. We introduce the following Lipschitz semi-norm for a suitable real-valued function f on $\mathbb{R}$:

$$||f||_L := \sup\{|f(x) - f(y)|/|x - y| : x \neq y \text{ in } \mathbb{R}\}.$$

From Kantorovich-Rubinstein theorem (cf. [14, Theorem 11.8.2]) and Assumption 2.2.3, it follows that for all $x \in \mathbb{R}$,

$$\begin{aligned}
&W(P_t(x, \cdot), \pi_0) \\
&= \sup \left\{ \left| \int_{\mathbb{R}} f(y) \{P_t(x, dy) - \pi_0(dy)\} \right| : ||f||_L \leq 1 \right\} \\
&= \sup \left\{ \left| \int_{\mathbb{R}} (f(y) - f(0)) \{P_t(x, dy) - \pi_0(dy)\} \right| : ||f||_L \leq 1 \right\} \\
&\leq \sup \left\{ \left| \int_{\mathbb{R}} h(y) \{P_t(x, dy) - \pi_0(dy)\} \right| : |h(y)| \leq 1 + |y|^q \right\} \\
&\leq C_q \exp(-at)(1 + |x|^q).
\end{aligned}$$

□

Proof of Proposition 2.3.5 It is enough to check the conditions of [64, Theorem 3.1.1 and Theorem 3.1.3] for $p_\gamma = p_\alpha = 1$. As was mentioned in the proof of Theorem 2.3.1,

$$g_1(x) := -\partial_\gamma c(x, \gamma^\star)(c^2(x, \gamma^\star) - C^2(x))/c^3(x, \gamma^\star),$$

and

$$g_2(x) := -\partial_\alpha a(x, \alpha^\star)(A(x) - a(x, \alpha^\star))/c^2(x, \gamma^\star)$$

are centered. In the following, we give the proof concerning g_1 and omit its index 1 for simplicity. The regularity conditions on the coefficients imply that there exist positive constants L and D such that

$$|g(x) - g(y)| \leq D(2 + |x|^L + |y|^L)|x - y|.$$

Making use of the trivial inequalities $|x-y|^l \le |x|^l + |y|^l$ and $|x|^l \le 1 \vee |x|^{L+l}$ for any $L>0$, $l \in (0,1)$ and $x, y \in \mathbb{R}$, we have

$$\sup_{x,y\in\mathbb{R}, x\neq y} \frac{|g(x)-g(y)|}{(2+|x|^{L+1-1/p}+|y|^{L+1-1/p})|x-y|^{1/p}} < \infty,$$

for any $p>1$. Recall that we put $h_L(x) = 1+|x|^L$ in Assumption 2.2.3. The inequality (2.2.2) gives

$$\begin{aligned}
&\int_{\mathbb{R}} h_L(y) P_t(x,dy) \\
&\le ||P_t(x,\cdot)-\pi_0(\cdot)||_{h_L} + \int_{\mathbb{R}} (1+|y|^L)\pi_0(dy) \\
&\le \left(C_L + \int_{\mathbb{R}} (1+|y|^L)\pi_0(dy)\right) h_L(x).
\end{aligned}$$

We write $L' = L+1-1/p$ for abbreviation. Building on this estimate and the previous lemma, the conditions of [64, Theorem 3.1.1 and Theorem 3.1.3] are satisfied with

$$\begin{aligned}
&p=p,\ q=\frac{p}{p-1},\ d(x,y)=|x-y|,\ r(t)=\exp(-at),\ \phi(x)=1+|x|^{L'}, \\
&\psi(x) = 2^{q-1}\left(C_{qL'} + \int_{\mathbb{R}} h_{qL'}(y)\pi_0(dy)\right) h_{qL'}(x), \\
&\chi(x) = 2^{q^2-1}\left(C_{qL'} + \int_{\mathbb{R}} h_{qL'}(y)\pi_0(dy)\right)^q \left(C_{q^2L'} + \int_{\mathbb{R}} h_{q^2L'}(y)\pi_0(dy)\right) h_{q^2L'}(x),
\end{aligned}$$

and here these symbols correspond to the ones used in [64]. As for g_2, the conditions can be checked as well. Hence the desired result follows. □

To derive the asymptotic normality of $\hat{\theta}_n$, the following CLT-type theorem for stochastic integrals with respect to Poisson random measures will come into the picture:

Lemma 2.5.2. *Let $N(ds,dz)$ be a Poisson random measure associated with one-dimensional Lévy process defined on a stochastic basis $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t>0}, P)$ whose Lévy measure is written as ν_0. Assume that a continuous vector-valued function f on $\mathbb{R}_+ \times \mathbb{R} \times \mathbb{R}$ and a $\mathcal{F}_t$-predictable process H_t satisfy:*

1. *For all $T>0$ and $k=2,4$,*

$$E\left[\int_0^T \int_{\mathbb{R}} |f(T,H_s,z)|^k \nu_0(dz)ds\right] < \infty,$$

and their exists a positive definite matrix Σ such that

$$E\left[\int_0^T \int_{\mathbb{R}} f(T,H_s,z)^{\otimes 2} \nu_0(dz)ds\right] \to \Sigma,$$

as $T\to\infty$;

2. there exists $\delta > 0$ such that

$$E\left[\int_0^T \int_{\mathbb{R}} |f(T, H_s, z)|^{2+\delta} \nu_0(dz)ds\right] \to 0,$$

as $T \to \infty$.

Then, for the associated compensated Poisson random measure $\tilde{N}(ds, dz)$, we have

$$\int_0^T \int_{\mathbb{R}} f(T, H_s, z)\tilde{N}(ds, dz) \xrightarrow{\mathcal{L}} N(0, \Sigma),$$

as $T \to \infty$.

Proof. By Cramer-Wold device, it is sufficient to show only one-dimensional case. This proof is almost the same as [13, Theorem 14. 5. I]. For notational brevity, we set

$$X_1(t) := \int_0^t \int_{\mathbb{R}} f(T, H_s, z)\tilde{N}(ds, dz),$$
$$X_2(t) := \int_0^t \int_{\mathbb{R}} |f(T, H_s, z)|^2 \nu_0(dz)ds,$$

Introduce a stopping time $S := \inf\{t > 0 : X_2(t) \geq \Sigma\}$. Note that $X_2(S) = \Sigma$ because $X_2(t)$ is continuous. Define a random function $\zeta(u, t)$ and $\Psi(u, t)$ by

$$\zeta(u, t) = \exp\left\{iuX_1(t \wedge S) + \frac{u^2}{2}X_2(t \wedge S)\right\},$$
$$\Psi(u, t) = \exp\{iuf(T, H_t, z)\} - 1 - iuf(T, H_t, z) + \frac{u^2}{2}|f(T, H_t, z)|^2.$$

Applying Itô's formula, we obtain

$$\begin{aligned}
&\zeta(u, T) \\
&= 1 + iu\int_0^{T\wedge S} \zeta(u, s-)dX_1(s) + \frac{u^2}{2}\int_0^{T\wedge S} \zeta(u, s-)dX_2(s) \\
&+ \sum_{0<s\leq T\wedge S} (\zeta(u, s-)\exp\{iu\Delta X_1(s)\} - \zeta(u, s-) - iu\zeta(u, s-)\Delta X_1(s)) \\
&= 1 + \int_0^{T\wedge S} \int_{\mathbb{R}} \zeta(u, s-)\left(\exp\{iuf(T, H_s, z)\} - 1\right)\tilde{N}(ds, dz) \\
&+ \int_0^{T\wedge S} \int_{\mathbb{R}} \zeta(u, s)\Psi(u, s)\nu_0(dz)ds.
\end{aligned}$$

For later use, we here present the following elementary inequality (cf. [15]): for all $u \in \mathbb{R}$ and $n \in \mathbb{N} \cup \{0\}$,

$$\left|\exp(iu) - \sum_{j=0}^n \frac{(iu)^j}{j!}\right| \leq \frac{|u|^{n+1}}{(n+1)!} \wedge \frac{2|u|^n}{n!}. \tag{2.5.5}$$

By the definition of S, we have

$$|\zeta(u,T)| \le \exp\left\{\frac{u^2\Sigma}{2}\right\}.$$

Since

$$\int_0^T \int_{\mathbb{R}} \zeta(u,s-)\left(\exp\{iuf(T,H_s,z)\}-1\right)\tilde{N}(ds,dz)$$

is an L_2-martingale (cf. [3, Section 4]) from these estimates, the optional sampling theorem implies that

$$E\left[\int_0^{T\wedge S} \int_{\mathbb{R}} \zeta(u,s-)\left(\exp\{iuf(T,H_s,z)\}-1\right)\tilde{N}(ds,dz)\right] = 0.$$

Next we show that

$$E\left[\int_0^{T\wedge S}\int_{\mathbb{R}} \zeta(u,s)\Psi(u,s)\nu_0(dz)ds\right] \to 0.$$

Again using the above estimates, we have

$$\begin{aligned}
&\left|E\left[\int_0^{T\wedge S}\int_{\mathbb{R}} \zeta(u,s)\Psi(u,s)\nu_0(dz)ds\right]\right| \\
&\le E\left[\int_0^{T\wedge S}\int_{\mathbb{R}} \exp\left\{\frac{u^2}{2}\Sigma\right\}\left(\frac{|uf(T,H_s,z)|^3}{6}\wedge|uf(T,H_s,z)|^2\right)\nu_0(dz)ds\right] \\
&\le C_\delta \exp\left\{\frac{u^2}{2}\Sigma\right\} E\left[\int_0^T\int_{\mathbb{R}} |uf(T,H_s,z)|^{2+\delta}\nu_0(dz)ds\right] \\
&\to 0,
\end{aligned}$$

where C_δ is a positive constant such that

$$\frac{|x|^3}{6}\wedge|x|^2 \le C_\delta|x|^{2+\delta}$$

for all $x\in\mathbb{R}$. At last we observe that

$$X_1(T\wedge S) - X_1(T) \xrightarrow{P} 0.$$

In view of Lenglart's inequality and the isometry property of stochastic integral with respect to Poisson random measure (cf. [3, Section 4]), it suffices to show

$$E\left[\int_{T\wedge S}^T\int_{\mathbb{R}} |f(T,H_s,z)|^2\nu_0(dz)ds\right] \to 0.$$

However the latter convergence is clear from Assumption (1). Hence the proof is complete. □

Next we show the following lemma which gives the fundamental small time moment estimate of X:

Lemma 2.5.3. *Under Assumptions 2.2.1-2.2.3, it follows that*

$$E_{j-1}[|X_s - X_{j-1}|^p] \lesssim h_n(1 + |X_{j-1}|^p), \tag{2.5.6}$$

for any positive constant $p \in (1 \vee \beta, 2)$ and $s \in (t_{j-1}, t_j]$.

Proof. Recall that $\int |z|^p \nu_0(dz) < \infty$ from Assumption 2.2.1. By Lipschitz continuity of the coefficients and [17, Theorem 1.1], it follows that

$$\begin{aligned}
&E_{j-1}[|X_s - X_{j-1}|^p] \\
&\lesssim E_{j-1}\left[\left|\int_{t_{j-1}}^s (A_u - A_{j-1})du + \int_{t_{j-1}}^s (C_{u-} - C_{j-1})dZ_u\right|^p\right] \\
&+ h_n^p |A_{j-1}|^p + h_n |C_{j-1}|^p \int_{\mathbb{R}} |z|^p \nu_0(dz) + o_p(h_n) \\
&\lesssim E_{j-1}\left[\left|\int_{t_{j-1}}^{t_j} (C_{s-} - C_{j-1})dZ_s\right|^p\right] + h_n^{p-1} \int_{t_{j-1}}^{t_j} E_{j-1}[|X_s - X_{j-1}|^p]ds \\
&+ h_n \left(1 + |X_{j-1}|^p + o_p(1)\right)
\end{aligned}$$

Applying Burkholder-Davis-Gundy's inequality (cf. [50, Theorem 48]), we have

$$\begin{aligned}
&E_{j-1}\left[\left|\int_{t_{j-1}}^{t_j} (C_{s-} - C_{j-1})dZ_s\right|^p\right] \\
&\lesssim E_{j-1}\left[\left(\int_{t_{j-1}}^{t_j}\int_{\mathbb{R}} (C_{s-} - C_{j-1})^2 z^2 N(ds, dz)\right)^{p/2}\right] \\
&= E_{j-1}\left[\left(\sum_{t_{j-1} \le s < t_j} (C_{s-} - C_{j-1})^2 (Z_s - Z_{s-})^2\right)^{p/2}\right] \\
&\le E_{j-1}\left[\sum_{t_{j-1} \le s < t_j} |C_{s-} - C_{j-1}|^p |Z_s - Z_{s-}|^p\right] \\
&= \int_{t_{j-1}}^{t_j} E_{j-1}[|X_s - X_{j-1}|^p]ds \int_{\mathbb{R}} |z|^p \nu_0(dz),
\end{aligned}$$

for the Poisson random measure $N(ds, dz)$ associated with Z. Hence Gronwall's inequality gives (2.5.6). □

Proof of Theorem 2.3.7 According to Cramer-Wold device, it is enough to show for $p_\gamma = p_\alpha = 1$. From a similar estimates used in Theorem 2.3.1, we have

$$\sqrt{T_n}\partial_\gamma \mathbb{G}_{1,n}(\gamma^\star) \tag{2.5.7}$$

$$
\begin{aligned}
&= -\frac{2}{\sqrt{T_n}}\sum_{j=1}^{n}\left\{\frac{\partial_\gamma c_{j-1}}{c_{j-1}^3}\left(h_n c_{j-1}^2 - (\Delta_j X)^2\right)\right\}\\
&= -\frac{2}{\sqrt{T_n}}\sum_{j=1}^{n}\left\{\frac{\partial_\gamma c_{j-1}}{c_{j-1}^3}(h_n c_{j-1}^2 - C_{j-1}^2(\Delta_j Z)^2)\right\} + o_p(1)\\
&= -\frac{2}{\sqrt{T_n}}\sum_{j=1}^{n}\left\{\frac{\partial_\gamma c_{j-1}}{c_{j-1}^3}C_{j-1}^2(h_n - (\Delta_j Z)^2)\right\} - \frac{2}{\sqrt{T_n}}\int_0^{T_n}\frac{\partial_\gamma c_s}{c_s^3}(c_s^2 - C_s^2)ds\\
&- \frac{2}{\sqrt{T_n}}\sum_{j=1}^{n}\int_{t_{j-1}}^{t_j}\left\{\frac{\partial_\gamma c_{j-1}}{c_{j-1}^3}(c_{j-1}^2 - C_{j-1}^2) - \frac{\partial_\gamma c_s}{c_s^3}(c_s^2 - C_s^2)\right\}ds + o_p(1)\\
&=: \mathbb{F}_{1,n} + \mathbb{F}_{2,n} + \mathbb{F}_{3,n} + o_p(1). \qquad (2.5.8)
\end{aligned}
$$

We evaluate each term separately below. Rewriting $\mathbb{F}_{1,n}$ in a stochastic integral form via Itô's formula, we have

$$
\begin{aligned}
\mathbb{F}_{1,n} &= -\frac{2}{\sqrt{T_n}}\sum_{j=1}^{n}\int_{t_{j-1}}^{t_j}\int_{\mathbb{R}}\frac{\partial_\gamma c_{s-}}{c_{s-}^3}C_{s-}^2 z^2\tilde{N}(ds,dz)\\
&- \frac{2}{\sqrt{T_n}}\sum_{j=1}^{n}\int_{t_{j-1}}^{t_j}\int_{\mathbb{R}}\left(\frac{\partial_\gamma c_{j-1}}{c_{j-1}^3}C_{j-1}^2 - \frac{\partial_\gamma c_{s-}}{c_{s-}^3}C_{s-}^2\right)z^2\tilde{N}(ds,dz)\\
&- \frac{4}{\sqrt{T_n}}\sum_{j=1}^{n}\frac{\partial_\gamma c_{j-1}}{c_{j-1}^3}C_{j-1}^2\int_{t_{j-1}}^{t_j}(Z_{s-} - Z_{j-1})dZ_s.
\end{aligned}
$$

for the compensated Poisson random measure $\tilde{N}(ds,dz)$ associated with Z. Using Burkholder's inequality and the isometry property, it follows that for a positive constant K,

$$
\begin{aligned}
&E\left[\left(\frac{1}{\sqrt{T_n}}\sum_{j=1}^{n}\int_{t_{j-1}}^{t_j}\int_{\mathbb{R}}\left(\frac{\partial_\gamma c_{j-1}}{c_{j-1}^3}C_{j-1}^2 - \frac{\partial_\gamma c_{s-}}{c_{s-}^3}C_{s-}^2\right)z^2\tilde{N}(ds,dz)\right)^2\right]\\
&\lesssim \frac{1}{T_n}\sum_{j=1}^{n}E\left[\left(\int_{t_{j-1}}^{t_j}\int_{\mathbb{R}}\left(\frac{\partial_\gamma c_{j-1}}{c_{j-1}^3}C_{j-1}^2 - \frac{\partial_\gamma c_{s-}}{c_{s-}^3}C_{s-}^2\right)z^2\tilde{N}(ds,dz)\right)^2\right]\\
&\lesssim \frac{1}{T_n}\sum_{j=1}^{n}\int_{t_{j-1}}^{t_j}E\left[\left(\int_0^1\partial_x\left(\frac{\partial_\gamma c}{c^3}C^2\right)(X_{j-1} + u(X_s - X_{j-1}))du\right)(X_s - X_{j-1})\right]ds\\
&\lesssim \frac{1}{T_n}\sum_{j=1}^{n}\int_{t_{j-1}}^{t_j}\sqrt{\sup_{t\in\mathbb{R}^+}E[1+|X_t|^K]}\sqrt{E[(X_s - X_{j-1})^2]}ds\\
&\lesssim \sqrt{h_n},
\end{aligned}
$$

and that

$$
E\left[\left|\int_{t_{j-1}}^{t_j}(Z_{s-} - Z_{j-1})dZ_s\right|^2\right] \lesssim \int_{t_{j-1}}^{t_j}E[|Z_{s-t_{j-1}}|^2]ds \le h_n^2.
$$

Hence

$$\mathbb{F}_{1,n} = -\frac{2}{\sqrt{T_n}} \sum_{j=1}^{n} \int_{t_{j-1}}^{t_j} \int_{\mathbb{R}} \frac{\partial_\gamma c_{s-}}{c_{s-}^3} C_{s-}^2 z^2 \tilde{N}(ds, dz) + o_p(1).$$

Let us turn to observe $\mathbb{F}_{2,n}$. Let $f_{i,t} := f_i(X_t)$ for $i = 1, 2$, and especially, let $f_{i,j} := f_i(X_{t_j})$. From Proposition 2.3.5, we obtain

$$\begin{aligned}
\mathbb{F}_{2,n} &= -\frac{2}{\sqrt{T_n}} \sum_{j=1}^{n} \left(f_{1,j} - f_{1,j-1} + \int_{t_{j-1}}^{t_j} \frac{\partial_\gamma c_s}{c_s^3} (c_s^2 - C_s^2) ds \right) - \frac{2}{\sqrt{T_n}} (f_{1,n} - f_{1,0}) \\
&= -\frac{2}{\sqrt{T_n}} \sum_{j=1}^{n} \left(f_{1,j} - f_{1,j-1} + \int_{t_{j-1}}^{t_j} \frac{\partial_\gamma c_s}{c_s^3} (c_s^2 - C_s^2) ds \right) + o_p(1).
\end{aligned}$$

For abbreviation, we simply write

$$\xi_{1,j}(t) = f_{1,t} - f_{1,j-1} + \int_{t_{j-1}}^{t} \frac{\partial_\gamma c_s (c_s^2 - C_s^2)}{c_s^3} ds.$$

According to Proposition 2.3.5, the weighted Hölder continuity of f, and Lemma 2.5.3, $\{\xi_{1,j}(t), \mathcal{F}_{t_{j-1}+t} : t \in [0, h_n]\}$ turns out to be an L_2-martingale. Thus the martingale representation theorem [24, Theorem Ⅲ. 4. 34] implies that there exists a predictable process $s \mapsto \tilde{\xi}_{1,j}(s, z)$ such that

$$\xi_{1,j}(t) = \int_{t_{j-1}}^{t} \int_{\mathbb{R}} \tilde{\xi}_{1,j}(s, z) \tilde{N}(ds, dz).$$

Hence the continuous martingale component of $\xi_{1,j}$ is 0. By the property of f_1, we can define the stochastic integral $\int_{t_{j-1}}^{t} \int_{\mathbb{R}} \left(f_1(X_{s-} + C_{s-} z) - f_1(X_{s-}) \right) \tilde{N}(ds, dz)$ on $t \in [t_{j-1}, t_j]$ and this process is also an L_2-martingale with respect to $\{\mathcal{F}_{t_{j-1}+t} : t \in [0, h_n]\}$. Utilizing [24, Theorem I. 4. 52] and [50, Corollary Ⅱ. 6. 3], we have

$$\begin{aligned}
&E\left[\left| \frac{1}{\sqrt{T_n}} \sum_{j=1}^{n} \left\{ \xi_{1,j}(t_j) - \int_{t_{j-1}}^{t_j} \int_{\mathbb{R}} \left(f_1(X_{s-} + C_{s-} z) - f_1(X_{s-}) \right) \tilde{N}(ds, dz) \right\} \right|^2 \right] \\
&\lesssim \frac{1}{T_n} \sum_{j=1}^{n} E\left[\left| \xi_{1,j}(t_j) - \int_{t_{j-1}}^{t_j} \int_{\mathbb{R}} \left(f_1(X_{s-} + C_{s-} z) - f_1(X_{s-}) \right) \tilde{N}(ds, dz) \right|^2 \right] \\
&= \frac{1}{T_n} \sum_{j=1}^{n} E\left[\left[\xi_{1,j}(\cdot) - \int_{t_{j-1}}^{\cdot} \int_{\mathbb{R}} \left(f_1(X_{s-} + C_{s-} z) - f_1(X_{s-}) \right) \tilde{N}(ds, dz) \right]_{t_j} \right] \\
&= 0.
\end{aligned}$$

Here $[Y.]_t$ denotes the quadratic variation for any semimartingale Y at time t, and we used Burkholder's inequality for a martingale difference between the first line and the second line. By similar estimates above, we have $\mathbb{F}_{3,n} = o_p(1)$. Having these arguments in hand, it turns out that

$$\begin{aligned}
&\sqrt{T_n}\partial_\gamma \mathbb{G}_{1,n}(\gamma^\star) \\
&= -\frac{2}{\sqrt{T_n}}\int_0^{T_n}\int_{\mathbb{R}}\left(\frac{\partial_\gamma c_{s-}}{c_{s-}^3}C_{s-}^2 z^2 + f_1(X_{s-}+C_{s-}z) - f_1(X_{s-})\right)\tilde{N}(ds,dz) + o_p(1).
\end{aligned}$$

We can deduce from Assumption 2.2.2 and Proposition 2.3.5 that there exist positive constants K, K', K'' and $\epsilon_0 < 1\wedge(2-\beta)$ such that for all $z\in\mathbb{R}$

$$\begin{aligned}
&\sup_t\left\{\frac{1}{t}\int_0^t E\left[\left(\frac{\partial_\gamma c_s}{c_s^3}C_s^2 z^2 + f_1(X_s - C_s z) - f_1(X_s)\right)^2\right]ds\right\} \\
&\lesssim \sup_t\left\{\frac{1}{t}\int_0^t \left(|z|^{2-\epsilon_0}\vee z^4\right)\left(1+\sup_t E\left[|X_t|^K\right] + \left(1+\sup_t E\left[|X_t|^{K'}\right]\right)|z|^{K''}\right)ds\right\} \\
&\lesssim (|z|^{2-\epsilon_0}\vee z^4)\left(1+|z|^{K''}\right),
\end{aligned}$$

and the last term is ν_0-integrable. Then, there exist positive constants K and K' (possibly take different values from the previous ones) such that for any $z\in\mathbb{R}$,

$$\begin{aligned}
&\left|\frac{1}{t}\int_0^t E\left[\left(\frac{\partial_\gamma c_s}{c_s^3}C_s^2 z^2 + f_1(X_s + C_s z) - f_1(X_s)\right)^2\right]ds\right. \\
&\qquad\qquad \left. - \int_{\mathbb{R}}\left(\frac{\partial_\gamma c(y,\gamma^\star)}{c^3(y,\gamma^\star)}C^2(y)z^2 + f_1(y+C(y)z) - f_1(y)\right)^2\pi_0(dy)\right| \\
&= \left|\frac{1}{t}\int_0^t\int_{\mathbb{R}}\int_{\mathbb{R}}\left(\frac{\partial_\gamma c(y,\gamma^\star)}{c^3(y,\gamma^\star)}C^2(y)z^2 + f_1(y+C(y)z) - f_1(y)\right)^2(P_s(x,dy)-\pi_0(dy))\eta(dx)ds\right| \\
&\lesssim (|z|^{2-\epsilon_0}\vee z^4)\left(1+|z|^{K'}\right)\frac{1}{t}\int_0^t\int_{\mathbb{R}}||P_s(x,\cdot)-\pi_0(\cdot)||_{h_K}\eta(dx)ds \\
&\to 0.
\end{aligned}$$

Thus the dominated convergence theorem and the isometry property give

$$\begin{aligned}
&\lim_{n\to\infty} E\left[\left(\frac{1}{\sqrt{T_n}}\int_0^{T_n}\int_{\mathbb{R}}\left(\frac{\partial_\gamma c_{s-}}{c_{s-}^3}C_{s-}^2 z^2 + f_1(X_{s-}+C_{s-}z) - f_1(X_{s-})\right)\tilde{N}(ds,dz)\right)^2\right] \\
&= \lim_{n\to\infty}\frac{1}{T_n}\int_0^{T_n}\int_{\mathbb{R}} E\left[\left(\frac{\partial_\gamma c_s}{c_s^3}C_s^2 z^2 + f_1(X_s + C_s z) - f_1(X_s)\right)^2\right]\nu_0(dz)ds \\
&= \frac{1}{4}\Sigma_\gamma.
\end{aligned}$$

It follows from Assumption 2.2.3 and Proposition 2.3.5 that

$$E\left[\int_0^{T_n}\int_{\mathbb{R}}\left|\frac{1}{\sqrt{T_n}}\left(\frac{\partial_\gamma c_s}{c_s^3}C_s^2 z^2 + f_1(X_s+C_s z) - f_1(X_s)\right)\right|^{2+K}\nu_0(dz)ds\right]\to 0.$$

From Taylor expansion around $\gamma^\star$, $\partial_\alpha \mathbb{G}_{2,n}(\alpha)$ is decomposed as:

$$\begin{aligned}
&\sqrt{T_n}\partial_\alpha \mathbb{G}_{2,n}(\alpha^\star)\\
&= \frac{2}{\sqrt{T_n}}\sum_{j=1}^n \frac{\partial_\alpha a_{j-1}}{c_{j-1}^2}(\Delta_j X - h_n a_{j-1}) + \frac{2}{T_n}\sum_{j=1}^n \partial_\alpha a_{j-1}(\Delta_j X - h_n a_{j-1})\partial_\gamma c_{j-1}^{-2}\left(\sqrt{T_n}(\hat{\gamma}_n-\gamma^\star)\right)\\
&+ \left(\int_0^1 \frac{2}{(T_n)^{3/2}}\sum_{j=1}^n \partial_\alpha a_{j-1}(\Delta_j X - h_n a_{j-1})\partial_\gamma^2 c_{j-1}^{-2}(\gamma^\star + u(\hat{\gamma}_n-\gamma^\star))du\right)\left(\sqrt{T_n}(\hat{\gamma}_n-\gamma^\star)\right)^2
\end{aligned}$$

Sobolev's inequality and the tail probability estimates of $\hat{\gamma}_n$ imply that the third term of the right-hand-side is $o_p(1)$. Hence a similar manner to the first half leads to

$$\begin{aligned}
&\sqrt{T_n}\partial_\alpha \mathbb{G}_{2,n}(\alpha^\star) - \frac{2}{T_n}\sum_{j=1}^n \partial_\alpha a_{j-1}(\Delta_j X - h_n a_{j-1})\partial_\gamma c_{j-1}^{-2}\left(\sqrt{T_n}(\hat{\gamma}_n-\gamma^\star)\right)\\
&= \frac{2}{\sqrt{T_n}}\sum_{j=1}^n \frac{\partial_\alpha a_{j-1}}{c_{j-1}^2}(\Delta_j X - h_n a_{j-1}) + o_p(1)\\
&= \frac{2}{\sqrt{T_n}}\sum_{j=1}^n \frac{\partial_\alpha a_{j-1}}{c_{j-1}^2}C_{j-1}\Delta_j Z + \frac{2}{\sqrt{T_n}}\int_0^{T_n}\frac{\partial_\alpha a_s}{c_s^2}(A_s - a_s)ds + o_p(1)\\
&= \frac{2}{\sqrt{T_n}}\sum_{j=1}^n \left(f_{2,j} - f_{2,j-1} + \int_{t_{j-1}}^{t_j}\frac{\partial_\alpha a_s}{c_s^2}(A_s - a_s)ds\right)\\
&+ \frac{2}{\sqrt{T_n}}\sum_{j=1}^n \int_{t_{j-1}}^{t_j}\int_{\mathbb{R}}\frac{\partial_\alpha a_s}{c_{s-}^2}C_{s-}\tilde{N}(ds,dz) + o_p(1)\\
&= \frac{2}{\sqrt{T_n}}\int_0^{T_n}\int_{\mathbb{R}}\left(\frac{\partial_\alpha a_s}{c_{s-}^2}C_{s-}z + f_2(X_{s-}+C_{s-}z) - f_2(X_{s-})\right)\tilde{N}(ds,dz) + o_p(1),
\end{aligned}$$

and we have

$$\begin{aligned}
&\lim_{n\to\infty} E\left[\left(\frac{1}{\sqrt{T_n}}\int_0^{T_n}\int_{\mathbb{R}}\left(\frac{\partial_\alpha a_s}{c_{s-}^2}C_{s-}z + f_2(X_{s-}+C_{s-}z) - f_2(X_{s-})\right)\tilde{N}(ds,dz)\right)^2\right] = \frac{1}{4}\Sigma_\alpha,\\
&\lim_{n\to\infty} E\left[\int_0^{T_n}\int_{\mathbb{R}}\left|\frac{1}{\sqrt{T_n}}\left(\frac{\partial_\alpha a_s}{c_s^2}C_s z + f_2(X_s+C_s z) - f_2(X_s)\right)\right|^{2+K}\nu_0(dz)ds\right] = 0.
\end{aligned}$$

From the isometry property and the trivial identity $xy = \left\{(x+y)^2 - (x-y)^2\right\}/4$ for any $x, y \in \mathbb{R}$, it follows that

$$\lim_{n\to\infty} E\left[\left(\frac{1}{\sqrt{T_n}}\int_0^{T_n}\int_{\mathbb{R}}\left(\frac{\partial_\gamma c_{s-}}{c_{s-}^3}C_{s-}^2 z^2 + f_1(X_{s-}+C_{s-}z) - f_1(X_{s-})\right)\tilde{N}(ds,dz)\right)\right.$$
$$\left.\times\left(\frac{1}{\sqrt{T_n}}\int_0^{T_n}\int_{\mathbb{R}}\left(\frac{\partial_\alpha a_s}{c_{s-}^2}C_{s-}z + f_2(X_{s-}+C_{s-}z) - f_2(X_{s-})\right)\tilde{N}(ds,dz)\right)\right] = -\frac{1}{4}\Sigma_{\alpha\gamma}.$$

Hence the moment estimates in the proof of Theorem 2.3.1, Lemma 2.5.2 and Taylor's formula yield that

$$\sqrt{T_n}\begin{pmatrix} -\partial_\gamma^2 \mathbb{G}_{1,n}(\gamma^\star) & 0 \\ -\frac{2}{T_n}\sum_{j=1}^n \partial_\alpha a_{j-1}(\Delta_j X - h_n a_{j-1})\partial_\gamma c_{j-1}^{-2} & -\partial_\alpha^2 \mathbb{G}_{2,n}(\alpha^\star)\end{pmatrix}\begin{pmatrix}\hat{\gamma}_n - \gamma^\star \\ \hat{\alpha}_n - \alpha^\star\end{pmatrix}$$
$$= \sqrt{T_n}\begin{pmatrix}\partial_\gamma \mathbb{G}_{1,n}(\gamma^\star) \\ \partial_\alpha \mathbb{G}_{2,n}(\alpha^\star)\end{pmatrix} + o_p(1) \xrightarrow{\mathcal{L}} N(0,\Sigma).$$

To achieve the desired result, it suffices to show

$$\partial_\gamma^2 \mathbb{G}_{1,n}(\gamma^\star) \xrightarrow{P} \Gamma_\gamma,\ \partial_\alpha^2 \mathbb{G}_{2,n}(\alpha^\star) \xrightarrow{P} \Gamma_\alpha,$$

and

$$\frac{2}{T_n}\sum_{j=1}^n \partial_\alpha a_{j-1}(\Delta_j X - h_n a_{j-1})\partial_\gamma c_{j-1}^{-2} \xrightarrow{P} \Gamma_{\alpha\gamma}.$$

However the first two convergence are straightforward from the proof of Theorem 2.3.1, and the last convergence follows from the ergodic theorem. Thus the proof is complete.

Chapter 3

Estimation method for ergodic jump diffusion models based on iterative Jarque-Bera type test

3.1 Introduction

Suppose that we are given discrete-time but high-frequency observation $(X_{t_j^n})_{j=0}^n$ from a solution to the one-dimensional ergodic stochastic differential equation (SDE) with jumps

$$dX_t = a(X_t, \alpha)dw_t + b(X_t, \beta)dt + c(X_{t-})dJ_t, \tag{3.1.1}$$

defined on a complete filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, P)$. Each ingredient is supposed to be as follows:

- The coefficients $a : \mathbb{R} \times \Theta_\alpha \mapsto \mathbb{R}$ and $b : \mathbb{R} \times \Theta_\beta \mapsto \mathbb{R}$ are Lipschitz continuous and known except for the p-dimensional parameter

$$\theta := (\alpha, \beta) \in \Theta_\alpha \times \Theta_\beta = \Theta,$$

  where Θ_α and Θ_β are bounded convex domains and subset of $\mathbb{R}^{p_\alpha}$ and $\mathbb{R}^{p_\beta}$, respectively.

- w is a standard Wiener process and J a compound Poisson process, that is, for a Poisson process N whose intensity parameter is $\lambda \in [0, \infty)$ and i.i.d random variables $\{\xi_i\}_{i\in\mathbb{N}}$, it is expressed as

$$J_t = \sum_{i=1}^{N_t} \xi_i.$$

- The sampling times fulfill that for a positive sequence (h_n), t_j^n can be written as
$$t_j = t_j^n = jh_n, \tag{3.1.2}$$
and that the terminal sampling time $T_n := t_n^n = nh_n \to \infty$.
- (w, J) is $\mathcal{F}_t$-adapted, and the initial variable X_0 is $\mathcal{F}_0$-adapted and independent of (w, J).

Throughout this chapter, we assume that there exists a true value $\theta_0 := (\alpha_0, \beta_0) \in \Theta$.

On the one hand, for diffusion models, many estimator of θ have been proposed, such as Gaussian quasi-likelihood estimator [27], adaptive estimator [61], multi-step estimator [26], to mention few. On the other hand, in the presence of the jump component, elimination of the effect of J is apparently crucial for an accurate estimation of θ. A well-known approach for it is the threshold based method independently proposed in [34], [57], and [48]. In the method, we regard that the increment
$$\Delta_j X := X_{t_j} - X_{t_{j-1}},$$
contains the jump component if $|\Delta_j X| > r_n$ for a fixed *jump-detection threshold* $r_n > 0$, and estimate θ after removing such increments. For a suitably chosen $r_n > 0$, it is shown that the estimator of θ has asymptotic normality at the same rate as diffusion models, while finite-sample performance of the threshold method strongly depends on the value of r_n. A data-adaptive and quantitative choice of the threshold in the jump-detection filter is a subtle and sensitive problem, and still remains as an annoying problem in practice; see [55], [56], as well as the references therein. This practical issue can also be seen in other jump detection methods such as [2].

The primary objective of this chapter is to formulate an intuitively easy-to-understand strategy, which can simultaneously estimate θ and detect jumps without any precise calibration of a jump-detection threshold. For this purpose, we utilize the approximate self-normalized residuals [39], which makes the classical Jarque-Bera test [25] adapted to our model. More specifically, the hypothesis test whose significance level is $\alpha \in (0, 1)$ is constructed by the following manner: let the null hypothesis be of "no jump component" :
$$\mathcal{H}_0 : \lambda = 0,$$
against the alternative hypothesis of "non-trivial jump component":
$$\mathcal{H}_1 : \lambda > 0.$$
Then, if the Jarque-Bera type statistic introduced later is larger than a given percentile of the chi-square distribution with 2 degrees of freedom, we reject the null hypothesis $\mathcal{H}_0$; and otherwise, we accept $\mathcal{H}_0$. For such a test, we can intuitively regard that

the largest increment contains at least one jump when the null hypothesis is rejected. Following this inspection, our proposed method will goes as follows: we iteratively conduct the test with removing the largest increments in the retained samples until rejection of $\mathcal{H}_0$ is stopped; after that, we construct the modified estimator of θ by the remaining samples. Our method enables us not only just to make a "pre-cleaning" of diffusion-like data sequence by removing large jumps which collapse the approximate Gaussianity of the self-normalized residuals, but also to approximately quantify jumps relative to continuous fluctuations in a natural way.

This chapter is organized as follows: in Section 3.2, we give a brief summary of the approximate self-normalized residuals, and the Jarque-Bera type test for jump diffusion models. Section 3.3 provides our strategy and some remarks for its practical use. In Section 3.4, we will focus on a least-squares type estimator and its one-step version for the following model:

$$dX_t = \left(\sum_{l=1}^{p_\alpha} \alpha^{(l)} a^{(l)}(X_t)\right)^{1/2} dw_t + \sum_{k=1}^{p_\beta} \beta^{(k)} b^{(k)}(X_t)dt + c(X_{t-})dJ_t, \tag{3.1.3}$$

with suitable functions $\{a^{(l)}(x)\}_{l=1}^{p_\alpha}$ and $\{b^{(k)}(x)\}_{k=1}^{p_\beta}$. It will be seen that in the calculation of the estimator we can sidestep optimization, and thus it is numerically tractable, retaining high representational power of the nonlinearity in the state variable. Moreover, the estimator has not only consistency but also an asymptotic equivalence to an good estimator based only on the unobserved continuous part of X. We show some numerical experiments result in Section 3.5. Finally, Appendix 3.6 presents the technical proofs of the result given in Section 3.4.

3.2 Jarque-Bera normality test for jump diffusion models

To see whether a working model fits data well or not, and/or whether data in hand have outliers or not, diagnosis based on residual analysis is often done. For jump diffusion models, [39] formulated a Jarque-Bera normality test based on self-normalized residuals. In this section, we briefly review the construction of the self-normalized residual, and the Jarque-Bera statistics with its asymptotic behavior.

For each $j \in \{1, \dots, n\}$, introduce the function

$$\epsilon_j(\alpha) = \epsilon_{n,j}(\alpha) := \frac{\Delta_j X}{\sqrt{a_{j-1}^2(\alpha) h_n}}. \tag{3.2.1}$$

Then, following [39] we introduce the self-normalized residual and the Jarque-Bera type statistic by

$$\hat{N}_j = \hat{S}_n^{-1/2}(\epsilon_j(\hat{\alpha}_n) - \bar{\hat{\epsilon}}_n),$$
$$\mathrm{JB}_n = \frac{1}{6n}\left(\sum_{j=1}^{n}(\hat{N}_j)^3 - 3\sqrt{h_n}\sum_{j=1}^{n}\partial_x a_{j-1}(\hat{\alpha}_n)\right)^2 + \frac{1}{24n}\left(\sum_{j=1}^{n}((\hat{N}_j)^4 - 3)\right)^2,$$

where

$$\bar{\hat{\epsilon}}_n := \frac{1}{n}\sum_{j=1}^{n}\epsilon_j(\hat{\alpha}_n), \quad \hat{S}_n := \frac{1}{n}\sum_{j=1}^{n}(\epsilon_j(\hat{\alpha}_n) - \bar{\hat{\epsilon}}_n)^2.$$

The following theorem gives the asymptotic behavior of JB_n, which ensures theoretical validity of the Jarque-Bera type test based on JB_n.

Theorem 3.2.1. *([39, Theorems 3.1 and 4.1])*

1. *Under $\mathcal{H}_0 : \lambda = 0$ and suitable regularity conditions, for any estimator $\hat{\alpha}_n$ of α satisfying*
$$\sqrt{n}(\hat{\alpha}_n - \alpha_0) = O_p(1), \tag{3.2.2}$$
we have
$$\mathrm{JB}_n \xrightarrow{\mathcal{L}} \chi^2(2).$$

2. *Under $\mathcal{H}_1 : \lambda > 0$ and suitable regularity conditions, we have*
$$\mathrm{JB}_n \xrightarrow{P} \infty.$$

Remark 3.2.2. *The residual defined by (3.2.1) is of the Euler type with ignoring the drift fluctuation; under the sampling conditions in Assumption 3.4.1 given later, we can ignore the presence of the drift term in construction of residuals. Indeed, as in [40], instead of (3.2.1) we could consider*

$$\epsilon_j(\theta) = \epsilon_{n,j}(\theta) := \frac{\Delta_j X - h_n b_{j-1}(\beta)}{\sqrt{a_{j-1}^2(\alpha)h_n}}.$$

This case may require more computation time, while we would then have a more or less stabilized performance under $\mathcal{H}_0$ compared with the case of (3.2.1).

In the rest of this section, suppose that the null hypothesis $\mathcal{H}_0$ is true; namely the underlying model is a diffusion process. Among choices of $\hat{\alpha}_n$, the Gaussian quasi-maximum likelihood estimator (GQMLE) is one of the most important candidates because it has the asymptotic efficiency. The GQMLE is defined as any maximizer of the

Gaussian quasi-likelihood (GQL)

$$\mathbb{H}_n(\theta) := \sum_{j=1}^{n} \log \left\{ \frac{1}{\sqrt{2\pi a_{j-1}^2(\alpha) h_n}} \phi \left(\frac{\Delta_j X - b_{j-1}(\beta) h_n}{\sqrt{a_{j-1}^2(\alpha) h_n}} \right) \right\},$$

where ϕ denotes the standard normal density. This quasi-likelihood is constructed based on the local-Gauss approximation of the transition probability $\mathcal{L}(X_{t_j}|X_{t_{j-1}})$ by $N(b_{j-1}(\beta)h_n, a_{j-1}^2(\alpha)h_n)$. It is well known that the asymptotic normality holds true under suitable regularity conditions [27]: For the GQMLE $\tilde{\theta}_n = (\tilde{\alpha}_n, \tilde{\beta}_n)$, we have

$$\left(\sqrt{n}(\tilde{\alpha}_n - \alpha_0),\ \sqrt{T_n}(\tilde{\beta}_n - \beta_0) \right) \xrightarrow{\mathcal{L}} N\left(0,\ \mathrm{diag}(I_1^{-1}(\alpha_0), I_2^{-1}(\beta_0))\right),$$

where

$$I_1(\alpha_0) = \frac{1}{2} \int \left(\frac{\partial_\alpha a^2}{a^2}(x, \alpha_0) \right)^{\otimes 2} \pi_0(dx),$$
$$I_2(\beta_0) = \int \left(\frac{\partial_\beta b}{a}(x, \beta_0) \right)^{\otimes 2} \pi_0(dx).$$

Here π_0 denotes the invariant measure of X. We note that the GQMLE is asymptotically efficient in Hájek-Le Cam sense (cf. [20]).

If the coefficients a and b are highly nonlinear and/or the number of the parameters is large, then the calculation of the GQMLE can be quite time-consuming. To deal with such a problem, it is effective to separate optimizations of α and β by utilizing the difference of the small-time stochastic orders of the dt- and dw_t-terms. To be specific we introduce the following stepwise version of the GQMLE $\check{\theta}_n := (\check{\alpha}_n, \check{\beta}_n)$:

$$\check{\alpha}_n \in \underset{\alpha \in \bar{\Theta}_\alpha}{\mathrm{argmax}} \sum_{j=1}^{n} \log \left\{ \frac{1}{\sqrt{2\pi a_{j-1}^2(\alpha) h_n}} \phi \left(\frac{\Delta_j X}{\sqrt{a_{j-1}^2(\alpha) h_n}} \right) \right\},$$
$$\check{\beta}_n \in \underset{\beta \in \bar{\Theta}_\beta}{\mathrm{argmax}} \sum_{j=1}^{n} \log \left\{ \frac{1}{\sqrt{2\pi a_{j-1}^2(\check{\alpha}_n) h_n}} \phi \left(\frac{\Delta_j X - b_{j-1}(\beta) h_n}{\sqrt{a_{j-1}^2(\check{\alpha}_n) h_n}} \right) \right\}.$$

Under suitable regularity condition, it is shown that the stepwise GQMLE has the same asymptotic distribution as the original GQMLE $\tilde{\theta}_n$ (cf. [61]). Hence its asymptotic efficiency, and the same result of Theorem 3.2.1 holds for it as well. Although we have to conduct two optimization for the stepwise estimation scheme, it reduces the numbers of the parameters to be simultaneously optimized, thus relieving the computational time.

3.3 Proposed strategy

Let $q \in (0,1)$ be a small number, which will later serve as the significance level. Based on the Jarque-Bera type test introduced in the previous section, we propose an iterative jump detection procedure. We implicitly suppose that we are given an estimator $\hat{\theta}_n$ of θ defined to be any element $\hat{\theta}_n \in \operatorname{argmax} M_n$ for some contrast function M_n of the from

$$M_n(\theta) := \sum_{j=1}^{n} m_{h_n}\left(X_{t_{j-1}}, \Delta_j X;\, \theta\right).$$

Then, our procedure is as follows. We denote by $\chi^2_q(2)$ the q-percent critical value of the chi-squared distribution with 2 degrees of freedom.

Step 0. Set $k = k_n = 0$, and let $\hat{\mathcal{J}}_n^0 := \emptyset$.

Step 1. Calculate the modified estimator $\hat{\theta}_n^k$ defined by

$$\hat{\theta}_n^k \in \operatorname*{argmax}_{\theta \in \Theta} \sum_{j \le n;\, \notin \hat{\mathcal{J}}_n^k} m_{h_n}\left(X_{t_{j-1}}, \Delta_j X;\, \theta\right),$$

then let

$$\bar{\hat{\epsilon}}_n^k := \frac{1}{n-k} \sum_{j \notin \hat{\mathcal{J}}_n^k} \epsilon_j(\hat{\alpha}_n^k), \qquad \hat{S}_n^k := \frac{1}{n-k} \sum_{j \notin \hat{\mathcal{J}}_n^k} (\epsilon_j(\hat{\alpha}_n^k) - \bar{\hat{\epsilon}}_n^k)^2,$$

and (re-)construct the following modified self-normalized residuals $(\hat{N}_j^k)_{j=1}^n$ and Jarque-Bera type statistics JB_n^k:

$$\begin{aligned} \hat{N}_j^k &:= (\hat{S}_n^k)^{-1/2}(\epsilon_j(\hat{\alpha}_n^k) - \bar{\hat{\epsilon}}_n^k), \\ \mathrm{JB}_n^k &:= \frac{1}{6(n-k)} \left(\sum_{j \notin \hat{\mathcal{J}}_n^k} (\hat{N}_j^k)^3 - 3\sqrt{h_n} \sum_{j \notin \hat{\mathcal{J}}_n} \partial_x b_{j-1}(\hat{\theta}_n^k) \right)^2 \\ &\quad + \frac{1}{24(n-k)} \left(\sum_{j \notin \hat{\mathcal{J}}_n^k} ((\hat{N}_j^k)^4 - 3) \right)^2. \end{aligned} \tag{3.3.1}$$

Step 2. If $\mathrm{JB}_n^k > \chi^2_q(2)$, then pick out the interval number

$$j(k+1) := \operatorname*{argmax}_{j \in \{1,\dots,n\} \setminus \hat{\mathcal{J}}_n^k} |\Delta_j X|$$

(We here implicitly assume that there is no tie among the values $|\Delta_1 X|, \dots, |\Delta_n X|$), add it to the set $\hat{\mathcal{J}}_n^k$:

$$\hat{\mathcal{J}}_n^{k+1} := \hat{\mathcal{J}}_n^k \cup \{j(k+1)\},$$

and then return to *Step 1.* If $\mathrm{JB}_n^k \le \chi_q^2(2)$, then set an estimated the number of jumps to be

$$k^\star = k^\star(\omega) := \min\left\{k \le n;\ \mathrm{JB}_n^k \le \chi_q^2(2)\right\}$$

and go to *Step 3.*

Step 3. If $k^\star = 0$, regard that there is no jump; otherwise, we regard that each of $\Delta_{j(1)}X, \dots, \Delta_{j(k^\star)}X$ contains one jump. Finally, set $\hat{\theta}_n^{k^\star}$ to be an estimator of θ.

The above-described method enables us to divide the set of the whole increments $(\Delta_j X)_{j=1}^n$ into the following two categories:

- "One-jump" group $(\Delta_j X)_{j \in \hat{\mathcal{J}}_n^{k^\star}} = \{\Delta_{j(1)}X, \dots, \Delta_{j(k^\star)}X\}$, and
- "No-jump" group $(\Delta_j X)_{j \notin \hat{\mathcal{J}}_n^{k^\star}} = (\Delta_j X)_{j=1}^n \setminus \{\Delta_{j(1)}X, \dots, \Delta_{j(k^\star)}X\}$.

Automatically entailed just after jump removals is stopped is the estimator $\hat{\theta}_n^{k^\star}$ of the drift and diffusion part of X, which is the maximizer of the modified Gaussian quasi-likelihood defined by

$$\theta \mapsto \sum_{j \notin \hat{\mathcal{J}}_n^{k^\star}} \log\left\{\frac{1}{\sqrt{2\pi a_{j-1}^2(\alpha)h_n}}\phi\left(\frac{\Delta_j X - b_{j-1}(\beta)h_n}{\sqrt{a_{j-1}^2(\alpha)h_n}}\right)\right\}.$$

Remark 3.3.1. *In the above-described procedure we simply remove the largest increments at each step, with keeping the positions of the remaining data. Note that in the construction of the modified estimator $\hat{\theta}_n^k$ it is incorrect to use the "shifted" samples $(Y_{t_j})_{j \notin \hat{\mathcal{J}}_n^{k_n}}$ defined by*

$$Y_{t_j} = X_{t_j} - \sum_{i \in \hat{\mathcal{J}}_n^{k_n} \cap \{1, \dots, j\}} \Delta_i X.$$

This is because one-step transition density of the original process X is spatially different from Y, so that the estimation result would not suitably reflect the information of data.

Remark 3.3.2. *In practice, the size of "last-removed" increment would be used as the threshold for detecting jumps for future observations: with the value $r_n(k) := |\Delta_{j(k)}X|$ in hand, for future observations $(Y_{t_j^n})_{j=0}^n$ we regard that a jump occurred over $[t_{j-1}, t_j]$ if*

$$|\Delta_j Y| > r_n(k).$$

Remark 3.3.3. *When the jump coefficient is parameterized as $c(x,\gamma)$ and a model of the common jump distribution, say F_J, of the compound Poisson process J is given, it might be possible to consider estimation of γ and F_J based on the sequence $\{\Delta_{j(k)}X/c_{j(k)-1}(\gamma)\}_k$, with supposing that they are i.i.d. random variables with common jump distribution F_J. This is beyond the scope of this chapter, and we leave it as a future study.*

Remark 3.3.4. *At k-th iteration, it can be regarded that we conduct the Jarque-Bera type test for the trimmed data $(X_{t_{j-1}},\Delta_j X)_{j\notin\hat{\mathcal{J}}_n^k}$. Hence the null hypothesis $\mathcal{H}_0^k$ and alternative hypothesis $\mathcal{H}_1^k$ of the test are formally written as follows:*

$$\mathcal{H}_0^k : \sharp\{j\in\{1,\dots,n\}\mid \Delta_j N\geq 1\}\leq k,$$
$$\mathcal{H}_1^k : \sharp\{j\in\{1,\dots,n\}\mid \Delta_j N\geq 1\}> k,$$

where $\sharp A$ denotes the cardinality of a set A. From this formulation, we have the inclusion relation

$$\mathcal{H}_0\subset\mathcal{H}_0^1\subset\mathcal{H}_0^2\subset\cdots\subset\mathcal{H}_0^k\subset\cdots.$$

This inclusion relation implicitly suggests that we can "skip" first some redundant stages when seemingly several jumps do exist. Such a situation may often occur because the expected number of jumps of the compound Poisson process J up to time t is λt, namely, number of the jumps becomes larger and larger as the terminal sampling time T_n increases.

3.4 Asymptotic results

As was mentioned in the previous section, we have a choice of an estimator of θ. As a matter of course, for each estimator $\hat{\theta}_n$, we need to study asymptotic behavior of its modified version $\hat{\theta}_n^{k*}$. In this section, we will derive asymptotic results for a numerically tractable least-squares type estimator and the corresponding one-step improved version, when the underlying SDE (3.1.1) is of the from:

$$dX_t=\left(\sum_{l=1}^{p_\alpha}\alpha^{(l)}a^{(l)}(X_t)\right)^{1/2}dw_t+\sum_{k=1}^{p_\beta}\beta^{(k)}b^{(k)}(X_t)dt+c(X_{t-})dJ_t,$$

where the real-valued functions $a^{(1)},\dots,a^{(p_\alpha)}$ and $b^{(1)},\dots,b^{(p_\beta)}$ are known. For simplicity, we write

$$\mathbb{A}(x)=(a^{(1)}(x),\dots,a^{(p_\alpha)}(x))^\top,\quad \mathbb{B}(x)=(b^{(1)}(x),\dots,b^{(p_\beta)}(x))^\top.$$

Then, we construct an estimator of θ in the following manner:

- Diffusion parameter
    1. Least square estimator (LSE):
    $$\tilde{\alpha}_n := \operatorname*{argmin}_\alpha \sum_{j=1}^n \left\{(\Delta_j X)^2 - h_n \mathbb{A}_{j-1}^\top \alpha\right\}^2.$$
    2. Scoring:
    $$\hat{\alpha}_n := \tilde{\alpha}_n - \left(\sum_{j=1}^n \frac{\mathbb{A}_{j-1}\mathbb{A}_{j-1}^\top}{(\mathbb{A}_{j-1}^\top \tilde{\alpha}_n)^2}\right)^{-1} \sum_{j=1}^n \left(\frac{1}{\mathbb{A}_{j-1}^\top \tilde{\alpha}_n} - \frac{(\Delta_j X)^2}{h_n (\mathbb{A}_{j-1}^\top \tilde{\alpha}_n)^2}\right) \mathbb{A}_{j-1}. \quad (3.4.1)$$
- Drift parameter

  Plug-in LSE:
  $$\hat{\beta}_n := \operatorname*{argmin}_\beta \sum_{j=1}^n \frac{(\Delta_j X - h_n \mathbb{B}_{j-1}^\top \beta)^2}{\mathbb{A}_{j-1}^\top \hat{\alpha}_n}.$$

From simple calculation, the estimators $\tilde{\alpha}_n$ and $\hat{\beta}_n$ are explicitly written as

$$\tilde{\alpha}_n = \frac{1}{h_n}\left(\sum_{j=1}^n \mathbb{A}_{j-1}\mathbb{A}_{j-1}^\top\right)^{-1} \sum_{j=1}^n (\Delta_j X)^2 \mathbb{A}_{j-1}, \quad (3.4.2)$$

$$\hat{\beta}_n = \frac{1}{h_n}\left(\sum_{j=1}^n \frac{\mathbb{B}_{j-1}\mathbb{B}_{j-1}^\top}{\mathbb{A}_{j-1}^\top \hat{\alpha}_n}\right)^{-1} \sum_{j=1}^n \frac{\Delta_j X}{\mathbb{A}_{j-1}^\top \hat{\alpha}_n} \mathbb{B}_{j-1}. \quad (3.4.3)$$

However, $\tilde{\alpha}_n$ is not asymptotically efficient while $\hat{\beta}_n$ is in case where the underlying process is a diffusion process. That is why we additionally consider the one-step estimator $\hat{\alpha}_n$ based on the stepwise GQL:

$$\mathbb{H}_n(\alpha) := -\frac{1}{2}\sum_{j=1}^n \left\{\log(2\pi) + \log(\mathbb{A}_{j-1}^\top \alpha) + \frac{(\Delta_j X)^2}{h_n \mathbb{A}_{j-1}^\top \alpha}\right\}.$$

If $J \equiv 0$, it is easy to see that $\hat{\alpha}_n$ is asymptotic efficient under appropriate regularity conditions.

Following Section 3.3, their modified versions are

$$\tilde{\alpha}_n^{k_n} = \frac{1}{h_n}\left(\sum_{j \notin \hat{\mathcal{J}}_n^{k_n}} \mathbb{A}_{j-1}\mathbb{A}_{j-1}^\top\right)^{-1} \sum_{j \notin \hat{\mathcal{J}}_n^{k_n}} (\Delta_j X)^2 \mathbb{A}_{j-1}, \quad (3.4.4)$$

$$\hat{\beta}_n^{k_n} = \frac{1}{h_n}\left(\sum_{j\notin\hat{\mathcal{J}}_n^{k_n}} \frac{\mathbb{B}_{j-1}\mathbb{B}_{j-1}^\top}{\mathbb{A}_{j-1}^\top\hat{\alpha}_n^{k_n}}\right)^{-1} \sum_{j\notin\hat{\mathcal{J}}_n^{k_n}} \frac{\Delta_j X}{\mathbb{A}_{j-1}^\top\hat{\alpha}_n^{k_n}}\mathbb{B}_{j-1}, \tag{3.4.5}$$

where $\hat{\alpha}_n^{k_n}$ is the modified one-step estimator:

$$\hat{\alpha}_n^{k_n} = \tilde{\alpha}_n^{k_n} - \left(\sum_{j\notin\hat{\mathcal{J}}_n^{k_n}} \frac{\mathbb{A}_{j-1}\mathbb{A}_{j-1}^\top}{(\mathbb{A}_{j-1}^\top\tilde{\alpha}_n^{k_n})^2}\right)^{-1} \sum_{j\notin\hat{\mathcal{J}}_n^{k_n}} \left(\frac{1}{\mathbb{A}_{j-1}^\top\tilde{\alpha}_n^{k_n}} - \frac{(\Delta_j X)^2}{h_n(\mathbb{A}_{j-1}^\top\tilde{\alpha}_n^{k_n})^2}\right)\mathbb{A}_{j-1}. \tag{3.4.6}$$

What is important from these expressions is that we calculate the modified estimators $\tilde{\alpha}_n^{k_n}$, $\hat{\beta}_n^{k_n}$, and $\hat{\alpha}_n^{k_n}$ simply by removing the corresponding indices from the sums without repetitive numerical optimizations, thus reducing the computational time to a large extent. Further, it should also be noted that we may proceed only with $\tilde{\alpha}_n^k$ without the one-step version $\hat{\alpha}_n^k$, if the asymptotically efficient estimator is not the first thing to have and quick-to-compute estimator is more needed.

To obtain our main result, we introduce some assumptions below.

Assumption 3.4.1. *(Sampling design). There exists a positive constant $\delta \in (0,1)$ such that*

$$T_n := nh_n \to \infty, \quad \frac{\log n}{T_n} \vee n^{1+\delta}h_n^{2+\delta}(\log n)^2 \to 0.$$

Assumption 3.4.2. *(Regularity).*

1. *For every $x, y \in \mathbb{R}$, there exists a positive constant C being independent of x and y such that*

$$\left|\sqrt{\mathbb{A}(x)^\top\alpha_0} - \sqrt{\mathbb{A}(y)^\top\alpha_0}\right| + |\mathbb{B}(x) - \mathbb{B}(y)| + |c(x) - c(y)| \le C|x-y|.$$

2. *The function $\mathbb{A}(x)$ and $c(x)$ fulfill the following estimates:*

$$0 < \inf_{x,\alpha}\mathbb{A}(x)^\top\alpha \le \sup_{x,\alpha}\mathbb{A}(x)^\top\alpha < \infty,$$
$$0 < \inf_x |c(x)| \le \sup_x |c(x)| < \infty.$$

3. *The function $\mathbb{A}(x)$ has continuous derivatives satisfying that for every $x \in \mathbb{R}$,*

$$|\partial_x^i\mathbb{A}(x)| \le C(1+|x|^C) \quad (i \in \{1,2\}),$$

where C is a positive constant independent of x.

Assumption 3.4.3. *(Stability).*

1. *There exists a unique invariant probability measure* π_0*, and for any function* $f \in L_1(\pi_0)$*, we have*

$$\frac{1}{T}\int_0^T f(X_t)dt \xrightarrow{p} \int_{\mathbb{R}} f(x)\pi_0(dx), \quad \text{as } T \to \infty.$$

2. *For any* $q > 0$*,*

$$\sup_{t\in\mathbb{R}^+} E[|X_t|^q] < \infty, \quad \int_{\mathbb{R}} |x|^q \pi_0(dx) < \infty.$$

Hereafter for any $x \in \mathbb{R}$, $\lfloor x \rfloor$ denotes the maximum integer which does not exceed x. Recall that the driving noise J can be expressed as

$$J_t = \sum_{i=1}^{N_t} \xi_i,$$

by a Poisson process N and i.i.d random variables (ξ_i) being independent of N. For the variables (ξ_i), we assume the following.

Assumption 3.4.4. *(Jump size).*

1. *We have* $E[|\xi_1|^q] < \infty$ *for any* $q > 0$*.*

2. *There exists a positive deterministic sequence* (a_n) *satisfying that for all* $M > 0$*,*

$$\max_{1\le j\le \lfloor T_n \rfloor} |\xi_j| = O_p(a_n), \quad a_n^3\sqrt{h_n \log n} = o(1),$$
$$P\left(|\xi_1| \le M\left(\sqrt{n}h_n \vee a_n^{\frac{3}{4}}(h_n \log n)^{\frac{1}{8}}\right)\right) = o\left(\frac{1}{T_n}\right).$$

Here are some technical comments about each assumption.

- Assumption 3.4.1 is a bit stronger than the so-called “rapidly increasing design" $nh_n \to \infty$ and $nh_n^2 \to 0$, which is one of standard conditions in the literature of statistical inference for ergodic processes based on high-frequency data; for example, it suffices that there exists a constant $\kappa \in (1/2, 1)$ for which $0 < \liminf_n n^\kappa h_n \le \limsup_n n^\kappa h_n < \infty$. This condition will be required for handling the extreme value of the solution process X, and asymptotically allowing the number of jump-removal operations to exceed the expectation of the number of jump times; for more details, see the comment after Theorem 3.4.8.

- Assumption 3.4.2 ensures the existence of the càdàg solution of (3.1.1), and its Markovian property (cf. [3, Section 6]).
- Under Assumption 3.4.3, by mimicking the proof of [27, Lemma 8] we see that

$$\frac{1}{n}\sum_{j=1}^{n} f_{j-1}(\alpha) \xrightarrow{p} \int_{\mathbb{R}} f(x,\alpha)\pi_0(dx)$$

uniformly in α, for each f which is differentiable with respect to (x,α) and such that each partial derivative is of at most polynomial growth uniformly in α. For an easy-to-check sufficient condition for Assumption 3.4.3, see [37].

- Concerning Assumptions 3.4.4-(2), we note that such a sequence (a_n) does exist. Here is an example.
  - First let us remark that simply taking

$$a_n = T_n^{1/q} \tag{3.4.7}$$

for a sufficiently large $q \geq 6$ is enough for the first two conditions. For any $\epsilon > 0$ we have

$$P\left(a_n^{-1}\max_{1\leq j\leq \lfloor T_n\rfloor}|\xi_j| > \epsilon\right) = 1-(1-P(|\xi_1|>a_n\epsilon))^{\lfloor T_n\rfloor},$$

the right-hand-side tending to 0 if

$$T_n P(|\xi_1| > a_n\epsilon) \to 0.$$

Grant the moment condition Assumption 3.4.4-(2), the last condition holds under (3.4.7). Hence we have $\max_{1\leq j\leq \lfloor T_n\rfloor}|\xi_j| = o_p(a_n)$, hence in particular the first one in Assumption 3.4.4-(2). Furthermore, under Assumption 3.4.1, the second one in Assumption 3.4.4-(2) follows from

$$a_n^3\sqrt{h_n\log n} = \left(\frac{a_n}{T_n^{1/6}}\right)^3\sqrt{nh_n^2\log n} \to 0. \tag{3.4.8}$$

  - For the third one in Assumption 3.4.4-(1), which will be used for detecting jumps, suppose that there exist constants $\epsilon, \delta > 0$ and $C \geq 0$ such that

$$P(|\xi_1| < x) \leq Cx^{8+\epsilon} \qquad \forall x \in [0,\delta).$$

Suppose also that $T_n(nh_n^2)^{4+\frac{\epsilon}{2}} \to 0$, which, in particular, holds in case where $h_n = cn^{-\kappa}$ for some $c > 0$ with $\kappa \in \left(\frac{10+\epsilon}{2(9+\epsilon)}, 1\right)$. Then, (3.4.7) and (3.4.8) lead to

$$\begin{aligned}
&T_n P\left(|\xi_1| \leq \sqrt{n}h_n \vee a_n^{\frac{3}{4}}(h_n \log n)^{\frac{1}{8}}\right) \\
&\lesssim T_n(nh_n^2)^{4+\frac{\epsilon}{2}} \vee T_n\left(a_n^3\sqrt{h_n \log n}\right)^{2+\epsilon/4} \\
&\lesssim T_n(nh_n^2)^{4+\frac{\epsilon}{2}} \vee T_n^{1+3(2+\epsilon/4)/q} h_n^{1+\epsilon/8}(\log n)^{1+\epsilon/8} \to 0,
\end{aligned}$$

upon taking $q \geq 6$ large enough.

To investigate the asymptotic property of our estimators, we introduce the unobserved continuous part of X defined by

$$X_t^{\mathrm{cont}} = X_t - X_0 - \int_0^t c(X_{s-})dJ_s = \int_0^t a(X_s, \alpha_0)dw_t + \int_0^t b(X_s, \beta_0)dt.$$

Let $\check{\alpha}_n$ be an estimator satisfying

$$\sqrt{n}(\check{\alpha}_n - \alpha_0) = O_p(1). \tag{3.4.9}$$

Taking Remark 3.3.1 into consideration, we define its one-step estimator $\tilde{\alpha}_n^{\mathrm{cont}}$ by

$$\hat{\alpha}_n^{\mathrm{cont}} = \check{\alpha}_n - \left(\sum_{j=1}^n \frac{\mathbb{A}_{j-1}\mathbb{A}_{j-1}^\top}{(\mathbb{A}_{j-1}^\top \check{\alpha}_n)^2}\right)^{-1} \sum_{j=1}^n \left(\frac{1}{\mathbb{A}_{j-1}^\top \check{\alpha}_n} - \frac{(\Delta_j X^{\mathrm{cont}})^2}{h_n(\mathbb{A}_{j-1}^\top \check{\alpha}_n)^2}\right)\mathbb{A}_{j-1}.$$

We also define a plug-in LSE by

$$\hat{\beta}_n^{\mathrm{cont}} := \frac{1}{h_n}\left(\sum_{j=1}^n \frac{\mathbb{B}_{j-1}\mathbb{B}_{j-1}^\top}{\mathbb{A}_{j-1}^\top \hat{\alpha}_n^{\mathrm{cont}}}\right)^{-1} \sum_{j=1}^n \frac{\Delta_j X^{\mathrm{cont}}}{\mathbb{A}_{j-1}^\top \hat{\alpha}_n^{\mathrm{cont}}}\mathbb{B}_{j-1}.$$

The following theorem shows that $(\hat{\alpha}_n^{\mathrm{cont}}, \hat{\beta}_n^{\mathrm{cont}})$ has asymptotic normality, and that $\hat{\beta}_n^{\mathrm{cont}}$ achieves the asymptotic efficiency.

Theorem 3.4.5. *Suppose that Assumptions 3.4.1-3.4.3, and Assumption 3.4.4-(1) hold. Then we have*

$$\left(\sqrt{n}(\hat{\alpha}_n^{\mathrm{cont}} - \alpha_0), \sqrt{T_n}(\hat{\beta}_n^{\mathrm{cont}} - \beta_0)\right) \xrightarrow{\mathcal{L}} N(0, \Sigma_0),$$

where

$$\Sigma_0 := \begin{pmatrix} 2\left\{\int \left(\frac{\mathbb{A}(x)}{(\mathbb{A}(x))^\top \alpha_0}\right)^{\otimes 2} \pi_0(dx)\right\}^{-1} & O \\ O & \left\{\int \frac{\mathbb{B}^{\otimes 2}(x)}{\mathbb{A}(x)^\top \alpha_0}\pi_0(dx)\right\}^{-1} \end{pmatrix}.$$

Remark 3.4.6. *The asymptotic efficiency of* $\hat{\beta}^{\mathrm{cont}}$ *follows from [28, Theorem 2.2]. Concerning the diffusion parameter, we note that* $\hat{\alpha}^{\mathrm{cont}}$ *has the same performance as the estimator in [57] and [48] based on a jump filter.*

The next theorem states that the modified LSE type diffusion estimator $\tilde{\alpha}_n^{k_n}$ has the $\sqrt{n}$-consistency as long as the number of jumps does not exceeds that of the number of jump removals.

Theorem 3.4.7. *Suppose that Assumptions 3.4.1-3.4.4 hold. Then, for any* $\epsilon > 0$, *and non-decreasing deterministic sequence* $(k_n) \subset \mathbb{N}$ *satisfying that*

$$k_n = o\left(\frac{\sqrt{n}}{\log n}\right), \tag{3.4.10}$$

we can find a sufficiently large $M > 0$ *and* $N \in \mathbb{N}$ *for which*

$$\sup_{n \geq N} P\left(\left\{|\sqrt{n}(\tilde{\alpha}_n^{k_n} - \alpha_0)| > M\right\} \cap \{1 \leq N_{T_n} \leq k_n\}\right) < \epsilon. \tag{3.4.11}$$

For convenience, we redefine $\tilde{\alpha}_n^{k_n}$ as

$$\tilde{\alpha}_n^{k_n} = \begin{cases} \tilde{\alpha}_n^{k_n} & \text{on} \quad \{1 \leq N_{T_n} \leq k_n\}, \\ \alpha_0 & \text{on} \quad \{1 \leq N_{T_n} \leq k_n\}^c. \end{cases} \tag{3.4.12}$$

Then obviously $\tilde{\alpha}_n^{k_n}$ satisfies (3.4.9), and we can apply the result of Theorem 3.4.5 to the estimators

$$\hat{\alpha}_n^{k_n,\mathrm{cont}} := \tilde{\alpha}_n^{k_n} - \left(\sum_{j=1}^n \frac{\mathbb{A}_{j-1}\mathbb{A}_{j-1}^\top}{(\mathbb{A}_{j-1}^\top \tilde{\alpha}_n^{k_n})^2}\right)^{-1} \sum_{j=1}^n \left(\frac{1}{\mathbb{A}_{j-1}^\top \tilde{\alpha}_n^{k_n}} - \frac{(\Delta_j X^{\mathrm{cont}})^2}{h_n(\mathbb{A}_{j-1}^\top \tilde{\alpha}_n^{k_n})^2}\right)\mathbb{A}_{j-1},$$

$$\hat{\beta}_n^{k_n,\mathrm{cont}} := \frac{1}{h_n}\left(\sum_{j=1}^n \frac{\mathbb{B}_{j-1}\mathbb{B}_{j-1}^\top}{\mathbb{A}_{j-1}^\top \hat{\alpha}_n^{k_n}}\right)^{-1} \sum_{j=1}^n \frac{\Delta_j X^{\mathrm{cont}}}{\mathbb{A}_{j-1}^\top \hat{\alpha}_n^{k_n}} \mathbb{B}_{j-1}.$$

Recall that we finish our procedure once we have $\mathrm{JB}_n^{k_n} \leq \chi_q^2(2)$. The following theorem gives the asymptotic equivalence between the estimator $(\hat{\alpha}_n^{k_n,\mathrm{cont}}, \hat{\beta}_n^{k_n,\mathrm{cont}})$, and the modified estimator $(\hat{\alpha}_n^{k_n}, \hat{\beta}_n^{k_n})$ on the set $\left\{\mathrm{JB}_n^{k_n} \leq \chi_q^2(2)\right\}$.

Theorem 3.4.8. *Suppose that Assumptions 3.4.1-3.4.4 hold. Then, for any* $\epsilon > 0$, $q \in (0,1)$, *and non-decreasing deterministic sequence* $(k_n) \subset \mathbb{N}$ *fulfilling (3.4.10), we have*

$$P\left(\left\{\left|\sqrt{n}(\hat{\alpha}_n^{k_n} - \hat{\alpha}_n^{k_n,\mathrm{cont}})\right| \vee \left|\sqrt{T_n}(\hat{\beta}_n^{k_n} - \hat{\beta}_n^{k_n,\mathrm{cont}})\right| > \epsilon\right\} \cap \left\{\mathrm{JB}_n^{k_n} \leq \chi_q^2(2)\right\}\right) \to 0. \tag{3.4.13}$$

Remark 3.4.9. *We should note that the number of jump removals is automatically determined by the iterative Jarque-Bera type test, and thus there is no need to choose* (k_n) *in practice.*

Table 3.1: The performance of our estimators is given in case (i). The mean is given with the standard deviation in parenthesis. In this table, $k_n^\star$ denotes the number of jumps.

T_n	n	h_n	$k_n^\star$	(i)Gamma distribution					
				$\hat{\alpha}_n^0$	$\hat{\beta}_n^0$	$\hat{\alpha}_n^{k_n}$	$\hat{\beta}_n^{k_n}$	$\hat{\alpha}_n^{k_n^\star}$	$\hat{\beta}_n^{k_n^\star}$
28.8	1000	0.03	15	18.80	0.62	3.38	0.99	3.38	1.00
				(4.31)	(0.13)	(0.20)	(0.09)	(0.20)	(0.09)
62.1	10000	0.006	30	17.7	0.63	3.07	1.00	3.08	1.00
				(2.91)	(0.09)	(0.05)	(0.06)	(0.04)	(0.06)

Table 3.2: The performance of our estimators is given in case (ii). The mean is given with the standard deviation in parenthesis. In this table, $k_n^\star$ denotes the number of jumps.

T_n	n	h_n	$k_n^\star$	(ii)Bilateral inverse Gaussian distribution					
				$\hat{\alpha}_n^0$	$\hat{\beta}_n^0$	$\hat{\alpha}_n^{k_n}$	$\hat{\beta}_n^{k_n}$	$\hat{\alpha}_n^{k_n^\star}$	$\hat{\beta}_n^{k_n^\star}$
28.8	1000	0.03	15	10.83	0.82	3.19	0.99	3.15	1.00
				(3.70)	(0.22)	(0.17)	(0.14)	(0.16)	(0.14)
62.1	10000	0.006	30	10.22	0.82	3.04	1.01	3.04	1.01
				(2.46)	(0.15)	(0.06)	(0.09)	(0.05)	(0.09)

3.5 Numerical experiments

In this section, we conduct Monte Carlo simulation in order to see the performance of our method. First we consider the following statistical model:

$$dX_t = \sqrt{\frac{\alpha}{1+\sin^2 X_t}}dw_t - \beta X_t dt + dJ_t \quad X_0 = 0, \tag{3.5.1}$$

with the true value $\theta_0 := (\alpha_0, \beta_0) = (3, 1)$. As the jump size distributions, we set (i) Gamma distribution $\Gamma(4, 1)$ (one-sided positive jumps) and (ii) bilateral inverse Gaussian distribution $bIG(2, 1, 4, 1)$ (two-sided jumps). The bilateral inverse Gaussian random variable $X \sim bIG(\delta_1, \gamma_1, \delta_2, \gamma_2)$ is defined as the difference of two independent inverse Gaussian random variable $X_1 \sim IG(\delta_1, \gamma_1)$ and $X_2 \sim IG(\delta_2, \gamma_2)$. Here we set number of jumps fixed just for numerical comparison purpose.

Based on independently simulated 1000 sample path, the mean and standard deviation of our estimator $(\hat{\alpha}_n^{k_n}, \hat{\beta}_n^{k_n})$ are tabulated in Table 3.1 and Table 3.2 with the estimators $(\hat{\alpha}_n^0, \hat{\beta}_n^0)$ and $(\hat{\alpha}_n^{k_n^\star}, \hat{\beta}_n^{k_n^\star})$. The first estimator $(\hat{\alpha}_n^0, \hat{\beta}_n^0)$ is constructed by the whole data, and the latter estimator $(\hat{\alpha}_n^{k_n^\star}, \hat{\beta}_n^{k_n^\star})$ is constructed by the true no-jump group.

These tables and figures indicate that:

- In both case, the modified estimators get closer and closer to the true value as jump removals proceed.
- Since the performance of $(\hat{\alpha}_n^{k_n}, \hat{\beta}_n^{k_n})$ and $(\hat{\alpha}_n^{k_n^\star}, \hat{\beta}_n^{k_n^\star})$ is almost the same, the jump detection by our method works well.
- Concerning the drift estimator, the degree of improvement is not large for (ii) relative to (i). It may be due to the two-sided jump structure of $bIG(2,1,4,1)$; thus the amount of improvement is generally expected to be much more significant when the jump distribution is skewed.
- In the estimator $(\hat{\alpha}_n^0, \hat{\beta}_n^0)$, the performance of $\hat{\alpha}_n^0$ is worse than $\hat{\beta}_n^0$. This is because the diffusion estimator is based on the square of the increments $(\Delta_j X)_j$, thus being heavily affected by jumps.
- Overall, the diffusion parameter are overestimated even by $\hat{\alpha}_n^{k_n^\star}$. Taking into consideration the fact that the mean-reverting point of X is 0, the magnitude of the increment should be larger after one jump occurs. Thus, although jumps are correctly picked, such overestimation can be seen.

3.6 Appendix

For abbreviation, we additionally use the following notations:

- $R(x)$ denotes a differentiable matrix-valued function on $\mathbb{R}$ for which there exists a constant $C>0$ such that $|R(x)| + |\partial_x R(x)| \lesssim (1+|x|)^C$, $x \in \mathbb{R}$.
- We often write $a(x,\alpha)$ and $b(x,\beta)$ instead of $\sqrt{(\mathbb{A}(x))^\top \alpha}$ and $(\mathbb{B}(x))^\top \beta$.

Throughout this section, Assumptions 3.4.1 to 3.4.4 are in force. To show our asymptotic result, we prove some fundamental lemmas. Let us recall that

$$X_t^{\mathrm{cont}} := X_t - X_0 - \int_0^t c(X_{s-}) dJ_s = \int_0^t a(X_s, \alpha_0) dw_t + \int_0^t b(X_s, \beta_0) dt.$$

Lemma 3.6.1. *We have*

$$\left(\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left(\frac{1}{\mathbb{A}_{j-1}^{\top}\alpha_0}-\frac{(\Delta_j X^{\mathrm{cont}})^2}{h_n(\mathbb{A}_{j-1}^{\top}\alpha_0)^2}\right)\mathbb{A}_{j-1}, \frac{1}{\sqrt{T_n}}\sum_{j=1}^{n}\frac{\Delta_j X^{\mathrm{cont}}-h_n\mathbb{B}_{j-1}^{\top}\beta_0}{\mathbb{A}_{j-1}^{\top}\alpha_0}\mathbb{B}_{j-1}\right)\xrightarrow{\mathcal{L}} N\left(0,\Sigma\right),$$

where

$$\Sigma := \begin{pmatrix} 2\int\left(\frac{\mathbb{A}(x)}{(\mathbb{A}(x))^{\top}\alpha_0}\right)^{\otimes 2}\pi_0(dx) & O \\ O & \int\frac{\mathbb{B}^{\otimes 2}(x)}{\mathbb{A}(x)^{\top}\alpha_0}\pi_0(dx)\end{pmatrix}.$$

Proof. By the Cramér-Wold device, it is enough to show the case where $p_\alpha = p_\beta = 1$. From the martingale central limit theorem, the desired result follows if we show

$$\frac{1}{\sqrt{n}}\sum_{j=1}^{n}E^{j-1}\left[\left(\frac{1}{a_{j-1}^2}-\frac{(\Delta_j X^{\mathrm{cont}})^2}{h_n a_{j-1}^4}\right)\mathbb{A}_{j-1}\right]\xrightarrow{p}0, \tag{3.6.1}$$

$$\frac{1}{n}\sum_{j=1}^{n}E^{j-1}\left[\left\{\left(\frac{1}{a_{j-1}^2}-\frac{(\Delta_j X^{\mathrm{cont}})^2}{h_n a_{j-1}^4}\right)\mathbb{A}_{j-1}\right\}^2\right]\xrightarrow{p}2\int\left(\frac{\mathbb{A}(x)}{a^2(x,\alpha_0)}\right)^2\pi_0(dx), \tag{3.6.2}$$

$$\frac{1}{n^2}\sum_{j=1}^{n}E^{j-1}\left[\left\{\left(\frac{1}{a_{j-1}^2}-\frac{(\Delta_j X^{\mathrm{cont}})^2}{h_n a_{j-1}^4}\right)\mathbb{A}_{j-1}\right\}^4\right]\xrightarrow{p}0, \tag{3.6.3}$$

$$\frac{1}{\sqrt{T_n}}\sum_{j=1}^{n}E^{j-1}\left[\frac{\Delta_j X^{\mathrm{cont}}-h_n b_{j-1}}{a_{j-1}^2}\mathbb{B}_{j-1}\right]\xrightarrow{p}0, \tag{3.6.4}$$

$$\frac{1}{T_n}\sum_{j=1}^{n}E^{j-1}\left[\left(\frac{\Delta_j X^{\mathrm{cont}}-h_n b_{j-1}}{a_{j-1}^2}\mathbb{B}_{j-1}\right)^2\right]\xrightarrow{p}\int\frac{\mathbb{B}^2(x)}{a^2(x,\alpha_0)}\pi_0(dx), \tag{3.6.5}$$

$$\frac{1}{(T_n)^2}\sum_{j=1}^{n}E^{j-1}\left[\left(\frac{\Delta_j X^{\mathrm{cont}}-h_n b_{j-1}}{a_{j-1}^2}\mathbb{B}_{j-1}\right)^4\right]\xrightarrow{p}0 \tag{3.6.6}$$

$$\frac{1}{n\sqrt{h_n}}\sum_{j=1}^{n}E^{j-1}\left[\left(\frac{1}{a_{j-1}^2}-\frac{(\Delta_j X^{\mathrm{cont}})^2}{h_n a_{j-1}^4}\right)\frac{\Delta_j X^{\mathrm{cont}}-h_n b_{j-1}}{a_{j-1}^2}\mathbb{A}_{j-1}\mathbb{B}_{j-1}\right]\xrightarrow{p}0. \tag{3.6.7}$$

By using the martingale property of the stochastic integral, Jensen's inequality, the Lipschitz continuity of b, and [40, Lemma 4.5], we have

$$E^{j-1}[\Delta_j X^{\mathrm{cont}}] = h_n b_{j-1} + \int_{t_{j-1}}^{t_j}E^{j-1}[b_s - b_{j-1}]ds = h_n b_{j-1} + h_n^{\frac{3}{2}}R_{j-1}. \tag{3.6.8}$$

Itô's formula and Fubini's theorem for conditional expectation yield that

$$E^{j-1}[(\Delta_j X^{\mathrm{cont}})^2]$$

$$= E^{j-1}\left[2\int_{t_{j-1}}^{t_j}(X_s^{\text{cont}} - X_{j-1}^{\text{cont}})dX_s^{\text{cont}} + \int_{t_{j-1}}^{t_j}(a_s^2 - a_{j-1}^2)ds + a_{j-1}^2 h_n\right]$$
$$= a_{j-1}^2 h_n + 2\int_{t_{j-1}}^{t_j}\left(\int_{t_{j-1}}^{s} E^{j-1}\left[b_u b_s\right]du\right)ds + \int_{t_{j-1}}^{t_j} E^{j-1}[a_s^2 - a_{j-1}^2]ds.$$

Again making use of the Lipschitz continuity of $b(x,\beta_0)$ and [40, Lemma 4.5], we get

$$\left|\int_{t_{j-1}}^{t_j}\left(\int_{t_{j-1}}^{s} E^{j-1}\left[b_u b_s\right]du\right)ds\right|$$
$$\lesssim \int_{t_{j-1}}^{t_j}\left(\int_{t_{j-1}}^{s} E^{j-1}\left[1 + |X_u| + |X_s| + |X_u||X_s|\right]ds\right)ds$$
$$\lesssim h_n^2(1 + |X_{j-1}|^2).$$

Since $\partial_x a^2(x,\alpha)$ and $\partial_x^2 a^2(x,\alpha)$ are of at most polynomial growth with respect to x uniformly in α, we can similarly deduce that

$$\left|\int_{t_{j-1}}^{t_j} E^{j-1}[a_s^2 - a_{j-1}^2]ds\right|$$
$$\lesssim \Bigg|\int_{t_{j-1}}^{t_j} E^{j-1}\Bigg[\partial_x a_{j-1}^2(X_s - X_{j-1})$$
$$+ \frac{1}{2}\int_0^1\int_0^1 \partial_x^2 a^2(X_{j-1} + uv(X_s - X_{j-1}),\alpha_0)dudv(X_s - X_{j-1})^2\Bigg]ds\Bigg|$$
$$\lesssim \left|\int_{t_{j-1}}^{t_j}\left(\int_{t_{j-1}}^{s} E^{j-1}[b_u]du\right)ds\partial_x a_{j-1}^2\right|$$
$$+ \int_{t_{j-1}}^{t_j} E^{j-1}\left[\left(1 + |X_{j-1}|^C + |X_s - X_{j-1}|^C\right)(X_s - X_{j-1})^2\right]ds$$
$$\lesssim h_n^2 R_{j-1}.$$

Hence

$$E^{j-1}[(\Delta_j X^{\text{cont}})^2] = h_n a_{j-1}^2 + h_n^2 R_{j-1}. \tag{3.6.9}$$

For any $q \geq 2$, Burkholder's inequality for conditional expectation gives

$$E^{j-1}\left[|\Delta_j X^{\text{cont}}|^q\right] \lesssim h_n^{\frac{q}{2}} R_{j-1}. \tag{3.6.10}$$

Repeatedly using Itô's formula and (3.6.10), we have

$$E^{j-1}[(\Delta_j X^{\text{cont}})^4]$$
$$= E^{j-1}\left[4\int_{t_{j-1}}^{t_j}(X_s^{\text{cont}} - X_{j-1}^{\text{cont}})^3 dX_s^{\text{cont}} + 6\int_{t_{j-1}}^{t_j}(X_s^{\text{cont}} - X_{j-1}^{\text{cont}})^2 a_s^2 ds\right]$$

$$
\begin{aligned}
&= 6E^{j-1}\Bigg[\int_{t_{j-1}}^{t_j}\left\{2\int_{t_{j-1}}^{s}(X_u^{\text{cont}} - X_{j-1}^{\text{cont}})dX_u^{\text{cont}} + \int_{t_{j-1}}^{s}(a_u^2 - a_{j-1}^2)du + a_{j-1}^2 s\right\} ds a_{j-1}^2 \\
&\quad + \int_{t_{j-1}}^{t_j}\left\{2\int_{t_{j-1}}^{s}(X_u^{\text{cont}} - X_{j-1}^{\text{cont}})dX_u^{\text{cont}} + \int_{t_{j-1}}^{s}(a_u^2 - a_{j-1}^2)du + a_{j-1}^2 s\right\}(a_s^2 - a_{j-1}^2)ds\Bigg] \\
&\quad + h_n^{\frac{5}{2}} R_{j-1} \\
&= 3h_n^2 a_{j-1}^4 + h_n^{\frac{5}{2}} R_{j-1}.
\end{aligned}
\tag{3.6.11}
$$

In particular, it follows from (3.6.8), (3.6.9), (3.6.11) that

$$
E^{j-1}\left[\left\{\left(\frac{1}{a_{j-1}^2} - \frac{(\Delta_j X^{\text{cont}})^2}{h_n a_{j-1}^4}\right)\mathbb{A}_{j-1}\right\}^2\right] = 2\frac{\mathbb{A}_{j-1}^2}{a_{j-1}^4} + h_n^{\frac{1}{2}} R_{j-1}. \tag{3.6.12}
$$

Now, the convergences (3.6.1)-(3.6.7) follow from the expressions (3.6.8)-(3.6.10) and the ergodic theorem

$$
\frac{1}{n}\sum_{j=1}^{n}\zeta(X_{j-1}) \xrightarrow{p} \int \zeta(x)\pi_0(dx),
$$

for each π_0-integrable function ζ. Thus we obtain the desired result. $\square$

Proof of Theorem 3.4.5. Applying Taylor's expansion, we have

$$
\frac{1}{n}\sum_{j=1}^{n}\frac{\mathbb{A}_{j-1}\mathbb{A}_{j-1}^\top}{(\mathbb{A}_{j-1}^\top\check{\alpha}_n)^2} = \frac{1}{n}\sum_{j=1}^{n}\frac{\mathbb{A}_{j-1}\mathbb{A}_{j-1}^\top}{(\mathbb{A}_{j-1}^\top\alpha_0)^2} + \left\{\frac{1}{n}\int_0^1\sum_{j=1}^{n}\partial_\alpha\left(\frac{\mathbb{A}_{j-1}\mathbb{A}_{j-1}^\top}{\left[\mathbb{A}_{j-1}^\top(\alpha_0 + u(\check{\alpha}_n - \alpha_0))\right]^2}\right)du\right\}[\check{\alpha}_n - \alpha_0].
$$

The ergodic theorem implies that the first term of the right-hand-side converges to $\int\left(\frac{\mathbb{A}(x)}{\mathbb{A}(x)^\top\alpha_0}\right)^{\otimes 2}\pi_0(dx)$ in probability. From Assumption 3.4.2 and $\sqrt{n}(\check{\alpha}_n - \alpha_0) = O_p(1)$, the second term of the right-hand-side is $o_p(1)$. We also have

$$
\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left(\frac{1}{\mathbb{A}_{j-1}^\top\check{\alpha}_n} - \frac{(\Delta_j X^{\text{cont}})^2}{h_n(\mathbb{A}_{j-1}^\top\check{\alpha}_n)^2}\right)\mathbb{A}_{j-1} \\
&= \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left(\frac{1}{\mathbb{A}_{j-1}^\top\alpha_0} - \frac{(\Delta_j X^{\text{cont}})^2}{h_n(\mathbb{A}_{j-1}^\top\alpha_0)^2}\right)\mathbb{A}_{j-1} \\
&\quad + \left\{\frac{1}{n}\sum_{j=1}^{n}\left(-\frac{1}{(\mathbb{A}_{j-1}^\top\alpha_0)^2} + 2\frac{(\Delta_j X^{\text{cont}})^2}{h_n(\mathbb{A}_{j-1}^\top\alpha_0)^3}\right)\mathbb{A}_{j-1}\mathbb{A}_{j-1}^\top\right\}[\sqrt{n}(\check{\alpha}_n - \alpha_0)] + o_p(1).
\end{aligned}
$$

By (3.6.9), (3.6.11), and [19, Lemma 9], it follows that

$$\frac{1}{n}\sum_{j=1}^{n}\left(-\frac{1}{(\mathbb{A}_{j-1}^{\top}\alpha_0)^2}+2\frac{(\Delta_j X^{\mathrm{cont}})^2}{h_n(\mathbb{A}_{j-1}^{\top}\alpha_0)^3}\right)\mathbb{A}_{j-1}\mathbb{A}_{j-1}^{\top}\xrightarrow{p}\int\left(\frac{\mathbb{A}(x)}{(\mathbb{A}(x))^{\top}\alpha_0}\right)^{\otimes 2}\pi_0(dx).$$

Hence we obtain

$$\sqrt{n}(\hat{\alpha}^{\mathrm{cont}}-\alpha_0)=-\left\{\int\left(\frac{\mathbb{A}(x)}{(\mathbb{A}(x))^{\top}\alpha_0}\right)^{\otimes 2}\pi_0(dx)\right\}^{-1}\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left(\frac{1}{\mathbb{A}_{j-1}^{\top}\alpha_0}-\frac{(\Delta_j X^{\mathrm{cont}})^2}{h_n(\mathbb{A}_{j-1}^{\top}\alpha_0)^2}\right)\mathbb{A}_{j-1}+o_p(1).$$

In the same way, we have

$$\sqrt{T_n}(\hat{\beta}_n^{\mathrm{cont}}-\beta_0)=\left(\int\frac{\mathbb{B}^{\otimes 2}(x)}{\mathbb{A}(x)^{\top}\alpha_0}\pi_0(dx)\right)^{-1}\frac{1}{\sqrt{T_n}}\sum_{j=1}^{n}\frac{\Delta_j X^{\mathrm{cont}}-h_n b_{j-1}}{\mathbb{A}_{j-1}^{\top}\alpha_0}\mathbb{B}_{j-1}+o_p(1),$$

and the desired result follows from Slutsky's lemma and Lemma 3.6.1. □

We now turn to proving Theorems 3.4.7 and 3.4.8.

Lemma 3.6.2. *Let $\{\tau_i\}_{i\in\mathbb{N}}$ denote jump times of N. Then we have*

$$P\left({}^{\exists}i\in\mathbb{N},{}^{\exists}j\in\{1,\dots,n\}\ \ s.t.\ \ \tau_i,\tau_{i+1}\in[t_{j-1},t_j)\right)\to 0,\qquad n\to\infty.$$

Proof. Since the random variables $\tau_1,\tau_2-\tau_1,\tau_3-\tau_2,\dots$ independently obey the exponential distribution with mean $1/\lambda$, it follows that

$$\begin{aligned}P\left({}^{\exists}i,j\ \ \text{s.t.}\ \ \tau_i,\tau_{i+1}\in[t_{j-1},t_j)\right)&\le\sum_{i=2}^{\infty}P\left({}^{\exists}j\in\{2,\dots,i\}\ \ \text{s.t.}\ \ \tau_j-\tau_{j-1}<h_n\right)P(N_{T_n}=i)\\&\le(1-e^{-\lambda h_n})\sum_{i=2}^{\infty}\frac{(\lambda T_n)^i}{(i-1)!}e^{-\lambda T_n}\lesssim nh_n^2\to 0.\end{aligned}$$

□

For convenience, we hereafter write

$$B_n=\left\{{}^{\exists}i\in\mathbb{N},{}^{\exists}j\in\{1,\dots,n\}\ \ \text{s.t.}\ \ \tau_i,\tau_{i+1}\in[t_{j-1},t_j)\right\}^c.$$

By Lemma 3.6.2 we have $P(B_n)\to 1$.

Let

$$\begin{aligned}C_{k_n,n}&:=\left\{{}^{\exists}i\in\mathbb{N},\ {}^{\exists}j\in\{1,\dots,n\}\ \ \text{s.t.}\ \ \tau_i\in[t_{j-1},t_j)\ \ \text{and}\ \ j\notin\hat{\mathcal{J}}_n^{k_n}\right\}^c\\&=\left\{{}^{\forall}i\in\mathbb{N},\ {}^{\forall}j\in\{1,\dots,n\},\ \ \tau_i\notin[t_{j-1},t_j)\ \ \text{or}\ \ j\in\hat{\mathcal{J}}_n^{k_n}\right\}.\end{aligned}$$

The next lemma shows an asymptotic negligibility of the failure-to-detection rate: we can correctly detect all jumps on $\{1\le N_{T_n}\le k_n\}\cap B_n$ with probability tending to 1:

Lemma 3.6.3. *We have*

$$P\Big(C_{k_n,n}^c \cap \{1 \le N_{T_n} \le k_n\} \cap B_n\Big) \to 0, \qquad n \to \infty.$$

Proof. Hereafter we use the following notations:

$$\begin{aligned} \mathcal{D}_n &= \{j \le n : {}^{\exists}\tau_i \in [t_{j-1}, t_j)\}, \\ \mathcal{C}_n &= \{1, \dots, n\} \setminus \mathcal{D}_n. \end{aligned}$$

Write $\eta_j = \frac{\Delta_j w}{\sqrt{h_n}}$ for $j \le n$. Recalling that the set $\hat{\mathcal{J}}_n^{k_n}$ of estimated jump times is constructed through picking up the first k_n-largest increments in magnitude, we have

$$\begin{aligned}
&P\Big(C_{k_n,n}^c \cap \{1 \le N_{T_n} \le k_n\} \cap B_n\Big) \\
&\le P(\{{}^{\exists}j' \in \mathcal{D}_n, j'' \in \mathcal{C}_n \ \text{ s.t. } \ |\Delta_{j'}X| < |\Delta_{j''}X|\} \cap \{1 \le N_{T_n} \le k_n\} \cap B_n) \\
&\le P\Bigg(\Bigg\{{}^{\exists}j' \in \mathcal{D}_n, j'' \in \mathcal{C}_n \ \text{ s.t. } \ \inf_x |c(x)| \min_{1 \le j \le N_{T_n}} |\xi_j| \\
&\qquad < \left|\int_{t_{j'-1}}^{t_{j'}} b_s ds + \int_{t_{j'-1}}^{t_{j'}} a_s dw_s\right| + \left|\int_{t_{j''-1}}^{t_{j''}} b_s ds + \int_{t_{j''-1}}^{t_{j''}} a_s dw_s\right|\Bigg\} \cap \{1 \le N_{T_n} \le k_n\} \cap B_n\Bigg) \\
&\le P\Bigg(\Bigg\{\inf_x |c(x)| \min_{1 \le j \le N_{T_n}} |\xi_j| < 2\sqrt{h_n} \sup_x |a(x)| \max_{1 \le j \le n} |\eta_j| \\
&\qquad + 2 \max_{1 \le j \le n} \left(\left|\int_{t_{j-1}}^{t_j} b_s ds\right| + \left|\int_{t_{j-1}}^{t_j} (a_s - a_{j-1}) dw_s\right|\right)\Bigg\} \cap \{1 \le N_{T_n} \le k_n\} \cap B_n\Bigg) \\
&\le P\left(\left\{\inf_x |c(x)|^2 \min_{1 \le j \le N_{T_n}} |\xi_j|^2 < r_{1,n} + r_{2,n}\right\} \cap \{1 \le N_{T_n} \le k_n\} \cap B_n\right), \qquad (3.6.13)
\end{aligned}$$

where

$$\begin{aligned} r_{1,n} &:= 4h_n \sup_x a^2(x) \max_{1 \le j \le n} |\eta_j|^2 \\ r_{2,n} &:= 8 \sum_{j=1}^n \left\{\left(\int_{t_{j-1}}^{t_j} b_s ds\right)^2 + \left(\int_{t_{j-1}}^{t_j} (a_s - a_{j-1}) dw_s\right)^2\right\} \end{aligned}$$

From extreme value theory (cf. [16, Table 3.4.4]), we have

$$\max_{1 \le j \le n} |\eta_i|^2 - \left(\log n - \frac{1}{2} \log\log n - \log\Gamma\left(\frac{1}{2}\right)\right) = O_p(1).$$

This together with Assumption 3.4.2 leads to

$$r_{1,n} = O_p(h_n \log n) = O_p(nh_n^2).$$

Jensen's inequality leads to

$$E\left[\left(\int_{t_{j-1}}^{t_j} b_s ds\right)^2\right] \leq h_n \int_{t_{j-1}}^{t_j} E[b_s^2] ds \lesssim h_n^2.$$

Applying Burkholder's inequality with [40, Lemma 4.5], we get

$$E\left[\left(\int_{t_{j-1}}^{t_j} (a_s - a_{j-1}) dw_s\right)^2\right] \lesssim \int_{t_{j-1}}^{t_j} E[|X_s - X_{j-1}|^2] ds \lesssim h_n^2,$$

so that

$$r_{2,n} = O_p(nh_n^2).$$

Hence, for any $\epsilon \in (0,1)$, we can pick sufficiently large N and K such that for all $n \geq N$,

$$P\left(r_{1,n} + r_{2,n} > Knh_n^2\right) < \epsilon.$$

It follows from these estimates and $E[N_{T_n}] = \lambda T_n$ that for every n large enough, the upper bound in (3.6.13) can be further bounded by

$$\begin{aligned}
&P\left(\left\{\min_{1\leq j\leq N_{T_n}} |\xi_j|^2 < \frac{K}{\inf_x |c(x)|^2} nh_n^2\right\} \cap \{1 \leq N_{T_n} \leq k_n\} \cap B_n\right) + \epsilon \\
&\leq \sum_{i=1}^{k_n} P\left(\left\{\min_{1\leq j\leq i} |\xi_j|^2 < \frac{K}{\inf_x |c(x)|^2} nh_n^2\right\} \cap \{N_{T_n} = i\}\right) + \epsilon \\
&\leq \sum_{i=1}^{k_n} iP\left(|\xi_1|^2 < \frac{K}{\inf_x |c(x)|^2} nh_n^2\right) P\left(N_{T_n} = i\right) + \epsilon \\
&\lesssim T_n P\left(|\xi_1|^2 < \frac{K}{\inf_x |c(x)|^2} nh_n^2\right) + \epsilon = o(1) + \epsilon. \qquad (3.6.14)
\end{aligned}$$

Since the choice of ϵ is arbitrary, Assumption 3.4.4 implies the desired result. □

Next, we show that on $\{N_{T_n} \geq k_n + 1\} \cap B_n$, indices which belongs to $\mathcal{C}_n$ are asymptotically outside $\hat{\mathcal{J}}_n^{k_n}$:

Lemma 3.6.4.

$$P\left(\{\mathcal{C}_n \cap \hat{\mathcal{J}}_n^{k_n} \neq \emptyset\} \cap \{N_{T_n} \geq k_n + 1\} \cap B_n\right) \to 0, \quad n \to \infty.$$

Proof. The lemma can be shown in a quite similar way to Lemma 3.6.3. Note that

$$\{\mathcal{C}_n \cap \hat{\mathcal{J}}_n^{k_n} \neq \emptyset\} \subset \left\{ {}^{\exists} j' \in \mathcal{D}_n, j'' \in \mathcal{C}_n \ \text{ s.t. } \ |\Delta_{j'} X| < |\Delta_{j''} X| \right\}.$$

Letting ϵ, N, and K be the same as in the proof of Lemma 3.6.3, as in (3.6.13) and (3.6.14) we have

$$\begin{aligned}
&P\left(\{\mathcal{C}_n \cap \hat{\mathcal{J}}_n^{k_n} \neq \emptyset\} \cap \{N_{T_n} \geq k_n + 1\} \cap B_n\right) \\
&\leq P\left(\left\{\min_{1\leq j \leq N_{T_n}} |\xi_j|^2 < \frac{K}{\inf_x |c(x)|^2} n h_n^2\right\} \cap \{N_{T_n} \geq k_n + 1\} \cap B_n\right) \\
&\lesssim T_n P\left(|\xi_1|^2 < \frac{K}{\inf_x |c(x)|^2} n h_n^2\right) + \epsilon = o(1) + \epsilon
\end{aligned}$$

for any $n \geq N$. Hence the proof is complete. □

Proof of Theorem 3.4.7. Thanks to Lemma 3.6.2 and Lemma 3.6.3, it suffices to show that for any $\epsilon > 0$ there correspond sufficiently large $M > 0$ and $N \in \mathbb{N}$ for which

$$\sup_{n \geq N} P\left(\left\{|\sqrt{n}(\tilde{\alpha}_n^{k_n} - \alpha_0)| > M\right\} \cap G_{k_n,n}\right) < \epsilon, \tag{3.6.15}$$

where

$$G_{k_n,n} := \{1 \leq N_{T_n} \leq k_n\} \cap B_n \cap C_{k_n,n}.$$

For any $q \geq 2$, Jensen's inequality, Burkholder's inequality and Assumption 3.4.2 imply that

$$\begin{aligned}
E[|\Delta_j X^{\text{cont}}|^q] &\lesssim E\left[\left(\int_{t_{j-1}}^{t_j} a_s^2 ds\right)^{\frac{q}{2}} + h_n^{q-1} \int_{t_{j-1}}^{t_j} |b_s|^q ds\right] \\
&\lesssim h_n^{\frac{q}{2}} \sup_t E[1 + |X_t|^q] = h_n^{\frac{q}{2}}.
\end{aligned} \tag{3.6.16}$$

Since $\Delta_j X = \Delta_j X^{\text{cont}}$ for each $j \notin \hat{\mathcal{J}}_n^{k_n}$ on $G_{k_n,n}$, we have

$$|\tilde{\alpha}_n^{k_n} - \alpha_0| \mathbb{1}_{G_{k_n,n}} \leq (|\kappa_{1,n}| + |\kappa_{2,n}| + |\kappa_{3,n}|) \mathbb{1}_{G_{k_n,n}}, \tag{3.6.17}$$

where

$$\kappa_{1,n} := \frac{1}{h_n} \left\{ \left(\sum_{j \notin \hat{\mathcal{J}}_n^{k_n}} \mathbb{A}_{j-1} \mathbb{A}_{j-1}^\top \right)^{-1} - \left(\sum_{j=1}^{n} \mathbb{A}_{j-1} \mathbb{A}_{j-1}^\top \right)^{-1} \right\} \sum_{j \notin \hat{\mathcal{J}}_n^{k_n}} \mathbb{A}_{j-1} (\Delta_j X^{\text{cont}})^2,$$

$$\kappa_{2,n} := -\frac{1}{h_n}\left(\sum_{j=1}^{n}\mathbb{A}_{j-1}\mathbb{A}_{j-1}^{\top}\right)^{-1}\sum_{j\in\hat{\mathcal{J}}_n^{k_n}}\mathbb{A}_{j-1}(\Delta_j X^{\mathrm{cont}})^2,$$
$$\kappa_{3,n} := \frac{1}{h_n}\left(\sum_{j=1}^{n}\mathbb{A}_{j-1}\mathbb{A}_{j-1}^{\top}\right)^{-1}\left\{\sum_{j=1}^{n}\mathbb{A}_{j-1}(\Delta_j X^{\mathrm{cont}})^2 - h_n\left(\sum_{j=1}^{n}\mathbb{A}_{j-1}\mathbb{A}_{j-1}^{\top}\right)\alpha_0\right\}.$$

Below we look at these three terms separately.

1. Evaluation of $\kappa_{1,n}$: From the ergodic theorem, we have

$$\frac{1}{n}\sum_{j=1}^{n}\mathbb{A}_{j-1}\mathbb{A}_{j-1}^{\top}\xrightarrow{p}\int\mathbb{A}(x)(\mathbb{A}(x))^{\top}\pi_0(dx) > 0.$$

Hence we can suppose that the inverse matrix of $\frac{1}{n}\sum_{j=1}^{n}\mathbb{A}_{j-1}\mathbb{A}_{j-1}^{\top}$ exists for large enough n. Since $\mathbb{A}(x)$ is bounded, we can also obtain

$$\frac{1}{n}\sum_{j\notin\hat{\mathcal{J}}_n^{k_n}}\mathbb{A}_{j-1}\mathbb{A}_{j-1}^{\top} = \frac{1}{n}\sum_{j=1}^{n}\mathbb{A}_{j-1}\mathbb{A}_{j-1}^{\top} - \frac{1}{n}\sum_{j\in\hat{\mathcal{J}}_n^{k_n}}\mathbb{A}_{j-1}\mathbb{A}_{j-1}^{\top} = \frac{1}{n}\sum_{j=1}^{n}\mathbb{A}_{j-1}\mathbb{A}_{j-1}^{\top} + O_p\left(\frac{k_n}{n}\right),$$
$$\left|\frac{1}{T_n}\sum_{j\notin\hat{\mathcal{J}}_n^{k_n}}\mathbb{A}_{j-1}(\Delta_j X^{\mathrm{cont}})^2\right| \lesssim \frac{1}{T_n}\sum_{j=1}^{n}(\Delta_j X^{\mathrm{cont}})^2 = O_p(1),$$

from (3.6.16). Hence it follows that

$$\begin{aligned}
&|\sqrt{n}\kappa_{1,n}|\mathbb{1}_{G_{k_n,n}} \\
&\lesssim \left(\frac{1}{n}\sum_{j=1}^{n}\mathbb{A}_{j-1}\mathbb{A}_{j-1}^{\top}\right)^{-1}\left|\sqrt{n}\left\{\left(\frac{1}{n}\sum_{j=1}^{n}\mathbb{A}_{j-1}\mathbb{A}_{j-1}^{\top}\right)\left(\frac{1}{n}\sum_{j\notin\hat{\mathcal{J}}_n^{k_n}}\mathbb{A}_{j-1}\mathbb{A}_{j-1}^{\top}\right)^{-1} - I_{p_\alpha}\right\}\right| \\
&\qquad\times\left(\frac{1}{T_n}\sum_{j=1}^{n}(\Delta_j X^{\mathrm{cont}})^2\right) \\
&= O_p\left(\frac{k_n}{\sqrt{n}}\right) = o_p(1).
\end{aligned} \tag{3.6.18}$$

2. Evaluation of $\kappa_{2,n}$: Recall that $\eta_j := \frac{\Delta_j w}{\sqrt{h_n}}$. Under Assumption 3.4.2, we can derive from the estimates of $r_{1,n}$ and $r_{2,n}$ in the proof of Lemma 3.6.3 that, on $G_{k_n,n}$,

$$\left|\frac{1}{T_n}\sum_{j\in\hat{\mathcal{J}}_n^{k_n}}\mathbb{A}_{j-1}(\Delta_j X^{\mathrm{cont}})^2\right|$$

$$\lesssim \frac{1}{T_n}\left\{\sum_{j=1}^{n}\left[\left(\int_{t_{j-1}}^{t_j}(a_s-a_{j-1})dw_s\right)^2+\left(\int_{t_{j-1}}^{t_j}b_s ds\right)^2\right]+k_nh_n\max_{1\le j\le n}|\eta_j|^2\right\}$$
$$=O_p\left(\frac{1}{T_n}(nh_n^2\vee k_nh_n\log n)\right)=O_p\left(h_n\vee\frac{k_n\log n}{n}\right)=o_p(1). \tag{3.6.19}$$

Thus we get

$$|\sqrt{n}\kappa_{2,n}|\mathbb{1}_{G_{k_n,n}}=O_p\left(\sqrt{nh_n^2}\vee\frac{k_n\log n}{\sqrt{n}}\right)=o_p(1). \tag{3.6.20}$$

3. Evaluation of $\kappa_{3,n}$: From (3.6.9), (3.6.11), (3.6.10), and the martingale central limit theorem (see the proof of Lemma 3.6.1), it follows that

$$\sqrt{n}\kappa_{3,n}=\left(\frac{1}{n}\sum_{j=1}^{n}\mathbb{A}_{j-1}\mathbb{A}_{j-1}^\top\right)^{-1}\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\mathbb{A}_{j-1}\left\{\left(\frac{\Delta_jX^{\mathrm{cont}}}{\sqrt{h_n}}\right)^2-\mathbb{A}_{j-1}^\top\alpha_0\right\}=O_p(1). \tag{3.6.21}$$

Substituting (3.6.18), (3.6.20) and (3.6.21) into (3.6.17) now yields that

$$\left|\sqrt{n}\left(\tilde{\alpha}_n^{k_n}-\alpha_0\right)\right|\mathbb{1}_{G_{k_n,n}}=O_p(1),$$

followed by (3.6.15). □

Proof of Theorem 3.4.8. By the Lindeberg-Feller theorem we have

$$\frac{N_{T_n}-\lambda T_n}{\sqrt{\lambda T_n}}\xrightarrow{\mathcal{L}}N(0,1),$$

so that for any positive nondecreasing sequence (l_n) satisfying $\frac{l_n-\lambda T_n}{\sqrt{\lambda T_n}}\to\infty$, we have

$$P\left(N_{T_n}\ge l_n\right)=P\left(\frac{N_{T_n}-\lambda T_n}{\sqrt{\lambda T_n}}\ge\frac{l_n-\lambda T_n}{\sqrt{\lambda T_n}}\right)\to 0; \tag{3.6.22}$$

in particular, this implies that we may focus on the case where $N_{T_n}\le(\lambda+1)T_n-1$ and $k_n\le(\lambda+1)T_n-1$ without loss of generality.

Let

$$D_{k_n,n}:=\left\{\mathcal{C}_n\cap\hat{\mathcal{J}}_n^{k_n}=\emptyset\right\}.$$

From (3.6.22) with $l_n=(\lambda+1)T_n+1$, Lemma 3.6.2, and Lemma 3.6.3, it follows that for any $\epsilon>0$,

$$P\left(\left\{\left|\sqrt{n}(\hat{\alpha}_n^{k_n}-\hat{\alpha}_n^{k_n,\mathrm{cont}})\right|\vee\left|\sqrt{T_n}(\hat{\beta}_n^{k_n}-\hat{\beta}_n^{k_n,\mathrm{cont}})\right|>\epsilon\right\}\cap\left\{\mathrm{JB}_n^{k_n}\le\chi_q^2(2)\right\}\right)$$

$$
\begin{aligned}
&\leq P\left(\left\{\left|\sqrt{n}(\hat{\alpha}_n^{k_n}-\hat{\alpha}_n^{k_n,\mathrm{cont}})\right|\vee\left|\sqrt{T_n}(\hat{\beta}_n^{k_n}-\hat{\beta}_n^{k_n,\mathrm{cont}})\right|>\epsilon\right\}\cap G_{k_n,n}\right)\\
&\qquad+P\left(\left\{\mathrm{JB}_n^{k_n}\leq\chi_q^2(2)\right\}\cap\{k_n+1\leq N_{T_n}\leq(\lambda+1)T_n\}\cap B_n\cap D_{k_n,n}\right)+o(1).
\end{aligned}
$$

We will complete the proof by showing that both of the first two terms in the upper bound vanish as $n\to\infty$.

First we verify

$$
P\left(\left\{\left|\sqrt{n}(\hat{\alpha}_n^{k_n}-\hat{\alpha}_n^{k_n,\mathrm{cont}})\right|\vee\left|\sqrt{T_n}(\hat{\beta}_n^{k_n}-\hat{\beta}_n^{k_n,\mathrm{cont}})\right|>\epsilon\right\}\cap G_{k_n,n}\right)=o(1). \tag{3.6.23}
$$

Recall that for any $j\notin\mathcal{J}_n^{k_n}$, $\Delta_j X=\Delta_j X^{\mathrm{cont}}$ on $G_{k_n,n}$. Making use of Assumption 3.4.2, (3.6.19), and a similar argument to the proof of Theorem 3.4.7, we get

$$
\begin{aligned}
&\left|\sqrt{n}(\hat{\alpha}_n^{k_n}-\hat{\alpha}_n^{k_n,\mathrm{cont}})\mathbb{1}_{G_{k_n,n}}\right|\\
&\leq\left|\left(\frac{1}{n}\sum_{j\notin\hat{\mathcal{J}}_n^{k_n}}\frac{\mathbb{A}_{j-1}\mathbb{A}_{j-1}^\top}{\mathbb{A}_{j-1}^\top\tilde{\alpha}_n^{k_n}}\right)^{-1}\left\{\frac{1}{\sqrt{n}}\sum_{j\in\hat{\mathcal{J}}_n^{k_n}}\left(\frac{1}{\mathbb{A}_{j-1}^\top\tilde{\alpha}_n^{k_n}}-\frac{(\Delta_j X^{\mathrm{cont}})^2}{h_n(\mathbb{A}_{j-1}^\top\tilde{\alpha}_n^{k_n})^2}\right)\mathbb{A}_{j-1}\right\}\right|\mathbb{1}_{G_{k_n,n}}\\
&\quad+\left|\left(\frac{1}{n}\sum_{j=1}^n\frac{\mathbb{A}_{j-1}\mathbb{A}_{j-1}^\top}{\mathbb{A}_{j-1}^\top\tilde{\alpha}_n^{k_n}}\right)^{-1}\right|\left|\left(\frac{1}{n}\sum_{j=1}^n\frac{\mathbb{A}_{j-1}\mathbb{A}_{j-1}^\top}{\mathbb{A}_{j-1}^\top\tilde{\alpha}_n^{k_n}}\right)\left(\frac{1}{n}\sum_{j\notin\hat{\mathcal{J}}_n^{k_n}}\frac{\mathbb{A}_{j-1}\mathbb{A}_{j-1}^\top}{\mathbb{A}_{j-1}^\top\tilde{\alpha}_n^{k_n}}\right)^{-1}-I_{p_\alpha}\right|\\
&\quad\times\left|\frac{1}{\sqrt{n}}\sum_{j=1}^n\left(\frac{1}{\mathbb{A}_{j-1}^\top\tilde{\alpha}_n^{k_n}}-\frac{(\Delta_j X^{\mathrm{cont}})^2}{h_n(\mathbb{A}_{j-1}^\top\tilde{\alpha}_n^{k_n})^2}\right)\mathbb{A}_{j-1}\right|\mathbb{1}_{G_{k_n,n}}\\
&=\left|O_p(1)\cdot\left\{O_p\left(\frac{k_n}{\sqrt{n}}\right)+O_p\left(\sqrt{nh_n^2}\vee\frac{k_n\log n}{\sqrt{n}}\right)\right\}\right|+|O_p(1)\cdot o_p(1)\cdot O_p(1)|\\
&=o_p(1).
\end{aligned}
$$

Let us turn to look at $\sqrt{T_n}(\hat{\beta}_n^{k_n}-\hat{\beta}_n^{k_n,\mathrm{cont}})$. It follows from Itô's formula that

$$
\begin{aligned}
&\left|\frac{1}{n}\sum_{j\in\hat{\mathcal{J}}_n^{k_n}}\frac{\mathbb{B}_{j-1}\mathbb{B}_{j-1}^\top}{\mathbb{A}_{j-1}^\top\hat{\alpha}_n^{k_n}}\right|\mathbb{1}_{G_{k_n,n}}\\
&\lesssim\frac{k_n}{n}\left(1+\sup_{0\leq t\leq T_n}X_t^2\right)\\
&=\frac{k_n}{n}\sup_{0\leq t\leq T_n}\left\{1+X_0^2+2\int_0^t X_{s-}dX_s+\int_0^t a_s^2ds+\sum_{0<s\leq t}(\Delta_s X)^2\right\}\\
&\lesssim\frac{k_n}{n}\left\{1+X_0^2+\int_0^{T_n}\left(a_s^2+|X_sb_s|+c_s^2+|X_sc_s\lambda E[\xi_1]|\right)ds\right.
\end{aligned}
$$

$$+\sup_{0\le t\le T_n}\left|\int_0^t X_s a_s dw_s+\int_0^t\int_{\mathbb{R}}\left(c_{s-}^2 z^2+X_{s-}c_{s-}z\right)\tilde{N}(ds,dz)\right|\Bigg\},\tag{3.6.24}$$

where $\tilde{N}(\cdot,\cdot)$ denotes the compensated Poisson random measure associated with J. Applying Assumption 3.4.3 and Burkholder's inequality to the last term, we get

$$\frac{1}{n}\sum_{j\in\hat{\mathcal{J}}_n^{k_n}}\frac{\mathbb{B}_{j-1}\mathbb{B}_{j-1}^\top}{\mathbb{A}_{j-1}^\top\hat{\alpha}_n^{k_n}}\mathbb{1}_{G_{k_n,n}}=O_p(h_nk_n)=O_p(nh_n^2)=o_p(1),$$

and thus

$$\frac{1}{n}\sum_{j=1}^n\frac{\mathbb{B}_{j-1}\mathbb{B}_{j-1}^\top}{\mathbb{A}_{j-1}^\top\hat{\alpha}_n^{k_n}}\mathbb{1}_{G_{k_n,n}}=\frac{1}{n}\sum_{j\notin\hat{\mathcal{J}}_n^{k_n}}\frac{\mathbb{B}_{j-1}\mathbb{B}_{j-1}^\top}{\mathbb{A}_{j-1}^\top\hat{\alpha}_n^{k_n}}\mathbb{1}_{G_{k_n,n}}+o_p(1).\tag{3.6.25}$$

Below, we show that

$$\begin{aligned}&\left|\frac{1}{\sqrt{T_n}}\sum_{j\in\hat{\mathcal{J}}_n^{k_n}}\frac{\Delta_j X^{\mathrm{cont}}-h_nb_{j-1}}{\mathbb{A}_{j-1}^\top\hat{\alpha}_n^{k_n}}\mathbb{B}_{j-1}\right|\\&=\left|\frac{1}{\sqrt{T_n}}\sum_{j\in\hat{\mathcal{J}}_n^{k_n}}\frac{\int_{t_{j-1}}^{t_j}(b_s-b_{j-1})ds+\int_{t_{j-1}}^{t_j}(a_s-a_{j-1})dw_s+a_{j-1}\Delta_j w}{\mathbb{A}_{j-1}^\top\hat{\alpha}_n^{k_n}}\mathbb{B}_{j-1}\right|=o_p(1).\end{aligned}\tag{3.6.26}$$

Utilizing the Lipschitz continuity of b and [40, Lemma 4.5], we have

$$\begin{aligned}E\left[\left|\frac{1}{\sqrt{T_n}}\sum_{j\in\hat{\mathcal{J}}_n^{k_n}}\frac{\int_{t_{j-1}}^{t_j}(b_s-b_{j-1})ds}{\mathbb{A}_{j-1}^\top\hat{\alpha}_n^{k_n}}\mathbb{B}_{j-1}\right|\right]&\lesssim\frac{1}{\sqrt{T_n}}\sum_{j=1}^n\int_{t_{j-1}}^{t_j}E\left[|(b_s-b_{j-1})\mathbb{B}_{j-1}|\right]ds\\&=O_p\left(\sqrt{nh_n^2}\right)=o_p(1).\end{aligned}$$

From the elementary inequality

$$|x|\le C+\frac{|x|^2}{C},\tag{3.6.27}$$

for any positive constant C and real number x, we get

$$\left|\frac{1}{\sqrt{T_n}}\sum_{j\in\hat{\mathcal{J}}_n^{k_n}}\frac{\int_{t_{j-1}}^{t_j}(a_s-a_{j-1})dw_s}{\mathbb{A}_{j-1}^\top\hat{\alpha}_n^{k_n}}\mathbb{B}_{j-1}\right|$$

$$
\begin{aligned}
&\lesssim \frac{1}{\sqrt{T_n}} \sum_{j \in \hat{\mathcal{J}}_n^{k_n}} \left| \int_{t_{j-1}}^{t_j} (a_s - a_{j-1}) dw_s \mathbb{B}_{j-1} \right| \\
&\lesssim \frac{1}{\sqrt{T_n}} \sum_{j \in \hat{\mathcal{J}}_n^{k_n}} \left\{ \frac{\sqrt{T_n}}{k_n (\log n)^2} + \frac{k_n (\log n)^2}{\sqrt{T_n}} \left| \int_{t_{j-1}}^{t_j} (a_s - a_{j-1}) dw_s \mathbb{B}_{j-1} \right|^2 \right\} \\
&\lesssim \frac{1}{(\log n)^2} + \frac{1}{n} \sum_{j=1}^{n} \left| \frac{1}{h_n} \int_{t_{j-1}}^{t_j} (a_s - a_{j-1}) dw_s \mathbb{B}_{j-1} \right|^2 n h_n^2 (\log n)^2 \\
&\lesssim \frac{1}{(\log n)^2} + O_p\left(n h_n^2 (\log n)^2 \right) = o_p(1).
\end{aligned}
$$

Here we used the condition $k_n \leq (\lambda + 1) T_n - 1$ and Burkholder's inequality. Under Assumption 3.4.3, for any $q > 2$ we have

$$
E\left[\sup_{0 \leq t \leq T_n} |X_t|^q \right] = O\left(T_n \right)
$$

through Itô's formula as in (3.6.24). This combined with Jensen's inequality gives

$$
E\left[\sup_{0 \leq t \leq T_n} |X_t| \right] = O(T_n^{\epsilon}), \tag{3.6.28}
$$

for any $\epsilon > 0$. With $\delta \in (0,1)$ given in Assumption 3.4.1, let $\epsilon = \frac{\delta}{3}$ and $\delta' = \frac{4}{3}\delta$, respectively. Then, making use of (3.6.28) with an application of (3.6.27), we have

$$
\begin{aligned}
&\left| \frac{1}{\sqrt{T_n}} \sum_{j \in \hat{\mathcal{J}}_n^{k_n}} \frac{a_{j-1} \Delta_j w}{\mathbb{A}_{j-1}^{\top} \hat{\alpha}_n^{k_n}} \mathbb{B}_{j-1} \right| \\
&\lesssim \frac{\max_{1 \leq j \leq n} |\mathbb{B}_{j-1}|}{\sqrt{T_n}} \sum_{j \in \hat{\mathcal{J}}_n^{k_n}} \left\{ \frac{T_n^{\frac{1-\delta'}{2}}}{k_n} + \frac{k_n}{T_n^{\frac{1-\delta'}{2}}} (\Delta_j w)^2 \right\} \\
&\lesssim O_p\left(T_n^{-\frac{\delta'}{2} + \epsilon} \vee T_n^{1 + \epsilon + \frac{\delta'}{2}} h_n \log n \right) \\
&= O_p\left(T_n^{-\frac{\delta}{3}} \vee n^{1+\delta} h_n^{2+\delta} \log n \right) = o_p(1),
\end{aligned}
$$

thus concluding (3.6.26). As in the proof of Theorem 3.4.5, it follows from (3.6.25) and (3.6.26) that

$$
\left| \sqrt{T_n} (\hat{\beta}_n^{k_n} - \hat{\beta}_n^{k_n, \mathrm{cont}}) \mathbb{1}_{G_{k_n,n}} \right|
$$

$$
\begin{aligned}
&\leq \left| \left(\frac{1}{n} \sum_{j \notin \hat{\mathcal{J}}_n^{k_n}} \frac{\mathbb{B}_{j-1}\mathbb{B}_{j-1}^\top}{\mathbb{A}_{j-1}^\top \hat{\alpha}_n^{k_n}} \right)^{-1} \frac{1}{\sqrt{T_n}} \sum_{j \in \hat{\mathcal{J}}_n^{k_n}} \frac{\Delta_j X^{\mathrm{cont}} - h_n b_{j-1}}{\mathbb{A}_{j-1}^\top \hat{\alpha}_n^{k_n}} \mathbb{B}_{j-1} \right| \mathbb{1}_{G_{k_n,n}} \\
&\quad + \left| \left(\frac{1}{n} \sum_{j=1}^{n} \frac{\mathbb{B}_{j-1}\mathbb{B}_{j-1}^\top}{\mathbb{A}_{j-1}^\top \hat{\alpha}_n^{k_n}} \right)^{-1} \right| \left| \left(\frac{1}{n} \sum_{j=1}^{n} \frac{\mathbb{B}_{j-1}\mathbb{B}_{j-1}^\top}{\mathbb{A}_{j-1}^\top \hat{\alpha}_n^{k_n}} \right) \left(\frac{1}{n} \sum_{j \notin \hat{\mathcal{J}}_n^{k_n}} \frac{\mathbb{B}_{j-1}\mathbb{B}_{j-1}^\top}{\mathbb{A}_{j-1}^\top \hat{\alpha}_n^{k_n}} \right)^{-1} - I_{p_\beta} \right| \\
&\quad \times \left| \frac{1}{\sqrt{T_n}} \sum_{j=1}^{n} \frac{\Delta_j X^{\mathrm{cont}} - h_n b_{j-1}}{\mathbb{A}_{j-1}^\top \hat{\alpha}_n^{k_n}} \mathbb{B}_{j-1} \right| \mathbb{1}_{G_{k_n,n}} \\
&= o_p(1),
\end{aligned}
$$

so that (3.6.23) holds true.

It remains to verify

$$
P\left(\left\{ \mathrm{JB}_n^{k_n} \leq \chi_q^2(2) \right\} \cap H_{k_n,n} \right) = o(1) \tag{3.6.29}
$$

where

$$
H_{k_n,n} := \{k_n + 1 \leq N_{T_n} \leq (\lambda+1)T_n\} \cap B_n \cap D_{k_n,n};
$$

recall that we are assuming that $k_n \leq (\lambda+1)T_n - 1$ without loss of generality. In view of the definition (3.3.1), (3.6.29) follows on showing that for any $M > 0$,

$$
P\left(\left\{ \frac{1}{\sqrt{n}} \sum_{j \notin \hat{\mathcal{J}}_n^{k_n}} \left(\left(\hat{N}_j^k \right)^4 - 3 \right) < M \right\} \cap H_{k_n,n} \right) = o(1); \tag{3.6.30}
$$

recall the notation $\hat{N}_j^k = (\hat{S}_n^k)^{-1/2}(\epsilon_j(\hat{\alpha}_n^k) - \bar{\hat{\epsilon}}_n^k)$.

First we will prove

$$
\bar{\hat{\epsilon}}_n^{k_n} \mathbb{1}_{H_{k_n,n}} = O_p\left(a_n \sqrt{h_n} \right). \tag{3.6.31}
$$

Decompose $\bar{\hat{\epsilon}}_n^{k_n}$ as

$$
\bar{\hat{\epsilon}}_n^{k_n} = \frac{1}{n - k_n} \left(\sum_{j=1}^{n} \epsilon_j(\hat{\alpha}_n^{k_n}) - \sum_{j \in \hat{\mathcal{J}}_n^{k_n}} \epsilon_j(\hat{\alpha}_n^{k_n}) \right).
$$

For any $j \in \mathbb{N}$, we define the indicator function $\chi_j(\cdot)$ as:

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

Since

$$\sum_{j=1}^{n}\epsilon_j(\hat{\alpha}_n^{k_n}) = \sum_{j=1}^{n}\frac{\int_{t_{j-1}}^{t_j} a_s dw_s}{\sqrt{h_n \mathbb{A}_{j-1}^\top \hat{\alpha}_n^{k_n}}} + \sum_{j=1}^{n}\frac{\int_{t_{j-1}}^{t_j}(b_s + c_s\lambda E[\xi_1])ds}{\sqrt{h_n \mathbb{A}_{j-1}^\top \hat{\alpha}_n^{k_n}}} + \int_0^{T_n}\sum_{j=1}^{n}\frac{c_{s-}\chi_j(s)}{\sqrt{h_n \mathbb{A}_{j-1}^\top \hat{\alpha}_n^{k_n}}}d\tilde{J}_s,$$

where $\tilde{J}_t := J_t - \lambda E[\xi_1]t$, Assumption 3.4.2, Sobolev's inequality and Burkholder's inequality imply that for any $q > p_\alpha$,

$$\begin{aligned} E\left[\sup_{\alpha\in\Theta_\alpha}\left|\frac{1}{n-k_n}\sum_{j=1}^{n}\epsilon_j(\hat{\alpha}_n^{k_n})\right|^q\right] &\lesssim \sup_{\alpha\in\Theta_\alpha}\left\{E\left[\left|\frac{1}{n}\sum_{j=1}^{n}\epsilon_j(\alpha)\right|^q\right] + E\left[\left|\frac{1}{n}\sum_{j=1}^{n}\partial_\alpha\epsilon_j(\alpha)\right|^q\right]\right\} \\ &= O_p\left(h_n^{\frac{q}{2}} \vee n^{-\frac{q}{2}}\right) = O_p\left(h_n^{\frac{q}{2}}\right) \end{aligned} \tag{3.6.32}$$

It also follows from Assumptions 3.4.1 and 3.4.2 that

$$\begin{aligned} \max_{1\le j\le n}\epsilon_j^2(\hat{\alpha}_n^{k_n})\mathbb{1}_{H_{k_n,n}} &\lesssim \frac{1}{h_n}\max_{1\le j\le n}(\Delta_j X)^2\mathbb{1}_{H_{k_n,n}} \\ &\lesssim \frac{1}{h_n}\sum_{j=1}^{n}\left\{\left(\int_{t_{j-1}}^{t_j} b_s ds\right)^2 + \left(\int_{t_{j-1}}^{t_j}(a_s - a_{j-1})dw_s\right)^2\right\} \\ &\quad + \max_{1\le j\le n}\eta_j^2 + \frac{1}{h_n}\max_{1\le j\le\lfloor(\lambda+1)T_n\rfloor}\xi_j^2 \\ &\lesssim O_p(T_n) + O_p(\log n) + O_p\left(\frac{a_n^2}{h_n}\right) \\ &= O_p\left(\frac{a_n^2}{h_n}\left(\frac{T_n h_n}{a_n^2} + 1\right)\right) = O_p\left(\frac{a_n^2}{h_n}\right). \end{aligned} \tag{3.6.33}$$

This gives

$$\left|\frac{1}{n-k_n}\sum_{j\in\hat{\mathcal{J}}_n^{k_n}}\epsilon_j(\hat{\alpha}_n^{k_n})\right|\mathbb{1}_{H_{k_n,n}} \lesssim \frac{k_n}{n}\sqrt{\max_{1\le j\le n}\epsilon_j^2(\hat{\alpha}_n^{k_n})\mathbb{1}_{H_{k_n,n}}} = O_p\left(a_n\sqrt{h_n}\right), \tag{3.6.34}$$

and (3.6.31) follows from (3.6.32) and (3.6.34).

Note that (3.6.31) under Assumption 3.4.4 entails

$$\begin{aligned} \hat{S}_n^{k_n}\mathbb{1}_{H_{k_n,n}} &= \frac{1}{n-k_n}\sum_{j\notin\hat{\mathcal{J}}_n^{k_n}}\epsilon_j^2(\hat{\alpha}_n^{k_n})\mathbb{1}_{H_{k_n,n}} + O_p(a_n^2 h_n) \\ &= \frac{1}{n-k_n}\sum_{j\notin\hat{\mathcal{J}}_n^{k_n}}\epsilon_j^2(\hat{\alpha}_n^{k_n})\mathbb{1}_{H_{k_n,n}} + o_p(1). \end{aligned} \tag{3.6.35}$$

From Assumption 3.4.2, the following relation holds:

$$\frac{1}{T_n}\sum_{j\notin\hat{\mathcal{J}}_n^{k_n}}(\Delta_j X)^2 \lesssim \frac{1}{n-k_n}\sum_{j\notin\hat{\mathcal{J}}_n^{k_n}}\epsilon_j^2(\hat{\alpha}_n^{k_n}) \lesssim \frac{1}{T_n}\sum_{j=1}^{n}(\Delta_j X)^2 \tag{3.6.36}$$

From Cauchy-Schwarz inequality, Burkholder's inequality and [40, Lemma 4.5], we derive

$$\begin{aligned}
E\left[(\Delta_j X)^2\right] &= E\Bigg[\bigg(\int_{t_{j-1}}^{t_j}(a_s-a_{j-1})dw_s+\int_{t_{j-1}}^{t_j}(b_s+\lambda E(\xi_1)c_s)ds \\
&\qquad +\int_{t_{j-1}}^{t_j}(c_{s-}-c_{j-1})d\tilde{J}_s+a_{j-1}\Delta_j w+c_{j-1}\Delta_j\tilde{J}\bigg)^2\Bigg] \\
&= E\left[\left(a_{j-1}\Delta_j w+c_{j-1}\Delta_j\tilde{J}\right)^2\right]+O\left(h_n^{\frac{3}{2}}\right) \\
&\lesssim h_n.
\end{aligned}$$

Hence the rightmost side in (3.6.36) is $O_p(1)$. In a similar manner, we have

$$\begin{aligned}
&\frac{1}{T_n}\sum_{j\notin\hat{\mathcal{J}}_n^{k_n}}(\Delta_j X)^2 \\
&=\frac{1}{T_n}\sum_{j\notin\hat{\mathcal{J}}_n^{k_n}}\left(a_{j-1}\Delta_j w+c_{j-1}\Delta_j\tilde{J}\right)^2+O_p\left(\sqrt{h_n}\right) \\
&=\frac{1}{T_n}\sum_{j\notin\hat{\mathcal{J}}_n^{k_n}}\left(c_{j-1}\Delta_j\tilde{J}\right)^2+\frac{1}{T_n}\sum_{j=1}^{n}\left\{(a_{j-1}\Delta_j w)^2+2a_{j-1}c_{j-1}\Delta_j w\Delta_j\tilde{J}\right\} \\
&\qquad -\frac{1}{T_n}\sum_{j\in\hat{\mathcal{J}}_n^{k_n}}\left\{(a_{j-1}\Delta_j w)^2+2a_{j-1}c_{j-1}\Delta_j w\Delta_j\tilde{J}\right\}.
\end{aligned}$$

The independence between w and J, [19, Lemma 9], and the ergodic theorem yield that

$$\frac{1}{T_n}\sum_{j=1}^{n}\left\{(a_{j-1}\Delta_j w)^2+2a_{j-1}c_{j-1}\Delta_j w\Delta_j\tilde{J}\right\}\xrightarrow{p}\int a^2(x,\alpha_0)\pi_0(dx)>0.$$

In a similar manner to (3.6.33), Assumption 3.4.4 implies that

$$\begin{aligned}
&\left|\frac{1}{T_n}\sum_{j\in\hat{\mathcal{J}}_n^{k_n}}\left\{(a_{j-1}\Delta_j w)^2+2a_{j-1}c_{j-1}\Delta_j w\Delta_j\tilde{J}\right\}\right|\mathbb{1}_{H_{k_n,n}} \\
&\qquad \le O_p\left(\frac{k_n\log n}{n}\right)+O_p\left(a_n\sqrt{h_n\log n}\right)=O_p\left(a_n\sqrt{h_n\log n}\right)=o_p(1),
\end{aligned}$$

where again we used $a_n \gtrsim 1$. Summarizing the last three displays leads to

$$\frac{1}{T_n}\sum_{j\notin\hat{\mathcal{J}}_n^{k_n}}(\Delta_j X)^2 \geq \int a^2(x,\alpha_0)\pi_0(dx) + o_p(1).$$

This combined with (3.6.31), (3.6.35) and (3.6.36) implies that for any $\epsilon>0$ there exist a positive constant $K>1$ and a positive integer N such that

$$\sup_{n\geq N} P\left[\left(\left\{\hat{S}_n^{k_n}<\frac{1}{K}\right\}\cup\left\{\hat{S}_n^{k_n}>K\right\}\right)\cap\left\{\bar{\hat{\epsilon}}_n^{k_n}>Ka_n\sqrt{h_n}\right\}\cap H_{k_n,n}\right]<\epsilon.$$

Therefore we may below focus on the event

$$D_{k_n,n,\epsilon} := \left\{\frac{1}{K}\leq \hat{S}_n^{k_n}\leq K\right\}\cap\left\{\bar{\hat{\epsilon}}_n^{k_n}\leq Ka_n\sqrt{h_n}\right\}\cap H_{k_n,n}.$$

From Assumption 3.4.2, we have the following estimates with positive constants $C = C(a,c)$ and $C' = C'(a,c)$ only depending on the coefficient (a,c):

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{j\notin\hat{\mathcal{J}}_n^{k_n}}\left(\left(\hat{N}_j^k\right)^4-3\right)\mathbb{1}_{D_{k_n,n,\epsilon}}\\
&\gtrsim \sqrt{n}\left(\frac{1}{n}\sum_{j\notin\hat{\mathcal{J}}_n^{k_n}}\left\{\left(\epsilon_j(\hat{\alpha}_n^{k_n})-\bar{\hat{\epsilon}}_n^{k_n}\right)^4-3\right\}\right)\mathbb{1}_{D_{k_n,n,\epsilon}}\\
&\gtrsim \sqrt{n}\left\{\frac{1}{n}\sum_{j\notin\hat{\mathcal{J}}_n^{k_n}}\left(\frac{(\Delta_j X)^4}{h_n^2}-C\frac{a_n}{h_n}|\Delta_j X|^3\right)+O_p(1)\right\}\mathbb{1}_{D_{k_n,n,\epsilon}}\\
&\gtrsim \sqrt{n}\left\{\frac{1}{n}\sum_{j\notin\hat{\mathcal{J}}_n^{k_n}}\left(\frac{(\Delta_j J)^4}{h_n^2}-C'|\Delta_j J|^3\left(\frac{|\eta_j|}{h_n^{\frac{3}{2}}}+\frac{a_n}{h_n}\right)\right)+O_p(1)\right\}\mathbb{1}_{D_{k_n,n,\epsilon}}\\
&\gtrsim \sqrt{n}\Bigg\{\frac{1}{nh_n^2}\Bigg(\min_{1\leq j\leq\lfloor(\lambda+1)T_n\rfloor}|\xi_j|^4\\
&\qquad -C'\max_{1\leq j\leq\lfloor(\lambda+1)T_n\rfloor}|\xi_j|^3\left(\sqrt{h_n}\max_{1\leq j\leq n}|\eta_j|+h_na_n\right)\Bigg)+O_p(1)\Bigg\}\mathbb{1}_{D_{k_n,n,\epsilon}}\\
&\gtrsim \sqrt{n}\left\{\frac{1}{nh_n^2}\left(\min_{1\leq j\leq\lfloor(\lambda+1)T_n\rfloor}|\xi_j|^4-O_p\left(a_n^3\sqrt{h_n\log n}\right)\right)+O_p(1)\right\}\mathbb{1}_{D_{k_n,n,\epsilon}}\\
&\gtrsim \sqrt{n}\left\{\frac{a_n^3\sqrt{\log n}}{nh_n^{3/2}}\left(\frac{\min_{1\leq j\leq\lfloor(\lambda+1)T_n\rfloor}|\xi_j|^4}{a_n^3\sqrt{h_n\log n}}-O_p(1)\right)+O_p(1)\right\}\mathbb{1}_{D_{k_n,n,\epsilon}}.
\end{aligned}\tag{3.6.37}$$

Now we note that Assumption 3.4.4 implies that for any $M>0$,

$$P\left(\min_{1\leq j\leq\lfloor(\lambda+1)T_n\rfloor}|\xi_j|^4>M^4a_n^3\sqrt{h_n\log n}\right)$$

$$= \left\{1 - P\left(|\xi_1| \leq M a_n^{3/4} (h_n \log n)^{1/8}\right)\right\}^{\lfloor(\lambda+1)T_n\rfloor} \to 1.$$

Hence, we conclude that on an event whose probability gets arbitrarily close to 1 as $n \to \infty$,

$$\frac{1}{\sqrt{n}} \sum_{j \notin \hat{\mathcal{J}}_n^{k_n}} \left(\left(\hat{N}_j^k\right)^4 - 3\right) \mathbb{1}_{D_{k_n,n,\epsilon}} \gtrsim \frac{a_n^3 \sqrt{n \log n}}{n h_n^{3/2}} \mathbb{1}_{D_{k_n,n,\epsilon}} \gtrsim \frac{\sqrt{n \log n}}{n h_n^{3/2}} \mathbb{1}_{D_{k_n,n,\epsilon}},$$

followed by (3.6.30), hence (3.6.29) as well. The proof of Theorem 3.4.8 is thus complete. $\square$

Chapter 4

Statistical Analysis on R

YUIMA package on R is for simulation and statistical analysis of stochastic processes, and still under development. The author implemented his estimation methods as function `qmleLevy` and `snr` in the package. In this chapter, we illustrate its usage with `R` codes.

4.1 Function `qmleLevy`

The function `qmleLevy` is for the Gaussian quasi-likelihood estimation of the parameter $\theta := (\alpha, \gamma)$ in the Lévy driven SDE models expressed as:

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

It is defined by the following form:

```
qmleLevy(yuima, start, lower, upper, joint = FALSE, third = FALSE)
```

The inputs are as follows:

- `yuima`: a yuima object (including the form of coefficients, data, timestamps, ...);
- `lower`: a named list for specifying lower bounds of parameters;
- `upper`: a named list for specifying upper bounds of parameters;
- `start`: initial values to be passed to the optimizer;
- `joint`: perform joint estimation or two stage estimation? by default `joint=FALSE`. If there exists an overlapping parameter, `joint=TRUE` does not work for the theoretical reason;

- `third`; perform adaptive estimation? by default `third=FALSE`. If there exists an overlapping parameter, `third=TRUE` does not work for the theoretical reason,

and as the output, the function gives the estimated value of parameters.

Below we give an example with R code. Consider the following SDE model:

$$dX_t = -\theta_0 X_t dt + \frac{\theta_1}{\sqrt{1+X_t^2}} dZ_t,$$

where the driving noise Z_t obeys bGamma$(t, \sqrt{2}, t, \sqrt{2})$. We set:

- Sample size: 10000;
- The size of observation interval: 0.01;
- Terminal time: 100;
- True value: $(\theta_{0,0}, \theta_{1,0}) = (1, 2)$;
- Parameter space: $(0.5, 4) \times (1, 4)$.

The initial values of optimization are random variables from the uniform distribution on the parameter space. The example code is shown below:

```
dri<-"-theta0*x" ## set drift

jum<-"theta1/(1+x^2)^(-1/2)" ## set jump

yuima<-setModel(drift = dri
                ,jump.coeff = jum
                ,solve.variable = "x",state.variable = "x"
                ,measure.type = "code"
                ,measure = list(df="rbgamma(z,1,sqrt(2),1,sqrt(2))"))

n<-100000 ## the number of total generation
tp<-0.1 ## the degree of thinning
N<-n*tp ## the number of samples
T<-100 ## terminal
hn<-T/N ## stepsize

sam<-setSampling(Terminal = T, n=n) ## set sampling scheme
subsam<-setSampling(Terminal = T, n=N)
```

```
yuima<-setYuima(model = yuima, sampling = sam) ## model

true<-list(theta0 = 1,theta1 = 2) ## true values
upper<-list(theta0 = 4, theta1 = 4) ## set upper bound
lower<-list(theta0 = 0.5, theta1 = 1) ## set lower bound

set.seed(123)
yuima<-simulate(yuima, xinit = 0, true.parameter = true,sampling = sam,
                subsampling = subsam) ## generate a path
start<-list(theta0 = runif(1,0.5,4),
           theta1 = runif(1,1,4)) ## set initial values
qmleLevy(yuima,start=start,lower=lower,upper=upper, joint = FALSE,
           third = TRUE)
```

```
## $first
##   theta1
## 1.965757
##
## $second
##    theta0
## 0.9774629
##
## $third
##   theta1
## 1.964678
```

4.2 Function snr

For the jump diffusion models (3.1.1), the function `snr` under develpment conducts the iterative Jarque-Bera normality test proposed in Chapter 3, and calculate the Gaussian quasi-likelihood estimator of the drift and diffusion parameters. In YUIMA package, `snr` is defined as

```
snr(yuima,start,upper,lower,q)
```

The inputs are as follows:

- `yuima`: a yuima object (jump diffusion models with data, time-stamps, ...);
- `lower`: a named list for specifying lower bounds of parameters;

- `upper`: a named list for specifying upper bounds of parameters;
- `start`: initial values to be passed to the optimizer;
- `q`: significance level of the (iterative) Jarque-Bera test for jump detection.

As its output, the followings are given:

- sample path with jump points;
- plot of the original self-normalized residuals;
- histogram of the self-normalized residuals after jump detection;
- transition of estimators and Jarque-Bera statistics;
- ordered absolute-value of increments with threshold;
- the value of the initial estimator and jump-removed estimator with jump times and sizes.

Below we demonstrate an example. Suppose that the following statistical model is given:

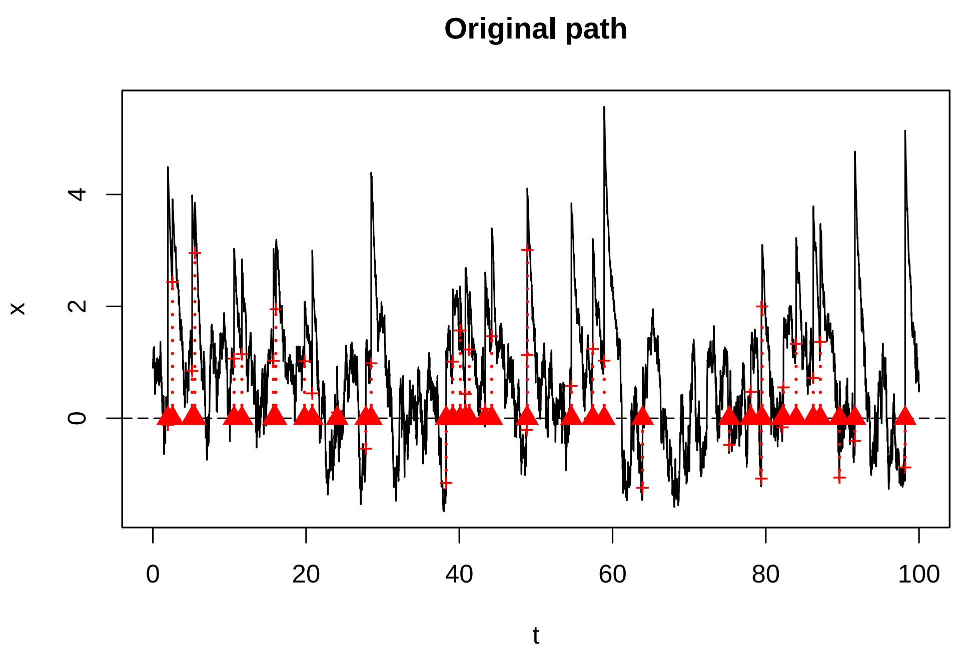
$$dX_t = \frac{\theta}{\sqrt{1+X_t^2}}dw_t - X_t dt + dJ_t,$$

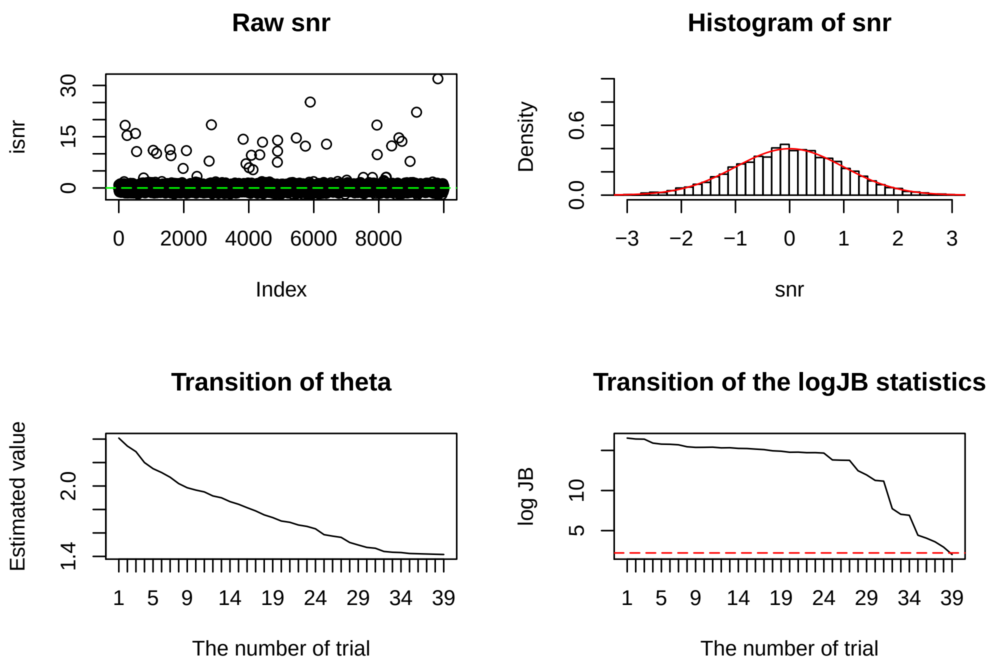
where the intensity and jump distribution of the driving compound Poisson process are 0.3 and $\Gamma(2,1)$, respectively. We set:

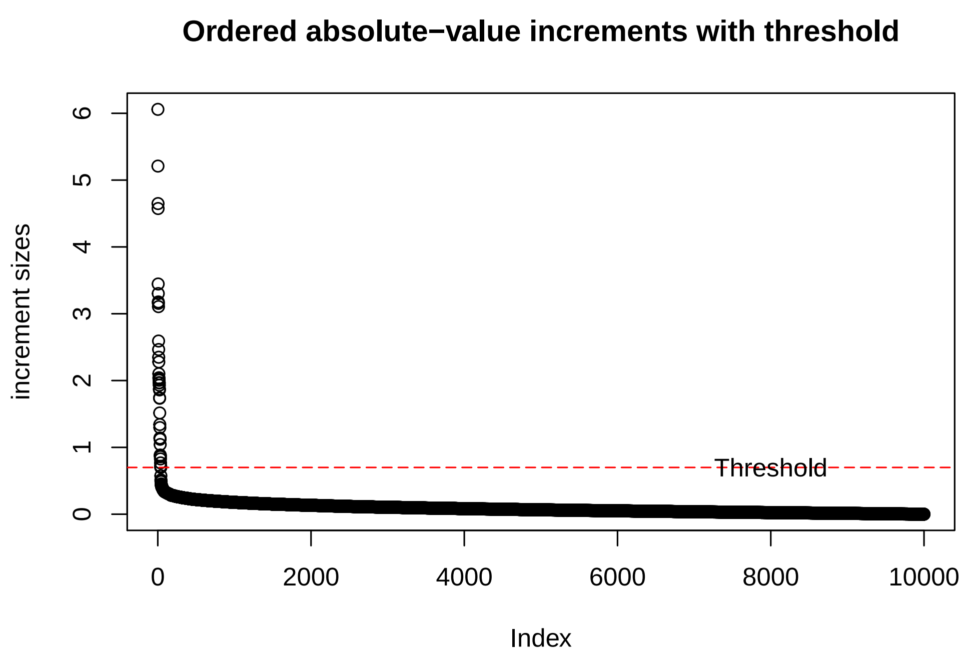
- Sample size: 10000;
- The size of observation interval: 0.01;
- Terminal time: 100;
- True value: $\sqrt{2}$;
- Parameter space: $(0.01, 100)$;
- Start value of the optimization: 0.5;
- Significance level: 0.01.

We can make use of the function `snr` by the following code, and the output is given after it:

```r
mod <- setModel(drift="-x",
                diffusion="theta/sqrt(1+x^2)",
                jump.coeff="1",
                measure=list(intensity="0.3",
                         df=list("dgamma(z, 2, 1)")),
                measure.type="CP")
T <- 100
n <- 10000
samp <- setSampling(Terminal=T, n=n)
yuima <- setYuima(model = mod, sampling = samp)
set.seed(123)
yuima <- simulate(yuima, xinit=1,true.parameter=list(theta=sqrt(2)),
sampling = samp)
snr(yuima,start=list(theta=0.5),upper=c(theta=100),lower=c(theta=0.01),
q=0.01)
```




Raw snr
isnr
0
15
30
0
2000
4000
6000
8000
Index
Histogram of snr
Density
0.0
0.6
−3
−2
−1
0
1
2
3
snr
Transition of theta
Estimated value
1.4
2.0
1
5
9
14
19
24
29
34
39
The number of trial
Transition of the logJB statistics
log JB
5
10
1
5
9
14
19
24
29
34
39
The number of trial


Ordered absolute−value increments with threshold
increment sizes
0
1
2
3
4
5
6
Threshold
0
2000
4000
6000
8000
10000
Index

```
$Removed
  Jump time        98.19        91.63         1.96
"Jump size"      "6.059"       "5.21"      "4.649"
      58.91         28.5        54.63         5.13
    "4.575"      "3.445"      "3.302"      "3.177"
      79.43         86.2        20.81        43.38
    "3.156"      "3.106"      "2.592"      "2.464"
      38.28         40.8        87.12        15.74
    "2.349"      "2.279"      "2.102"      "2.042"
      63.92        57.42        44.23         10.6
    "2.025"      "2.005"      "1.968"      "1.966"
      83.96         48.8        48.86         27.8
    "1.931"      "1.875"       "1.86"      "1.741"
      11.62         2.56        89.63        39.15
    "1.735"      "1.515"      "1.341"      "1.296"
      48.88        79.53        16.06         19.8
     "1.14"      "1.135"      "1.122"      "1.042"
       5.48        41.29        24.06        40.12
    "0.885"      "0.877"       "0.86"      "0.827"
      82.19        78.04        82.31        75.25
    "0.768"      "0.717"      "0.709"      "0.699"

$OGQMLE
  theta
2.54405

$JRGQMLE
   theta
1.416278
```

Bibliography

[1] R. A. Adams. Some integral inequalities with applications to the imbedding of Sobolev spaces defined over irregular domains. *Trans. Amer. Math. Soc.*, 178:401–429, 1973.

[2] Y. Aït-Sahalia and J. Jacod. Testing for jumps in a discretely observed process. *Ann. Statist.*, 37(1):184–222, 2009.

[3] D. Applebaum. *Lévy processes and stochastic calculus*, volume 116 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, second edition, 2009.

[4] F. M. Bandi and T. H. Nguyen. On the functional estimation of jump-diffusion models. *J. Econometrics*, 116(1-2):293–328, 2003. Frontiers of financial econometrics and financial engineering.

[5] G. Barles, R. Buckdahn, and E. Pardoux. Backward stochastic differential equations and integral-partial differential equations. *Stochastics Stochastics Rep.*, 60(1-2):57–83, 1997.

[6] G. Barles and C. Imbert. Second-order elliptic integro-differential equations: viscosity solutions' theory revisited. *Ann. Inst. H. Poincaré Anal. Non Linéaire*, 25(3):567–585, 2008.

[7] O. E. Barndorff-Nielsen. Processes of normal inverse Gaussian type. *Finance Stoch.*, 2(1):41–68, 1998.

[8] O. E. Barndorff-Nielsen, N. Shephard, et al. Basics of lévy processes. *Draft Chapter from a book by the authors on Lévy Driven Volatility Models*, 2012.

[9] D. Belomestny and M. Reiß. Estimation and calibration of Lévy models via Fourier methods. In *Lévy matters. IV*, volume 2128 of *Lecture Notes in Math.*, pages 1–76. Springer, Cham, 2015.

[10] R. H. Berk. Limiting behavior of posterior distributions when the model is incorrect. *Ann. Math. Statist. 37 (1966), 51–58; correction, ibid*, 37:745–746, 1966.

[11] R. N. Bhattacharya. On the functional central limit theorem and the law of the iterated logarithm for Markov processes. *Z. Wahrsch. Verw. Gebiete*, 60(2):185–201, 1982.

[12] A. Brouste, M. Fukasawa, H. Hino, S. M. Iacus, K. Kamatani, Y. Koike, H. Masuda, R. Nomura, T. Ogihara, Y. Shimuzu, et al. The yuima project: a computational framework for simulation and inference of stochastic differential equations. *Journal of Statistical Software*, 57(4):1–51, 2014.

[13] D. J. Daley and D. Vere-Jones. *An introduction to the theory of point processes. Vol. II.* Probability and its Applications (New York). Springer, New York, second edition, 2008. General theory and structure.

[14] R. M. Dudley. *Real analysis and probability*, volume 74 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2002. Revised reprint of the 1989 original.

[15] R. Durrett. *Probability: theory and examples*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, fourth edition, 2010.

[16] P. Embrechts, C. Klüppelberg, and T. Mikosch. *Modelling extremal events*, volume 33 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin, 1997. For insurance and finance.

[17] J. E. Figueroa-López. Small-time moment asymptotics for Lévy processes. *Statist. Probab. Lett.*, 78(18):3355–3365, 2008.

[18] J. E. Figueroa-López. Nonparametric estimation of Lévy models based on discrete-sampling. In *Optimality*, volume 57 of *IMS Lecture Notes Monogr. Ser.*, pages 117–146. Inst. Math. Statist., Beachwood, OH, 2009.

[19] V. Genon-Catalot and J. Jacod. On the estimation of the diffusion coefficient for multi-dimensional diffusion processes. In *Annales de l'institut Henri Poincaré (B) Probabilités et Statistiques*, volume 29, pages 119–151. Gauthier-Villars, 1993.

[20] E. Gobet. LAN property for ergodic diffusions with discrete observations. *Ann. Inst. H. Poincaré Probab. Statist.*, 38(5):711–737, 2002.

[21] S. Hamadène. Viscosity solutions of second order integral–partial differential equations without monotonicity condition: A new result. *Nonlinear Anal.*, 147:213–235, 2016.

[22] S. Hamadène and M.-A. Morlais. Viscosity solutions for second order integro-differential equations without monotonicity condition: the probabilistic approach. *Stochastics*, 88(4):632–649, 2016.

[23] P. J. Huber. The behavior of maximum likelihood estimates under nonstandard conditions. In *Proc. Fifth Berkeley Sympos. Math. Statist. and Probability (Berkeley, Calif., 1965/66), Vol. I: Statistics*, pages 221–233. Univ. California Press, Berkeley, Calif., 1967.

[24] J. Jacod and A. N. Shiryaev. *Limit theorems for stochastic processes*, volume 288 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, second edition, 2003.

[25] C. M. Jarque and A. K. Bera. A test for normality of observations and regression residuals. *Internat. Statist. Rev.*, 55(2):163–172, 1987.

[26] K. Kamatani and M. Uchida. Hybrid multi-step estimators for stochastic differential equations based on sampled data. *Stat. Inference Stoch. Process.*, 18(2):177–204, 2015.

[27] M. Kessler. Estimation of an ergodic diffusion from discrete observations. *Scand. J. Statist.*, 24(2):211–229, 1997.

[28] A. Kohatsu-Higa, E. Nualart, and N. K. Tran. Lan property for an ergodic diffusion with jumps. *Statistics*, 51(2):419–454, 2017.

[29] T. Komorowski and A. Walczuk. Central limit theorem for Markov processes with spectral gap in the Wasserstein metric. *Stochastic Process. Appl.*, 122(5):2155–2184, 2012.

[30] U. Küchler and S. Tappe. Bilateral gamma distributions and processes in financial mathematics. *Stochastic Process. Appl.*, 118(2):261–283, 2008.

[31] A. M. Kulik. Exponential ergodicity of the solutions to SDE's with a jump noise. *Stochastic Process. Appl.*, 119(2):602–632, 2009.

[32] Y. A. Kutoyants. The asymptotics of misspecified MLEs for some stochastic processes: a survey. *Stat. Inference Stoch. Process.*, 20(3):347–367, 2017.

[33] H. Long, C. Ma, and Y. Shimizu. Least squares estimators for stochastic differential equations driven by small Lévy noises. *Stochastic Processes and their Applications*, 2016.

[34] C. Mancini. Estimation of the characteristics of the jumps of a general Poisson-diffusion model. *Scand. Actuar. J.*, (1):42–52, 2004.

[35] R. Martin, C. Ouyang, and F. Domagni. 'Purposely misspecified' posterior inference on the volatility of a jump diffusion process. *Statist. Probab. Lett.*, 134:106–113, 2018.

[36] H. Masuda. Ergodicity and exponential β-mixing bounds for multidimensional diffusions with jumps. *Stochastic Process. Appl.*, 117(1):35–56, 2007.

[37] H. Masuda. On stability of diffusions with compound-Poisson jumps. *Bull. Inform. Cybernet.*, 40:61–74, 2008.

[38] H. Masuda. Approximate self-weighted LAD estimation of discretely observed ergodic Ornstein-Uhlenbeck processes. *Electron. J. Stat.*, 4:525–565, 2010.

[39] H. Masuda. Asymptotics for functionals of self-normalized residuals of discretely observed stochastic processes. *Stochastic Process. Appl.*, 123(7):2752–2778, 2013.

[40] H. Masuda. Convergence of gaussian quasi-likelihood random fields for ergodic Lévy driven SDE observed at high frequency. *Ann. Statist.*, 41(3):1593–1641, 2013.

[41] H. Masuda. Parametric estimation of Lévy processes. In *Lévy matters. IV*, volume 2128 of *Lecture Notes in Math.*, pages 179–286. Springer, Cham, 2015.

[42] H. Masuda. Non-gaussian quasi-likelihood estimation of SDE driven by locally stable Lévy process. *Stochastic Processes and their Applications, to appear*, 2018.

[43] H. Masuda and Y. Uehara. On stepwise estimation of Lévy driven stochastic differential equation (japanese). *Proc. Inst. Statist. Math.*, 65(1):21–38, 2017.

[44] H. Masuda and Y. Uehara. Two-step estimation of ergodic Lévy driven SDE. *Stat. Inference Stoch. Process*, 20:105–137, 2017.

[45] I. W. McKeague. Estimation for diffusion processes under misspecified models. *J. Appl. Probab.*, 21(3):511–520, 1984.

[46] S. P. Meyn and R. L. Tweedie. Stability of Markovian processes. III. Foster-Lyapunov criteria for continuous-time processes. *Adv. in Appl. Probab.*, 25(3):518–548, 1993.

[47] R. Nickl, M. Reiß, J. Söhl, and M. Trabs. High-frequency Donsker theorems for Lévy measures. *Probab. Theory Related Fields*, 164(1-2):61–108, 2016.

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

[49] E. Pardoux and A. Y. Veretennikov. On the Poisson equation and diffusion approximation. I. *Ann. Probab.*, 29(3):1061–1085, 2001.

[50] P. E. Protter. *Stochastic integration and differential equations*, volume 21 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin, second edition, 2004. Stochastic Modelling and Applied Probability.

[51] D. Revuz and M. Yor. *Continuous martingales and Brownian motion*, volume 293 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, third edition, 1999.

[52] K.-i. Sato. *Lévy processes and infinitely divisible distributions.* Cambridge university press, 1999.

[53] K.-i. Sato and M. Yamazato. Operator-self-decomposable distributions as limit distributions of processes of Ornstein-Uhlenbeck type. *Stochastic Process. Appl.*, 17(1):73–100, 1984.

[54] E. Schmisser. Non-parametric adaptive estimation of the drift for a jump diffusion process. *Stochastic Process. Appl.*, 124(1):883–914, 2014.

[55] Y. Shimizu. A practical inference for discretely observed jump-diffusions from finite samples. *J. Japan Statist. Soc.*, 38(3):391–413, 2008.

[56] Y. Shimizu. Threshold estimation for jump-type stochastic processes from discrete observations. *Proc. Inst. Statist. Math.*, 57(1):97–118, 2009.

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

[58] F. W. Steutel. Infinite divisibility in theory and practice. *Scand. J. Statist.*, 6(2):57–64, 1979.

[59] F. W. Steutel and K. van Harn. *Infinite divisibility of probability distributions on the real line*, volume 259 of *Monographs and Textbooks in Pure and Applied Mathematics.* Marcel Dekker, Inc., New York, 2004.

[60] M. Uchida and N. Yoshida. Estimation for misspecified ergodic diffusion processes from discrete observations. *ESAIM Probab. Stat.*, 15:270–290, 2011.

[61] M. Uchida and N. Yoshida. Adaptive estimation of an ergodic diffusion process based on sampled data. *Stochastic Process. Appl.*, 122(8):2885–2924, 2012.

[62] Y. Uehara. Statistical inference for misspecified ergodic lévy driven stochastic differential equation models. *Stochastic Processes and their Applications, to appear*, 2018.

[63] A. W. Van der Vaart. *Asymptotic statistics*, volume 3. Cambridge university press, 2000.

[64] A. Y. Veretennikov and A. M. Kulik. The extended Poisson equation for weakly ergodic Markov processes. *Teor. Ĭmovīr. Mat. Stat.*, (85):22–38, 2011.

[65] A. Y. Veretennikov and A. M. Kulik. Diffusion approximation of systems with weakly ergodic Markov perturbations. II. *Teor. Ĭmovīr. Mat. Stat.*, (88):1–16, 2013.

[66] F.-Y. Wang. Derivative formula and Harnack inequality for SDEs driven by Lévy processes. *Stoch. Anal. Appl.*, 32(1):30–49, 2014.

[67] F.-Y. Wang, L. Xu, and X. Zhang. Gradient estimates for SDEs driven by multiplicative Lévy noise. *J. Funct. Anal.*, 269(10):3195–3219, 2015.

[68] H. White. Maximum likelihood estimation of misspecified models. *Econometrica*, 50(1):1–25, 1982.

[69] N. Yoshida. Polynomial type large deviation inequalities and quasi-likelihood analysis for stochastic differential equations. *Ann. Inst. Statist. Math.*, 63(3):431–479, 2011.