Non-Gaussian quasi-likelihood estimation of locally stable SDE

Masuda, Hiroki
Faculty of Mathematics, Kyushu University

出版情報：MI Preprint Series. 2016-4, 2016-04-21. 九州大学大学院数理学研究院
バージョン：
権利関係：

MI Preprint Series

Mathematics for Industry
Kyushu University

Non-Gaussian quasi-likelihood estimation of locally stable SDE

Hiroki Masuda

MI 2016-4

(Received April 21, 2016)

Institute of Mathematics for Industry
Graduate School of Mathematics
Kyushu University
Fukuoka, JAPAN

NON-GAUSSIAN QUASI-LIKELIHOOD ESTIMATION OF LOCALLY STABLE SDE

HIROKI MASUDA

Abstract. We address parametric estimation of both trend and scale coefficients of a pure-jump Lévy driven univariate stochastic differential equation (SDE) model based on high-frequency data over a fixed time period. It is known from the previous study [35] that the conventional Gaussian quasi-maximum likelihood estimator is inconsistent. In this paper, under the assumption that the driving Lévy process is locally stable, we propose a novel quasi-likelihood function based on the small-time non-Gaussian stable approximation of the unknown transition density. The resulting estimator is shown to be asymptotically mixed-normally distributed and remarkably more efficient than the Gaussian quasi-maximum likelihood estimator. We need neither ergodicity nor existence of finite moments. Compared with the existing methods for estimating SDE models, the proposed quasi-likelihood enables us to achieve better performance in a unified manner for a wide range of the driving Lévy processes.

1. Introduction

Stochastic differential equation (SDE) is a basic model to describe time-varying physical and natural phenomena. It is a common knowledge that, when considering Wiener process as a driving noise, the small-time Gaussian approximation of increments very often leads to a good results, such as asymptotic efficiency of estimators and so on; the same can be said to more general diffusion-type models such as continuous Itô semimartingales. Nevertheless, there do exist quite a lot of situations where strong non-Gaussian feature of distributions of small-time increments of data sequence is dominant, making the Gaussianity assumption inappropriate to reflect reality. In particular, at high-frequency time scales such a character often may not be described by a diffusion with compound-Poisson jumps as well, since jumps are then very sparse so that increments may be approximately Gaussian except intervals where a jump occurred. In order to reflect non-Gaussianity, which is one of the stylized facts often observed in real data such as financial returns [1] and to build up a more versatile statistical model, it is of great significance to incorporate a non-Gaussian noise distribution. The feature calls for a more tailor-made estimation procedure when the driving Lévy process is of pure-jump type, for which the approximate Gaussianity in small-time no longer holds true and its statistical inference of which becomes generally more complicated. In this paper we will propose and analyze a new class of SDE models driven by a "stable-like" Lévy process, forming a broad class of Lévy processes, which can even approximate a Wiener process.

1.1. Objective.

Given an underlying complete filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in\mathbb{R}_+}, \mathbb{P})$, we consider a solution to the univariate Markovian SDE

$$dX_t = a(X_t, \alpha)dt + c(X_{t-}, \gamma)dJ_t, \tag{1.1}$$

where we assume:

- The initial random variable X_0 is $\mathcal{F}_0$-measurable;
- J is a pure-jump (càdlàg) Lévy process adapted to the filtration $(\mathcal{F}_t)$, independent of X_0, and having the Lévy-Khintchine representation

$$\mathbb{E}(e^{iuJ_t}) = \exp\left\{t\left(\int_{|z|\le 1}(e^{iuz} - 1 - iuz)\nu(dz) + \int_{|z|>1}(e^{iuz} - 1)\nu(dz)\right)\right\} \tag{1.2}$$

  for $t \in \mathbb{R}_+$ and $u \in \mathbb{R}$;

Date: April 21, 2016.

Key words and phrases. Asymptotic mixed-normality, high-frequency sampling, locally stable Lévy process, stable quasi-likelihood function, stochastic differential equations.

[1] A quotation from [15]: *"The apparent paradox, which has puzzled many a researcher, is that the tails appear to become less heavy for less frequent (e.g., monthly) returns than for more frequent (e.g., daily) returns, a phenomenon not easily explainable by the standard models."*

- The trend coefficient $a : \mathbb{R} \times \Theta_\alpha \to \mathbb{R}$ and scale coefficient $c : \mathbb{R} \times \Theta_\gamma \to \mathbb{R}$ are assumed to be known except for the p-dimensional parameter
$$\theta := (\alpha, \gamma) \in \Theta_\alpha \times \Theta_\gamma = \Theta,$$
with $\Theta_\alpha \in \mathbb{R}^{p_\alpha}$ and $\Theta_\gamma \in \mathbb{R}^{p_\gamma}$ being bounded convex domains.

We will assume that the distribution $\mathcal{L}(h^{-1/\beta} J_h)$ weakly tends to the symmetric stable distribution with index $\beta \in [1, 2)$ (the assumption will be made rigorous in Assumption 2.1), and that the process X is observed only at discrete but high-frequency time instants $t_j^n = jh_n$, $j = 0, 1, \dots, n$, with nonrandom sampling step size $h = h_n \to 0$; it is a trivial matter to remove the equidistance assumption on the sampling times, as long as the ratios of $\min_{j \le n}(t_j - t_{j-1})$ and $\max_{j \le n}(t_j - t_{j-1})$ are bounded in an appropriate order. This paper focuses on the so-called bounded-domain asymptotics:
$$T_n \equiv T \text{ for a fixed terminal sampling time } T \in (0, \infty).$$
This amounts to observing not the complete path $(X_t)_{t \le T}$ but the discretized step process
$$X_t^{(n)} := X_{\lfloor t/h \rfloor h}, \qquad t \in [0, T]. \tag{1.3}$$
We are concerned here with estimation of θ, assuming that the true value $\theta_0 = (\alpha_0, \gamma_0) \in \Theta$ does exist. Due to the lack of a closed-form formula for the transition distribution, a feasible approach based on the *genuine* likelihood function is rarely available. In this paper, we will introduce a novel *non-Gaussian quasi-likelihood function* [2], much extending the prototype mentioned in [33] and [36]. More specifically, under some conditions we will provide a quasi-likelihood estimator $\hat{\theta}_n = (\hat{\alpha}_n, \hat{\gamma}_n)$ such that
$$\left(\sqrt{n} h_n^{1-1/\beta} (\hat{\alpha}_n - \alpha_0),\ \sqrt{n}(\hat{\gamma}_n - \gamma_0) \right)$$
is asymptotically mixed-normally distributed, entailing that the "activity" index β (see (2.2) below) determines the rate of convergence of estimating the trend parameter α. Most notably, even when T_n is fixed we can estimate not only the scale parameter γ but also the drift parameter α, with the explicit asymptotic distribution in hand. To prove the asymptotic mixed normality, we will take a doubly approximate procedure based on the Euler-Maruyama scheme combined with the stable approximation of $\mathcal{L}(h^{-1/\beta} J_h)$ for small h: see Section 3.1.

The model is semiparametric in the sense that we do not completely specify the Lévy measure of $\mathcal{L}(J)$, while supposing the parametric coefficients; of course, the Lévy measure is an infinite-dimensional parameter, so that β never solely determines the distribution $\mathcal{L}(J)$. In estimation of $\mathcal{L}(X)$, it seems desirable (whenever possible) to estimate (α, γ) with leaving the remaining parameters contained in Lévy measure as much as unknown. The proposed quasi-likelihood provides us with a widely applicable tool for this purpose, extending the preceding results on diffusion processes. It gives an estimator having much better asymptotic behavior compared with the Gaussian maximum quasi-likelihood estimator, which was previously studied by [35] and is known to be inconsistent when the target sampling time period is fixed (see below for a literature review). Our results will clarify several interesting phenomena that cannot be shared by the case of diffusion process where J is a standard Wiener process. Also we should mention that use of the Gaussian quasi-likelihood can result in a rather inefficient and even inconsistent estimation, see, e.g., [3] and [35].

Note that we assume from the very beginning that J contains no Gaussian factor. Normally, the simultaneous presence of a non-degenerate diffusion plus a non-null jump part makes parametric-estimation problem much more complicated. Some recent studies have revealed utility of pure-jump models. See the recent papers [24] and [29], which are especially concerned with financial context, however, it is obvious that pure-jump models should be useful for modeling in many application fields where non-Gaussianity of data should be more appropriate, such as signal processing (detection, estimation, etc.), population dynamics, hydrology, radiophysics, turbulence, biological molecule movement, noise-contaminated biosignals, and so on. We also refer to, among many others, [2] and [53] for recent related works in this direction. [3]

[2] Non-Gaussian quasi-likelihoods have not received much attention compared with the popular Gaussian quasi-likelihood; among others, we refer to the recent paper [11] for a certain non-Gaussian quasi-likelihood estimation of possibly heavy-tailed GARCH models, and also to [60] for self-weighted Laplace quasi-likelihood in a time series context.

[3] See also the recent preprint: Klebanov, L. B. and Volchenkova, I. V. (2015), Heavy-tailed distributions in finance: reality or myth? Amateurs viewpoint. arXiv:1507.07735.

Formally, our model (1.1) is a continuous-time analogue to the discrete-time model

$$X_j = a(X_{j-1}, \alpha) + c(X_{j-1}, \gamma)\epsilon_j, \qquad j = 1, \dots, n,$$

where ϵ_j are i.i.d. random variables. By making use of the small-time non-Gaussian stable structure, our model setup enables us to formulate a flexible and unified estimation paradigm, which cannot be shared with the discrete-time counterpart. In particular, our estimation procedure is not effected by heavy-tail property of the noise distribution $\mathcal{L}(J_1)$.

We end this introductory subsection with some remarks on the high-frequency-sampling asymptotics.

- The present bounded-domain asymptotics enables us to "localize" the event, sidestepping *stability (such as the ergodicity)* and *moment-condition* issues on $\mathcal{L}(J_1)$, which is quite often inevitable for developing asymptotic theory for $T_n \to \infty$. To develop an infill asymptotics without ergodicity, however, we need much more than the (martingale) central limit theorem with Gaussian limit: a mixed-normal limit theory for LAQ statistical experiments plays an essential role. Fortunately, we have a very general tool which can cover a setting where an underlying space is of Poisson type carrying a pure-jump Lévy process: Jacod's characterization of the conditionally Gaussian martingales (see [12] and [19]).
- It should be noted that there is no correspondence between actual-time scale and the model-time scale; virtually, we may always set the terminal sampling time T_n to be a fixed value $T > 0$, so that T may represent one day, one month, one year, and whatever it be. However, it is well-known that observed information corresponding to some parameters is stochastically bounded, and thereby cannot be estimated consistently in theory; see [26], and [37] and the the references therein.

1.2. **Some background.** The high-frequency data setting is quite beneficial from statistical point of view, since it can enable us to: formulate explicit approximate estimation procedures; total observing period can be fixed, say $T = 1$ being day, one week, and so forth (and it leads to a realized result, here referring more precisely to the fact that we have not normal but mixed-normal asymptotic distribution in parametric estimation); and, quite often, to keep the model structure rather general, making use of fine continuous-time structure of the model.

In the rest of this paper we will suppress the dependence on n from the notations t_j^n and h_n, and denote by $(\mathbb{P}_\theta)_{\theta\in\Theta}$ the family of the image measures of X given by (1.1) in $\mathbb{D}(\mathbb{R}_+;\mathbb{R})$, the Skorokhod space of càdlàg functions from $\mathbb{R}_+$ to $\mathbb{R}$. For any process Y we will denote by

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

the jth increments, and

$$g_{j-1}(v) = g(X_{t_{j-1}}, v)$$

for a function g having two components x and v, such as $a_{j-1}(\alpha) = a(X_{t_{j-1}}, \alpha)$. Below we will give a brief overview on the related existing literature.

First, concerning the small-time Gaussian approximation, let us recall some basic results in the case of an ergodic diffusion model

$$dX_t = a(X_t, \alpha)dt + c(X_t, \gamma)dw_t$$

with true invariant distribution $\pi(dx;\theta_0)$. Under appropriate conditions we can deduce the asymptotic normality of the Gaussian quasi-maximum likelihood estimator (GQMLE) defined to be any maximizer of

$$\theta \mapsto \sum_{j=1}^n \log\left\{\frac{1}{\sqrt{2\pi c_{j-1}^2(\gamma)h_n}} \exp\left(-\frac{(\Delta_j X - a_{j-1}(\alpha)h_n)^2}{2c_{j-1}^2(\gamma)h_n}\right)\right\}, \tag{1.4}$$

which comes from the "fake" small-time Gaussian approximation the transition probability:

$$\mathcal{L}(X_{t_j}|X_{t_{j-1}} = x) \approx N\left(X_{t_{j-1}} + a_{j-1}(\alpha)h_n,\, c_{j-1}^2(\gamma)h_n\right). \tag{1.5}$$

Then, under appropriate conditions we have the asymptotic normality

$$\begin{aligned} &\left(\sqrt{T_n}(\hat{\alpha}_n - \alpha_0), \sqrt{n}(\hat{\gamma}_n - \gamma_0)\right) \\ &\xrightarrow{\mathcal{L}} N\left(0,\, \mathrm{diag}\left[\left\{\int \left(\frac{\partial_\alpha a}{c}\right)^{\otimes 2}(x,\theta_0)\pi(dx;\theta_0)\right\}^{-1},\, 2\left\{\int \left(\frac{\partial_\gamma (c^2)}{c^2}\right)^{\otimes 2}(x,\theta_0)\pi(dx;\theta_0)\right\}\right]\right), \end{aligned} \tag{1.6}$$

where $\xrightarrow{\mathcal{L}}$ denotes the convergence in distribution. (1.6) clarifies that we can estimate the diffusion parameter γ faster than the drift one α, where we have to let $T_n \to \infty$ for α while we do not for γ (cf. [27], [55], and [58]); we should note that the simple form (1.4) which works under the sampling-frequency condition $nh_n^2 \to 0$ is just for simplicity of exposition, and incorporating the higher-order Itô-Taylor expansion of the one-step conditional mean and variance into the quasi-likelihood enables us to deduce an estimator having the same asymptotic normality as in (1.6). The resulting phenomenon is known to be asymptotically efficient [14]. Also known in the literature is that, even when $T_n \equiv T$ we may estimate γ in an asymptotically efficient manner as

$$\sqrt{n}(\hat{\gamma}_n - \gamma_0) \xrightarrow{\mathcal{L}} MN\left(0,\ 2\left\{\frac{1}{T}\int_0^T \left(\frac{\partial_\gamma(c^2)}{c^2}\right)^{\otimes 2}(X_t,\gamma_0)dt\right\}^{-1}\right) \tag{1.7}$$

by making use of the variant of (1.4):

$$\gamma \mapsto \sum_{j=1}^n \log\left\{\frac{1}{\sqrt{2\pi c_{j-1}^2(\gamma)h_n}}\exp\left(-\frac{(\Delta_j X)^2}{2c_{j-1}^2(\gamma)h_n}\right)\right\}, \tag{1.8}$$

where the drift coefficient is now a non-estimable nuisance element: see [12], [13], and [56] for details. Here and in the sequel, the symbol "MN" stands for the "mixed normal".

Obviously, the above-mentioned features is already in force for the scaled Wiener process with drift $X_t = \alpha t + \gamma w_t$, where the Gaussian quasi-likelihood becomes the genuine likelihood, so that the asymptotics of the MLE becomes trivial: in the independent-increment case where $a(x,\alpha) = \alpha$ and $c(x,\gamma) = \gamma > 0$, (1.6) formally reduces to $(\sqrt{T_n}(\hat{\alpha}_n - \alpha_0),\ \sqrt{n}(\hat{\gamma}_n - \gamma_0)) \xrightarrow{\mathcal{L}} N(0,\ \gamma_0^2 \mathrm{diag}(1,\ 1/2))$. In this paper we will extend the notion of the "local-Gauss" contrast function well-known for diffusions can be extended to the "local-non-Gaussian-stable" contrast function, resulting in an essentially much more efficient estimator. Nevertheless, it should be mentioned that the GQMLE can be also used for the Lévy driven case. Indeed, it turned out by the previous work [35], where the Gaussian part may or may not be present, that adopting the Gaussian quasi-maximum likelihood estimator based on the local-Gauss approximation (1.5) leads to the asymptotic normality only at rate $\sqrt{T_n}$ for both of α and γ, possibly resulting in significant efficiency loss with inevitably requiring that $T_n \to \infty$.

Turning to the pure-jump cases, we proceed along remarks.

Remark 1.1. Consider the Lévy process $X_t = \alpha t + \gamma J_t$ for a standard β-stable Lévy process J associated with the characteristic function $u \mapsto \exp(-|u|^\beta)$. Then the jth increment is $h^{-1/\beta}\Delta_j X = \alpha h^{1-1/\beta} + \gamma h^{-1/\beta}\Delta_j J$, from which we see that the model shows different feature depending on the value β:

- for $\beta \in (1,2)$, the noise part $\gamma h^{-1/\beta}\Delta_j J$ is dominant compared with the drift part $\alpha h^{1-1/\beta}$;
- for the critical case $\beta = 1$ (the Cauchy case), the drift and the noise parts are of the same stochastic order;
- for $\beta \in (0,1)$, the drift part $\alpha h^{1-1/\beta}$ is dominant compared with the noise part $\gamma h^{-1/\beta}\Delta_j J$.

These phenomena turns out to remain the same even when both drift and scale coefficients are randomly time-varying. Our results can cover the case $\beta \in [1,2)$ for much more general non-linear SDE of the form (1.1), revealing an analogous phenomena. □

Remark 1.2. For an explicit example, let us consider estimation of $\theta = (\alpha,\gamma)$ of a Lévy process $X_t = \alpha t + \gamma J_t$ where J is a normal-inverse Gaussian Lévy process, which is locally Cauchy (see [26]). Just for comparison with the GQMLE, $\mathcal{L}(J_t) = NIG(\delta,0,\delta t,0)$ for some $\delta > 0$, so that $\mathbb{E}(J_t) = 0$, $\mathbb{E}(J_t^2) = t$, and $\mathcal{L}(X_t) = NIG(\delta/\gamma, 0, \gamma\delta t, \alpha t)$; see [5] for the details of NIG Lévy processes. Then we have the following.

(1) The GQMLE is asymptotically normal:

$$\left(\sqrt{T_n}(\hat{\alpha}_n - \alpha_0),\ \sqrt{T_n}(\hat{\gamma}_n - \gamma_0)\right) \xrightarrow{\mathcal{L}} N\left(0,\ \gamma_0^2 \mathrm{diag}\left(1,\ \frac{3}{2\delta^2}\right)\right).$$

This can be deduced in a direct manner, following an analogous way to [35]. Indeed, for the GQMLE a simple computation gives the identity

$$(\sqrt{T_n}(\hat{\alpha}_n - \alpha_0),\ \sqrt{T_n}(\hat{\gamma}_n^2 - \gamma_0^2)) = T_n^{-1/2}\sum_{j=1}^n (v_j,\ v_j^2 - \gamma_0^2 h),$$

where $v_j := \Delta_j X - \alpha_0 h$. Even for a general centered and standardized J (i.e. $\mathbb{E}(J_t) = 0$, $\mathbb{E}(J_t^2) = t$) it readily follows from the fact $h^{-1}E(J_h^k) \to \nu_k := \int z^k \nu(dz)$ for $k \geq 2$, where ν denotes the Lévy measure of J, and the Lindeberg-Feller theorem together with the delta method that

$$(\sqrt{T_n}(\hat{\alpha}_n - \alpha_0),\ \sqrt{T_n}(\hat{\gamma}_n - \gamma_0)) \xrightarrow{\mathcal{L}} N_2\left(0,\ \gamma_0^2 \begin{pmatrix} 1 & \nu_3/2 \\ \nu_3/2 & \nu_4/2 \end{pmatrix}\right).$$

In this case, we may leave δ unknown while the value does affect the asymptotic covariance matrix.

(2) Let $\delta = 1$. The QMLE based on the Cauchy likelihood (a special case of the proposed quasi-likelihood) satisfies that

$$(\sqrt{n}(\hat{\alpha}_n - \alpha_0),\ \sqrt{n}(\hat{\gamma}_n - \gamma_0)) \xrightarrow{\mathcal{L}} N_2\left(0,\ 2\gamma_0^2 I_2\right).$$

This is asymptotically efficient, that is, our estimator makes it possible to estimate (α, γ). We emphasize that the asymptotic normality holds even when $\mathcal{L}(J_t) = NIG(\delta', 0, t, 0)$ with leaving $\delta' > 0$ unknown.

Thus very different asymptotic behaviors can occur, our estimator being much better. □

Remark 1.3. We note that consistent estimation of β is possible by two data only. Consider a Lévy process $X_t = \alpha t + \gamma J_t$. Let $\mathcal{L}(h^{-1/\beta}(X_h - h\alpha)) \Rightarrow S_\beta(\gamma)$, and suppose that we have $\delta_n := h^{1-1/\beta}(\hat{\alpha}_n - \alpha) \xrightarrow{p} 0$. Fix any $j \leq n$, and write $\mathcal{S}_j = h^{-1/\beta}(\Delta_j X - \alpha h)$ and

$$\tilde{\beta}_n = \frac{-\log(1/h)}{\log|\Delta_j X - \hat{\alpha}_n h|}.$$

Then we have under $\mathbb{P}_\theta$,

$$\log(1/h)(\tilde{\beta}_n^{-1} - \beta^{-1}) = -\log|\mathcal{S}_j - \delta_n|.$$

It follows from the assumption $\delta_n \xrightarrow{p} 0$ that the right-hand side is $O_p(1)$. To see this, fix any $\epsilon > 0$, and pick an $M' > 0$ such that $\sup_n \mathbb{P}_\theta(|\delta_n| > M') < \epsilon/2$. With this $M' > 0$, we have for $M > M'$ large enough and $\kappa \in (0,1)$,

$$\begin{aligned}
\sup_n \mathbb{P}_\theta\left(\big|\log|\mathcal{S}_j - \delta_n|\big| > M\right) &\leq \epsilon/2 + \sup_n \mathbb{P}_\theta\left(\big|\log|\mathcal{S}_j - \delta_n|\big| > M \ \bigwedge\ |\delta_n| \leq M'\right) \\
&\leq \epsilon/2 + \sup_n \mathbb{P}_\theta\left(|\mathcal{S}_j - \delta_n|^\kappa + |\mathcal{S}_j - \delta_n|^{-\kappa} \gtrsim M \ \bigwedge\ |\delta_n| \leq M'\right) \\
&\leq \epsilon/2 + \sup_n \mathbb{P}_\theta\left(|\mathcal{S}_1|^\kappa \gtrsim M/2\right) + \sup_n \mathbb{P}_\theta\left(|\mathcal{S}_j - \delta_n|^{-\kappa} \gtrsim M/2\right) \\
&\leq \epsilon/2 + \epsilon/4 + \epsilon/4 = \epsilon.
\end{aligned}$$

It may be checked that the joint distribution $\mathcal{L}(\mathcal{S}_j, \delta_n)$ is tight with bounded density. The above implies that $\log(1/h)(\tilde{\beta}_n - \beta)$ is asymptotically $\mathcal{L}(\beta^2 \log|S_\beta|)$-distributed, while the rate $\log(1/h)$ is quite unsatisfactory. □

Remark 1.4. Contrary to the diffusion case, very little is known about asymptotic efficiency phenomenon for the Lévy driven (1.1) with observing (1.3). For local asymptotic normality results when X is a Lévy process, i.e. when $a(x,\alpha)$ and $c(x,\gamma)$ are constants, we refer to [37] for several explicit case studies for Lévy processes, and to [18] for a general locally stable Lévy processes. Recently, [8] proved the LAMN property about the drift parameter α especially when $c(x,\gamma)$ is a given constant and the support of the Lévy measure ν is bounded. The asymptotic efficiency in the sense of Hajék-Jeganathan-Le Cam of the β-stable quasi-likelihood estimator is assured by their LAMN result. Just like that the Gaussian quasi-likelihood is asymptotically efficient for diffusions, concerning the SDE (1.1) driven by a locally β-stable Lévy process we conjecture that our estimator is asymptotically efficient (see also the discussion in Section 3.3). The detailed study of which is the scope of this paper and one of important future works. □

Remark 1.5. For general locally β-stable pure-jump Itô-semimartingale models, there exist many results on asymptotic behavior of the power-variation statistics of the form $\frac{1}{n}\sum_{j=1}^{[nt]} |n^{1/\beta}\Delta_j X|^p$ (here $h_n = 1/n$) in estimation of the integrated(-powered) scale process [50], [51], [52], and [53]. Application of the law of large numbers and the stable convergence in law available in the p-variation literature seems attractive due to their computational simplicity, while it does not seem to be of direct use for our parametric-estimation purpose. □

Section 2 describes our basic model setup. The main results are presented in Section 3. Section 4 is devoted to the proof of the main results.

We end this section with some basic notation. For a variable $x = (x_i)i$, ∂_x^k denotes the kth partial-differentiation operators with respect to the components of x (e.g. $\partial_x = \{\frac{\partial}{\partial x_i}\}_i$ and $\partial_x^2 = \{\frac{\partial^2}{\partial x_i \partial x_j}\}_{i,j}$); given a function $f = f(s_1, \dots, s_k) : S_1 \times \cdots \times S_k \to \mathbb{R}^m$ with $S_i \subset \mathbb{R}^{d_i}$, we write $\partial_{s_1}^{j_1} \dots \partial_{s_k}^{j_k} f$ for the array of partial derivatives of dimension $m \times (\prod_{i=1}^k d_i j_i)$; φ_ξ denotes the characteristic function of a random variable ξ; $M^{\otimes 2} := MM^\top$ for any matrix M with $\top$ denoting the transpose; C denotes a generic positive constant which may vary at each appearance; $a_n \lesssim b_n$ and $a_n \sim b_n$ mean that $a_n \le C b_n$ for every n large enough and that $a_n/b_n \to 1$ for $n \to \infty$, respectively; and finally, the symbols $\xrightarrow{p}$ and $\Rightarrow$ denotes the convergence in $\mathbb{P}$-probability and the weak convergence, respectively.

2. Setup and assumptions

2.1. **Locally stable Lévy process: weak convergence and L^1-local limit theorem.** Recall the Lévy-Khintchine form (1.2) of J. In our study, asymptotic behavior of the distribution $\mathcal{L}(J_h)$ in small-time will play an essential role. As was mentioned in the introduction, the small-time Gaussian approximation (1.5) efficiently works in case of diffusions, where J is a standard Wiener process so that $\mathcal{L}(h^{-1/2} J_h) = N(0,1)$ exactly. The construction of our quasi-likelihood (Section 3.1) will be based on a non-Gaussian-stable counterpart to this fact.

The infinitely divisibility is a vital concept in statistical modeling, building the most general class stemming from cumulate asymptotically negligible independent noises: the celebrated Lindeberg-Feller central limit theorem describes a special case of this phenomenon. The locally infinitely divisible approximation of the likelihood function seems work well, but it is too diverse to make up a reasonably unified estimation procedure. Fortunately and importantly, we know that only strictly stable distribution can occur as a possible asymptotic distribution of a linearly scaled small-time increment of the driving Lévy process (see [7, Proposition 1]). Supposing its locally stable property, we can give a unified approximation procedure for a quasi-likelihood estimation of the model. Specifically, we know from [7, Proposition 1] that if $\mathcal{L}(\kappa_h^{-1} J_h)$ weakly converges as $h \to 0$ to a non-trivial distribution for some positive nonrandom sequence κ_h such that $\kappa_h \to 0$ as $h \to 0$, then it necessarily follows that:

- κ is regularly varying of index $1/\beta$ with $\beta \in (0, 2]$;
- F is strictly β-stable; and $\mathcal{L}(\kappa_h^{-1} J_h)$ admits a bounded continuous Lebesgue density.

We call any Lévy process satisfying the above a *locally stable Lévy process*. We here look at symmetric ν and the choice $\kappa_h = h^{1/\beta}$ with $\beta \in [1, 2)$, the weak limit being the standard symmetric β-stable distribution corresponding to the characteristic function

$$u \mapsto \exp(-|u|^\beta), \quad u \in \mathbb{R}.$$

Denote this distribution by S_β. We refer to [23], [45], and [61] for a comprehensive account of the general theory of stable distributions and/or processes.

Now we assume that

$$\nu \text{ is symmetric and } \mathcal{L}(h^{-1/\beta} J_h) \Rightarrow S_\beta \text{ as } h \to 0; \tag{2.1}$$

we will assume that $\beta \in [1, 2)$ later on. The value β equals the Blumenthal-Getoor index:

$$\beta := \inf\left\{ b \ge 0 : \int_{|z| \le 1} |z|^b \nu(dz) < \infty \right\}, \tag{2.2}$$

which measures degree of J's jump activity. We note that many locally stable Lévy processes with finite variance can exhibit both small-time non-Gaussianity (e.g., heavy-tailed property (excess kurtosis)) and large-time Gaussianity (i.e. central-limit effect), which are consistent with stylized facts observed in some actual phenomena; see, e.g., [40] and [44] and the references therein. [4]

[4] See also the recent preprint: Klebanov, L. B. and Volchenkova, I. V. (2015), Heavy-tailed distributions in finance: reality or myth? Amateurs viewpoint. arXiv:1507.07735.

The locally stable property in (2.1) can be characterized by the β-stable-like behavior of ν around the origin. Let us briefly discuss how to verify it. By (1.2) with the symmetry of ν, the random variable $h^{-1/\beta}J_h$ has no drift and its Lévy measure is given by

$$\nu_h(B) := h\nu\big(\{z;\, h^{-1/\beta}z \in B\}\big).$$

Then, according to [46, Theorem 8.7] we have $\mathcal{L}(h^{-1/\beta}J_h) \Rightarrow S_\beta$ as $h \to 0$ if and only if

$$\int f(y)\nu_h(dy) \to \int f(y)\nu_0(dy), \quad h \to 0,$$

for every continuous bounded function f vanishing in a neighborhood of the origin, where ν_0 is the Lévy measure of S_β, namely $\nu_0(dy) = g_0(y)dy$ for $g_0(y) = c_\beta|z|^{-1-\beta}$ with (cf. [46, Lemma 14.11])

$$c_\beta := \frac{1}{2}\left\{\frac{1}{\beta}\Gamma(1-\beta)\cos\left(\frac{\beta\pi}{2}\right)\right\}^{-1}.$$

($\lim_{\beta\to 1} c_\beta = \pi^{-1}$).

A convenient sufficient condition can be given in terms of the Lévy density of the most active part of J: for example, it is enough that for a neighborhood U of the origin ν can be bounded below on $U \setminus \{0\}$ by a β-stable-like absolute continuous part. Specifically, let the Lévy measure ν be symmetric and decomposed as

$$\nu(dz) = \nu^\sharp(dz) + \nu^\flat(dz), \tag{2.3}$$

where $\nu^\sharp(dz) = g(z)dz$ in a neighborhood of the origin where

$$g(z) = \frac{c_\beta}{|z|^{1+\beta}}\bar{g}(z) \tag{2.4}$$

for a bounded continuous non-negative function $\bar{g}$ satisfying that $\lim_{|z|\to 0}\bar{g}(z) = 1$, and where

$$\nu^\flat\big(\{z \neq 0; \epsilon \le |z| \le 1\}\big) \lesssim \epsilon^{-\beta'}, \quad \epsilon \in (0,1], \tag{2.5}$$

for some $\beta' < \beta$. Then (2.1) is satisfied. The condition (2.5) means that the $\nu^\flat$-part of J is strictly less active than the $\nu^\sharp$-part; equivalently, writing $J = J^\sharp + J^\flat$ with independent Lévy processes $J^\sharp$ and $J^\flat$ corresponding to the Lévy measures $\nu^\sharp$ and $\nu^\flat$, respectively, we have $h^{-1/\beta}J_h^\sharp \Rightarrow S_\beta$ and $h^{-1/\beta}J_h^\flat \xrightarrow{p} 0$ as $h \to 0$. In particular, the infinitely divisible distribution $\mathcal{L}(h^{-1/\beta}J_h^\sharp)$ admits the Lévy density $g_h(z) := c_\beta|z|^{-1-\beta}\bar{g}(h^{1/\beta}z)$.

The condition (2.4) is satisfied by many concrete Lévy processes for which $\mathcal{L}(J_1)$ is generalized hyperbolic (except for the normal gamma), Student-t, Meixner, stable, and the (normal) tempered stable distributions. Under (2.4) it is not difficult to show that $|\varphi_{h^{-1/\beta}J_h}(\cdot)| \in \bigcap_{q>0} L^q(du)$, so that, thanks to Sharpe's criterion [48], f_h is everywhere positive and $\log f_h$ is always well-defined.

As a matter of fact, the mere weak convergence $\mathcal{L}(h^{-1/\beta}J_h) \Rightarrow S_\beta$ is not enough for our purpose. Under (2.1), we denote by f_h the Lebesgue density of $\mathcal{L}(h^{-1/\beta}J_h)$:

$$f_h(y) = \frac{1}{2\pi}\int e^{-iuy}\varphi_{h^{-1/\beta}J_h}(u)du.$$

Denote by ϕ_β the bounded smooth Lebesgue density of S_β. [5] We now describe the assumptions on J given by (1.2), requiring an L^1-local limit theorem with some convergence rate:

Assumption 2.1 (Structure of J)**.** (i) *The Lévy measure ν is symmetric and $\mathcal{L}(h^{-1/\beta}J_h) \Rightarrow S_\beta$ as $h \to 0$ for some $\beta \in [1,2)$.*

(ii) *There exist positive constants C_0 and C_ν such that $\int_{|z|>y}\nu(dz) \le C_\nu y^{-\beta}$ for $y \in (0, C_0]$.*

(iii) $\limsup_{n\to\infty}\sqrt{n}\int|f_{h_n}(y) - \phi_\beta(y)|dy < \infty$.

Remark 2.2. • Assumption 2.1(ii) roughly says that ν behaves like the Lévy measure of S_β distribution near the origin and that the tail of ν is equivalent to or lighter than that of S_β; hence it is trivial when $\mathcal{L}(J_1) = S_\beta$. In particular, if the tail of ν is as just described, then Assumption 2.1(ii) holds under (2.3), (2.4) and (2.5).

[5] Some asymptotic behavior of ϕ_β can be found in [4, Eq. (2.5)–(2.10)] and the references therein.

- Under Assumption 2.1(ii) we can apply [30, Theorem 2(a) and (c)] to conclude that
$$\mathbb{E}\left(\sup_{t\le h}|J_t|^r\right)\lesssim h^{r/\beta}.$$
- If $\|f_h-\phi_\beta\|_\infty\lesssim h^{a_\nu}$ for some $a_\nu>0$ (see Lemma 2.4 below), then Assumptions 2.1(ii) and (iii) together imply that
$$\limsup_{n\to\infty}\int|y|^\kappa|f_h(y)-\phi_\beta(y)|dy=0 \tag{2.6}$$
for some $\kappa\in(0,\beta)$ small enough. Indeed, noting that $\sup_{h\in(0,1]}\int|y|^\kappa f_h(y)dy+\int|y|^\kappa\phi_\beta(y)dy<\infty$, we have $\int|y|^\kappa|f_h(y)-\phi_\beta(y)|dy\lesssim h^{-\kappa/2}\|f_h-\phi_\beta\|_\infty+\int_{|y|\ge h^{-1/2}}|f_h(y)-\phi_\beta(y)|dy\lesssim h^{\kappa/2}+\int_{|y|\ge h^{-1/2}}|y|^\kappa f_h(y)dy+\int_{|y|\ge h^{-1/2}}|y|^\kappa\phi_\beta(y)dy\to0$.

□

We will need Assumption 2.1(iii) for proving the central limit theorem for the quasi-score function evaluated at the true value. Unfortunately, contrary to Assumption 2.1(i) and (ii) verification of Assumption 2.1(iii) seems non-trivial even when we know the explicit form of ν; obviously, it is automatic if $\mathcal{L}(J_1)=S_\beta$. A trivial sufficient conditions for Assumption 2.1(iii) is that there exist a constant $c_\nu>0$ and a positive sequence (ϵ_n) such that
$$\sqrt{n}\epsilon_n\to0\quad\text{and}\quad|f_{h_n}(y)-\phi_\beta(y)|\le\epsilon_n(1\wedge|y|^{-(1+c_\nu)}); \tag{2.7}$$
then, $\lim_{n\to\infty}\int\{\log(1+|y|)\}^{1+\epsilon}|f_{h_n}(y)-\phi_\beta(y)|dy=0$ and $\lim_{n\to\infty}\sqrt{n}\int|f_{h_n}(y)-\phi_\beta(y)|dy=0$. However, this still seems not so simple to verify.

Here is an explicit example where Assumption 2.1(iii) holds.

Example 2.3 (Normal inverse-Gaussian Lévy process)**.** Let J be an NIG Lévy process such that $\mathcal{L}(J_t)=NIG(\eta,0,t,0)$, where $\eta>0$ may be unknown. The probability density f_h and Lévy density g_h of $\mathcal{L}(h^{-1}J_h)$ are given by
$$\begin{aligned}f_h(y)&=\frac{e^{\eta h}}{\pi(1+y^2)}\eta h\sqrt{1+y^2}K_1\big(\alpha h\sqrt{1+y^2}\big),\\ g_h(z)&=\frac{1}{\pi|z|^2}\eta h|z|K_1(\eta h|z|),\end{aligned}$$
respectively. It follows that $\mathcal{L}(h^{-1}J_h)$ weakly tends to the standard Cauchy distribution, whose probability density and Lévy density are given by $y\mapsto\pi^{-1}(1+y^2)^{-1}$ and $z\mapsto\pi^{-1}|z|^{-2}$, respectively, hence Assumption 2.1(i) follows. Assumption 2.1(ii) is obvious. For Assumption 2.1(iii), we note that $f_h(y)-\phi_1(y)=\phi_1(y)u_h(y)$, where
$$u_h(y):=A(\eta h\sqrt{1+y^2})e^{\eta h}+\{(e^{\eta h}-1)/(\eta h)\}\eta h$$
with $A(\epsilon):=\epsilon K_1(\epsilon)-1$. We deduce from [16, 9.6.8, 9.6.26, 9.7.2] that:
- $K_1(z)\sim\sqrt{\pi/(2z)}e^{-z}$ as $z\to\infty$;
- $\partial A'(\epsilon)=-\epsilon K_0(\epsilon)\sim-\epsilon\log(1/\epsilon)\to0$ as $\epsilon\to0$, hence $\sup_{\epsilon>0}\epsilon^{-k}|A(\epsilon)|<\infty$ for $k\in(0,1]$ by L'Hopital's Rule and the smoothness of the Bessel function on $(0,\infty)$.

Then it follows that $|u_h(y)|\lesssim h^k(1+y^2)^{k/2}+h\lesssim h^k(1+|y|^k)$ for any $k\in(0,1]$. Since $\sqrt{n}h_n^k\lesssim\sqrt{T_n}h_n^{k-1/2}\lesssim h_n^{k-1/2}$ (here (T_n) is presupposed to be bounded), the condition (2.7) follows on taking any $k\in(1/2,1)$. □

2.2. **Verification of Assumption 2.1(iii).** We now discuss how to verify Assumption 2.1(iii) in terms of $\nu(dz)$. For this purpose, we refer to the following lemma, which provides us with, under the symmetry of ν, some easy conditions under which we can specify the rate of convergence in the sup-norm local limit theorem.

Lemma 2.4. *Assume the decomposition (2.3) with (2.4) and (2.5), and assume that there exists a constant $\delta>0$ such that*
$$\bar{g}(z)=1+O(|z|^\delta),\qquad|z|\to0. \tag{2.8}$$

Then there exits a constant $a_\nu > 0$ such that [6]

$$\|f_h - \phi_\beta\|_\infty \lesssim h^{a_\nu} \tag{2.9}$$

with the value a_ν being given as follows:

- *if $\nu^\flat(\mathbb{R}) < \infty$, $a_\nu = \delta'/\beta$ for any $\delta' \in (0,\beta)\cap(0,\delta]$;*
- *if $\nu^\flat(\mathbb{R}) = \infty$, $a_\nu = (\delta'/\beta) \wedge (1-\beta_{1+}/\beta)$ for any $\delta' \in (0,\beta)\cap(0,\delta]$ and $\beta_{1+} > \beta_1$, where $\beta_1 < \beta$ denotes the Blumental-Getoor index of the Lévy measure $\nu^\flat$.*

See [32, Lemma 4.4(b)-(iv),(v)] for the proof of Lemma 2.4.

Lemma 2.5. *Let Assumptions 2.1(i) and (ii) and (2.9) hold, and suppose that $h_n \lesssim n^{-c}$ for some $c > 0$. Then Assumption 2.1(iii) hold with $\limsup_{n\to\infty} \sqrt{n}\int |f_{h_n}(y) - \phi_\beta(y)|dy = 0$ if*

$$\frac{1}{2c}\left(\frac{1}{\beta} + 1\right) < a_\nu. \tag{2.10}$$

Proof. Let $\kappa > 0$ be a constant and divide the domain of integration into $\{y;\, |y| \le n^\kappa\}$ and its complement, to deduce that $\overline{\Delta}_n := \sqrt{n}\int |f_h(y) - \phi_\beta(y)|dy \lesssim n^{1/2+\kappa}\|f_h - \phi_\beta\|_\infty + \sqrt{n}\int_{|y|>n^\kappa}(f_h + \phi_\beta)(y)dy \lesssim n^{1/2+\kappa-ca_\nu} + \sqrt{n}\int_{|y|>n^\kappa}|y|^{-1-\beta}dy \lesssim n^{(1/2+\kappa-ca_\nu)\vee(1/2-\kappa\beta)}$. Hence, to conclude that $\overline{\Delta}_n \to 0$ it suffices to pick κ such that both $1/2+\kappa-ca_\nu < 0$ and $1/2-\kappa\beta < 0$ hold. Such a κ does exist under (2.10). □

In particular, if $c = 1$ and $a_\nu = \delta'/\beta < 1$ under the assumptions of Lemmas 2.4 and 2.5, then the condition (2.10) reduces to

$$\frac{1}{2}(1+\beta) < \delta'.$$

This condition entails $\beta > 1$, preferring a bigger $\delta > 0$; no restriction on β arises if we can take, e.g., $\delta = 2$ such as the case of $\bar{g}(z) = \exp(-cz^2)$ for $c > 0$.

As a seemingly different way to verify Assumption 2.1(iii), we refer to the following inequality, which states that the $L^1(dy)$-norm estimate can be deduced from the sup-norm estimate, with a slight loss in convergence rate.

Lemma 2.6. *Let $\mathfrak{p}_n$ and $\mathfrak{p}$ be probability densities on $\mathbb{R}^d$, and $r > 0$ a number such that*

$$\sup_{u>0} u^r \int_{|y|>u} \mathfrak{p}(y)dy < \infty.$$

Then we have

$$\int |\mathfrak{p}_n(y) - \mathfrak{p}(y)|\, dy = O\left(\|\mathfrak{p}_n - \mathfrak{p}\|_\infty^{r/(r+d)}\right), \qquad n \to \infty.$$

This lemma was given in [47], the proof being simple: for any $A > 0$ we have $\int |\mathfrak{p}_n(y) - \mathfrak{p}(y)|\, dy = 2\int (\mathfrak{p} - \mathfrak{p}_n)_+(y)dy \le 2\{\int_{|x|\le A}(\mathfrak{p} - \mathfrak{p}_n)_+(y)dy + \int_{|x|>A}(\mathfrak{p} - \mathfrak{p}_n)_+(y)dy\} \lesssim A^d\|\mathfrak{p}_n - \mathfrak{p}\|_\infty + \int_{|y|>A}\mathfrak{p}(y)dy \lesssim A^d\|\mathfrak{p}_n - \mathfrak{p}\|_\infty + A^{-r}$, hence optimizing the upper bound with respect to A leads to $\int |\mathfrak{p}_n(y) - \mathfrak{p}(y)|\, dy \lesssim \|\mathfrak{p}_n - \mathfrak{p}\|_\infty^{r/(r+d)}$.

It holds that $\sup_u u^\beta \int_{|y|>u}\phi_\beta(y)dy < \infty$, hence if f_{h_n} and ϕ_β fulfils (2.9) and if $h_n \lesssim n^{-\kappa}$ for $\kappa > 0$, then we can apply Lemma 2.6 with $r = \beta$:

$$\sqrt{n}\int |f_{h_n}(y) - \phi_\beta(y)|dy \lesssim \sqrt{n} h_n^{a_\nu\beta/(\beta+1)} \lesssim n^{1/2 - a_\nu\kappa\beta/(\beta+1)}.$$

Thus Assumption 2.1(iii) holds for $a_\nu > (\beta+1)/(2\beta\kappa)$ with $\sqrt{n}\int |f_h(y) - \phi_\beta(y)|dy \to 0$; in particular, this is the case if $\kappa = 1$, $\beta > 1$, and $a_\nu \le 1$ can be arbitrarily close to 1.

Note that Lemma 2.6 becomes unworkable if $d \ge 2$ were large because of the severely stringent condition on the rate of $\|\mathfrak{p}_n - \mathfrak{p}\|_\infty \to 0$.

Remark 2.7. The criterion Lemma 2.6 is not sharp, as seen from Example 2.3: there, we can show that $\sup_{y\in\mathbb{R}}|f_h(y) - \phi_1(y)| \lesssim h$, so that (2.9) holds with $a_\nu = 1$, but then Lemma 2.6 only tells us that $\sqrt{n}\int |f_{h_n}(y) - \phi_1(y)|dy = O(1)$. □

[6] A seemingly related result is [28], which studied the rate of convergence in the locally stable-limit theorem for triangular array of random variables.

There seems to be no trivial inclusion relation between the criteria based on Lemmas 2.5 and 2.6.

2.3. **Locally stable stochastic differential equation.** Our objective is the SDE (1.1):

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

We are concerned only with the coefficients smooth enough.

Assumption 2.8 (Regularity of the coefficients)**.** (1) *The functions $a(\cdot, \alpha_0)$ and $c(\cdot, \gamma_0)$ are globally Lipschitz and of class $\mathcal{C}^2(\mathbb{R})$.*

(2) *$a(x, \cdot) \in \mathcal{C}^3(\Theta_\alpha)$ and $c(x, \cdot) \in \mathcal{C}^3(\Theta_\gamma)$ for each $x \in \mathbb{R}$.*

(3) $\sup_{\theta\in\overline{\Theta}} \left\{ \max_{0\le k\le 3} \max_{0\le l\le 2} \left(\left|\partial_\alpha^k \partial_x^l a(x,\alpha)\right| + \left|\partial_\gamma^k \partial_x^l c(x,\gamma)\right| \right) + c^{-1}(x,\gamma) \right\} \lesssim 1 + |x|^C.$

The standard theory (e.g. [22, III §2c.]) ensures that the SDE admits a unique strong solution as a functional of X_0 and the Poisson random measure driving J.

Assumption 2.9 (Model identifiability)**.** *The random functions $t \mapsto (a(X_t, \alpha), c(X_t, \gamma))$ and $t \mapsto (a(X_t, \alpha_0), c(X_t, \gamma_0))$ on $[0, T]$ a.s. coincide if and only if $\theta = \theta_0$.*

3. Stable quasi-likelihood estimation

3.1. **Heuristic for construction.** To motivate our quasi-likelihood, we here present a formal heuristic argument. In what follows we will abbreviate $\int_j$ as $\int_j$. In view of the Euler approximation, we have under $\mathbb{P}_\theta$

$$\begin{aligned} X_{t_j} &= X_{t_{j-1}} + \int_j a(X_s, \alpha)ds + \int_j c(X_{s-}, \gamma)dJ_s \\ &\approx X_{t_{j-1}} + a_{j-1}(\alpha)h + c_{j-1}(\gamma)\Delta_j J, \end{aligned}$$

from which we may expect that

$$\epsilon_j(\theta) = \epsilon_{n,j}(\theta) := \frac{\Delta_j X - h a_{j-1}(\alpha)}{h^{1/\beta} c_{j-1}(\gamma)} \approx h^{-1/\beta}\Delta_j J \tag{3.1}$$

in an appropriate sense. Then it follows from (2.1) that for each n the sequence $\{\epsilon_j(\theta)\}_{j\le n}$ under $\mathbb{P}_\theta$ will approximately form a S_β-i.i.d. random variables.

Assume that the process X admits the time-homogeneous transition Lebesgue density under $\mathbb{P}_\theta$, say $p_h(x, y; \theta)dy = \mathbb{P}_\theta(X_h \in dy | X_0 = x)$, and let $\mathbb{E}_\theta^{j-1}$ denote the expectation operator under $\mathbb{P}_\theta$ conditional on $\mathcal{F}_{t_{j-1}}$. Then, we proceed with the following *twofold* approximation of the conditional distribution $\mathcal{L}(X_{t_j} | X_{t_{j-1}})$ under $\mathbb{P}_\theta$:

$$\begin{aligned} p_h(X_{t_{j-1}}, X_{t_j}; \theta) &= \frac{1}{2\pi}\int \exp(-iuX_{t_j})\mathbb{E}_\theta^{j-1}\{\exp(iuX_{t_j})\}du \\ &\approx \frac{1}{2\pi}\int \exp(-iuX_{t_j})\mathbb{E}_\theta^{j-1}\left[\exp\left\{iu(X_{t_{j-1}} + a_{j-1}(\alpha)h + c_{j-1}(\gamma)\Delta_j J)\right\}\right]du \qquad (3.2) \\ &= \frac{1}{2\pi}\int \exp\left\{-iu(\Delta_j X - a_{j-1}(\alpha)h)\right\}\varphi_{h^{-1/\beta}J_h}(c_{j-1}(\gamma)h^{1/\beta}u)du \\ &= \frac{1}{c_{j-1}(\gamma)h^{1/\beta}}\frac{1}{2\pi}\int \exp\{-iu\epsilon_j(\theta)\}\varphi_{h^{-1/\beta}J_h}(u)du \\ &= \frac{1}{c_{j-1}(\gamma)h^{1/\beta}} f_h\left(\epsilon_j(\theta)\right) \\ &\approx \frac{1}{c_{j-1}(\gamma)h^{1/\beta}}\phi_\beta\left(\epsilon_j(\theta)\right), \qquad (3.3) \end{aligned}$$

where, concerning the two approximations, we note that:

- (3.2) becomes exact if and only if both a and c are constant (i.e. X is a Lévy process);
- (3.3), where the locally stable property comes into the picture, becomes exact if $\mathcal{L}(J_1) = S_\beta$.

Since the genuine likelihood function equals $\theta \mapsto \sum_{j=1}^n \log p_h(X_{t_{j-1}}, X_{t_j}; \theta)$, the last observation suggests to estimate θ_0 by a maximizer of the random function (ignoring the known factor "$\log(h^{-1/\beta})$")

$$\mathbb{H}_n(\theta) := \sum_{j=1}^n \Big(-\log c_{j-1}(\gamma) + \log\phi_\beta\left(\epsilon_j(\theta)\right) \Big), \tag{3.4}$$

which is a.s. well-defined thanks to the positivity of ϕ_β. We call this $\mathbb{H}_n$ the stable quasi-likelihood function, and then define the *stable quasi-maximum likelihood estimator (SQMLE)* by any

$$\hat{\theta}_n = (\hat{\alpha}_n, \hat{\gamma}_n) \in \operatorname*{argmax}_{\theta\in\overline{\Theta}} \mathbb{H}_n(\theta), \tag{3.5}$$

where $\overline{\Theta}$ denotes the closure of Θ, hence there always exists at least one such $\hat{\theta}_n$; obviously, the SQMLE is the non-Gaussian-stable counterpart to the Gaussian quasi-likelihood.

Our contrast function involves, in addition to the activity index β, the computationally demanding β-stable density ϕ_β. In a subsequent versions, we will discuss about a possibility of how to handle the case of unknown β and more tractable variants of $\mathbb{H}_n$.

Remark 3.1. It may happen that the density f_h is explicit for each $h > 0$ for some J not exactly β-stable. Then we could sidestep the β-stable approximation (3.3), considering instead

$$p_h(X_{t_{j-1}}, X_{t_j}; \theta) \approx \frac{1}{c_{j-1}(\gamma)h^{1/\beta}} f_h\left(\epsilon_j(\theta)\right).$$

The normal-inverse Gaussian J (Example 2.3) is such an example: if $\mathcal{L}(J_1) = NIG(\alpha, 0, \delta, 0)$, then f_h is the explicit density of the $NIG(\alpha h, 0, \delta, 0)$-distribution. Nevertheless and obviously, such an "exact $\mathcal{L}(h^{-1/\beta}J_h)$-case" consideration much diminishes the class of admissible J, and going in this direction entails individual case studies. □

3.2. **Asymptotics of SQMLE: main results.** Building on what we have seen above, we now state the asymptotic behavior of the SQMLE defined through (3.4) and (3.5). Recall that we are assuming that $\beta \in [1, 2)$ and that the terminal sampling time $T_n \equiv T$. For $\mathcal{F}$-measurable random variables $\mu = \mu(\omega) \in \mathbb{R}^p$ and $\Sigma = \Sigma(\omega) \in \mathbb{R}^p \otimes \mathbb{R}^p$, we denote by $MN_p(\mu, \Sigma)$ the p-dimensional mixed-normal distribution corresponding to the characteristic function $v \mapsto \mathbb{E}[\exp\{i\mu[v] - (1/2)\Sigma[v, v]\}]$. That is to say, when $Y \sim MN_p(\mu, \Sigma)$, Y is defined on an appropriate extension of the original probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and is equivalent in distribution to a random variable $\mu + \Sigma^{1/2}Z$ for $Z \sim N_p(0, I_p)$ independent of $\mathcal{F}$.

We let

$$g_\beta(y) := \frac{\partial}{\partial y} \log \phi_\beta(y) = \frac{\partial \phi_\beta}{\phi_\beta}(y), \qquad k_\beta(y) := 1 + y g_\beta(y),$$

both being finite as g_β and k_β are bounded. Note that $\int g_\beta(y)\phi_\beta(y)dy = \int k_\beta(y)\phi_\beta(y)dy = 0$, and also that $\int g_\beta(y) f_h(y)dy = 0$ because of the symmetry of f_h. Let further

$$\mathfrak{b}_{\beta,n}(\nu) := \sqrt{n}\int k_\beta(z)\{f_h(y) - \phi_\beta(y)\}dy = \int k_\beta(z) f_h(y)dy, \qquad \mathfrak{b}_\beta(\nu) := \lim_{n\to\infty} \mathfrak{b}_{\beta,n}(\nu). \tag{3.6}$$

We also write:

$$C_\alpha(\beta) = \int g_\beta^2(y)\phi_\beta(y)dy, \qquad C_\gamma(\beta) = \int k_\beta^2(y)\phi_\beta(y)dy,$$

$$\mu_{T,\gamma}(\theta_0; \beta) = -\frac{1}{T}\int_0^T \frac{\partial_\gamma c(X_t, \gamma_0)}{c(X_t, \gamma_0)} dt, \tag{3.7}$$

$$\Sigma_{T,\alpha}(\theta_0) = \frac{1}{T}\int_0^T \frac{\{\partial_\alpha a(X_t, \alpha_0)\}^{\otimes 2}}{c^2(X_t, \gamma_0)} dt, \tag{3.8}$$

$$\Sigma_{T,\gamma}(\gamma_0) = \frac{1}{T}\int_0^T \frac{\{\partial_\gamma c(X_t, \gamma_0)\}^{\otimes 2}}{c^2(X_t, \gamma_0)} dt. \tag{3.9}$$

The next theorem shows the asymptotic mixed normality of the SMQLE, the main claim of this paper.

Theorem 3.2. *Suppose that Assumptions 2.1, 2.8, and 2.9 hold. Then we have*

$$\left(\sqrt{n}h_n^{1-1/\beta}(\hat{\alpha}_n - \alpha_0),\ \sqrt{n}(\hat{\gamma}_n - \gamma_0)\right) \xrightarrow{\mathcal{L}} MN_p\left(m_T(\theta_0; \beta),\ \Gamma_T(\theta_0; \beta)^{-1}\right), \tag{3.10}$$

where

$$m_T(\theta_0; \beta) := \left(0,\ \{C_\gamma(\beta)\Sigma_{T,\gamma}(\gamma_0)\}^{-1}\mu_{T,\gamma}(\theta_0; \beta)\mathfrak{b}_\beta(\nu)\right),$$
$$\Gamma_T(\theta_0; \beta) := \operatorname{diag}\left\{C_\alpha(\beta)\Sigma_{T,\alpha}(\theta_0),\ C_\gamma(\beta)\Sigma_{T,\gamma}(\gamma_0)\right\}.$$

We gave some sufficient conditions for $\mathfrak{b}_\beta(\nu)=0$ in Section 2.2. Unfortunately, we do not know how we can evaluate $\mathfrak{b}_\beta(\nu)$ in a unified manner. Indeed, precise estimation of f_h may generally require not only the locally stable property but also the full information of the Lévy measure ν. It should be noted that even when we have a full parametric form of ν, it may contain a parameter which can be consistently estimated only when $T_n\to\infty$; see [37] for several such examples.

According to the continuity of the random mapping $\theta\mapsto(\Sigma_{T,\alpha}(\theta),\Sigma_{T,\gamma}(\gamma))$, we can readily deduce by applying the uniform law of large numbers presented in Lemma 4.4 that

$$\begin{aligned}\hat{\mu}_{T,\gamma,n} &:= -\frac{1}{n}\sum_{j=1}^n \frac{\partial_\gamma c_{j-1}(\hat{\gamma}_n)}{c_{j-1}(\hat{\gamma}_n)} \xrightarrow{p} \mu_{T,\gamma}(\theta_0;\beta),\\ \hat{\Sigma}_{T,\alpha,n} &:= \frac{1}{n}\sum_{j=1}^n \frac{\{\partial_\alpha a_{j-1}(\hat{\alpha}_n)\}^{\otimes 2}}{c_{j-1}^2(\hat{\gamma}_n)} \xrightarrow{p} \Sigma_{T,\alpha}(\theta_0),\\ \hat{\Sigma}_{T,\gamma,n} &:= \frac{1}{n}\sum_{j=1}^n \frac{\{\partial_\gamma c_{j-1}(\hat{\gamma}_n)\}^{\otimes 2}}{c_{j-1}^2(\hat{\gamma}_n)} \xrightarrow{p} \Sigma_{T,\gamma}(\gamma_0).\end{aligned}$$

It turns out that the quasi-score function ($\mathcal{F}$-)stably converges in distribution, so that the Studentization via the continuous-mapping theorem is straightforward, making the asymptotic distributional result feasible:

Corollary 3.3. *Under the assumptions of Theorem 3.2, we have*

$$\left(\hat{\Gamma}_{T,\alpha,n}^{1/2}\sqrt{n}h_n^{1-1/\beta}(\hat{\alpha}_n-\alpha_0),\,\hat{\Gamma}_{T,\gamma,n}^{1/2}\left\{\sqrt{n}(\hat{\gamma}_n-\gamma_0)-\hat{m}_{T,\gamma,n}\right\}\right)\xrightarrow{\mathcal{L}} N_p(0,I_p), \tag{3.11}$$

where $\hat{m}_{T,\gamma,n} := \{C_\gamma(\beta)\hat{\Sigma}_{T,\gamma,n}\}^{-1}\hat{\mu}_{T,\gamma,n}\mathfrak{b}_{\beta,n}(\nu)$, $\hat{\Gamma}_{T,\alpha,n} := C_\alpha(\beta)\hat{\Sigma}_{T,\alpha,n}$ *and* $\hat{\Gamma}_{T,\gamma,n} := C_\gamma(\beta)\hat{\Sigma}_{T,\gamma,n}$. *Especially if* $\mathfrak{b}_\beta(\nu)=0$, *then*

$$\left(\hat{\Gamma}_{T,\alpha,n}^{1/2}\sqrt{n}h_n^{1-1/\beta}(\hat{\mu}_n-\mu_0),\,\hat{\Gamma}_{T,\gamma,n}^{1/2}\sqrt{n}(\hat{\sigma}_n-\sigma_0)\right)\xrightarrow{\mathcal{L}} N_p(0,I_p).$$

Again we emphasize that the SQMLE is consistent and asymptotically mixed normal for any fixed terminal sampling time T, while, of course, finite-sample performances of the SQMLE depends on the value T. We may deduce a large-time counterpart of Theorem 3.2 and Corollary 3.3 under the the ergodicity, resulting in a asymptotic (never "mixed-") normality with completely analogous form. The details will be given elsewhere.

3.3. **Remarks and discussion.** Theorem 3.2 reveals several interesting phenomena, including some essential difference between the Gaussian and non-Gaussian stable quasi-likelihood estimators. [7]

(1) The asymptotic distribution of $\hat{\gamma}_n$ (resp. $\hat{\alpha}_n$) is normal if the mapping $x\mapsto\frac{\partial_\gamma c(x,\gamma_0)}{c(x,\gamma_0)}$ (resp. $x\mapsto\frac{\partial_\alpha a(x,\alpha_0)}{c(x,\gamma_0)}$) is constant; in particular, this is the case if X is a Lévy process.

(2) The estimators $\hat{\alpha}_n$ and $\hat{\gamma}_n$ are *asymptotically orthogonal* whereas not necessarily independent due possible non-Gaussianity in the limit. The orthogonality is theoretically beneficial in view of adaptive estimation (cf. [9] and [25]).

(3) Let $\mathfrak{b}_\beta(\nu)=0$ for simplicity, and consider the case of $\beta\in(1,2)$. We can rewrite (3.10) as

$$\left(n^{1/\beta-1/2}(\hat{\alpha}_n-\alpha_0),\ \sqrt{n}(\hat{\gamma}_n-\gamma_0)\right)\xrightarrow{\mathcal{L}} MN_p\left(0,\,\mathrm{diag}\big(T^{-2(1-1/\beta)}\Sigma_{T,\alpha}^{-1}(\theta_0;\beta),\,\Sigma_{T,\gamma}^{-1}(\gamma_0;\beta)\big)\right). \tag{3.12}$$

If fluctuation of X is virtually stable along time in the sense that both of $\Sigma_{T,\alpha}(\theta_0;\beta)$ and $\Sigma_{T,\gamma}(\gamma_0;\beta)$ do not vary so much with the terminal sampling time T, then the asymptotic covariance matrix of $\hat{\alpha}_n$ will tend to get smaller (resp. larger) in magnitude for a larger (resp. smaller) T. We emphasize that this feature with respect to T is *non-asymptotic*.

(4) Taking $\beta=2$ in the factors $C_\alpha(\beta)\Sigma_{T,\alpha}(\theta_0)$ and $C_\gamma(\beta)\Sigma_{T,\gamma}(\gamma_0)$ results in those of the diffusion case [27], also [55]. In this respect, our locally stable approximation methodology formally generalizes the local-Gauss approximation. Since the latter one for the diffusion case is known to be asymptotically efficient (see [13]), it is expected that the stable quasi-likelihood is asymptotically efficient as well for general locally stable SDE; it is the case for some particular cases, see [8], [18], and the references therein.

[7] We remark that some of the items below go for the ergodic case as well.

(5) The locally Cauchy case, where $\beta = 1$ and $\mathbb{H}_n$ is fully explicit, may be of special interest:

$$\left(\sqrt{n}(\hat{\alpha}_n - \alpha_0),\ \sqrt{n}(\hat{\gamma}_n - \gamma_0)\right)$$
$$\xrightarrow{\mathcal{L}} MN_p\left(0,\ 2\,\mathrm{diag}\left(\frac{1}{T}\int_0^T \frac{\{\partial_\alpha a(X_t,\alpha_0)\}^{\otimes 2}}{c(X_t,\gamma_0)^2}dt,\ \frac{1}{T}\int_0^T \frac{\{\partial_\gamma c(X_t,\gamma_0)\}^{\otimes 2}}{c(X_t,\gamma_0)^2}dt\right)^{-1}\right).$$

This formally extends the i.i.d. model from the Cauchy population. The Cauchy quasi-likelihood has been also investigated in the robust-regression literature; see e.g. [38] and [39] for a breakdown-point result in some relevant models. It would be interesting to study their SDE-model counterparts.

(6) Table 1 summarizes the rates of convergence concerning the three quasi-likelihood functions when the SDE is

$$dX_t = a(X_t,\alpha)dt + c(X_{t-},\gamma)dZ_t, \tag{3.13}$$

for a driving Lévy process Z, where the coefficient (a,c) is correctly specified.

Quasi-likelihood	Driving Lévy process Z	Rates of convergence $\hat{\alpha}_n$	$\hat{\gamma}_n$	
(i) Gauss	Wiener process	$\sqrt{nh_n}$	$\sqrt{n}$	Ref. [27]
(ii) Gauss	Lévy process with jumps	$\sqrt{nh_n}$	$\sqrt{nh_n}$	Ref. [35]
(iii) Non-Gaussian stable	Locally β-stable ($\beta < 2$)	$\sqrt{n}h_n^{1-1/\beta}$	$\sqrt{n}$	

TABLE 1. A comparison of the Gaussian and non-Gaussian stable quasi-likelihood functions for the SDE (3.13), where the coefficient (a,c) is correctly specified.

We refer to [34] for a handy test statistic for distinguishing the cases (i) and (ii) in Table 1. The statistics is constructed through partial sums of the self-normalized powered residuals and possesses the desirable asymptotic properties: asymptotically distribution-free under the null model with no jumps, and consistent against the presence of arbitrary jump part.

(7) Taking *higher-order increments* in the time-series literature is a simple classical device to remove the trend effect. It will be also beneficial in high-frequency data models. We refer to [50] for some asymptotics of the power variation statistics for a pure-jump Itô semimartingale, and also to [6] for the multipower-variation statistics for a Brownian semi-stationary process. In the present context, we may make use of the second-order increments

$$\Delta_j^{(2)}X := \Delta_j X - \Delta_{j-1}X, \qquad j = 2,\dots,n,$$

in order to diminish the small-time trend effect, resulting in a broader admissible range of the value of the activity index β, which we are assuming to be equal to or greater than one [8]; in the case $\beta \in (0,1)$, the naive Euler scheme spoils because small-time variation of X is governed by that of the trend coefficient $a(x,\alpha)$. Further, by using $(\Delta_j^{(2)}X)_j$ we may effectively get rid of the asymmetry of $\mathcal{L}(J_1)$. Very roughly, since we have

$$\begin{aligned}h^{-1/\beta}\Delta_j^{(2)}X &= h^{-1/\beta}\left(\int_{t_{j-1}} a(X_s,\alpha_0)ds - \int_{t_{j-2}} a(X_s,\alpha_0)ds\right)\\ &\quad + h^{-1/\beta}\left(\int_{t_{j-1}} c(X_{s-},\gamma_0)dJ_s - \int_{t_{j-2}} c(X_{s-},\gamma_0)dJ_s\right),\end{aligned}$$

the drift-part fluctuation of $h^{-1/\beta}\Delta_j^{(2)}X$ is of order $o_p(h^{1-1/\beta})$, while the noise-part fluctuation is of order $O_p(1)$; also, the principal part of the latter is $c_{j-2}(\gamma_0)h^{-1/\beta}(\Delta_{j-1}J - \Delta_{j-2}J)$, which has (after the localization if necessary; see Section 4.1) the $\mathcal{F}_{t_{j-2}}$-conditional mean zero. The resulting asymptotic behavior of the corresponding SQMLE will be almost the same as in the

[8]Another possibility is to use the trajectory-fitting type contrast function as in [31].

case of first-order increments, the only price we have to pay being a (slightly) more complicated forms of the asymptotic random covariance matrix.

(8) In case of $\beta > 1$, the following *stepwise-estimation* strategy may be used [9]:
 (a) First, we estimate γ via the SQMLE $\hat{\gamma}_n$ with regarding $a(x,\alpha) \equiv 0$;
 (b) Second, we estimate α by the modified quasi-likelihood function $\alpha \mapsto \mathbb{H}_n(\alpha, \hat{\gamma}_n)$.
 The resulting asymptotic distribution of the two-step SMQLE would be the same. It is expected that this strategy will enable us to handle the mixed-parameter-coefficient case, by which we mean the SDE model having the parametric form slightly more general than (1.1):
$$dX_t = a(X_t, \alpha, \gamma)dt + c(X_{t-}, \gamma)dJ_t.$$
 This strategy is the non-Gaussian counterpart of (the first two steps of) the adaptive estimation developed by [55] for diffusions. [10]

(9) Sometimes it is possible to refine and/or modify the Euler-approximation in construction of the SQMLE. It could be possible to follow the proof of the main claims with placing the Euler-type residual $\epsilon_j(\theta)$ by
$$\epsilon_j(\theta;\beta) := \frac{X_{t_j} - \mu_{j-1}^n(\alpha)}{h^{1/\beta}\sigma_{j-1}(\gamma)},$$
 for some functions $\mu^n(x,\alpha)$ and $\sigma^n(x,\gamma)$ on $\mathbb{R} \times \Theta_\alpha$; here, the random functions $\mu_{j-1}^n(\alpha)$ and $\sigma_{j-1}^n(\gamma)$ are roughly of $O_p(1)$ and will serve as instantaneous location and scale of X_{t_j}, respectively. Then, we would look at the f-quasi-likelihood function
$$\mathbb{H}_n(\theta;\beta) := \sum_{j=1}^n \log\left\{\frac{1}{\sigma_{j-1}^n(\gamma)} f\left(\epsilon_j(\theta;\beta)\right)\right\}. \tag{3.14}$$
 Trivially, the naive Euler-type residual mentioned before corresponds to the choices $\mu^n(x,\alpha) = x + h_n a(x,\alpha)$ and $\sigma^n(x,\gamma) = c(x,\gamma)$, but adopting (3.14) allows us to encompass some formal modifications:
 - We can make a martingale in a directly when linear-in-state drift, such as $\mu^n(x,\alpha) = e^{-\alpha}x$ when $a(x,\alpha) = -\alpha x$.
 - We may take $\sigma(x,\gamma) \equiv 1$ with regarding γ as a nuisance parameter, as was done for the lease-squares type estimation in [31].
 - As was mentioned in the previous item, we would take $\mu^n \equiv 0$ in case of $\beta > 1$; in this case, we may regard the trend coefficient as an infinite-dimensional nuisance parameter, and of course even may set it to be non-Markovian type. For example, the point-delay SDE model studied in [43] and [49] would be covered.

 Although the generic f enables us to take quasi-likelihoods other than ϕ_β into account, we have to be careful about how asymptotic behaviors get changed accordingly.

(10) Because of the bounded-domain asymptotics, we may deal with without essential change the following more general setting: we observe a possibly non-Markovian sample $\{(X_{t_j}, Y_{t_j})\}_{j=0}^n$ from the $(\mathcal{F}_t)$-adapted process described by the SDE
$$dX_t = a(X_t, Y_t; \alpha)dt + c(X_{t-}, Y_{t-}; \gamma)dJ_t, \tag{3.15}$$
 where $Y = (Y^k)$ is a multivariate covariate process. If, for example, Y is driven by J plus another Lévy process J' independent of J, then the martingale-representation argument used in Sections 4.2 and 4.4.1 may remain valid with trivial modifications (even when J' has non-null Gaussian part, of course). The presence of the process Y makes it possible to incorporate both exogenous- and endogenous-random effect. Then we may obtain a new class of *non-ergodic stochastic-regression models*, for which it seems worth studying associated covariate-process selection problems.

(11) We may consider even more general class than (3.15): the driving noise may be no longer a Lévy process as long as the locally stable structure stays valid. It can be vary as n increases, say

[9] It seems to be the case also for $\beta = 1$, when making use of the second-order increments.

[10] The stepwise estimation for a class of general Lévy driven SDE can be done based on the Gaussian quasi-likelihood function: Masuda, H. and Uehara, Y. (2016), On stepwise estimation of Lévy driven stochastic differential equation. In preparation.

J^n, with a slight formal extension of the underlying statistical experiments which varies along n, denoted by $(\Omega^n, \mathcal{A}^n, \{\mathbb{P}^n_\theta; \theta \in \Theta\})$. Then, the locally stable property is read as the convergence

$$\mathcal{L}(h_n^{-1/\beta} J^n_{h_n}) \Rightarrow \mathcal{S}_\beta(1), \qquad n \to \infty.$$

For example, this allows us to consider $J^n_t = \epsilon_n w_t + J_t$ with $h_n^{-1/\beta}\epsilon_n\sqrt{h_n} \to 0$. Even more generally, J^n may be no longer a Lévy process. For example, it may be a semimartingale, the local characteristic of $h_n^{-1/\beta} J^n$ being possibly (X, Y)-dependent, say

$$(B^n(X_{t-}, Y_{t-}),\, C^n(X_{t-}, Y_{t-}),\, \nu^n(X_{t-}, Y_{t-}; dz))\,,$$

where, in order to derive the S_β-distribution in small time, we should have both

$$h_n^{-1/\beta}|B^n(x_\cdot, y_\cdot)| + h_n^{-2/\beta}|C^n(x_\cdot, y_\cdot)| \to 0,$$
$$\nu^n(x_\cdot, y_\cdot; dz) \to c_\beta |z|^{-(1+\beta)}dz.$$

in an appropriate sense. For example: many types of point-processes noise with randomly perturbed intensity can be considered; also, one can consider a randomly perturbed "Ornstein-Uhlenbeck" type processes given by the SDE $dX_t = \{-\alpha X_t + \epsilon_n^\alpha \mu(X_{t-})\}dt + \{\gamma + \epsilon_n^\gamma \sigma(X_{t-})\}dJ_t$, where J is a locally β-stable Lévy process and where $|\epsilon_n^\alpha| \vee |\epsilon_n^\gamma| \to 0$ as $n \to \infty$ at a speed fast enough, so that we can suitably approximate the statistical experiment by the "skeleton" model $dX_t = -\alpha X_t dt + \gamma dJ_t$.

(12) We here focus on cases of correctly specified coefficient. If we remove this constraint, the asymptotic result might drastically get changed; this issue of *misspecified coefficients* will be studied elsewhere. See [54] for the case of diffusion: if the model for diffusion coefficient is truly misspecified in their sense, the rate of convergence of the GQMLE is no longer $\sqrt{n}$ but the slower $\sqrt{nh_n}$, thereby implying that the usual infill asymptotics for estimating diffusion coefficient over a fixed time domain breaks down.

(13) The *uniform tail-probability estimate* of the form

$$\sup_n \mathbb{P}\left\{ \left| \left(\sqrt{n} h_n^{1-1/\beta}(\hat{\alpha}_n - \alpha_0),\ \sqrt{n}(\hat{\gamma}_n - \gamma_0) \right) \right| > r \right\} \le C\delta(r), \qquad r > 0,$$

for some sequence $\delta(r) \to 0$ as $r \to \infty$ is a very important tool for various statistical analyses including prediction, higher-order statistics, and model assessment such as AIC-type (convergence of moments) and BIC-type (expansion of the marginal quasi-likelihood). In this respect, analysis of the statistical random fields associated with the stable quasi-likelihood is one of important future works. We refer to [56] for the detailed Gaussian quasi-likelihood analysis for volatility estimation of a class of continuous stochastic-regression models; see also [57] and [59] for related previous studies.

4. Proofs

Throughout this section, Assumptions 2.1, 2.8, and 2.9 are in force.

4.1. Localization: elimination of big jumps.

Most of the key moment estimates involved in the proofs, such as Burkholder's inequality, fail to hold if $\mathcal{L}(J_1)$ is heavy-tailed. We begin with a localization of the underlying probability space by eliminating possible big jumps of J, enabling us to proceed as if $\mathbb{E}(|J_1|^q) < \infty$ for every $q > 0$. This is a simple yet very powerful technique to sidestep a series of probability and/or moment estimates when dealing with cases of fixed T. The point here is that, since our main results are concerned with the weak properties we may conveniently focus on a subset $\Omega_{K,T} (\in \mathcal{F}) \subset \Omega$ if we can control the probability $\mathbb{P}(\Omega_{K,T})$ to be arbitrarily close to 1. This "localization" procedure is nowadays standard in the context of limit theory for (multi)power-variation statistics, and has been considered for quite general semimartingale models. We refer the interested reader to [21, Section 4.4.1].

We here proceed without resorting to the general localization result. Recall the Lévy-Khintchine representation (1.2). Let $\mu(dt, dz)$ denote the Poisson random measure having the intensity measure $dt \otimes \nu(dz)$, and $\tilde{\mu}(dt, dz)$ its compensated version. Fix any $K > 0$. Then, $\int_{1<|z|\le K} z\nu(dz) = 0$ since ν is assumed to be symmetric, and the Lévy-Itô decomposition of J takes the form

$$J_t = \int_{|z|\le K} z\tilde{\mu}(ds, dz) + \int_{|z|>K} z\mu(ds, dz) =: M^K_t + A^K_t,$$

where M^K is a purely-discontinuous martingale and A^K the compound-Poisson process independent of M^K; that is to say, the symmetry assumption of ν makes the parametric form of the drift coefficient unaffected by elimination of big jumps of J. Since $\sup_t |\Delta M_t^K| \le K$, we have $\mathbb{E}(|M_t^K|^q) < \infty$ for any $t \in \mathbb{R}_+$ and $q > 0$; more precisely, if we have $K = \inf\{a > 0;\ \mathrm{supp}(\nu) \subset \{z;\ |z| \le a\}\}$, then by [46, Thm.26.1] we have $\mathbb{E}\{\exp(r|J_t|\log|J_t|)\} < \infty$ for each $t > 0$ and $r \in (0, K^{-1})$. [11] Further, the event

$$\Omega_{K,T} := \big\{\omega \in \Omega;\ \mu\big((0,T],\ \{z; |z| > K\}\big) = 0\big\} \cap \{\omega \in \Omega;\ |X_0| \le K\}$$

has the probability $\exp\{-T\int_{|z|>K}\nu(dz)\}$, which gets arbitrarily close to 1 with K large enough. Let $(X_t^K)_{t\in[0,T]}$ be given by a solution process to the SDE

$$dX^K = a(X_t^K, \alpha_0)dt + c(X_{t-}^K, \gamma_0)dM_t^K,$$

which obviously admits a strong solution as a functional of (X_0, M^K). We have $X_t(\omega) = X_t^K(\omega)$ for $t \in [0,T]$ and $\omega \in \Omega_{K,T}$.

To state the localization lemma, we need further notation. Let $\zeta_n : \overline{\Theta} \to \mathbb{R}$ be $\mathcal{F}_T^X := \sigma(X_t : t \le T)$-measurable random functions. For clarity we write $\zeta_n(\theta; X)$, specifying the dependence on X. We introduce the extended probability space of the form

$$\big(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}}(d\omega, d\omega')\big) = \big(\Omega \times \Omega', \mathcal{F} \otimes \mathcal{F}', \mathbb{P}(d\omega)\mathbb{Q}(\omega, d\omega')\big)$$

with $\mathbb{Q}$ denoting a transition probability from $(\Omega, \mathcal{F})$ to $(\Omega', \mathcal{F}')$; see [21, Section 2.1.4] for details. Let $\xi_n(X) \in \mathbb{R}^k$ be $\mathcal{F}_T^X$-measurable random variables defined on $(\Omega, \mathcal{F})$, and $\xi_0(X, \eta) \in \mathbb{R}^k$ a random variable defined on $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})$ where η is a random element independent of $\mathcal{F}$: more specifically, $\xi_0(X,\eta)(\omega,\omega') = \xi_0(X(\omega), \eta(\omega'))$.

Finally, let us recall that, given random variables G_n and G_∞ taking their values in a some metric space E, where the latter is defined on $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})$, we say that G_n converges stably in law to G_∞, denoted by $G_n \xrightarrow{\mathcal{L}_s} G_\infty$, if $\mathbb{E}\{f(G_n)U\} \to \tilde{\mathbb{E}}\{f(G_\infty)U\}$ for every bounded $\mathcal{F}$-measurable random variable $U \in \mathbb{R}$ and every continuous function $f : E \to \mathbb{R}$; see [22, Chapter VI] for details. This mode of convergence entails that $(G_n, H_n) \xrightarrow{\mathcal{L}} (G_\infty, H_\infty)$ for every random variables H_n and H_∞ such that $H_n \xrightarrow{p} H_\infty$. We refer to [20], [21], [22], and also [17] for comprehensive accounts of the stable convergence.

Lemma 4.1. *With the aforementioned setting, we have the following.*

(i) *If* $\sup_\theta |\zeta_n(\theta; X^K)| \xrightarrow{p} 0$ *for every* $K > 0$ *large enough, then* $\sup_\theta |\zeta_n(\theta; X)| \xrightarrow{p} 0$.

(ii) *If* $\xi_n(X^K) \xrightarrow{\mathcal{L}_s} \xi_0(X^K, \eta)$ *for every* $K > 0$ *large enough, then* $\xi_n(X) \xrightarrow{\mathcal{L}} \xi_0(X, \eta)$.

Proof. (i) Suppose that $\sup_\theta |\zeta_n(\theta; X^K)| \xrightarrow{p} 0$ for any $K > 0$. Given any $\epsilon, \epsilon' > 0$ we may take a $K > 0$ so large that $\mathbb{P}(\Omega_{K,T}^c) < \epsilon'$, so that $\limsup_n \mathbb{P}\{\sup_\theta |\zeta_n(\theta; X)| > \epsilon\} \le \epsilon' + \limsup_n \mathbb{P}\{\sup_\theta |\zeta_n(\theta; X^K)| > \epsilon\} \le \epsilon'$ and the convergence $\sup_\theta |\zeta_n(\theta; X)| \xrightarrow{p} 0$ follows.

(ii) Supposing that $\xi_n(X^K) \xrightarrow{\mathcal{L}_s} \xi_0(X^K, \eta)$ for every $K > 0$, we want to deduce that $\xi_n(X) \xrightarrow{\mathcal{L}} \xi_0(X, \eta)$. Pick any $\epsilon > 0$ and continuous bounded function f, and then take sufficiently large $K > 0$ so that $\mathbb{P}(\Omega_{K,T}^c) < \epsilon/(2\|f\|_\infty)$. The term $|\mathbb{E}\{f(\xi_n(X))\} - \tilde{\mathbb{E}}\{f(\xi_0(X,\eta))\}|$ is bounded by

$$\begin{aligned}
&\big|\mathbb{E}\{f(\xi_n(X)); \Omega_{K,T}\} - \tilde{\mathbb{E}}\{f(\xi_0(X,\eta)); \Omega_{K,T}\}\big| + \big|\mathbb{E}\{f(\xi_n(X)); \Omega_{K,T}^c\}\big| + \big|\tilde{\mathbb{E}}\{f(\xi_0(X,\eta)); \Omega_{K,T}^c\}\big| \\
&\le \big|\mathbb{E}\{f(\xi_n(X^K)); \Omega_{K,T}\} - \tilde{\mathbb{E}}\{f(\xi_0(X^K,\eta)); \Omega_{K,T}\}\big| + 2\|f\|_\infty \mathbb{P}(\Omega_{K,T}),
\end{aligned}$$

hence $\limsup_n |\mathbb{E}\{f(\xi_n(X))\} - \tilde{\mathbb{E}}\{f(\xi_0(X,\eta))\}| \le \epsilon$ as required. ☐

Based on Lemma 4.1, we may and do suppose that

$$\exists K > 0, \quad \mathbb{P}\left(\forall t \in [0,T],\ |\Delta J_t| \le K\right) = 1 \tag{4.1}$$

in what follows. For notational convenience, we keep using the notation X instead of X^K.

[11] The moment estimate for Lévy processes in small time is interesting in its own right. Several authors have studied asymptotic behavior of the moment $\mathbb{E}\{f(J_h)\}$ as $h \to 0$ for a suitable function f, for which we refer to the following recent papers as well as the references therein for recent developments on this subject: Deng, C. S. and Schilling, R. L. On shift Harnack inequalities for subordinate semigroups and moment estimates for Lévy processes. *Stochastic Process. Appl.*, to appear; Kühn, F. (2015), Existence and estimates of moments for Lévy-type processes. arXiv:1507.07907.

Following the argument [21, Section 2.1.5] together with Gronwall's inequality under the global Lipschitz condition of $(a(\cdot,\alpha_0),c(\cdot,\gamma_0))$, we see that for any $q\ge 2$ and $s\in[0,T]$,

$$\mathbb{E}\left(\sup_{t\le T}|X_t|^q\right)\le C(T,K,q),\qquad \sup_{t\in[s,s+h]\cap[0,T]}\mathbb{E}(|X_t-X_s|^q|\mathcal{F}_s)\lesssim h(1+|X_s|^C). \tag{4.2}$$

In particular, $\limsup_{\delta\to 0}\sup_{|t-s|\le\delta}\mathbb{E}\left(|X_t-X_s|^q\right)=0$.

4.2. **Preliminary asymptotics.** Throughout this section, we focus on the random function

$$U_n(\theta):=\sum_{j=1}^n \pi_{j-1}(\theta)\eta(\epsilon_j(\theta)),$$

where $\pi:\mathbb{R}\times\overline{\Theta}\to\mathbb{R}^k\otimes\mathbb{R}^m$ and $\eta:\mathbb{R}\to\mathbb{R}^m$ are measurable functions; this form of $U_n(\theta)$ will appear in common in the proofs of the consistency and asymptotic mixed normality of the SQMLE, and the results in this section will be repeatedly used later.

We abbreviate $\mathbb{E}(\cdot|\mathcal{F}_{t_{j-1}})$ as $\mathbb{E}^{j-1}(\cdot)$. Write $U_n(\theta)$ as the sum of

$$\begin{aligned}U_{1,n}(\theta)&:=\sum_{j=1}^n \pi_{j-1}(\theta)\Big(\eta(\epsilon_j(\theta))-\mathbb{E}^{j-1}\{\eta(\epsilon_j(\theta))\}\Big),\\ U_{2,n}(\theta)&:=\sum_{j=1}^n \pi_{j-1}(\theta)\mathbb{E}^{j-1}\{\eta(\epsilon_j(\theta))\}.\end{aligned}$$

Given a doubly indexed random function $F_{nj}(\theta)$ on $\overline{\Theta}$ and a positive sequence (a_n) we write:

$$F_{nj}(\theta)=\begin{cases} o_p^*(a_n) & \text{if}\quad \sup_{j\le n}\sup_\theta|F_{nj}(\theta)|=o_p(a_n);\\ O_p^*(a_n) & \text{if}\quad \sup_{j\le n}\sup_\theta|F_{nj}(\theta)|=O_p(a_n);\\ O_{L^q}^*(a_n) & \text{if}\quad \sup_n\sup_{j\le n}\mathbb{E}\left(\sup_\theta|a_n^{-1}F_{nj}(\theta)|^q\right)<\infty.\end{cases}$$

4.2.1. *Uniform estimate of the martingale part.* We begin with the martingale part $U_{1,n}$.

Lemma 4.2. *Suppose that:*

(i) $\pi\in\mathcal{C}^1(\mathbb{R}\times\Theta)$ *and* $\sup_\theta\left(|\pi(x,\theta)|+|\partial_\theta\pi(x,\theta)|\right)\lesssim 1+|x|^C$ *for every* $x\in\mathbb{R}$;

(ii) $\eta\in\mathcal{C}^1(\mathbb{R})$ *and* $|\eta(y)|+|y||\partial\eta(y)|\lesssim 1+\log(1+|y|)$.

Then we have $U_{1,n}(\theta)=O_{L^q}^*(\sqrt{n})$ *for every* $q>0$*, hence in particular* $U_{1,n}(\theta)=O_p^*(\sqrt{n})$.

Proof. Since we are assuming that the parameter space Θ is a bounded convex domain, the Sobolev inequality [1, p.415] is in force: for each $q>p$, we have $\mathbb{E}\left(\sup_\theta|n^{-1/2}U_{1,n}(\theta)|^q\right)\lesssim\sup_\theta\mathbb{E}\left(|n^{-1/2}U_{1,n}(\theta)|^q\right)+\sup_\theta\mathbb{E}\left(|n^{-1/2}\partial_\theta U_{1,n}(\theta)|^q\right)$. To achieve the proof, it therefore suffices to show that both $\{n^{-1/2}U_{1,n}(\theta)\}$ and $\{n^{-1/2}\partial_\theta U_{1,n}(\theta)\}$ are L^q-bounded for each θ and $q>0$. We fix any $q>0$ and θ in the rest of this proof.

Put $\chi_j(\theta)=\pi_{j-1}(\theta)\left(\eta(\epsilon_j(\theta))-\mathbb{E}^{j-1}\{\eta(\epsilon_j(\theta))\}\right)$, so that $U_{1,n}(\theta)=\sum_{j=1}^n\chi_j(\theta)$. Under the present regularity conditions we may pass the differentiation χ_j with respect to θ under the operator $\mathbb{E}^{j-1}$:

$$\begin{aligned}\partial_\theta\chi_j(\theta)&=\partial_\theta\pi_{j-1}(\theta)\Big(\eta(\epsilon_j(\theta))-\mathbb{E}^{j-1}\{\eta(\epsilon_j(\theta))\}\Big)\\ &\quad+\pi_{j-1}(\theta)\Big(\partial\eta(\epsilon_j(\theta))\partial_\theta\epsilon_j(\theta)-\mathbb{E}^{j-1}\{\partial\eta(\epsilon_j(\theta))\partial_\theta\epsilon_j(\theta)\}\Big).\end{aligned}\tag{4.3}$$

For each n the sequences $\{\chi_j(\theta)\}_j$ and $\{\partial_\theta\chi_j(\theta)\}_j$ form a martingale difference array with respect to $(\mathcal{F}_{t_j})$, hence, in view of Burkholder's inequality for martingale difference arrays, the required L^q-boundedness of $\{n^{-1/2}U_{1,n}(\theta)\}$ and $\{n^{-1/2}\partial_\theta U_{1,n}(\theta)\}$ follows on showing that $\sup_{j\le n}\mathbb{E}(|\chi_j(\theta)|^q)\lesssim 1$ and $\sup_{j\le n}\mathbb{E}(|\partial_\theta\chi_j(\theta)|^q)\lesssim 1$.

Observe that for $\beta\ge 1$,

$$\begin{aligned}|\epsilon_j(\theta)|^r&=\left|h^{-1/\beta}c_{j-1}^{-1}(\gamma)\{\Delta_j X-ha_{j-1}(\alpha)\}\right|^r\\ &\lesssim(1+|X_{t_{j-1}}|^C)\left\{|h^{-1/\beta}\Delta_j X|^r+h^{r'(1-1/\beta)}(1+|X_{t_{j-1}}|^C)\right\}\\ &\lesssim(1+|X_{t_{j-1}}|^C)\left(|h^{-1/\beta}\Delta_j X|^r+1\right).\end{aligned}\tag{4.4}$$

Pick an $r\in(0,\beta)$. Under the local β-stable property $\mathcal{L}(h^{-1/\beta}J_h)\Rightarrow S_\beta$ the family $(|h^{-1/\beta}J_h|^r)_{h\in(0,1]}$ for each $r\in(0,\beta)$ is uniformly integrable, so that $\sup_{h\in(0,1]}\mathbb{E}(|h^{-1/\beta}J_h|^r)<\infty$. Then it follows from (4.2) and the last estimate combined with the linear growth property of $a(\cdot,\alpha_0)$, Burkholder's inequality for the stochastic integral with respect to J, and the global Lipschitz property of $c(\cdot,\gamma_0)$ that

$$\begin{aligned}
\mathbb{E}^{j-1}\left(|h^{-1/\beta}\Delta_j X|^r\right) &\lesssim h^{r(1-1/\beta)}\left(\frac{1}{h}\int_j \mathbb{E}^{j-1}\{|a(X_s,\alpha_0)|^2\}ds\right)^{r/2}\\
&\quad + h^{-r/\beta}\mathbb{E}^{j-1}\left(\left|\int_j (c(X_s,\gamma_0)-c_{j-1}(\gamma_0))dJ_s\right|^r\right)\\
&\quad + (1+|X_{t_{j-1}}|^C)\mathbb{E}(|h^{-1/\beta}J_h|^r)\\
&\lesssim (1+h^{r(1-1/\beta)})(1+|X_{t_{j-1}}|^C)+h^{-r/\beta}\left(\int_j \mathbb{E}^{j-1}(|X_s-X_{t_{j-1}}|^2)ds\right)^{r/2}\\
&\lesssim (1+h^{r(1-1/\beta)})(1+|X_{t_{j-1}}|^C)+h^{-r/\beta}\left\{h^2(1+|X_{t_{j-1}}|^C)\right\}^{r/2}\\
&\lesssim 1+|X_{t_{j-1}}|^C. \qquad (4.5)
\end{aligned}$$

Using (4.4) and (4.5) together with the disintegration, we arrive at the estimate $\mathbb{E}\{(1+|X_{t_{j-1}}|^C)|\epsilon_j(\theta)|^r\}\lesssim 1+\sup_{t\le T}\mathbb{E}(|X_t|^C)\lesssim 1$ valid for $r\in(0,\beta)$. Now it is easy to deduce the estimate $\sup_{j\le n}\mathbb{E}(|\chi_j(\theta)|^q)\lesssim 1$.

Turning to the proof of $\sup_{j\le n}\mathbb{E}(|\partial_\theta\chi_j(\theta)|^q)\lesssim 1$, we note that $\partial_\alpha\epsilon_j(\theta)=-h^{1-1/\beta}\frac{\partial_\alpha a_{j-1}(\alpha)}{c_{j-1}(\gamma)}$ and $\partial_\gamma\epsilon_j(\theta)=-\frac{\partial_\gamma c_{j-1}(\gamma)}{c_{j-1}(\gamma)}\epsilon_j(\theta)$. By (4.3) that the components of $\partial_\theta\chi_j(\theta)$ consists of the terms

$$\begin{aligned}
&\pi^{(1)}_{j-1}(\theta)\Big(\eta(\epsilon_j(\theta))-\mathbb{E}^{j-1}\{\eta(\epsilon_j(\theta))\}\Big),\\
&\pi^{(2)}_{j-1}(\theta)\Big(\partial\eta(\epsilon_j(\theta))-\mathbb{E}^{j-1}\{\partial\eta(\epsilon_j(\theta))\}\Big),\\
&\pi^{(3)}_{j-1}(\theta)\Big(\epsilon_j(\theta)\partial\eta(\epsilon_j(\theta))-\mathbb{E}^{j-1}\{\epsilon_j(\theta)\partial\eta(\epsilon_j(\theta))\}\Big)
\end{aligned}$$

for some $\pi^{(i)}(x,\theta)$, $i=1,2,3$, all satisfying the conditions imposed on $\pi(x,\theta)$. Also taking the conditions on η into account, we can exactly follow the previous proof to deduce the estimate $\sup_{j\le n}\mathbb{E}(|\partial_\theta\chi_j(\theta)|^q)\lesssim 1$. The proof is complete. □

Under the assumptions of Lemma 4.2 we have

$$\frac{1}{nh^{1-1/\beta}}U_{1,n}(\theta)=O_p^*\left((\sqrt{n}h^{1-1/\beta})^{-1}\right)=o_p^*(1). \qquad (4.6)$$

4.2.2. *Uniform moment-order estimate of the predictable part.* Next we turn to the predictable (compensator) part $U_{2,n}$. Let us introduce the notation:

$$\begin{aligned}
&\delta'_j(\gamma)=\frac{c_{j-1}(\gamma_0)}{c_{j-1}(\gamma)}h^{-1/\beta}\Delta_j J, \qquad \mathfrak{b}(x,\theta)=c^{-1}(x,\gamma)\{a(x,\alpha_0)-a(x,\alpha)\},\\
&a^{\Delta}_{j-1}(s)=a(X_s,\alpha_0)-a_{j-1}(\alpha_0), \qquad c^{\Delta}_{j-1}(s)=c(X_s,\alpha_0)-c_{j-1}(\alpha_0),\\
&r_j(\gamma)=\frac{h^{-1/\beta}}{c_{j-1}(\gamma)}\int_j a^{\Delta}_{j-1}(s)ds+\frac{h^{-1/\beta}}{c_{j-1}(\gamma)}\int_j c^{\Delta}_{j-1}(s)dJ_s,
\end{aligned}$$

so that

$$\epsilon_j(\theta)=\delta'_j(\gamma)+h^{1-1/\beta}\mathfrak{b}_{j-1}(\theta)+r_j(\gamma).$$

By the Taylor expansion we can write

$$U_{2,n}(\theta)=U^0_{2,n}(\theta)+U'_{2,n}(\theta)+U''_{2,n}(\theta), \qquad (4.7)$$

where, with $\overline{r}_j(\theta;\eta):=\int_0^1\partial\eta(\delta'_j(\gamma)+h^{1-1/\beta}\mathfrak{b}_{j-1}(\theta)+sr_j(\gamma))ds$ and $\pi'(x,\theta):=\pi(x,\theta)c^{-1}(x,\gamma)$,

$$\begin{aligned}
U^0_{2,n}(\theta)&=\sum_{j=1}^n \pi_{j-1}(\theta)\mathbb{E}^{j-1}\left\{\eta\left(\delta'_j(\gamma)+h^{1-1/\beta}\mathfrak{b}_{j-1}(\theta)\right)\right\},\\
U'_{2,n}(\theta)&=h^{-1/\beta}\sum_{j=1}^n \pi'_{j-1}(\theta)\mathbb{E}^{j-1}\left(\overline{r}_j(\theta;\eta)\int_j a^{\Delta}_{j-1}(s)ds\right),
\end{aligned}$$

$$U''_{2,n}(\theta) = h^{-1/\beta} \sum_{j=1}^{n} \pi'_{j-1}(\theta) \mathbb{E}^{j-1}\left(\overline{r}_j(\theta;\eta) \int_j c^{\Delta}_{j-1}(s-)dJ_s \right).$$

One of the key ingredients in the proofs of the main results is a uniform law of large numbers for $(nh^{1-1/\beta})^{-1}U_{2,n}(\theta)$; Lemma 4.3 below reveals that the terms $U'_{2,n}(\theta)$ and $U''_{2,n}(\theta)$ have no contribution.

For later reference, let us state Itô's formula for $t > s$:

$$\begin{aligned} \psi(X_t) &= \psi(X_s) + \int_s^t \partial\psi(X_{u-})dX_u \\ &\quad + \int_s^t \int \{\psi(X_{u-} + c(X_{u-},\gamma_0)z) - \psi(X_{u-}) - \partial\psi(X_{u-})c(X_{u-},\gamma_0)z\}\, \mu(du,dz), \end{aligned} \tag{4.8}$$

which is valid for any $\mathcal{C}^\beta$-function [12] ψ; see [21, Theorems 3.2.1b) and 3.2.2a)] for details. Let $\mathcal{A}$ denote the (formal) infinitesimal generator of X given by

$$\mathcal{A}\psi(x) = \partial\psi(x)a(x,\alpha_0) + \int \{\psi(x + c(x,\gamma_0)z) - \psi(x) - \partial\psi(x)c(x,\gamma_0)z\}\, \nu(dz),$$

where the second term in the right-hand side is well-defined. It follows from (4.8) that

$$\psi(X_t) = \psi(X_s) + \int_s^t \mathcal{A}\psi(X_u)du + \int_s^t \int \{\psi(X_{u-} + c(X_{u-},\gamma_0)z) - \psi(X_{u-})\}\, \tilde{\mu}(du,dz). \tag{4.9}$$

Obviously, we have $|\mathcal{A}\psi(x)| \lesssim 1 + |x|^C$ for ψ such that the derivatives $\partial^k\psi$ for $k \in \{0,1,2\}$ exist and are bounded by a polynomial.

Lemma 4.3. *Suppose that:*

(i) $\pi \in \mathcal{C}^1(\mathbb{R} \times \overline{\Theta})$ *and* $\sup_\theta\{|\pi(x,\theta)| + |\partial_\theta\pi(x,\theta)|\} \lesssim 1 + |x|^C$ *for every* $x \in \mathbb{R}$*;*
(ii) $\eta \in \mathcal{C}^1(\mathbb{R})$ *with bounded first derivative.*

Then we have $U'_{2,n}(\theta) = O^*_{L^q}(nh^{2-1/\beta})$ *and* $U''_{2,n}(\theta) = O^*_{L^q}(nh^{2-1/\beta})$ *for every* $q > 0$*. In particular, if* $\beta > 2/3$ *we have both* $U'_{2,n}(\theta) = o^*_p(\sqrt{n})$ *and* $U''_{2,n}(\theta) = o^*_p(\sqrt{n})$*.*

Proof. We begin with $U'_{2,n}(\theta)$. Applying (4.9) with $\psi(x) = a(x,\alpha_0)$ and then taking the conditional expectation, we get

$$\begin{aligned} \left| \mathbb{E}^{j-1}\left(\int_j a^{\Delta}_{j-1}(s)ds \right) \right| &= \left| \int_j \mathbb{E}^{j-1}\{a^{\Delta}_{j-1}(s)\}ds \right| \\ &\le \int_j \int_{t_{j-1}}^s \mathbb{E}^{j-1}\{|\mathcal{A}a(X_u,\alpha_0)|\}\, duds \\ &\lesssim \int_j \int_{t_{j-1}}^s \{1 + \mathbb{E}^{j-1}(|X_u|^C)\}\, duds \\ &\lesssim \int_j \int_{t_{j-1}}^s (1 + |X_{t_{j-1}}|^C)duds = O^*_{L^q}(h^2). \end{aligned} \tag{4.10}$$

Write $m_j(\theta;\eta) = \overline{r}_j(\theta;\eta) - \mathbb{E}^{j-1}\{\overline{r}_j(\theta;\eta)\}$ and $\tilde{a}^{\Delta}_{j-1}(s) = a^{\Delta}_{j-1}(s) - \mathbb{E}^{j-1}\{a^{\Delta}_{j-1}(s)\}$. Then, using (4.10) and noting that $\overline{r}_j(\theta;\eta)$ is essentially bounded, we get

$$\begin{aligned} U'_{2,n}(\theta) &= h^{-1/\beta} \sum_{j=1}^{n} \pi'_{j-1}(\theta) \mathbb{E}^{j-1}\left(m_j(\theta;\eta) \int_j \tilde{a}^{\Delta}_{j-1}(s)ds \right) + O^*_{L^q}(nh^{2-1/\beta}) \\ &= h^{-1/\beta} \sum_{j=1}^{n} \pi'_{j-1}(\theta) \int_j \mathbb{E}^{j-1}\left(m_j(\theta;\eta)\tilde{a}^{\Delta}_{j-1}(s) \right) ds + O^*_{L^q}(nh^{2-1/\beta}). \end{aligned} \tag{4.11}$$

In view of the expression (4.11) and Jensen's inequality, the claim $U'_{2,n}(\theta) = O^*_{L^q}(nh^{2-1/\beta})$ follows from

$$\frac{1}{h}\mathbb{E}^{j-1}\left\{ m_j(\theta;\eta)\tilde{a}^{\Delta}_{j-1}(s) \right\} = O^*_{L^q}(1), \qquad s \in [t_{j-1},t_j]. \tag{4.12}$$

By (4.9) we may write

$$\tilde{a}^{\Delta}_{j-1}(s) = \int_{t_{j-1}}^s f(X_{t_{j-1}},X_u)du + \int_{t_{j-1}}^s \int g(X_{u-},z)\tilde{\mu}(du,dz), \tag{4.13}$$

[12] In case of $\beta \in (1,2)$, this means that ψ is $\mathcal{C}^1$ and the derivative $\partial\psi$ is locally Hölder continuous with index $\beta - [\beta]$.

where $\mathbb{E}^{j-1}\{f(X_{t_{j-1}},X_u)\}=0$ with $f(x,x')$ being at most polynomial-growth in (x,x'), and where $g(x,z):=a(x+zc(x,\gamma_0),\alpha_0)-a(x,\alpha_0)$; by the regularity conditions on $a(x,\alpha_0)$, we have $|g(x,z)|\lesssim |z|(1+|z|^C)(1+|x|^C)$. Hence, (4.12) follows upon showing that

$$\frac{1}{h}\mathbb{E}^{j-1}\bigg(m_j(\theta;\eta)\int_{t_{j-1}}^s\int g(X_{u-},z)\tilde{\mu}(du,dz)\bigg)=O^*_{L^q}(1). \tag{4.14}$$

Let $H_{j,t}(\theta;\eta):=\mathbb{E}\left\{m_j(\theta;\eta)|\,\mathcal{F}_t\right\}$ for $t\in[t_{j-1},t_j]$; then we have $H_{j,t_j}(\theta;\eta)=m_j(\theta;\eta)$. Since $\mathcal{F}_t=\sigma(X_0)\vee\sigma(J_s;s\le t)$, invoking [41, Theorem I.32] we see that $\{H_{j,t}(\theta;\eta),\mathcal{F}_{t_{j-1}}\vee\sigma(J_t);\,t\in[t_{j-1},t_j]\}$ is a (essentially bounded) martingale. According to the martingale representation theorem [22, Theorem III.4.34], the process $H_{j,t}(\theta)$ can be represented as a stochastic integral of the from

$$H_{j,t}(\theta;\eta)=\int_{t_{j-1}}^t\int \xi_j(s,z;\theta)\tilde{\mu}(ds,dz),\qquad t\in[t_{j-1},t_j], \tag{4.15}$$

with a bounded predictable process $s\mapsto\xi_j(s,z;\theta)$ such that $\int_j\int|\xi_j(s,z;\theta)|^2\nu(dz)ds\lesssim 1$. Then, substituting the last expressions into (4.14), using the martingale property of the stochastic integrals (take the conditioning with respect to $\mathcal{F}_s$ inside the sign "$\mathbb{E}^{j-1}$"), and then applying the integration-by-parts formula, we see that the left-hand side of (4.14) equals

$$\frac{1}{h}\mathbb{E}^{j-1}\bigg(\int_{t_{j-1}}^s\int\xi_j(u,z;\theta)g(X_{u-},z)\nu(dz)du\bigg).$$

By Jensen and Cauchy-Schwarz inequalities and the upper bounded of $|g(x,z)|$ mentioned before, we can bound the qth-absolute moment of the last quantity as follows:

$$\begin{aligned}
&\mathbb{E}\bigg\{\sup_\theta\bigg|\frac{1}{h}\mathbb{E}^{j-1}\bigg(\int_{t_{j-1}}^s\int\xi_j(u,z;\theta)g(X_{u-},z)\nu(dz)du\bigg)\bigg|^q\bigg\}\\
&\lesssim\frac{1}{h}\int_{t_{j-1}}^s\mathbb{E}\bigg\{\sup_\theta\mathbb{E}^{j-1}\bigg(\bigg|\int\xi_j(u,z;\theta)g(X_{u-},z)\nu(dz)\bigg|\bigg)^q\bigg\}du\\
&\lesssim\frac{1}{h}\int_{t_{j-1}}^s\mathbb{E}\left(1+|X_u|^C\right)du\lesssim 1.
\end{aligned}$$

Thus we obtain (4.14), concluding that $U'_{2,n}(\theta)=O^*_{L^q}(nh^{2-1/\beta})$.

Next we consider $U''_{2,n}(\theta)$. Using the martingale representation (4.11) again,

$$\begin{aligned}
U''_{2,n}(\theta)&=h^{-1/\beta}\sum_{j=1}^n\pi'_{j-1}(\theta)\mathbb{E}^{j-1}\bigg(m_j(\theta;\eta)\int_j c^\Delta_{j-1}(s-)dJ_s\bigg)\\
&\quad+h^{-1/\beta}\sum_{j=1}^n\pi'_{j-1}(\theta)\mathbb{E}^{j-1}\{\overline{r}_j(\theta;\eta)\}\mathbb{E}^{j-1}\bigg(\int_j c^\Delta_{j-1}(s-)dJ_s\bigg)\\
&=h^{-1/\beta}\sum_{j=1}^n\pi'_{j-1}(\theta)\mathbb{E}^{j-1}\bigg(\Delta_jH_j(\theta;\eta)\int_j c^\Delta_{j-1}(s-)dJ_s\bigg)\\
&=h^{-1/\beta}\sum_{j=1}^n\pi'_{j-1}(\theta)\mathbb{E}^{j-1}\bigg(\int_j\int\xi_j(s,z;\theta)\tilde{\mu}(ds,dz)\int_j\int c^\Delta_{j-1}(s-)z\tilde{\mu}(ds,dz)\bigg)\\
&=h^{-1/\beta}\sum_{j=1}^n\pi'_{j-1}(\theta)\mathbb{E}^{j-1}\bigg(\int_j\int\xi_j(s,z;\theta)zc^\Delta_{j-1}(s)\nu(dz)ds\bigg).
\end{aligned}\tag{4.16}$$

As in the case of a^Δ_{j-1} we have $|\mathbb{E}^{j-1}\{c^\Delta_{j-1}(s)\}|\le\int_{t_{j-1}}^s\mathbb{E}^{j-1}\{|\mathcal{A}c(X_u,\gamma_0)|\}du=O^*_{L^q}(h)$.

The process $\tilde{\Xi}_{j,s}(\theta):=\int\xi_j(s,z;\theta)z\nu(dz)-\mathbb{E}^{j-1}\{\int\xi_j(s,z;\theta)z\nu(dz)\}$ for $s\in[t_{j-1},t_j]$ satisfies that $|\tilde{\Xi}_{j,s}(\theta)|^2\le\int|z|^2\nu(dz)\int|\xi_j(s,z;\theta)|^2\nu(dz)\lesssim 1$. By (4.16), we then have

$$U''_{2,n}(\theta)=h^{-1/\beta}\sum_{j=1}^n\pi'_{j-1}(\theta)\int_j\mathbb{E}^{j-1}\left(\tilde{\Xi}_{j,s}(\theta)\tilde{c}^\Delta_{j-1}(s)\right)ds+O^*_{L^q}(nh^{2-1/\beta}), \tag{4.17}$$

with $\tilde{c}^\Delta_{j-1}(s):=c^\Delta_{j-1}(s)-\mathbb{E}^{j-1}\{c^\Delta_{j-1}(s)\}$; trivially, $\tilde{c}^\Delta_{j-1}(s)$ admits a similar representation to (4.13). Now we make a further application of the representation theorem for (4.17). For each j, the processes $M'^j_u(\theta):=\mathbb{E}^{j-1}\{\Xi^j_s(\theta)|\mathcal{F}_u\}$ and $M''^j_u:=\mathbb{E}^{j-1}\{\tilde{c}^\Delta_{j-1}(s)|\mathcal{F}_u\}$ for $u\in[t_{j-1},s]$ are martingales with respect to the filtration $\{\mathcal{F}_{t_{j-1}}\vee\sigma(J_u)\,:\,u\in[t_{j-1},s]\}$. Hence there correspond predictable processes $m_u'^j(z;\theta)$ and $m_u''^j(z)$ bounded for each z, such that $M_s'^j(\theta) = \int_{t_{j-1}}^s \int m_u'^j(z;\theta)\tilde{\mu}(du,dz)$ and $M_s''^j = \int_{t_{j-1}}^s \int m_u''^j(z)\tilde{\mu}(du,dz)$. Thus, applying the integration by parts formula as before we can continue (4.17) as follows:

$$U_{2,n}''(\theta) = h^{-1/\beta}\sum_{j=1}^n \pi_{j-1}'(\theta)\int_j\int_{t_{j-1}}^s \mathbb{E}^{j-1}\Big(\int m_u'^j(z;\theta)m_u''^j(z)\nu(dz)\Big)duds + O_{L^q}^*(nh^{2-1/\beta}).$$

We see that the first term in the right-hand side is $O_{L^q}^*(nh^{2-1/\beta})$, hence so is $U_{2,n}''(\theta)$.

Since $\sqrt{n}h^{2-1/\beta} \lesssim h^{3/2-1/\beta}$, the last part of the lemma is trivial. The proof is complete. □

4.2.3. *Uniform law of large numbers.* Building on the above arguments, we now look at uniform asymptotic behavior of $U_n(\theta)$. To this end we first note the following basic law of large numbers:

Lemma 4.4. *For any measurable function $f : \mathbb{R}\times\overline{\Theta} \to \mathbb{R}$ such that $\sup_\theta (|f(x,\theta)| + |\partial_\theta f(x,\theta)|) \lesssim 1+|x|^C$, we have*

$$\sup_\theta \sup_{t\le T}\left|\frac{1}{n}\sum_{j=1}^{[nt/T]} f(X_{t_{j-1}},\theta) - \frac{1}{T}\int_0^t f(X_s,\theta)ds\right| \xrightarrow{p} 0.$$

Proof. The target quantity can be bounded by

$$\begin{aligned}&\sup_{t\le T}\frac{1}{n}\sum_{j=1}^{[nt/T]}\frac{1}{h}\int_j \sup_\theta |f(X_s,\theta) - f_{j-1}(\theta)|ds + \frac{h}{T}\sup_\theta\sup_{t\le T}|f(X_t,\theta)| \\ &\lesssim \frac{1}{n}\sum_{j=1}^n \frac{1}{h}\int_j (1+|X_{t_{j-1}}|^C + |X_s|^C)|X_s - X_{t_{j-1}}|ds + \frac{h}{T}\Big(1+\sup_{t\le T}|X_t|^C\Big).\end{aligned}$$

By (4.2) the expectation of the right-most side is $o(1)$, hence the claim follows. □

Proposition 4.5. *Assume that the conditions (i) and (ii) in Lemma 4.2 hold.*

(1) *For $\beta = 1$, we have*

$$\frac{1}{n}U_n(\theta) = \frac{1}{T}\int_0^T \pi(X_t,\theta)\int \eta\Big(\frac{c(X_t,\gamma_0)}{c(X_t,\gamma)}z + \mathfrak{b}(X_t,\theta)\Big)\phi_1(z)dzdt + o_p^*(1).$$

(2) *For $\beta\in(1,2)$, we have*

$$\frac{1}{n}U_n(\theta) = \frac{1}{T}\int_0^T \pi(X_t,\theta)\eta\Big(\frac{c(X_t,\gamma_0)}{c(X_t,\gamma)}z\Big)\phi_\beta(dz)dzdt + o_p^*(1).$$

(3) *For $\beta\in(1,2)$, if further η is odd, then we have*

$$\frac{1}{nh^{1-1/\beta}}U_n(\theta) = O_p^*(1).$$

Proof. Let

$$\overline{U}_{2,n}^0(\theta) := \frac{1}{nh^{1-1/\beta}}U_{2,n}^0(\theta) = \frac{1}{n}\sum_{j=1}^n \pi_{j-1}(\theta)\frac{1}{h^{1-1/\beta}}\mathbb{E}^{j-1}\left\{\eta\left(\delta_j'(\gamma) + h^{1-1/\beta}\mathfrak{b}_{j-1}(\theta)\right)\right\}.$$

(1) Write $\overline{U}_{2,n}^0(\theta)$ as the sum of

$$\begin{aligned}\frac{1}{n}\sum_{j=1}^n f_{j-1}^1(\theta) &:= \frac{1}{n}\sum_{j=1}^n \pi_{j-1}(\theta)\int\eta\Big(\frac{c_{j-1}(\gamma_0)}{c_{j-1}(\gamma)}z + \mathfrak{b}_{j-1}(\theta)\Big)\phi_1(z)dz, \\ \frac{1}{n}\sum_{j=1}^n f_{j-1}^2(\theta) &:= \frac{1}{n}\sum_{j=1}^n \pi_{j-1}(\theta)\int\eta\Big(\frac{c_{j-1}(\gamma_0)}{c_{j-1}(\gamma)}z + \mathfrak{b}_{j-1}(\theta)\Big)\{f_h(z) - \phi_1(z)\}dz.\end{aligned}$$

Pick any $\kappa\in(0,1)$ small enough to make (2.6) valid, and observe that $|\eta(y)|\lesssim 1+|y|^\kappa$. Then,

$$\sup_\theta\left|\eta\Big(\frac{c_{j-1}(\gamma_0)}{c_{j-1}(\gamma)}z + h^{1-1/\beta}\mathfrak{b}_{j-1}(\theta)\Big)\right| \lesssim (1+|X_{t_{j-1}}|^C)(1+|z|^\kappa). \tag{4.18}$$

Hence we have the bounds: $|f_{j-1}^1(\theta)| \lesssim (1+|X_{t_{j-1}}|^C)\int(1+|z|^\kappa)\phi_1(y)dy \lesssim 1+|X_{t_{j-1}}|^C$ and $|f_{j-1}^2(\theta)| \lesssim (1+|X_{t_{j-1}}|^C)\int(1+|z|^\kappa)|f_h(y)-\phi_1(y)|dy = o_p^*(1)$; in particular,

$$\frac{1}{n}\sum_{j=1}^n f_{j-1}^2(\theta) = o_p^*(1).$$

Likewise, noting that the same upper bound as in (4.18) is available for the function $y \mapsto y\partial\eta(y)$, we see that $\partial_\theta f_{j-1}^1(\theta)$ can by bounded by a sum of constant multiples of the terms $1+|X_{t_{j-1}}|^C$ (coming from the term involving $\partial_\theta \pi_{j-1}(\theta)$) and

$$\begin{aligned}
&\left|\pi_{j-1}(\theta)\int \partial\eta\biggl(\frac{c_{j-1}(\gamma_0)}{c_{j-1}(\gamma)}z+\mathfrak{b}_{j-1}(\theta)\biggr)\biggl\{-\frac{\partial_\gamma c_{j-1}(\gamma)}{c_{j-1}(\gamma)}\left(\frac{c_{j-1}(\gamma_0)}{c_{j-1}(\gamma)}z\right)+\partial_\theta\mathfrak{b}_{j-1}(\theta)\biggr\}\phi_1(z)dz\right| \\
&\lesssim (1+|X_{t_{j-1}}|^C)\int\biggl\{\biggl|\biggl(\frac{c_{j-1}(\gamma_0)}{c_{j-1}(\gamma)}z+\mathfrak{b}_{j-1}(\theta)\biggr)\partial\eta\biggl(\frac{c_{j-1}(\gamma_0)}{c_{j-1}(\gamma)}z+\mathfrak{b}_{j-1}(\theta)\biggr)\biggr|+\|\partial\eta\|_\infty\biggr\}\phi_1(z)dz \\
&\lesssim 1+|X_{t_{j-1}}|^C.
\end{aligned}$$

The claim now follows on applying Lemmas 4.2, 4.3, and 4.4.

(2) For $\beta \in (1,2)$, the Taylor expansion gives

$$\begin{aligned}
&h^{1-1/\beta}\overline{U}_{2,n}^0(\theta) \\
&= \frac{1}{n}\sum_{j=1}^n \pi_{j-1}(\theta)\mathbb{E}^{j-1}\left\{\eta\left(\delta_j'(\gamma)+h^{1-1/\beta}\mathfrak{b}_{j-1}(\theta)\right)\right\} \\
&= \frac{1}{n}\sum_{j=1}^n \pi_{j-1}(\theta)\int\eta\biggl(\frac{c_{j-1}(\gamma_0)}{c_{j-1}(\gamma)}z\biggr)f_h(z)dz \\
&\quad + h^{1-1/\beta}\frac{1}{n}\sum_{j=1}^n \pi_{j-1}(\theta)\mathfrak{b}_{j-1}(\theta)\int_0^1 \mathbb{E}^{j-1}\left\{\partial\eta\left(\delta_j'(\gamma)+sh^{1-1/\beta}\mathfrak{b}_{j-1}(\theta)\right)\right\}ds.
\end{aligned} \tag{4.19}$$

Following a similar line to the case of $\beta=1$, we see that the first term in the rightmost side of (4.19) equals $\frac{1}{T}\int_0^T \pi(X_t,\theta)\eta(\frac{c(X_t,\gamma_0)}{c(X_t,\gamma)}z)\phi_\beta(dz)dzdt + o_p^*(1)$. By the boundedness of $\partial\eta$ and the estimate $|\pi_{j-1}(\theta)\mathfrak{b}_{j-1}(\theta)| \lesssim 1+|X_{t_{j-1}}|^C$, the last term on the right-hand side is $O_p^*(h^{1-1/\beta}) = o_p^*(1)$. Hence the claim follows from Lemmas 4.2 and 4.3.

(3) Recalling (4.7), under the conditions in Lemma 4.3 we have

$$\frac{1}{nh^{1-1/\beta}}U_{2,n}(\theta) = \overline{U}_{2,n}^0(\theta) + O_p^*(h).$$

If further η is odd in addition to $\beta \in (1,2)$, then by the symmetry of $\mathcal{L}(h^{-1/\beta}\Delta_j J)$ we have $\mathbb{E}^{j-1}\{\eta(\delta_j'(\gamma))\} = \int\eta(\frac{c_{j-1}(\gamma_0)}{c_{j-1}(\gamma)}z)f_h(z)dz = 0$ a.s. and the identity (4.19) becomes $\overline{U}_{2,n}^0(\theta) = O_p^*(1)$. □

The next corollary, which will be used in the proof of the consistency of $\hat{\alpha}_n$, is obvious from the first identity in (4.19) and the fact $\mathfrak{b}_{j-1}(\alpha_0,\gamma) \equiv 0$ (also recall (4.6)).

Corollary 4.6. *Assume that the conditions (i) and (ii) in Lemma 4.2 hold, and let $\beta \in (1,2)$ and η be odd. Then we have*

$$\frac{1}{nh^{1-1/\beta}}\sum_{j=1}^n \pi_{j-1}(\theta)\mathbb{E}^{j-1}\left\{\eta(\epsilon_j(\alpha_0,\gamma))\right\} = o_p^*(1),$$

hence we also have

$$\frac{1}{nh^{1-1/\beta}}\sum_{j=1}^n \pi_{j-1}(\theta)\eta(\epsilon_j(\alpha_0,\gamma)) = o_p^*(1).$$

4.3. **Proof of the consistency.** The statistical random field associated with the stable quasi-likelihood has a multi-scaling structure. We first prove the following lemma[13].

Lemma 4.7 (Consistency under possible multi-scaling). *Let $K_1 \subset \mathbb{R}^{p_1}$ and $K_2 \subset \mathbb{R}^{p_2}$ be compact sets, and let $H_n : K_1 \times K_2 \to \mathbb{R}$ be a random function of the form*

$$H_n(u_1, u_2) = k_{1,n} H_{1,n}(u_1) + k_{2,n} H_{2,n}(u_1, u_2)$$

for some positive non-random sequences $(k_{1,n})$ and $(k_{2,n})$ and some continuous random functions $H_{1,n} : K_1 \to \mathbb{R}$ and $H_{2,n} : K_1 \times K_2 \to \mathbb{R}$. Let $(u_{1,0}, u_{2,0}) \in K_1^\circ \times K_2^\circ$ be a non-random vector. Assume the following conditions:

- $k_{2,n} = o(k_{1,n})$;
- $\sup_{u_1} |H_{1,n}(u_1) - H_{1,0}(u_1)| \xrightarrow{p} 0$ *and* $\sup_{(u_1,u_2)} |H_{2,n}(u_1, u_2) - H_{2,0}(u_1, u_2)| \xrightarrow{p} 0$ *for some continuous random functions $H_{1,0}$ and $H_{2,0}$;*
- $\{u_{1,0}\} = \operatorname{argmax} H_{1,0}$ *and* $\{u_{2,0}\} = \operatorname{argmax} H_{2,0}(u_{1,0}, \cdot)$ *a.s.*

Then, for any $(\hat{u}_{1,n}, \hat{u}_{2,n}) \in K_1 \times K_2$ such that $H_n(\hat{u}_{1,n}, \hat{u}_{2,n}) \geq \sup H_n - o_p(k_{2,n})$ a.s. we have $(\hat{u}_{1,n}, \hat{u}_{2,n}) \xrightarrow{p} (u_{1,0}, u_{2,0})$.

Proof. The claim is a special case of [42, Theorem 1]; in our setting we do not need the a.s. representation theorem. For convenience, we sketch the proof.

The assumption implies that $(H_{1,n}, H_{2,n}) \xrightarrow{\mathcal{L}} (H_{1,0}, H_{2,0})$ in $\mathcal{C}(K_1 \times K_2)$. Let $H'_n(u_1) := k_{1,n}^{-1} H_n(u_1, \hat{u}_{2,n}) = H_{1,n}(u_1) + k_{2,n} k_{1,n}^{-1} H_{2,n}(u_1, \hat{u}_{2,n})$. The second term in the rightmost side is $o_p(1)$ uniformly in $u_1 \in K_1$, so that $H'_n(\cdot) \xrightarrow{\mathcal{L}} H_{1,0}(\cdot)$ in $\mathcal{C}(K_1)$ with the limit a.s. uniquely maximized at $\hat{u}_{1,0}$. Since $H'_n(\hat{u}_{1,n}) \geq \sup_{u_1} k_{1,n}^{-1} H_n(u_1, \hat{u}_{2,n}) - o_p(k_{2,n} k_{1,n}^{-1}) = \sup H'_n - o_p(1)$, the argmax theorem gives $\hat{u}_{1,n} \xrightarrow{p} u_{1,0}$. We can follow a similar way to deduce $\hat{u}_{2,n} \xrightarrow{p} u_{2,0}$ along with replacing H'_n by $H''_n(u_2) := k_{2,n}^{-1}\{H_n(\hat{u}_{1,n}, u_2) - H_n(\hat{u}_{1,n}, u_{2,0})\} = H_{2,n}(\hat{u}_{1,n}, u_2) - H_{2,n}(\hat{u}_{1,n}, u_{2,0})$; H''_n has the continuous limit process $H_{2,0}(u_{1,0}, \cdot) - H_{2,0}(u_{1,0}, u_{2,0})$ in $\mathcal{C}(K_2)$, which is a.s. uniquely maximized at $\hat{u}_{2,0}$. □

Before proceeding, we make a couple of remarks concerning Assumption 2.9. For $\beta > 1$, we introduce the random functions $\mathbb{Y}_{\beta,1}(\cdot) = \mathbb{Y}_{\beta,1}(\cdot; \gamma_0) : \Theta_\gamma \to \mathbb{R}$ and $\mathbb{Y}_{\beta,2}(\cdot) = \mathbb{Y}_{\beta,2}(\cdot; \theta_0) : \Theta \to \mathbb{R}$ be given by

$$\mathbb{Y}_{\beta,1}(\gamma) = \frac{1}{T} \int_0^T \int \left[\log \left\{ \frac{c(X_t, \gamma_0)}{c(X_t, \gamma)} \phi_\beta \left(\frac{c(X_t, \gamma_0)}{c(X_t, \gamma)} z \right) \right\} - \log \phi_\beta(z) \right] \phi_\beta(z) dz dt, \tag{4.20}$$

$$\mathbb{Y}_{\beta,2}(\theta) = \frac{1}{2T} \int_0^T \mathfrak{b}^2(X_t, \theta) \int \partial g_\beta \left(\frac{c(X_t, \gamma_0)}{c(X_t, \gamma)} z \right) \phi_\beta(z) dz dt. \tag{4.21}$$

We also define $\mathbb{Y}_1(\cdot) = \mathbb{Y}_1(\cdot; \theta_0) : \Theta \to \mathbb{R}$ by

$$\mathbb{Y}_1(\theta) = \frac{1}{T} \int_0^T \int \left[\log \left\{ \frac{c(X_t, \gamma_0)}{c(X_t, \gamma)} \phi_1 \left(\frac{c(X_t, \gamma_0)}{c(X_t, \gamma)} z + \mathfrak{b}(X_t, \theta) \right) \right\} - \log \phi_1(z) \right] \phi_1(z) dz dt.$$

These three functions are a.s. continuous in θ. Assumptions 2.8 and 2.9 together with Jensen's inequality (applied ω-wise) imply that both $\mathbb{Y}_{\beta,1}$ and $\mathbb{Y}_1$ are non-negative functions with $\{\gamma_0\} = \operatorname{argmax} \mathbb{Y}_{\beta,1}$ and $\{\theta_0\} = \operatorname{argmax} \mathbb{Y}_1$ a.s. Moreover,

$$\mathbb{Y}_{\beta,2}(\alpha, \gamma_0) = -\frac{1}{2} \int \frac{\{\partial \phi_\beta(z)\}^2}{\phi_\beta(z)} dz \cdot \frac{1}{T} \int_0^T \mathfrak{b}^2(X_t, \theta) dt \leq 0,$$

the maximum 0 being attained if and only if $\alpha = \alpha_0$.

4.3.1. *Case of $\beta = 1$.* Let

$$\mathbb{Y}_{1,n}(\theta) := \frac{1}{n} (\mathbb{H}_n(\theta) - \mathbb{H}_n(\theta_0)) = \frac{1}{n} \sum_{j=1}^n \left(\log \frac{c_{j-1}(\gamma_0)}{c_{j-1}(\gamma)} + \log \phi_1(\epsilon_j(\theta)) - \log \phi_1(\epsilon_j) \right). \tag{4.22}$$

Under Assumption 2.9 Jensen's inequality implies that $\{\theta_0\} = \mathbb{Y}_1$ a.s. Hence, by means of Lemma 4.7 the consistency of $\hat{\theta}_n$ ($\in \operatorname{argmax} \mathbb{Y}_{1,n}$) follows on showing that $\sup_\theta |\mathbb{Y}_{1,n}(\theta) - \mathbb{Y}_1(\theta)| \xrightarrow{p} 0$. This readily follows from Lemma 4.4 and Proposition 4.5(1) with $\pi(x, \theta) \equiv 1$ and $\eta = \log \phi_1$.

[13] A formal extension to multi-scaling of more than two factors is trivial.

4.3.2. *Case of* $\beta \in (1,2)$. Observe that $\mathbb{H}_n(\theta) - \mathbb{H}_n(\theta_0) = k_n \mathbb{Y}_{\beta,1,n}(\gamma) + l_n \mathbb{Y}_{\beta,2,n}(\alpha,\gamma)$, where $k_n := n$, $k_n := nh^{2(1-1/\beta)}$, and

$$\mathbb{Y}_{\beta,1,n}(\gamma) := \frac{1}{n}\{\mathbb{H}_n(\alpha_0,\gamma) - \mathbb{H}_n(\alpha_0,\gamma_0)\},$$
$$\mathbb{Y}_{\beta,2,n}(\alpha,\gamma) := \frac{1}{nh^{2(1-1/\beta)}}\{\mathbb{H}_n(\alpha,\gamma) - \mathbb{H}_n(\alpha_0,\gamma)\}.$$

Recall the definitions (4.20) and (4.21) of $\mathbb{Y}_{\beta,1}$ and $\mathbb{Y}_{\beta,2}$, respectively. By applying Lemma 4.7 under Assumption 2.9, the consistency of the SQMLE follows from the uniform convergences:

$$\sup_{\gamma} |\mathbb{Y}_{\beta,1,n}(\gamma) - \mathbb{Y}_{\beta,1}(\gamma;\gamma_0)| \xrightarrow{p} 0, \tag{4.23}$$

$$\sup_{\theta} |\mathbb{Y}_{\beta,2,n}(\theta) - \mathbb{Y}_{\beta,2}(\theta;\theta_0)| \xrightarrow{p} 0. \tag{4.24}$$

The proof of (4.23) is much the same as in the case of $\beta = 1$, hence we only prove (4.24).

Recall the notation $g_\beta(y) = \frac{\partial\phi_\beta(y)}{\phi_\beta(y)}$, which is bounded smooth and satisfies that

$$\sup_{y} |y|^{k+1}\left|\partial^k g_\beta(y)\right| < \infty, \qquad k \in \mathbb{Z}_+. \tag{4.25}$$

Since we are now having the vanishing factor "$h^{2(1-1/\beta)}$" in the denominator, a slightly different care than the case of (4.23) is necessary.

Observe that

$$\begin{aligned}
\mathbb{Y}_{\beta,2,n}(\theta) &= \frac{1}{nh^{2(1-1/\beta)}}\sum_{j=1}^{n}\bigg(\log\phi_\beta(\epsilon_j(\theta)) - \log\phi_\beta(\epsilon_j(\alpha_0,\gamma))\bigg) \\
&= \frac{1}{nh^{1-1/\beta}}\sum_{j=1}^{n}\mathfrak{b}_{j-1}(\theta) g_\beta(\epsilon_j(\alpha_0,\gamma)) + \frac{1}{2n}\sum_{j=1}^{n}\mathfrak{b}_{j-1}^2(\theta)\partial g_\beta(\epsilon_j(\alpha_0,\gamma)) \\
&\quad + \frac{1}{2n}\sum_{j=1}^{n}\mathfrak{b}_{j-1}^2(\theta)\left\{\partial g_\beta(\tilde{\epsilon}_j(\theta)) - \partial g_\beta(\epsilon_j(\alpha_0,\gamma))\right\} \\
&=: \mathbb{Y}'_{\beta,2,n}(\theta) + \mathbb{Y}^0_{\beta,2,n}(\theta) + \mathbb{Y}''_{\beta,2,n}(\theta),
\end{aligned} \tag{4.26}$$

where $\tilde{\epsilon}_j(\theta)$ is a random point on the segment connecting $\epsilon_j(\theta)$ and $\epsilon_j(\alpha_0,\gamma)$. We have $\mathbb{Y}'_{\beta,2,n}(\theta) = o_p^*(1)$ by Corollary 4.6. Since $|\tilde{\epsilon}_j(\theta) - \epsilon_j(\alpha_0,\gamma)| \le |\epsilon_j(\theta) - \epsilon_j(\alpha_0,\gamma)| \lesssim (1+|X_{t_{j-1}}|^C)h^{1-1/\beta} = O_p^*(h^{1-1/\beta}) = o_p^*(1)$, we also have $\mathbb{Y}''_{\beta,2,n}(\theta) = o_p^*(1)$. To deduce that $\mathbb{Y}^0_{\beta,2,n}(\theta) = \frac{1}{T}\int_0^T \mathfrak{b}^2(X_t,\theta)\int \partial g_\beta\left(\frac{c(X_t,\gamma_0)}{c(X_t,\gamma)}z\right)\phi_\beta(z)dzdt + o_p^*(1)$, we can apply Proposition 4.5(2) with $\pi(x,\theta) = \frac{1}{2}\mathfrak{b}^2(x,\theta)$ and $\eta = \partial g_\beta$ under the quite trivial modification that we have "$\epsilon_j(\alpha_0,\gamma)$" instead of "$\epsilon_j(\theta)$" inside the η.

4.4. **Proof of the asymptotic mixed normality.** Having verified the consistency of the SQMLE $\hat{\theta}_n$, we turn to the proof of the asymptotic mixed normality. For convenience we introduce the rate matrix:

$$D_n = (D_{n,k})_{k=1}^p = \mathrm{diag}\left(\sqrt{n}h^{1-1/\beta}I_{p_\alpha},\, \sqrt{n}I_{p_\gamma}\right) \in \mathbb{R}^p \otimes \mathbb{R}^p.$$

We also write $\hat{u}_n = \left(\sqrt{n}h^{1-1/\beta}(\hat{\alpha}_n - \alpha_0),\, \sqrt{n}(\hat{\gamma}_n - \gamma_0)\right)$. The consistency allows us to focus on the event $\{\partial_\theta \mathbb{H}_n(\hat{\theta}_n) = 0\}$, on which the Taylor formula gives

$$\bigg(-D_n^{-1}\partial_\theta^2\mathbb{H}_n(\theta_0)D_n^{-1} + \hat{r}_n\bigg)[\hat{u}_n] = D_n^{-1}\partial_\theta\mathbb{H}_n(\theta_0), \tag{4.27}$$

where $\hat{r}_n = \{\hat{r}_n^{kl}\}_{k,l}$ is a bilinear form such that

$$|\hat{r}_n| \le \sum_{k,l,m=1}^{p}\bigg(D_{n,k}^{-1}D_{n,l}^{-1}\sup_{\theta=(\theta_i)_{i=1}^p}|\partial_{\theta_k}\partial_{\theta_l}\partial_{\theta_m}\mathbb{H}_n(\theta)|\bigg)|\hat{\theta}_{m,n} - \theta_{0,m}|. \tag{4.28}$$

If we have

$$(\Delta_{n,T},\, \Gamma_{n,T}) \xrightarrow{\mathcal{L}} (\Delta_T,\, \Gamma_T(\theta_0;\beta)) \quad \text{where } \Delta_T \sim MN_p\left(\mu_T(\theta_0;\beta),\, \Gamma_T(\theta_0;\beta)\right), \tag{4.29}$$

$$r^*_{n,T} := \max_{1\le k,l,m\le p} D_{n,k}^{-1}D_{n,l}^{-1}\sup_{\theta=(\theta_i)_{i=1}^p}|\partial_{\theta_k}\partial_{\theta_k}\partial_{\theta_m}\mathbb{H}_n(\theta)| = O_p(1), \tag{4.30}$$

then $\hat{r}_n = o_p(1)$ and hence

$$\begin{aligned}\hat{u}_n &= \left(\Gamma_T(\theta_0;\beta) + o_p(1)\right)^{-1} \Delta_{n,T} \\ &= \Gamma_T^{-1}(\theta_0;\beta)\Delta_{n,T} + o_p(1) \\ &\xrightarrow{\mathcal{L}} \Gamma_T^{-1}(\theta_0;\beta)\Delta_T \sim MN_p\left(\Gamma_T^{-1}(\theta_0;\beta)\mu_T(\theta_0;\beta),\, \Gamma_T^{-1}(\theta_0;\beta)\right),\end{aligned}$$

concluding the proof. Since $\Gamma_T(\theta_0;\beta)$ is random, the joint weak convergence (4.29) is far from being obvious and we cannot deduce it from a direct application of the usual martingale central limit theorem for triangular arrays of random variables (e.g. [10]).

The stable convergence is the right mode of convergence to deduce (4.29). In order to complete the proof it suffices to prove (4.30) and the following two convergences:

$$\Delta_{n,T} := D_n^{-1}\partial_\theta \mathbb{H}_n(\theta_0) \xrightarrow{\mathcal{L}_s} \Delta_T \sim MN_p(\mu_T(\theta_0;\beta), \Sigma_T(\theta_0;\beta)); \tag{4.31}$$

$$\Gamma_{n,T} := -D_n^{-1}\partial_\theta^2 \mathbb{H}_n(\tilde{\theta}_n)D_n^{-1} \xrightarrow{p} \Gamma_T(\theta_0;\beta). \tag{4.32}$$

4.4.1. *Proof of* (4.30). We may and do let $p_\alpha = p_\gamma = 1$. Let $R(x,\theta)$ denote any matrix-valued function on $\mathbb{R}$ such that $\sup_\theta |R(x,\theta)| \lesssim 1+|x|^C$; it may change at each appearance. By straightforward computations,

$$\begin{aligned}\frac{1}{nh^{2(1-1/\beta)}}\partial_\alpha^3 \mathbb{H}_n(\theta) &= \frac{1}{nh^{1-1/\beta}}\sum_{j=1}^n R_{j-1}(\theta)g_\beta(\epsilon_j(\theta)) \\ &\quad + \frac{1}{n}\sum_{j=1}^n \Big(R_{j-1}(\theta)\partial g_\beta(\epsilon_j(\theta)) + h^{1-1/\beta}R_{j-1}(\theta)\partial^2 g_\beta(\epsilon_j(\theta))\Big), \\ \frac{1}{nh^{2(1-1/\beta)}}\partial_\alpha^2\partial_\gamma \mathbb{H}_n(\theta) &= \frac{1}{nh^{1-1/\beta}}\sum_{j=1}^n \Big(R_{j-1}(\theta)g_\beta(\epsilon_j(\theta)) + R_{j-1}(\theta)\epsilon_j(\theta)\partial g_\beta(\epsilon_j(\theta))\Big) \\ &\quad + \frac{1}{n}\sum_{j=1}^n \Big(R_{j-1}(\theta)\epsilon_j(\theta)\partial^2 g_\beta(\epsilon_j(\theta)) + R_{j-1}(\theta)\partial g_\beta(\epsilon_j(\theta))\Big), \\ \frac{1}{n}\partial_\gamma^3 \mathbb{H}_n(\theta) &= \frac{1}{n}\sum_{j=1}^n \Big(R_{j-1}(\theta) + R_{j-1}(\theta)\epsilon_j(\theta)g_\beta(\epsilon_j(\theta)) \\ &\quad + R_{j-1}(\theta)\epsilon_j^2(\theta)\partial g_\beta(\epsilon_j(\theta)) + R_{j-1}(\theta)\epsilon_j^3(\theta)\partial^2 g_\beta(\epsilon_j(\theta))\Big), \\ \frac{1}{nh^{1-1/\beta}}\partial_\alpha\partial_\gamma^2 \mathbb{H}_n(\theta) &= \frac{1}{n}\sum_{j=1}^n \Big(R_{j-1}(\theta)g_\beta(\epsilon_j(\theta)) + R_{j-1}(\theta)\epsilon_j(\theta)\partial g_\beta(\epsilon_j(\theta)) \\ &\quad + R_{j-1}(\theta)\epsilon_j^2(\theta)\partial^2 g_\beta(\epsilon_j(\theta))\Big).\end{aligned}$$

Because of (4.25), all the terms having the factor "$1/n$" in front of the summation sign in the above right-hand sides are $O_p^*(1)$. Hence we only need to take care of the remaining terms. But the functions $y \mapsto g_\beta(y)$ and $y \mapsto y\partial g_\beta(y)$ are odd, so that Proposition 4.5(3) concludes that both $\frac{1}{nh^{1-1/\beta}}\sum_{j=1}^n R_{j-1}(\theta)g_\beta(\epsilon_j(\theta))$ and $\frac{1}{nh^{1-1/\beta}}\sum_{j=1}^n R_{j-1}(\theta)\epsilon_j(\theta)\partial g_\beta(\epsilon_j(\theta))$ are $O_p^*(1)$. These observations are enough to deduce (4.30).

4.4.2. *Proof of* (4.31). We will apply the general stable central limit theorem due to Jacod [19], the crucial finding in which is the characterization result for conditionally Gaussian continuous-time martingales defined on an extended probability space. Nowadays it is one of established fundamental tools to derive asymptotic distributional results for high-frequency over a fixed time period, the technique essentially dating back to [12] and later formulated by [19] for a much more general model. The foremost important point is that Jacod's results not only can deal with very general triangular arrays of random variables, but also do not require the nesting condition on the underlying filtration, which is assumed in most of the existing stable convergence results, and fails to hold for high-frequency data models.

Let $\epsilon_j := \epsilon_j(\theta_0)$. We have

$$\Delta_{n,T} = \left(\frac{1}{\sqrt{n}h^{1-1/\beta}}\partial_\alpha \mathbb{H}_n(\theta_0),\ \frac{1}{\sqrt{n}}\partial_\gamma \mathbb{H}_n(\theta_0)\right)$$

$$= \left(-\frac{1}{\sqrt{n}} \sum_{j=1}^{n} \frac{\partial_\alpha a_{j-1}(\alpha_0)}{c_{j-1}(\gamma_0)} g_\beta(\epsilon_j),\, -\frac{1}{\sqrt{n}} \sum_{j=1}^{n} \frac{\partial_\gamma c_{j-1}(\gamma_0)}{c_{j-1}(\gamma_0)} \{1 + \epsilon_j g_\beta(\epsilon_j)\} \right).$$

For $t \in [0,T]$ we introduce the partial sum process in $\mathbb{D}([0,T];\mathbb{R}^p)$:

$$\Delta_{n,t} := \left(-\frac{1}{\sqrt{n}} \sum_{j=1}^{[nt/T]} \frac{\partial_\alpha a_{j-1}(\alpha_0)}{c_{j-1}(\gamma_0)} g_\beta(\epsilon_j),\, -\frac{1}{\sqrt{n}} \sum_{j=1}^{[nt/T]} \frac{\partial_\gamma c_{j-1}(\gamma_0)}{c_{j-1}(\gamma_0)} \{1 + \epsilon_j g_\beta(\epsilon_j)\} \right).$$

We set

$$\pi(x) = \mathrm{diag}\left(-\frac{\partial_\alpha a_{j-1}(\alpha_0)}{c_{j-1}(\gamma_0)},\, -\frac{\partial_\gamma c_{j-1}(\gamma_0)}{c_{j-1}(\gamma_0)} \right) \in \mathbb{R}^p \otimes \mathbb{R}^2,$$
$$\eta(y) = (g_\beta(y),\, 1 + y g_\beta(y)) = (g_\beta(y),\, k_\beta(y)) \in \mathbb{R}^2,$$

so that $\Delta_{n,t} = n^{-1/2} \sum_{j=1}^{n} \pi_{j-1}\eta(\epsilon_j)$. Write $\Gamma_t(\theta_0;\beta)$ for $\Gamma_T(\theta_0;\beta)$ with the integral sign "$\int_0^T$" replaced by "$\int_0^t$". By means of [19, Theorem 3-2] (or [22, Theorem IX.7.28]), the stable convergence (4.31) is implied by the following statements: for each $t \in [0,T]$ and for any bounded $(\mathcal{F}_t)$-adapted martingale M,

$$\sum_{j=1}^{[nt/T]} \mathbb{E}^{j-1}\left(\left| \frac{1}{\sqrt{n}} \pi_{j-1}\eta(\epsilon_j) \right|^4 \right) \xrightarrow{p} 0, \tag{4.33}$$

$$\frac{1}{n} \sum_{j=1}^{[nt/T]} \pi_{j-1} \mathbb{E}^{j-1}\Big\{ \Big(\eta(\epsilon_j) - \mathbb{E}^{j-1}\{\eta(\epsilon_j)\} \Big)^{\otimes 2} \Big\} \pi_{j-1}^\top \xrightarrow{p} \Gamma_t(\theta_0;\beta), \tag{4.34}$$

$$\sup_{t\in[0,T]} \left| \frac{1}{\sqrt{n}} \sum_{j=1}^{[nt/T]} \pi_{j-1} \mathbb{E}^{j-1}\{\eta(\epsilon_j)\} - \big(0,\, \mu_{T,\gamma}(\theta_0;\beta)\mathfrak{b}_\beta(\nu)\big) \right| \xrightarrow{p} 0, \tag{4.35}$$

$$\sum_{j=1}^{[nt/T]} \mathbb{E}^{j-1}\left(\frac{1}{\sqrt{n}} \pi_{j-1}\eta(\epsilon_j)\Delta_j M \right) \xrightarrow{p} 0. \tag{4.36}$$

The Lyapunov condition (4.33) trivially holds since η is bounded and $|\pi(x)| \lesssim 1 + |x|^C$. For (4.34), arguing as in the proof of Lemma 4.3 with $\int \eta(z)\phi_\beta(z)dz = 0$ and $|\int \eta(z)\{f_h(z) - \phi_\beta(z)\}dz| = O(n^{-1/2})$, we see that

$$\begin{aligned} \mathbb{E}^{j-1}\{\eta(\epsilon_j)\} &= \int \eta(z) f_h(z) dz + O_p^*(h^{2-1/\beta}) \\ &= \int \eta(z)\phi_\beta(z)dz + O_p^*(n^{-1/2}) = O_p^*(n^{-1/2}), \\ \mathbb{E}^{j-1}\left\{\eta^{\otimes 2}(\epsilon_j)\right\} &= \int \eta^{\otimes 2}(z) f_h(z) dz + O_p^*(h^{2-1/\beta}) \\ &= \int \eta^{\otimes 2}(z)\phi_\beta(z)dz + O_p^*(n^{-1/2}). \end{aligned} \tag{4.37}$$

Then the left-hand side of (4.34) equals

$$\frac{1}{n} \sum_{j=1}^{[nt/T]} \pi_{j-1} \left(\int \eta^{\otimes 2}(z)\phi_\beta(z)dz \right) \pi_{j-1}^\top + O_p(n^{-1/2}),$$

and by means of Lemma 4.4 the first term converges in probability to $\Gamma_t(\theta_0;\beta)$.

The uniform convergence (4.35) follows on applying (4.37) and Lemma 4.4:

$$\begin{aligned} \frac{1}{\sqrt{n}} \sum_{j=1}^{[nt/T]} \pi_{j-1}\mathbb{E}^{j-1}\{\eta(\epsilon_j)\} &= \frac{1}{n} \sum_{j=1}^{[nt/T]} \pi_{j-1} \left(\sqrt{n} \int \eta(z) f_h(z) dz \right) + O_p(\sqrt{n} h^{2-1/\beta}) \\ &= \frac{1}{n} \sum_{j=1}^{[nt/T]} \pi_{j-1} \left\{ (0,\, \mathfrak{b}_\beta(\nu)) + o(1) \right\} + O_p(h^{3/2-1/\beta}) \\ &= \left(0,\, \frac{1}{T} \int_0^t \pi(X_s,\theta_0) ds \mathfrak{b}_\beta(\nu) \right) + o_p(1) \\ &\xrightarrow{p} \big(0,\, \mu_{T,\gamma}(\theta_0;\beta)\mathfrak{b}_\beta(\nu)\big) \end{aligned}$$

uniformly in $t \in [0,T]$.

Finally we turn to (4.36). Recall that we are assuming that $\mathcal{F}_t = \sigma(X_0) \vee \sigma(J_s; s \le t)$. By means of the decomposition theorem [22, Theorem I.4.18] for local martingales, we may write $M = M^c + M^d$ for a continuous part M^c and the associated purely discontinuous part M^d. Our underlying probability space supports no Wiener process, hence in view of the martingale representation theorem [22, Theorem III.4.34] for M, we may set $M^c = 0$. To show (4.35) we will follow a similar way to [52] with successive use of general theory of martingales convergence.

It suffices to prove the claim when both π and η are real-valued. The jumps of M over $[0,T]$ are bounded, and we have $M_t^n := \sum_{j=1}^{[nt/T]} \Delta_j M \xrightarrow{a.s.} M_t$ in $\mathbb{D}([0,T];\mathbb{R})$. Let

$$N_t^n := \sum_{j=1}^{[nt/T]} \frac{1}{\sqrt{n}} \pi_{j-1}\eta(\epsilon_j),$$

then for each n, N^n is a local martingale with respect to $(\mathcal{F}_t)$; then (4.36) equals that $\langle M^n, N^n\rangle_t \to 0$ for each $t \le T$. Following the same route as in the proof of (4.34), we see that the angle-bracket process

$$\langle N^n\rangle_t = \frac{1}{n}\sum_{j=1}^{[nt/T]} \pi_{j-1}^2 \mathbb{E}^{j-1}\{\eta(\epsilon_j)\}$$

is C-tight, that is, it is tight in $\mathbb{D}([0,T];\mathbb{R})$ and any weak limit process has a.s. continuous sample paths. By means of [22, Theorem VI.4.13] we then deduce that (N^n) is tight in $\mathbb{D}([0,T];\mathbb{R})$. We also observe that $\lim_n \mathbb{P}(\sup_{t\le T}|\Delta N_t^n| > \epsilon) = 0$ for every $\epsilon > 0$, which automatically holds under the Lyapunov condition: (4.33) remains valid also for $\mathbb{E}^{j-1}$ replaced by $\mathbb{E}$, and hence

$$\mathbb{P}\left(\sup_{t\le T}|\Delta N_t^n| > \epsilon\right) \le \epsilon^{-4}\sum_{j=1}^{[nt/T]} \mathbb{E}\left(|\Delta_j N^n|^4\right) \lesssim \frac{1}{n}.$$

Thus we conclude from [22, Theorem VI.3.26(iii)] that (N^n) is C-tight.

Fix any $\{n'\} \subset \mathbb{N}$. By [22, Theorem VI.3.33] the process $H^n := (M^n, N^n)$ is tight in $\mathbb{D}([0,T];\mathbb{R})$, so that by Prokhorov's theorem we can pick a subsequence $\{n''\} \subset \{n'\}$ for which there exists a process $H = (M,N)$ such that M and N are purely discontinuous and continuous, respectively, and that $H^{n''} \xrightarrow{\mathcal{L}} H$ along $\{n''\}$ in $\mathbb{D}([0,T];\mathbb{R})$. We have

$$\sup_n \mathbb{E}\left(\max_{j\le n}|\Delta_j N^n|\right) \lesssim \sup_n \frac{1}{\sqrt{n}}\mathbb{E}\left(1 + \sup_{t\le T}|X_t|^C\right) < \infty,$$

hence it follows from [22, Corollary VI.6.30] that the sequence $(H^{n''})$ is predictably uniformly tight, in particular, $(H^{n''}, [H^{n''}]) \xrightarrow{\mathcal{L}} (H, [H])$, with the limit quadratic-variation process $[H] = [M,N] = \langle M^c, N^c\rangle + \sum_{s\le\cdot}(\Delta M_s)(\Delta N_s) = 0$ a.s. identically. We have seen that given any $\{n'\} \subset \mathbb{N}$ we can find $\{n''\} \subset \{n'\}$ for which $[H^{n''}] \xrightarrow{\mathcal{L}} 0$ along $\{n''\}$, from which we conclude that

$$[H^n]_t = [M^n, N^n]_t = \sum_{j=1}^{[nt/T]} \frac{1}{\sqrt{n}}\pi_{j-1}\eta(\epsilon_j)\Delta_j M \xrightarrow{p} 0 \tag{4.38}$$

in $\mathbb{D}([0,T];\mathbb{R})$. It remains to show that $[M^n,N^n]_t$ and $\langle M^n, N^n\rangle_t$ are asymptotically equivalent for each $t \le T$; then, (4.38) yields that $\langle M^n, N^n\rangle_t \xrightarrow{p} 0$, hence (4.36). This can be seen as follows: since the function η and the squared-jump sum process $\sum_{0<s\le t}(\Delta M_s^n)^2$ are bounded,

$$\begin{aligned}
\mathbb{E}\left\{\left([M^n,N^n]_t - \langle M^n,N^n\rangle_t\right)^2\right\} &\lesssim \mathbb{E}\left(\sum_{0<s\le t}(\Delta M_s^n \Delta N_s^n)^2\right) \\
&\lesssim \frac{1}{n}\mathbb{E}\left\{\left(1 + \sup_{t\le T}|X_t|^C\right)\sum_{0<s\le t}(\Delta M_s^n)^2\right\} \\
&\lesssim \frac{1}{n}\mathbb{E}\left(1 + \sup_{t\le T}|X_t|^C\right) \lesssim \frac{1}{n} \to 0.
\end{aligned}$$

Remark 4.8. The present setting of $\mathcal{F}_t$ is not essential. We may enlarge it as long as the martingale-representation arguments stay valid. Even when we have a Wiener process in our model, we can still follow the martingale-representation argument as in [52]. □

4.4.3. *Proof of* (4.32). The components of $\Gamma_{n,T}$ consist of

$$-\frac{1}{nh^{2(1-1/\beta)}}\partial_\alpha^2\mathbb{H}_n(\theta_0)=\frac{1}{nh^{1-1/\beta}}\sum_{j=1}^n\frac{\partial_\alpha^2 a_{j-1}(\alpha_0)}{c_{j-1}(\gamma_0)}g_\beta(\epsilon_j)-\frac{1}{n}\sum_{j=1}^n\frac{\{\partial_\alpha a_{j-1}(\alpha_0)\}^{\otimes 2}}{c_{j-1}^2(\gamma_0)}\partial g_\beta(\epsilon_j), \tag{4.39}$$

$$\begin{aligned}-\frac{1}{n}\partial_\gamma^2\mathbb{H}_n(\theta_0)=&-\frac{1}{n}\sum_{j=1}^n\frac{\partial_\gamma^2 c_{j-1}(\gamma_0)}{c_{j-1}(\gamma_0)}\{1+\epsilon_j g_\beta(\epsilon_j)\}\\&-\frac{1}{n}\sum_{j=1}^n\frac{\{\partial_\gamma c_{j-1}(\gamma_0)\}^{\otimes 2}}{c_{j-1}^2(\gamma_0)}\left\{1+2\epsilon_j g_\beta(\epsilon_j)+\epsilon_j^2\partial g_\beta(\epsilon_j)\right\},\end{aligned} \tag{4.40}$$

$$-\frac{1}{nh^{1-1/\beta}}\partial_\alpha\partial_\gamma\mathbb{H}_n(\theta_0)=-\frac{1}{n}\sum_{j=1}^n\frac{\{\partial_\alpha a_{j-1}(\alpha_0)\}\otimes\{\partial_\gamma c_{j-1}(\gamma_0)\}}{c_{j-1}^2(\gamma_0)}\{g_\beta(\epsilon_j)+\epsilon_j\partial g_\beta(\epsilon_j)\}. \tag{4.41}$$

By (4.6) and Corollary 4.6, the first term in the right-hand side of (4.39) is $o_p(1)$. It follows from Proposition 4.5 that the second term equals

$$\begin{aligned}-\frac{1}{nh^{2(1-1/\beta)}}\partial_\alpha^2\mathbb{H}_n(\theta_0)&=-\frac{1}{n}\sum_{j=1}^n\frac{\{\partial_\alpha a_{j-1}(\alpha_0)\}^{\otimes 2}}{c_{j-1}^2(\gamma_0)}\int\partial g_\beta(z)\phi_\beta(z)dz+o_p(1)\\&=\frac{1}{n}\sum_{j=1}^n\frac{\{\partial_\alpha a_{j-1}(\alpha_0)\}^{\otimes 2}}{c_{j-1}^2(\gamma_0)}\int g_\beta^2(y)\phi_\beta(z)dz+o_p(1)\\&=C_\alpha(\beta)\Sigma_{T,\alpha}(\theta_0)+o_p(1).\end{aligned}$$

For $-n^{-1}\partial_\gamma^2\mathbb{H}_n(\theta_0)$, by Proposition 4.5 and $\int k_\beta(y)\phi_\beta(y)dy=0$ we see that the first term in the right-hand side of (4.40) is $o_p(1)$. As for the second term, noting the function $l_\beta(y):=1+2yg_\beta(y)+y^2\partial g_\beta(y)$ satisfies that $\int l_\beta(y)\phi_\beta(y)dy=-\int k_\beta^2(y)\phi_\beta(y)dy=-C_\gamma(\beta)$, we get

$$\begin{aligned}-\frac{1}{n}\partial_\gamma^2\mathbb{H}_n(\theta_0)&=-\frac{1}{n}\sum_{j=1}^n\frac{\{\partial_\gamma c_{j-1}(\gamma_0)\}^{\otimes 2}}{c_{j-1}^2(\gamma_0)}\int l_\beta(z)\phi_\beta(z)dz+o_p(1)\\&=C_\gamma(\beta)\Sigma_{T,\gamma}(\gamma_0)+o_p(1).\end{aligned}$$

Finally, since $y\mapsto g_\beta(y)+y\partial g_\beta(y)=g_\beta(y)k_\beta(y)$ is odd, Corollary 4.6 concludes that $-(nh^{1-1/\beta})^{-1}\partial_\alpha\partial_\gamma\mathbb{H}_n(\theta_0)=o_p(1)$, completing the proof of (4.32).

4.4.4. *Proof of Corollary 3.3.* The convergence (3.11) follows from (4.27), (4.29), and (4.30).

Acknowledgement. This work was partly supported by JSPS KAKENHI Grant Number 26400204 and JST, CREST.

References

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

[2] Y. Aït-Sahalia and J. Jacod. *High-Frequency Financial Econometrics*. Princeton University Press, 2014.

[3] A. Akharif and M. Hallin. Efficient detection of random coefficients in autoregressive models. *Ann. Statist.*, 31(2):675–704, 2003. Dedicated to the memory of Herbert E. Robbins.

[4] B. Andrews, M. Calder, and R. A. Davis. Maximum likelihood estimation for α-stable autoregressive processes. *Ann. Statist.*, 37(4):1946–1982, 2009.

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

[6] O. E. Barndorff-Nielsen, J. M. Corcuera, and M. Podolskij. Limit theorems for functionals of higher order differences of Brownian semi-stationary processes. In *Prokhorov and contemporary probability theory*, volume 33 of *Springer Proc. Math. Stat.*, pages 69–96. Springer, Heidelberg, 2013.

[7] J. Bertoin and R. A. Doney. Spitzer's condition for random walks and Lévy processes. *Ann. Inst. H. Poincaré Probab. Statist.*, 33(2):167–178, 1997.

[8] E. Clément and A. Gloter. Local asymptotic mixed normality property for discretely observed stochastic differential equations driven by stable Lévy processes. *Stochastic Process. Appl.*, 125(6):2316–2352, 2015.

[9] D. R. Cox and N. Reid. Parameter orthogonality and approximate conditional inference. *J. Roy. Statist. Soc. Ser. B*, 49(1):1–39, 1987. With a discussion.

[10] A. Dvoretzky. Asymptotic normality of sums of dependent random vectors. In *Multivariate analysis, IV (Proc. Fourth Internat. Sympos., Dayton, Ohio, 1975)*, pages 23–34. North-Holland, Amsterdam, 1977.

[11] J. Fan, L. Qi, and D. Xiu. Quasi-maximum likelihood estimation of garch models with heavy-tailed likelihoods. *J. Bus. Econom. Statist.*, 32(2):178–191, 2014.

[12] V. Genon-Catalot and J. Jacod. On the estimation of the diffusion coefficient for multi-dimensional diffusion processes. *Ann. Inst. H. Poincaré Probab. Statist.*, 29(1):119–151, 1993.

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

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

[15] M. Grabchak and G. Samorodnitsky. Do financial returns have finite or infinite variance? A paradox and an explanation. *Quant. Finance*, 10(8):883–893, 2010.

[16] I. S. Gradshteyn and I. M. Ryzhik. *Table of integrals, series, and products*. Elsevier/Academic Press, Amsterdam, seventh edition, 2007. Translated from the Russian, Translation edited and with a preface by Alan Jeffrey and Daniel Zwillinger, With one CD-ROM (Windows, Macintosh and UNIX).

[17] E. Häusler and H. Luschgy. *Stable convergence and stable limit theorems*, volume 74 of *Probability Theory and Stochastic Modelling*. Springer, Cham, 2015.

[18] D. Ivanenko, A. M. Kulik, and H. Masuda. Uniform lan property of locally stable lévy process observed at high frequency. *ALEA Lat. Am. J. Probab. Math. Stat.*, 12(2):835–862, 2015.

[19] J. Jacod. On continuous conditional Gaussian martingales and stable convergence in law. In *Séminaire de Probabilités, XXXI*, volume 1655 of *Lecture Notes in Math.*, pages 232–246. Springer, Berlin, 1997.

[20] J. Jacod. Statistics and high-frequency data. In *Statistical methods for stochastic differential equations*, volume 124 of *Monogr. Statist. Appl. Probab.*, pages 191–310. CRC Press, Boca Raton, FL, 2012.

[21] J. Jacod and P. Protter. *Discretization of processes*, volume 67 of *Stochastic Modelling and Applied Probability*. Springer, Heidelberg, 2012.

[22] J. Jacod and A. N. Shiryaev. *Limit theorems for stochastic processes*. Springer-Verlag, Berlin, second edition, 2003.

[23] A. Janicki and A. Weron. *Simulation and chaotic behavior of α-stable stochastic processes*, volume 178 of *Monographs and Textbooks in Pure and Applied Mathematics*. Marcel Dekker Inc., New York, 1994.

[24] B.-Y. Jing, X.-B. Kong, and Z. Liu. Modeling high-frequency financial data by pure jump processes. *Ann. Statist.*, 40(2):759–784, 2012.

[25] B. Jørgensen and S. J. Knudsen. Parameter orthogonality and bias adjustment for estimating functions. *Scand. J. Statist.*, 31(1):93–114, 2004.

[26] R. Kawai and H. Masuda. Local asymptotic normality for normal inverse Gaussian Lévy processes with high-frequency sampling. *ESAIM Probab. Stat.*, 17:13–32, 2013.

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

[28] V. P. Knopova. On the speed of convergence in the local limit theorem for triangular arrays of random variables. *Theory Stoch. Process.*, 18(2):24–32, 2012.

[29] X.-B. Kong, Z. Liu, and B.-Y. Jing. Testing for pure-jump processes for high-frequency data. *Ann. Statist.*, 43(2):847–877, 2015.

[30] H. Luschgy and G. Pagès. Moment estimates for Lévy processes. *Electron. Commun. Probab.*, 13:422–434, 2008.

[31] H. Masuda. Simple estimators for parametric Markovian trend of ergodic processes based on sampled data. *J. Japan Statist. Soc.*, 35(2):147–170, 2005.

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

[33] H. Masuda. On quasi-likelihood analyses for stochastic differential equations with jumps. In *Int. Statistical Inst.: Proc. 58th World Statistical Congress, 2011, Dublin (Session IPS007)*, pages 83–91, 2011.

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

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

[36] H. Masuda. Estimating an ergodic process driven by non-Gaussian noise. *J. Jpn. Stat. Soc. Jpn. Issue*, 44(2):471–495, 2015.

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

[38] I. Mizera and C. H. Müller. Breakdown points and variation exponents of robust M-estimators in linear models. *Ann. Statist.*, 27(4):1164–1177, 1999.

[39] I. Mizera and C. H. Müller. Breakdown points of Cauchy regression-scale estimators. *Statist. Probab. Lett.*, 57(1):79–89, 2002.

[40] V. Pipiras and M. S. Taqqu. Small and large scale asymptotics of some Lévy stochastic integrals. *Methodol. Comput. Appl. Probab.*, 10(2):299–314, 2008.

[41] P. E. Protter. *Stochastic integration and differential equations*, volume 21 of *Stochastic Modelling and Applied Probability*. Springer-Verlag, Berlin, 2005. Second edition. Version 2.1, Corrected third printing.

[42] P. Radchenko. Mixed-rates asymptotics. *Ann. Statist.*, 36(1):287–309, 2008.

[43] M. Reiß, M. Riedle, and O. van Gaans. Delay differential equations driven by Lévy processes: stationarity and Feller properties. *Stochastic Process. Appl.*, 116(10):1409–1432, 2006.

[44] J. Rosiński. Tempering stable processes. *Stochastic Process. Appl.*, 117(6):677–707, 2007.

[45] G. Samorodnitsky and M. S. Taqqu. *Stable non-Gaussian random processes*. Stochastic Modeling. Chapman & Hall, New York, 1994. Stochastic models with infinite variance.

[46] K.-i. Sato. *Lévy processes and infinitely divisible distributions*, volume 68 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1999. Translated from the 1990 Japanese original, Revised by the author.

[47] R. J. Serfling. A variation on scheffe's theorem, with application to nonparametric density estimation. Technical report, DTIC Document, Statistics Report M502 ONR Technical Report No. 137, 1979.

[48] M. Sharpe. Zeroes of infinitely divisible densities. *Ann. Math. Statist.*, 40:1503–1505, 1969.

[49] I. Stojkovic and O. van Gaans. Invariant measures and a stability theorem for locally Lipschitz stochastic delay equations. *Ann. Inst. Henri Poincaré Probab. Stat.*, 47(4):1121–1146, 2011.

[50] V. Todorov. Power variation from second order differences for pure jump semimartingales. *Stochastic Process. Appl.*, 123(7):2829–2850, 2013.

[51] V. Todorov and G. Tauchen. Limit theorems for power variations of pure-jump processes with application to activity estimation. *Ann. Appl. Probab.*, 21(2):546–588, 2011.

[52] V. Todorov and G. Tauchen. Realized Laplace transforms for pure-jump semimartingales. *Ann. Statist.*, 40(2):1233–1262, 2012.

[53] V. Todorov and G. Tauchen. Limit theorems for the empirical distribution function of scaled increments of Itô semimartingales at high frequencies. *Ann. Appl. Probab.*, 24(5):1850–1888, 2014.

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

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

[56] M. Uchida and N. Yoshida. Quasi likelihood analysis of volatility and nondegeneracy of statistical random field. *Stochastic Process. Appl.*, 123(7):2851–2876, 2013.

[57] N. Yoshida. Asymptotic behavior of M-estimator and related random field for diffusion process. *Ann. Inst. Statist. Math.*, 42(2):221–251, 1990.

[58] N. Yoshida. Estimation for diffusion processes from discrete observation. *J. Multivariate Anal.*, 41(2):220–242, 1992.

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

[60] K. Zhu and S. Ling. Global self-weighted and local quasi-maximum exponential likelihood estimators for ARMA-GARCH/IGARCH models. *Ann. Statist.*, 39(4):2131–2163, 2011.

[61] V. M. Zolotarev. *One-dimensional stable distributions*, volume 65 of *Translations of Mathematical Monographs*. American Mathematical Society, Providence, RI, 1986. Translated from the Russian by H. H. McFaden, Translation edited by Ben Silver.

Faculty of Mathematics, Kyushu University, 744 Motooka Nishi-ku Fukuoka 819-0395, Japan
E-mail address: hiroki@math.kyushu-u.ac.jp
URL: http://www2.math.kyushu-u.ac.jp/~hiroki/hmhp.html

List of MI Preprint Series, Kyushu University

The Global COE Program
Math-for-Industry Education & Research Hub

MI

MI2008-1 Takahiro ITO, Shuichi INOKUCHI & Yoshihiro MIZOGUCHI
Abstract collision systems simulated by cellular automata

MI2008-2 Eiji ONODERA
The intial value problem for a third-order dispersive flow into compact almost Hermitian manifolds

MI2008-3 Hiroaki KIDO
On isosceles sets in the 4-dimensional Euclidean space

MI2008-4 Hirofumi NOTSU
Numerical computations of cavity flow problems by a pressure stabilized characteristic-curve finite element scheme

MI2008-5 Yoshiyasu OZEKI
Torsion points of abelian varieties with values in nfinite extensions over a p-adic field

MI2008-6 Yoshiyuki TOMIYAMA
Lifting Galois representations over arbitrary number fields

MI2008-7 Takehiro HIROTSU & Setsuo TANIGUCHI
The random walk model revisited

MI2008-8 Silvia GANDY, Masaaki KANNO, Hirokazu ANAI & Kazuhiro YOKOYAMA
Optimizing a particular real root of a polynomial by a special cylindrical algebraic decomposition

MI2008-9 Kazufumi KIMOTO, Sho MATSUMOTO & Masato WAKAYAMA
Alpha-determinant cyclic modules and Jacobi polynomials

MI2008-10 Sangyeol LEE & Hiroki MASUDA
Jarque-Bera Normality Test for the Driving Lévy Process of a Discretely Observed Univariate SDE

MI2008-11 Hiroyuki CHIHARA & Eiji ONODERA
A third order dispersive flow for closed curves into almost Hermitian manifolds

MI2008-12 Takehiko KINOSHITA, Kouji HASHIMOTO and Mitsuhiro T. NAKAO
On the L^2 a priori error estimates to the finite element solution of elliptic problems with singular adjoint operator

MI2008-13 Jacques FARAUT and Masato WAKAYAMA
Hermitian symmetric spaces of tube type and multivariate Meixner-Pollaczek polynomials

MI2008-14 Takashi NAKAMURA
Riemann zeta-values, Euler polynomials and the best constant of Sobolev inequality

MI2008-15 Takashi NAKAMURA
Some topics related to Hurwitz-Lerch zeta functions

MI2009-1 Yasuhide FUKUMOTO
Global time evolution of viscous vortex rings

MI2009-2 Hidetoshi MATSUI & Sadanori KONISHI
Regularized functional regression modeling for functional response and predictors

MI2009-3 Hidetoshi MATSUI & Sadanori KONISHI
Variable selection for functional regression model via the L_1 regularization

MI2009-4 Shuichi KAWANO & Sadanori KONISHI
Nonlinear logistic discrimination via regularized Gaussian basis expansions

MI2009-5 Toshiro HIRANOUCHI & Yuichiro TAGUCHII
Flat modules and Groebner bases over truncated discrete valuation rings

MI2009-6 Kenji KAJIWARA & Yasuhiro OHTA
Bilinearization and Casorati determinant solutions to non-autonomous 1+1 dimensional discrete soliton equations

MI2009-7 Yoshiyuki KAGEI
Asymptotic behavior of solutions of the compressible Navier-Stokes equation around the plane Couette flow

MI2009-8 Shohei TATEISHI, Hidetoshi MATSUI & Sadanori KONISHI
Nonlinear regression modeling via the lasso-type regularization

MI2009-9 Takeshi TAKAISHI & Masato KIMURA
Phase field model for mode III crack growth in two dimensional elasticity

MI2009-10 Shingo SAITO
Generalisation of Mack's formula for claims reserving with arbitrary exponents for the variance assumption

MI2009-11 Kenji KAJIWARA, Masanobu KANEKO, Atsushi NOBE & Teruhisa TSUDA
Ultradiscretization of a solvable two-dimensional chaotic map associated with the Hesse cubic curve

MI2009-12 Tetsu MASUDA
Hypergeometric τ-functions of the q-Painlevé system of type $E_8^{(1)}$

MI2009-13 Hidenao IWANE, Hitoshi YANAMI, Hirokazu ANAI & Kazuhiro YOKOYAMA
A Practical Implementation of a Symbolic-Numeric Cylindrical Algebraic Decomposition for Quantifier Elimination

MI2009-14 Yasunori MAEKAWA
On Gaussian decay estimates of solutions to some linear elliptic equations and its applications

MI2009-15 Yuya ISHIHARA & Yoshiyuki KAGEI
Large time behavior of the semigroup on L^p spaces associated with the linearized compressible Navier-Stokes equation in a cylindrical domain

MI2009-16 Chikashi ARITA, Atsuo KUNIBA, Kazumitsu SAKAI & Tsuyoshi SAWABE
Spectrum in multi-species asymmetric simple exclusion process on a ring

MI2009-17 Masato WAKAYAMA & Keitaro YAMAMOTO
Non-linear algebraic differential equations satisfied by certain family of elliptic functions

MI2009-18 Me Me NAING & Yasuhide FUKUMOTO
Local Instability of an Elliptical Flow Subjected to a Coriolis Force

MI2009-19 Mitsunori KAYANO & Sadanori KONISHI
Sparse functional principal component analysis via regularized basis expansions and its application

MI2009-20 Shuichi KAWANO & Sadanori KONISHI
Semi-supervised logistic discrimination via regularized Gaussian basis expansions

MI2009-21 Hiroshi YOSHIDA, Yoshihiro MIWA & Masanobu KANEKO
Elliptic curves and Fibonacci numbers arising from Lindenmayer system with symbolic computations

MI2009-22 Eiji ONODERA
A remark on the global existence of a third order dispersive flow into locally Hermitian symmetric spaces

MI2009-23 Stjepan LUGOMER & Yasuhide FUKUMOTO
Generation of ribbons, helicoids and complex scherk surface in laser-matter Interactions

MI2009-24 Yu KAWAKAMI
Recent progress in value distribution of the hyperbolic Gauss map

MI2009-25 Takehiko KINOSHITA & Mitsuhiro T. NAKAO
On very accurate enclosure of the optimal constant in the a priori error estimates for H_0^2-projection

MI2009-26 Manabu YOSHIDA
Ramification of local fields and Fontaine's property (Pm)

MI2009-27 Yu KAWAKAMI
Value distribution of the hyperbolic Gauss maps for flat fronts in hyperbolic three-space

MI2009-28 Masahisa TABATA
Numerical simulation of fluid movement in an hourglass by an energy-stable finite element scheme

MI2009-29 Yoshiyuki KAGEI & Yasunori MAEKAWA
Asymptotic behaviors of solutions to evolution equations in the presence of translation and scaling invariance

MI2009-30 Yoshiyuki KAGEI & Yasunori MAEKAWA
On asymptotic behaviors of solutions to parabolic systems modelling chemotaxis

MI2009-31 Masato WAKAYAMA & Yoshinori YAMASAKI
Hecke's zeros and higher depth determinants

MI2009-32 Olivier PIRONNEAU & Masahisa TABATA
Stability and convergence of a Galerkin-characteristics finite element scheme of lumped mass type

MI2009-33 Chikashi ARITA
Queueing process with excluded-volume effect

MI2009-34 Kenji KAJIWARA, Nobutaka NAKAZONO & Teruhisa TSUDA
Projective reduction of the discrete Painlevé system of type$(A_2 + A_1)^{(1)}$

MI2009-35 Yosuke MIZUYAMA, Takamasa SHINDE, Masahisa TABATA & Daisuke TAGAMI
Finite element computation for scattering problems of micro-hologram using DtN map

MI2009-36 Reiichiro KAWAI & Hiroki MASUDA
Exact simulation of finite variation tempered stable Ornstein-Uhlenbeck processes

MI2009-37 Hiroki MASUDA
On statistical aspects in calibrating a geometric skewed stable asset price model

MI2010-1 Hiroki MASUDA
Approximate self-weighted LAD estimation of discretely observed ergodic Ornstein-Uhlenbeck processes

MI2010-2 Reiichiro KAWAI & Hiroki MASUDA
Infinite variation tempered stable Ornstein-Uhlenbeck processes with discrete observations

MI2010-3 Kei HIROSE, Shuichi KAWANO, Daisuke MIIKE & Sadanori KONISHI
Hyper-parameter selection in Bayesian structural equation models

MI2010-4 Nobuyuki IKEDA & Setsuo TANIGUCHI
The Itô-Nisio theorem, quadratic Wiener functionals, and 1-solitons

MI2010-5 Shohei TATEISHI & Sadanori KONISHI
Nonlinear regression modeling and detecting change point via the relevance vector machine

MI2010-6 Shuichi KAWANO, Toshihiro MISUMI & Sadanori KONISHI
Semi-supervised logistic discrimination via graph-based regularization

MI2010-7 Teruhisa TSUDA
UC hierarchy and monodromy preserving deformation

MI2010-8 Takahiro ITO
Abstract collision systems on groups

MI2010-25 Toshimitsu TAKAESU
On the Essential Self-Adjointness of Anti-Commutative Operators

MI2010-26 Reiichiro KAWAI & Hiroki MASUDA
On the local asymptotic behavior of the likelihood function for Meixner Lévy processes under high-frequency sampling

MI2010-27 Chikashi ARITA & Daichi YANAGISAWA
Exclusive Queueing Process with Discrete Time

MI2010-28 Jun-ichi INOGUCHI, Kenji KAJIWARA, Nozomu MATSUURA & Yasuhiro OHTA
Motion and Bäcklund transformations of discrete plane curves

MI2010-29 Takanori YASUDA, Masaya YASUDA, Takeshi SHIMOYAMA & Jun KOGURE
On the Number of the Pairing-friendly Curves

MI2010-30 Chikashi ARITA & Kohei MOTEGI
Spin-spin correlation functions of the q-VBS state of an integer spin model

MI2010-31 Shohei TATEISHI & Sadanori KONISHI
Nonlinear regression modeling and spike detection via Gaussian basis expansions

MI2010-32 Nobutaka NAKAZONO
Hypergeometric τ functions of the q-Painlevé systems of type $(A_2 + A_1)^{(1)}$

MI2010-33 Yoshiyuki KAGEI
Global existence of solutions to the compressible Navier-Stokes equation around parallel flows

MI2010-34 Nobushige KUROKAWA, Masato WAKAYAMA & Yoshinori YAMASAKI
Milnor-Selberg zeta functions and zeta regularizations

MI2010-35 Kissani PERERA & Yoshihiro MIZOGUCHI
Laplacian energy of directed graphs and minimizing maximum outdegree algorithms

MI2010-36 Takanori YASUDA
CAP representations of inner forms of $Sp(4)$ with respect to Klingen parabolic subgroup

MI2010-37 Chikashi ARITA & Andreas SCHADSCHNEIDER
Dynamical analysis of the exclusive queueing process

MI2011-1 Yasuhide FUKUMOTO& Alexander B. SAMOKHIN
Singular electromagnetic modes in an anisotropic medium

MI2011-2 Hiroki KONDO, Shingo SAITO & Setsuo TANIGUCHI
Asymptotic tail dependence of the normal copula

MI2011-3 Takehiro HIROTSU, Hiroki KONDO, Shingo SAITO, Takuya SATO, Tatsushi TANAKA & Setsuo TANIGUCHI
Anderson-Darling test and the Malliavin calculus

MI2011-4 Hiroshi INOUE, Shohei TATEISHI & Sadanori KONISHI
Nonlinear regression modeling via Compressed Sensing

MI2011-5 Hiroshi INOUE
Implications in Compressed Sensing and the Restricted Isometry Property

MI2011-6 Daeju KIM & Sadanori KONISHI
Predictive information criterion for nonlinear regression model based on basis expansion methods

MI2011-7 Shohei TATEISHI, Chiaki KINJYO & Sadanori KONISHI
Group variable selection via relevance vector machine

MI2011-8 Jan BREZINA & Yoshiyuki KAGEI
Decay properties of solutions to the linearized compressible Navier-Stokes equation around time-periodic parallel flow
Group variable selection via relevance vector machine

MI2011-9 Chikashi ARITA, Arvind AYYER, Kirone MALLICK & Sylvain PROLHAC
Recursive structures in the multispecies TASEP

MI2011-10 Kazunori YASUTAKE
On projective space bundle with nef normalized tautological line bundle

MI2011-11 Hisashi ANDO, Mike HAY, Kenji KAJIWARA & Tetsu MASUDA
An explicit formula for the discrete power function associated with circle patterns of Schramm type

MI2011-12 Yoshiyuki KAGEI
Asymptotic behavior of solutions to the compressible Navier-Stokes equation around a parallel flow

MI2011-13 Vladimír CHALUPECKÝ & Adrian MUNTEAN
Semi-discrete finite difference multiscale scheme for a concrete corrosion model: approximation estimates and convergence

MI2011-14 Jun-ichi INOGUCHI, Kenji KAJIWARA, Nozomu MATSUURA & Yasuhiro OHTA
Explicit solutions to the semi-discrete modified KdV equation and motion of discrete plane curves

MI2011-15 Hiroshi INOUE
A generalization of restricted isometry property and applications to compressed sensing

MI2011-16 Yu KAWAKAMI
A ramification theorem for the ratio of canonical forms of flat surfaces in hyperbolic three-space

MI2011-17 Naoyuki KAMIYAMA
Matroid intersection with priority constraints

MI2012-1 Kazufumi KIMOTO & Masato WAKAYAMA
Spectrum of non-commutative harmonic oscillators and residual modular forms

MI2012-2 Hiroki MASUDA
Mighty convergence of the Gaussian quasi-likelihood random fields for ergodic Levy driven SDE observed at high frequency

MI2012-3 Hiroshi INOUE
A Weak RIP of theory of compressed sensing and LASSO

MI2012-4 Yasuhide FUKUMOTO & Youich MIE
Hamiltonian bifurcation theory for a rotating flow subject to elliptic straining field

MI2012-5 Yu KAWAKAMI
On the maximal number of exceptional values of Gauss maps for various classes of surfaces

MI2012-6 Marcio GAMEIRO, Yasuaki HIRAOKA, Shunsuke IZUMI, Miroslav KRAMAR, Konstantin MISCHAIKOW & Vidit NANDA
Topological Measurement of Protein Compressibility via Persistence Diagrams

MI2012-7 Nobutaka NAKAZONO & Seiji NISHIOKA
Solutions to a q-analog of Painlevé III equation of type $D_7^{(1)}$

MI2012-8 Naoyuki KAMIYAMA
A new approach to the Pareto stable matching problem

MI2012-9 Jan BREZINA & Yoshiyuki KAGEI
Spectral properties of the linearized compressible Navier-Stokes equation around time-periodic parallel flow

MI2012-10 Jan BREZINA
Asymptotic behavior of solutions to the compressible Navier-Stokes equation around a time-periodic parallel flow

MI2012-11 Daeju KIM, Shuichi KAWANO & Yoshiyuki NINOMIYA
Adaptive basis expansion via the extended fused lasso

MI2012-12 Masato WAKAYAMA
On simplicity of the lowest eigenvalue of non-commutative harmonic oscillators

MI2012-13 Masatoshi OKITA
On the convergence rates for the compressible
Navier- Stokes equations with potential force

MI2013-1 Abuduwaili PAERHATI & Yasuhide FUKUMOTO
A Counter-example to Thomson-Tait-Chetayev's Theorem

MI2013-2 Yasuhide FUKUMOTO & Hirofumi SAKUMA
A unified view of topological invariants of barotropic and baroclinic fluids and their application to formal stability analysis of three-dimensional ideal gas flows

MI2013-3 Hiroki MASUDA
Asymptotics for functionals of self-normalized residuals of discretely observed stochastic processes

MI2013-4 Naoyuki KAMIYAMA
On Counting Output Patterns of Logic Circuits

MI2013-5 Hiroshi INOUE
RIPless Theory for Compressed Sensing

MI2013-6 Hiroshi INOUE
Improved bounds on Restricted isometry for compressed sensing

MI2013-7 Hidetoshi MATSUI
Variable and boundary selection for functional data via multiclass logistic regression modeling

MI2013-8 Hidetoshi MATSUI
Variable selection for varying coefficient models with the sparse regularization

MI2013-9 Naoyuki KAMIYAMA
Packing Arborescences in Acyclic Temporal Networks

MI2013-10 Masato WAKAYAMA
Equivalence between the eigenvalue problem of non-commutative harmonic oscillators and existence of holomorphic solutions of Heun's differential equations, eigenstates degeneration, and Rabi's model

MI2013-11 Masatoshi OKITA
Optimal decay rate for strong solutions in critical spaces to the compressible Navier-Stokes equations

MI2013-12 Shuichi KAWANO, Ibuki HOSHINA, Kazuki MATSUDA & Sadanori KONISHI
Predictive model selection criteria for Bayesian lasso

MI2013-13 Hayato CHIBA
The First Painleve Equation on the Weighted Projective Space

MI2013-14 Hidetoshi MATSUI
Variable selection for functional linear models with functional predictors and a functional response

MI2013-15 Naoyuki KAMIYAMA
The Fault-Tolerant Facility Location Problem with Submodular Penalties

MI2013-16 Hidetoshi MATSUI
Selection of classification boundaries using the logistic regression

MI2014-1 Naoyuki KAMIYAMA
Popular Matchings under Matroid Constraints

MI2014-2 Yasuhide FUKUMOTO & Youichi MIE
Lagrangian approach to weakly nonlinear interaction of Kelvin waves and a symmetry-breaking bifurcation of a rotating flow

MI2014-3 Reika AOYAMA
Decay estimates on solutions of the linearized compressible Navier-Stokes equation around a Parallel flow in a cylindrical domain

MI2014-4 Naoyuki KAMIYAMA
The Popular Condensation Problem under Matroid Constraints

MI2014-5 Yoshiyuki KAGEI & Kazuyuki TSUDA
Existence and stability of time periodic solution to the compressible Navier-Stokes equation for time periodic external force with symmetry

MI2014-6 This paper was withdrawn by the authors.

MI2014-7 Masatoshi OKITA
On decay estimate of strong solutions in critical spaces for the compressible Navier-Stokes equations

MI2014-8 Rong ZOU & Yasuhide FUKUMOTO
Local stability analysis of azimuthal magnetorotational instability of ideal MHD flows

MI2014-9 Yoshiyuki KAGEI & Naoki MAKIO
Spectral properties of the linearized semigroup of the compressible Navier-Stokes equation on a periodic layer

MI2014-10 Kazuyuki TSUDA
On the existence and stability of time periodic solution to the compressible Navier-Stokes equation on the whole space

MI2014-11 Yoshiyuki KAGEI & Takaaki NISHIDA
Instability of plane Poiseuille flow in viscous compressible gas

MI2014-12 Chien-Chung HUANG, Naonori KAKIMURA & Naoyuki KAMIYAMA
Exact and approximation algorithms for weighted matroid intersection

MI2014-13 Yusuke SHIMIZU
Moment convergence of regularized least-squares estimator for linear regression model

MI2015-1 Hidetoshi MATSUI & Yuta UMEZU
Sparse regularization for multivariate linear models for functional data

MI2015-2 Reika AOYAMA & Yoshiyuki KAGEI
Spectral properties of the semigroup for the linearized compressible Navier-Stokes equation around a parallel flow in a cylindrical domain

MI2015-3 Naoyuki KAMIYAMA
Stable Matchings with Ties, Master Preference Lists, and Matroid Constraints

MI2015-4 Reika AOYAMA & Yoshiyuki KAGEI
Large time behavior of solutions to the compressible Navier-Stokes equations around a parallel flow in a cylindrical domain

MI2015-5 Kazuyuki TSUDA
Existence and stability of time periodic solution to the compressible Navier-Stokes-Korteweg system on R^3

MI2015-6 Naoyuki KAMIYAMA
Popular Matchings with Ties and Matroid Constraints

MI2015-7 Shoichi EGUCHI & Hiroki MASUDA
Quasi-Bayesian model comparison for LAQ models

MI2015-8 Yoshiyuki KAGEI & Ryouta OOMACHI
Stability of time periodic solution of the Navier-Stokes equation on the half-space under oscillatory moving boundary condition

MI2016-1 Momonari KUDO
Analysis of an algorithm to compute the cohomology groups of coherent sheaves and its applications

MI2016-2 Yoshiyuki KAGEI & Masatoshi OKITA
Asymptotic profiles for the compressible Navier-Stokes equations on the whole space

MI2016-3 Shota ENOMOTO & Yoshiyuki KAGEI
Asymptotic behavior of the linearized semigroup at space-periodic stationary solution of the compressible Navier-Stokes equation

MI2016-4 Hiroki MASUDA
Non-Gaussian quasi-likelihood estimation of locally stable SDE