

# CONVERGENCE OF GAUSSIAN QUASI-LIKELIHOOD RANDOM FIELDS FOR ERGODIC LÉVY DRIVEN SDE OBSERVED AT HIGH FREQUENCY

BY HIROKI MASUDA<sup>1</sup>

*Kyushu University*

This paper investigates the Gaussian quasi-likelihood estimation of an exponentially ergodic multidimensional Markov process, which is expressed as a solution to a Lévy driven stochastic differential equation whose coefficients are known except for the finite-dimensional parameters to be estimated, where the diffusion coefficient may be degenerate or even null. We suppose that the process is discretely observed under the rapidly increasing experimental design with step size  $h_n$ . By means of the polynomial-type large deviation inequality, convergence of the corresponding statistical random fields is derived in a mighty mode, which especially leads to the asymptotic normality at rate  $\sqrt{nh_n}$  for all the target parameters, and also to the convergence of their moments. As our Gaussian quasi-likelihood solely looks at the local-mean and local-covariance structures, efficiency loss would be large in some instances. Nevertheless, it has the practically important advantages: first, the computation of estimates does not require any fine tuning, and hence it is straightforward; second, the estimation procedure can be adopted without full specification of the Lévy measure.

**1. Introduction.** Let  $X = (X_t)_{t \in \mathbb{R}_+}$  be a solution to the stochastic differential equation (SDE)

$$(1.1) \quad dX_t = a(X_t, \alpha) dt + b(X_t, \beta) dW_t + c(X_{t-}, \beta) dJ_t,$$

where the ingredients involved are as follows:

- the finite-dimensional unknown parameter

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

where, for simplicity, the parameter spaces  $\Theta_\alpha \subset \mathbb{R}^{p_\alpha}$  and  $\Theta_\beta \subset \mathbb{R}^{p_\beta}$  are supposed to be bounded convex domains; the parameter  $\alpha$  (resp.,  $\beta$ ) affects local trend (resp., local dispersion);

- an  $r'$ -dimensional standard Wiener process  $W$  and an  $r''$ -dimensional centered pure-jump Lévy process  $J$ , whose Lévy measure is denoted by  $\nu$ ;

---

Received February 2012; revised April 2013.

<sup>1</sup>Supported in part by JSPS KAKENHI Grant Number 20740061 and 23740082.

*MSC2010 subject classifications.* 62M05.

*Key words and phrases.* Exponential ergodicity, Gaussian quasi-likelihood estimation, high-frequency sampling, Lévy driven stochastic differential equation, polynomial-type large deviation inequality.

- the initial variable  $X_0$  independent of  $(W, J)$ , with  $\eta := \mathcal{L}(X_0)$  possibly depending on  $\theta$ ;
- the measurable functions  $a: \mathbb{R}^d \times \Theta_\alpha \rightarrow \mathbb{R}^d$ ,  $b: \mathbb{R}^d \times \Theta_\beta \rightarrow \mathbb{R}^d \otimes \mathbb{R}^{r'}$ , and  $c: \mathbb{R}^d \times \Theta_\beta \rightarrow \mathbb{R}^d \otimes \mathbb{R}^{r''}$ .

Incorporation of the jump part extends the continuous-path diffusion parametric model, which are nowadays widely used in many application fields. We denote by  $P_\theta$  the image measure of a solution process  $X$  associated with  $\theta \in \Theta \subset \mathbb{R}^p$ , where  $p := p_\alpha + p_\beta$ . Suppose that the true parameter  $\theta_0 = (\alpha_0, \beta_0) \in \Theta$  does exist, with  $P_0$  denoting the shorthand for the true image measure  $P_{\theta_0}$ , and that  $X$  is not completely (continuously) observed but only discretely at high frequency under the condition for the rapidly increasing experimental design: we are given a sample  $(X_{t_0}, X_{t_1}, \dots, X_{t_n})$ , where  $t_j = t_j^n = jh_n$  for some  $h_n > 0$  such that

$$(1.2) \quad T_n := nh_n \rightarrow \infty \quad \text{and} \quad nh_n^2 \rightarrow 0$$

for  $n \rightarrow \infty$ . The main objective of this paper is to estimate  $\theta_0$  under the exponential ergodicity of  $X$ ; the equidistant sampling assumption can be weakened to some extent as long as the long-term and high-frequency framework is concerned; however, it is just a technical extension making the presentation notationally messy, and hence we do not deal with it in the main context to make the presentation more clear.

It is common knowledge that the maximum likelihood estimation is generally infeasible, since the transition probability is most often unavailable in a closed form. This implies that the conventional statistical analyses based on the genuine likelihood have no utility. For this reason, we have to resort to some other feasible estimation procedure, which could be a lot of things. Among several possibilities, we are concerned here with the *Gaussian quasi-likelihood (GQL)* function defined as if the conditional distributions of  $X_{t_j}$  given  $X_{t_{j-1}}$  are Gaussian with approximate but explicit mean vector and covariance matrix; see (2.9) below.

The terminology ‘‘quasi-likelihood’’ has originated as the pioneering work of Wedderburn [46], the concept of which formed a basis of the generalized linear regression. The GQL-based estimation has been known to have the advantage of computational simplicity and robustness for misspecification of the noise distribution, and is well-established as a fundamental tool in estimating possibly non-Gaussian and dependent statistical models. Just to be a little more precise, consider a time-series  $Y_1, \dots, Y_n$  in  $\mathbb{R}$  with a fixed  $Y_0$ , and denote by  $m_{j-1}(\theta) \in \mathbb{R}$  and  $v_{j-1}(\theta) > 0$  the conditional mean and conditional variance of  $Y_j$  given  $(Y_0, \dots, Y_{j-1})$ , where  $\theta$  is an unknown parameter of interest. Then, the *Gaussian quasi maximum likelihood estimator (GQMLE)* is defined to be a maximizer of the function

$$\theta \mapsto \sum_{j=1}^n \log \left\{ \frac{1}{\sqrt{2\pi v_{j-1}(\theta)}} \exp \left( -\frac{(Y_j - m_{j-1}(\theta))^2}{2v_{j-1}(\theta)} \right) \right\}.$$

Namely, we compute the likelihood of  $(Y_1, Y_2, \dots, Y_n)$  as if the conditional law of  $Y_j$  given  $(Y_1, \dots, Y_{j-1})$  is Gaussian with mean  $m_{j-1}(\theta)$  and variance  $v_{j-1}(\theta)$ , so that only the structures of the conditional mean and variance do matter. Although it is not asymptotically efficient in general, it can serve as a widely applicable estimation procedure. One can consult Heyde [12] for an extensive and systematic account of statistical inference based on the GQL. The GQL has been a quite popular tool for (semi)parametric estimation, and especially there exists a vast amount of literature concerning asymptotics of the GQL for time series models with possibly non-Gaussian error sequence; among others, we refer to Straumann and Mikosch [41] for a class of conditionally heteroscedastic time series models, and Bardet and Wintenburger [3] for multidimensional causal time series, as well as the references therein.

Let us return to our framework. On one hand, for the diffusion case (where  $c \equiv 0$ ), the GQL-estimation issue has been solved under some regularity conditions, especially the GQL, which leads to an asymptotically efficient estimator, where the crucial point is that the optimal rates of convergence for estimating  $\alpha$  and  $\beta$  are different and given by  $\sqrt{T_n}$  and  $\sqrt{n}$ , respectively; see Gobet [11] for the local asymptotic normality of the corresponding statistical experiments. For how to construct an explicit contrast function, we refer to Yoshida [47] and Kessler [18] as well as the references therein; specifically, they employed a discretized version of the continuous-observation likelihood process, and a higher order local-Gauss approximation of the transition density, respectively. Sørensen [40] includes an extensive bibliography of many existing results, including explicit martingale estimating functions for discretely observed diffusions (not necessarily at high frequency). On the other hand, the issue has not been addressed enough in the presence of jumps (possibly of infinite variation). The question we should then ask is what will occur when one adopts the GQL function. In this paper, we will provide sufficient conditions under which the GQL random field associated with our statistical experiments converges in a mighty mode; see Section 3. We will apply Yoshida [48] to derive the mighty convergence with the limit being shifted Gaussian. As results, we will obtain an asymptotically normally distributed estimator at rate  $\sqrt{T_n}$  for both  $\alpha$  and  $\beta$  and also, very importantly, the convergence of their moments to the corresponding ones of the limit centered Gaussian distribution. Different from the diffusion case, the GQL does not lead to an asymptotically efficient estimator in the presence of jumps, and is not even rate-efficient for  $\beta$ : for instance, in the case where  $X$  is a diffusion with compound-Poisson jumps, the information loss in the QMLE of  $\alpha$  can be large if the jump part is much larger than the diffusion part; see Section 2.3.2. That is to say, the performance of our QMLE may strongly depend on the structure of the jump part and its relation to the possibly nondegenerate diffusion one, which may be considered as a possible major drawback of our estimation procedure. Nevertheless, it has the practically important advantages: first, the computation of estimates does not require any fine tuning, hence is straightforward; second, the estimation procedure can be adopted

without full specification of the Lévy measure  $\nu$ . Further, our numerical experiments in Section 2.4 reveal that, when the diffusion part is absent, it can happen that the finite-sample performance of  $\hat{\theta}_n$  becomes as good as the diffusion case if  $J$  “distributionally” close to the Wiener process.

We should mention that the convergence of moments especially serves as a fundamental tool when analyzing asymptotic behavior of the expectations of statistics depending on the estimator, for example, asymptotic bias and mean squared prediction error, model-selection devices (information criteria) and remainder estimation in higher-order inference. In the past, several authors have investigated such a strong mode of convergence of estimators; see Bhansali and Papangelou [5], Chan and Ing [6], Findley and Wei [8], Inagaki and Ogata [14], Jegannathan [16, 17], Ogata and Inagaki [35], Sieders and Dzhaparidze [39] and Uchida [42], as well as Ibragimov and Has’minski [13], Kutoyants [22, 23] and Yoshida [48]. See also the recent paper Uchida and Yoshida [43] for an adaptive parametric estimation of diffusions with moment convergence of estimators under the sampling design  $nh_n^k \rightarrow 0$  for arbitrary integer  $k \geq 2$ .

The rest of this paper is organized as follows. Section 2 introduces our GQL random field and presents its asymptotic behavior, together with a small numerical example for observing finite-sample performance of the GQMLE. Section 3 provides a somewhat general result concerning the mighty convergence, based on which we prove our main result in Section 4. In Section 5, we prove a fairly simple criterion for the exponential ergodicity assumption in dimension one, only in terms of the coefficient  $(a, b, c)$  and the Lévy measure  $\nu(dz)$ .

Throughout this paper, asymptotics are taken for  $n \rightarrow \infty$  unless otherwise mentioned, and the following notation is used:

- $I_r$  denotes the  $r \times r$ -identity matrix;
- given a multilinear form  $M = \{M^{(i_1 i_2 \dots i_K)} : i_k = 1, \dots, d_k; k = 1, \dots, K\} \in \mathbb{R}^{d_1} \otimes \dots \otimes \mathbb{R}^{d_K}$  and variables  $u_k = \{u_k^{(i)}\}_{i \leq d_k} \in \mathbb{R}^{d_k}$ , we write

$$M[u_1, \dots, u_K] = \sum_{i_1=1}^{d_1} \dots \sum_{i_K=1}^{d_K} M^{(i_1 i_2 \dots i_K)} u_1^{(i_1)} \dots u_K^{(i_K)}.$$

The correspondences of indices of  $M$  and  $u_k$  will be clear from each context. Some of  $u_k$  may be missing in “ $M[u_1, \dots, u_K]$ ” so that the resulting form again defines a multilinear form, for example,  $M[u_3, \dots, u_K] \in \mathbb{R}^{d_1} \otimes \mathbb{R}^{d_2}$ . In particular, given two multilinear forms  $M^{(j)} = \{M^{(i_1 i_2 \dots i_{K(j)})}\}$ ,  $j = 1, 2$ , we often use the notation  $M^{(1)} \otimes M^{(2)}$  for the tensor product

$$\begin{aligned} & (M^{(1)} \otimes M^{(2)})[u_1, \dots, u_{K(1)}, v_1, \dots, v_{K(2)}] \\ & := (M^{(1)}[u_1, \dots, u_{K(1)}])(M^{(2)}[v_1, \dots, v_{K(2)}]). \end{aligned}$$

When  $K \leq 2$ , identifying  $M$  as a vector or matrix, we write  $M^{\otimes 2} = MM^\top$  with  $\top$  denoting the transpose; furthermore,  $|M|$  denotes either, depending on the context,  $\det(M)$  when  $d_1 = d_2$ , or any matrix norm of  $M$ .

- $\partial_a^m$  stands for the bundled  $m$ th partial differential operator with respect to  $a = \{a^{(i)}\}$ .
- $C$  denotes generic positive constant possibly varying from line to line, and we write  $x_n \lesssim y_n$  if  $x_n \leq Cy_n$  a.s. for every  $n$  large enough.

**2. Gaussian quasi-likelihood estimation.** We denote by  $(\Omega, \mathcal{F}, \mathbf{F} = (\mathcal{F}_t)_{t \in \mathbb{R}_+}, P)$  a complete filtered probability space on which the process  $X$  given by (1.1) is defined: the initial variable  $X_0$  being  $\mathcal{F}_0$ -measurable, and  $(W, J)$  being  $\mathbf{F}$ -adapted.

2.1. Assumptions.

ASSUMPTION 2.1 (Moments).  $E[J_1] = 0$ ,  $E[J_1^{\otimes 2}] = I_{r''}$ , and  $E[|J_1|^q] < \infty$  for all  $q > 0$ .

We introduce the function  $V : \mathbb{R}^d \times \Theta_\beta \rightarrow \mathbb{R}^d \otimes \mathbb{R}^d$  by

$$V = b^{\otimes 2} + c^{\otimes 2}.$$

For each  $\theta$ , the function  $x \mapsto V(x, \beta)$  can be viewed as an approximate local covariance matrix of the law of  $h_n^{-1/2}(X_{h_n} - x)$  under  $P_\theta[\cdot | X_0 = x]$ .

Let  $\bar{\Theta}$  denote the closure of  $\Theta$ .

ASSUMPTION 2.2 (Smoothness). (a) The coefficient  $(a, b, c)$  has the extension in  $\mathcal{C}(\mathbb{R}^d \times \bar{\Theta})$ , and has partial derivatives such that  $(\partial_\alpha a, \partial_\beta b, \partial_\beta c)$  admits the extension in  $\mathcal{C}(\mathbb{R}^d \times \bar{\Theta})$ , that

$$\sup_{(x, \theta) \in \mathbb{R}^d \times \Theta} \{|\partial_x a(x, \alpha)| + |\partial_x b(x, \beta)| + |\partial_x c(x, \beta)|\} < \infty,$$

and that, for each  $k \in \{0, 1, 2\}$  and  $l \in \{0, 1, \dots, 5\}$ , there exists a constant  $C(k, l) \geq 0$  for which

$$\sup_{(x, \theta) \in \mathbb{R}^d \times \Theta} (1 + |x|)^{-C(k, l)} \{|\partial_x^k \partial_\alpha^l a(x, \alpha)| + |\partial_x^k \partial_\beta^l b(x, \beta)| + |\partial_x^k \partial_\beta^l c(x, \beta)|\} < \infty.$$

(b)  $V(x, \beta)$  is invertible for each  $(x, \beta)$ , and there exists a constant  $C(V) \geq 0$  such that

$$\sup_{(x, \beta) \in \mathbb{R}^d \times \Theta_\beta} (1 + |x|)^{-C(V)} |V^{-1}(x, \beta)| < \infty.$$

When considering large-time asymptotics, the stability property of  $X$  much affects the statistical analysis in essential ways. A typical situation to be considered is that  $X$  is ergodic. We impose here a stronger stability condition. Let  $(P_t)$  denote the transition semigroup of  $X$ . Given a function  $\rho : \mathbb{R}^d \rightarrow \mathbb{R}_+$  and a signed measure  $m$  on the  $d$ -dimensional Borel space, we define

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

ASSUMPTION 2.3 (Stability). (a) There exists a probability measure  $\pi_0$  such that for every  $q > 0$  we can find a constant  $a > 0$  for which

$$(2.1) \quad \sup_{t \in \mathbb{R}_+} e^{at} \|P_t(x, \cdot) - \pi_0(\cdot)\|_g \lesssim g(x), \quad x \in \mathbb{R}^d,$$

where  $g(x) := 1 + |x|^q$ .

(b) For every  $q > 0$ ,

$$(2.2) \quad \sup_{t \in \mathbb{R}_+} E_0[|X_t|^q] < \infty.$$

Here and in the sequel,  $E_0$  denotes the expectation operator with respect to  $P_0$ . Condition (2.1) with  $g$  replaced by the constant 1 is the *exponential ergodicity*, which in particular entails the ergodic theorem: the limit  $\pi_0$  is a unique invariant distribution such that, for every  $f \in L^1(\pi_0)$ ,

$$(2.3) \quad \frac{1}{T_n} \int_0^{T_n} f(X_t) dt \rightarrow^P \int f(x)\pi_0(dx),$$

where  $\rightarrow^P$  stands for the convergence in  $P_0$ -probability; we see that

$$\frac{1}{n} \sum_{j=1}^n f(X_{t_{j-1}}) \rightarrow^P \int f(x)\pi_0(dx)$$

for continuously differentiable  $f$  with  $\partial f$  at most polynomial order, since

$$(2.4) \quad \begin{aligned} & E_0 \left[ \left| \frac{1}{T_n} \int_0^{T_n} f(X_t) dt - \frac{1}{n} \sum_{j=1}^n f(X_{t_{j-1}}) \right| \right] \\ & \lesssim \frac{1}{n} \sum_{j=1}^n \sup_{t_{j-1} \leq s \leq t_j} \sqrt{E_0[|X_s - X_{t_{j-1}}|^2]} \rightarrow 0. \end{aligned}$$

We also note that Assumption 2.3 entails the *exponential absolute regularity*, also referred to as the *exponential  $\beta$ -mixing property*. This means that  $\beta_X(t) = O(e^{-at})$  as  $t \rightarrow \infty$  for some  $a > 0$ , where  $\beta_X$  denotes the  $\beta$ -mixing coefficient

$$\beta_X(t) := \sup_{s \in \mathbb{R}_+} \int \|P_t(x, \cdot) - \eta P_{s+t}(\cdot)\| \eta P_s(dx),$$

where  $\eta P_t := \mathcal{L}(X_t)$  and  $\|m\| := \|m\|_1$ . Let us recall that the exponential absolute regularity implies the *exponential strong-mixing property*, which plays an essential role in Yoshida [48], Lemma 4, which we will apply in the proof of Theorem 2.7.

Several sufficient conditions for Assumption 2.3 are known; for diffusion processes, see the references of Masuda [28, 29] for some details. In the presence of the jump component, verification of (2.1) can become much more involved. Especially if the coefficients are nonlinear and the Lévy process  $J$  is of infinite variation, the verification may be far from being a trivial matter. We refer to Kulik [19,

20], Maruyama and Tanaka [26], Menaldi and Robin [33], Meyn and Tweedie [34] and Wang [45] as well as Masuda [28, 29] for some general results concerning the (exponential) ergodicity. For the sake of convenience, focusing on the univariate case and setting ease of verification above generality, we will provide in Proposition 5.4 sufficient conditions for Assumption 2.3, in a form enabling us to deal with cases of nonlinear coefficients and infinite-variation  $J$ ; see also Remark 5.6.

Define  $\mathbb{G}_\infty(\theta) = (\mathbb{G}_\infty^\alpha(\theta), \mathbb{G}_\infty^\beta(\beta)) \in \mathbb{R}^p$  by

$$(2.5) \quad \mathbb{G}_\infty^\alpha(\theta) = \int \partial_\alpha a(x, \alpha) [V^{-1}(x, \beta) [a(x, \alpha_0) - a(x, \alpha)]] \pi_0(dx),$$

$$(2.6) \quad \mathbb{G}_\infty^\beta(\beta) = \int \{V^{-1}(\partial_\beta V) V^{-1}(x, \beta)\} [V(x, \beta_0) - V(x, \beta)] \pi_0(dx).$$

[In (2.6), we regarded “ $V^{-1}(\partial_\beta V) V^{-1}(x, \beta)$ ” as a bilinear form with dimensions of indices being  $p_\beta$  and  $d^2$ .] Further, let  $\mathbb{G}'_\infty(\theta_0) := \text{diag}\{\mathbb{G}'_\infty^\alpha(\theta_0), \mathbb{G}'_\infty^\beta(\theta_0)\} \in \mathbb{R}^p \otimes \mathbb{R}^p$ , where, for each  $v'_1, v'_2 \in \mathbb{R}^{p_\alpha}$  and  $v''_1, v''_2 \in \mathbb{R}^{p_\beta}$ ,

$$(2.7) \quad \begin{aligned} &\mathbb{G}'_\infty^\alpha(\theta_0)[v'_1, v'_2] \\ &= - \int V^{-1}(x, \beta_0) [\partial_\alpha a(x, \alpha_0)[v'_1], \partial_\alpha a(x, \alpha_0)[v'_2]] \pi_0(dx), \end{aligned}$$

$$(2.8) \quad \begin{aligned} &\mathbb{G}'_\infty^\beta(\theta_0)[v''_1, v''_2] \\ &= - \int \text{trace}[\{(V^{-1} \partial_\beta V) \otimes (V^{-1} \partial_\beta V)\}(x, \beta_0)[v''_1, v''_2]] \pi_0(dx). \end{aligned}$$

ASSUMPTION 2.4 (Identifiability). There exist positive constants  $\chi_\alpha = \chi_\alpha(\theta_0)$  and  $\chi_\beta = \chi_\beta(\theta_0)$  such that  $|\mathbb{G}_\infty^\alpha(\theta)|^2 \geq \chi_\alpha |\alpha - \alpha_0|^2$  and  $|\mathbb{G}_\infty^\beta(\beta)|^2 \geq \chi_\beta |\beta - \beta_0|^2$  for every  $\theta \in \Theta$ .

ASSUMPTION 2.5 (Nondegeneracy). Both  $\mathbb{G}'_\infty^\alpha(\theta_0)$  and  $\mathbb{G}'_\infty^\beta(\theta_0)$  are invertible.

Assumptions 2.4 and 2.5 are quite typical in statistical estimation. In Lemma 2.6 below, both assumptions are implied by a kind of uniform nonsingularity. Define two bilinear forms  $\bar{A}(\alpha', \alpha'', \beta')$  and  $\bar{B}(\beta', \beta'')$  by, just like (2.7) and (2.8),

$$\bar{A}(\alpha', \alpha'', \beta')[v'_1, v'_2] = \int V^{-1}(x, \beta') [\partial_\alpha a(x, \alpha')[v'_1], \partial_\alpha a(x, \alpha'')[v'_2]] \pi_0(dx),$$

$$\begin{aligned} &\bar{B}(\beta', \beta'')[v''_1, v''_2] \\ &= \int \text{trace}[\{(V^{-1}(\partial_\beta V) V^{-1})(x, \beta') \otimes \partial_\beta V(x, \beta'')\} [v''_1, v''_2]] \pi_0(dx). \end{aligned}$$

LEMMA 2.6. Suppose that  $\bar{A}(\alpha', \alpha'', \beta')$  and  $\bar{B}(\beta', \beta'')$  are nonsingular uniformly in  $\alpha', \alpha'' \in \Theta_\alpha$  and  $\beta', \beta'' \in \Theta_\beta$ . Then both Assumptions 2.4 and 2.5 hold true.

PROOF. It is obvious that Assumption 2.5 follows. The mean-value theorem applied to (2.5) and (2.6) leads to  $\mathbb{G}_\infty^\alpha(\theta) = \bar{A}(\alpha, \tilde{\alpha}, \beta)[\alpha_0 - \alpha]$  for some  $\tilde{\alpha}$  lying on the segment connecting  $\alpha$  and  $\alpha_0$ , with a similar form for  $\mathbb{G}_\infty^\beta(\beta)$ ; recall that  $\Theta_\alpha$  and  $\Theta_\beta$  are presupposed to be convex. Since  $\inf_{\alpha', \alpha'', \beta'} \|\bar{A}(\alpha', \alpha'', \beta')\| > 0$  and  $\inf_{\beta', \beta''} \|\bar{B}(\beta', \beta'')\| > 0$  under the assumption, the matrices  $\bar{A}^{\otimes 2}$  and  $\bar{B}^{\otimes 2}$  are uniformly positive definite, hence Assumption 2.4 follows.  $\square$

2.2. *Asymptotics: Main results.* In what follows, we write

$$\Delta_j Y = Y_{t_j} - Y_{t_{j-1}}$$

for any process  $Y$ , and

$$f_{j-1}(a) = f(X_{t_{j-1}}, a)$$

for a variable  $a$  in some set  $A$  and a measurable function  $f$  on  $\mathbb{R}^d \times A$ . The Euler approximation for SDE (1.1) is formally

$$X_{t_j} \approx X_{t_{j-1}} + a_{j-1}(\alpha)h_n + b_{j-1}(\beta)\Delta_j W + c_{j-1}(\beta)\Delta_j J$$

under  $P_\theta$ , which leads us to consider the local-Gauss distribution approximation

$$(2.9) \quad \mathcal{L}(X_{t_j} | X_{t_{j-1}}) \approx \mathcal{N}_d(X_{t_{j-1}} + a_{j-1}(\alpha)h_n, h_n V_{j-1}(\beta)).$$

Put

$$\chi_j(\alpha) = \Delta_j X - h_n a_{j-1}(\alpha).$$

Based on (2.9), we define our GQL by

$$(2.10) \quad \mathbb{Q}_n(\theta) = - \sum_{j=1}^n \left\{ \log |V_{j-1}(\beta)| + \frac{1}{h_n} V_{j-1}^{-1}(\beta) [\chi_j(\alpha)^{\otimes 2}] \right\},$$

and the corresponding QMLE by any element

$$\hat{\theta}_n = (\hat{\alpha}_n, \hat{\beta}_n) \in \operatorname{argmax}_{\theta \in \Theta} \mathbb{Q}_n(\theta).$$

Under Assumption 2.1 we have  $\int z^{(k)} z^{(l)} \nu(dz) = \delta_{kl}$  for  $k, l \in \{1, \dots, r''\}$ . We need some further notation in this direction. For  $i_1, \dots, i_m \in \{1, \dots, r''\}$  with  $m \geq 3$ , we write  $\nu(m)$  for the  $m$ th mixed moments of  $\nu$ ,

$$\nu(m) = \{v_{i_1 \dots i_m}(m)\}_{i_1, \dots, i_m} := \left\{ \int z^{(i_1)} \dots z^{(i_m)} \nu(dz) \right\}_{i_1, \dots, i_m}.$$

Let  $c^{(k)}(x, \beta) \in \mathbb{R}^d$  denote the  $k$ th column of  $c(x, \beta)$ . We introduce the matrix

$$(2.11) \quad \mathbb{V}(\theta_0) := \begin{pmatrix} \mathbb{G}_\infty^\alpha(\theta_0) & \mathbb{V}_{\alpha\beta} \\ \mathbb{V}_{\alpha\beta}^\top & \mathbb{V}_{\beta\beta} \end{pmatrix},$$



where, for each  $v' \in \mathbb{R}^{P\alpha}$  and  $v''_1, v''_2 \in \mathbb{R}^{P\beta}$ ,

$$\begin{aligned} \mathbb{V}_{\alpha\beta}[v', v''_1] &:= - \int \sum_{k', l', s'} v_{k'l's'}(3) V^{-1}(x, \beta_0) [\partial_{\alpha} a(x, \alpha_0)[v'], c^{(s')}(x, \beta_0)] \\ &\quad \times \{ \partial_{\beta} V^{-1}(x, \beta_0) \} [v''_1, c^{(k')}(x, \beta_0), c^{(l')}(x, \beta_0)] \pi_0(dx), \\ \mathbb{V}_{\beta\beta}[v''_1, v''_2] &:= \int \sum_{s, t, s', t'} v_{sts't'}(4) \{ \partial_{\beta} V^{-1}(x, \beta_0) [v''_1, c^{(s)}(x, \beta_0), c^{(t)}(x, \beta_0)] \} \\ &\quad \times \{ \partial_{\beta} V^{-1}(x, \beta_0) [v''_2, c^{(s')}(x, \beta_0), c^{(t')}(x, \beta_0)] \} \pi_0(dx). \end{aligned}$$

Finally, put

$$\Sigma_0 = \begin{pmatrix} (-\mathbb{G}'_{\infty\alpha})^{-1}(\theta_0) & \{ (\mathbb{G}'_{\infty\alpha})^{-1} \mathbb{V}_{\alpha\beta} (\mathbb{G}'_{\infty\beta})^{-1} \}(\theta_0) \\ \text{Sym.} & \{ (\mathbb{G}'_{\infty\beta})^{-1} \mathbb{V}_{\beta\beta} (\mathbb{G}'_{\infty\beta})^{-1} \}(\theta_0) \end{pmatrix}.$$

Now we can state our main result, the proof of which is deferred to Section 4.1.

**THEOREM 2.7.** *Suppose Conditions 2.1, 2.2, 2.3, 2.4 and 2.5. Then we have*

$$E_0[f(\sqrt{T_n}(\hat{\theta}_n - \theta_0))] \rightarrow \int f(u)\phi(u; 0, \Sigma_0) du, \quad n \rightarrow \infty$$

for every continuous function  $f : \mathbb{R}^p \rightarrow \mathbb{R}$  of at most polynomial growth, where  $\phi(\cdot; 0, \Sigma_0)$  denotes the centered Gaussian density with covariance matrix  $\Sigma_0$ .

The following two remarks are immediate:

- The estimators  $\hat{\alpha}_n$  and  $\hat{\beta}_n$  are asymptotically independent if  $\nu(3) = 0$ , implying that  $\hat{\alpha}_n$  and  $\hat{\beta}_n$  may not be asymptotically independent if  $\nu$  is skewed. If  $c \equiv 0$  so that  $X$  is a diffusion, then  $\nu(4) = 0$ , so that  $\mathbb{V}_{\beta\beta} = 0$  and  $\sqrt{T_n}(\hat{\beta}_n - \beta_0)$  is asymptotically degenerate at 0. This is in accordance with the case of diffusion, where the GQMLE of  $\beta$  is  $\sqrt{n}$ -consistent. See Section 2.3.2 for a discussion on the efficiency issue.
- The revealed convergence rate  $\sqrt{T_n}$  of the GQMLE  $\hat{\beta}_n$  alerts us to take precautions against the presence of jumps. For instance, suppose that one has adopted the parametric diffusion model [i.e., (1.1) with  $c \equiv 0$ ] although there actually does exist a nonnull jump part. Then one takes  $\sqrt{n}$  for the convergence rate of  $\hat{\beta}_n$ , although the true one is  $\sqrt{T_n}$ , which may lead to a seriously inappropriate confidence zone. This point can be sufficient grounds for importance of testing the presence of jumps. In case of one-dimensional  $X$ , Masuda [31], Section 4, constructed an analogue to Jarque–Bera normality test and studied its asymptotic behavior. See Masuda [32] for a multivariate extension.

In order to construct confidence regions for  $\theta_0$  as well as to perform statistical tests, we need a consistent estimator of the asymptotic covariance matrix  $\Sigma_0$ . Although  $\Sigma_0$  contains unknown third and fourth mixed moments of  $\nu$ , it turns out to

be possible to provide a consistent estimator of  $\Sigma_0$  without any specific knowledge of  $\nu$  other than Assumption 2.1. Let

$$\hat{\Sigma}_n = \begin{pmatrix} (-\hat{G}'_n)^{-1} & (\hat{G}'_n)^{-1} \hat{V}_{\alpha\beta,n} (\hat{G}'_n)^{-1} \\ \text{Sym.} & (\hat{G}'_n)^{-1} \hat{V}_{\beta\beta,n} (\hat{G}'_n)^{-1} \end{pmatrix},$$

where, for each  $v'_1, v'_2 \in \mathbb{R}^{p_\alpha}$  and  $v''_1, v''_2 \in \mathbb{R}^{p_\beta}$ ,

$$\hat{G}'_n{}^\alpha[v'_1, v'_2] := -\frac{1}{n} \sum_{j=1}^n V_{j-1}^{-1}(\hat{\beta}_n) [\partial_\alpha a_{j-1}(\hat{\alpha}_n)[v'_1], \partial_\alpha a_{j-1}(\hat{\alpha}_n)[v'_2]],$$

$$\hat{G}'_n{}^\beta[v''_1, v''_2] := -\frac{1}{n} \sum_{j=1}^n \text{trace}\{(V_{j-1}^{-1} \partial_\beta V_{j-1}) \otimes (V_{j-1}^{-1} \partial_\beta V_{j-1})\}(\hat{\beta}_n)[v''_1, v''_2],$$

$$\begin{aligned} \hat{V}_{\alpha\beta,n}[v'_1, v''_1] &:= -\sum_{j=1}^n \frac{1}{T_n} (V_{j-1}^{-1} \otimes \partial_\beta V_{j-1}^{-1})(\hat{\beta}_n) \\ &\quad \times [(\partial_\alpha a_{j-1}(\hat{\alpha}_n)[v'_1], \chi_j(\hat{\alpha}_n)), (v''_1, \chi_j(\hat{\alpha}_n)^{\otimes 2})], \end{aligned}$$

$$\begin{aligned} \hat{V}_{\beta\beta,n}[v''_1, v''_2] &:= \sum_{j=1}^n \frac{1}{T_n} (\partial_\beta V_{j-1}^{-1} \otimes \partial_\beta V_{j-1}^{-1})(\hat{\beta}_n) [(v''_1, \chi_j(\hat{\alpha}_n)^{\otimes 2}), (v''_2, \chi_j(\hat{\alpha}_n)^{\otimes 2})]. \end{aligned}$$

We will denote by  $\rightarrow^{\mathcal{L}}$  the weak convergence under  $P_0$ .

COROLLARY 2.8. *Under the conditions of Theorem 2.7, we have  $\hat{\Sigma}_n \rightarrow^P \Sigma_0$ , and hence*

$$(2.12) \quad \hat{\Sigma}_n^{-1/2} \sqrt{T_n}(\hat{\theta}_n - \theta_0) \rightarrow^{\mathcal{L}} \mathcal{N}_p(0, I_p)$$

holds true.

The proof of Corollary 2.8 is given in Section 4.2.

The primary objective of this paper is to derive the  $L^q(P_0)$ -boundedness of  $\sqrt{T_n}(\hat{\theta}_n - \theta_0)$  for every  $q > 0$ , for which the moment conditions [Assumptions 2.1 plus 2.3(b)] seem indispensable. Nevertheless, as pointed out by the anonymous referee, the existence of the moments of all orders is too much to ask in Corollary 2.8. Let us discuss a possibility of relaxing the moment condition in some detail; to make the exposition more clear, we here do not seek the greatest generality.

Clearly, the really necessary order (of  $J$ , hence  $X$  too) partly depends on the growth of the coefficients  $(a, b, c)$  and its partial derivatives with respect to  $\theta$ . We



for some  $W \in \mathbb{R}^p \otimes \mathbb{R}^m$  and  $\mathbb{R}^m$ -valued function  $g$  on  $\mathbb{R}^d \times \mathbb{R}^d \times \Theta$ . We would have a wide choice of  $W$  and  $g$ . When the conditional expectations involved do not admit closed forms, then the leading-term approximation of them via the Itô–Taylor expansion can be used. In view of this, as in Kessler [18], it would be formally possible to relax the condition  $nh_n^2 \rightarrow 0$  in (1.2) by gaining the order of the Itô–Taylor expansions of the conditional mean and conditional covariance,

$$\begin{aligned} E_\theta[X_{t_j}|\mathcal{F}_{t_{j-1}}] &= X_{t_{j-1}} + a_{j-1}(\alpha)h_n + \dots, \\ V_\theta[X_{t_j}|\mathcal{F}_{t_{j-1}}] &= V_{j-1}(\beta)h_n + \dots, \end{aligned}$$

which we have implicitly used up to the  $h_n$ -order terms to build  $\mathbb{Q}_n$  of (2.10). However, we then need specific moment structures of  $v$ , which appear in the higher orders of the above Itô–Taylor expansion. Moreover, we should note that the convergence rate  $\sqrt{T_n}$  can never be improved for both  $\alpha$  and  $\beta$ , even if  $E_\theta[X_{t_j}|\mathcal{F}_{t_{j-1}}]$  and  $V_\theta[X_{t_j}|\mathcal{F}_{t_{j-1}}]$  have closed forms, such as the case of linear drifts, so that the rate of  $h_n \rightarrow 0$  may not matter as long as  $T_n \rightarrow \infty$ . See also Remark 4.1.

- As was mentioned in the Introduction, the sampling points  $t_1, \dots, t_n$  may be irregularly spaced to some extent. Let  $0 \equiv t_0 < t_1 < \dots < t_n =: T_n$ , and put  $\Delta_j t := t_j - t_{j-1}$ . We claim that it is possible to remove the equidistance condition, while retaining that  $h_n := \max_{1 \leq j \leq n} \Delta_j t \rightarrow 0$ . We need the additional condition about asymptotic behavior of the spacing

$$(2.16) \quad \frac{1}{h_n} \min_{1 \leq j \leq n} \Delta_j t \rightarrow 1,$$

which obviously entails that  $T_n \sim nh_n$  (the ratio of both sides tends to 1). Then the same statements as in Theorem 2.7, Corollary 2.8 and Theorem 2.9 remain valid under (2.16). For this point, we only note that estimate (2.4) remains true even under (2.16): noting that

$$\begin{aligned} k_n &:= \max_{j \leq n} \left| \left( \frac{1}{n\Delta_j t} - \frac{1}{T_n} \right) n\Delta_j t \right| \\ &\leq \left( 1 - \frac{1}{h_n} \min_{j \leq n} \Delta_j t \right) + \left( \frac{nh_n}{T_n} - 1 \right) = o(1), \end{aligned}$$

we have, for any  $f$  such that both  $f$  and  $\partial f$  are of at most polynomial growth,

$$\begin{aligned} \delta_n &:= \left| \frac{1}{T_n} \int_0^{T_n} f(X_t) dt - \frac{1}{n} \sum_{j=1}^n f(X_{t_{j-1}}) \right| \\ &= \left| \sum_{j=1}^n \frac{1}{T_n} \int_{t_{j-1}}^{t_j} f(X_t) dt - \sum_{j=1}^n \frac{1}{n\Delta_j t} \int_{t_{j-1}}^{t_j} f(X_{t_{j-1}}) dt \right| \end{aligned}$$

$$\begin{aligned}
&\leq k_n \frac{1}{n} \sum_{j=1}^n \frac{1}{\Delta_j t} \int_{t_{j-1}}^{t_j} |f(X_t)| dt \\
&\quad + \frac{1}{n} \sum_{j=1}^n \frac{1}{\Delta_j t} \int_{t_{j-1}}^{t_j} |f(X_t) - f(X_{t_{j-1}})| dt \\
&\lesssim k_n \frac{1}{n} \sum_{j=1}^n \frac{1}{\Delta_j t} \int_{t_{j-1}}^{t_j} (1 + |X_t|)^C dt \\
&\quad + \frac{1}{n} \sum_{j=1}^n \frac{1}{\Delta_j t} \int_{t_{j-1}}^{t_j} (1 + |X_t|)^C |X_t - X_{t_{j-1}}| dt
\end{aligned}$$

for some  $C > 0$ . Therefore, Schwarz's inequality together with Lemma 4.5 leads to the estimate  $E_0[\delta_n] \lesssim k_n + \sqrt{h_n} = o(1)$ , enabling us to use  $n^{-1} \times \sum_{j=1}^n f(X_{t_{j-1}}) \rightarrow^p \int f(x) \pi_0(dx)$  as in the case of the equally-spaced sample. With this in mind, we can deduce the same estimates and limit results in the proofs given in Sections 4.2 to 4.3 in an entirely analogous way, the details being omitted.

### 2.3. Discussion.

2.3.1. *On the identifiability of the dispersion parameter.* Suppose that the coefficients  $b(x, \beta)$  and  $c(x, \beta)$  depend on  $\beta$  only through  $\beta_1$  and  $\beta_2$ , respectively, where  $\beta = (\beta_1, \beta_2)$ . On the one hand, it should be theoretically possible to identify  $\beta_1$  and  $\beta_2$  individually by the (intractable) likelihood function; for example, see Aït-Sahalia and Jacod [2] for the precise asymptotic behavior of the Fisher information matrix for  $\beta$  in case of univariate Lévy processes. We also refer to Aït-Sahalia and Jacod [1] for how to construct an asymptotically efficient estimator of  $\beta_1$  through the use of a truncated power-variation statistics, regarding  $\beta_2$  as a nuisance parameter. To perform individual estimation for more general diffusions with jumps, it is unadvised to resort to the likelihood based estimation. Instead, we may adopt a threshold-type estimator utilizing only relatively small (resp., large) increments of  $X$  for estimating  $\beta_1$  (resp.,  $\beta_2$ ), which makes it possible to extract information of the diffusion and jump parts separately, in compensation for a non-trivial fine tuning of the threshold; see Shimizu and Yoshida [38] and Ogihara and Yoshida [36] in case of compound-Poisson jumps and Shimizu [37] in the presence of infinitely many small jumps of finite variation.

On the other hand, our identifiability condition on  $\beta$  in Assumption 2.4 can be unfortunately stringent in the simultaneous presence of nondegenerate diffusion and jump components. Let us look at the assumption in the multiplicative-parameter case  $b(x, \beta) = \beta_1 b_0(x)$  and  $c(x, \beta) = \beta_2 c_0(x)$ , where  $b_0$  and  $c_0$  are

known positive functions and where we set  $d = r' = r'' = p_\beta = 1$  for simplicity; we implicitly suppose that the function equals 1 if it is constant because the constant then can be absorbed into  $\beta$ . Further, we here suppose that  $\bar{\Theta}_\beta \subset (0, \infty) \times (0, \infty)$ , so that  $X$  admits both nonnull diffusion and jump parts. Then direct computation gives  $\mathbb{G}_\infty^\beta(\beta) = M(\beta)[\beta_0 - \beta]$ , where

$$M(\beta) := \begin{pmatrix} 2\beta_1(\beta_{10} + \beta_1)I_{bb} & 2\beta_1(\beta_{20} + \beta_2)I_{bc} \\ 2\beta_2(\beta_{10} + \beta_1)I_{bc} & 2\beta_2(\beta_{20} + \beta_2)I_{cc} \end{pmatrix}$$

with  $I_{bb} := \int b_0^4(x)V^{-2}(x, \beta)\pi_0(dx)$ ,  $I_{bc} := \int b_0^2(x)c_0^2(x)V^{-2}(x, \beta)\pi_0(dx)$ , and  $I_{cc} := \int c_0^4(x)V^{-2}(x, \beta)\pi_0(dx)$ . We have  $|M(\beta)| = C(\beta)|I_{bb}I_{cc} - I_{bc}|$  for some constant  $C(\beta)$  depending on  $\beta$  such that  $\inf_\beta C(\beta) > 0$ , so that the identifiability condition on  $\beta$  is satisfied if  $|I_{bb}I_{cc} - I_{bc}| > 0$ . In view of Schwarz's inequality, we always have  $I_{bb}I_{cc} - I_{bc} \geq 0$ , the equality holding only when there exists an  $r \in \mathbb{R}$  such that  $b_0(x) = rc_0(x)$  for every  $x \in \mathbb{R}$ . That is, the GQMLE fails to be consistent as soon as  $b_0$  and  $c_0$  are proportional to each other; especially if both  $b_0$  and  $c_0$  are constant (hence 1, as was presupposed), then the GQMLE indeed cannot identify  $\beta_1$  and  $\beta_2$  individually, for there do exist infinitely many  $\beta = (\beta_1, \beta_2)$  such that

$$V(x, \beta) - V(x, \beta_0) = (\beta_1^2 + \beta_2^2) - (\beta_{10}^2 + \beta_{20}^2) = 0$$

for every  $x$ . This seems to be unavoidable as our contrast function  $\mathbb{M}_n$  is constructed solely based on fitting local conditional mean and covariance matrix. Although our estimation procedure cannot in general separate information of diffusion and jump variances, it should be noted that, when both  $b_0$  and  $c_0$  are constant, we may instead consistently estimate the "local variance"  $\beta_1^2 + \beta_2^2$ .

Finally, we remark that the identifiability condition " $|\mathbb{G}_\infty^\beta(\beta)|^2 \geq \chi_\beta|\beta - \beta_0|^2$ " becomes much simpler when we know that  $b(\cdot, \cdot) \equiv 0$  from the very beginning; then, in view of expression (2.6) and Assumption 2.2(b), it would suffice to have  $|\partial_\beta c^2(x, \beta)| > 0$  over a domain.

2.3.2. *On the asymptotic efficiency.* The efficiency issue for model (1.1) based on high-frequency sampling is a difficult problem and has been left unsolved over the years, which hinders us to do quantitative study on how much information loss occurs on our GQMLE; as a matter of fact, we do not know any Hajék bound on the asymptotic covariances especially when  $J$  is of infinite activity. This general issue is beyond the scope of this paper, but instead we give some remarks in this direction.

- Overall, the amount of efficiency loss in using our GQMLE may strongly depend on the structure of the jump part and on its relation to the possibly nondegenerate diffusion part; this would be a major drawback of our GQMLE. We do know the theoretical minimal asymptotic covariance matrix when  $X$  is a diffusion with compound-Poisson jumps with nondegenerate diffusion part, where,

in particular, the optimal rate of convergence in estimating  $\alpha$  is  $\sqrt{T_n}$ , achieved by our GQMLE  $\hat{\alpha}_n$ ; for details, see Shimizu and Yoshida [38] and Ogihara and Yoshida [36], as well as the references therein. In order to observe the effect of the jump part in estimation of  $\alpha$  in a concise way, let us look at the univariate  $X$  given by

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

where  $\alpha \in \mathbb{R}$ ,  $\inf_x b(x) \wedge \inf_x c(x) > 0$ , and  $J$  is a centered compound-Poisson process. The asymptotic variance of  $\hat{\alpha}_n$  is then given by the inverse of

$$-\mathbb{G}'_{\infty}{}^{\alpha}(\theta_0) = \int \{b^2(x) + c^2(x)\}^{-1} \{\partial_{\alpha} a(x, \alpha_0)\}^2 \pi_0(dx),$$

while the minimal asymptotic variance of the asymptotically efficient estimator is the inverse of  $A_0^* := \int b^{-2}(x) \{\partial_{\alpha} a(x, \alpha_0)\}^2 \pi_0(dx)$ . Hence, it would be natural to measure amount of efficiency loss in using  $\hat{\alpha}_n$  by the quantity

$$A_0^* - \{-\mathbb{G}'_{\infty}{}^{\alpha}(\theta_0)\} = \int \frac{\{\partial_{\alpha} a(x, \alpha_0)\}^2}{b^2(x)} \left( \frac{c^2(x)}{b^2(x) + c^2(x)} \right) \pi_0(dx).$$

From this expression, we may expect that the efficiency loss may be large (resp., not so significant) when the jump part is much larger (resp., smaller) compared with the diffusion part. This point comes into focus by looking at the Ornstein–Uhlenbeck process

$$dX_t = -\alpha_0 X_t dt + \beta_1 dW_t + \beta_2 dJ_t,$$

where  $\alpha_0, \beta_1, \beta_2 > 0$ . In this case, by means of the special relation  $m\alpha_0\kappa(m) = \kappa_Z(m)$  for  $m \in \mathbb{N}$ , where  $\kappa(m)$  and  $\kappa_Z(m)$ , respectively, denote the  $m$ th cumulants of  $\pi_0$  and  $\mathcal{L}(\beta_1 W_1 + \beta_2 J_1)$  (cf. Barndorff-Nielsen and Shephard [4]), we have

$$A_0^* - \{-\mathbb{G}'_{\infty}{}^{\alpha}(\theta_0)\} = \frac{\beta_2^2}{\beta_1^2(\beta_1^2 + \beta_2^2)} \int x^2 \pi_0(dx) = \frac{1}{2\alpha_0} \left( \frac{\beta_2}{\beta_1} \right)^2,$$

which becomes larger (resp., smaller) with increasing (resp., decreasing)  $\beta_2^2/\beta_1^2$ , the ratio of the jump-part variance to the diffusion-part one.

Furthermore, if  $X$  is supposed to be of pure-jump driven type (i.e.,  $b \equiv 0$ ) from the very beginning, the optimal rate of convergence in estimating  $\alpha$  may be faster than  $\sqrt{T_n}$ . For example, if  $X$  is the Ornstein–Uhlenbeck-type process  $dX_t = -\alpha X_t dt + dJ_t$  and if  $\mathcal{L}(h^{-1/\gamma} J_h)$  for small  $h$  behaves like the non-Gaussian  $\gamma$ -stable distribution [ $\gamma \in (0, 2)$ ], then the least absolute deviation (LAD)-type estimator has asymptotic normality at rate  $\sqrt{nh_n^{1-1/\gamma}}$ , which is faster than  $\sqrt{T_n} = \sqrt{nh_n}$ ; see Masuda [30] for details. Unfortunately, it is not clear whether or not it is possible to generalize the LAD-type estimation method to deal with  $X$  of (1.1) with nonlinear coefficients.

- Let us consider

$$(2.17) \quad dX_t = a(X_t, \alpha) dt + c(X_{t-}, \beta) dJ_t,$$

where  $J$  is a centered pure-jump Lévy process of infinite activity [i.e.,  $\nu(\mathbb{R}) = \infty$ ] such that  $E[J_1^2] = 1$ . Sometimes, a pure-jump Lévy process  $J$  can be approximated by a standard Wiener process if the parameter contained in the Lévy measure  $\nu(dz)$  behaves suitably; for instance,  $\mathcal{L}(J_1) \rightarrow \mathcal{N}_1(0, 1)$  as  $\delta \rightarrow \infty$  if  $\mathcal{L}(J_1)$  obeys the symmetric centered normal inverse-Gaussian distribution  $\text{NIG}(\delta, 0, \delta, 0)$ . Although the rate of convergence  $\sqrt{T_n}$  of our GQMLE  $\hat{\beta}_n$  can be never improved as long as we have a nonnull jump part, it is expected, in general, that if  $\mathcal{L}(J_1)$  in (2.17) gets “closer” to the normal distribution [i.e., if both  $|\nu(3)|$  and  $\nu(4)$  become small], our GQMLE will exhibit better performance; see Table 1 in Section 2.4 for some simulation results in this setting. As a matter of fact, Theorem 2.7 verifies that

$$\sup_{n \in \mathbb{N}} V_0[\sqrt{T_n}(\hat{\beta}_n - \beta_0)] \lesssim \nu(4).$$

[Recall that  $\mathbb{V}_{\beta\beta}$  depends on  $\nu(4)$  linearly.] It is worth mentioning that, even though  $\hat{\beta}_n$  is here  $\sqrt{T_n}$ -consistent,  $\sqrt{n}(\hat{\beta}_n - \beta_0)$  behaves like a tight sequence if  $\kappa_n := \nu(4)$  gets smaller as  $\kappa_n = O(h_n)$ .

2.4. *A numerical example.* For simulation purposes, we consider the following concrete model:

$$(2.18) \quad dX_t = \frac{-\alpha X_t}{\sqrt{1 + X_t^2}} dt + \sqrt{\beta} dJ_t, \quad X_0 = 0,$$

where the true value is  $(\alpha_0, \beta_0) = (1, 1)$ , the driving process is the normal inverse Gaussian Lévy process such that  $\mathcal{L}(J_t) = \text{NIG}(\delta, 0, \delta t, 0)$ , where  $\delta = 1, 10$  or  $20$ . It holds that  $E[J_t] = 0$ ,  $E[J_t^2] = t$ , and  $\mathcal{L}(J_t) \rightarrow \mathcal{N}(0, t)$  in total variation as  $\delta \rightarrow \infty$ , and that  $\nu(3) = 0$  and  $\nu(4) = 3/\delta^2$ . Model (2.18) is a normal-inverse Gaussian counterpart to the hyperbolic diffusion, for which  $J$  is replaced by a standard Wiener process. For this  $X$ , we can verify all the assumptions; see Proposition 5.4 for the verification of the stability conditions.

We simulated 1000 independent paths by Euler scheme with sufficiently fine step size to obtain 1000 independent estimates  $\hat{\theta}_n = (\hat{\alpha}_n, \hat{\alpha}_n)$ , and then computed their empirical mean and standard deviations.

Figure 1 shows typical sample paths of  $X$  for  $\delta = 1, 10$ , and  $20$ , with a diffusion corresponding to  $X$  with  $J$  replaced with a standard Wiener process, just for comparison.

Table 1 reports the results; just for comparison, we included the case of diffusion, where  $J$  is a standard Wiener process. From the table, we can observe the following:

- the performance of  $\hat{\alpha}_n$  are rather similar for all the three cases;



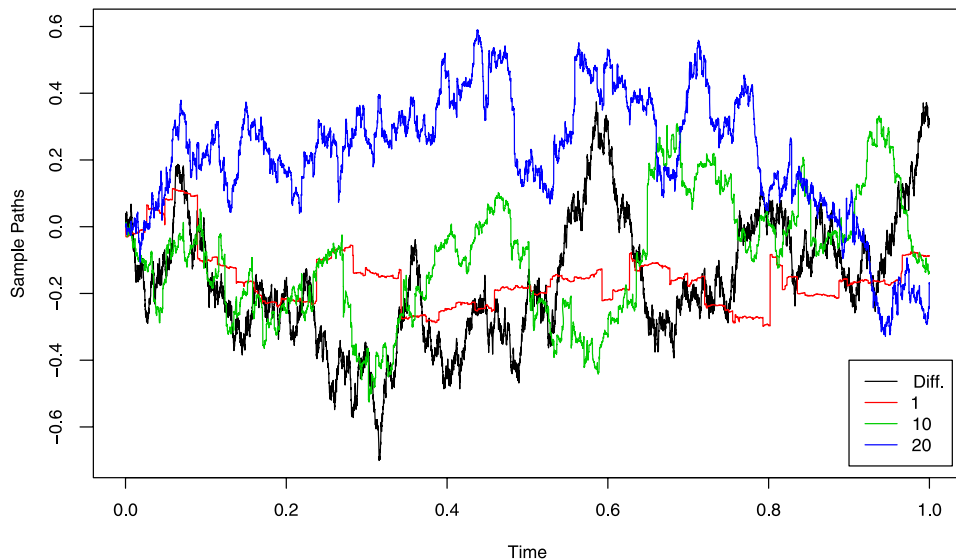


FIG. 1. Plots of sample paths of  $X$  of (2.18) for  $\delta = 1, 10$  and  $20$ , with a diffusion corresponding to  $X$  with  $J$  replaced by a standard Wiener process.

- the performance of  $\hat{\beta}_n$  gets better for larger  $\delta$ , which can be expected from the fact that the asymptotic variance of  $\hat{\beta}_n$  is a constant multiple of  $v(4) = 3\delta^{-2}$ ; we have  $\nabla_{\beta\beta} \rightarrow 0$  as  $\delta \rightarrow \infty$ .

**3. Mighty convergence of a class of continuous random fields.** In this section, we prove a fundamental result concerning the “single-norming” mighty con-

TABLE 1  
 Finite sample performance of  $\hat{\theta}_n$  concerning the model (2.18); just for comparison, the case of diffusion is also included. In each case, the sample mean is given with the sample standard deviation in parenthesis

$T_n$	$h_n$	Diffusion		$\delta = 1$		$\delta = 10$		$\delta = 20$	
		$\alpha$	$\beta$	$\alpha$	$\beta$	$\alpha$	$\beta$	$\alpha$	$\beta$
10	0.05	1.16 (0.63)	0.96 (0.10)	1.15 (0.62)	0.98 (0.58)	1.18 (0.65)	0.97 (0.11)	1.18 (0.65)	0.96 (0.10)
	0.01	1.19 (0.67)	0.99 (0.04)	1.17 (0.64)	0.97 (0.48)	1.21 (0.66)	0.99 (0.07)	1.19 (0.68)	0.99 (0.05)
100	0.05	1.00 (0.18)	0.97 (0.03)	1.00 (0.19)	0.98 (0.17)	1.00 (0.18)	0.97 (0.04)	1.01 (0.17)	0.97 (0.03)
	0.01	1.02 (0.18)	0.99 (0.01)	1.02 (0.19)	1.00 (0.17)	1.02 (0.18)	0.99 (0.02)	1.03 (0.19)	1.00 (0.02)

vergence of a continuous statistical random fields associated with general vector-valued estimating functions; here, the “single-norming” means that the rates of convergence are the same for all the arguments of the corresponding estimator. Theorem 3.5 below will serve as a fundamental tool in the proof of Theorem 2.7; the content of this section can be read independently of the main body.

To proceed, we need some notation. Denote by  $\{\mathcal{X}_n, \mathcal{A}_n, (P_\theta)_{\theta \in \Theta}\}_{n \in \mathbb{N}}$  underlying statistical experiments, where  $\Theta \subset \mathbb{R}^p$  is a bounded convex domain. Let  $\theta_0 \in \Theta$ , and write  $P_0 = P_{\theta_0}$ . Let  $\mathbb{G}_n = (\mathbb{G}_{j,n})_{j=1}^p : \mathcal{X}_n \times \Theta \rightarrow \mathbb{R}^p$  be vector-valued random functions; as usual, we will simply write  $\mathbb{G}_n(\theta)$ , dropping the argument of  $\mathcal{X}_n$ . Our target “contrast” function is

$$(3.1) \quad \mathbb{M}_n(\theta) := -\frac{1}{T_n} |\mathbb{G}_n(\theta)|^2,$$

where  $(T_n)$  is a nonrandom positive real sequence such that  $T_n \rightarrow \infty$ . The corresponding “ $M$ -estimator” is defined to be any measurable mapping  $\hat{\theta}_n : \mathcal{X}_n \rightarrow \overline{\Theta}$  such that

$$\hat{\theta}_n \in \operatorname{argmax}_{\theta \in \overline{\Theta}} \mathbb{M}_n(\theta).$$

Due to the compactness of  $\overline{\Theta}$  and the continuity of  $\mathbb{M}_n$  imposed later on, we can always find such a  $\hat{\theta}_n$ . The estimate  $\hat{\theta}_n$  can be any root of  $\mathbb{G}_n(\theta) = 0$  as soon as it exists.

Set  $U_n(\theta_0) := \{u \in \mathbb{R}^p : \theta_0 + T_n^{-1/2}u \in \Theta\}$  and define random fields  $\mathbb{Z}_n : U_n(\theta_0) \rightarrow (0, \infty)$  by

$$(3.2) \quad \mathbb{Z}_n(u) = \mathbb{Z}_n(u; \theta_0) := \exp\{\mathbb{M}_n(\theta_0 + T_n^{-1/2}u) - \mathbb{M}_n(\theta_0)\}.$$

Obviously, it holds that

$$\hat{u}_n := \sqrt{T_n}(\hat{\theta}_n - \theta_0) \in \operatorname{argmax}_{\theta \in \overline{\Theta}} \mathbb{Z}_n(\theta).$$

We consider the following two conditions for the random fields  $\mathbb{Z}_n$ .

- [Polynomial type Large Deviation Inequality (PLDI)]. For every  $M > 0$ , we have

$$(3.3) \quad \sup_{r>0} \left\{ r^M \sup_{n \in \mathbb{N}} P_0 \left[ \sup_{|u|>r} \mathbb{Z}_n(u) \geq e^{-r} \right] \right\} < \infty.$$

- (Weak convergence on compact sets). There exists a random field  $\mathbb{Z}_0(\cdot) = \mathbb{Z}_0(\cdot; \theta_0)$  such that  $\mathbb{Z}_n \rightarrow^{\mathcal{L}} \mathbb{Z}_0$  in  $\mathcal{C}(\overline{B(R)})$  for each  $R > 0$ , where  $\overline{B(R)} := \{u \in \mathbb{R}^p; |u| \leq R\}$ .

Under these conditions, the mode of convergence of  $\mathbb{Z}_n(\cdot)$  is mighty enough to deduce that the maximum-point sequence  $(\hat{u}_n)_n$  is  $L^q(P_0)$ -bounded for every  $q > 0$ ,

which especially implies that  $(\hat{u}_n)_n$  is tight: indeed, if (3.3) is in force,

$$\begin{aligned} \sup_{n \in \mathbb{N}} P_0[|\hat{u}_n| > r] &\leq \sup_{n \in \mathbb{N}} P_0\left[\sup_{|u|>r} \mathbb{Z}_n(u) \geq \mathbb{Z}_n(0)\right] \\ &= \sup_{n \in \mathbb{N}} P_0\left[\sup_{|u|>r} \mathbb{Z}_n(u) \geq 1\right] \lesssim \frac{1}{r^M} \end{aligned}$$

for every  $r > 0$ , so that

$$\sup_{n \in \mathbb{N}} E_0[|\hat{u}_n|^q] = \int_0^\infty \sup_{n \in \mathbb{N}} P_0[|\hat{u}_n| > s^{1/q}] ds \lesssim 1 + \int_1^\infty s^{-M/q} ds < \infty.$$

If  $u \mapsto \mathbb{Z}_0(u)$  is a.s. maximized at a unique point  $\hat{u}_\infty$ , then it follows from the tightness of  $(\hat{u}_n)_{n \in \mathbb{N}}$  that  $\hat{u}_n \xrightarrow{\mathcal{L}} \hat{u}_\infty$ ; let us remind the reader that the weak convergence on any compact set alone is not enough to deduce the weak convergence of  $\hat{u}_n$ , since  $U_n(\theta_0) \uparrow \mathbb{R}^p$  and we have no guarantee that  $(\hat{u}_n)$  is tight. Moreover, owing to the PLDI, the moment of  $f(\hat{u}_n)$  converges to that of  $f(\hat{u}_\infty)$  for every continuous function  $f$  on  $\mathbb{R}^p$  of at most polynomial growth. In our framework,  $\log \mathbb{Z}_0$  admits a quadratic structure with a normally distributed linear term and a nonrandom positive definite quadratic term, so that  $\hat{u}_\infty$  is asymptotically normally distributed.

We now introduce regularity conditions.

**ASSUMPTION 3.1 (Smoothness).** The functions  $\theta \mapsto \mathbb{G}_n(\theta)$  are continuously extended to the boundary of  $\Theta$ , and belong to  $\mathcal{C}^3(\Theta)$ ,  $P_0$ -a.s.

**ASSUMPTION 3.2 (Bounded moments).** For every  $K > 0$ ,

$$\sup_{n \in \mathbb{N}} E_0 \left[ \left| \frac{1}{\sqrt{T_n}} \mathbb{G}_n(\theta_0) \right|^K \right] + \max_{k \in \{0, 1, 2, 3\}} \sup_{n \in \mathbb{N}} E_0 \left[ \sup_{\theta \in \Theta} \left| \frac{1}{T_n} \partial_\theta^k \mathbb{G}_n(\theta) \right|^K \right] < \infty.$$

Let  $M > 0$  be a given constant.

**ASSUMPTION 3.3 (Limits).** (a) There exist a nonrandom function  $\mathbb{G}_\infty : \Theta \rightarrow \mathbb{R}^p$  and positive constants  $\chi = \chi(\theta_0)$  and  $\varepsilon$  such that:  $\mathbb{G}_\infty(\theta_0) = 0$ ;  $\sup_\theta |\mathbb{G}_\infty(\theta)| < \infty$ ;  $|\mathbb{G}_\infty(\theta)|^2 \geq \chi |\theta - \theta_0|^2$  for every  $\theta \in \Theta$ ; and

$$\sup_{n \in \mathbb{N}} E_0 \left[ \sup_{\theta \in \Theta} \left| \sqrt{T_n} \left( \frac{1}{T_n} \mathbb{G}_n(\theta) - \mathbb{G}_\infty(\theta) \right) \right|^{M+\varepsilon} \right] < \infty.$$

(b) There exists a nonrandom  $\mathbb{G}'_\infty(\theta_0) \in \mathbb{R}^p \otimes \mathbb{R}^p$  of rank  $p$  such that

$$\sup_{n \in \mathbb{N}} E_0 \left[ \left| \sqrt{T_n} \left( \frac{1}{T_n} \partial_\theta \mathbb{G}_n(\theta_0) - \mathbb{G}'_\infty(\theta_0) \right) \right|^M \right] < \infty.$$

ASSUMPTION 3.4 (Weak convergence).  $T_n^{-1/2}\mathbb{G}_n(\theta_0) \rightarrow^{\mathcal{L}} \mathcal{N}_p(0, \mathbb{V}(\theta_0))$  for some positive definite  $\mathbb{V}(\theta_0) \in \mathbb{R}^p \otimes \mathbb{R}^p$ .

Let  $\Sigma(\theta_0) := (\mathbb{G}'_\infty)^{-1}\mathbb{V}(\mathbb{G}'_\infty)^{-1\top}(\theta_0)$ . The main claim of this section is the following.

THEOREM 3.5. *Let  $M > 0$ .*

(a) *Suppose that Assumptions 3.1, 3.2 and 3.3 hold. Then the PLDI (3.3) holds true.*

(b) *If Assumption 3.4 is additionally met, then*

$$E_0[f(\hat{u}_n)] \rightarrow \int f(u)\phi(u; 0, \Sigma(\theta_0)) du$$

for every continuous function  $f: \mathbb{R}^p \rightarrow \mathbb{R}$  satisfying that  $\limsup_{|u| \rightarrow \infty} |u|^{-q} \times |f(u)| < \infty$  for some  $q \in (0, M)$ .

PROOF. Applying Taylor’s expansion to (3.2), we get

$$(3.4) \quad \log \mathbb{Z}_n(u) = \Delta_n(\theta_0)[u] - \frac{1}{2}\Gamma(\theta_0)[u, u] + \xi_n(u),$$

where  $\Delta_n(\theta_0) := T_n^{-1/2} \partial_\theta \mathbb{M}_n(\theta_0)$ ,  $\Gamma_n(\theta_0) := -T_n^{-1} \partial_\theta^2 \mathbb{M}_n(\theta_0)$ ,  $\Gamma(\theta_0) := 2\mathbb{G}'_\infty(\theta_0)^\top \mathbb{G}'_\infty(\theta_0)$  and

$$(3.5) \quad \begin{aligned} \xi_n(u) &:= \frac{1}{2} \{ \Gamma(\theta_0) - \Gamma_n(\theta_0) \} [u, u] \\ &\quad - \int_0^1 (1-s) \int \partial_\theta \Gamma_n(\theta_0 + stT_n^{-1/2}u) [sT_n^{-1/2}u, u^{\otimes 2}] dt ds. \end{aligned}$$

We will prove (a) by making use of Yoshida [48], Theorem 3(c). The task is then to verify conditions [A1''], [A4'], [A6], [B1] and [B2] of that paper. For convenience and clarity, we will list them in a reduced form with our notation. First we look at [B1] and [B2]:

[B1] the matrix  $\Gamma(\theta_0)$  is positive definite;

[B2] there exists a constant  $\chi > 0$  such that  $\mathbb{Y}(\theta) \leq -\chi^2|\theta - \theta_0|^2$  for each  $\theta \in \Theta$ .

Here  $\mathbb{Y}(\theta) := -|\mathbb{G}_\infty(\theta)|^2$ , where  $\mathbb{G}_\infty(\theta)$  is the one appearing in Assumption 3.3. Obviously, Assumption 3.3 assures [B1] and [B2] (the identifiability); in particular, we have the convergence  $T_n^{-1}\mathbb{M}_n(\theta) \rightarrow^p -|\mathbb{G}_\infty(\theta)|^2$  for each  $\theta \in \Theta$ , so that

$$\mathbb{Y}_n(\theta) := \frac{1}{T_n} \{ \mathbb{M}_n(\theta) - \mathbb{M}_n(\theta_0) \} = \frac{1}{T_n} \log \mathbb{Z}_n(\sqrt{T_n}(\theta - \theta_0)) \rightarrow^p \mathbb{Y}(\theta).$$

Next, given constants  $M > 0$  [the number in (3.3)] and  $\alpha \in (0, 1)$ , conditions [A6], [A1''] and [A4'] read as follows:

- [A6] (i)  $\sup_n E_0[|\Delta_n(\theta_0)|^{M_1}] < \infty$  for  $M_1 := M/(1 - \rho_1)$ .
- (ii)  $\sup_n E_0[\sup_{\theta} |T_n^{1/2-\beta_2}(\mathbb{Y}_n(\theta) - \mathbb{Y}(\theta))|^{M_2}] < \infty$ , for  $M_2 := M/(1 - 2\beta_2 - \rho_2)$ .
- [A1''] (i)  $\sup_n E_0[\sup_{\theta} |T_n^{-1} \partial_{\theta}^3 \mathbb{M}_n(\theta)|^{M_3}] < \infty$  for  $M_3 := M/\{\alpha/(1 - \alpha) - \rho_1\}$ .
- (ii)  $\sup_n E_0[|T_n^{\beta_1}(\Gamma_n(\theta_0) - \Gamma(\theta_0))|^{M_4}] < \infty$  for  $M_4 := M/\{2\beta_1/(1 - \alpha) - \rho_1\}$ .
- [A4'] The parameters  $\alpha, \beta_1, \beta_2, \rho_1$  and  $\rho_2$  fulfil the inequalities

$$0 < \beta_1 < 1/2, \quad 0 < \rho_1 < \min\left(1, \frac{\alpha}{1 - \alpha}, \frac{2\beta_1}{1 - \alpha}\right),$$

$$2\alpha < \rho_2, \quad \beta_2 \geq 0, \quad 1 - 2\beta_2 - \rho_2 > 0.$$

These conditions involve several “moment-index” parameters to be controlled, which do not seem straightforward to handle. Nevertheless, under our assumptions we can provide a rather simplified version. Instead of “[A1’], [A4’] and [A6]” we will verify the following “[A1’#] and [A6#]”:

- [A1’#] (i)  $\sup_n E_0[\sup_{\theta} |T_n^{-1} \partial_{\theta}^3 \mathbb{M}_n(\theta)|^K] < \infty$  for every  $K > 0$ .
- (ii)  $\sup_n E_0[|\sqrt{T_n}(\Gamma_n(\theta_0) - \Gamma(\theta_0))|^{M-\varepsilon_1}] < \infty$  for every  $\varepsilon_1 > 0$  small enough.
- [A6#] (i)  $\sup_n E_0[|\Delta_n(\theta_0)|^K] < \infty$  for every  $K > 0$ .
- (ii)  $\sup_n E_0[\sup_{\theta} |\sqrt{T_n}(\mathbb{Y}_n(\theta) - \mathbb{Y}(\theta))|^{M+\varepsilon/2}] < \infty$ , for  $\varepsilon$  given in Assumption 3.3.

Let us show that “[A1’#] and [A6#]” imply “[A1’], [A4’] and [A6].” First, by [A1’#](i) and [A6#](i), the numbers  $M_1$  and  $M_3$  can be arbitrarily large, so that we may in particular take  $\alpha$  and  $\rho_1$  arbitrarily small (i.e., nearly zero). Then we have [A1’](i) and [A6](i). Next, we note that in [A1’#](ii) the exponent of “ $T_n$ ” is 1/2, hence we may let  $\beta_2$  be sufficiently close to 1/2. Then, taking  $\alpha$  and  $\rho_1$  small enough with  $\rho_1 < \alpha/(1 - \alpha)$ , we can obtain the first two inequalities in [A4’]. Next, in view of [A6#](ii), we can take  $\beta_2 = 0$  and  $\rho_2$  small enough to make [A6](ii) and the last three ones in [A4’] valid. Finally, as for  $M_4$ , we note that a suitable control of  $(\alpha, \rho_1, \beta_1)$  leads to

$$\frac{2\beta_1}{1 - \alpha} - \rho_1 = 1 + \left(\frac{\alpha}{1 - \alpha} - \rho_1\right) + \frac{2\beta_1 - 1}{1 - \alpha} > 1,$$

so that [A1’](ii) follows. In sum, under “[A1’#] and [A6#]”, we can pick  $\rho_1, \rho_2, \alpha \approx 0$  and  $\beta_2 = 0$ , and then  $\beta_1 \approx 1/2$ , in order to make all of “[A1’], [A4’] and [A6]” valid. Thus we are left to proving [A1’#] and [A6#] above.

We begin with [A1’#]. Since  $|T_n^{-1} \partial_{\theta}^3 \mathbb{M}_n(\theta)| \lesssim |T_n^{-1} \mathbb{G}_n(\theta)| |T_n^{-1} \partial_{\theta}^3 \mathbb{G}_n(\theta)| + |T_n^{-1} \partial_{\theta} \mathbb{G}_n(\theta)| |T_n^{-1} \partial_{\theta}^2 \mathbb{G}_n(\theta)|$ , we have for every  $K > 0$ ,

$$\sup_{n \in \mathbb{N}} E_0 \left[ \sup_{\theta \in \Theta} \left| \frac{1}{T_n} \partial_{\theta}^3 \mathbb{M}_n(\theta) \right|^K \right] < \infty.$$

Noting that  $\partial_{\theta_i} \partial_{\theta_j} \mathbb{M}_n = -2T_n^{-1} \{ \partial_{\theta_i} \partial_{\theta_j} \mathbb{G}_n[\mathbb{G}_n] + \partial_{\theta_i} \mathbb{G}_n[\partial_{\theta_j} \mathbb{G}_n] \}$ , we also have

$$\begin{aligned} & \sqrt{T_n} |\Gamma_n(\theta_0) - \Gamma(\theta_0)| \\ & \lesssim \left| \frac{1}{\sqrt{T_n}} \mathbb{G}_n(\theta_0) \right| \left| \frac{1}{T_n} \partial_{\theta}^2 \mathbb{G}_n(\theta_0) \right| \\ & \quad + \left( |\Gamma(\theta_0)| + \left| \frac{1}{T_n} \partial_{\theta} \mathbb{G}_n(\theta_0) \right| \right) \left| \sqrt{T_n} \left( \frac{1}{T_n} \partial_{\theta} \mathbb{G}_n(\theta_0) - \mathbb{G}'_{\infty}(\theta_0) \right) \right|. \end{aligned}$$

Therefore, Assumptions 3.2 and 3.3 combined with Hölder’s inequality yield that for  $\varepsilon_1 \in (0, M)$ ,

$$\begin{aligned} & \sup_{n \in \mathbb{N}} E_0 [ |\sqrt{T_n} (\Gamma_n(\theta_0) - \Gamma(\theta_0))|^{M-\varepsilon_1} ] \\ & \lesssim 1 + \left\{ \sup_{n \in \mathbb{N}} E_0 \left[ \left| \sqrt{T_n} \left( \frac{1}{T_n} \partial_{\theta} \mathbb{G}_n(\theta_0) - \mathbb{G}'_{\infty}(\theta_0) \right) \right|^M \right] \right\}^{(M-\varepsilon_1)/M} < \infty. \end{aligned}$$

Thus [A1''#] follows.

Next we prove [A6#]. Statement (i) is obvious from Assumption 3.2,

$$(3.6) \quad \sup_{n \in \mathbb{N}} E_0 [ |\Delta_n(\theta_0)|^K ] \lesssim \sup_{n \in \mathbb{N}} E_0 \left[ \left| \frac{1}{T_n} \partial_{\theta} \mathbb{G}_n(\theta_0) \right|^K \left| \frac{1}{\sqrt{T_n}} \mathbb{G}_n(\theta_0) \right|^K \right] < \infty.$$

Using the estimate

$$\begin{aligned} & |\sqrt{T_n} (\mathbb{Y}_n(\theta) - \mathbb{Y}(\theta))| \\ & \leq \frac{1}{\sqrt{T_n}} \left| \frac{1}{\sqrt{T_n}} \mathbb{G}_n(\theta) \right|^2 \\ & \quad + \left( |\mathbb{G}_{\infty}(\theta)| + \left| \frac{1}{T_n} \mathbb{G}_n(\theta) \right| \right) \left| \sqrt{T_n} \left( \frac{1}{T_n} \mathbb{G}_n(\theta) - \mathbb{G}_{\infty}(\theta) \right) \right|, \end{aligned}$$

it follows under Assumptions 3.2 and 3.3 that

$$\begin{aligned} & \sup_{n \in \mathbb{N}} E_0 \left[ \sup_{\theta \in \Theta} |\sqrt{T_n} (\mathbb{Y}_n(\theta) - \mathbb{Y}(\theta))|^{M+\varepsilon/2} \right] \\ & \lesssim 1 + \sup_{n \in \mathbb{N}} E_0 \left[ \sup_{\theta \in \Theta} \left| \sqrt{T_n} \left( \frac{1}{T_n} \mathbb{G}_n(\theta) - \mathbb{G}_{\infty}(\theta) \right) \right|^{M+\varepsilon} \right]^{(M+\varepsilon/2)/(M+\varepsilon)} < \infty. \end{aligned}$$

Thus [A6#] is ensured, and the proof of (a) is complete.

We now turn to the proof of (b). Fix any  $R > 0$ . Since we know that the sequence  $(\hat{u}_n)$  is  $L^q(P_0)$ -bounded for each  $q \in (0, M)$  and that the set  $\operatorname{argmax}_u \log \mathbb{Z}_{\infty}(u)$  a.s. consists of the only point

$$\hat{u}_{\infty} := \Gamma(\theta_0)^{-1} \Delta_{\infty}(\theta_0) \sim \mathcal{N}_p(0, \Sigma(\theta_0)),$$

it suffices to show that  $\log \mathbb{Z}_n \rightarrow^{\mathcal{L}} \log \mathbb{Z}_\infty$  in  $\mathcal{C}(\overline{B(R)})$ , where

$$\begin{aligned} \log \mathbb{Z}_\infty(u) &:= \Delta_\infty(\theta_0)[u] - \frac{1}{2} \Gamma(\theta_0)[u, u], \\ \Delta_\infty(\theta_0) &\sim \mathcal{N}_p(0, 4\mathbb{G}'_\infty(\theta_0)^\top \mathbb{V}(\theta_0) \mathbb{G}'_\infty(\theta_0)) \end{aligned}$$

(e.g., Yoshida [48], Theorem 5). We have  $T_n^{-1} \partial_\theta \mathbb{G}_n(\theta_0) \rightarrow^p \mathbb{G}'_\infty(\theta_0)$  from Assumption 3.3, hence Slutsky’s lemma and Assumption 3.4 imply that

$$\Delta_n(\theta_0) = -\frac{2}{T_n} \partial_\theta \mathbb{G}_n(\theta_0) \left[ \frac{1}{\sqrt{T_n}} \mathbb{G}_n(\theta_0) \right] \rightarrow^{\mathcal{L}} \Delta_\infty(\theta_0).$$

Also, we have

$$(3.7) \quad |\xi_n(u)| \lesssim |u|^2 |\Gamma_n(\theta_0) - \Gamma(\theta_0)| + \frac{|u|^3}{\sqrt{T_n}} \sup_{\theta \in \Theta} \left| \frac{1}{T_n} \partial_\theta^3 \mathbb{M}_n(\theta) \right| = o_p(1)$$

for every  $u \in \overline{B(R)}$ . Thus, recalling expression (3.4), we get  $\log \mathbb{Z}_n(u) \rightarrow^{\mathcal{L}} \log \mathbb{Z}_0(u)$  for every  $u \in \overline{B(R)}$ , and moreover, due to the linearity in  $u$  of the weak convergence term  $\Delta_n(\theta_0)[u]$ , the Cramér–Wold device ensures the finite-dimensional convergence. Therefore, it remains to check the tightness of  $\{\log \mathbb{Z}_n(u)\}_{u \in \overline{B(R)}}$ . In view of the classical Kolmogorov tightness criterion for continuous random fields (e.g., Kunita [21], Theorem 1.4.7), it suffices to show that there exists a constant  $\gamma > p (= \dim \Theta)$  such that

$$(3.8) \quad \sup_{|u| \leq R} \sup_{n \in \mathbb{N}} E_0[|\log \mathbb{Z}_n(u)|^\gamma] + \sup_{n \in \mathbb{N}} E_0 \left[ \sup_{|u| \leq R} |\partial_u \log \mathbb{Z}_n(u)|^\gamma \right] < \infty.$$

In view of the estimates in (3.6) and (3.7) as well as the expressions (3.4) and (3.5),

$$\begin{aligned} &\sup_{u \in \overline{B(R)}} \sup_{n \in \mathbb{N}} E_0[|\log \mathbb{Z}_n(u)|^\gamma] \\ &\lesssim \sup_{n \in \mathbb{N}} E_0[|\Delta_n(\theta_0)|^\gamma] + 1 + \sup_{u \in \overline{B(R)}} \sup_{n \in \mathbb{N}} E_0[|\xi_n(u)|^\gamma] \\ &\lesssim 1 + E_0[|\Gamma_n(\theta_0) - \Gamma(\theta_0)|^\gamma] + \sup_{n \in \mathbb{N}} E_0 \left[ \sup_{\theta \in \Theta} \left| \frac{1}{T_n} \partial_\theta^3 \mathbb{M}_n(\theta) \right|^\gamma \right] < \infty. \end{aligned}$$

Furthermore, since

$$\begin{aligned} \partial_u \log \mathbb{Z}_n(u) &= \partial_u \left\{ \mathbb{M}_n \left( \theta_0 + \frac{1}{\sqrt{T_n}} u \right) - \mathbb{M}_n(\theta_0) \right\} \\ &= \frac{1}{\sqrt{T_n}} \partial_\theta \mathbb{M}_n \left( \theta_0 + \frac{1}{\sqrt{T_n}} u \right) \\ &= \frac{1}{\sqrt{T_n}} \left\{ \partial_\theta \mathbb{M}_n(\theta_0) + \frac{1}{\sqrt{T_n}} \int_0^1 \partial_\theta^2 \mathbb{M}_n \left( \theta_0 + \frac{s}{\sqrt{T_n}} u \right) [u] ds \right\}, \end{aligned}$$

the finiteness of  $\sup_n E_0[\sup_{|u| \leq R} |\partial_u \log \mathbb{Z}_n(u)|^\gamma]$  follows on applying Assumption 3.2 to the estimate

$$\begin{aligned} & \sup_{|u| \leq R} |\partial_u \log \mathbb{Z}_n(u)| \\ & \lesssim \left| \frac{1}{\sqrt{T_n}} \mathbb{G}_n(\theta_0) \right| \left| \frac{1}{T_n} \partial_\theta \mathbb{G}_n(\theta_0) \right| + \sup_{\theta \in \Theta} \left| \frac{1}{T_n} \partial_\theta^2 \mathbb{M}_n(\theta) \right| \\ & \lesssim \left| \frac{1}{\sqrt{T_n}} \mathbb{G}_n(\theta_0) \right| \left| \frac{1}{T_n} \partial_\theta \mathbb{G}_n(\theta_0) \right| \\ & \quad + \sup_{\theta \in \Theta} \left\{ \left| \frac{1}{T_n} \mathbb{G}_n(\theta) \right| \left| \frac{1}{T_n} \partial_\theta^2 \mathbb{G}_n(\theta) \right| + \left| \frac{1}{T_n} \partial_\theta \mathbb{G}_n(\theta) \right|^2 \right\}. \end{aligned}$$

Thus we have obtained (3.8), thereby achieving the proof of (b).  $\square$

REMARK 3.6. We have confined ourselves to the “single-norming (i.e., scalar- $T_n$ )” case for the squared quasi-score function. Nevertheless, as in the original formulation of Yoshida [48], Theorem 1, it would be also possible to deal with “multi-norming” cases where elements of  $\hat{\theta}_n$  possibly converge at different rates, that is, cases of a matrix norming instead of the scalar norming  $\sqrt{T_n}$ . This would require somewhat more complicated arguments, but we do not need such an extension in this paper.

**4. Proofs of Theorem 2.7 and Corollary 2.8.**

4.1. *Proof of Theorem 2.7.* The proof of Theorem 2.7 is achieved by applying Theorem 3.5. When we have a reasonable estimating function  $\theta \mapsto \mathbb{G}_n(\theta)$  with which an estimator of  $\theta$  is defined by a random root of the estimating equation  $\mathbb{G}_n(\theta) = 0$ , it may be unclear what is the “single” associated contrast function to be maximized or minimized; for example, it would be often the case when  $\mathbb{G}_n$  is constructed via a kind of (conditional-) moment fittings. The setup (4.3) below provides a way of converting the situation from  $Z$ -estimation to  $M$ -estimation.

4.1.1. *Introductory remarks.* At first glance, it seems that, in order to investigate the asymptotic behavior of  $\hat{\theta}_n$ , we may proceed as in the case of diffusions, expanding the GQL  $\mathbb{Q}_n$  of (2.10) and then investigating asymptotic behaviors of the derivatives  $\partial_\theta^k \mathbb{Q}_n$ ; see Yoshida [48], Section 6, for details. Following this route, however, leads to an inconvenience, essentially due to the fact that  $(h_n^{-1/2} \Delta_j X)_{j \leq n}$  is not  $L^q(P_0)$ -bounded for  $q > 2$ . To see this more precisely, let us take a brief look at the simple one-dimensional Lévy process  $X_t = \alpha t + \sqrt{\beta} J_t$ , with  $\theta = (\alpha, \beta) \in \mathbb{R} \times (0, \infty)$  and  $\mathcal{L}(J_1)$  admitting finite moments. In this case,



$$\begin{aligned} \mathbb{Q}_n(\theta) &= -\sum_j \{(\log \beta) + (\beta h_n)^{-1}(\Delta_j X - \alpha h_n)^2\}, \\ \partial_\alpha \mathbb{Q}_n(\theta) &= \sum_{j=1}^n \frac{2}{\beta} (\Delta_j X - \alpha h_n), \\ \partial_\beta \mathbb{Q}_n(\theta) &= \sum_{j=1}^n \frac{1}{\beta^2 h_n} \{(\Delta_j X - \alpha h_n)^2 - \beta h_n\}, \\ \partial_\alpha^2 \mathbb{Q}_n(\theta) &= \frac{-2T_n}{\beta}, \quad \partial_\alpha \partial_\beta \mathbb{Q}_n(\theta) = -\sum_{j=1}^n \frac{2}{\beta^2} (\Delta_j X - \alpha h_n), \\ \partial_\beta^2 \mathbb{Q}_n(\theta) &= -\sum_{j=1}^n \frac{2}{\beta^3 h_n} \left\{ (\Delta_j X - \alpha h_n)^2 - \frac{\beta h_n}{2} \right\}. \end{aligned}$$

We can deduce the convergences

$$\begin{aligned} \frac{1}{T_n} \partial_\alpha^2 \mathbb{Q}_n(\theta_0) &\rightarrow^P -2\beta_0^{-1}, & \frac{1}{\sqrt{n}\sqrt{T_n}} \partial_\alpha \partial_\beta \mathbb{Q}_n(\theta_0) &\rightarrow^P 0, \\ \frac{1}{n} \partial_\beta^2 \mathbb{Q}_n(\theta_0) &\rightarrow^P -\beta_0^{-2}, \end{aligned}$$

so that the normalized quasi observed-information matrix  $-D_n^{-1} \partial_\theta^2 \mathbb{Q}_n(\theta_0) D_n^{-1} \rightarrow^P \text{diag}(2\beta_0^{-1}, \beta_0^{-2})$ , where  $D_n := \text{diag}(\sqrt{T_n}, \sqrt{n})$ . In view of the classical Cramér-type method for  $M$ -estimation, we should then have a central limit theorem for the normalized quasi-score  $\{T_n^{-1/2} \partial_\alpha \mathbb{Q}_n(\theta_0), n^{-1/2} \partial_\beta \mathbb{Q}_n(\theta_0)\}$  for an asymptotic normality at rate  $D_n$  to be valid for the  $M$ -estimator associated with  $\mathbb{Q}_n$ . However, different from the drifted Wiener process, the sequence  $\{n^{-1/2} \partial_\beta \mathbb{Q}_n(\theta_0)\}$  does *not* converge because  $(h_n^{-1/2} \Delta_j X)_{j \leq n}$  cannot be  $L^q$ -bounded for large  $q > 2$  as can be seen from the moment structure of Lévy processes; see Luschgy and Pagès [24] for general moment estimates in small time with several concrete examples. Although we only mentioned the Lévy process with diagonal norming, the situation remains the same even when  $X$  is actually an ergodic solution to (1.1).

The observation made in the last paragraph says that the situation is different from the case of diffusions, when developing asymptotic theory concerning the Gaussian quasi-likelihood for model (1.1) under high-frequency sampling framework; it is also different from the case of time series models, where the usual  $\sqrt{n}$ -consistency holds in most cases (see the references cited in the [Introduction](#)). Earlier attempts to tackle this point have been made by Mancini [25], Shimizu and Yoshida [38], Ogihara and Yoshida [36], where they incorporated jump-detection filters in defining a contrast function. The filter approach has its own advantage such as  $\sqrt{n}$ -rate estimation of the diffusion parameter, even in the presence of jumps; however, we should have in mind that its implementation involves fine-tuning parameters, thereby possibly preventing us from a straightforward use of the approach.

In order to prove Theorem 2.7, we will look at not  $\theta \mapsto \mathbb{Q}_n(\theta)$ , but

$$\theta \mapsto \mathbb{G}_n(\theta) = \{\mathbb{G}_n^\alpha(\theta), \mathbb{G}_n^\beta(\theta)\},$$

where  $\mathbb{G}_n^\alpha : \Theta \rightarrow \mathbb{R}^{p_\alpha}$  and  $\mathbb{G}_n^\beta : \Theta \rightarrow \mathbb{R}^{p_\beta}$  are defined by

$$(4.1) \quad \mathbb{G}_n^\alpha(\theta) = \sum_{j=1}^n \partial_\alpha a_{j-1}(\alpha) [V_{j-1}^{-1}(\beta) [\chi_j(\alpha)]],$$

$$(4.2) \quad \mathbb{G}_n^\beta(\theta) = \sum_{j=1}^n \left( \{-\partial_\beta V_{j-1}^{-1}(\beta)\} [\chi_j(\alpha)^{\otimes 2}] - h_n \frac{\partial_\beta |V_{j-1}(\beta)|}{|V_{j-1}(\beta)|} \right).$$

Our contrast function  $\mathbb{M}_n(\theta)$  is then defined to be the “squared quasi-score” as in (3.1),

$$(4.3) \quad \mathbb{M}_n(\theta) = -\frac{1}{T_n} |\mathbb{G}_n(\theta)|^2.$$

Trivially,  $\mathbb{G}_n : \Theta \rightarrow \mathbb{R}^p$  fulfil that  $\mathbb{G}_n(\theta) = \{(1/2)\partial_\alpha \mathbb{Q}_n(\theta), h_n \partial_\beta \mathbb{Q}_n(\theta)\}$ . The difference is that we put the factor “ $h_n$ ” in front of  $\partial_\beta \mathbb{Q}_n(\theta)$ ; our estimating procedure is formally not the usual  $M$ -estimation based on the Taylor expansion of  $\theta \mapsto \mathbb{Q}_n(\theta)$  around  $\theta_0$ , but rather a kind of minimum distance estimation concerning the Gaussian quasi-score function. The optimization with respect to  $\theta$  is asymptotically the same for both of  $\mathbb{Q}_n$  and  $\mathbb{M}_n$ : if there is no root  $\theta \in \Theta$  for  $\mathbb{G}_n(\theta) = 0$ , then we may assign any value (e.g., any element of  $\Theta$ ) to  $\hat{\theta}_n$ , upholding the claim of Theorem 2.7.

REMARK 4.1. More general cases than (4.1) and (4.2) can be treated, such as

$$\begin{aligned} \mathbb{G}_n^\alpha(\theta) &= \sum_{j=1}^n \bar{W}_{j-1}^\alpha(\theta) \{X_{t_j} - m_{j-1}(\theta)\}, \\ \mathbb{G}_n^\beta(\theta) &= \sum_{j=1}^n (\bar{W}_{j-1}^{\beta,1}(\theta) [\{X_{t_j} - m_{j-1}(\theta)\}^{\otimes 2}] - h_n \bar{W}_{j-1}^{\beta,2}(\theta)) \end{aligned}$$

for some measurable  $m : \mathbb{R}^d \times \Theta \rightarrow \mathbb{R}^d$ ,  $\bar{W}^\alpha : \mathbb{R}^d \times \Theta \rightarrow \mathbb{R}^{p_\alpha} \otimes \mathbb{R}^d$ ,  $\bar{W}^{\beta,1} : \mathbb{R}^d \times \Theta \rightarrow \mathbb{R}^{p_\beta} \otimes (\mathbb{R}^d \otimes \mathbb{R}^d)$  and  $\bar{W}^{\beta,2} : \mathbb{R}^d \times \Theta \rightarrow \mathbb{R}^{p_\beta}$ . This may be called a GQMLE as well, for we are still solely fitting the local mean vectors and covariance matrices. This setting allows us to deal with, for example, the parametric model

$$dX_t = a(X_t, \theta) dt + b(X_t, \theta) dW_t + c(X_{t-}, \theta) dJ_t$$

with possibly degenerate  $b$  and  $c$ , the resulting GQMLE  $\hat{\theta}_n$  still being asymptotically normal at rate  $\sqrt{T_n}$  under suitable conditions. To avoid unnecessarily messy notation and regularity conditions without losing essence, we have decided to treat (1.1) in this paper.

For later use, we here introduce some convention and recall a couple of basic facts that we will make use often without notice:

- We will often suppress “ $(\theta_0)$ ” from the notation:  $\chi_j := \chi_j(\alpha_0)$ ,  $a_{j-1} := a_{j-1}(\alpha_0)$ ,  $\mathbb{G}_n^\alpha = \mathbb{G}_n^\alpha(\theta_0)$ , and so forth.
- $f_j$  denotes a shorthand for  $f_{t_{j-1}}^{t_j}$ .
- $M'_{j-1}(\theta) := \partial_\alpha a_{j-1}(\alpha)^\top V_{j-1}^{-1}(\beta) \in \mathbb{R}^{p_\alpha} \otimes \mathbb{R}^d$ .
- $M''_{j-1}(\beta) := -\partial_\beta V_{j-1}^{-1}(\beta) = \{V_{j-1}^{-1}(\partial_\beta V_{j-1})V_{j-1}^{-1}\}(\beta) \in \mathbb{R}^{p_\beta} \otimes \mathbb{R}^d \otimes \mathbb{R}^d$ .
- $d_{j-1}(\beta) := |V_{j-1}(\beta)|^{-1} \partial_\beta |V_{j-1}(\beta)| \in \mathbb{R}^{p_\beta}$ .
- Given real sequence  $a_n$  and random variables  $Y_n$  possibly depending on  $\theta$ , we write  $Y_n = O_p^*(a_n)$  if  $\sup_{n,\theta} E_0[|a_n^{-1} Y_n|^K] < \infty$  for every  $K > 0$ .
- $E_0^{j-1}[\cdot] := E_0[\cdot | \mathcal{F}_{t_{j-1}}]$ .
- $R$  denotes a generic function on  $\mathbb{R}^d$ , possibly depending on  $n$  and  $\theta$ , for which there exists a constant  $C \geq 0$  such that  $\sup_{n,\theta} |R(x)| \leq C(1 + |x|)^C$  for every  $x \in \mathbb{R}^d$ .
- Burkholder’s inequality: for a martingale difference array  $(\zeta_{nj})_{j \leq n}$  and every  $q \geq 2$ ,

$$E_0 \left[ \max_{k \leq n} \left| \sum_{j \leq k} \frac{1}{\sqrt{n}} \zeta_{nj} \right|^q \right] \lesssim E_0 \left[ \left( \frac{1}{n} \sum_{j \leq n} \zeta_{nj}^2 \right)^{q/2} \right] \lesssim \frac{1}{n} \sum_{i \leq n} E[|\zeta_{nj}|^q].$$

Moreover, if  $b$ . and  $c$ . are sufficiently integrable adapted processes, then

$$E_0 \left[ \left| \int_0^T b_{s-} dW_s \right|^q \right] \lesssim T^{q/2-1} \int_0^T E_0[|b_s|^q] ds,$$

$$E_0 \left[ \left| \int_0^T c_{s-} dJ_s \right|^q \right] \lesssim (1 \vee T)^{q/2-1} \int_0^T E_0[|c_s|^q] ds$$

for every  $T > 0$  and  $q \geq 2$  such that  $E[|J_1|^q] < \infty$ .

- Sobolev’s inequality (e.g., Friedman [10], Section 10.2),

$$E_0 \left[ \sup_{\theta \in \Theta} |u(\theta)|^q \right] \lesssim \sup_{\theta \in \Theta} \{ E_0[|u(\theta)|^q] + E_0[|\partial_\theta u(\theta)|^q] \}$$

for  $q > p$  and a random field  $u \in \mathcal{C}^1(\Theta)$ ; recall that  $p$  denotes the dimension of  $\theta$  and that we are presupposing the boundedness and convexity of  $\Theta$ . We will make use of this type of inequality to derive some uniform-in- $\theta$  moment estimates for martingale terms.

We now turn to the proof of Theorem 2.7 by verifying the conditions of Theorem 3.5.

4.1.2. *Verification of the conditions on  $\mathbb{G}_n$ .* We rewrite  $\mathbb{G}_n$  as follows:

$$(4.4) \quad \mathbb{G}_n^\alpha(\theta) = \sum_{j=1}^n M'_{j-1}(\theta)[\chi_j] - h_n \sum_{j=1}^n M'_{j-1}(\theta)[a_{j-1}(\alpha) - a_{j-1}],$$

$$(4.5) \quad \begin{aligned} \mathbb{G}_n^\beta(\theta) &= \sum_{j=1}^n \{M''_{j-1}(\beta)[\chi_j^{\otimes 2}] - h_n d_{j-1}(\beta)\} \\ &\quad + 2h_n \sum_{j=1}^n M''_{j-1}(\beta)[\chi_j, a_{j-1} - a_{j-1}(\alpha)] \\ &\quad + h_n^2 \sum_{j=1}^n M''_{j-1}(\beta)[\{a_{j-1} - a_{j-1}(\alpha)\}^{\otimes 2}]. \end{aligned}$$

We have

$$(4.6) \quad \chi_j = \zeta_j + r_j,$$

where

$$(4.7) \quad \zeta_j := \int_j \tilde{a}_{j-1}(s) ds + \int_j b(X_s, \beta_0) dW_s + \int_j c(X_{s-}, \beta_0) dJ_s,$$

$$(4.8) \quad r_j := \int_j \{E_0^{j-1}[a(X_s, \alpha_0)] - a_{j-1}\} ds,$$

with  $\tilde{a}_{j-1}(s) := a(X_s, \alpha_0) - E_0^{j-1}[a(X_s, \alpha_0)]$ . Obviously,  $(\zeta_j)_{j \leq n}$  forms a martingale difference array with respect to the discrete-time filtration  $(\mathcal{F}_{t_j})_{j \leq n}$ .

Itô's formula and the present integrability condition lead to

$$(4.9) \quad E_0^{j-1}[a(X_s, \alpha_0)] - a_{j-1} = \int_j E_0^{j-1}[\mathcal{A}a(X_u, \alpha_0)] du = h_n R_{j-1},$$

where  $\mathcal{A}$  denotes the (extended) generator associated with  $X$  under  $P_0$ , that is, for  $f \in \mathcal{C}^2(\mathbb{R}^d)$

$$\begin{aligned} \mathcal{A}f(x) &= \partial f(x)[a(x, \alpha_0)] + \frac{1}{2} \partial^2 f(x)[b(x, \beta_0)^{\otimes 2}] \\ &\quad + \int \{f(x + c(x, \beta_0)z) - f(x) - \partial f(x)[c(x, \beta_0)z]\} \nu(dz). \end{aligned}$$

Putting (4.8) and (4.9) together gives  $r_j = h_n^2 R_{j-1}$ , therefore

$$(4.10) \quad \chi_j = \zeta_j + h_n^2 R_{j-1}.$$

Assumption 3.1 obviously holds under the present differentiability conditions. We begin with verifying Assumption 3.2.

LEMMA 4.2. *For every  $K > 0$ , we have*

$$\sup_{n \in \mathbb{N}} E_0 \left[ \left| \frac{1}{\sqrt{T_n}} \mathbb{G}_n(\theta_0) \right|^K \right] + \sup_{n \in \mathbb{N}} E_0 \left[ \sup_{\theta \in \Theta} \left| \frac{1}{T_n} \mathbb{G}_n(\theta) \right|^K \right] < \infty.$$

PROOF. By substituting (4.10) in (4.4) and (4.5) and then rearranging the resulting terms, we have

$$\begin{aligned} \mathbb{G}_n^\alpha(\theta) &= \sum_{j=1}^n M'_{j-1}(\theta) \zeta_j + h_n \sum_{j=1}^n M'_{j-1}(\theta) \{a_{j-1} - a_{j-1}(\alpha)\} \\ (4.11) \quad &+ h_n^2 \sum_{j=1}^n M'_{j-1}(\theta) R_{j-1}, \end{aligned}$$

$$\begin{aligned} \mathbb{G}_n^\beta(\theta) &= \sum_{j=1}^n \{M''_{j-1}(\beta) [\zeta_j^{\otimes 2}] - h_n d_{j-1}(\beta)\} \\ (4.12) \quad &+ 2h_n \sum_{j=1}^n M''_{j-1}(\beta) [\zeta_j, a_{j-1} - a_{j-1}(\alpha)] + h_n^2 \sum_{j=1}^n R_{j-1}. \end{aligned}$$

To achieve the proof, we will separately look at  $T_n^{-1/2} \mathbb{G}_n^\alpha$ ,  $T_n^{-1/2} \mathbb{G}_n^\beta$ ,  $T_n^{-1} \mathbb{G}_n^\alpha(\theta)$  and  $T_n^{-1} \mathbb{G}_n^\beta(\theta)$ . Fix any integer  $K > (2 \vee p)$  in the sequel.

First we prove  $T_n^{-1/2} \mathbb{G}_n^\alpha = O_p^*(1)$ . Observe that

$$\begin{aligned} \frac{1}{\sqrt{T_n}} \mathbb{G}_n^\alpha &= \sum_{j=1}^n \frac{1}{\sqrt{T_n}} M'_{j-1} \zeta_j + \sqrt{T_n h_n^2} \frac{1}{n} \sum_{j=1}^n M'_{j-1} R_{j-1} \\ &= \sum_{j=1}^n \frac{1}{\sqrt{T_n}} M'_{j-1} \zeta_j + O_p^*(\sqrt{T_n h_n^2}). \end{aligned}$$

By (4.7),

$$\begin{aligned} \sum_{j=1}^n \frac{1}{\sqrt{T_n}} M'_{j-1} \zeta_j &= \sum_{j=1}^n \frac{1}{\sqrt{n}} \left( M'_{j-1} \frac{1}{\sqrt{h_n}} \int_j b(X_s, \beta_0) dW_s \right) \\ (4.13) \quad &+ \sqrt{h_n} \sum_{j=1}^n \frac{1}{\sqrt{n}} \left( M'_{j-1} \frac{1}{h_n} \int_j \tilde{a}_{j-1}(s) ds \right) \\ &+ \sum_{j=1}^n \frac{1}{\sqrt{T_n}} M'_{j-1} \int_j c(X_{s-}, \beta_0) dJ_s. \end{aligned}$$

Burkholder’s inequality implies that the first and second term on the right-hand side are  $O_p^*(1)$  and  $O_p^*(\sqrt{h_n})$ , respectively. As for the last term, by writing

$\mathbf{1}_j : (0, \infty) \rightarrow \{0, 1\}$  for the identity function of the interval  $(t_{j-1}, t_j]$ ,

$$\begin{aligned}
 & E_0 \left[ \left| \sum_{j=1}^n \frac{1}{\sqrt{T_n}} M'_{j-1} \int_j c(X_{s-}, \beta_0) dJ_s \right|^K \right] \\
 & \lesssim T_n^{-K/2} E_0 \left[ \left| \int_0^{T_n} \sum_{j=1}^n \mathbf{1}_j(s) M'_{j-1} c(X_{s-}, \beta_0) dJ_s \right|^K \right] \\
 (4.14) \quad & \lesssim T_n^{-K/2} T_n^{K/2-1} \int_0^{T_n} E_0 \left[ \left( \sum_{j=1}^n \mathbf{1}_j(s) |M'_{j-1} c(X_{s-}, \beta_0)| \right)^K \right] ds \\
 & = \frac{1}{T_n} \int_0^{T_n} \sum_{j=1}^n \mathbf{1}_j(s) E_0 [ |M'_{j-1} c(X_{s-}, \beta_0)|^K ] ds \\
 & \lesssim \frac{1}{T_n} \sum_{j=1}^n \int_j ds = 1,
 \end{aligned}$$

and hence we are done.

We now prove  $T_n^{-1/2} \mathbb{G}_n^\beta = O_p^*(1)$ . In the sequel, we may and do suppose that  $d = p_\beta = r' = r'' = 1$ : this reduction is possible because of the polarization identity

$$[S', S''] = \frac{1}{4} ([S' + S''] - [S' - S'']),$$

which is valid for any two semimartingales  $S'$  and  $S''$ . By (4.10) and (4.5),

$$\frac{1}{\sqrt{T_n}} \mathbb{G}_n^\beta = \sum_{j=1}^n \frac{1}{\sqrt{T_n}} (M''_{j-1} \zeta_j^2 - h_n d_{j-1}) + O_p^*(\sqrt{T_n h_n^2}),$$

so that it remains to verify

$$(4.15) \quad \sum_{j=1}^n \frac{1}{\sqrt{T_n}} M''_{j-1} (\zeta_j^2 - h_n V_{j-1}) = O_p^*(1).$$

Define  $\zeta_j(t)$  for  $t \in (t_{j-1}, t_j]$  by

$$\begin{aligned}
 \zeta_j(t) &= \int_{t_{j-1}}^t \tilde{a}_{j-1}(s) ds + \int_{t_{j-1}}^t b(X_s, \beta_0) dW_s \\
 &\quad + \int_{t_{j-1}}^t c(X_{s-}, \beta_0) dJ_s.
 \end{aligned}$$

Let  $N(ds, dz)$  denote the Poisson random measure associated with  $J$ , and  $\tilde{N}$  its compensated version [i.e.,  $J_t = \int_0^t \int z \tilde{N}(ds, dz)$ ]. The quadratic variation at time  $t$

is then given as follows (cf. Jacod and Shiryaev [15], I.4.49(d), I.4.55(b)):

$$\begin{aligned} [\zeta_j(\cdot)]_t &= \int_{t_{j-1}}^t b^2(X_{s-}, \beta_0) ds + \int_{t_{j-1}}^t \int c^2(X_{s-}, \beta_0) z^2 N(ds, dz) \\ &= (t - t_{j-1})V_{j-1} + \int_{t_{j-1}}^t \int c^2(X_{s-}, \beta_0) \tilde{N}(ds, dz) + \int_{t_{j-1}}^t g_{j-1}(s) ds, \end{aligned}$$

where we used the assumption  $\int z^2 \nu(dz) = 1$  (with the temporary assumption  $r'' = 1$ ) and  $g_{j-1}(s) := b^2(X_s, \beta_0) - b_{j-1}^2 + c^2(X_{s-}, \beta_0) - c_{j-1}^2$ . Applying the integration-by-parts formula, we get

$$\begin{aligned} \zeta_j^2 - h_n V_{j-1} &= \left\{ 2 \int_j \zeta_j(s-) d\zeta_j(s) + \int_j \int c^2(X_{s-}, \beta_0) z^2 \tilde{N}(ds, dz) \right. \\ &\quad \left. + \int_j (g_{j-1}(s) - E_0^{j-1}[g_{j-1}(s)]) ds \right\} \\ &\quad + \int_j E_0^{j-1}[g_{j-1}(s)] ds \\ &=: \zeta_j^{(0)} + \zeta_j^{(1)} \quad \text{say.} \end{aligned}$$

We can deduce that  $\sum_{j=1}^n T_n^{-1/2} M''_{j-1} \zeta_j^{(0)} = O_p^*(1)$ , as is the case in the proof of  $\sum_{j=1}^n T_n^{-1/2} M'_{j-1} \zeta_j = O_p^*(1)$  via the expression (4.13). Moreover, we can apply Itô's formula to get  $\zeta_j^{(1)} = h_n^2 R_{j-1}$  under the  $C^2$  property of  $x \mapsto (b(x, \beta_0), c(x, \beta_0))$ , from which it follows that  $\sup_n E_0[|\sum_{j=1}^n T_n^{-1/2} M''_{j-1} \times \zeta_j^{(1)}|^K] \lesssim \sup_n (T_n h_n^2)^{K/2} < \infty$ . We thus get (4.15).

Let us turn to prove  $\sup_\theta |T_n^{-1} \mathbb{G}_n^\alpha(\theta)| = O_p^*(1)$ . In the same way as in the proof of  $T_n^{-1/2} \mathbb{G}_n^\alpha = O_p^*(1)$ , we can prove  $\sum_{j=1}^n T_n^{-1/2} M'_{j-1}(\theta) \zeta_j = O_p^*(T_n^{-1/2})$  for each  $\theta \in \Theta$ , since the explicit dependence on  $\theta$  is only through the predictable parts  $M'_{j-1}(\theta)$ ; similar arguments will apply in some places below. Therefore, it follows from (4.11) that, for each  $\theta \in \Theta$ ,

$$\begin{aligned} \frac{1}{T_n} \mathbb{G}_n^\alpha(\theta) &= \frac{1}{\sqrt{T_n}} \left( \sum_{j=1}^n \frac{1}{\sqrt{T_n}} M'_{j-1}(\theta) \zeta_j \right) + h_n \left( \frac{1}{n} \sum_{j=1}^n M'_{j-1}(\theta) R_{j-1} \right) \\ &\quad + \frac{1}{n} \sum_{j=1}^n M'_{j-1}(\theta) \{a_{j-1} - a_{j-1}(\alpha)\} \\ (4.16) \quad &= O_p^* \left( \frac{1}{\sqrt{T_n}} \vee h_n \right) + \frac{1}{n} \sum_{j=1}^n M'_{j-1}(\theta) \{a_{j-1} - a_{j-1}(\alpha)\} \\ &= O_p^* \left( \frac{1}{\sqrt{T_n}} \right) + \frac{1}{n} \sum_{j=1}^n M'_{j-1}(\theta) \{a_{j-1} - a_{j-1}(\alpha)\}, \end{aligned}$$

so that  $T_n^{-1}\mathbb{G}_n^\alpha(\theta) = O_p^*(1)$ . In a quite similar manner, we obtain [see (4.32) and (4.33) below]

$$(4.17) \quad \begin{aligned} & \frac{1}{T_n} \partial_\theta \mathbb{G}_n^\alpha(\theta) \\ &= O_p^* \left( \frac{1}{\sqrt{T_n}} \right) + \frac{1}{n} \sum_{j=1}^n \partial_\theta [M'_{j-1}(\theta) \{a_{j-1} - a_{j-1}(\alpha)\}] = O_p^*(1). \end{aligned}$$

Therefore, we arrive at  $\sup_\theta |T_n^{-1}\mathbb{G}_n^\alpha(\theta)| = O_p^*(1)$  by means of the Sobolev inequality.

It remains to prove  $\sup_\theta |T_n^{-1}\mathbb{G}_n^\beta(\theta)| = O_p^*(1)$ ; we remind the reader that we are supposing that  $d = p_\beta = r' = r'' = 1$ . As in the proof of (4.15), we can prove

$$\sum_{j=1}^n \frac{1}{\sqrt{T_n}} \partial_\theta^k M''_{j-1}(\beta) (\zeta_j^2 - h_n V_{j-1}) = O_p^*(1)$$

for each  $k = 0, 1$  and  $\beta$ , so that the Sobolev inequality gives  $\sum_{j=1}^n T_n^{-1/2} \times M''_{j-1}(\beta) (\zeta_j^2 - h_n V_{j-1}) = O_p^*(1)$ . Therefore, it follows from (4.12) and simple manipulation that

$$(4.18) \quad \begin{aligned} \frac{1}{T_n} \mathbb{G}_n^\beta(\theta) &= \frac{1}{\sqrt{T_n}} \left( \sum_{j=1}^n \frac{1}{\sqrt{T_n}} M''_{j-1}(\beta) (\zeta_j^2 - h_n V_{j-1}) \right) \\ &\quad + \frac{2\sqrt{T_n}}{n} \sum_{j=1}^n \frac{1}{\sqrt{T_n}} M''_{j-1}(\beta) \{a_{j-1} - a_{j-1}(\alpha)\} \zeta_j \\ &\quad + \frac{h_n}{n} \sum_{j=1}^n R_{j-1} + \frac{1}{n} \sum_{j=1}^n M''_{j-1}(\beta) \{V_{j-1} - V_{j-1}(\beta)\} \\ &= O_p^* \left( \frac{1}{\sqrt{T_n}} \vee \frac{\sqrt{T_n}}{n} \vee h_n \right) + \frac{1}{n} \sum_{j=1}^n M''_{j-1}(\beta) \{V_{j-1} - V_{j-1}(\beta)\} \\ &= O_p^* \left( \frac{1}{\sqrt{T_n}} \right) + \frac{1}{n} \sum_{j=1}^n M''_{j-1}(\beta) \{V_{j-1} - V_{j-1}(\beta)\}. \end{aligned}$$

Thus  $T_n^{-1}\mathbb{G}_n^\beta(\theta) = O_p^*(1)$ . Quite similarly, we get  $T_n^{-1} \partial_\theta \mathbb{G}_n^\beta(\theta) = O_p^*(1)$ ,

$$(4.19) \quad \begin{aligned} & \frac{1}{T_n} \partial_\theta \mathbb{G}_n^\beta(\theta) \\ &= O_p^* \left( \frac{1}{\sqrt{T_n}} \right) + \frac{1}{n} \sum_{j=1}^n \partial_\theta [M''_{j-1}(\beta) \{V_{j-1} - V_{j-1}(\beta)\}] = O_p^*(1), \end{aligned}$$



completing the proof.  $\square$

Next we turn to verifying the uniform moment estimates in Assumptions 3.3. To this end, we prove a preliminary lemma.

LEMMA 4.3. *Suppose the following conditions:*

- the measurable function  $f : \mathbb{R}^d \times \Theta \rightarrow \mathbb{R}$  fulfils that  $\theta \mapsto f(x, \theta)$  is differentiable for each  $x$  and that

$$g(x) := \sup_{\theta \in \Theta} \{ |f(x, \theta)| \vee |\partial_\theta f(x, \theta)| \}$$

is of at most polynomial growth;

- there exist a probability measure  $\pi_0$  and a constant  $a > 0$  such that  $\|P_t(x, \cdot) - \pi_0(\cdot)\|_g \lesssim e^{-at} g(x)$ ;
- $\sup_t E_0[|X_t|^q] < \infty$  for every  $q > 0$ .

Then, for every  $K > 0$  we have

$$\sup_{n \in \mathbb{N}} E_0 \left[ \sup_{\theta \in \Theta} \left| \sqrt{T_n} \left( \frac{1}{n} \sum_{j=1}^n f_{j-1}(\theta) - \int f(x, \theta) \pi_0(dx) \right) \right|^K \right] < \infty.$$

PROOF. Put  $n^{-1} \sum_{j=1}^n f_{j-1}(\theta) - \int f(x, \theta) \pi_0(dx) = \Lambda'_n(f; \theta) + \Lambda''_n(f; \theta)$ , where  $\Lambda'_n(f; \theta) := n^{-1} \sum_{j=1}^n \{f_{j-1}(\theta) - E_0[f_{j-1}(\theta)]\}$  and  $\Lambda''_n(f; \theta) := n^{-1} \times \sum_{j=1}^n \{E_0[f_{j-1}(\theta)] - \int f(x, \theta) \pi_0(dx)\}$ . Under the present assumptions, we can apply Yoshida [48], Lemma 4, to get  $E_0[|\partial_\theta^k \Lambda'_n(f; \theta)|^K] \lesssim T_n^{-K/2} + T_n^{1-K} \lesssim T_n^{-K/2}$  for  $k \in \{0, 1\}$  and  $K \geq 2$ , yielding that  $\max_{k=0,1} \sup_\theta \sup_n E_0[|\sqrt{T_n} \partial_\theta^k \Lambda'_n(f; \theta)|^K] < \infty$ . The Sobolev inequality then gives

$$\sup_{n \in \mathbb{N}} E_0 \left[ \sup_{\theta \in \Theta} |\sqrt{T_n} \Lambda'_n(f; \theta)|^K \right] < \infty.$$

As for  $\Lambda''_n(f; \theta)$ , we have for  $k \in \{0, 1\}$ ,

$$\begin{aligned} & |\sqrt{T_n} \partial_\theta^k \Lambda''_n(f; \theta)| \\ &= \left| \frac{\sqrt{T_n}}{n} \sum_{j=1}^n \left( \iint \partial_\theta^k f(y, \theta) P_{t_{j-1}}(x, dy) \eta(dx) \right. \right. \\ &\quad \left. \left. - \iint \partial_\theta^k f(y, \theta) \pi_0(dy) \eta(dx) \right) \right| \\ &= \left| \frac{\sqrt{T_n}}{n} \sum_{j=1}^n \int \left( \int \partial_\theta^k f(y, \theta) \{P_{t_{j-1}}(x, dy) - \pi_0(dy)\} \eta(dx) \right) \right| \end{aligned}$$

$$\begin{aligned} &\leq \frac{\sqrt{T_n}}{n} \sum_{j=1}^n \int \|P_{t_{j-1}}(x, \cdot) - \pi_0(\cdot)\|_g \eta(dx) \\ &\lesssim \frac{\sqrt{T_n}}{n} \sum_{j=1}^n \exp(-at_{j-1}) \lesssim \frac{1}{\sqrt{T_n}}. \end{aligned}$$

This completes the proof.  $\square$

COROLLARY 4.4. *Assumption 3.3(a) holds true.*

PROOF. Again we may and do suppose that  $d = p_\beta = r' = r'' = 1$ . Recalling (4.16), (4.17), (4.18) and (4.19), we apply Lemma 4.3 with  $f(x, \theta) = M'(x, \theta)\{a(x, \alpha_0) - a(x, \alpha)\}$  and  $f(x, \theta) = M''(x, \beta)\{V(x, \beta_0) - V(x, \beta)\}$  to conclude

$$\sup_{n \in \mathbb{N}} E_0 \left[ \sup_{\theta \in \Theta} \left| \sqrt{T_n} \left( \frac{1}{T_n} \mathbb{G}_n(\theta) - \mathbb{G}_\infty(\theta) \right) \right|^K \right] < \infty$$

for every  $K > 0$ , where  $\mathbb{G}_\infty(\theta) := (\mathbb{G}_\infty^\alpha(\theta), \mathbb{G}_\infty^\beta(\theta))$  are given by (2.5) and (2.6), the integrals in which are finite by the assumptions. Trivially  $\mathbb{G}_\infty(\theta_0) = 0$ , and Assumption 3.3(a) is verified with  $\chi = \chi_\alpha \wedge \chi_\beta$ .  $\square$

Let us mention the fundamental fact concerning conditional size of  $X$ 's increments. For the convenience of reference we include a sketch of the proof.

LEMMA 4.5. *Let  $g(x) := |a(x, \alpha_0)| \vee |b(x, \beta_0)| \vee |c(x, \beta_0)|$ , and fix any  $q \geq 2$  such that  $E[|J_t|^q] < \infty$ . Then*

$$E_0^{j-1} \left[ \sup_{s \in [t_{j-1}, t_j]} |X_s - X_{t_{j-1}}|^q \right] \lesssim \begin{cases} h_n^{q/2} g^q(X_{t_{j-1}}), & \text{if } c \equiv 0, \\ h_n g^q(X_{t_{j-1}}), & \text{otherwise.} \end{cases}$$

*In particular, the left-hand side is essentially bounded if so is  $g$ .*

PROOF. Let  $c \not\equiv 0$ . Given a constant  $M > 0$ , we let  $\tau_{j-1, M} := \inf\{s \geq t_{j-1} : |X_s| \geq M\}$  and  $\xi_{j-1, M}(s) := E_0^{j-1}[\sup\{|X_u - X_{t_{j-1}}|^q : u \in [t_{j-1}, s \wedge \tau_{j-1, M}]\}]$ . We can make use of the Lipschitz property of the coefficients and Masuda [27], Lemma E.1, to derive  $\xi_{j-1, M}(t_j) \lesssim \int_{t_{j-1}}^{t_j} \xi_{j-1, M}(s) ds + h_n g^q(X_{t_{j-1}})$ , the upper bound being  $P_0$ -a.s. finite according to the definition of  $\tau_{j-1, M}$ . Hence the claim follows on applying Gronwall's inequality and then letting  $M \uparrow \infty$ . The case of  $c \equiv 0$  is similar.  $\square$

We now prove the central limit theorem required in Assumption 3.4.

LEMMA 4.6. *We have*

$$(4.20) \quad \frac{1}{\sqrt{T_n}} \mathbb{G}_n(\theta_0) \rightarrow^{\mathcal{L}} \mathcal{N}_p(0, \mathbb{V}(\theta_0)),$$

where  $\mathbb{V}(\theta_0)$  is given by (2.11).

PROOF. We begin with extracting the leading martingale terms of the sequences  $T_n^{-1/2} \mathbb{G}_n^\alpha$  and  $T_n^{-1/2} \mathbb{G}_n^\beta$ ; recall the expressions (4.11) and (4.12). Let us rewrite (4.7) as

$$(4.21) \quad \zeta_j = m_j + r'_j,$$

where

$$m_j := b_{j-1} \Delta_j W + c_{j-1} \Delta_j J,$$

$$r'_j := \int_j \tilde{a}_{j-1}(s) ds + \int_j (b(X_s, \beta_0) - b_{j-1}) dW_s + \int_j (c(X_{s-}, \beta_0) - c_{j-1}) dJ_s.$$

We claim that it suffices to prove that

$$(4.22) \quad \sum_{j=1}^n \frac{1}{\sqrt{T_n}} \begin{pmatrix} \tilde{\gamma}_j^\alpha \\ \tilde{\gamma}_j^\beta \end{pmatrix} \rightarrow^{\mathcal{L}} \mathcal{N}_p(0, \mathbb{V}(\theta_0)),$$

where  $\tilde{\gamma}_j^\alpha := M'_{j-1} m_j$  and  $\tilde{\gamma}_j^\beta := M''_{j-1} [m_j^{\otimes 2}] - h_n d_{j-1}$ , both of which form martingale difference arrays with respect to  $(\mathcal{F}_{t_j})_{j \leq n}$ ; we can verify that  $E_0^{j-1}[\tilde{\gamma}_j^\beta[u]] = 0$  for each  $u \in \mathbb{R}^{p\beta}$ , making use of the identity  $\text{trace}\{A(x)^{-1} \partial_x \times A(x)\} = \partial_x |A(x)| / |A(x)|$  for a differentiable square-matrix function  $A$ . In fact, recalling what we have seen in the proof of Lemma 4.2, we observe the following:

- We have

$$\begin{aligned} \frac{1}{\sqrt{T_n}} \mathbb{G}_n^\alpha &= \sum_{j=1}^n \frac{1}{\sqrt{T_n}} M'_{j-1} \left( \int_j b(X_s, \beta_0) dW_s + \int_j c(X_{s-}, \beta_0) dJ_s \right) + o_p(1) \\ &= \sum_{j=1}^n \frac{1}{\sqrt{T_n}} \tilde{\gamma}_j^\alpha + \sum_{j=1}^n \frac{1}{\sqrt{T_n}} M'_{j-1} \int_j (b(X_s, \beta_0) - b_{j-1}) dW_s \\ &\quad + \sum_{j=1}^n \frac{1}{\sqrt{T_n}} M'_{j-1} \int_j (c(X_{s-}, \beta_0) - c_{j-1}) dJ_s + o_p(1). \end{aligned}$$

By means of Burkholder’s inequality and Lemma 4.5 combined with the conditioning argument,

$$\begin{aligned} E_0 \left[ \left| \sum_{j=1}^n \frac{1}{\sqrt{T_n}} M'_{j-1} \int_j (b(X_s, \beta_0) - b_{j-1}) dW_s \right|^2 \right] \\ \lesssim E_0 \left[ \sum_{j=1}^n \frac{1}{T_n} |M'_{j-1}|^2 |R_{j-1}| \int_j h_n ds \right] \lesssim h_n. \end{aligned}$$

Following the same line as in (4.14), we also get

$$E_0 \left[ \left| \sum_{j=1}^n T_n^{-1/2} M'_{j-1} \int_j (c(X_s, \beta_0) - c_{j-1}) dJ_s \right|^2 \right] \lesssim h_n.$$

Therefore, it follows that

$$(4.23) \quad \frac{1}{\sqrt{T_n}} \mathbb{G}_n^\alpha = \sum_{j=1}^n \frac{1}{\sqrt{T_n}} \tilde{\gamma}_j^\alpha + o_p(1).$$

- Put  $B'_n = 2 \sum_{j=1}^n T_n^{-1/2} M''_{j-1}[m_j, r'_j]$  and  $B''_n = \sum_{j=1}^n T_n^{-1/2} M''_{j-1}[r'_j, r'_j]$ , then

$$\begin{aligned} \frac{1}{\sqrt{T_n}} \mathbb{G}_n^\beta &= \sum_{j=1}^n \frac{1}{\sqrt{T_n}} (M''_{j-1}[\zeta_j^{\otimes 2}] - h_n d_{j-1}) + o_p(1) \\ &= \sum_{j=1}^n \frac{1}{\sqrt{T_n}} \tilde{\gamma}_j^\beta + B'_n + B''_n + o_p(1). \end{aligned}$$

Since  $\sup_{j \leq n} E_0[|r'_j|^q] \lesssim h_n^2$  for every  $q \geq 2$  and  $E_0^{j-1}[|m_j|^2] \lesssim |R_{j-1}|^2 h_n$ , the Cauchy–Schwarz inequality leads to

$$\begin{aligned} E_0[|B'_n|] &\lesssim \frac{1}{n} \sum_{j=1}^n \sqrt{\frac{n}{h_n}} E_0[|R_{j-1}|^2 E_0^{j-1}[|m_j|^2]]^{1/2} E_0[|r'_j|^2]^{1/2} \\ &\lesssim \sqrt{nh_n^2} \rightarrow 0. \end{aligned}$$

Moreover, for any  $\varepsilon \in (0, 1/3)$ , Hölder’s inequality gives

$$\begin{aligned} E_0[|B''_n|] &\lesssim \frac{1}{n} \sum_{j=1}^n \sqrt{\frac{n}{h_n}} E_0[|R_{j-1}| |r'_j|^2] \\ &\lesssim \frac{1}{n} \sum_{j=1}^n \sqrt{\frac{n}{h_n}} E_0[|R_{j-1}|^{(1+\varepsilon)/\varepsilon}]^{\varepsilon/(1+\varepsilon)} E_0[|r'_j|^{2(1+\varepsilon)}]^{1/(1+\varepsilon)} \\ &\lesssim \frac{1}{n} \sum_{j=1}^n \sqrt{\frac{n}{h_n}} E_0[|r'_j|^{2(1+\varepsilon)}]^{1/(1+\varepsilon)} \lesssim \sqrt{nh_n^{4/(1+\varepsilon)-1}} \\ &\lesssim \sqrt{nh_n^2} \rightarrow 0. \end{aligned}$$

Hence we have derived

$$(4.24) \quad \frac{1}{\sqrt{T_n}} \mathbb{G}_n^\beta = \sum_{j=1}^n \frac{1}{\sqrt{T_n}} \tilde{\gamma}_j^\beta + o_p(1).$$

Having (4.23) and (4.24) in hand, it remains to verify (4.22). We are going to apply the classical martingale central limit theorem (e.g., Dvoretzky [7]).

Put  $\tilde{\gamma}_j = (\tilde{\gamma}_j^\alpha, \tilde{\gamma}_j^\beta)$ . It is easy to verify the Lyapunov condition: in fact, we have  $E_0^{j-1}[|\tilde{\gamma}_j|^K] \lesssim h_n |R_{j-1}|$  for any  $K > 2$ , so that  $\sum_{j=1}^n E_0[|T_n^{-1/2} \tilde{\gamma}_j|^K] \lesssim T_n^{1-K/2} \rightarrow 0$ . It remains to compute the convergence of the quadratic characteristics:  $\sum_{j=1}^n E_0^{j-1}[\tilde{\gamma}_j^{\otimes 2}] \rightarrow^P \mathbb{V}(\theta_0)$ . By means of the Cramér–Wold device, it suffices to prove that for each  $v'_1, v'_2 \in \mathbb{R}^{p_\alpha}$  and  $v''_1, v''_2 \in \mathbb{R}^{p_\beta}$ ,

$$(4.25) \quad \sum_{j=1}^n \frac{1}{T_n} E_0^{j-1}[(\tilde{\gamma}_j^\alpha)^{\otimes 2}][v'_1, v'_2] \rightarrow^P \mathbb{G}'_\infty[v'_1, v'_2],$$

$$(4.26) \quad \mathbb{V}_{\alpha\beta,n}[v'_1, v''_1] := \sum_{j=1}^n \frac{1}{T_n} E_0^{j-1}[\tilde{\gamma}_j^\alpha \otimes \tilde{\gamma}_j^\beta][v'_1, v''_1] \rightarrow^P \mathbb{V}_{\alpha\beta}[v'_1, v''_1],$$

$$(4.27) \quad \mathbb{V}_{\beta\beta,n}[v''_1, v''_2] := \sum_{j=1}^n \frac{1}{T_n} E_0^{j-1}[(\tilde{\gamma}_j^\beta)^{\otimes 2}][v''_1, v''_2] \rightarrow^P \mathbb{V}_{\beta\beta}[v''_1, v''_2].$$

First, (4.25) readily follows by noting  $E_0^{j-1}[m_j^{\otimes 2}] = h_n V_{j-1}$  and applying the ergodic theorem (2.3). Next,

$$(4.28) \quad \begin{aligned} & \mathbb{V}_{\alpha\beta,n}[v'_1, v''_1] \\ &= \frac{1}{n} \sum_{j=1}^n \frac{1}{h_n} E_0^{j-1}[M'_{j-1}[m_j] \otimes M''_{j-1}[m_j^{\otimes 2}]] [v'_1, v''_1] \\ &= \frac{1}{n} \sum_{j=1}^n \frac{1}{h_n} \sum_{k,l,s} E_0^{j-1}[m_j^{(k)} m_j^{(l)} m_j^{(s)}] \{M'_{j-1}^{(s)} \otimes M''_{j-1}^{(kl)}\} [v'_1, v''_1]. \end{aligned}$$

For later use, we here note that, as  $h \rightarrow 0$ ,

$$E[J_h^{(i_1)} \dots J_h^{(i_m)}] = \begin{cases} h v_{i_1 i_2 i_3}(3), & m = 3, \\ h v_{i_1 i_2 i_3 i_4}(4) + O(h^2), & m = 4; \end{cases}$$

this can be easily seen through the relation between the mixed moments and cumulants of  $J_h$ , where the latter can be computed as the values at 0 of the partial derivatives of the cumulant function  $u \mapsto \log E[\exp(i J_h[u])] = h \int \{\exp(iu[z]) - 1 - iu[z]\} \nu(dz)$ . In view of the expression

$$m_j^{(k)} = \sum_{k'} b_{j-1}^{(kk')} \Delta_j w^{(k')} + \sum_{k''} c_{j-1}^{(kk'')} \Delta_j J^{(k'')}$$

together with the orthogonalities between the increments of  $w$  and  $J$ , we get

$$\begin{aligned}
 E_0^{j-1}[m_j^{(k)} m_j^{(l)} m_j^{(s)}] &= \sum_{k',l',s'} c_{j-1}^{(kk')} c_{j-1}^{(ll')} c_{j-1}^{(ss')} E[\Delta_j J^{(k')} \Delta_j J^{(l')} \Delta_j J^{(s')}] \\
 (4.29) \qquad &= \sum_{k',l',s'} c_{j-1}^{(kk')} c_{j-1}^{(ll')} c_{j-1}^{(ss')} E[J_{h_n}^{(k')} J_{h_n}^{(l')} J_{h_n}^{(s')}] \\
 &= h_n \sum_{k',l',s'} c_{j-1}^{(kk')} c_{j-1}^{(ll')} c_{j-1}^{(ss')} \nu_{k'l's'}(3).
 \end{aligned}$$

(Since  $E[J_1] = 0$ , the 3rd mixed cumulants and the 3rd mixed moments of  $J_{h_n}$  coincides.) Substituting (4.29) in (4.28), we get (4.26)

$$\begin{aligned}
 &\mathbb{V}_{\alpha\beta,n}[v'_1, v''_1] \\
 &= \frac{1}{n} \sum_{j=1}^n \sum_{k,l,s} \sum_{k',l',s'} c_{j-1}^{(kk')} c_{j-1}^{(ll')} c_{j-1}^{(ss')} \nu_{k'l's'}(3) \{M_{j-1}'^{(s)} \otimes M_{j-1}''^{(kl)}\} [v'_1, v''_1] \\
 &= \frac{1}{n} \sum_{j=1}^n \sum_{k',l',s'} \nu_{k'l's'}(3) \{M_{j-1}'[v'_1, c_{j-1}^{(s')}]\} \{M_{j-1}''[v''_1, c_{j-1}^{(k')}, c_{j-1}^{(l')}] \} \\
 &\rightarrow^p \mathbb{V}_{\alpha\beta}[v'_1, v''_1].
 \end{aligned}$$

Finally, we look at  $\mathbb{V}_{\beta\beta,n}$ . Direct computation gives

$$\begin{aligned}
 &\mathbb{V}_{\beta\beta,n}[v''_1, v''_2] \\
 &= \frac{1}{n} \sum_{j=1}^n \frac{1}{h_n} E_0^{j-1} [(M_{j-1}'' \otimes M_{j-1}'')(v''_1, m_j^{\otimes 2}), (v''_2, m_j^{\otimes 2})] \\
 &\quad - \frac{1}{n} \sum_{j=1}^n E_0^{j-1} [(d_{j-1} \otimes M_{j-1}'')(v''_1, (v''_2, m_j^{\otimes 2}))] \\
 &\quad - \frac{1}{n} \sum_{j=1}^n E_0^{j-1} [(d_{j-1} \otimes M_{j-1}'')(v''_2, (v''_1, m_j^{\otimes 2}))] \\
 (4.30) \qquad &\quad + h_n \left( \frac{1}{n} \sum_{j=1}^n d_{j-1}^{\otimes 2}[v''_1, v''_2] \right) \\
 &= \frac{1}{n} \sum_{j=1}^n \frac{1}{h_n} E_0^{j-1} [\{M_{j-1}''[v''_1, m_j^{\otimes 2}]\} \{M_{j-1}''[v''_2, m_j^{\otimes 2}]\}] + O_p(h_n) \\
 &= \frac{1}{n} \sum_{j=1}^n \frac{1}{h_n} \sum_{k,l,k',l'} M_{j-1}''^{(kl)} [v''_1] M_{j-1}''^{(k'l')} [v''_2] E_0^{j-1} [m_j^{(k)} m_j^{(l)} m_j^{(k')} m_j^{(l')}] \\
 &\quad + O_p(h_n).
 \end{aligned}$$

Using the orthogonality as before and noting the fact that  $E[|w_{h_n}|^4] = O(h_n^2)$ , we get

$$\begin{aligned}
 & E_0^{j-1} [m_j^{(k)} m_j^{(l)} m_j^{(k')} m_j^{(l')}] \\
 &= \sum_{s,t,s',t'} c_{j-1}^{(ks)} c_{j-1}^{(lt)} c_{j-1}^{(k's')} c_{j-1}^{(l't')} E[J_{h_n}^{(s)} J_{h_n}^{(t)} J_{h_n}^{(s')} J_{h_n}^{(t')}] + R_{j-1} h_n^2 \\
 (4.31) \quad &= h_n \sum_{s,t,s',t'} c_{j-1}^{(ks)} c_{j-1}^{(lt)} c_{j-1}^{(k's')} c_{j-1}^{(l't')} \{v_{sts't'}(4) + O(h_n)\} + R_{j-1} h_n^2 \\
 &= h_n \sum_{s,t,s',t'} c_{j-1}^{(ks)} c_{j-1}^{(lt)} c_{j-1}^{(k's')} c_{j-1}^{(l't')} v_{sts't'}(4) + R_{j-1} h_n^2.
 \end{aligned}$$

By putting (4.30) and (4.31) together, we get (4.27)

$$\begin{aligned}
 & \mathbb{V}_{\beta\beta,n}[v_1'', v_2''] \\
 &= \frac{1}{n} \sum_{j=1}^n \sum_{s,t,s',t'} v_{sts't'}(4) \{M''_{j-1}[v_1'', c_{j-1}^{(s)}, c_{j-1}^{(t)}]\} \{M''_{j-1}[v_2'', c_{j-1}^{(s')}, c_{j-1}^{(t')}] \} \\
 &\quad + O_p(h_n) \\
 &\rightarrow^p \mathbb{V}_{\beta\beta}[v_1'', v_2''].
 \end{aligned}$$

The proof is thus complete.  $\square$

4.1.3. *Verification of the conditions on the derivatives of  $\mathbb{G}_n$ .* Based on (4.4) and (4.5), we derive the following bilinear forms:

$$\begin{aligned}
 (4.32) \quad \partial_\alpha \mathbb{G}_n^\alpha(\theta) &= \sum_{j=1}^n \partial_\alpha M'_{j-1}(\theta) [\chi_j] - h_n \sum_{j=1}^n \partial_\alpha M'_{j-1}(\theta) [a_{j-1}(\alpha) - a_{j-1}] \\
 &\quad - h_n \sum_{j=1}^n M'_{j-1}(\theta) \partial_\alpha a_{j-1}(\alpha),
 \end{aligned}$$

$$\begin{aligned}
 (4.33) \quad \partial_\beta \mathbb{G}_n^\alpha(\theta) &= \sum_{j=1}^n \partial_\beta M'_{j-1}(\theta) [\chi_j] \\
 &\quad - h_n \sum_{j=1}^n \partial_\beta M'_{j-1}(\theta) [a_{j-1}(\alpha) - a_{j-1}],
 \end{aligned}$$

$$\begin{aligned}
 (4.34) \quad \partial_\alpha \mathbb{G}_n^\beta(\theta) &= -2h_n \sum_{j=1}^n \{M''_{j-1}(\beta) \partial_\alpha a_{j-1}(\alpha)\} \\
 &\quad \times [\chi_j - h_n \{a_{j-1}(\alpha) - a_{j-1}\}],
 \end{aligned}$$

$$\begin{aligned}
 \partial_\beta \mathbb{G}_n^\beta(\theta) &= \sum_{j=1}^n \{ \partial_\beta M''_{j-1}(\beta) [\chi_j^{\otimes 2}] - h_n \partial_\beta d_{j-1}(\beta) \} \\
 (4.35) \quad &- 2h_n \sum_{j=1}^n \partial_\beta M''_{j-1}(\beta) [\chi_j, a_{j-1}(\alpha) - a_{j-1}] \\
 &+ h_n^2 \sum_{j=1}^n \partial_\beta M''_{j-1}(\beta) [\{a_{j-1}(\alpha) - a_{j-1}\}^{\otimes 2}].
 \end{aligned}$$

We can prove the following lemma in a similar way to the proof of Lemma 4.2.

LEMMA 4.7. *For every  $K > 0$ ,*

$$\sup_n E_0 \left[ \sup_\theta \left| \frac{1}{T_n} \partial_\theta^k \mathbb{G}_n(\theta) \right|^K \right] < \infty, \quad k = 1, 2, 3.$$

Recall that the matrix  $\mathbb{G}'_\infty(\theta_0) = \text{diag}\{\mathbb{G}'_\infty{}^\alpha(\theta_0), \mathbb{G}'_\infty{}^\beta(\theta_0)\}$  is given by (2.7) and (2.8).

LEMMA 4.8. *For every  $K > 0$ ,*

$$\sup_{n \in \mathbb{N}} E_0 \left[ \left| \sqrt{T_n} \left( \frac{1}{T_n} \partial_\theta \mathbb{G}_n(\theta_0) - \mathbb{G}'_\infty(\theta_0) \right) \right|^K \right] < \infty.$$

PROOF. First, concerning the off-diagonal parts, we have

$$\begin{aligned}
 \frac{1}{T_n} \partial_\beta \mathbb{G}_n^\alpha &= \frac{1}{\sqrt{T_n}} \sum_{j=1}^n \frac{1}{\sqrt{T_n}} \partial_\beta M'_{j-1}[\chi_j] = O_p^* \left( \frac{1}{\sqrt{T_n}} \right), \\
 \frac{1}{T_n} \partial_\alpha \mathbb{G}_n^\beta &= -2 \frac{h_n}{\sqrt{T_n}} \sum_{j=1}^n \frac{1}{\sqrt{T_n}} M''_{j-1}[\partial_\alpha a_{j-1}, \chi_j] = O_p^* \left( \frac{h_n}{\sqrt{T_n}} \right),
 \end{aligned}$$

where the moment estimates for the martingale terms will be proved in an analogous way to the proof of Lemma 4.2. Next, we observe

$$\begin{aligned}
 \frac{1}{T_n} \partial_\alpha \mathbb{G}_n^\alpha - \mathbb{G}'_\infty{}^\alpha &= \frac{1}{\sqrt{T_n}} \sum_{j=1}^n \frac{1}{\sqrt{T_n}} \partial_\alpha M'_{j-1}[\chi_j] - \frac{1}{n} \sum_{j=1}^n M'_{j-1} \partial_\alpha a_{j-1} - \mathbb{G}'_\infty{}^\alpha(\theta_0) \\
 &= O_p^* \left( \frac{1}{\sqrt{T_n}} \right) + \frac{1}{\sqrt{T_n}} \left\{ \sqrt{T_n} \left( -\frac{1}{n} \sum_{j=1}^n M'_{j-1} \partial_\alpha a_{j-1} - \mathbb{G}'_\infty{}^\alpha(\theta_0) \right) \right\} \\
 &= O_p^* \left( \frac{1}{\sqrt{T_n}} \right),
 \end{aligned}$$



where we used Lemma 4.3 for the last equality. It remains to look at  $T_n^{-1} \partial_\beta \mathbb{G}_n^\beta$ . Plugging in the identity  $\chi_j = m_j + r'_j + h_n^2 R_{j-1}$  and making use of what we have seen in the first half of the proof of Lemma 4.6, we proceed as follows:

$$\begin{aligned}
 & \frac{1}{T_n} \partial_\beta \mathbb{G}_n^\beta \\
 &= \frac{1}{T_n} \sum_{j=1}^n (\partial_\beta M''_{j-1} [(m_j + r'_j)^{\otimes 2}] - h_n \partial_\beta d_{j-1}) + O_p^*(h_n) \\
 &= \frac{1}{T_n} \sum_{j=1}^n (\partial_\beta M''_{j-1} [m_j^{\otimes 2}] - h_n \partial_\beta d_{j-1}) + O_p^*(\sqrt{h_n}) \\
 (4.36) \quad &= \frac{1}{\sqrt{T_n}} \left\{ \sum_{j=1}^n \frac{1}{\sqrt{T_n}} (\partial_\beta M''_{j-1} [m_j^{\otimes 2}] - E_0^{j-1} [\partial_\beta M''_{j-1} [m_j^{\otimes 2}]]) \right\} \\
 &+ \frac{1}{T_n} \sum_{j=1}^n (E_0^{j-1} [\partial_\beta M''_{j-1} [m_j^{\otimes 2}]] - h_n \partial_\beta d_{j-1}) + O_p^*(\sqrt{h_n}) \\
 &= \frac{1}{T_n} \sum_{j=1}^n (E_0^{j-1} [\partial_\beta M''_{j-1} [m_j^{\otimes 2}]] - h_n \partial_\beta d_{j-1}) + O_p^*\left(\frac{1}{\sqrt{T_n}}\right) \\
 &= \frac{1}{n} \sum_{j=1}^n [\text{trace}\{(-\partial_{\beta_l} \partial_{\beta_{l'}} V_{j-1}^{-1}) V_{j-1}\} - \partial_{\beta_l} \partial_{\beta_{l'}} \log |V_{j-1}|]_{l,l'=1}^{p\beta} \\
 &+ O_p^*\left(\frac{1}{\sqrt{T_n}}\right).
 \end{aligned}$$

The  $(l, l')$ th component of the first term in (4.36) tends in probability to

$$\begin{aligned}
 & \int [\text{trace}\{-\partial_{\beta_l} \partial_{\beta_{l'}} V^{-1} V(x, \beta_0)\} - \partial_{\beta_l} \partial_{\beta_{l'}} \log |V|(x, \beta_0)] \pi_0(dx) \\
 &= - \int \text{trace}\{(V^{-1}(\partial_{\beta_l} V) V^{-1}(\partial_{\beta_{l'}} V))(x, \beta_0)\} \pi_0(dx).
 \end{aligned}$$

Accordingly, a reduced version of Lemma 4.3 with  $\Theta = \{\theta_0\}$  applies to conclude that  $T_n^{-1} \partial_\beta \mathbb{G}_n^\beta(\theta_0) - \mathbb{G}_\infty^\beta(\theta_0) = O_p^*(T_n^{-1/2})$ . The proof is complete.  $\square$

4.2. *Proof of Corollary 2.8.* By Theorem 2.7, we know that  $\sqrt{T_n}(\hat{\alpha}_n - \alpha_0) = O_p(1)$  and  $\sqrt{T_n}(\hat{\beta}_n - \beta_0) = O_p(1)$ . It is easy to see from Taylor expansion that  $\hat{\mathbb{G}}_n^{\prime\alpha} \rightarrow^p \mathbb{G}_\infty^{\prime\alpha}(\theta_0)$  and  $\hat{\mathbb{G}}_n^{\prime\beta} \rightarrow^p \mathbb{G}_\infty^{\prime\beta}(\theta_0)$ . Turning to  $\hat{V}_{\alpha\beta,n}$  and  $\hat{V}_{\beta\beta,n}$ , we plug the expression  $\chi_j(\hat{\alpha}_n) = \chi_j + \sqrt{h_n/n} R_{j-1}[\sqrt{T_n}(\hat{\alpha}_n - \alpha_0)]$  into their definitions and

then apply Taylor expansion with respect to  $\hat{\theta}_n$  around  $\theta_0$  as before, to obtain

$$\begin{aligned}
 \hat{\mathbb{V}}_{\alpha\beta,n}[v'_1, v''_1] &= - \sum_{j=1}^n \frac{1}{T_n} (V_{j-1}^{-1} \otimes \partial_\beta V_{j-1}^{-1}) [(\partial_\alpha a_{j-1}[v'_1], \chi_j), (v''_1, \chi_j^{\otimes 2})] \\
 &\quad + O_p\left(\frac{1}{\sqrt{T_n}}\right), \\
 \hat{\mathbb{V}}_{\beta\beta,n}[v''_1, v''_1] &= \sum_{j=1}^n \frac{1}{T_n} (\partial_\beta V_{j-1}^{-1} \otimes \partial_\beta V_{j-1}^{-1}) [(v''_1, \chi_j^{\otimes 2}), (v''_1, \chi_j^{\otimes 2})] \\
 &\quad + O_p\left(\frac{1}{\sqrt{T_n}}\right).
 \end{aligned}
 \tag{4.37}$$

We only show that  $\hat{\mathbb{V}}_{\alpha\beta,n}[v'_1, v''_1] \rightarrow^p \mathbb{V}_{\alpha\beta}[v'_1, v''_1]$ , for the case of  $\hat{\mathbb{V}}_{\beta\beta,n}$  is similar.

Write  $\sum_{j=1}^n \eta_j$  for the first term in the right-hand side of (4.37). We can show that

$$\sum_{j=1}^n E_0^{j-1}[\eta_j] \rightarrow^p \mathbb{V}_{\alpha\beta}[v'_1, v''_1]$$

in a similar manner to show the convergence of the quadratic characteristics in the proof of Lemma 4.6. Noting that  $E_0^{j-1}[|\chi_j|^q] \leq h_n R_{j-1}$  for every  $q \geq 2$ , we also have

$$\sum_{j=1}^n E_0[(\eta_j - E_0^{j-1}[\eta_j])^2] \lesssim \sum_{j=1}^n E_0[\eta_j^2] \lesssim \frac{1}{T_n} \rightarrow 0.$$

Applying the Lenglart domination property for the martingale  $\sum_{j=1}^n (\eta_j - E_0^{j-1}[\eta_j])$  (cf. Jacod and Shiryaev [15], I.3.30), we conclude that  $\sum_{j=1}^n \eta_j \rightarrow^p \mathbb{V}_{\alpha\beta}[v'_1, v''_1]$ , hence  $\hat{\mathbb{V}}_{\alpha\beta,n}[v'_1, v''_1] \rightarrow^p \mathbb{V}_{\alpha\beta}[v'_1, v''_1]$ .

4.3. *Proof of Theorem 2.9.* First, we mention an auxiliary estimate. Recall (4.6) and (4.21):  $\chi_j := \Delta_j X - h_n a_{j-1}(\alpha_0) = m_j + (r_j + r'_j)$ . Using Birkholder’s inequality and then the Lipschitz continuity of the coefficients, we see that

$$E_0[|r_j + r'_j|^{q'}] \lesssim \int_j E_0[|X_s - X_{t_{j-1}}|^{q'}] ds \lesssim h_n^2 \|g\|^{q'} \lesssim h_n^2$$

for  $q' \in [2, q]$ , where  $g$  is the one given in Lemma 4.5. In this proof,  $R$  denotes a generic essentially bounded function on  $\mathbb{R}^d$  possibly depending on  $n$  and  $\theta$ .

By means of the classical  $M$ -estimation theory (e.g., van der Vaart [44], Chapter 5), it is crucial to have the uniform convergence

$$\sup_{\theta \in \Theta} \left| \frac{1}{T_n} \mathbb{G}_n(\theta) - \mathbb{G}_\infty(\theta) \right| + \sup_{\theta \in \Theta} \left| \frac{1}{T_n} \partial_\theta \mathbb{G}_n(\theta) - \mathbb{G}'_\infty(\theta) \right| \rightarrow^p 0.
 \tag{4.38}$$

Most key materials to prove this have been obtained in the proof of Theorem 2.7, so we only give a sketch.

Note that the variables  $M'_{j-1}(\theta)$  and  $M''_{j-1}(\beta)$  are now essentially bounded uniformly in  $\theta$ . Substituting  $\chi_j = m_j + h_n^2 R_{j-1}$  in the expressions (4.4) and (4.5) about  $\mathbb{G}_n$ , and also (4.32), (4.33), (4.34) and (4.35) about  $\partial_\theta \mathbb{G}_n$ , it is not difficult to deduce (4.38); as was in the proof of Theorem 2.7, for the estimate to be valid uniformly in  $\theta$  we applied Sobolev inequality in part, where it was needed that  $E[|J_1|^q] < \infty$  for some  $q > p$ .

Now, the consistency of  $\hat{\theta}_n$  follows from (4.38):  $\hat{\theta}_n \xrightarrow{P} \theta_0$ . Since  $P[\omega : \mathbb{G}_n(\hat{\theta}_n(\omega)) = 0] \rightarrow 1$ , we may and do suppose that  $\mathbb{G}_n(\hat{\theta}_n) = 0$ . In view of (4.38) and the Taylor expansion  $0 = T_n^{-1/2} \mathbb{G}_n(\theta_0) + T_n^{-1} \partial_\theta \mathbb{G}_n(\tilde{\theta}_n)[\sqrt{T_n}(\hat{\theta}_n - \theta_0)]$ , where the point  $\tilde{\theta}_n$  lies on the segment connecting  $\hat{\theta}_n$  and  $\theta_0$ , it suffices to have the central limit theorem (4.20). By close inspection of the proof of Lemma 4.6, we note that the present assumption [especially  $q > (4 \vee p)$  about the moment order] is enough to conclude (4.20). The proof is complete.

**5. A criterion for the exponential ergodicity in dimension one.** In this section, we set  $d = r' = r'' = 1$  and suppress dependence on the parameter from the notation

$$(5.1) \quad dX_t = a(X_t) dt + b(X_t) dW_t + c(X_{t-}) dJ_t.$$

We here forget Assumptions 2.1 to 2.5, and instead introduce the following set of conditions.

ASSUMPTION 5.1.  $(a, b, c)$  is of class  $C^1(\mathbb{R})$  and globally Lipschitz, and  $(b, c)$  is bounded.

ASSUMPTION 5.2. Either one of the following conditions holds true:

- (i)  $b(x') \neq 0$  for some  $x'$ ,  $c(x'') \neq 0$  for every  $x''$ , and there exists a constant  $\bar{\varepsilon} > 0$  such that  $\nu(-\varepsilon, 0) \wedge \nu(0, \varepsilon) > 0$  for every  $\varepsilon \in (0, \bar{\varepsilon})$ ;
- (ii)  $b \equiv 0$ ,  $c(x'') \neq 0$  for every  $x''$ , and we have the decomposition

$$\nu = \nu_\star + \nu_\natural$$

for two Lévy measures  $\nu_\star$  and  $\nu_\natural$ , where the restriction of  $\nu_\star$  to some open set of the form  $(-\bar{\varepsilon}, 0) \cup (0, \bar{\varepsilon})$  admits a continuously differentiable positive density  $g_\star$ .

ASSUMPTION 5.3.

- (i)  $E[J_1] = 0$  and  $\int_{|z|>1} |z|^q \nu(dz) < \infty$  for some  $q \geq 1$ , and

$$\limsup_{|x| \rightarrow \infty} \frac{a(x)}{x} < 0.$$

- (ii)  $E[J_1] = 0$  and  $\int_{|z|>1} \exp(q|z|) \nu(dz) < \infty$  for some  $q > 0$ , and
 
$$\limsup_{|x| \rightarrow \infty} \operatorname{sgn}(x) a(x) < 0.$$

The next proposition gives a pretty simple criterion for Assumption 2.3.

PROPOSITION 5.4. *The following holds true:*

(a) *Suppose conditions 5.1, 5.2, 5.3(i), and that  $E[|X_0|^q] < \infty$ . Then, there exist a probability measure  $\pi$  and a constant  $a > 0$  such that (2.1) holds true for a  $C^2$ -function  $g$  satisfying that  $g(x) = 1 + |x|^q$  outside a neighborhood of the origin. Further, (2.2) holds true for the  $q$  given in 5.3(i).*

(b) *Suppose 5.1, 5.2, 5.3(ii), and that  $E[\exp(q|X_0|)] < \infty$ . Then, there exist a probability measure  $\pi$  and constants  $a, \varepsilon > 0$  such that (2.1) holds true for a  $C^2$ -function  $g$  satisfying that  $g(x) = 1 + \exp(\varepsilon|x|)$  outside a neighborhood of the origin. Further, (2.2) holds true for arbitrary  $q > 0$ .*

PROOF. The Lipschitz continuity implies that the SDE (5.1) admits a unique strong solution. We consider the following conditions:

- (I) there exists a constant  $\Delta > 0$  for which every compact sets are petite for the Markov chain  $(X_{j\Delta})_{j \in \mathbb{Z}_+}$ ;
- (II) the exponential Lyapunov-drift criterion

$$(5.2) \quad \mathcal{A}\varphi \leq -c\varphi + d$$

holds true for some constants  $c, d > 0$  and some  $\varphi : \mathbb{R} \rightarrow \mathbb{R}_+$  belonging to the domain of  $\mathcal{A}$  such that  $\lim_{|x| \rightarrow \infty} \varphi(x) = \infty$ , where  $\mathcal{A}$  denotes the extended generator of  $X$ .

As in the proof of Masuda [28], the proof of Theorem 2.2, in each of (a) and (b) the exponential ergodicity (2.1) follows from (I) and (II), and the moment bound (2.2) from (II) alone. In order to prove (I), we will first verify the *Local Doeblin (LD) condition* (see Kulik [19] for details); we note that the LD condition implies (I) for any  $\Delta > 0$ . Then we will verify the drift condition (II) with different choices of  $\varphi$  under Assumptions 5.3(i) and 5.3(ii).

*Verification of (I): the LD condition.*

First, we verify the LD condition under Assumption 5.2(i). Let  $\Pi_x(A) := \nu(\{z \in \mathbb{R} : c(x)z \in A\})$ , and refer to Kulik’s condition (S) in the reduced form

$$(S) \quad \forall x \in \mathbb{R} \ \forall v \in \{-1, 1\} \ \exists \rho \in (-1, 1) \ \forall \delta > 0 : \\ \Pi_x(\{y \in \mathbb{R} : yv \geq \rho|y|\} \cap \{y \in \mathbb{R} : |y| \leq \delta\}) > 0.$$

Under Assumption 5.2(i), it follows from Kulik [19], Theorem 1.3, Proposition A.2 and Proposition 4.7, that the condition (S) above implies the LD condition. Simple

manipulation shows that the last condition is equivalent to the following:

$$\forall x \in \mathbb{R} \forall \delta > 0:$$

$$\nu(\{z \in \mathbb{R} : 0 \leq c(x)z \leq \delta\}) \wedge \nu(\{z \in \mathbb{R} : -\delta \leq c(x)z \leq 0\}) > 0.$$

Since  $\nu(\mathbb{R}) > 0$ , it suffices to look at  $x$  such that  $c(x) \neq 0$ . However, for such  $x$ , the condition obviously holds true under Assumption 5.2(i).

Next we verify the LD condition under Assumption 5.2(ii). If  $c$  is constant, then we can apply Kulik [19], Proposition 0.1, to verify the LD condition. Therefore, we suppose that  $\partial_x c \neq 0$  in what follows. We smoothly truncate the support of  $\nu_*$  as follows: pick any  $\underline{\varepsilon} \in (0, \bar{\varepsilon})$ , let  $\psi : \mathbb{R} \rightarrow [0, 1]$  be given by<sup>2</sup>

$$\psi(z) := \begin{cases} \exp\{-(z - \underline{\varepsilon})^{-1} - (\bar{\varepsilon} - z)^{-1}\}, & (\underline{\varepsilon} < z < \bar{\varepsilon}), \\ 0, & (\text{otherwise}) \end{cases}$$

and set

$$\nu_1(dz) := \{\psi(z) + \psi(-z)\} \nu_*(dz) = \{\psi(z) + \psi(-z)\} g_*(z) dz.$$

Then we have the decomposition  $\nu = \nu_1 + \nu_2$ , where  $\nu_2(dz) := [1 - \{\psi(z) + \psi(-z)\}] \nu_*(dz) + \nu_{\dagger}(dz)$  defines a Lévy measure. The function  $z \mapsto \{\psi(z) + \psi(-z)\} g_*(z)$  is smooth and supported by  $[-\bar{\varepsilon}, -\underline{\varepsilon}] \cup [\underline{\varepsilon}, \bar{\varepsilon}]$ . With this truncation in hand, we can apply Kulik [19], Proposition A.1, which states that, when the diffusion part is absent, the LD condition is implied by the conditions (S) plus  $(\hat{N})$ ,

$$(\hat{N}) \quad \exists x'' \in \mathbb{R} \exists t'' > 0: P_{x''}[\hat{S}_{t''} = \mathbb{R}] > 0,$$

where  $\hat{S}_t := \{u \mathcal{E}_t^u c(X_{\tau-}); u \in \mathbb{R}, \tau \in \mathcal{D}_1 \cap (0, t)\}$ , with  $\mathcal{D}_1$  and  $(\mathcal{E}_s^t)_{0 \leq s \leq t}$ , respectively, denoting the domain of the point process  $N_1$  associated with  $\nu_1$  and a right-continuous solution to

$$\mathcal{E}_s^t = 1 + \int_s^t \partial_x a(X_u) \mathcal{E}_s^u du + \int_s^t \partial_x c(X_{u-}) \mathcal{E}_s^{u-} dJ_u.$$

As (S) has been already verified in the previous paragraph, it remains to prove  $(\hat{N})$ ; obviously, if  $\nu$  fulfils Assumption 5.2(ii), then it does Assumption 5.2(i) too. The stochastic-exponential formula leads to

$$\mathcal{E}_s^t = \exp(Y_t - Y_s) \prod_{s < u \leq t} (1 + \Delta Y_u) \exp(-\Delta Y_u), \quad s \leq t,$$

where  $Y_u := \int_0^u \partial_x a(X_v) dv + \int_0^u \partial_x c(X_{v-}) dJ_v$ . We now introduce the two auxiliary sets

$$A'(t) := \{\omega \in \Omega : \mathcal{D}_1 \cap (0, t) \neq \emptyset\},$$

$$A''(t) := \{\omega \in \Omega : N((0, t], \{z \in \mathbb{R}; |z| \geq \|\partial_x c\|_{\infty}^{-1}\}) = 0\},$$

---

<sup>2</sup>The author owes Professor A. M. Kulik for this clear-cut choice of  $\psi$ .

where  $N(dt, dz)$  denotes the Poisson random measure associated with  $J$ . According to the implications

$$\begin{aligned} \{|\Delta J_u| < \|\partial_x c\|_\infty^{-1}, u \in (0, t]\} &\subset \{|\partial_x c(X_{u-})\Delta J_u| < 1, u \in (0, t]\} \\ &= \{|\Delta Y_u| < 1, u \in (0, t]\} \\ &\subset \{\mathcal{E}_s^t \neq 0, s \in [0, t]\}, \end{aligned}$$

the process  $(\mathcal{E}_s^t)_{0 \leq s \leq t}$  stays positive a.s. on  $A''(t)$ . Since  $P[A'(t) \cap A''(t)] > 0$  for every  $t > 0$  and  $c$  is nonvanishing on  $\mathbb{R}$ , we observe that for every  $x \in \mathbb{R}$  and  $t > 0$

$$\begin{aligned} P_x[\hat{S}_t = \mathbb{R}] &\geq P_x[\{\hat{S}_t = \mathbb{R}\} \cap A'(t) \cap A''(t)] \\ &\geq P_x[\{\mathcal{E}_s^t c(X_{s-}) \neq 0 \text{ for some } s \in (0, t)\} \cap A'(t) \cap A''(t)] \\ &= P_x[\{c(X_{s-}) \neq 0 \text{ for some } s \in (0, t)\} \cap A'(t) \cap A''(t)] \\ &= P_x[A'(t) \cap A''(t)] > 0, \end{aligned}$$

hence the LD condition.

*Verification of (II): the drift condition.* Now we turn to the verification of (5.2). For verification under Assumption 5.3(i), one can refer to Kulik [19] and Masuda [28, 29]; in this case, we may set  $\varphi(x) = |x|^q$  outside a sufficiently large neighborhood of the origin. We are left to showing (5.2) under Assumption 5.3(ii), where, compared with Assumption 5.3(i), we impose a weaker condition on the drift function  $a$  while a stronger moment condition on  $\nu$ . We will achieve the proof in a somewhat similar manner to the proof of Masuda [29], Theorem 1.2.

Fix any  $\varepsilon \in (0, q\|c\|_\infty^{-1} \wedge 1)$  and pick a  $\varphi = \varphi_\varepsilon \in \mathcal{C}^2(\mathbb{R})$  fulfilling:

- $\varphi(x) = \exp(\varepsilon|x|)$  for  $|x| \geq \varepsilon^{-1}$ ;
- $\varphi(x) \leq \exp(\varepsilon|x|)$  for every  $x$ ;
- $|\partial_x^2 \varphi(x)| \leq C\varepsilon^2 \varphi(x)$  for every  $x$ .

We can write  $\mathcal{A}\varphi = \mathcal{G}\varphi + \mathcal{J}\varphi$ , where

$$\begin{aligned} \mathcal{G}\varphi(x) &:= \partial_x \varphi(x)a(x) + \frac{1}{2} \partial_x^2 \varphi(x)b^2(x), \\ \mathcal{J}\varphi(x) &:= \int \{\varphi(x + c(x)z) - \varphi(x) - \partial_x \varphi(x)c(x)z\} \nu(dz). \end{aligned}$$

According to the local boundedness of  $x \mapsto \mathcal{A}\varphi(x)$ , we may and do concentrate on  $x$  with  $|x|$  large enough. Direct algebra gives

$$(5.3) \quad \mathcal{G}\varphi(x) \leq \varepsilon \varphi(x) \{\text{sgn}(x)a(x) + C\varepsilon\}.$$

Further, by means of Taylor's theorem and the property of  $\varphi$ ,

$$\begin{aligned}
 |\mathcal{J}\varphi(x)| &\lesssim |c(x)|^2 \int |z|^2 \left( \sup_{0 \leq s \leq 1} |\partial_x^2 \varphi(x + sc(x)z)| \right) \nu(dz) \\
 (5.4) \qquad &\lesssim \varepsilon^2 \exp(\varepsilon|x|) \int |z|^2 \exp(\varepsilon\|c\|_\infty|z|) \nu(dz) \\
 &\lesssim \varepsilon^2 \varphi(x).
 \end{aligned}$$

By putting (5.3) and (5.4) together and by taking  $\varepsilon$  small enough, we can find a constant  $c_0 > 0$  for which  $\mathcal{A}\varphi(x) \leq -c_0\varphi(x)$  for every  $|x|$  large enough. The proof of Proposition 5.4 is complete.  $\square$

**REMARK 5.5.** If the condition on  $\nu$  in Assumption 5.2(i) fails to hold, then  $J$  is necessarily a compound-Poisson process. In this case, we can utilize the criteria given in Masuda [29].

**REMARK 5.6.** By combining the results of the LD-condition argument and general stability theory for Markov processes, it is possible to formulate subexponential- and polynomial-ergodicity versions, as well as the ergodicity version (without rate specification); see, for example, Meyn and Tweedie [34] and Fort and Roberts [9]. Especially, as in Masuda [29], the conditions on  $(a, b, c)$  in Proposition 5.4 can be considerably relaxed in case of the ergodicity version, because the Lyapunov condition required then becomes much weaker.

**Acknowledgments.** The author is grateful to two anonymous referees for careful reading and for several valuable and constructive comments, which led to substantial improvement of the paper. He also thanks Professor A. M. Kulik for his helpful comments concerning the verification of the LD condition appearing in the proof of Proposition 5.4.

## REFERENCES

- [1] AÏT-SAHALIA, Y. and JACOD, J. (2007). Volatility estimators for discretely sampled Lévy processes. *Ann. Statist.* **35** 355–392. [MR2332279](#)
- [2] AÏT-SAHALIA, Y. and JACOD, J. (2008). Fisher's information for discretely sampled Lévy processes. *Econometrica* **76** 727–761. [MR2433480](#)
- [3] BARDET, J.-M. and WINTENBERGER, O. (2009). Asymptotic normality of the quasi-maximum likelihood estimator for multidimensional causal processes. *Ann. Statist.* **37** 2730–2759. [MR2541445](#)
- [4] BARNDORFF-NIELSEN, O. E. and SHEPHARD, N. (2001). Non-Gaussian Ornstein–Uhlenbeck-based models and some of their uses in financial economics. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **63** 167–241. [MR1841412](#)
- [5] BHANSALI, R. J. and PAPANGELOU, F. (1991). Convergence of moments of least squares estimators for the coefficients of an autoregressive process of unknown order. *Ann. Statist.* **19** 1155–1162. [MR1126319](#)

- [6] CHAN, N. H. and ING, C.-K. (2011). Uniform moment bounds of Fisher's information with applications to time series. *Ann. Statist.* **39** 1526–1550. [MR2850211](#)
- [7] DVORETZKY, A. (1977). Asymptotic normality of sums of dependent random vectors. In *Multivariate Analysis, IV (Proc. Fourth Internat. Sympos., Dayton, Ohio, 1975)* 23–34. North-Holland, Amsterdam. [MR0451351](#)
- [8] FINDLEY, D. F. and WEI, C.-Z. (2002). AIC, overfitting principles, and the boundedness of moments of inverse matrices for vector autoregressions and related models. *J. Multivariate Anal.* **83** 415–450. [MR1945962](#)
- [9] FORT, G. and ROBERTS, G. O. (2005). Subgeometric ergodicity of strong Markov processes. *Ann. Appl. Probab.* **15** 1565–1589. [MR2134115](#)
- [10] FRIEDMAN, A. (2006). *Stochastic Differential Equations and Applications*. Dover Publications, Mineola, NY. [MR2295424](#)
- [11] GOBET, E. (2002). LAN property for ergodic diffusions with discrete observations. *Ann. Inst. Henri Poincaré Probab. Stat.* **38** 711–737. [MR1931584](#)
- [12] HEYDE, C. C. (1997). *Quasi-Likelihood and Its Application: A General Approach to Optimal Parameter Estimation*. Springer, New York. [MR1461808](#)
- [13] IBRAGIMOV, I. A. and HAS'MINSKIĬ, R. Z. (1981). *Statistical Estimation: Asymptotic Theory. Applications of Mathematics 16*. Springer, New York. [MR0620321](#)
- [14] INAGAKI, N. and OGATA, Y. (1975). The weak convergence of likelihood ratio random fields and its applications. *Ann. Inst. Statist. Math.* **27** 391–419. [MR0405689](#)
- [15] JACOD, J. and SHIRYAEV, A. N. (2003). *Limit Theorems for Stochastic Processes*, 2nd ed. *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]* **288**. Springer, Berlin. [MR1943877](#)
- [16] JEGANATHAN, P. (1982). On the convergence of moments of statistical estimators. *Sankhyā Ser. A* **44** 213–232. [MR0688801](#)
- [17] JEGANATHAN, P. (1989). A note on inequalities for probabilities of large deviations of estimators in nonlinear regression models. *J. Multivariate Anal.* **30** 227–240. [MR1015369](#)
- [18] KESSLER, M. (1997). Estimation of an ergodic diffusion from discrete observations. *Scand. J. Stat.* **24** 211–229. [MR1455868](#)
- [19] KULIK, A. M. (2009). Exponential ergodicity of the solutions to SDE's with a jump noise. *Stochastic Process. Appl.* **119** 602–632. [MR2494006](#)
- [20] KULIK, A. M. (2011). Asymptotic and spectral properties of exponentially  $\phi$ -ergodic Markov processes. *Stochastic Process. Appl.* **121** 1044–1075. [MR2775106](#)
- [21] KUNITA, H. (1990). *Stochastic Flows and Stochastic Differential Equations. Cambridge Studies in Advanced Mathematics 24*. Cambridge Univ. Press, Cambridge. [MR1070361](#)
- [22] KUTOYANTS, Y. A. (1984). *Parameter Estimation for Stochastic Processes. Research and Exposition in Mathematics 6*. Heldermann, Berlin. [MR0777685](#)
- [23] KUTOYANTS, Y. A. (2004). *Statistical Inference for Ergodic Diffusion Processes*. Springer, London. [MR2144185](#)
- [24] LUSCHGY, H. and PAGÈS, G. (2008). Moment estimates for Lévy processes. *Electron. Commun. Probab.* **13** 422–434. [MR2430710](#)
- [25] MANCINI, C. (2004). Estimation of the characteristics of the jumps of a general Poisson-diffusion model. *Scand. Actuar. J.* **1** 42–52. [MR2045358](#)
- [26] MARUYAMA, G. and TANAKA, H. (1959). Ergodic property of  $N$ -dimensional recurrent Markov processes. *Mem. Fac. Sci. Kyushu Univ. Ser. A* **13** 157–172. [MR0112175](#)
- [27] MASUDA, H. (2005). Simple estimators for parametric Markovian trend of ergodic processes based on sampled data. *J. Japan Statist. Soc.* **35** 147–170. [MR2328423](#)
- [28] MASUDA, H. (2007). Ergodicity and exponential  $\beta$ -mixing bounds for multidimensional diffusions with jumps. *Stochastic Process. Appl.* **117** 35–56. [MR2287102](#)
- [29] MASUDA, H. (2008). On stability of diffusions with compound-Poisson jumps. *Bull. Inform. Cybernet.* **40** 61–74. [MR2484446](#)



- [30] MASUDA, H. (2010). Approximate self-weighted LAD estimation of discretely observed ergodic Ornstein–Uhlenbeck processes. *Electron. J. Stat.* **4** 525–565. [MR2660532](#)
- [31] MASUDA, H. (2011). Approximate quadratic estimating function for discretely observed Lévy driven SDEs with application to a noise normality test. *RIMS Kôkyûroku* **1752** 113–131.
- [32] MASUDA, H. (2013). Asymptotics for functionals of self-normalized residuals of discretely observed stochastic processes. *Stochastic Process. Appl.* **123** 2752–2778. [MR3054544](#)
- [33] MENALDI, J. L. and ROBIN, M. (1999). Invariant measure for diffusions with jumps. *Appl. Math. Optim.* **40** 105–140. [MR1686598](#)
- [34] MEYN, S. P. and TWEEDIE, R. L. (1993). Stability of Markovian processes. III. Foster–Lyapunov criteria for continuous-time processes. *Adv. in Appl. Probab.* **25** 518–548. [MR1234295](#)
- [35] OGATA, Y. and INAGAKI, N. (1977). The weak convergence of the likelihood ratio random fields for Markov observations. *Ann. Inst. Statist. Math.* **29** 165–187. [MR0458760](#)
- [36] OGIHARA, T. and YOSHIDA, N. (2011). Quasi-likelihood analysis for the stochastic differential equation with jumps. *Stat. Inference Stoch. Process.* **14** 189–229. [MR2832818](#)
- [37] SHIMIZU, Y. (2006).  $M$ -estimation for discretely observed ergodic diffusion processes with infinitely many jumps. *Stat. Inference Stoch. Process.* **9** 179–225. [MR2249182](#)
- [38] SHIMIZU, Y. and YOSHIDA, N. (2006). Estimation of parameters for diffusion processes with jumps from discrete observations. *Stat. Inference Stoch. Process.* **9** 227–277. [MR2252242](#)
- [39] SIEDERS, A. and DZHAPARIDZE, K. (1987). A large deviation result for parameter estimators and its application to nonlinear regression analysis. *Ann. Statist.* **15** 1031–1049. [MR0902244](#)
- [40] SØRENSEN, M. (2008). Efficient estimation for ergodic diffusions sampled at high frequency. Preprint.
- [41] STRAUMANN, D. and MIKOSCH, T. (2006). Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series: A stochastic recurrence equations approach. *Ann. Statist.* **34** 2449–2495. [MR2291507](#)
- [42] UCHIDA, M. (2010). Contrast-based information criterion for ergodic diffusion processes from discrete observations. *Ann. Inst. Statist. Math.* **62** 161–187. [MR2577445](#)
- [43] UCHIDA, M. and YOSHIDA, N. (2012). Adaptive estimation of an ergodic diffusion process based on sampled data. *Stochastic Process. Appl.* **122** 2885–2924. [MR2931346](#)
- [44] VAN DER VAART, A. W. (1998). *Asymptotic Statistics. Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge Univ. Press, Cambridge. [MR1652247](#)
- [45] WANG, J. (2008). Criteria for ergodicity of Lévy type operators in dimension one. *Stochastic Process. Appl.* **118** 1909–1928. [MR2454470](#)
- [46] WEDDERBURN, R. W. M. (1974). Quasi-likelihood functions, generalized linear models, and the Gauss–Newton method. *Biometrika* **61** 439–447. [MR0375592](#)
- [47] YOSHIDA, N. (1992). Estimation for diffusion processes from discrete observation. *J. Multivariate Anal.* **41** 220–242. [MR1172898](#)
- [48] YOSHIDA, N. (2011). Polynomial type large deviation inequalities and quasi-likelihood analysis for stochastic differential equations. *Ann. Inst. Statist. Math.* **63** 431–479. [MR2786943](#)

INSTITUTE OF MATHEMATICS FOR INDUSTRY  
KYUSHU UNIVERSITY  
744 MOTOOKA NISHI-KU FUKUOKA 819-0395  
JAPAN  
E-MAIL: [hiroki@imi.kyushu-u.ac.jp](mailto:hiroki@imi.kyushu-u.ac.jp)