Partially observed stochastic system における推定に関する一思案

シンポジウム「非線形モデリングと現象解析」2001 年 11 月 29 日
東京大学大学院数理科学研究科博士課程　増田 弘毅*

1 Introduction

Recently, Genon-Catalot, Jeantheau and Laredo [10] (henceforth designated GJL) have considered a two-dimensional simple stochastic volatility model (X, Y) which satisfies a dynamics

$$\begin{cases} dX_t &= V_0(X_t, \theta)dt + V(X_t, \theta)dW_t, \quad X_0 = \eta, \\ dY_t &= \sqrt{X_t}dB_t, \quad Y_0 = 0 \end{cases}$$

where W and B are two independent Wiener processes, X and Y are an unobserved strictly stationary hidden diffusion process and an observable process, respectively, and an initial random variable $\eta > 0$ a.s. is independent of (W, B) and distributed with the stationary distribution of X. From this continuous time model, GJL constructed a discrete time hidden Markov model in the sense of Bickel et al. [4], and they also presented a moment estimation for an unknown parameter θ included in a hidden diffusion process based on discretely sampled data with an equidistance. It seems advantageous that a diffusion coefficient of *hidden* diffusion process may contain unknown parameters since, so far, for estimation of unknown parameters in diffusion coefficient, many authors have considered an asymptotic rule such that the time mesh of sampled data tends to 0 (one of interesting alternative method based on martingale estimating functions has been developed by Bibby and Sørensen [3]). GJL focused on mixing property of the model and applied classical limit theorems based on the property.

Here, motivated by GJL, we follow the similar argument to them and extend their model so that it may contain general Lévy processes as driving (noise) processes of a latent process and an observed one. These two Lévy processes may be correlated. Moreover, instead of restriction to simple stochastic volatility models in GJL, we will assume only a certain measurability condition for dynamics of our models which looks somewhat abstract. Our model includes the case of GJL. In particular, a latent process may possess *the ε-Markov structure* which covers many type of stochastic processes such as strong solutions of stochastic differential equations (with time delay) and higher order discrete time Markov processes. Anyway, mixing property of a latent strictly stationary Markov process is also the essential assumption for us as GJL [10].

2 A class of partially observed stochastic systems

We will consider stochastic processes $S = \{(X_t, Y_t)\}_{t \in \mathbf{R}_+}$ and $L = \{(L_t^{(1)}, L_t^{(2)})\}_{t \in \mathbf{R}_+}$ on a given probability space $(\Omega, \mathcal{F}, P)$ where X and Y are $\mathbf{R}^{d_1}$- and $\mathbf{R}^{d_2}$-dimensional càdlàg stochastic processes, respectively. We assume that only Y is observable while X is unobservable. Moreover, $L^{(1)}$ and $L^{(2)}$ are $\mathbf{R}^{r_1}$ and $\mathbf{R}^{r_2}$-dimensional Lévy processes starting at the origin, respectively. Both of distributions of X and Y depend on unknown parameter $\theta \in \Theta \subseteq \mathbf{R}^p$, $p \geq 1$, which we want to estimate. Let P_θ and E_θ denote the probability measure corresponding to each $\theta \in \Theta$ and the expectation under P_θ, and let P_0 and E_0 denotes similar symbols corresponding to the true value $\theta_0 \in \Theta$. For $I \subseteq \mathbf{R}_+$, let $\mathcal{F}_I^X := \sigma[X_t : t \in I] \vee \mathcal{N}$,

*E-mail : flstf@ms.u-tokyo.ac.jp

$\mathcal{F}_I^{dL} := \sigma[L_t - L_s : s, t \in I] \vee \mathcal{N}$ and $\mathcal{F}_I^{X,dL} := \mathcal{F}_I^X \vee \mathcal{F}_I^{dL}$ where $\mathcal{N}$ denotes the σ-field generated by all the P-null sets. Finally, for a Euclidean space $\mathcal{E}$, let $\mathcal{B}(\mathcal{E})$ denote the Borel σ-field of subsets of $\mathcal{E}$.

We assume the following three conditions about stochastic structure of our models, and we will later assume a further one for moment estimation.

A1 (X_0, Y_0) and $(L_t^{(1)}, L_t^{(2)})_{t\in\mathbf{R}_+}$ are mutually independent while X_0 and Y_0, and $L^{(1)}$ and $L^{(2)}$ may be correlated, respectively.

A2 Under P_0, X is strictly stationary and α-mixing, that is,

$$\alpha_X(t) := \sup\{|P_0(A \cap B) - P_0(A)P_0(B)| : A \in \mathcal{F}_{[0]}^X, B \in \mathcal{F}_{[t,\infty)}^X\} \to 0$$

as t tends to infinity.

A3 X satisfies that $\sigma[X_t] \subseteq \mathcal{F}_{[s-\varepsilon,s]}^X \vee \mathcal{F}_{[s,t]}^{dL}$ for some $\varepsilon \geq 0$ and any s and t such that $0 \leq s \leq t$. For this ε and any s and t such that $\varepsilon \geq 0$, $s, t \in \mathbf{R}_+$, it is assumed that there exists a nonrandom functional F (not depending on the time) such that $Y_t - Y_s = F(X_{[s-\varepsilon,s]}, L_u - L_v; s \leq u, v \leq t)$. Moreover, the initial process $(X_t)_{t\in[-\varepsilon,0]}$ is independent of L.

Under above assumptions, we shall define a discrete time observable process y from Y by

$$y_j = Y_{j\Delta} - Y_{(j-1)\Delta}, \qquad j \in \mathbf{Z}_+^* \tag{1}$$

where $\Delta > 0$ is a fixed deterministic sampling interval and $\mathbf{Z}_+^* = \{1, 2, \dots\}$. Our estimators will be represented in terms of n data $\{y_j : j = 1, 2, \dots, n\}$.

A typical non-delayed (i.e. $\varepsilon = 0$) example is a strong solution of a system of stochastic differential equations

$$\begin{cases} dX_t = V_X(X_t, \theta)dL_t^{(1)}, & X_0 = x_0, \\ dY_t = V_Y(X_t, \theta)dL_t^{(2)}, & Y_0 = y_0, \end{cases}$$

where $V_X : \mathbf{R}^{d_1} \times \Theta \to \mathbf{R}^{d_1} \otimes \mathbf{R}^{r_1}$ and $V_Y : \mathbf{R}^{d_1} \times \Theta \to \mathbf{R}^{d_2} \otimes \mathbf{R}^{r_2}$ are non-anticipative functionals. Apart from diffusion models, the Ornstein-Uhlenbeck type process is an example for X with a view toward concrete statistical application to finance, see a recent stimulating work Barndorff-Nielsen [2]. Masuda [18] investigated mixing property for those processes.

Remark 1. The time index set $\mathbf{R}_+$ may be $\mathbf{N}$ since a discrete time processes $X = (X_n)_{n\in\mathbf{N}}$ can be embedded into a corresponding continuous time process $X = (X_t)_{t\in\mathbf{R}_+}$ with $X_t = X_{[t]}$. A1, A2 and A3 are then modified by an obvious way. In discrete time framework, we do not have to consider the construction (1) of y. We will assume that y is just a observed process. For example, consider a simple discrete Kalman-Bucy filter model:

$$\begin{cases} X_n = \phi X_{n-1} + \xi_n \\ y_n = \psi X_n + \eta_n \end{cases}$$

where ϕ and ψ are unknown parameters taking values in $\mathbf{R}$ appropriately, and $\xi = (\xi_j)_{j\in\mathbf{Z}_+^*}$ and $\eta = (\eta_j)_{j\in\mathbf{Z}_+^*}$ are any independent and identically distributed random variables which may be mutually correlated. ∎

3 Inherited strict stationarity and mixing property

The key feature of our model is that y inherits ergodicity from X under our assumptions. This is the essential property for our purpose. Namely, based on this fact, we can apply classical limit theorems in the next section under an additional assumption A4 below.

Theorem 1 (Masuda[17], submitted).
Under A1∼A3, $y = (y_j)_{j \in \mathbf{Z}_+^*}$ y is strict stationary and α-mixing satisfying

$$\alpha_y(k) \leq 4\alpha_X\big((k-1)\Delta - 2\varepsilon\big) \tag{2}$$

for $k \in \{u \in \mathbf{Z}_+^* : (k-1)\Delta - 2\varepsilon \geq 0\}$, where $\alpha_y(\cdot)$ denotes the α-mixing coefficient of y.

Since α-mixing means ergodicity for strictly stationary stochastic processes, Theorem 1 indeed asserts that y inherits ergodicity from X.

4 Applying classical limit theorems for estimating functions

4.1 Limit theorems

Following Theorem 1, we can adapt classical limit theorems for a function of observations $\{y_j : j = 1, 2, \dots, n\}$ with the same spirit to GJL. They are well-known Birkhoff's ergodic theorem and a central limit theorem for mixing processes.

Here we introduce some notations. Let $f : \mathbf{R}^{md_2} \to \mathbf{R}^p$ be an estimating function where $1 \leq m \leq n$. We will write $f = (f_1, f_2, \dots, f_p)$ so that $f_j : \mathbf{R}^{md_2} \to \mathbf{R}$. For such f_js, define a matrix $\Sigma = [\Sigma(f_k, f_l; m)]_{1 \leq k,l \leq p}$ with

$$\begin{aligned}\Sigma(f_k, f_l; m) \quad := \quad & Cov_0[f_k(y_1, \dots, y_m), f_l(y_1, \dots, y_m)] \\ & + \sum_{j=1}^{\infty} \Big\{ Cov_0[f_k(y_1, \dots, y_m), f_l(y_{j+1}, \dots, y_{j+m})] \\ & \quad + Cov_0[f_k(y_{j+1}, \dots, y_{j+m}), f_l(y_1, \dots, y_m)] \Big\}. \end{aligned} \tag{3}$$

We need a further assumption which is mainly necessary for the mixing central limit theorem:

A4. There exists a constant $\delta > 0$ such that $E_0[|f|^{2+\delta}] < \infty$ and $\sum_{k=1}^{\infty} \alpha_X(k)^{\delta/(\delta+2)} < \infty$.

Remark 2. Under the additional assumption A4, Σ is well-defined. ■

Remark 3. Suppose that y is one-dimensional. Then a typical case for f_j is of polynomial type like

$$f(y_{j+1}, \dots, y_{j+m}) = y_{j+1}^{k_1} \cdots y_{j+m}^{k_m}, \quad k_i \in \mathbf{N}.$$

We will not pursue the question of "which type of estimators is best in the sense of minimizing asymptotic variances?". However, these polynomial types are natural choices in terms of computational tractability.
■

The limit theorems on our focus are as follows.

Theorem 2 (Ergodic theorem). *Assume that A1∼A4 hold. Then we have*

$$\frac{1}{n} \sum_{j=1}^{n-m+1} f(y_j, y_{j+1}, \dots, y_{j+m-1}) \xrightarrow{P_0-a.s.} E_0[f(y_1, y_2, \dots, y_m)] \tag{4}$$

as n tends to infinity.

Theorem 3 (Central limit theorem). *Assume that A1∼A4 hold. If Σ is positive-definite, we have*

$$\frac{1}{\sqrt{n}} \sum_{j=1}^{n-m+1} \big\{ f(y_j, \dots, y_{j+m-1}) - E_0[f(y_1, \dots, y_m)] \big\} \xrightarrow{P_0-weakly} N_p(\mathbf{0}, \Sigma). \tag{5}$$

as n tends to infinity where $N_p(\mathbf{0}, \Sigma)$ denotes p-dimensional normal distribution with a mean vector $\mathbf{0}$ and a covariance matrix Σ.

See, e.g., Durrett [6] for proofs.

4.2 Construction of moment estimators

The limit in (4) for each f is a function of the true parameter θ_0. Then, in terms of Theorem 2 and Theorem 3, we can construct strong consistent and asymptotically normal estimating functions based on $\{y_j : j = 1, \dots, n\}$ according to the moment method.

Define two functions $\mathbf{P}_n f : \mathbf{R}^{nd_2} \to \mathbf{R}^p$ and $H : \Theta \to \mathbf{R}^p$ as follows:

$$
\begin{aligned}
(\mathbf{P}_n f)(\cdot, \dots, \cdot) &= \frac{1}{n} \sum_{j=1}^{n-m+1} f(\cdot, \dots, \cdot), & (6)\\
H(\theta) &= E_\theta[f(y_1, \dots, y_m)]. & (7)
\end{aligned}
$$

Henceforth we omit $(y_1, \dots, y_n)$ of a random variable $(\mathbf{P}_n f)(y_1, \dots, y_n)$.

For a large n, moment estimator is defined as a solution of

$$
\mathbf{P}_n f = H(\hat{\theta}_n) \tag{8}
$$

When one solve a system of equations (8) with respect to each component of $\hat{\theta}_n$, there is a possibility such that several solutions exist per each component. Thus, in general, it is inevitable to be accompanied by non-identifiability. But there are cases where it does not emerge or affect. We will consider such concrete examples later.

Apart from non-identifiability, here we concentrate on the case where $\hat{\theta}_n$ is determined uniquely. In this case, of course, we need some additional regularity conditions. If H is one-to-one on some domain $\overline{\Theta} \subseteq \Theta$ such that $\theta_0 \in \overline{\Theta}$, then $\hat{\theta}_n$ is uniquely determined by (8), that is, $\hat{\theta}_n = H^{-1}(\mathbf{P}_n f)$.

Theorem 4. *Let A1∼A4 hold, $\overline{\Theta} \subseteq \Theta$ be open and $\theta_0 \in \overline{\Theta}$. Furthermore, assume that H satisfies followings.*

1. *H is one-to-one on $\overline{\Theta}$.*
2. *H is continuously differentiable at $\theta_0 \in \overline{\Theta}$.*
3. *H has a nonsingular derivative at θ_0, that is, $\det\big(\frac{d}{d\theta} H(\theta)\big)\big|_{\theta=\theta_0} \neq 0$ where, for a matrix A, $\det(A)$ denotes the determinant of A.*

Then, a moment estimator $\hat{\theta}_n$ uniquely exists with a probability tending to 1, and for each θ_0,

$$
\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{P_0-weakly} N_p\big(\mathbf{0}, (H'(\theta_0)^{-1})\Sigma(H'(\theta_0)^{-1})^T\big) \tag{9}
$$

as n tends to infinity, where $'$ and T denotes a differentiation with respect to θ and matrix transposition, respectively.

By Theorem 4, if A1∼A4 holds true and H is invertible, then $\mathbf{P}_n \varphi$ is a consistent and asymptotically normal estimator for $H(\theta_0)$.

5 Examples of application

In this section, several concrete examples are presented for concrete study [1]. All of stochastic processes are assumed to be one-dimensional.

5.1 Random trend models with a hidden Gaussian process

Here we consider two continuous-time linear state space model. Two different cases about $L^{(1)}$ and $L^{(2)}$ are presented.

[1] The results of numerical experiments will be given in the presentation.

5.1.1 Jump type observation noise

Consider the model

$$\begin{cases} dX_t &= p(q - X_t)dt + \sqrt{r}dW_t, \\ dY_t &= X_t dt + dL_t \end{cases} \tag{10}$$

where W is an Wiener process and L is a one-dimensional normal inverse Gaussian Lévy motion (NIGLM) which is independent of W. In this example, $X = (X_t)_{t \in \mathbf{R}_+}$ expresses a trend varying along time, and $Y = (Y_t)_{t \in \mathbf{R}_+}$ does a noisy observation. Normal inverse Gaussian distributions forms a very flexible class containing, for example, normal and scaled-Cauchy as a limiting case with respect to parameters. In general, a NIGLM depends on four parameters $(\alpha, \beta, \delta, \mu)$ where α, β, δ and μ has specified meanings, namely, steepness, degree of asymmetry, scale and location, respectively. These parameters satisfy $\alpha \geq 0$, $\alpha^2 \geq \beta^2$, $\delta > 0$ and $\mu \in \mathbf{R}$. See Eberlein [7], Barndorff-Nielsen [1] and their references for more analytic fact about NIGLMs. Here we assume $L = (L_t)_{t \in \mathbf{R}_+}$ is symmetric and centered, that is, $\beta = \mu = 0$ so that $L_1 \sim NIG(\alpha, 0, \delta, 0)$ where $\alpha, \delta > 0$ and $NIG(\alpha, 0, \delta, 0)$ denotes the normal inverse Gaussian distribution with four parameters $(\alpha, 0, \delta, 0)$. In this case, L is a purely discontinuous local martingale (see Jacod and Shiryaev [13] for this terminology). Furthermore we assume α is a known for simplicity. Then the unknown parameter is $\theta = (p, q, r, \delta)$ where all of p, q, r and δ are positive.

A hidden Gaussian Ornstein-Uhlenbeck process X is strictly stationary with stationary distribution $N_1\big(q, r/(2p)\big)$ and mixing with exponentially decreasing α-mixing coefficient. Thus all of A1∼A4 are satisfied if X_0 is distributed as $N\big(q, r/(2p)\big)$ and independent of W and L (Y_0 may be any random variables which satisfies A1). After tedious calculation, we derive the corresponding estimators for θ based on polynomial type estimating functions $f(y_j, y_{j+1}, y_{j+2}) = (y_j, y_j^2, y_j y_{j+1}, y_j y_{j+2})$ as

$$\begin{aligned} \hat{q}_n &= \frac{1}{n\Delta}\sum_{j=1}^n y_j, \\ \hat{p}_n &= \frac{1}{\Delta}\log\Big(\frac{\mathcal{C}_{1,2}}{\mathcal{C}_{1,3}}\Big), \\ \hat{r}_n &= \frac{2\{\mathcal{C}_{1,2}\log(\mathcal{C}_{1,2}/\mathcal{C}_{1,3})\}^3}{\Delta(\mathcal{C}_{1,2} - \mathcal{C}_{1,3})^2}, \\ \hat{\delta}_n &= \frac{\alpha}{\Delta}\left[\frac{1}{n}\sum_{j=1}^n y_j^2 - \Big(\frac{1}{n}\sum_{j=1}^n y_j\Big)^2 - \frac{2\mathcal{C}_{1,2}^3}{(\mathcal{C}_{1,2} - \mathcal{C}_{1,3})^2}\Big\{\log\Big(\frac{\mathcal{C}_{1,2}}{\mathcal{C}_{1,3}}\Big) - 1 + \frac{\mathcal{C}_{1,3}}{\mathcal{C}_{1,2}}\Big\}\right] \end{aligned}$$

where, for $k = 1, 2$,

$$\mathcal{C}_{1,k+1} = \frac{1}{n}\sum_{j=1}^{n-k} y_j y_{j+k} - \Big(\frac{1}{n}\sum_{j=1}^n y_j\Big)^2.$$

To estimate α simultaneously using polynomial type estimating functions, we shall use a further estimator $\frac{1}{n}\sum_{j=1}^n y_j^4$ and compute its limit

$$E_0[y_1^4] = E_0\left[\left(\int_0^\Delta X_t dt + L_\Delta\right)^4\right]$$

Although this is relatively complicated, it is also possible to derive the explicit estimators for $\theta = (p, q, r, \alpha, \delta)$.

For simulation purpose, let us consider a reduced model from (10) described as

$$\begin{cases} dX_t &= -X_t dt + \sqrt{r}dW_t, \quad X_0 = \eta, \\ dY_t &= X_t dt + dL_t, \quad Y_0 = 0 \end{cases} \tag{11}$$

Δ	obs.	mean of $\hat{r}_n$	s.d. of $\hat{r}_n$	mean of $\hat{\alpha}_n$	s.d. of $\hat{\alpha}_n$
2.0	50	-	-	-	-
	100	1.4200	1.0744	1.1472	16.6568
	300	1.4236	0.3065	1.0406	0.2106
	500	1.4803	0.2836	1.0786	0.1817
1.0	50	-	-	-	-
	100	1.4017	0.7899	1.0802	0.2012
	300	1.4503	0.2645	1.0200	0.0468
	500	1.4851	0.2068	1.0257	0.0378
0.5	50	1.3372	2.0152	1.1748	0.4589
	100	1.3355	1.0415	1.1223	0.1843
	300	1.4814	0.3054	1.0576	0.3054
	500	1.4835	0.2070	1.0110	0.0296

Table 1: Jump type measurement noise case : Teh values of parameters are $(r_0, \alpha_0) = (1.5, 1.0)$, $\delta = 1.0$ and $X_0 = Y_0 = 0.0$. The symbol $-$ in this table means discarded cases, that is, the cases where estimated value is negative.

where $L_1 \sim NIG(\alpha, 0, \delta, 0)$ with known δ and unknown α. The estimators for (r, α) are as follows.

$$
\begin{aligned}
\hat{r}_n &= \frac{2}{(1-e^{-\Delta})^2}\left(\frac{1}{n-1}\sum_{j=1}^{n-1} y_j y_{j+1}\right) \\
\hat{\alpha}_n &= \frac{\delta\Delta}{\frac{1}{n}\sum_{j=1}^{n} y_j^2 - \hat{r}_n D}.
\end{aligned}
$$

Table 1 shows the corresponding simulation result.

5.1.2 Two correlated Wiener processes

Next let us consider a case where $L^{(1)}$ and $L^{(2)}$ are correlated Wiener processes. The model is the following.

$$
\begin{cases} dX_t &= (q - X_t)dt + dW_t, \\ dY_t &= X_t dt + \sqrt{\sigma} dB_t \end{cases} \tag{12}
$$

where two Wiener processes W and B have a known correlated coefficient ρ and the unknown parameter is $\theta = (q, \sigma) \in \mathbf{R} \times (0, \infty)$. In this case, X is a Gaussian of the form

$$
X_t = X_0 e^{-t} + q(1 - e^{-t}) + \int_0^t e^{-(t-s)} dW_s
$$

with a stationary distribution $N(q, 1/2)$. Corresponding polynomial type estimating equations are as follows.

$$
\begin{aligned}
\frac{1}{n}\sum_{j=1}^{n} y_j &= \hat{q}_n \Delta \\
\frac{1}{n}\sum_{j=1}^{n} y_j^2 &= \hat{q}_n^2 \Delta^2 + \hat{\sigma}_n \Delta + (\Delta + e^{-\Delta} - 1)(1 + 2\rho\sqrt{\hat{\sigma}_n})
\end{aligned}
$$

Δ	obs.	mean of $\hat{q}_n$	s.d. of $\hat{q}_n$	mean of $\hat{\sigma}_n$	s.d. of $\hat{\sigma}_n$
2.0	50	0.0060	0.0386	1.8550	0.2968
	100	0.0034	0.0209	1.9637	0.1398
	200	0.0236	0.0109	1.9884	0.0967
1.0	50	0.0540	0.1084	1.9567	0.2750
	100	0.0318	0.0421	1.9340	0.1446
	200	-0.0096	0.0198	1.9864	0.0749
0.5	50	-0.0803	0.1704	1.9276	0.2279
	100	0.0235	0.0844	1.9236	0.1164
	200	0.0124	0.0393	1.9880	0.0552

Table 2: The two correlated Wiener case : values of parameters are $(q_0, \sigma_0) = (0.0, 2.0)$, $\rho = 0.5$ and $X_0 = Y_0 = 0.0$.

Note that, although a $\hat{\sigma}_n$ has two solutions, it is uniquely determined by the restriction $\sigma > 0$. Then we can derive estimators for q and σ as

$$\hat{q}_n = \frac{1}{n\Delta}\sum_{j=1}^{n} y_j \tag{13}$$

$$\hat{\sigma}_n = \frac{1}{\Delta^2}\left\{-\rho D + \sqrt{\rho^2 D - \Delta D - \hat{q}_n^2 \Delta^3 + \frac{\Delta}{n}\sum_{j=1}^{n} y_j^2}\right\}^2 \tag{14}$$

where $D := \Delta + e^{-\Delta} - 1$. It follows that moment estimators are affected by a correlation among $L^{(1)}$ and $L^{(2)}$ in this case. Table 2 shows the simulation result for this model.

More generally, we can construct estimators while reserving unknown parameters contained in L in the following model

$$\begin{cases} dX_t &= V_0(X_t, \theta)dt + V(X_t, \theta)dW_t \\ dY_t &= S(X_t, \theta)dt + dL_t \end{cases} \tag{15}$$

where all of X, Y, W and L are one-dimensional, W is an Wiener process, L is a Lévy process whose distribution contains unknown parameters and W and L are independent. Namely, we can regard the parameters contained in L as nuisance parameters. Of course, $y = (y_j)_{j \in \mathbf{Z}_+^*}$ is of (1). Define two functions $\mathcal{K} : \mathbf{R}^l \to \mathbf{R}$ and $\mathcal{K}^* : \mathbf{R}^l \to \mathbf{R}$ as

$$\mathcal{K}(u_1, \dots, u_l) = i^{-l}\frac{\partial^l}{\partial u_1 \cdots \partial u_l}\Big\{\log E_0\Big[\exp\Big(i\sum_{j=1}^{s}(u_{1+l_{j-1}} + \cdots + u_{l_j})y_{t_j}\Big)\Big]\Big\} \tag{16}$$

and

$$\mathcal{K}^*(u_1, \dots, u_l) = i^{-l}\frac{\partial^l}{\partial u_1 \cdots \partial u_l}\log E_0\Big[\exp\Big(i\sum_{j=1}^{s}(u_{1+l_{j-1}} + \cdots + u_{l_j})\int_{(t_j-1)\Delta}^{t_j\Delta} S_0(X_t, \theta)dt\Big)\Big] \tag{17}$$

where $\sum_{j=1}^{s} l_j = l$, $l_0 = 0$ and l_j, $1 \le j \le s$, are positive integers. Then, we have the following statement.

Lemma 1 (Masuda[17], submitted).
If the model is of the form (15), then we have $\mathcal{K}(u_1, \dots, u_l) = \mathcal{K}^*(u_1, \dots, u_l)$ for any $(u_1, \dots, u_l) \in \mathbf{R}^l$.

Based on Lemma 1, we can do systematic partial estimation for the model (15) in the sense stated above by considering $\mathcal{K}(0, \dots, 0)$. For example, in the case of $l = 2, 3$, $l_j = 1$ for all $j \ge 1$, and $t_j = j$, we have

$$\mathcal{K}(0, 0) = E_0[y_1 y_2] - E_0[y_1]E_0[y_2]$$

and

$$\mathcal{K}(0,0,0) = E_0[y_1y_2y_3] - 2E_0[y_1y_2]E_0[y_1] - E_0[y_1y_3]E_0[y_1] + 2(E_0[y_1])^3.$$

Recalling original empirical convergent sequences, it follows that

$$\frac{1}{n}\sum_{j=1}^{n-1} y_jy_{j+1} - \Big(\frac{1}{n}\sum_{j=1}^{n} y_j^2\Big)^2$$

and

$$\frac{1}{n}\sum_{j=1}^{n-2} y_jy_{j+1}y_{j+2} - \Big\{2\Big(\frac{1}{n}\sum_{j=1}^{n-1} y_jy_{j+1}\Big) + \Big(\frac{1}{n}\sum_{j=1}^{n-2} y_jy_{j+2}\Big)\Big\}\Big(\frac{1}{n}\sum_{j=1}^{n} y_j\Big) + 2\Big(\frac{1}{n}\sum_{j=1}^{n} y_j\Big)^3$$

are estimators for

$$E_0\Big[\Big(\int_0^{\Delta} S_0(X_t,\theta)dt\Big)\Big(\int_{\Delta}^{2\Delta} S_0(X_t,\theta)dt\Big)\Big] - (\Delta E_0[S_0(X_0,\theta)])^2$$

and

$$\begin{aligned}
&E_0\Big[\Big(\int_0^{\Delta} S_0(X_t,\theta)dt\Big)\Big(\int_{\Delta}^{2\Delta} S_0(X_t,\theta)dt\Big)\Big(\int_{2\Delta}^{3\Delta} S_0(X_t,\theta)dt\Big)\Big] \\
&\quad - 2\Delta E_0[S_0(X_0,\theta)]E_0\Big[\Big(\int_0^{\Delta} S_0(X_t,\theta)dt\Big)\Big(\int_{\Delta}^{2\Delta} S_0(X_t,\theta)dt\Big)\Big] \\
&\quad - \Delta E_0[S_0(X_0,\theta)]E_0\Big[\Big(\int_0^{\Delta} S_0(X_t,\theta)dt\Big)\Big(\int_{2\Delta}^{3\Delta} S_0(X_t,\theta)dt\Big)\Big] + 2(\Delta E_0[S_0(X_0,\theta)])^3
\end{aligned}$$

respectively.

5.2 Stochastic volatility models

Here we consider an example in which, when $L^{(1)}$ and $L^{(2)}$ are correlated, explicit expressions for estimating functions are not available while the ergodicity of X is theoretically inherited.

GJL [10] gave an example of a simple stochastic volatility model

$$\begin{cases} dX_t &= p(q - X_t)dt + \sqrt{rX_t}dW_t, \quad X_0 = \eta, \\ dY_t &= \sqrt{X_t}dB_t, \quad Y_0 = 0. \end{cases} \tag{18}$$

where $W = (W_t)_{t\in\mathbf{R}_+}$ and $B = (B_t)_{t\in\mathbf{R}_+}$ are independent Wiener processes, and η is an almost sure positive random variable whose distribution (under P_0) is the stationary one of one-dimensional hidden diffusion X. $X = (X_t)_{t\in\mathbf{R}_+}$ and $Y = (Y_t)_{t\in\mathbf{R}_+}$ are a squared volatility process and a log-asset price process, respectively. The unknown parameter is $\theta = (p, q, r)$ with $2pq \geq r > 0$ so that X_t stays positive almost surely. In this example, X has an α-mixing coefficient which decays exponentially fast, and a stationary distribution $\Gamma(2pq/r, 2p/r)$. Moreover, X has a unique strong solution so that the measurability condition for X in A3 is satisfied by the definition of strong solutions of stochastic differential equations. GJL adopted

$$f(y_j, y_{j+1}, y_{j+2}) = (y_j^2, y_j^2y_{j+1}^2, y_j^4) \tag{19}$$

for the estimating function $f = (f_1, f_2, f_3)$. Proposition 3.1 of GJL [10] allows us to compute $E_0[y_1^2]$, $E_0[y_1^2y_2^2]$ and $E_0[y_1^4]$ explicitly. After that, strongly consistent and asymptotically normal estimators based on $\{y_j; j = 1, 2, \ldots, n\}$ are derived by (possibly numerically) solving three estimating equations associated with (19).

Next, using the model (18), let us consider a case where W and B are correlated as $B_t = \rho W_t + \sqrt{1-\rho^2}\tilde{W}_t$ where ρ is a correlation coefficient and $\tilde{W}$ is a Wiener process which is independent of

W. Although our argument can be theoretically applied to this case, explicit expression for some of corresponding estimating equations are not valid apart from the independent case, according to correlation between W and B. For example, we have

$$E_0[y_1^4] = 3\left\{q^2\Delta^2 + \frac{2Var_0[X_0]}{p^2}(e^{-p\Delta} + p\Delta - 1)\right\} + 12\rho^2 \int_0^\Delta E_0\left[X_s\left(\int_0^s M_u dM_u\right)\right] ds \tag{20}$$

where $M_u := \int_0^u \sqrt{X_v} dW_v$. The explicit expression of second term of (20) in terms of θ and data seems impossible by straightforward way.

However, it is possible according to the Burkholder-Gundy-Davis's inequality to give the useful criterion for the first condition in A4 with the same spirit to Proposition 2.3 in GJL even in the case where W and B is correlated.

Lemma 2. *Assume (18) for the structure of the model. Suppose that there exists positive finite constants C and r with $r \geq 1$ such that*

$$|f(y_1, y_2, \dots, y_m)| \leq C\left(1 + \sum_{j=1}^m |y_j|^r\right). \tag{21}$$

Then, if it holds that

$$E_0[X_0^{r(1+\delta/2)}] < \infty \tag{22}$$

for some $\delta > 0$, then (21) holds true for that δ.

References

[1] Barndorff-Nielsen, O.E. (1998): Processes of Normal inverse Gaussian type. *Finance and Stochastics.* **2**, 41-68.

[2] Barndorff-Nielsen, O.E. and Shephard, N. (2001): Non-Gaussian OU based models and some of their uses in financial economics (with discussion). *Journal of the Royal Statistical Society, Series B 63*, 167-241.

[3] Bibby, M. and Sørensen, M. (1995): Martingale estimation functions for discretely observed diffusion processes. *Bernoulli*, **1**, (1), 17-39.

[4] Bickel, P.J., Ritov, Y. and Rydén, T. (1998): Asymptotic normality of the maximum-likelihood estimator for general hidden Markov models. *Ann. Statist.* **26** , no.4, 1614-1635.

[5] Doukhan, P. (1994): *Mixing. Properties and examples.* Lecture Notes in Statistics, 85. Springer-Verlag, New York.

[6] Durrett, R. (1996): *Probability: theory and examples. Second edition.* Duxbury Press.

[7] Eberlein, E. (1999): Application of generalized hyperbolic Lévy motions to finance. in " *Levy Processes: Theory and Applications* ", O.E. Barndorff-Nielsen, T. Mikosch, and S. Resnick (Eds.), Birkhauser, 319-336.

[8] Genon-Catalot, V., Jeantheau, T. and Laredo, C. (1998): Limit theorems for discretely observed stochastic volatility models. *Bernoulli*, **4**, no.3, 283-303.

[9] Genon-Catalot, V., Jeantheau, T. and Laredo, C. (1999): Parameter estimation for discretely observed stochastic volatility models. *Bernoulli*, **5**, no.5, 855-872.

[10] Genon-Catalot, V., Jeantheau, T. and Laredo, C.(2000): Stochastic volatility madels as hidden Markov models and statistical applications. *Bernoulli.* **6**, 1051-1079.

[11] Hall, P. and Heyde, C.C. (1980): *Martingale limit theory and its application.* Probability and Mathematical Statistics. Academic Press, New York-London.

[12] Ikeda, N. and Watanabe, S. (1989): *Stochastic differential equations and diffusion processes. Second edition.* North-Holland Mathematical Library, 24.

[13] Jacod, J, and Shiryaev, A.N. (1987): *Limit theorems for stochastic processes.* Springer-Verlag, Berlin-New York.

[14] Jensen, J.L. and Petersen, N.V. (1999): Asymptotic normality of the maximum likelihood estimator in state space models. *Ann. Statist.* **27** , no.2, 514-535.

[15] Kusuoka, S. and Yoshida, N. (2000): Malliavin calculus, geometric mixing, and expansion of diffusion functionals. *Probab. Theory Related Fields* **116**, no.4, 457-484.

[16] Masuda, H. (2001): On parameter estimation in a hidden Markov model. *Master Thesis. Graduate School of Mathematical Science, University of Tokyo.*

[17] Masuda, H. (2001a): On moment estimation in a class of partially observed state space models. (Submitted)

[18] Masuda, H. (2001b): On some properties of multidimensional Ornstein-Uhlenbeck type processes. (Submitted)

[19] Shephard, N. and Pitt, M.K. (1997): Likelihood analysis of non-Gaussian measurement time series. *Biometrika.* **84**. no.3, 653-667.

[20] van der Vaart, A.W. (1998): *Asymptotic statistics.* Cambridge Series in Statistical and Probabilistic Mathematics.