On Quasi-Likelihood Analyses for Stochastic Differential Equations with Jumps *†

Hiroki Masuda ‡

August 3, 2011

1 Main objective

We consider parametric estimation of *Stochastic Differential Equations* (*SDE* for short), assuming that the solution process is observed not fully but only at high-frequency discrete time points. The objective here is a solution process $X = (X_t)_{t\in\mathbb{R}_+}$ to the Markovian SDE

$$dX_t = a(X_t, \alpha)dt + b(X_t, \beta)dw_t + c(X_{t-}, \beta)dJ_t, \tag{1}$$

where the ingredients involved are: the unknown finite-dimensional parameter

$$\theta = (\alpha, \beta) \in \Theta_\alpha \times \Theta_\beta = \Theta \subset \mathbb{R}^p,$$

with $\Theta_\alpha \subset \mathbb{R}^{p_\alpha}$ and $\Theta_\beta \subset \mathbb{R}^{p_\beta}$ being bounded convex domains; a standard Wiener process w and a centered *pure-jump Lévy process* J, both being one-dimensional, and the latter characterized by the Lévy measure ν such that $\nu(\mathbb{R}\backslash\{0\}) \in (0, \infty]$; the initial variable X_0, with its law being possibly unknown, independent of the driving process (w, J); and finally, the real-valued measurable functions a on $\mathbb{R} \times \Theta_\alpha$, and b and c on $\mathbb{R} \times \Theta_\beta$, all of which are supposed to be known up to θ.

We want to estimate the true parameter $\theta_0 = (\alpha_0, \beta_0) \in \Theta$, supposed to exist, based on a high-frequency sample

$$(X_{t_0}, X_{t_1}, \ldots, X_{t_n}),$$

where $t_j = t_j^n = jh_n$, $j \le n$, for some sampling mesh h_n fulfilling $nh_n^2 \to 0$ as $n \to \infty$. We may say that α controls the trend structure, while β the noise-(martingale-)part structure. Some asymptotic distribution results will be given, based on which we can construct confidence regions of the parameter, and also immediately perform the Wald-type test. Although our main interest lies in the case of non-constant coefficients, it could be also possible to deal with Lévy processes.

2 Background and Wedderburn's quasi likelihood

The Lévy process is *the* continuous-time random walk, including Wiener and Poisson processes as special cases and serving as a building block for constructing a more general and flexible model; see Sato [30] for a systematic and extensive account of Lévy processes. SDE driven by a Lévy process can form a versatile-model class capturing time-varying phenomena observed in the real and

*This note was written for the presentation at the ISI 2011 Dublin (IPS007); http://www.isi2011.ie/content/index.php. The author is grateful to all the members of the IPS007, especially, to Prof. Nakahiro Yoshida for his invitation.

†*Keywords.* Asymptotic inference, Gaussian and non-Gaussian quasi-likelihood, high-frequency sampling, Lévy process, stochastic differential equation.

‡Institute of Mathematics for Industry, Kyushu University, 744 Motooka, Nishi-ku, Fukuoka 819-0395, Japan

nature worlds, hence we have high demand for developing statistical inference for them in order to extract valuable information from complicated time-evolution phenomena as well as to understand their essential skeleton structure. The application fields includes, for example, signal processing [27, 29], control and optimization through time-scale separation for reducing the system complexity [40].

It is the following common knowledge that makes the statistical problems pretty difficult: The exact maximum likelihood estimation is mostly infeasible for the statistical model (1), since the transition probability associated with X is not available in a closed form except for a few very special cases. Therefore the conventional statistical analyses based on the likelihood seem to have no utility, and for this reason we have to resort to some other feasible estimation procedure. Instead of using the "genuine" likelihood, one might think of M-estimation procedures. Among several possibilities, we are concerned here with some kinds of *Quasi Likelihood (QL) estimations*. The concept of QL was originally introduced by Wedderburn [39]; see also McCullagh [14]. The estimator stemming from a QL is called *Quasi Maximum Likelihood Estimator (QMLE)*, which is known to have the advantage of computational simplicity and robustness for model misspecification, in compensation for some amount of information loss. The estimation procedure is, roughly speaking, based on the "Gaussian approximation" of the transition distribution. We particularly call the QL of [39] the *Gaussian Quasi Likelihood (GQL)* and the resulting estimator the *Gaussian Quasi Maximum Likelihood Estimator (GQMLE)*, in order to distinguish it from the forthcoming other types of the quasi likelihoods.

It has been well-established that the GQMLE is a fundamental tool in estimating possibly *non-Gaussian and dependent* statistical models. Consider a time-series model $Y_1, \dots, Y_n$ in $\mathbb{R}$ with a fixed Y_0, for which we not have the exact knowledge of the transition probabilities $\mathcal{L}(Y_j | Y_0, Y_1, \dots, Y_{j-1})$, but we instead know the conditional mean and conditional variance, say $m_{j-1}(\theta) \in \mathbb{R}$ and $v_{j-1}(\theta) > 0$, where $\theta \in \Theta$ is an unknown parameter of interest. Then, the GQMLE is defined to be any maximizer of

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

over Θ^-, the closure of Θ; namely, we compute the likelihood of $(Y_1, Y_2, \dots, Y_n)$ as if the conditional law is the normal:

$$\mathcal{L}(Y_j | Y_0, Y_1, \dots, Y_{j-1}) \approx \mathcal{N}\left(m_{j-1}(\theta), v_{j-1}(\theta)\right).$$

Although it is not asymptotically efficient in general, it can serve as a widely applicable estimation procedure, and moreover, the GQL based estimation has a merit of robustness to model misspecification (of the Lévy measure, in our framework). Especially in the context of time series analysis, the GQL has been a quite popular tool for semiparametric estimation, and there exists vast amounts of literature concerning asymptotics of the GQMLE for models with possibly non-Gaussian error sequence; among others, we refer to Hall and Yao [8] and Straumann and Mikosch [33] for a class of conditionally heteroscedastic time series models, and Bardet and Wintenburger [2] for multidimensional causal time series, as well as the references therein.

One of our goal is to clarify what will occur if we follow the GQL based estimation procedure in our model in the presence of jumps. On the one hand, there already exist efficiency results if X is a diffusion (where $c(x, \beta) \equiv 0$); see Yoshida [41, 42], Kessler [11], and Gobet [6] as well as the references therein. One point is that the optimal rates of convergences of $\hat{\beta}_n$ is $\sqrt{n}$, while $\sqrt{nh_n}$ for $\hat{\alpha}_n$. Making use of this fact, Uchida and Yoshida [37] have proposed an adaptive estimation procedure based on a kind of GQMLE. Estimation of diffusions is still an active research area. See also Sørensen [32], whose bibliography includes many existing results concerning martingale estimating functions for discretely observed diffusions (not necessarily at high frequency). However, on the other hand, the issue has not been addressed enough in the presence of (possibly of infinite-variation) jumps, our main concern here: well, one can expect the GQL estimation is then far from being optimal, nevertheless, its implementation ease is worth being mentioned. It can be deduced that the GQMLE for (1) is then asymptotically normally distributed at rate $\sqrt{nh_n}$ for both α and β, implying that existence of any jump component slows down the rate of convergence of $\hat{\beta}_n$. Moreover, by means of the *Polynomial type Large Deviation Inequality (PLDI)* of Yoshida [42], we could derive the convergence of its moments to corresponding ones of the limit centered Gaussian distribution. The latter convergence especially serves as a fundamental tool when studying bias, variance, and etc., of statistics depending on the estimator; see Uchida and Yoshida [34, 35, 36] for the issue of model assessment.

As alternatives to the GQL, we will also present other types of QL doing better jobs than the GQMLE under some regularity conditions, when the continuous part of (1) vanishes.

3 Quasi likelihoods for the SDE observed at high-frequency

3.1 An overview

Here we describe the underlying idea of the QL framework for the model (1). Let us write

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

for any process Y, and $g_{j-1}(a) = g(X_{t_{j-1}}, a)$ for a function $g(\cdot,\cdot)$. We then think of the Euler-Maruyama approximation of X_{t_j} given $X_{t_{j-1}}$:

$$\begin{aligned} X_{t_j} &= X_{t_{j-1}} + \int_{t_{j-1}}^{t_j} a(X_s,\alpha)ds + \int_{t_{j-1}}^{t_j} b(X_s,\beta)dw_s \int_{t_{j-1}}^{t_j} c(X_{s-},\beta)dJ_s \\ &\approx X_{t_{j-1}} + a_{j-1}(\alpha)h_n + b_{j-1}(\beta)\Delta_j w + c_{j-1}(\beta)\Delta_j J \end{aligned} \tag{2}$$

under P_θ, the image measure of X associated with θ. *Utlizing (2) more or less, we try to fit some specific distribution for the transition law under P_θ, say $\mathcal{L}_\theta(X_{t_j}|X_{t_{j-1}})$.* If $c(x,\beta)\equiv 0$, so that X is a diffusion, then the local-Gauss approximation is an *efficient* choice (cf. Section 2 for relevant references):

$$\mathcal{L}_\theta(X_{t_j}|X_{t_{j-1}}) \approx \mathcal{N}(X_{t_{j-1}} + a_{j-1}(\alpha)h_n, b_{j-1}(\beta)^2 h_n). \tag{3}$$

But we are now focusing on X having jumps. Below, we will provide a framework including some *non-Gaussian* QL estimation.

Indeed, how to construct the QL could be a lot of things. There exist much less literature concerning non-Gaussian QL, compared with the Gaussian one. We refer to Fan et al. [4] for an adaptive non-Gaussian quasi likelihood estimation of a time seres model.

To give a unified account of our QL estimation procedures, let us introduce

$$v(x,\beta) := b(x,\beta)^2 + c(x,\beta)^2,$$

which may be regarded as a local-variance function. Our QL then takes the form

$$\mathbb{M}_n(\theta) = \sum_{j=1}^n \log\left\{\frac{1}{h_n^{1/\gamma} v_{j-1}(\beta)^{1/2}} f\left(\frac{\Delta_j X - a_{j-1}(\alpha)h_n}{h_n^{1/\gamma} v_{j-1}(\beta)^{1/2}}\right)\right\}, \tag{4}$$

for some probability density function f on $\mathbb{R}$, which is smooth enough and θ-free. Here γ denotes the activity index of the driving noise defined as follows: $\gamma = 2$ if $b(x,\beta)\not\equiv 0$, while $\gamma\in(0,2)$ if $b(x,\beta)\equiv 0$, $c(x,\beta)\not\equiv 0$, and J admits the local stable approximation in small time (see Section 3.3).

The corresponding QMLE is then defined to be any measurable mapping

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

By imposing some regularity conditions on f (smoothness, etc.), we could derive an asymptotic distribution results for the QL estimator corresponding to (4). However, needless to say, (4) is nothing but a formal setting and what is crucial is *how to pick a specific f and to make the estimation procedure implementable*.

The table summarizes the rates of convergence concerning the three QL, which will be mentioned a bit later: in the all cases, the asymptotic distribution of the QMLE is normal as soon as $nh_n \to \infty$.

Quasi likelihood	Estimation speeds α	β
Gauss (when $c(x,\beta) \equiv 0$)	$\sqrt{nh_n}$	$\sqrt{n}$
Gauss (when $c(x,\beta) \not\equiv 0$)	$\sqrt{nh_n}$	$\sqrt{nh_n}$
Non-Gaussian stable	$\sqrt{n}h_n^{1-1/\gamma}$	$\sqrt{n}$
Laplace (for Lévy OU process only, for now)	$\sqrt{n}h_n^{1-1/\gamma}$	$\times$

From the table one may immediately notice the importance of testing the presence of jump part, because the rate of convergence of the GQMLE $\hat{\beta}_n$ changes so that we cannot set $\limsup_{n\to\infty} nh_n < \infty$ in order to derive its consistency. Even when $nh_n \to \infty$, apart from the rate of convergence, the asymptotic covariance matrix takes a different form according as jumps do or do not exist. This directly leads to improper constructions of confidence zones. Fortunately, we do have a handy test statistic for the noise normality based on partial sums of the self-normalized powered residuals [22], and we have now developed an extension. The proposed test turned out to be asymptotically distribution-free under null and consistent against the presence of arbitrary jump part. The full details of the above results will be reported in due course in the papers prepared for publication: [23, 24, 25, 26].

As mentioned before, we could prove not only the asymptotic distribution (weak convergence) results but also the PLDI for some QL random fields in the scope of deriving the convergence of moments of (normalized) $\hat{\theta}_n$. More specifically, the PLDI is formulated as follows (see Yoshida [42] for details). We define random fields $\mathbb{Z}_n : U_n(\theta_0) := \{u \in \mathbb{R}^p : \theta_0 + r_n^{-1/2}u \in \Theta\} \to (0,\infty)$ by

$$\mathbb{Z}_n(u) = \mathbb{Z}_n(u;\theta_0) := \exp\left\{\mathbb{M}_n(\theta_0 + r_n^{-1/2}u) - \mathbb{M}_n(\theta_0)\right\}.$$

Then $\hat{u}_n := r_n^{1/2}(\hat{\theta}_n - \theta_0) \in \mathrm{argmax}_{\theta\in\Theta}\, \mathbb{Z}_n(\theta)$. We say that the PLDI holds if, given an $M > 0$, there exists a constant $C_M > 0$ such that the following estimate holds true:

$$\sup_{n\in\mathbb{N}} P_0\left[\sup_{|u|>r} \mathbb{Z}_n(u) \geq e^{-r}\right] \leq \frac{C_M}{r^M}, \quad r > 0, \tag{5}$$

where $P_0 := P_{\theta_0}$. Under (5), the random sequence $(\hat{u}_n)_{n\in\mathbb{N}}$ is $L^q(P_0)$-bounded for any $q \in (0, M)$ since

$$\sup_{n\in\mathbb{N}} P_0[|\hat{u}_n| > r] \leq \sup_{n\in\mathbb{N}} P_0\left[\sup_{|u|>r} \mathbb{Z}_n(u) \geq \mathbb{Z}_n(0;\theta_0)\right] = \sup_{n\in\mathbb{N}} P_0\left[\sup_{|u|>r} \mathbb{Z}_n(u) \geq 1\right] \leq \frac{C_M}{r^M}$$

for every $r > 0$. Therefore, as soon as $\hat{u}_n \to \hat{u}_0$ (P_0-weakly) we have $E[\varphi(\hat{u}_n)] \to E[\varphi(\hat{u}_0)]$ for every continuous φ such that $|\varphi(x)| \leq C(1+|x|)^q$. Taking φ in various ways, we can verify the asymptotic behavior of the bias, the mean squared error, etc., of $\hat{u}_n$. See also Chan and Ing [3].

3.2 Gaussian quasi-likelihood and the least-squares estimation

In case of non-degenerate diffusion coefficient, we assume that $E[J_t] = 0$ and $E[J_t^2] = t$ for each $t \in \mathbb{R}_+$, and that X is exponentially ergodic; see Masuda [16, 18, 19] and Kulik [12] for how to verify the ergodicity. Based on (2), we then consider the local-Gauss approximation (3), so that the QL (4) then becomes the GQML by taking $\gamma = 2$ and f to be the standard normal density:

$$\mathbb{M}_n(\theta) = -\sum_{j=1}^{n} \left\{ \log v_{j-1}(\theta) + \frac{(\Delta_j X - a_{j-1}(\alpha)h_n)^2}{v_{j-1}(\beta)h_n} \right\}.$$

In this case the Lévy measure ν of J need not to be specified. Note that *we may use the GQL even when* $b(x,\beta) \equiv 0$.

Also, let us note that the least-squares (and trajectory-fitting) type estimation formally corresponds to, say, the GQL estimation with known (constant) one-step conditional variance. See Masuda [17] for such drift-estimation procedures for possibly non-Markovian ergodic X. The least-squares type does not pay much attention to the martingale term, i.e., the diffusion coefficient in case of diffusions, thereby being particularly practicable while inefficient.

In case of compound-Poisson jump part, there exist efficient threshold estimation (jump-detection filter) approaches; see Mancini [15], Shimizu and Yoshida [31], and Ogihara and Yoshida [28] for details.

Once again, note that we are inevitably forced to make the assumption $nh_n \to \infty$ as long as using the GQL. Some time-varying phenomena might indeed seem more likely to be a diffusion process over "long-time" scale rather than over "short-time" scale. However, we generally do not know any specific relation between the model- and actual-time scales. However, as is mentioned in the next sections, the condition $nh_n \to \infty$ may be removed when $b(x,\beta) \equiv 0$.

3.3 Non-Gaussian stable quasi-likelihood

In case of the pure-jump case, so that (1) becomes

$$dX_t = a(X_t,\alpha)dt + c(X_{t-},\beta)dJ_t, \tag{6}$$

we assume that ν behaves like the symmetric stable Lévy measure near the origin. It can be shown under suitable conditions on ν that $\mathcal{L}(h_n^{-1/\gamma} J_{h_n})$, the law of normalized increment of J *in small time*, admits a density, say p_{h_n}, which is approximately the strict γ-stable density ϕ_γ corresponding to the characteristic function $u \mapsto \exp(-|u|^\gamma)$, where $\gamma \in (0,2)$. More specifically, we can find a constant $a_\nu > 0$ for which the uniform local-limit result holds true (the proof is similar to Masuda [21, Lemma 4.4]):

$$\sup_{y\in\mathbb{R}} \left| p_{h_n}(y) - \phi_\gamma(y) \right| \le C h_n^{a_\nu}.$$

In this case, the parameter $\gamma \in (0,2)$ denotes the Blumenthal-Getoor index of J. Many popular pure-jump Lévy processes admits the local-stable approximation, each suitable γ depending on the degree of ν-mass around the origin: for example, the generalized hyperbolic (except for the normal gamma), Meixner, tempered stable Lévy processes, and so on. Let us emphasize that the tail behavior of ν does *not* matter.

Note that, since (2) says

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

under P_θ, one may expect that $\{\epsilon_j(\theta_0)\}_{j\le n}$ can be approximated in some suitable sense under P_0 by the i.i.d. sequence $\{h_n^{-1/\gamma}\Delta_j J\}_{j\le n}$. This in turn implies that, as soon as f is good enough we can apply the general M-estimation theory together with the martingale limit theory, as was done in the diffusion case (see the references cited in Section 3.2). We may call (4) with $f = \phi_\gamma$ the (non-Gaussian) *Stable Quasi Likelihood (SQL for short)*, including the fully explicit *Cauchy quasi likelihood*. Under suitable conditions, we could deduce the (mixed-)normal asymptotic law of

$$\left(\sqrt{n}h_n^{1-1/\gamma}(\hat{\alpha}_n - \alpha_0), \sqrt{n}(\hat{\beta}_n - \beta_0)\right)$$

with the specific form of the asymptotic (random) covariance matrix.

We conjecture that the SQL estimator of (α, β) thus obtained is rate-optimal. Indeed, Kawai and Masuda [10] revealed that it is true for the normal inverse Gaussian Lévy process. However, we do not yet have the local asymptotic (mixed-)normality in a general model (1).

The SQL can target a rather rich class of infinite-variation J. We believe that the SQL-based approach serves as one of fundamental devices for estimating (6) based on high-frequency data.

3.4 Laplace quasi-likelihood

There is yet another type of non-Gaussian QL method called *Laplace Quasi Likelihood (LQL)* estimation, which may be more familiar under the name of the least absolute deviation estimation (L^1-minimum-distance estimation). It amounts to considering the QL associated with $x \mapsto f(x) := c\exp(-c|x-\mu|)/2$, $x \in \mathbb{R}$, $\mu \in \mathbb{R}$, with $c > 0$ being known. Unfortunately, it seems hard to deal with (6) in full generality because of the non-smoothness in θ of the QL. In this case we could target only the drift parameter $\alpha = (\alpha_1, \alpha_2)$ of the Lévy-Ornstein-Uhlenbeck process having the continuous-time first-order autoregressive structure. We here only mention the pure-jump case, keeping the local γ-stable approximation for the distributions $\mathcal{L}(h_n^{-1/\gamma} J_{h_n})$ in force. The SDE in question is then

$$dX_t = (\alpha_1 - \alpha_2 X_t)dt + \beta dJ_t.$$

Under suitable conditions, we can derive the asymptotic (mixed-)normality for

$$\sqrt{n}h_n^{1-1/\gamma}(\hat{\alpha}_n - \alpha_0).$$

The asymptotics is carried out by utilizing the convexity argument. We refer to Masuda [21] for details of the ergodic case.

One may more generally think of L^r-estimation ($r > 0$), corresponding to the quasi likelihood associated with the probability density $x \mapsto f(x) := rc^{1/r}\exp(-c|x-\mu|^r)/\{2\Gamma(1/r)\}$, $x \in \mathbb{R}$. However, in view of analytical tractability and optimization-programming convenience, the aforementioned Laplace QL is of special interest, and we could not find any great advantage for studying the case of $r \notin \{1, 2\}$.

4 Discussions

Let us give some remarks on our results and mention some ongoing subjects.

- For the GQL, it is straightforward to extend the result so as to subsume multivariate X. However, multivariate analogues for the other QL treated here would not be straightforward to formulate.

- Concerning the pure-jump case, we have seen that the SQL or LQL estimators of the drift parameter under $h_n = 1/n$ exhibits different rate of convergence from the diffusion case, so that the case of fixed-domain infill asymptotics is allowed.

  The asymptotic *mixed-normal* results in that case have been derived through the continuous-time martingale characterization of the conditionally Gaussian martingale. See Jacod [9] for the general limit theorem, the main idea of which has already appeared in Genon-Catalot and Jacod [5]. The rate of convergence and simple rate-optimal estimation were studied in Masuda [20] for stable Lévy processes; see also Aït-Sahalia and Jacod [1] for the asymptotic behavior of the Fisher information in a more general model setting. Let us note that we can deal with time-inhomogeneous case without essential change if $\limsup_{n\to\infty} nh_n < \infty$.

- In order to derive the asymptotic normality results even for heavy-tailed X (containing cases of heavy-tailed and long-memory diffusions such as a stationary diffusion process with Student marginal law, e.g., Hairer [7] and Veretennikov [38]), we could consider the *predictably tapered (self-weighted)* versions; see Masuda [21] for the case of LQL. This amounts to just putting in (4) a suitable non-negative weight function π, so that

$$\mathbb{M}_n(\theta) = -\sum_{j=1}^{n} \pi_{j-1} \left\{ \log v_{j-1}(\theta) + \frac{(\Delta_j X - a_{j-1}(\alpha)h_n)^2}{v_{j-1}(\beta)h_n} \right\}.$$

Although this is a trivial change, the tapering effect of the weight can be essential.

References

[1] Aït-Sahalia, Y. and Jacod, J. (2008), Fisher's information for discretely sampled Lévy processes. *Econometrica* **76**, 727–761.

[2] Bardet, J. M. and Wintenberger, O. (2009), Asymptotic normality of the quasi-maximum likelihood estimator for multidimensional causal processes. *Ann. Statist.* **37**, 2730–2759.

[3] Chan, N. H. and Ing, C. K. (2011) Uniform moment bounds of Fisher's information with applications to time series. *Ann. Statist.* **39**, 1526–1550.

[4] Fan, J., Qi, L. and Xiu, D. (2010), Non-Gaussian quasi maximum likelihood estimation of GARCH models. (January 22, 2010). Available at SSRN: http://ssrn.com/abstract=1540363

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

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

[7] Hairer, M. (2009), How hot can a heat bath get? *Comm. Math. Phys.* **292**, 131–177.

[8] Hall, P. and Yao, Q. (2003), Inference in ARCH and GARCH models with heavy-tailed errors. *Econometrica* **71**, 285–317.

[9] Jacod, J. (1997), On continuous conditional Gaussian martingales and stable convergence in law. Séminaire de Probabilités, XXXI, 232–246, *Lecture Notes in Math.*, **1655**, Springer, Berlin.

[10] Kawai, R. and Masuda, H. Local asymptotic normality for normal inverse Gaussian Lévy processes with high-frequency sampling. To appear in *ESAIM Probab. Stat.*

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

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

[13] Qi, L., Xiu, D., and Jianqing, F. (2010), Non-Gaussian Quasi Maximum Likelihood Estimation of GARCH Models (January 22, 2010). Available at SSRN: http://ssrn.com/abstract=1540363

[14] McCullagh, P. (1983), Quasilikelihood functions. *Ann. Statist.* **11**, 59–67.

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

[16] Masuda, H. (2004), On multidimensional Ornstein-Uhlenbeck processes driven by a general Lévy process. *Bernoulli* **10**, 1–24.

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

[18] Masuda, H. (2007), Ergodicity and exponential β-mixing bound for multidimensional diffusions with jumps. *Stochastic Processes Appl.* **117**, 35–56. [Erratum to this paper: (2009) *Stochastic Processes Appl.* **119**, 676–678.]

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

[20] Masuda, H. (2009), Joint estimation of discretely observed stable Levy processes with symmetric Levy density. Journal of the Japan Statistical Society 39 (2009), no.1, 49–75.

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

[22] Masuda, H. (2010), Approximate quadratic estimating function for discretely observed Levy driven SDEs with application to a noise normality test. MI Preprint Series, Kyushu University, 2010-21. Available at http://gcoe-mi.jp/temp/publish/bdfa1e74d21a7ee1a912e6b8877a576b.pdf

[23] Masuda, H., Mighty convergence of the Gaussian quasi-likelihood random fields for discretely observed multidimensional ergodic diffusions with jumps. In preparation.

[24] Masuda, H., On self-normalized residuals of discretely observed stochastic differential equations. In preparation.

[25] Masuda, H., Asymptotic mixed normality in the self-weighted LAD estimation for Lévy OU processes. In preparation.

[26] Masuda, H., Non-Gaussian quasi likelihood analysis for jump SDE observed at high frequency. In preparation.

[27] Nikias, C. L. and Shao, M. (1995), *Signal Processing with Alpha-Stable Distributions and Applications.* John Wiley & Sons, Inc., New York.

[28] Ogihara, T. and Yoshida, N., Quasi-likelihood analysis for the stochastic differential equation with jumps. To appear in *Stat. Inference Stoch. Process.*

[29] Poor, H. V. (1994), *An Introduction to Signal Detection and Estimation. Second Edition.* Springer-Verlag, New York.

[30] Sato, K. (1999), *Lévy Processes and Infinitely Divisible Distributions*. Cambridge University Press, Cambridge.

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

[32] Sørensen, M. (2008), Efficient estimation for ergodic diffusions sampled at high frequency. Preprint.

[33] Straumann, D. and Mikosch, T. (2006), Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series: a stochastic recurrence equations approach. *Ann. Statist.* **34**, 2449–2495.

[34] Uchida, M. and Yoshida, N. (2001), Information criteria in model selection for mixing processes. *Stat. Inference Stoch. Process.* **4**, 73–98.

[35] Uchida, M. and Yoshida, N. (2006), Asymptotic expansion and information criteria. *SUT J. Math.* **42**, 31–58.

[36] Uchida, Masayuki (2010), Contrast-based information criterion for ergodic diffusion processes from discrete observations. *Ann. Inst. Statist. Math.* **62**, 161–187.

[37] Uchida, M. and Yoshida, N. (2010), Adaptive estimation of an ergodic diffusion process based on sampled data. Preprint.

[38] Veretennikov, A. Y. (2006), On lower bounds for mixing coefficients of Markov diffusions. *From Stochastic Calculus to Mathematical Finance*, 623–633, Springer, Berlin.

[39] Wedderburn, R. W. M. (1974), Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. *Biometrika* **61**, 439–447.

[40] Yin, G. G. and Zhu, C. (2010), *Hybrid Switching Diffusions. Properties and Applications.* Springer, New York.

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

[42] Yoshida, N. (2005), Polynomial type large deviation inequalities and quasi-likelihood analysis for stochastic differential equations. To appear in *Ann. Inst. Statist. Math.*