Asymptotic properties of the penalized quasi-likelihood estimators and their applications

Yoshiki Kinoshita

Graduate School of Mathematical Sciences, University of Tokyo [1]

[1] Graduate School of Mathematical Sciences, University of Tokyo: 3-8-1 Komaba, Meguro-ku, Tokyo 153-8914, Japan.

Acknowledgement

The author wishes to express his sincere thanks to his supervisor Professor Nakahiro Yoshida for giving many valuable advices.

Contents

Chapter 1

Introducition

1.1 Quasi-likelihood analysis

Stochastic processes are useful to describe phenomena which depend on time. Ibragimov and Has'minskii (1981) developed the scheme to verify the asymptotic properties of maximum likelihood estimators. They reduced the asymptotic properties of the maximum likelihood estimator to the convergence of the random field corresponding to likelihood ratios. They also studied large deviation inequality which provides useful asymptotic properties of the maximum likelihood estimator including moment convergence.

While the likelihood estimator has good asymptotic properties, it is hard to calculate it in practice, since we can't often gain the complete data. Therefore we consider the quasi-(log) likelihood function $\mathbb{H}_T$ alternatively. Here, $T = \mathbb{N}$ or $\mathbb{R}_+$ and $\mathbb{H}_T : \Omega \times \Theta \to \mathbb{R}$ is some sequence of random fields where $\Theta \in \mathbb{R}^{\mathsf{p}}$ is a parameter space and $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$ is a stochastic basis. Even in this case, it is useful to consider the quasi-likelihood ratio

$$\mathbb{Z}_T = \exp(\mathbb{H}_T(\theta^* + a_T u) - \mathbb{H}_T(\theta^*))$$

where $a_T \in \mathrm{GL}(\mathsf{p})$ is a deterministic sequence in the general linear group over $\mathbb{R}$ of degree p.

Yoshida (2011) studied $\mathbb{Z}_T$ and proposed the polynomial type large deviation inequality (PLDI)

$$P\Big[\sup_{u;|u|\geq r} \mathbb{Z}_T(u) \geq r^{-N}\Big] \leq \frac{C_N}{r^N} \qquad (r > 0).$$

This inequality plays a crucial role in quasi likelihood analysis (QLA) because it implies the uniform boundedness

$$\sup_{T>0} E[|a_T^{-1}(\hat{\theta}_T - \theta^*)|^m] < \infty$$

and moment convergence

$$E[|a_T^{-1}(\hat{\theta}_T - \theta^*)|^{m'}] \to E[|\hat{u}_\infty|^{m'}]$$

for some large $m > 0$, every $m' \in (0, m)$ and a random variable $\hat{u}_\infty$ such that $a_T^{-1}(\hat{\theta}_T - \theta^*) \xrightarrow{d} \hat{u}_\infty$. These properties are useful to investigate an asymptotic behavior of statistics which depends on the moment of $a_T^{-1}(\hat{\theta}_T - \theta^*)$; see e.g. Chan and Ing (2011), Shimizu (2017), Suzuki and Yoshida (2018) and Umezu et al. (2019).

Yoshida (2011) also provided the tractable conditions to derive PLDI under the locally asymptotically quadratic (LAQ) setting (Yoshida 2011). Actually, PLDI is obtained with LAQ on many kinds of models, e.g. Clinet and Yoshida (2017), Masuda (2013), Ogihara and Yoshida (2014) and Uchida and Yoshida (2013).

1.2 Variable selection via regularization methods

Regularization methods, that impose a penalty term on a loss function, provide a tool for variable selection. The method is useful because it performs estimation and variable selection simultaneously. Penalized estimators are generally expressed in the following form

$$\hat{\theta}_{\text{penalty}} \in \underset{\theta \in \overline{\Theta}}{\operatorname{argmin}} \big\{ -L_n(\theta) + p(\theta) \big\},$$

where L_n is a log likelihood function or $-L_n$ is equal to the sum of squared residuals and p is a penalty term. One of the most simple regularization methods is the Bridge (Frank and Friedman 1993) that imposes the penalty term

$$p_\lambda^{\text{Bridge}}(\theta) = \lambda \sum_{i=1}^{\mathsf{p}} |\theta_i|^q$$

on the least square loss function, where $q > 0$ is a constant, p is a dimension of an unknown parameter θ and $\lambda > 0$ is a tuning parameter. For $q \leq 1$, the estimator performs variable selection. Especially, when $q = 1$, the estimator is called the Lasso (Tibshirani 1996). Other than Bridge, various regularization methods have been proposed, e.g. the smoothly clipped absolute deviation (SCAD; Fan and Li 2001) and the minimax concave penalty (MCP; Zhang 2010). These methods are widely studied and extended in the regression analysis. Knight and Fu (2000) derived a $\sqrt{n}$-consistency of the Bridge estimator $\hat{\theta}_{\text{Bridge}}$ and studied the limit distribution of $\sqrt{n}(\hat{\theta}_{\text{Bridge}} - \theta^*)$ where θ^* is the true value of θ. Zou (2006) proposed the adaptive Lasso and derived its oracle property. These results clarified the advantage of Bridge estimator with $q < 1$ and adaptive Lasso estimator compared to Lasso estimator $\hat{\theta}_{\text{Lasso}}$ in the sense of asymptotic efficiency because the limit distribution of $\sqrt{n}(\hat{\theta}_{\text{Lasso}} - \theta^*)$ has a redundant term.

Applications of regularization methods to the quasi-likelihood analysis (QLA) for stochastic models have been recently studied. The penalized quasi-maximum likelihood estimator is defined by

$$\hat{\theta}_T \in \underset{\theta \in \overline{\Theta}}{\operatorname{argmin}} \big\{ -\mathbb{H}_T(\theta) + p(\theta) \big\} \tag{1.1}$$

for a given quasi-likelihood function $\mathbb{H}_T$ in these situations. These approaches works well for various kinds of quasi-likelihood functions; see e.g. Belomestny and Trabs (2018), De Gregorio and Iacus (2012) and Gaïffas and Matulewicz (2019). Masuda and Shimizu (2017) studied the moment convergence of the Lasso estimator under more general settings. They derived the polynomial type large deviation inequality (PLDI) for the L^1-penalized contrast functions under suitable conditions.

1.3 Organization of this thesis

In this thesis, we consider the quasi-likelihood function $\mathbb{H}_T$ with LAQ property and PLDI, and we study the penalized maximum likelihood estimator defined in (1.1). Our penalty term can deal with many kinds of penalties including the Lasso, the Bridge and the adaptive Lasso. The objective in this thesis consists of two parts. One is to derive a polynomial type large deviation inequality for the penalized quasi-likelihood random field and another is to study asymptotic behavior of the penalized quasi-maximum likelihood estimators. The main results in Chapter 2 and Chapter 3 are based on the work in Kinoshita and Yoshida (2019).

The rest part of this thesis is organized as follows. In Chapter 2, we derive asymptotic properties of the penalized quasi-likelihood estimator. The PLDI is given in Section 2.2 and limit theorem is given in Section 2.4. We will apply the results in Chapter 2 to Itó Process in Chapter 3. In order to treat jumps, we consider the global filter in Chapter 4.

Chapter 2

Penalized quasi-likelihood estimator

2.1 Penalized quasi-likelihood estimator

Let Θ be a bounded open set in $\mathbb{R}^{\mathsf{p}}$. We denote by $\theta^* \in \Theta$ the true value of an unknown parameter $\theta \in \Theta$. Given a probability space $(\Omega, \mathcal{F}, P)$, we consider a sequence of random fields $\mathbb{H}_T : \Omega \times \overline{\Theta} \to \mathbb{R}$, $T \in \mathbb{T}$, where $\mathbb{T}$ is a subset of $\mathbb{R}_{\geq 0}$ with $\sup \mathbb{T} = \infty$ and $\overline{\Theta}$ is a closure of Θ. We assume $\mathbb{H}_T(\theta)$ is continuous for all $\omega \in \Omega$, where $\mathbb{H}_T(\theta)$ denotes the mapping $\overline{\Theta} \ni \theta \to \mathbb{H}_T(\theta, \omega)$ for each $\omega \in \Omega$. We call $\mathbb{H}_T(\theta)$ a quasi-likelihood function and define the quasi-maximum likelihood estimator (QMLE) $\hat{\theta}_T^{\mathrm{QMLE}}$ by

$$\hat{\theta}_T^{\mathrm{QMLE}} \in \operatorname*{argmax}_{\theta \in \overline{\Theta}} \mathbb{H}_T(\theta).$$

Here we use this expression in the sense that $\hat{\theta}_T^{\mathrm{QMLE}} : \Omega \to \overline{\Theta}$ is a measurable mapping satisfying

$$\mathbb{H}_T(\hat{\theta}_T^{\mathrm{QMLE}}) = \max_{\theta \in \overline{\Theta}} \mathbb{H}_T(\theta)$$

for all $\omega \in \Omega$.

If θ^* has sparsity, we can construct an estimator which performs parameter estimation and variable selection simultaneously by adding a penalty term to the quasi-likelihood function. Let us consider the penalized quasi-likelihood function

$$\mathbb{H}_T^{\dagger}(\theta) = \mathbb{H}_T(\theta) - p_T(\theta) \tag{2.1}$$

and the penalized estimator

$$\hat{\theta}_T \in \operatorname*{argmax}_{\theta \in \overline{\Theta}} \mathbb{H}_T^{\dagger}(\theta),$$

where $p_T : \overline{\Theta} \to \mathbb{R}_{\geq 0}$ is a penalty function for every $T \in \mathbb{T}$. In this thesis, we assume that p_T has the following expression

$$p_T(\theta) = \sum_{j=1}^{\mathsf{p}} \xi_T^j p(\theta_j), \tag{2.2}$$

where ξ_T^j are (possibly random) positive sequences and $p : \mathbb{R} \to \mathbb{R}_{\geq 0}$ is a function satisfying $p(0) = 0$. Indeed, this function is defined on $\mathbb{R}^{\mathsf{p}}$, but we consider the restriction to $\overline{\Theta}$.

In the following sections, we will denote $\{j; \theta_j^* = 0\}$ and $\{j; \theta_j^* \neq 0\}$ by $\mathcal{J}^{(0)}$ and $\mathcal{J}^{(1)}$, respectively. Furthermore, for a vector $x \in \mathbb{R}^{\mathsf{p}}$ and a matrix $A \in \mathbb{R}^{\mathsf{p}\times\mathsf{p}}$, the vector $(x_j)_{j\in\mathcal{J}^{(k)}}$ and the matrix $(A_{ij})_{i\in\mathcal{J}^{(k)}, j\in\mathcal{J}^{(l)}}$ will be denoted by $x^{(k)}$ and $A^{(kl)}$, respectively, and we will express x as $(x^{(0)}, x^{(1)})$. We write $s(A, x) = Ax$, $s_j(A, x) = (Ax)_j$ and $s^{(k)}(A, x) = (Ax)^{(k)}$. For tensors $A = (A_{i_1,\dots,i_d})$ and $B = (B_{i_1,\dots,i_d})$, we denote $A[B] = \sum_{i_1,\dots,i_d} A_{i_1,\dots,i_d} B_{i_1,\dots,i_d}$. Moreover we write $A[u_1, \dots, u_d] = A[u_1 \otimes \cdots \otimes u_d] = \sum_{i_1,\dots,i_d} A_{1_1,\dots,i_d} u_1^{i_1} \cdots u_d^{i_d}$ for vectors $u_1 = (u_1^{i_1})_{i_1}, \dots, u_d = (u_d^{i_d})_{i_d}$. We denote by $u^{\otimes r} = u \otimes \cdots \otimes u$ the r times tensor product of u.

2.2 Polynomial type large deviation inequality

We make use of the quasi-likelihood analysis (QLA) of Yoshida (2011) to examine the moment convergence of estimators for θ and to derive a limit theorem of it. Let $a_T \in \mathrm{GL}(\mathsf{p})$ be a deterministic sequence satisfying $||a_T|| \to 0$ as $T \to \infty$ and $\mathbb{U}_T = \{u \in \mathbb{R}^{\mathsf{p}}; \theta^* + a_T u \in \overline{\Theta}\}$. Here $||A||$ denotes the square root of the maximum eigenvalue of $A'A$ for $A \in \mathbb{R}^{\mathsf{p}\times\mathsf{p}}$ and A' is the transpose of A. Based on QLA, we define the random fields $\mathbb{Z}_T$ and $\mathbb{Z}_T^\dagger$ on $\mathbb{U}_T$ by

$$\mathbb{Z}_T(u) = \exp\biggl(\mathbb{H}_T(\theta^* + a_T u) - \mathbb{H}_T(\theta^*)\biggr)$$

and

$$\mathbb{Z}_T^\dagger(u) = \exp\biggl(\mathbb{H}_T^\dagger(\theta^* + a_T u) - \mathbb{H}_T^\dagger(\theta^*)\biggr).$$

Let $L > 0$ and $V_T(r) = \{u \in \mathbb{U}_T; r \leq u\}$ for $r > 0$. We assume a polynomial type large deviation inequality (PLDI) in Yoshida (2011) for $\mathbb{Z}_T$.

[A1] There exist constants $C_L > 0$ and $\varepsilon_L \in (0, 1)$ such that

$$P\left[\sup_{u\in V_T(r)} \mathbb{Z}_T(u) \geq \exp(-r^{2-\varepsilon_L})\right] \leq \frac{C_L}{r^L} \tag{2.3}$$

for all $r > 0, T > 0$.

Here the supremum of the empty set should read $-\infty$ by convention. In this thesis we assume that a_T is a diagonal matrix, and write

$$a_T = \begin{pmatrix} \alpha_T^1 & & & 0 \\ & \alpha_T^2 & & \\ & & \ddots & \\ 0 & & & \alpha_T^{\mathsf{p}} \end{pmatrix}.$$

Moreover we suppose $\prod_{1 \le j \le \mathsf{p}} \alpha_T^j \neq 0$ for all $T \in \mathbb{T}$. Let c_0 be a positive constant. In order to estimate $\mathbb{Z}_T^\dagger$, we consider the following three conditions for the penalty term.

[A2] p is differentiable except the origin.

[A3] For some positive constant ε,

$$\sup_{-\varepsilon < x < \varepsilon} p(x) < \infty.$$

[A4] For all $j \in \mathcal{J}^{(1)}$,

$$\sup_{T \in \mathbb{T}} |\alpha_T^j \xi_T^j| \le c_0$$

almost surely.

Remark 2.2.1. Given ξ_T, we can construct ξ_T' satisfying [A4] by taking ξ_T' such that $\xi_T'^j = \min(\xi_T^j, (\alpha^j)_T^{-1} c_0)$.

Example 2.2.2 (LASSO)**.** Define ξ_T^j by $\xi_T^j = |\alpha_T^j|^{-1}$ and p by $p(x) = |x|$, then the penalty term $p_T(\theta) = \sum_{j=1}^{\mathsf{p}} |\alpha_T^j|^{-1} |\theta_j|$ satisfies [A2]-[A4].

In the above setting, we can derive the PLDI for $\mathbb{Z}_T^\dagger$.

Theorem 2.2.3. Given $L > 0$, assume Conditions [A1]-[A4]. Then there exist constants $C_L' > 0$ and $\varepsilon_L' \in (0,1)$ such that

$$P\left[\sup_{u \in V_T(r)} \mathbb{Z}_T^\dagger(u) \ge \exp(-r^{2-\varepsilon_L'})\right] \le \frac{C_L'}{r^L} \tag{2.4}$$

for all $r > 0, T > 0$.

Proof. By [A1], there exist constants $C_L > 0$ and $\varepsilon_L \in (0,1)$ satisfying (2.3) for all $r > 0, T > 0$. Let $\varepsilon'_L \in (\varepsilon_L, 1)$. For every $T > 0$ and $r > 0$, we have

$$
\begin{aligned}
&P\left[\sup_{u \in V_T(r)} \mathbb{Z}_T^{\dagger}(u) \geq \exp(-r^{2-\varepsilon'_L})\right] \\
&\quad \leq P\left[\sup_{u \in V_T(r)} \mathbb{Z}_T^{\dagger}(u) \exp\left\{\sum_{j \in \mathcal{J}^{(0)}} \xi_T^j p\big(s_j(a_T, u)\big)\right\} \geq \exp(-r^{2-\varepsilon'_L})\right] \\
&\quad \leq \sum_{n=0}^{\infty} P\left[\sup_{\substack{2^n r \leq |u| \leq 2^{n+1} r \\ u \in V_T(r)}} \mathbb{Z}_T(u) \exp(B_1) \geq \exp(-r^{2-\varepsilon'_L})\right],
\end{aligned}
$$

where

$$
B_1 = -\sum_{j \in \mathcal{J}^{(1)}} \xi_T^j \left[p\Big(\theta_j^* + s_j(a_T, u)\Big) - p(\theta_j^*)\right].
$$

Conditions [A2] and [A3] imply

$$
\sup_{x \in U \setminus \{0\}} \frac{p(\theta + x) - p(\theta)}{x} < \infty
$$

for every $\theta \in \mathbb{R} \setminus \{0\}$ and every compact set $U \subset \mathbb{R}$. Moreover, by definition of $\mathbb{U}_T$, we observe that $\sup_{T \in \mathbb{T}} \sup_{u \in \mathbb{U}_T} |a_T u| < \infty$. Therefore, from [A4], we have

$$
\begin{aligned}
|B_1| &\leq \sum_{j \in \mathcal{J}^{(1)}} \xi_T^j |s_j(a_T, u)| \left|\frac{p(\theta_j^* + s_j(a_T, u)) - p(\theta_j^*)}{s_j(a_T, u)}\right| \\
&\leq c_0 K |u|
\end{aligned}
$$

for some $K > 0$ which does not depend on T and r. Then we have

$$
\begin{aligned}
&\sum_{n=0}^{\infty} P\left[\sup_{\substack{2^n r \leq |u| \leq 2^{n+1} r \\ u \in V_T(r)}} \mathbb{Z}_T(u) \exp(B_1) \geq \exp(-r^{2-\varepsilon'_L})\right] \\
&\qquad \leq \sum_{n=0}^{\infty} P\left[\sup_{\substack{2^n r \leq |u| \leq 2^{n+1} r \\ u \in V_T(r)}} \mathbb{Z}_T(u) \geq \exp\left(-r^{2-\varepsilon'_L} - 2^{n+1} c_0 K r\right)\right].
\end{aligned}
$$

Since $1 < 2 - \varepsilon'_L < 2 - \varepsilon_L$, there exists a constant $R_1 > 0$ such that

$$
-r^{2-\varepsilon'_L} - 2^{n+1} c_0 K r \geq -(2^n r)^{2-\varepsilon_L}
$$

for all $n \in \mathbb{N}$ and $r \geq R_1$. By this inequality and [A1], we obtain

$$\sum_{n=0}^{\infty} P\left[\sup_{\substack{2^n r \leq |u| \leq 2^{n+1} r \\ u \in V_T(r)}} \mathbb{Z}_T(u) \geq \exp\left(-r^{2-\varepsilon'_L} - 2^{n+1} c_0 K r\right)\right]$$
$$\leq \sum_{n=0}^{\infty} P\left[\sup_{\substack{|u| \geq 2^n r \\ u \in V_T(r)}} \mathbb{Z}_T(u) \geq \exp\left(-(2^n r)^{2-\varepsilon_L}\right)\right]$$
$$\leq \sum_{n=0}^{\infty} \frac{C_L}{(2^n r)^L} = \frac{1}{r^L} \frac{2^L C_L}{2^L - 1}$$

for every $r \geq R_1$. Let $C'_L = \max\{R_1^L, \frac{2^L C_L}{2^L - 1}\}$, we complete the proof. □

Let $\hat{u}_T = a_T^{-1}(\hat{\theta}_T - \theta^*)$ then

$$\hat{u}_T \in \operatorname*{argmax}_{u \in \mathbb{U}_T} \mathbb{Z}_T^{\dagger}(u).$$

PLDI derives the L^m-boundedness of $\hat{u}_T$ (Proposition 1 of Yoshida (2011)).

Proposition 2.2.4. Let $L > m > 0$. Suppose that there exists a constant C_L such that

$$P\left[\sup_{u \in V_T(r)} \mathbb{Z}_T^{\dagger}(u) \geq 1\right] \leq \frac{C_L}{r^L}$$

for all $T > 0$ and $r > 0$. Then it holds that

$$\sup_{T>0} E[|\hat{u}_T|^m] < \infty. \tag{2.5}$$

In particular, $\hat{u}_T = O_p(1)$ as $T \to \infty$ (i.e., for every $\epsilon > 0$, there exist $\mathcal{T} \in \mathbb{T}$ and $M > 0$ such that $P(|\hat{u}_T| > M) < \epsilon$ for all $T \geq \mathcal{T}$), under Conditions [A1]-[A4].

2.3 Consistency of variable selection

In this section, we will derive the selection consistency of $\hat{\theta}_T$. Let $q \in (0, 1]$, we consider the conditions for p.

[A5] There exists $\lambda > 0$ such that

$$\lim_{x \to 0} \frac{p(x)}{|x|^q} = \lambda.$$

[A6] For every $j \in \mathcal{J}^{(0)}$,

$$(\xi_T^j)^{-\frac{1}{q}} |\alpha_T^j|^{-1} \xrightarrow{p} 0$$

as $T \to \infty$.

LASSO penalty in Example 2.2.2 derives PLDI, however, it does not satisfy [A6]. We give another example for [A6].

Example 2.3.1 (Bridge type)**.** Let $q < 1$ and $q' \in (q, 1]$. Define ξ_T^j by $\xi_T^j = |\alpha_T^j|^{-q'}$ and p by $p(x) = |x|^q$, then the penalty term $p(\theta) = \sum_{j=1}^{\mathsf{p}} |\alpha_T^j|^{-q'} |\theta_j|^q$ satisfies [A2]-[A6].

Let $\tilde{a}_T$ be a diagonal matrix in $\mathbb{R}^{\mathsf{p}\times\mathsf{p}}$ satisfying $(\tilde{a}_T)_{jj} = (\xi_T^j)^{-\frac{1}{q}}$ for $j \in \mathcal{J}^{(0)}$ and $(\tilde{a}_T)_{jj} = a_T^j$ for $j \in \mathcal{J}^{(1)}$. Denote $a_T^{-1}\tilde{a}_T$ by G_T.

[A7] For every $M > 0$,

$$\sup_{\substack{u,v \in \mathbb{U}_T \\ |u|,|v|<M \\ u \neq v}} \frac{|\mathbb{H}_T(\theta^* + a_T u) - \mathbb{H}_T(\theta^* + a_T v)|}{|u-v|^q} ||G_T^{(00)}||^q \xrightarrow{p} 0$$

as $T \to \infty$.

Remark 2.3.2. Condition [A6] implies that

$$||G_T^{(00)}|| \xrightarrow{p} 0 \tag{2.6}$$

as $T \to \infty$. We usually assume [A6] to ensure this convergence in this thesis.

Remark 2.3.3. Condition [A7] is a technical one, however, we can derive it easily from the differentiability of $\mathbb{H}_T$.

[A7′] For some $R > 0$, the following conditions hold:

- **(i)** For every $T \in \mathbb{T}$, $\mathbb{H}_T$ is almost surely thrice differentiable with respect to θ on $B = B_R(\theta^*, \Theta) = \{\theta \in \Theta; |\theta - \theta^*| < R\}$,
- **(ii)** $||a_T|| |\partial_\theta \mathbb{H}_T(\theta^*) = O_p(1)$,
- **(iii)** $||a_T||^2 \sup_{\theta \in B} |\partial_\theta^2 \mathbb{H}_T(\theta)| = O_p(1)$,
- **(iv)** $||a_T||^2 \sup_{\theta \in B} |\partial_\theta^3 \mathbb{H}_T(\theta)| = O_p(1)$.

Proposition 2.3.4. Assume [A6] and [A7′], then [A7] holds.

Proof. Take $R' < R$ satisfying $\overline{B_{R'}(\theta^*)} = \{\theta \in \mathbb{R}^{\mathsf{p}}; |\theta - \theta^*| \leq R'\} \subset \Theta$. For $M > 0$, there exists a constant $\mathcal{T}_M \in \mathbb{T}$ such that $\theta^* + a_T u \in B_{R'}(\theta^*)$ for every $T > \mathcal{T}_M$ and $u \in \mathbb{R}$ satisfying $|u| < M$. Therefore, by Taylor's theorem

$$\begin{aligned}
&\big|\mathbb{H}_T(\theta^* + a_T u) - \mathbb{H}_T(\theta^* + a_T v)\big| \\
&\quad \leq \big|\partial_\theta \mathbb{H}_T(\theta^*)[a_T u] - \partial_\theta \mathbb{H}_T(\theta^*)[a_T v]\big| \\
&\qquad + \bigg|\int_0^1 (1-s)\partial_\theta^2 \mathbb{H}_T(\theta^* + s a_T u)[(a_T u)^{\otimes 2}] ds \\
&\qquad\qquad - \int_0^1 (1-s)\partial_\theta^2 \mathbb{H}_T(\theta^* + s a_T v)[(a_T v)^{\otimes 2}] ds\bigg| \\
&\quad \leq A_1 + A_2 + A_3.
\end{aligned}$$

for every $T > \mathcal{T}_M$ and every $u, v \in \mathbb{R}$ satifying $|u|, |v| < M$, where

$$\begin{aligned}
A_1 &= ||a_T|| \cdot \big|\partial_\theta \mathbb{H}_T(\theta^*)\big| \cdot |u - v|, \\
A_2 &= \bigg|\int_0^1 (1-s)\partial_\theta^2 \mathbb{H}_T(\theta^* + s a_T u)[(a_T u)^{\otimes 2}] ds \\
&\qquad - \int_0^1 (1-s)\partial_\theta^2 \mathbb{H}_T(\theta^* + s a_T u)[(a_T v)^{\otimes 2}] ds\bigg|
\end{aligned}$$

and

$$\begin{aligned}
A_3 = &\bigg|\int_0^1 (1-s)\partial_\theta^2 \mathbb{H}_T(\theta^* + s a_T u)[(a_T v)^{\otimes 2}] ds \\
&\qquad - \int_0^1 (1-s)\partial_\theta^2 \mathbb{H}_T(\theta^* + s a_T v)[(a_T v)^{\otimes 2}] ds\bigg|.
\end{aligned}$$

However, [A7′](ii) and (2.6) implies

$$\sup_{\substack{u,v \in \mathbb{U}_T \\ |u|,|v| < M \\ u \neq v}} \frac{A_1}{|u - v|^q} ||G_T^{(00)}||^q \xrightarrow{p} 0 \tag{2.7}$$

as $T \to \infty$. If $\theta^* + a_T u \in B_{R'}(\theta^*)$, then

$$A_2 \leq ||a_T||^2 \cdot |u + v| \cdot |u - v| \cdot \sup_{\theta \in B} |\partial_\theta^2 \mathbb{H}_T(\theta)|.$$

Therfore [A7′](iii) and (2.6) implies

$$\sup_{\substack{u,v \in \mathbb{U}_T \\ |u|,|v| < M \\ u \neq v}} \frac{A_2}{|u - v|^q} ||G_T^{(00)}||^q \xrightarrow{p} 0 \tag{2.8}$$

as $T \to \infty$. From Taylor's theorem, we have

$$\partial_\theta^2 \mathbb{H}_T(\theta^* + sa_T u) = \int_0^1 \partial_\theta^3 \mathbb{H}_T\big(\theta^* + sa_T u + s'(sa_T u - sa_T v)\big) ds' [sa_T u - sa_T v]$$

for every $s \in [0,1]$, $T > \mathcal{T}_M$ and every $u, v \in \mathbb{R}$ satifying $|u|, |v| < M$. Since $\theta^* + sa_T u + s'(sa_T u - sa_T v) \in B_{R'}(\theta^*) \subset B$, it follows that

$$A_3 \le ||a_T||^3 \cdot |v|^2 \cdot |u - v| \cdot \sup_{\theta \in B} |\partial_\theta^3 \mathbb{H}_T(\theta)|$$

for every $T > \mathcal{T}_M$ and every $u, v \in \mathbb{R}$ satifying $|u|, |v| < M$. Therefore [A7′](iv) and (2.6) implies

$$\sup_{\substack{u,v \in \mathbb{U}_T \\ |u|,|v|<M \\ u \neq v}} \frac{A_3}{|u - v|^q} ||G_T^{(00)}||^q \xrightarrow{p} 0 \tag{2.9}$$

as $T \to \infty$. From (2.7), (2.8) and (2.9), we have the desired result. □

The following theorem ensures that $\hat{\theta}_T$ enjoys the consistency of variable selection.

Theorem 2.3.5. Assume Conditions [A5]-[A7]. If $\hat{u}_T = O_p(1)$, then

$$P\Big(\hat{\theta}_T^{(0)} = 0\Big) \to 1 \tag{2.10}$$

as $T \to \infty$.

Proof. Let $\mathcal{S}_T^1 = \{(0, \hat{u}_T^{(1)}) \in \mathbb{U}_T\}$. For $M > 0$ and $T \in \mathbb{T}$, define $\mathcal{S}_T^2$ by

$$\mathcal{S}_{T,M}^2 = \Big\{ |\hat{u}_T^{(0)}| < M, \sum_{j \in \mathcal{J}^{(0)}} \xi_T^j p\big(s_j(a_T, \hat{u}_T)\big) \ge \frac{\lambda}{2} |(G_T^{(00)})^{-1} \hat{u}_T^{(0)}|^q \Big\}$$

and define $\mathtt{C}_{T,M}$ by

$$\mathtt{C}_{T,M} = \sup_{\substack{u,v \in \mathbb{U}_T \\ |u|,|v|<M \\ u \neq v}} \frac{|\mathbb{H}_T(\theta^* + a_T u) - \mathbb{H}_T(\theta^* + a_T v)|}{|u - v|^q}.$$

By definition,

$$\mathbb{Z}_T^\dagger(\hat{u}_T^{(0)}, \hat{u}_T^{(1)}) - \mathbb{Z}_T^\dagger(0, \hat{u}_T^{(1)}) = \exp\Big(\mathbb{H}^\dagger\big(\theta^* + a_T \hat{u}_T\big) - \mathbb{H}^\dagger\big(\theta^* + a_T(0, \hat{u}_T^{(1)})\big)\Big)$$

$$
\begin{aligned}
&= \exp\Big(\mathbb{H}\big(\theta^* + a_T\hat{u}_T\big) - \mathbb{H}\big(\theta^* + a_T(0, \hat{u}_T^{(1)})\big)\Big) \\
&\quad \times \exp\Big(-\sum_{j\in\mathcal{J}^{(0)}} \xi_T^j p\big(s_j(a_T, \hat{u}_T)\big)\Big). \qquad (2.11)
\end{aligned}
$$

Therefore we have

$$
\begin{aligned}
P(\hat{\theta}_T^{(0)} \neq 0) &\leq P\Big(\mathbb{Z}_T^\dagger(\hat{u}_T^{(0)}, \hat{u}_T^{(1)}) \geq \mathbb{Z}_T^\dagger(0, \hat{u}_T^{(1)}), \hat{u}_T^{(0)} \neq 0, (0, \hat{u}_T^{(1)}) \in \mathbb{U}_T\Big) \\
&\quad + P((\mathcal{S}_T^1)^c) \\
&\leq P\Big(\mathtt{C}_{T,M}|\hat{u}_T^{(0)}|^q \geq \frac{\lambda}{2}|(G_T^{(00)})^{-1}\hat{u}_T^{(0)}|^q, \hat{u}_T^{(0)} \neq 0\Big) \\
&\quad + P((\mathcal{S}_T^1)^c) + P((\mathcal{S}_{T,M}^2)^c).
\end{aligned}
$$

Therefore it suffices to estimate the following three probabilities:

$$
\begin{aligned}
P_1 &:= P\Big(\mathtt{C}_{T,M}||G_T^{(00)}||^q \geq \frac{\lambda}{2}\Big), \\
P_2 &:= P((\mathcal{S}_T^1)^c)
\end{aligned}
$$

and

$$
P_3 := P((\mathcal{S}_{T,M}^2)^c).
$$

However, by [A7] we have $P_1 \to 0$ as $T \to \infty$. Take $R > 0$ satisfying $B_R(\theta^*) = \{\theta \in \mathbb{R}^{\mathsf{p}}; |\theta - \theta^*|\} \subset \Theta$. Since $\hat{u}_T = O_p(1)$, $|a_T\hat{u}_T| < R$ for large T with large probability, therefore $P_2 \to 0$ as $T \to \infty$. Moreover, from [A5] and [A6], for every $\epsilon > 0$, there exist constants $M > 0$ and $\mathcal{T} \in \mathbb{T}$ such that $P_3 < \epsilon$ for every $T > \mathcal{T}$. □

2.4 Limit distribution

In this section, we consider the limit theorem of $\hat{\theta}_T$. Let $\tilde{u}_T = \tilde{a}_T^{-1}(\hat{\theta}_T - \theta^*)$ and $\widetilde{\mathbb{U}}_T(= \widetilde{\mathbb{U}}_T(\omega)) = \{u \in \mathbb{R}^{\mathsf{p}}; \theta^* + \tilde{a}_T u \in \overline{\Theta}\}$. Define the random field $\widetilde{\mathbb{Z}}_T^\dagger$ on $\widetilde{\mathbb{U}}_T$ by

$$
\widetilde{\mathbb{Z}}_T^\dagger(u) = \exp\Big(\mathbb{H}_T^\dagger(\theta^* + \tilde{a}_T u) - \mathbb{H}_T^\dagger(\theta^*)\Big),
$$

then

$$
\tilde{u}_T \in \operatorname*{argmax}_{u\in\widetilde{\mathbb{U}}_T} \widetilde{\mathbb{Z}}_T^\dagger(u)
$$

and

$$\hat{u}_T = G_T \tilde{u}_T.$$

For convenience, we extend $\widetilde{\mathbb{Z}}_T$ to $\mathbb{R}^{\mathsf{p}}$ so that the extension has a compact support and $\sup_{u \in \mathbb{R}^{\mathsf{p}} \setminus \widetilde{\mathbb{U}}_T} \widetilde{\mathbb{Z}}_T^{\dagger}(u) \leq \max_{u \in \partial \widetilde{\mathbb{U}}_T} \widetilde{\mathbb{Z}}_T^{\dagger}(u)$. In order to describe the limit distribution of $\tilde{u}_T$, we consider the following two conditions.

[A8] For all $M > 0$,

$$\sup_{\substack{u \in \mathbb{U}_T \\ |u| < M}} \left| \mathbb{H}_T(\theta^* + a_T u) - \mathbb{H}_T(\theta^*) \right| = O_p(1)$$

as $T \to \infty$.

[A9] For all $M > 0$,

$$\sup_{\substack{u,v \in \mathbb{U}_T \\ |u|,|v| < M \\ u \neq v}} \frac{|\mathbb{H}_T(\theta^* + a_T u) - \mathbb{H}_T(\theta^* + a_T v)|}{|u - v|^q} = O_p(1)$$

as $T \to \infty$.

Proposition 2.4.1. Assume Condition [A7′] is fulfilled, then [A8] and [A9] hold.

Proof. Similar to the proof of Proposition 2.3.4. □

Theorem 2.4.2. Assume [A1], [A5], [A6], [A8] and [A9]. If $\hat{u}_T = O_p(1)$, then

$$\tilde{u}_T = O_p(1)$$

as $T \to \infty$.

Proof. By assumption and the definition of $\tilde{u}_T$,

$$\tilde{u}_T^{(1)} = \hat{u}_T^{(1)} = O_p(1), \tag{2.12}$$

therefore it suffices to prove that $\tilde{u}_T^{(0)} = O_p(1)$. For $T \in \mathbb{T}$, $R > 0$, $M > 0$ and ε_L as in [A1], define $\mathcal{S}^1_{T,R,M}$ by

$$\mathcal{S}^1_{T,R,M} = \{|\hat{u}_T| < R, |\mathbb{Z}_T(0, \hat{u}_T^{(1)})| > \exp(-M^{2-\varepsilon_L})\}.$$

Moreover define P_1, P_2 and $P_3(R)$ by

$$P_1 = P\left(\sup_{\substack{|G_T^{(00)}u^{(0)}|\ge M\\ (G_T^{(00)}u^{(0)},\hat{u}_T^{(1)})\in\mathbb{U}_T}} \mathbb{Z}_T^\dagger(G_T^{(00)}u^{(0)},\hat{u}_T^{(1)}) \ge \mathbb{Z}_T^\dagger(0,\hat{u}_T^{(1)}), \mathcal{S}_{T,R,M}^1\right),$$

$$P_2 = P\left(\sup_{\substack{0<|G_T^{(00)}u^{(0)}|\le M\\ (G_T^{(00)}u^{(0)},\hat{u}_T^{(1)})\in\mathbb{U}_T}} \mathbb{Z}_T^\dagger(G_T^{(00)}u^{(0)},\hat{u}_T^{(1)}) \ge \mathbb{Z}_T^\dagger(0,\hat{u}_T^{(1)}), \mathcal{S}_{T,R,M}^1\right),$$

and

$$P_3(R) = P\big((\mathcal{S}_{T,R,M}^1)^c\big).$$

Then for every $M>0$ and $T\in\mathbb{T}$,

$$\begin{aligned} &P(|\tilde{u}_T^{(0)}|>M)\\ &\le P\left(\sup_{\substack{|u^{(0)}|\ge M\\ (u^{(0)},\tilde{u}_T^{(1)})\in\widetilde{\mathbb{U}}_T}} \widetilde{\mathbb{Z}}_T^\dagger(u^{(0)},\tilde{u}_T^{(1)}) \ge \widetilde{\mathbb{Z}}_T^\dagger(0,\tilde{u}_T^{(1)})\right)\\ &\le P_1+P_2+P_3(R). \end{aligned}$$

By [A1],

$$\begin{aligned} P_1 &\le P\left(\sup_{\substack{|G_T^{(00)}u^{(0)}|\ge M\\ (G_T^{(00)}u^{(0)},\hat{u}_T^{(1)})\in\mathbb{U}_T}} \mathbb{Z}_T(G_T^{(00)}u^{(0)},\hat{u}_T^{(1)}) \ge \mathbb{Z}_T(0,\hat{u}_T^{(1)}), \mathcal{S}_{T,R,M}^1\right)\\ &\le P\left[\sup_{u\in V_T(M)} \mathbb{Z}_T(u) \ge \exp(-M^{2-\varepsilon_L})\right]\\ &\le \frac{C_L}{M^L} \end{aligned} \tag{2.13}$$

for every $R>0$, $M>0$ and $T\in\mathbb{T}$. Take $R'>0$ satisfying $\overline{B_{R'}(\theta^*)}\subset\Theta$ and take $\mathcal{T}_R$ satisfying $\sup_{T>\mathcal{T}_R}||a_T||R<R'$, then for every $R>0$, $M>0$ and $T>\mathcal{T}_R$, $(0,\hat{u}_T^{(1)})\in\mathbb{U}_T$ on $\mathcal{S}_{T,R,M}^1$. Similarly to (2.11), by definition of $\mathtt{C}_{T,R}$, for every $u^{(0)}$ and $\hat{u}_T^{(1)}$ such that $(0,\hat{u}_T^{(1)})$ and $(G_T^{(00)}u^{(0)},\hat{u}_T^{(1)})$ belong to $\mathbb{U}_T$,

$$\mathbb{Z}_T^\dagger(G_T^{(00)}u^{(0)},\hat{u}_T^{(1)}) - \mathbb{Z}_T^\dagger(0,\hat{u}_T^{(1)}) \le \mathtt{C}_{T,R}|G_T^{(00)}u^{(0)}|^q - \sum_{j\in\mathcal{J}^{(0)}} \xi_T^j p\big((\xi_T^j)^{-\frac{1}{q}}u_j\big).$$

Denote $B_1 = -\sum_{j\in\mathcal{J}^{(0)}} \xi_T^j p\big((\xi_T^j)^{-\frac{1}{q}}u_j\big)$. For $M>0$ and $T\in\mathbb{T}$, define $\mathcal{S}^2_{T,M}$ by

$$\mathcal{S}^2_{T,M} = \left\{\sup_{0<|G_T^{(00)}u^{(0)}|\le M} \frac{B_1}{|u^{(0)}|^q} < -\frac{\lambda}{2}\right\}.$$

Then we have

$$\begin{aligned} P_2 &\le P\Big(\sup_{\substack{0<|G_T^{(00)}u^{(0)}|\le M\\ (G_T^{(00)}u^{(0)},\hat{u}_T^{(1)})\in\mathbb{U}_T}} (\mathtt{C}_{T,R}||G_T^{(00)}||^q - \frac{\lambda}{2})|u^{(0)}|^q > 0\Big) \\ &\le P\Big(\mathtt{C}_{T,R}||G_T^{(00)}||^q > \frac{\lambda}{2}\Big) + P\big((\mathcal{S}^2_{T,M})^c\big) \end{aligned}$$

for every $R>0$, $M>0$ and $T>\mathcal{T}_R$. By [A9] and (2.6), for every $\delta>0$, $R>0$ and $M>0$, there exists a constant $T_1(\delta,R,M)>\mathcal{T}_R$ such that

$$P\Big(\mathtt{C}_{T,R}||G_T^{(00)}||^q > \frac{\lambda}{2}\Big) < \delta$$

for every $T \ge T_1(\delta,R,M)$. Moreover [A5] implies that for every $\delta>0$ and $M>0$, there exists a constant $T_2(\delta,M)$ such that

$$P\big((\mathcal{S}^2_{T,M})^c\big) < \delta$$

for every $T > T_2(\delta,M)$. Therefore, for every $\delta>0$, $R>0$ and $M>0$, there exists a constant $T_3(\delta,R,M)$ such that

$$P_2 < 2\delta \tag{2.14}$$

for every $T>T_3(\delta,R,M)$. From (2.12), for every $\delta>0$, there exist $R_1>0$ and $T_4>0$ such that

$$P(|\hat{u}_T| \ge R_1) < \delta$$

for every $T > T_4$. Moreover, [A8] implies that for every $\delta>0$, there exist constants $T_5>0$ and $M_1>0$ such that

$$P_3(R_1) < \delta \tag{2.15}$$

for every $T>T_5$ and $M>M_1$. We have the desired result from (2.13), (2.14) and (2.15). □

We write $B(R) = \{u\in\mathbb{R}^{\mathsf{p}}; |u|\le R\}$. In order to describe the limit distribution of $\tilde{u}_T$, we introduce the local asymptotic quadraticity of $\mathbb{H}_T$.

Definition 2.4.3. The family $\mathbb{H}_T$ is called locally asymptotically quadratic (LAQ) at θ^* if there exist random vectors $\Delta_T, \Delta \in \mathbb{R}^{\mathsf{p}}$, random matrices $\Gamma_T, \Gamma \in \mathbb{R}^{\mathsf{p}\times\mathsf{p}}$ and random fields $r_T : \Omega \times \mathbb{U}_T \to \mathbb{R}$ such that

[A10] (i) for every $T \in \mathbb{T}$ and $u \in \mathbb{U}_T$

$$\mathbb{H}_T(\theta^* + a_T u) - \mathbb{H}_T(\theta^*) = \Delta_T' u - \frac{1}{2} u' \Gamma_T u + r_T(u),$$

(ii) Γ is almost surely positive definite,

(iii) $(\Delta_T, \Gamma_T) \xrightarrow{d} (\Delta, \Gamma)$ as $T \to \infty$,

(iv) For all $R > 0$, $\sup_{u \in B(R)} |r_T(u)| \xrightarrow{p} 0$ as $T \to \infty$.

Remark 2.4.4. One needs a certain global non-degeneracy of the random fields $\mathbb{H}_T$ as well as the LAQ property to prove the PLDI. Therefore [A1] is not redundant under [A10]. Moreover, the LAQ property will be used to identify the limit distribution of the estimators.

Let

$$\mathbb{Z}(u) = \exp\left(\Delta' u - \frac{1}{2} u' \Gamma u\right) \qquad (u \in \mathbb{R}^{\mathsf{p}})$$

and let $\hat{C}(\mathbb{R}^{\mathsf{p}}) = \{f \in C(\mathbb{R}^{\mathsf{p}}); \lim_{|u|\to\infty} |f(u)| = 0\}$. Equip $\hat{C}(\mathbb{R}^{\mathsf{p}})$ with the supremum norm. It is possible to extend $\mathbb{Z}_T$ from $\mathbb{U}_T$ to $\mathbb{R}^{\mathsf{p}}$ in such a way that the extended $\mathbb{Z}_T$ takes values in $\hat{C}(\mathbb{R}^{\mathsf{p}})$ and $0 \le \mathbb{Z}_T(u) \le \max_{v \in \partial \mathbb{U}_T} \mathbb{Z}_T(v)$ for all $u \in \mathbb{R}^{\mathsf{p}} \setminus \mathbb{U}_T$. We will write $\mathbb{Z}_T$ for the extended random field on $\mathbb{R}^{\mathsf{p}}$.

Proposition 2.4.5. Given $L > 0$, suppose that [A1] and [A10] are fulfilled. Let $m \in (0, L)$, then

$$E\big[f\big(a_T^{-1}(\hat{\theta}_{\mathrm{ML}} - \theta^*)\big)\big] \quad \to \quad E[\big[f\big(\Gamma^{-1}\Delta\big)\big]$$

as $T \to \infty$ for any continuous function $f : \mathbb{R}^{\mathsf{p}} \to \mathbb{R}$ satisfying $\limsup_{|u|\to\infty} |f(u)||u|^m < \infty$.

Proof. The finite-dimensional convergence $\mathbb{Z}_T \to^{d_f} \mathbb{Z}$ is obvious. By [A10], we see that for any $\epsilon > 0$,

$$\lim_{\delta\to\infty} \limsup_{T\to\infty} P\big(w_T(\delta, R) \ge \epsilon\big) = 0 \tag{2.16}$$

where

$$w_T(\delta, R) = \sup_{\substack{u^1, u^2 \in B(R) \\ |u^1 - u^2| \le \delta}} \big|\log \mathbb{Z}_T(u^1) - \log \mathbb{Z}_T(u^2)\big|.$$

Now the desired result follows from Theorem 4 of Yoshida (2011). □

Remark 2.4.6. As a matter of fact, for Proposition 2.4.5, the inequality of [A1] can be weakened. See Yoshida (2011) for details.

[A11] For every $j \in \mathcal{J}^{(1)}$, there exists a constant $\beta_j \in \mathbb{R}$ such that

$$\xi_T^j \alpha_T^j \xrightarrow{p} \beta_j$$

as $T \to \infty$.

Example 2.4.7. Bridge type penalty $p(\theta) = \sum_{j=1}^{\mathsf{p}} |\alpha_T^j|^{-q'} |\theta_j|^q$ as in Example 2.3.1 satisfies [A11]. Especially, if $q' < 1$, then $\beta_j = 0$ for all $j \in \mathcal{J}^{(1)}$.

Define the random field $\widetilde{Z}^\dagger$ on $\mathbb{R}^{\mathsf{p}}$ by

$$\widetilde{Z}^\dagger(u) = \exp\Big((\Delta^{(1)})' u^{(1)} - \frac{1}{2} (u^{(1)})' \Gamma^{(11)} u^{(1)} - \sum_{j \in \mathcal{J}^{(0)}} \lambda |u_j|^q - \sum_{j \in \mathcal{J}^{(1)}} \beta_j \frac{d}{dx} p(\theta_j^*) u_j \Big)$$

then $\widetilde{Z}^\dagger$ has an unique maximizer $\tilde{u}_\infty = \operatorname*{argmax}_{u \in \mathbb{R}^{\mathsf{p}}} \widetilde{Z}^\dagger(u)$ where $\tilde{u}_\infty^{(0)} = 0$ and $\tilde{u}_\infty^{(1)} = (\Gamma^{(11)})^{-1} (\Delta^{(1)} - \boldsymbol{\psi}^{(1)})$. Here $\boldsymbol{\psi}$ is some p-dimensional vector such that $\boldsymbol{\psi}_j = \beta_j \frac{d}{dx} p(\theta_j^*)$ for $j \in \mathcal{J}^{(1)}$. In the above setting, we estimate the asymptotic distribution of $\tilde{u}_T$.

Theorem 2.4.8. Assume Conditions [A2], [A5], [A6], [A10] and [A11]. If $\tilde{u}_T = O_p(1)$, then

$$(\tilde{a}_T^{(0)})^{-1} (\hat{\theta}_T^{(0)} - \theta^{*(0)}) \xrightarrow{p} 0$$

and

$$(\tilde{a}_T^{(1)})^{-1} (\hat{\theta}_T^{(1)} - \theta^{*(1)}) \xrightarrow{d} (\Gamma^{(11)})^{-1} (\Delta^{(1)} - \boldsymbol{\psi}^{(1)})$$

as $T \to \infty$.

Proof. It suffices to verify $\tilde{u}_T \xrightarrow{d} \tilde{u}_\infty$ as $T \to \infty$. From [A2], [A5], [A6], [A10] and [A11], it follows that

$$\big(\widetilde{\mathbb{Z}}_T^\dagger(u^1), \ldots, \widetilde{\mathbb{Z}}_T^\dagger(u^n) \big) \xrightarrow{d} \big(\widetilde{Z}^\dagger(u^1), \ldots, \widetilde{Z}^\dagger(u^n) \big) \tag{2.17}$$

as $T \to \infty$, for every $n \in \mathbb{N}$ and $u^1, \ldots, u^n \in \mathbb{R}$.
For $\delta > 0$ and $R > 0$, define $\tilde{w}_T(\delta, R)$ by

$$\tilde{w}_T(\delta, R) = \sup_{\substack{u^1, u^2 \in B(R) \\ |u^1 - u^2| \le \delta}} \big| \log \widetilde{\mathbb{Z}}_T^\dagger(u^1) - \log \widetilde{\mathbb{Z}}_T^\dagger(u^2) \big|.$$

Then, from [A2], [A5], [A6], [A10] and [A11], we have

$$\lim_{\delta\to 0}\limsup_{T\to\infty} P\big(\tilde{w}_T(\delta, R) > \epsilon\big) = 0 \tag{2.18}$$

for each $R > 0$ and $\epsilon > 0$. From (2.17) and (2.18), it follows that $\widetilde{\mathbb{Z}}_T^{\dagger}|_{B(R)} \xrightarrow{d} \widetilde{Z}^{\dagger}|_{B(R)}$ in $C(B(R))$ for every $R > 0$, where $\widetilde{\mathbb{Z}}_T^{\dagger}|_{B(R)}$ and $\widetilde{Z}|_{B(R)}$ denote the restriction of $\widetilde{\mathbb{Z}}_T^{\dagger}$ and $\widetilde{Z}^{\dagger}$ on $B(R)$ respectively. Since $\tilde{u}_T = O_p(1)$, we have $\tilde{u}_T \xrightarrow{d} \tilde{u}_\infty$ as $T \to \infty$. □

Remark 2.4.9. The convergence of distribution of the result of Theorem 2.4.8 can be extended to the stable convergence, if we replace the convergence of distribution in [A10](iii) by the stable convergence.

2.5 Probability of variable selection

PLDI provides uniformly boundedness of $\hat{u}_T$ as mentioned in (2.5). It enable us to estimate a probability that a correct model is selected. Let $\eta \in (0, 1]$. For $T \in \mathbb{T}$ and $R > 0$, define $\mathsf{c}_{T,R}$ by

$$\mathsf{c}_{T,R} = \sup_{\substack{u,v\in\mathbb{U}_T \\ |a_T u|,|a_T v|<R||a_T||^{1-\eta} \\ u\neq v}} \frac{|\mathbb{H}_T(\theta^* + a_T u) - \mathbb{H}_T(\theta^* + a_T v)|}{|u - v|^q}.$$

For $m > 0$, we denote $E[|\mathsf{c}_{T,R}|^m||G_T^{(00)}||^{qm}]$ by $c_T(m, R)$. If $E[|\mathsf{c}_{T,R}|^m||G_T^{(00)}||^{qm}] = \infty$, we define $c_T(m, R) = \infty$.

Remark 2.5.1. The sequence $c_T(m, R)$ is expected to be small as $T \to \infty$. We will estimate it at the end of this section.

Theorem 2.5.2. Given $m > 0$, suppose that the inequality (2.5) is fulfilled. Moreover assume [A5]. Then for every $R > 0$ and $m_0 > 0$ there exists a positive constant $D_{m,m_0,R}$ such that

$$P(\hat{\theta}_T^{(0)} \neq 0) < D_{m,m_0,R}(||a_T||^{m\eta} + c_T(m_0, R)) \tag{2.19}$$

for every $T \in \mathbb{T}$.

Proof. By [A5], there exists a positive constant R_1 such that $p(x) > \lambda|x|^q/2$ for every x satisfying $|x| < R_1$. Take $R_2 > 0$ satisfying $\overline{B_{R_2}(\theta^*)} = \{\theta \in \mathbb{R}^{\mathsf{p}}; |\theta - \theta^*| \leq R_2\} \subset \Theta$ and let $R_3 = \min\{R_1, R_2\}$. For $T \in \mathbb{T}$ define $\mathcal{S}_T$ by

$$\mathcal{S}_T = \{|a_T\hat{u}_T| < R_3, |a_T\hat{u}_T| < R||a_T||^{-\eta}\}.$$

By (2.11) and definition of R_3 and $\mathsf{c}_{T,R}$,

$$P(\hat{\theta}_T^{(0)} \neq 0) \leq P\Big(\mathbb{Z}_T^{\dagger}(\hat{u}_T^{(0)}, \hat{u}_T^{(1)}) \geq \mathbb{Z}_T^{\dagger}(0, \hat{u}_T^{(1)}), \hat{u}_T^{(0)} \neq 0\Big)$$
$$\leq P\Big(\mathsf{c}_T|\hat{u}_T^{(0)}|^q \geq \frac{\lambda}{2}|(G_T^{(00)})^{-1}\hat{u}_T^{(0)}|^q, \hat{u}_T^{(0)} \neq 0\Big) + P(\mathcal{S}_T^c).$$

Therefore it suffices to estimate the following two probabilities:

$$P_1 := P\Big(\mathsf{c}_T||G_T^{(00)}||^q \geq \frac{\lambda}{2}\Big),$$

and

$$P_2 := P(\mathcal{S}_T^c).$$

By Markov's inequality,

$$P_1 \leq \Big(\frac{2}{\lambda}\Big)^{m_0} c_T(m_0, R) \tag{2.20}$$

and

$$P_2 \leq (R^{-m}||a_T||^{m\eta} + R_3^{-m}||a_T||^m) \sup_{T\in\mathbb{T}} E[|\hat{u}_T|^m] \tag{2.21}$$

for every $T \in \mathbb{T}$.

From (2.5), (2.20) and (2.21), we have (2.19) for some $D_{m,m_0,R}$.

□

Theorem 2.5.2 gives an upper bound of the probability of overfitting, however, we need to estimate the probability of underfitting. Let $\underline{\hat{\theta}}_T^{(1)} = \min_{j\in\mathcal{J}^{(1)}} |\hat{\theta}_{T,j}|$.

Theorem 2.5.3. Given $m > 0$, suppose that the inequality (2.5) is fulfilled. Then there exists a positive constant D_m such that

$$P(\underline{\hat{\theta}}_T^{(1)} = 0) < D_m||a_T||^m \tag{2.22}$$

for every $T \in \mathbb{T}$.

Proof. Let $\underline{\theta}^{*(1)} = \min_{j\in\mathcal{J}^{(1)}} |\theta_j^*|$, then

$$P(\underline{\hat{\theta}}_T^{(1)} = 0) \leq P(|a_T\hat{u}_T| \geq \underline{\theta}^{*(1)}).$$

Moreover, by Markov's inequality,

$$P(\underline{\hat{\theta}}_T^{(1)} = 0) \leq \frac{||a_T||^m}{|\underline{\theta}^{*(1)}|^m} \sup_{T\in\mathbb{T}} E[|\hat{u}_T|^m]$$

for all $T \in \mathbb{T}$. Therefore from assuption, we have (2.22). □

We obtain the following corollary from Theorem 2.5.2 and Theorem 2.5.3:

Corollary 2.5.4. Given $m > 0$, suppose that the inequality (2.5) is fulfilled. Moreover assume that Condition [A5] holds. Then for every $R > 0$ and $m_0 > 0$, there exists a positive constant $D_{m,m_0,R}$ such that

$$P\big(\{j; \hat{\theta}_{T,j} = 0\} \neq \mathcal{J}^{(0)}\big) < D_{m,m_0,R}(||a_T||^{m\eta} + c_T(m_0, R))$$

for every $T \in \mathbb{T}$.

2.5.1 Estimation of $c_T(m_0, R)$

In this subsection, we assume that G_T is deterministic for simplicity and denote $G_T^{(00)}$ by g_T. Let $R_0^* > 0$ and $m_1 > 0$.

[A12] (i) For every $T \in \mathbb{T}$, $\mathbb{H}_T$ is almost surely thrice differentiable with respect to θ on $B^* = B_{R_0^*}(\theta^*, \Theta) = \{\theta \in \Theta; |\theta - \theta^*| < R_0^*\}$,

(ii) $\sup_{T\in\mathbb{T}} E\Big[||a_T||^{m_1}|\partial_\theta \mathbb{H}_T(\theta^*)|^{m_1}\Big] < \infty$,

(iii) $\sup_{T\in\mathbb{T}} E\Big[||a_T||^{2m_1} \sup_{\theta\in B^*} |\partial_\theta^2 \mathbb{H}_T(\theta)|^{m_1}\Big] < \infty$,

(iv) $\sup_{T\in\mathbb{T}} E\Big[||a_T||^{2m_1} \sup_{\theta\in B^*} |\partial_\theta^3 \mathbb{H}_T(\theta)|^{m_1}\Big] < \infty$.

Proposition 2.5.5. Assume that Condition [A12] holds. Then there exist positive constants $R_0 > 0$ and $K > 0$ such that

$$c_T(m_1, R_0) \leq K(||a_T||^{-\eta(2-q)} g_T^q)^{m_1}$$

for every $T \in \mathbb{T}$ satisfying $||a_T|| \leq 1$.

Proof. Take $R_0 \leq R_0^*$ satisfying that $\overline{B_{R_0}(\theta^*)} \subset \Theta$, then $\theta^* + a_T u \in B^*$ for every u satisfying that $|a_T u| < R_0$. Similarly to the proof of Proposition 2.3.4, we have

$$\big|\mathbb{H}_T(\theta^* + a_T u) - \mathbb{H}_T(\theta^* + a_T v)\big| \leq A_1 + A_2 + A_3$$

for every $u, v \in \mathbb{R}$ satisfying $|a_T u|, |a_T v| < R_0$ where

$$A_1 = |a_T(u - v)| \cdot |\partial_\theta \mathbb{H}_T(\theta^*)|,$$
$$A_2 = |a_T(u + v)| \cdot |a_T(u - v)| \cdot \sup_{\theta\in B^*} |\partial_\theta^2 \mathbb{H}_T(\theta)|$$

and

$$A_3 = |a_T v|^2 \cdot |a_T(u - v)| \cdot \sup_{\theta\in B^*} |\partial_\theta^3 \mathbb{H}_T(\theta)|.$$

If both $|a_T u|$ and $|a_T v|$ are less than $||a_T||^{1-\eta}R_0$, then

$$\begin{aligned}\frac{|a_T(u-v)|}{|u-v|^q} &\leq |a_T(u-v)|^{1-q}\frac{||a_T||^q|u-v|^q}{|u-v|^q}\\ &\leq (2R_0||a_T||^{1-\eta})^{1-q}||a_T||^q\\ &= (2R_0)^{1-q}||a_T||^{(1-\eta)(1-q)+q}.\end{aligned}$$

Therefore

$$\begin{aligned}E[|\mathsf{c}_{T,R_0}|^{m_1}] &\leq E\left[\sup_{\substack{u,v\in\mathbb{U}_T\\ |a_T u|,|a_T v|<R_0||a_T||^{1-\eta}\\ u\neq v}}\left(\frac{A_1+A_2+A_3}{|u-v|^q}\right)^{m_1}\right]\\ &\leq E\left[\sup_{\substack{u,v\in\mathbb{U}_T\\ |a_T u|,|a_T v|<R_0||a_T||^{1-\eta}\\ u\neq v}}\frac{3^{m_1-1}(A_1^{m_1}+A_2^{m_1}+A_3^{m_1})}{|u-v|^{qm_1}}\right]\\ &\leq 3^{m_1-1}(A_1'+A_2'+A_3')\end{aligned}$$

for every $T\in\mathbb{T}$ satisfying $||a_T||\leq 1$ where

$$\begin{aligned}A_1' &= E\left[\left((2R_0)^{1-q}||a_T||^{(1-\eta)(1-q)+q}|\partial_\theta\mathbb{H}_T(\theta^*)|\right)^{m_1}\right],\\ A_2' &= E\left[\left((2R_0)^{2-q}||a_T||^{(1-\eta)(2-q)+q}\sup_{\theta\in B^*}|\partial_\theta^2\mathbb{H}_T(\theta)|\right)^{m_1}\right]\end{aligned}$$

and

$$A_3' = E\left[\left(2^{1-q}R_0^{3-q}||a_T||^{(1-\eta)(3-q)+q}\sup_{\theta\in B^*}|\partial_\theta^3\mathbb{H}_T(\theta)|\right)^{m_1}\right].$$

If $||a_T||\leq 1$, then Conditions [A12](ii), (iii) and (iv) imply

$$\begin{aligned}A_1' &\leq K'||a_T||^{-\eta(1-q)m_1}\leq K'||a_T||^{-\eta(2-q)m_1},\\ A_2' &\leq K'||a_T||^{-\eta(2-q)m_1}\end{aligned}$$

and

$$A_3'\leq K'||a_T||^{(1-\eta(3-q))m_1}\leq K'||a_T||^{-\eta(2-q)m_1}$$

for some $K'>0$, respectively. Let $K>3^{m_1}K'$, we have the desired result. □

Example 2.5.6. Define $p(\theta) = \sum_{j=1}^{\mathsf{p}} |\alpha_T^j|^{-q'}|\theta_j|^q$ as in Example 2.3.1, then we have $g_T = ||a_T||^{(q'-q)/q}$. Let $m, m_1 > 0$ and suppose that the inequality (2.5) is fulfilled. Moreover assume that Conditions [A5] and [A12] hold. Let $\eta = (q'-q)m_1/(m+2(1-q)m_1)$. Then by Proposition 2.5.5 and Corollary 2.5.4, there exists a constant D_{m,m_1} such that

$$P\big(\{j; \hat{\theta}_{T,j} = 0\} \neq \mathcal{J}^{(0)}\big) < D_{m,m_1}||a_T||^{\frac{(q'-q)mm_1}{m+2(1-q)m_1}}$$

for every $T \in \mathbb{T}$.

2.5.2 The case of random $G_T^{(00)}$

Now we turn to the estimation of $c_T(m_0, R)$ in the case where $G_T^{(00)}$ is random. Let $m_2 > 0$.

Proposition 2.5.7. Assume that Condition [A12] holds. Then for every $m_2 > 0$, there exist positive constants $R_0 > 0$ and $K > 0$ such that

$$c_T\Big(\frac{m_1 m_2}{qm_1 + m_2}, R_0\Big) \leq K||a_T||^{-\frac{\eta(2-q)m_1m_2}{qm_1+m_2}}\big(E[||G_T^{(00)}||^{m_2}]\big)^{\frac{qm_1}{qm_1+m_2}}$$

for every $T \in \mathbb{T}$ satisfying $||a_T|| \leq 1$.

Proof. Similarly to the proof of Proposition 2.5.5, there exist constants $R_0 > 0$ and $K' > 0$ such that

$$E[|\mathsf{c}_{T,R_0}|^{m_1}] \leq K'||a_T||^{-\eta(2-q)m_1}$$

for every $T \in \mathbb{T}$ satisfying that $||a_T|| \leq 1$. Therefore from Hölder's inequality, we have

$$\begin{aligned} E[|\mathsf{c}_{T,R_0}|^{m_0}||G_T^{(00)}||^{qm_0}] &\leq \big(E[|\mathsf{c}_{T,R_0}|^{m_1}]\big)^{\frac{m_2}{qm_1+m_2}}\big(E[||G_T^{(00)}||^{m_2}]\big)^{\frac{qm_1}{qm_1+m_2}} \\ &\leq K'^{\frac{m_2}{qm_1+m_2}}||a_T||^{-\frac{\eta(2-q)m_1m_2}{qm_1+m_2}}\big(E[||G_T^{(00)}||^{m_2}]\big)^{\frac{qm_1}{qm_1+m_2}} \end{aligned}$$

where $m_0 = m_1m_2/(qm_1 + m_2)$. □

Example 2.5.8. In example 2.5.6, we have $g_T = ||a_T||^{(q'-q)/q}$. Here we assume that there exist positive constants m_2 and K' such that $E[||G_T^{(00)}||^{m_2}] \leq K'||a_T||^{(q'-q)m_2/q}$ for every $T \in \mathbb{T}$. Suppose that Condition [A12] holds. Then by proposition 2.5.7, there exists a constant $K > 0$ such that

$$c_T \leq K||a_T||^{\frac{(q'+q\eta-2\eta-q)m_1m_2}{qm_1+m_2}}$$

for every $T \in \mathbb{T}$ satisfying that $||a_T|| \leq 1$. Let $m > 0$ and suppose that the inequality (2.5) is fulfilled. Moreover assume that Condition [A5] holds. Let $\eta = (q' - q)m_1m_2/(qmm_1 + mm_2 + 2m_1m_2 - qm_1m_2)$, then by Corollary 2.5.4, there exists a constant D_m such that

$$P\big(\{j; \hat{\theta}_{T,j} = 0\} \neq \mathcal{J}^{(0)}\big) < D_{m,m_1,m_2}||a_T||^{\frac{(q'-q)mm_1m_2}{qmm_1+mm_2+(2-q)m_1m_2}}$$

for every $T \in \mathbb{T}$

2.6 Moment convergence

In this section, we will study the moment convergence of $\hat{u}_T$. The following theorem is a consequence of PLDI:

Theorem 2.6.1. Given $m > 0$, suppose that (2.5) holds. Moreover assume that the conclusion of Theorem 2.4.8 holds. Then we have

$$E[f(\hat{u}_T)] \to E[f(\tilde{u}_\infty)]$$

as $T \to \infty$ for any continuous function $f : \mathbb{R}^{\mathsf{p}} \to \mathbb{R}$ satisfying $\limsup_{|u|\to\infty} |f(u)||u|^{-m} = 0$.

Proof. Condition (2.5) implies an uniform integrability of $\{f(\hat{u}_T)\}_{T\in\mathbb{T}}$. By assumption, $\hat{u}_T \xrightarrow{d} \tilde{u}_\infty$ as $T \to \infty$. Therefore we can obtain the desired result. □

Theorem 2.6.1 suggests that $\lim_{T\to\infty} E\big[|(a_T^{(00)})^{-1}\hat{\theta}_T^{(0)}|^m\big] = 0$ for $m \in (0, L)$. From Theorem 2.5.2, we have another estimation of $\hat{\theta}_T^{(0)}$. Let $\Psi_T \in \mathrm{GL}(|\mathcal{J}^{(0)}|)$ be a deterministic sequence of positive matrices. For $m > 0$, we consider the condition

[A13] $||\Psi_T||^{m^*}(||a_T||^{m\eta} + c_T(m_0, R)) \to 0$ as $T \to \infty$.

Theorem 2.6.2. Given $m, m^*, m_0 > 0$, suppose that Condition [A13] and the inequality (2.19) hold. Then

$$E[|\Psi_T\hat{\theta}_T^{(0)}|^{m^*}] \to 0$$

as $T \to \infty$.

Proof. Let $\Theta_{\max} = \sup_{\theta\in\overline{\Theta}} |\theta|$, then

$$E[|\Psi_T\hat{\theta}_T^{(0)}|^{m^*}] \leq \Theta_{\max}^{m^*}||\Psi_T||^{m^*}P(\hat{\theta}_T^{(0)} \neq 0).$$

Therefore, from [A13], we have the desired result. □

2.7 Group lasso

In this section, we divide the unknown parameter into g different groups for example factor-level indicators of categorical data. Let $\mathsf{p}_1 + \cdots + \mathsf{p}_\mathsf{g} = \mathsf{p}$ and ${}^g\theta \in \mathbb{R}^{\mathsf{p}_g}$ for $g = 1, \ldots, \mathsf{g}$. We write $\theta = ({}^1\theta, \ldots, {}^\mathsf{g}\theta)$ and $a_T = \mathrm{diag}({}^1\alpha_T, \ldots, {}^\mathsf{g}\alpha_T)$. Moreover we denote $\mathrm{diag}({}^g\alpha_T) \in \mathbb{R}^{\mathsf{g}\times\mathsf{g}}$ by ${}^g a_T$. In this situation, the group lasso provides group-wise sparsity (Simon et al. 2013, Yuan and Lin 2006). The group lasso has the following penalty:

$$p_T(\theta) = \sum_{g=1}^{\mathsf{g}} {}^g\xi_T p_g({}^g\theta)$$

where ${}^g\xi_T$ is a (possibly random) positive sequence and $p_g : \mathbb{R}^{\mathsf{p}_g} \to \mathbb{R}_{\geq 0}$ is a function satisfying $p_g(0) = 0$ for $g = 1, \ldots, \mathsf{g}$.

Example 2.7.1. Let $K_1 \in \mathbb{R}^{\mathsf{p}^1\times\mathsf{p}^1}, \ldots, K_\mathsf{g} \in \mathbb{R}^{\mathsf{p}^\mathsf{g}\times\mathsf{p}^\mathsf{g}}$ be positive definite matrices. Define p_g by $p_g({}^g\theta) = ||{}^g\theta||_{K_g}$. Here $||{}^g\theta||_{K_g}$ denotes $({}^g\theta' K_g {}^g\theta)^{1/2}$. Then we have

$$p_T(\theta) = \sum_{g=1}^{\mathsf{g}} {}^g\xi_T ||{}^g\theta||_{K_g}.$$

We denote $\{g; {}^g\theta^* = 0\}$ and $\{g; {}^g\theta^* \neq 0\}$ by $\mathcal{G}^{(0)}$ and $\mathcal{G}^{(1)}$, respectively.

[A2′] p_g is differentiable except the origin $(g = 1, \ldots, \mathsf{g})$.

[A3′] For some positive constant ε,

$$\max_{1\leq g\leq \mathsf{g}} \sup_{|x|<\varepsilon} p_g(x) < \infty.$$

[A4′] For all $g \in \mathcal{G}^{(1)}$,

$$\sup_{T\in\mathbb{T}} |{}^g\alpha_T| {}^g\xi_T \leq c_0$$

almost surely.

Theorem 2.7.2. Given $L > 0$, assume Conditions [A1] and [A2′]-[A4′]. Then there exist constants $C'_L > 0$ and $\varepsilon'_L \in (0, 1)$ such that

$$P\left[\sup_{u\in V_T(r)} \mathbb{Z}_T^\dagger(u) \geq \exp(-r^{2-\varepsilon'_L})\right] \leq \frac{C'_L}{r^L}$$

for all $r > 0, T > 0$.

Proof. The proof is similarly to the proof of Theorem 2.2.3. By [A1], there exist constants $C_L > 0$ and $\varepsilon_L \in (0,1)$ satisfying (2.3) for all $r > 0, T > 0$. Let $\varepsilon'_L \in (\varepsilon_L, 1)$. Similarly to the proof of Theorem 2.2.3, for every $T > 0$ and $r > 0$, we have

$$P\left[\sup_{u\in V_T(r)} \mathbb{Z}_T^{\dagger}(u) \geq \exp(-r^{2-\varepsilon'_L})\right] \leq \sum_{n=0}^{\infty} P\left[\sup_{\substack{2^n r\leq |u|\leq 2^{n+1}r \\ u\in V_T(r)}} \mathbb{Z}_T(u)\exp(B_2) \geq \exp(-r^{2-\varepsilon'_L})\right].$$

where

$$B_2 = -\sum_{g\in\mathcal{G}^{(1)}} {}^g\xi_T\left[p_g\Big({}^g\theta^* + {}^g a_T{}^g u)\Big) - p_g({}^g\theta^*)\right].$$

Conditions [A2′] and [A3′] imply

$$\max_{1\leq g\leq \mathrm{g}} \sup_{x\in U_g\setminus\{0\}} \frac{p_g({}^g\theta + {}^g x) - p_g({}^g\theta)}{x} < \infty$$

for every ${}^g\theta \in \mathbb{R}^g \setminus \{0\}$ and every compact set $U_g \subset \mathbb{R}^g$. Moreover, by definition of $\mathbb{U}_T$, we observe that $\sup_{T\in\mathbb{T}}\sup_{u\in\mathbb{U}_T} |a_T u| < \infty$. Therefore, from [A4′], we have

$$|B_2| \leq c_0 K|u|$$

for some $K > 0$ which does not depend on T and r. Therefore, similarly to the proof of Theorem 2.2.3, we have the desired result. □

As above, grouped penalty dose not disturb PLDI. We now turn to the selection consistency of the grouped penalized estimator.

[A5′] There exists positive constants $\{\lambda_g\}_{g=1,\ldots,\mathrm{g}}$ such that

$$\lim_{{}^g x\to 0} \frac{p_g({}^g x)}{|{}^g x|^q} = \lambda_g$$

for every $g = 1, \ldots, \mathrm{g}$.

[A6′] For every $g \in \mathcal{G}^{(0)}$,

$$({}^g\xi_T)^{-\frac{1}{q}} |{}^g\alpha_T|^{-1} \xrightarrow{p} 0$$

as $T \to \infty$.

We write $\tilde{\lambda}_{\mathtt{g}} = \min_{1 \le g \le \mathtt{g}} \lambda_g$.

Theorem 2.7.3. Assume Conditions [A5′], [A6′] and [A7]. If $\hat{u}_T = O_p(1)$, then

$$P\Big(\hat{\theta}_T^{(0)} = 0\Big) \to 1$$

as $T \to \infty$.

Proof. Let $\mathcal{S}_T^1 = \{(0, \hat{u}_T^{(1)}) \in \mathbb{U}_T\}$. For $M > 0$ and $T \in \mathbb{T}$, define $\mathcal{S}_T^2$ by

$$\mathcal{S}_{T,M}^2 = \Big\{ |\hat{u}_T^{(0)}| < M, \sum_{g \in \mathcal{G}^{(0)}} {}^g\xi_T p\big({}^g a_T {}^g \hat{u}_T\big) \ge \frac{\tilde{\lambda}_{\mathtt{g}}}{2} |(G_T^{(00)})^{-1} \hat{u}_T^{(0)}|^q \Big\}$$

and define $\mathtt{C}_{T,M}$ by

$$\mathtt{C}_{T,M} = \sup_{\substack{u,v \in \mathbb{U}_T \\ |u|,|v| < M \\ u \ne v}} \frac{|\mathbb{H}_T(\theta^* + a_T u) - \mathbb{H}_T(\theta^* + a_T v)|}{|u - v|^q}.$$

Similarly to the proof of Theorem 2.3.5, we have

$$\begin{aligned} P(\hat{\theta}_T^{(0)} \ne 0) \le P\Big(\mathtt{C}_{T,M} |\hat{u}_T^{(0)}|^q \ge \frac{\tilde{\lambda}_{\mathtt{g}}}{2} |(G_T^{(00)})^{-1} \hat{u}_T^{(0)}|^q, \hat{u}_T^{(0)} \ne 0\Big) \\ + P((\mathcal{S}_T^1)^c) + P((\mathcal{S}_{T,M}^2)^c). \end{aligned}$$

Therefore it suffices to estimate the following three probabilities:

$$\begin{aligned} P_1 &:= P\Big(\mathtt{C}_{T,M} ||G_T^{(00)}||^q \ge \frac{\lambda}{2}\Big), \\ P_2 &:= P((\mathcal{S}_T^1)^c) \end{aligned}$$

and

$$P_3 := P((\mathcal{S}_{T,M}^2)^c).$$

However, by [A7] we have $P_1 \to 0$ as $T \to \infty$. Take $R > 0$ satisfying $B_R(\theta^*) = \{\theta \in \mathbb{R}^{\mathsf{p}}; |\theta - \theta^*|\} \subset \Theta$. Since $\hat{u}_T = O_p(1)$, $|a_T \hat{u}_T| < R$ for large T with large probability, therefore $P_2 \to 0$ as $T \to \infty$. Moreover, from [A5′] and [A6′], for every $\epsilon > 0$, there exist constants $M > 0$ and $\mathcal{T} \in \mathbb{T}$ such that $P_3 < \epsilon$ for every $T > \mathcal{T}$. □

The last part of this section is the limit theorem of grouped penalized estimator.

Theorem 2.7.4. Assume [A1], [A5′], [A6′], [A8] and [A9]. If $\hat{u}_T = O_p(1)$, then

$$\tilde{u}_T = O_p(1)$$

as $T \to \infty$.

Proof. For $T \in \mathbb{T}$, $R > 0$, $M > 0$ and ε_L as in [A1], define $\mathcal{S}^1_{T,R,M}$ as in the proof of Theorem 2.4.2. Similarty to the proof of Theorem2.4.2, it suffice to calculate P_1, P_2 and $P_3(R)$ where

$$P_1 = P\left(\sup_{\substack{|G_T^{(00)}u^{(0)}|\geq M \\ (G_T^{(00)}u^{(0)},\hat{u}_T^{(1)})\in\mathbb{U}_T}} \mathbb{Z}_T^\dagger(G_T^{(00)}u^{(0)},\hat{u}_T^{(1)}) \geq \mathbb{Z}_T^\dagger(0,\hat{u}_T^{(1)}), \mathcal{S}^1_{T,R,M}\right),$$

$$P_2 = P\left(\sup_{\substack{0<|G_T^{(00)}u^{(0)}|\leq M \\ (G_T^{(00)}u^{(0)},\hat{u}_T^{(1)})\in\mathbb{U}_T}} \mathbb{Z}_T^\dagger(G_T^{(00)}u^{(0)},\hat{u}_T^{(1)}) \geq \mathbb{Z}_T^\dagger(0,\hat{u}_T^{(1)}), \mathcal{S}^1_{T,R,M}\right),$$

and

$$P_3(R) = P\big((\mathcal{S}^1_{T,R,M})^c\big).$$

By [A1],

$$P_1 \leq \frac{C_L}{M^L} \tag{2.23}$$

for every $R > 0$, $M > 0$ and $T \in \mathbb{T}$. Take $R' > 0$ satisfying $\overline{B_{R'}(\theta^*)} \subset \Theta$ and take $\mathcal{T}_R$ satisfying $\sup_{T>\mathcal{T}_R} ||a_T||R < R'$, then for every $R > 0$, $M > 0$ and $T > \mathcal{T}_R$, $(0,\hat{u}_T^{(1)}) \in \mathbb{U}_T$ on $\mathcal{S}^1_{T,R,M}$. Similarly to (2.11), by definition of $\mathtt{C}_{T,R}$, for every $u^{(0)}$ and $\hat{u}_T^{(1)}$ such that $(0,\hat{u}_T^{(1)})$ and $(G_T^{(00)}u^{(0)},\hat{u}_T^{(1)})$ belong to $\mathbb{U}_T$,

$$\mathbb{Z}_T^\dagger(G_T^{(00)}u^{(0)},\hat{u}_T^{(1)}) - \mathbb{Z}_T^\dagger(0,\hat{u}_T^{(1)}) \leq \mathtt{C}_{T,R}|G_T^{(00)}u^{(0)}|^q - \sum_{g\in\mathcal{G}^{(0)}} {}^g\xi_T p_g\big(({}^g\xi_T^j)^{-\frac{1}{q}}{}^g u\big).$$

Denote $B_2 = -\sum_{g\in\mathcal{G}^{(0)}} {}^g\xi_T p_g\big(({}^g\xi_T)^{-\frac{1}{q}}{}^g u\big)$. For $M > 0$ and $T \in \mathbb{T}$, define $\mathcal{S}^2_{T,M}$ by

$$\mathcal{S}^2_{T,M} = \left\{\sup_{0<|G_T^{(00)}u^{(0)}|\leq M} \frac{B_2}{|u^{(0)}|^q} < -\frac{\tilde{\lambda}_{\mathrm{g}}}{2}\right\}.$$

Then we have

$$P_2 \leq P\Big(\sup_{\substack{0<|G_T^{(00)}u^{(0)}|\leq M \\ \big(G_T^{(00)}u^{(0)},\hat{u}_T^{(1)}\big)\in\mathbb{U}_T}} (\mathtt{C}_{T,R}||G_T^{(00)}||^q - \frac{\tilde{\lambda}_{\mathtt{g}}}{2})|u^{(0)}|^q > 0\Big)$$
$$\leq P\Big(\mathtt{C}_{T,R}||G_T^{(00)}||^q > \frac{\tilde{\lambda}_{\mathtt{g}}}{2}\Big) + P\big((\mathcal{S}_{T,M}^2)^c\big)$$

for every $R > 0$, $M > 0$ and $T > \mathcal{T}_R$. By [A9] and (2.6), for every $\delta > 0$, $R > 0$ and $M > 0$, there exists a constant $T_1(\delta, R, M) > \mathcal{T}_R$ such that

$$P\Big(\mathtt{C}_{T,R}||G_T^{(00)}||^q > \frac{\tilde{\lambda}_{\mathtt{g}}}{2}\Big) < \delta$$

for every $T \geq T_1(\delta, R, M)$. Moreover [A5′] implies that for every $\delta > 0$ and $M > 0$, there exists a constant $T_2(\delta, M)$ such that

$$P\big((\mathcal{S}_{T,M}^2)^c\big) < \delta$$

for every $T > T_2(\delta, M)$. Therefore, for every $\delta > 0$, $R > 0$ and $M > 0$, there exists a constant $T_3(\delta, R, M)$ such that

$$P_2 < 2\delta \tag{2.24}$$

for every $T > T_3(\delta, R, M)$. From (2.12), for every $\delta > 0$, there exist $R_1 > 0$ and $T_4 > 0$ such that

$$P(|\hat{u}_T| \geq R_1) < \delta$$

for every $T > T_4$. Moreover, [A8] implies that for every $\delta > 0$, there exist constants $T_5 > 0$ and $M_1 > 0$ such that

$$P_3(R_1) < \delta \tag{2.25}$$

for every $T > T_5$ and $M > M_1$. We have the desired result from (2.23), (2.24) and (2.25). □

[A11′] For every $g \in \mathcal{J}^{(1)}$, there exists a constant ${}^g\beta \in \mathbb{R}$ such that

$${}^g\xi_T|{}^g\alpha_T| \xrightarrow{p} {}^g\beta$$

as $T \to \infty$.

Define the random field $\widetilde{Z}_g^\dagger$ on $\mathbb{R}^{\mathsf{p}}$ by

$$\widetilde{Z}_g^\dagger(u) = \exp\Big((\Delta^{(1)})'u^{(1)} - \frac{1}{2}(u^{(1)})'\Gamma^{(11)}u^{(1)} \\ - \sum_{g\in\mathcal{G}^{(0)}} \lambda_g|{}^g u|^q - \sum_{g\in\mathcal{G}^{(1)}} {}^g\beta\frac{d}{d({}^g x)}p({}^g\theta^*){}^g u\Big)$$

then $\widetilde{Z}^\dagger$ has an unique maximizer $\tilde{u}_\infty = \underset{u\in\mathbb{R}^{\mathsf{p}}}{\mathrm{argmax}}\ \widetilde{Z}_g^\dagger(u)$ where $\tilde{u}_\infty^{(0)} = 0$ and $\tilde{u}_\infty^{(1)} = (\Gamma^{(11)})^{-1}(\Delta^{(1)} - \boldsymbol{\psi}_{\mathrm{g}}^{(1)})$. Here $\boldsymbol{\psi}_{\mathrm{g}}$ is some p-dimensional vector such that ${}^g\boldsymbol{\psi}_{\mathrm{g}} =$
${}^g\beta\frac{d}{d({}^g x)}p_g({}^g\theta^*)$ for $g \in \mathcal{G}^{(1)}$. In the above setting, we estimate the asymptotic distribution of $\tilde{u}_T$.

Theorem 2.7.5. Assume Conditions [A2′], [A5′], [A6′], [A10] and [A11′]. If $\tilde{u}_T = O_p(1)$, then

$$(\tilde{a}_T^{(0)})^{-1}(\hat{\theta}_T^{(0)} - \theta^{*(0)}) \xrightarrow{p} 0$$

and

$$(\tilde{a}_T^{(1)})^{-1}(\hat{\theta}_T^{(1)} - \theta^{*(1)}) \xrightarrow{d} (\Gamma^{(11)})^{-1}(\Delta^{(1)} - \boldsymbol{\psi}_{\mathrm{g}}^{(1)})$$

as $T \to \infty$.

Proof. It suffices to verify $\tilde{u}_T \xrightarrow{d} \tilde{u}_\infty$ as $T \to \infty$. From [A2′], [A5′], [A6′], [A10] and [A11′], it follows that

$$\big(\widetilde{\mathbb{Z}}_T^\dagger(u^1), \ldots, \widetilde{\mathbb{Z}}_T^\dagger(u^n)\big) \xrightarrow{d} \big(\widetilde{Z}_g^\dagger(u^1), \ldots, \widetilde{Z}_g^\dagger(u^n)\big) \tag{2.26}$$

as $T \to \infty$, for every $n \in \mathbb{N}$ and $u^1, \ldots, u^n \in \mathbb{R}$.
For $\delta > 0$ and $R > 0$, define $\tilde{w}_T(\delta, R)$ by

$$\tilde{w}_T(\delta, R) = \sup_{\substack{u^1,u^2\in B(R) \\ |u^1-u^2|\le\delta}} \big|\log\widetilde{\mathbb{Z}}_T^\dagger(u^1) - \log\widetilde{\mathbb{Z}}_T^\dagger(u^2)\big|.$$

Then, from [A2′], [A5′], [A6′], [A10] and [A11′], we have

$$\lim_{\delta\to 0}\limsup_{T\to\infty} P\big(\tilde{w}_T(\delta, R) > \epsilon\big) = 0 \tag{2.27}$$

for each $R > 0$ and $\epsilon > 0$. From (2.26) and (2.27), it follows that $\widetilde{\mathbb{Z}}_T^\dagger|_{B(R)} \xrightarrow{d} \widetilde{Z}^\dagger|_{B(R)}$ in $C(B(R))$ for every $R > 0$, where $\widetilde{\mathbb{Z}}_T^\dagger|_{B(R)}$ and $\widetilde{Z}|_{B(R)}$ denote the restriction of $\widetilde{\mathbb{Z}}_T^\dagger$ and $\widetilde{Z}_g^\dagger$ on $B(R)$ respectively. Since $\tilde{u}_T = O_p(1)$, we have $\tilde{u}_T \xrightarrow{d} \tilde{u}_\infty$ as $T \to \infty$. $\square$

Chapter 3

Application to volatility estimation

3.1 Volatility model

In this section, we apply our results to an Itó Process. Consider a stochastic regression model specified by the stochastic integral equation

$$Y_t = Y_0 + \int_0^t b_s ds + \int_0^t \sigma(X_s, \theta) dw_s, \quad t \in [0, T]. \tag{3.1}$$

Here given a stochastic basis $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T]}, P)$, w is an r-dimensional standard Wiener process, and b and X are progressively measurable processes taking values in $\mathbb{R}^{\mathsf{m}}$ and $\mathbb{R}^{\mathsf{d}}$, respectively. The function $\sigma : \mathbb{R}^{\mathsf{d}} \times \Theta \to \mathbb{R}^{\mathsf{m}} \otimes \mathbb{R}^{\mathsf{r}}$ has an unknown parameter $\theta \in \Theta$, a bounded open set of $\mathbb{R}^{\mathsf{p}}$. If $b_t = b(Y_t, t)$ and $X_t = (Y_t, t)$, then Y can be a time-inhomogeneous diffusion process. We want to estimate θ from the observations $(X_{t_j}, Y_{t_j})_{j=0,\ldots,n}$, $t_j = jh$ for $h = h_n = T/n$. No data of b_t is available.

High frequency data under finite time horizon will be treated, that is, T is fixed and n tends to ∞. This is a standard setting in finance. We will consider the penalized quasi-likelihood analysis for the volatility parameter θ. To apply the results in Sections 2.1-2.6, we will use n for T of Section 2.1, while T denotes the fixed terminal of the observations in what follows.

Let $S(x, \theta) = \sigma(x, \theta)^{\otimes 2} = \sigma(x, \theta)\sigma(x, \theta)'$. For estimation, we use the quasi-log likelihood function

$$\mathbb{H}_n(\theta) = -\frac{1}{2}\sum_{j=1}^{n} \left\{ \log\det S(X_{t_{j-1}}, \theta) + h^{-1} S^{-1}(X_{t_{j-1}}, \theta)[(\Delta_j Y)^{\otimes 2}] \right\},$$

where $\Delta_j Y = Y_{t_j} - Y_{t_{j-1}}$. Then the quasi-maximum likelihood estimator (QMLE) $\hat{\theta}_n^M$ is any estimator that satisfies

$$\hat{\theta}_n^M \in \text{argmax}_{\theta\in\overline{\Theta}} \mathbb{H}_n(\theta). \tag{3.2}$$

The quasi-Bayesian estimator (QBE) $\hat{\theta}_n^B$ with respect to the quadratic loss and a prior density $\pi : \Theta \to \mathbb{R}_+$ is given by

$$\hat{\theta}_n^B = \left(\int_\Theta \exp(\mathbb{H}_n(\theta))\pi(\theta)d\theta \right)^{-1} \int_\Theta \theta \exp(\mathbb{H}_n(\theta))\pi(\theta)d\theta. \tag{3.3}$$

The prior density π is assumed to be continuous and to satisfy $0 < \inf_{\theta\in\Theta} \pi(\theta) \leq \sup_{\theta\in\Theta} \pi(\theta) < \infty$.

Definition 3.1.1. The space $C_\uparrow^{a,b}(\mathbb{R}^{\mathsf{d}} \times \Theta; \mathbb{R}^{\mathsf{m}} \times \mathbb{R}^{\mathsf{r}})$ is the set of continuous functions $f : \mathbb{R}^{\mathsf{d}} \times \Theta \to \mathbb{R}^{\mathsf{m}} \times \mathbb{R}^{\mathsf{r}}$ that satisfy the following conditions.

(i) f has continuous derivatives $\partial_{s_1} \cdots \partial_{s_l} f$ for all $(s_1, \ldots, s_l) \in \{\theta, x\}^l$ such that $\#\{i \in \{1, \ldots, l\}; s_i = x\} \leq a$ and $\#\{i \in \{1, \ldots, l\}; s_i = \theta\} \leq b$.

(ii) Each derivative appearing in (i) satisfies

$$\sup_{\theta\in\Theta} |\partial_{s_1} \cdots \partial_{s_l} f(x, \theta)| \leq C(s_1, \ldots, s_l)(1 + |x|^{C(s_1,\ldots,s_l)}) \quad (x \in \mathbb{R}^{\mathsf{d}})$$

for some positive constant $C(s_1, \ldots, s_l)$.

We denote by $\to^{d_s(\mathcal{F})}$ the $\mathcal{F}$-stable convergence in distribution. Suppose that Θ has a Lipschitz boundary.

The following Condition [H1] is Condition [H1$^\sharp$] of Uchida and Yoshida (2013).

[H1] **(i)** $\sup_{0\leq t\leq T} \|b_t\|_p < \infty$ for all $p > 1$.

(ii) $\sigma \in C_\uparrow^{2,4}(\mathbb{R}^{\mathsf{d}} \times \Theta; \mathbb{R}^{\mathsf{m}} \otimes \mathbb{R}^{\mathsf{r}})$ and $\inf_{x,\theta} \det S(x, \theta) > 0$.

(iii) The process X has a representation

$$X_t = X_0 + \int_0^t \tilde{b}_s ds + \int_0^t a_s dw_s + \int_0^t \tilde{a}_s d\tilde{w}_s,$$

where $\tilde{b}$, a and $\tilde{a}$ are progressively measurable processes taking values in $\mathbb{R}^{\mathsf{d}}$, $\mathbb{R}^{\mathsf{d}} \otimes \mathbb{R}^{\mathsf{r}}$ and $\mathbb{R}^{\mathsf{d}} \otimes \mathbb{R}^{\mathsf{r}_1}$, respectively, and satisfy

$$\|X_0\|_p + \sup_{t\in[0,T]} (\|\tilde{b}_t\|_p + \|a_t\|_p + \|\tilde{a}_t\|_p) < \infty$$

for every $p > 1$. $\tilde{w}$ is an r_1-dimensional Wiener process independent of w, Let

$$\mathbb{Y}(\theta) = -\frac{1}{2T}\int_0^T \left\{\log\left(\frac{\det S(X_t,\theta)}{\det S(X_t,\theta^*)}\right) + \mathrm{Tr}\left(S^{-1}(X_t,\theta)S(X_t,\theta^*) - I_d\right)\right\} dt.$$

A key index χ_0 is defined by

$$\chi_0 = \inf_{\theta\neq\theta^*} \frac{-\mathbb{Y}(\theta)}{|\theta-\theta^*|^2}. \tag{3.4}$$

Non-degeneracy of χ_0 plays an important role in the discussion.

[H2] For every $L > 0$, there exists a constant c_L such that

$$P\left[\chi_0 \leq r^{-1}\right] \leq \frac{c_L}{r^L}$$

for all $r > 0$.

Define the random field $\mathbb{Z}_n$ on $\mathbb{U}_n$ by

$$\mathbb{Z}_n(u) = \exp\left\{\mathbb{H}_n\left(\theta^* + \frac{1}{\sqrt{n}}u\right) - \mathbb{H}_n(\theta^*)\right\} \tag{3.5}$$

for $u \in \mathbb{U}_n$. Then following the proof of Theorem 3 of Uchida and Yoshida (2013), we see that Condition [H2] together with [H1] implies that for every $L > 0$,

$$P\left[\sup_{u\in V_n(r)} \mathbb{Z}_n(u) \geq e^{-r^{2-\epsilon}}\right] \leq \frac{C_L}{r^L} \qquad (r > 0,\ n \in \mathbb{N}) \tag{3.6}$$

for some constant C_L and some $\epsilon \in (0,1)$. Thus Condition [A1] is fulfilled for $a_n = n^{-1/2}I_{\mathsf{p}\times\mathsf{p}}$ in the present situation.

Let

$$\Gamma(\theta^*)[u,u] = \frac{1}{2T}\int_0^T \mathrm{Tr}\left((\partial_\theta S)S^{-1}(\partial_\theta S)S^{-1}(X_t,\theta^*)[u^{\otimes 2}]\right) dt,$$

Now we have

Theorem 3.1.2. (Theorems 4 and 5 of Uchida and Yoshida (2013)) Suppose that $[H1]$ and $[H2]$ are satisfied. Then, for $A = M$ and B, $\sqrt{n}(\hat{\theta}_n^A - \theta^*) \to^{d_s(\mathcal{F})} \Gamma(\theta^*)^{-1/2}\zeta$ and

$$E\left[\mathsf{f}(\sqrt{n}(\hat{\theta}_n^A - \theta^*))\right] \to \mathbb{E}\left[\mathsf{f}(\Gamma(\theta^*)^{-1/2}\zeta)\right]$$

as $n \to \infty$ for all continuous functions f of at most polynomial growth, where ζ is a p-dimensional standard Gaussian vector independent of $\mathcal{F}$.

Let

$$
\begin{aligned}
\Delta_n[u] &= \frac{1}{\sqrt{n}}\partial_\theta \mathbb{H}_n(\theta^*)[u] \\
&= -\frac{1}{2\sqrt{n}}\sum_{j=1}^{n}\Bigg\{(\partial_\theta \log\det S(X_{t_{j-1}},\theta^*))[u] \\
&\qquad\qquad +h^{-1}(\partial_\theta S^{-1})(X_{t_{j-1}},\theta^*)[u,(\Delta_k Y)^{\otimes 2}]\Bigg\}, \\
\Gamma_n(\theta)[u,u] &= -\frac{1}{n}\partial_\theta^2 \mathbb{H}_n(\theta)[u,u] \\
&= \frac{1}{2n}\sum_{j=1}^{n}\Bigg\{(\partial_\theta^2 \log\det S(X_{t_{j-1}},\theta))[u^{\otimes 2}] \\
&\qquad\qquad +h^{-1}(\partial_\theta^2 S^{-1})(X_{t_{j-1}},\theta))[u^{\otimes 2},(\Delta_k Y)^{\otimes 2}]\Bigg\}
\end{aligned}
$$

and

$$
r_n(u) = \int_0^1 (1-s)\left\{\Gamma(\theta^*) - \Gamma_n(\theta^* + sn^{-1/2}u)\right\}[u,u]ds.
$$

Then, for $u \in \mathbb{R}^{\mathsf{p}}$ and large n, we have

$$
\mathbb{Z}_n(u) = \exp\left(\Delta_n[u] - \frac{1}{2}\Gamma(\theta^*)[u,u] + r_n(u)\right).
$$

Lemma 3.1.3. (Lemma 7 of Uchida and Yoshida (2013)) Assume $[H1]$. Then, for every $q > 0$,

(i) $\sup_{n\in\mathbf{N}} E\left[(\sqrt{n}\,|\Gamma_n(\theta^*) - \Gamma(\theta^*)|)^q\right] < \infty$,

(ii) $\sup_{n\in\mathbf{N}} E\left[\left(\frac{1}{n}\sup_{\theta\in\Theta}|\partial_\theta^3 \mathbb{H}_n(\theta)|\right)^q\right] < \infty$.

Then [A12](iv) is verified by using Lemma 3.1.3 for $\Gamma = \Gamma(\theta^*)$ under the Condition [A6]. It is not difficult to check [A10](iii) and (iv) with stability of the convergence if one follows the proof of Lemma 9 of Uchida and Yoshida (2013). The L^p boundedness of $\{\Delta_n\}_n$ is obvious for all $p > 1$. Almost sure positive definiteness of Γ (i.e., [A10](ii)) and L^p integrability of $\det\Gamma^{-1}$ follow from [H2]. Thus all the conditions in Condition [A10] are satisfied. L^p integrability of Γ is obvious, therefore $|\Gamma^{-1}|$ is L^p integrable, which implies Conditions [A12](ii) and (iii). Thus all the conditions in [A12] are satisfied. Condition [A7'] holds obviously under [A12]. Condtions [A2-6], [A11] and [A13] are fulfilled easily for some $\Psi_n \in \mathrm{GL}(|\mathcal{J}^{(0)}|)$. (See Example 2.3.1 in Section 2.3 for instance.) Consequently, the results in Sections 2.1-2.6 about the penalized estimators for (2.1) are valid.

Condition [H2] can be easily verified if we apply the analytic criterion or the geometric criterion of Uchida and Yoshida (2013).

3.2 Simulation of volatility model

In this section, we report the resutl of the simulation study to check the performance of the variable selection based on our penalized method. The model is a volatility regression model in Section 3.1. Let $\mathsf{p} = d$, $a_n = n^{-1/2} I_{\mathsf{p}\times\mathsf{p}}$, $q < 1$, $\sigma(\theta, x) = \exp\left(\sum_{k=1}^{\mathsf{p}} \theta_k \sin(x_s^k)\right)$ and

$$X_t^k = \int_0^t \frac{\sin(2k\pi s)}{(1 + (X_s^k)^2)} dw_s^k \quad k = (1, \dots, d),$$

where $w^1, \dots, w^d$ are independent standard Brownian motions.

Obviously, Condition [H1] is fulfilled. Following the Section 3.1, we define $\mathbb{H}_n$ by

$$\mathbb{H}_n(\theta) = -\frac{1}{2} \sum_{j=1}^{n} \left\{ \log \det S(X_{t_{j-1}}, \theta) + h^{-1} S^{-1}(X_{t_{j-1}}, \theta)[(\Delta_j Y)^{\otimes 2}] \right\}.$$

Similarly to the proof of Theorems 5 of Uchida and Yoshida (2013)), we have [A1]. Moreover we have [A10] and [A13] as discussed in Section 3.1. Define p by $p(x) = |x|^q$ and ξ_n^j by $\xi_n^j = n^{q'/2}$. By definition, we have [A2-6], [A11] and [A14]. We set $q = 0.3$, $q' = 2/3$, $\mathsf{p} = d = 10$ and $T = 1$. The true value θ^* of an unknown parameter θ is $\theta^* = (0, 1, 0, 1, 2, 0, 1, 1, 1, 0)'$. Four cases of n are considered: $n = 1000, 2000, 3000$ and 10000. We used the local quadratic approximation in Fan and Li (2001) for the optimization of the penalized quasi-likelihood function.

Table 4.1 compares the averages and standard deviations (parentheses) of quasi-maximum likelihood estimator (QMLE) and penalized estimator (p-QL) over 1000 iterations for each cases. Table 4.1 also shows the probability that correct model is selected is selected:

$$P(\hat{\theta}_{n.j} = 0)$$

Table 3.1: Simulation results for the volatility regression model.

	True		1000	2000	3000	10000
		QMLE	0.023(0.260)	-0.006(0.181)	0.007(0.152)	0.001(0.083)
$\hat{\theta}_1$	0	p-QL	-0.005(0.081)	0.002(0.051)	0.000(0.022)	0.000(0)
		prob	0.981	0.989	0.997	1
		QMLE	0.989(0.271)	1.002(0.180)	0.999(0.148)	1.000(0.085)
$\hat{\theta}_2$	1	p-QL	0.792(0.396)	0.893(0.244)	0.937(0.160)	0.972(0.077)
		prob	0.869	0.971	0.996	1
		QMLE	0.003(0.263)	-0.004(0.181)	-0.009(0.146)	0.007(0.084)
$\hat{\theta}_3$	0	p-QL	0.003(0.088)	0.000(0.030)	-0.002(0.042)	0.000(0)
		prob	0.982	0.992	0.995	1
		QMLE	0.986(0.263)	1.007(0.176)	1.003(0.146)	0.998(0.083)
$\hat{\theta}_4$	1	p-QL	0.808(0.392)	0.898(0.241)	0.932(0.160)	0.968(0.073)
		prob	0.88	0.977	0.993	1
		QMLE	1.997(0.258)	2.000(0.184)	2.006(0.148)	1.996(0.084)
$\hat{\theta}_5$	2	p-QL	1.912(0.328)	1.940(0.231)	1.965(0.139)	1.980(0.076)
		prob	0.999	0.999	1	1
		QMLE	-0.010(0.272)	0.001(0.185)	-0.004(0.155)	-0.001(0.083)
$\hat{\theta}_6$	0	p-QL	0.001(0.123)	-0.003(0.059)	0.000(0.028)	0.000(0)
		prob	0.968	0.989	0.995	1
		QMLE	0.997(0.262)	0.996(0.179)	0.998(0.153)	0.998(0.081)
$\hat{\theta}_7$	1	p-QL	0.789(0.390)	0.892(0.246)	0.927(0.172)	0.971(0.078)
		prob	0.867	0.967	0.992	1
		QMLE	1.000(0.267)	0.991(0.187)	1.002(0.150)	1.000(0.818)
$\hat{\theta}_8$	1	p-QL	0.811(0.399)	0.883(0.248)	0.936(0.162)	0.972(0.076)
		prob	0.881	0.968	0.995	1
		QMLE	1.006(0.269)	0.997(0.182)	1.002(0.152)	0.999(0.083)
$\hat{\theta}_9$	1	p-QL	0.788(0.394)	0.891(0.241)	0.937(0.167)	0.973(0.071)
		prob	0.871	0.971	0.994	1
		QMLE	0.022(0.264)	-0.008(0.181)	0.007(0.145)	0.000(0.081)
$\hat{\theta}_{10}$	0	p-QL	0.000(0.089)	-0.004(0.048)	0.000(0.027)	0.000(0.007)
		prob	0.975	0.986	0.993	0.998
		Under model	0.91	0.958	0.98	0.998
Total		Over model	0.606	0.9	0.979	1
		True model	0.591	0.878	0.962	0.998

for $j \in \mathcal{J}^{(0)}$ and

$$P(\hat{\theta}_{n.j} \neq 0)$$

for $j \in \mathcal{J}^{(1)}$. Under model is the probability that the estimator selects an under model:

$$P\Big(\{j; \hat{\theta}_{n,j} = 0\} \supset \mathcal{J}^{(0)}\Big)$$

and Over model is the probability that the estimator selects an over model:

$$P\Big(\{j; \hat{\theta}_{n,j} = 0\} \subset \mathcal{J}^{(0)}\Big).$$

True model is the probability that the true model is selected:

$$P\Big(\{j; \hat{\theta}_{n,j} = 0\} = \mathcal{J}^{(0)}\Big).$$

From Table 4.1, it can seen that when sample size n is large , penalized method performs variable selection very well. Moreover, the bias of non-zero parameters decrease as the sample size increases.

Chapter 4

Volatility estimation with a global jump filter

4.1 Volatility with a jump global filter

Given a stochastic basis $(\Omega, \mathcal{F}, \mathbf{F}, P)$ with a filtration $\mathbf{F} = (\mathcal{F}_t)_{t\in[0,T]}$, we suppose that an m-dimensional semimartingale $Y = (Y_t)_{t\in[0,T]}$

$$Y_t = Y_0 + \int_0^t b_s ds + \int_0^t \sigma(X_s, \theta) dw_s + J_t, t \in [0, T]. \tag{4.1}$$

depending on an unknown parameter θ in the closure of a bounded open set Θ in $\mathbb{R}^{\mathsf{p}}$. Here $w = (w_t)_{t\in[0,T]}$ is an r-dimensional standard $\mathbf{F}$-Wiener process, and the diffusion coefficient $\sigma : \mathbb{R}^{\mathsf{d}} \times \overline{\Theta} \to \mathbb{R}^{\mathsf{m}} \otimes \mathbb{R}^{\mathsf{r}}$ is a continuous function. The process $b = (b_t)_{t\in[0,T]}$ is an m-dimensional progressively measurable process. $J = (J_t)_{t\in[0,T]}$ is the jump part of Y with $J_0 = 0$. In this thesis, we will assume that J is finitely activate in that $\sum_{t\in[0,T]} 1_{\{\Delta J_t \neq 0\}} < \infty$ a.s. and $J_t = \sum_{s\in[0,t]} \Delta_s Y_s$, where $\Delta Y_s = Y_s - Y_{s-}$ and $\Delta Y_0 = 0$. The model (4.1) is a stochastic regression model with a covariate process $X = (X_t)_{t\in[0,T]}$ that is supposed to be a d-dimensional cádlág adapted process.

We will discuss estimation of the value of the unknown parameter θ based on the high frequency data $(X_{t_j}, Y_{t_j})_{j=0,1,\ldots,n,} t_j = T_j^n = jT/n$. In estimation, any structure of b is unknown and the data of b is not available. The jump part J is also structurally unknown and unobservable. In this situation, we cannot tell each increment $\Delta_j Y = Y_{t_j} - Y_{t_{j-1}}$ has jumps or not. Estimation of the volatility parameter θ is usually carried out by using a weighted sum of $(\Delta_j Y)^{\otimes 2}$. However, the existence of jumps ΔJ severely biases the estimate. To avoid the effects of contamination by ΔJ, threshold methods were developed; see e.g. Shimizu and Yoshida (2006) and Ogihara and Yoshida (2011). They proved asymptotic optimality of their estimators for the volatility parameter. However, in practice, it is known that the performance of a threshold method heavily depends on a choice of the tuning parameter involved in the threshold. The same kind fault appears even if the threshold is tuned by an estimator of the spot volatility. It is because, for optimal estimation, the threshold should be set at a relatively high level to catch all Brownian increments, and it causes incorrect acceptance of non-negligible jumps. These threshold methods are local filters, since the threshold for $\Delta_j Y$ is determined by the data in a local neighborhood of t_j. Differently from the local filters, Inatsugu and Yoshida (2018) recently proposed new jump filters. Their filters use a kind of order statistics of $\{\Delta_j Y\}_{j=1,\cdots,n}$. Due to time-global dependency of the filter, the martingale structure in the likelihood function is destroyed but they showed it can be re-caputured asymptotically and proved asymptotic optimality of the estimator. The resulting threshold becomes smaller because it is determined by the data endogenously.

The aim of this section is to formulate the quasi-likelihood analysis for sparse estimation of volatility with the global jump filters.

Some of the components of the jump part can vanish in some situations by nature of the phenomena. Then we do not need filtering jumps for such components, and component-wise construction of jump filters is necessary. It will be assumed that σ is a block matrix:

$$\sigma = \mathrm{diag}[\sigma^{(1)}(x,\theta),\dots,\sigma^{(\mathsf{k})}(x,\theta)] \tag{4.2}$$

where $\sigma^{(k)}(x,\theta)$ is a $\mathsf{m}_k \times \mathsf{m}_k$ matrices, $k = 1,\dots,\mathsf{k}$. Let

$$S(x,\theta) = \mathrm{diag}[S^{(1)}(x,\theta),\dots,S^{(\mathsf{k})}(x,\theta)]$$

for the $\mathsf{m}_k \times \mathsf{m}_k$ matrices $S^{(k)}(x,\theta) = \sigma^{(k)}(\sigma^{(k)})'(x,\theta)$, $k = 1,\dots,\mathsf{k}$. According to the decomposition (4.2), we denote $v = (v^{(k)})_{k=1,\dots,\mathsf{k}}$ for r-dimensional matrix v with $\mathsf{r} = \sum_k \mathsf{m}_k = \mathsf{m}$, and

$$Y_t = \begin{pmatrix} Y_t^{(1)} \\ \vdots \\ Y_t^{(\mathsf{k})} \end{pmatrix}, \quad b_t = \begin{pmatrix} b_t^{(1)} \\ \vdots \\ b_t^{(\mathsf{k})} \end{pmatrix}, \quad w_t = \begin{pmatrix} w_t^{(1)} \\ \vdots \\ w_t^{(\mathsf{k})} \end{pmatrix}, \quad J_t = \begin{pmatrix} J_t^{(1)} \\ \vdots \\ J_t^{(\mathsf{k})} \end{pmatrix}.$$

4.1.1 The global jump filter and the quasi-likelihood function

For estimation of θ, we will use the quasi-log likelihood function $\mathbb{H}_n(\theta)$ defined by

$$\mathbb{H}_n(\theta) = -\frac{1}{2}\sum_{k=1}^{\mathsf{k}} \sum_{j\in\mathcal{K}_n^{(k)}} \Big\{ (q_n^{(k)})^{-1} h^{-1} S^{(k)}(X_{t_{j-1}},\theta)^{-1}[(\Delta_j Y^{(k)})^{\otimes 2}] K_{n,j}^{(k)}$$

$$+ (p_n^{(k)})^{-1} \log\det S^{(k)}(X_{t_{j-1}}, \theta)\Big\} \tag{4.3}$$

when all $S^{(k)}(T_{t_{j-1}})$ are almost surely invertible. The index sets $\mathcal{K}_n^{(k)}$ in $\{1, \dots, n\}$ select the indices j for which we judge $J^{(k)}$ did not jump on the interval $(t_{j-1}, t_j]$. These index sets form our jump filter. The function $\mathbb{H}_n$ has scales $q_n^{(k)}$ and $p_n^{(k)}$ to avoid the bias caused by removing data by $\mathcal{K}_n^{(k)}$ in $\{1, \dots, n\}$. However, at least theoretically, we do not need to be nervous about choices of $q_n^{(k)}$ and $p_n^{(k)}$. In fact, the forthcoming asymptotic theory is valid even for $q_n^{(k)} = p_n^{(k)} = 1$. We will return to this point in Remark 4.1.3.

The factors $K_{n,j}^{(k)}$ are defined by

$$K_{n,j}^{(k)} = 1_{\left\{|\Delta_j Y^{(k)}| < C_*^{(k)} n^{\frac{1}{4}}\right\}} \tag{4.4}$$

where $C_*^{(k)}$ are arbitrarily given positive constants. The threshold of each $K_{n,j}^{(k)}$ is much larger than ordinary thresholds used in existing local jump filters. As a matter of fact, the factors $K_{n,j}^{(k)}$ are not necessary if we know boundedness of moments of ΔJ_t. Thanks to $K_{n,j}^{(k)}$, we can remove such a condition we cannot verify in practice.

Now our concern is how to construct the global filter $\mathcal{K}_n^{(k)}$. Fix $\delta_1^{(k)} \in (0, 1/2)$ for $k = 1, \dots, \mathsf{k}$. Let $s_n^{(k)} = n - B^{(k)} \lfloor n^{\delta_1^{(k)}} \rfloor$ with positive constants $B^{(k)}$. In what follows, we will only consider sufficiently large n. Moreover, let $\mathfrak{S}_{n,j-1}^{(k)}$ be an $\mathsf{m} \times \mathsf{m}$ positive definite symmetric random matrix. Let

$$V_j^{(k)} = |(\mathfrak{S}_{n,j-1}^{(k)})^{-1/2} \Delta_j Y^{(k)}|.$$

We denote by $V_{(j)}^{(k)}$ the j-th order statistic of $V_1^{(k)}, \dots, V_n^{(k)}$. Then the global filter we will work with is defined by

$$\mathcal{K}_n^{(k)} = \{j \in \{1, \dots, n\}; V_j^{(k)} < V_{(s_n^{(k)})}^{(k)}\}.$$

Let

$$\alpha_n = (\alpha_n^{(1)}, \dots, \alpha_n^{(\mathsf{k})}) \quad \text{with} \quad \alpha_n^{(k)} = 1 - s_n^{(k)}/n. \tag{4.5}$$

Then the numbers $(s_n^{(k)})_{k=1,\dots,\mathsf{k}}$ are determined by α_n. We can use α_n to specify the jump filter $\mathcal{K}_n^{(k)}$.

Remark 4.1.1. The filter $\mathcal{K}_n^{(k)}$ uses all the increments $\{\Delta_j Y\}_{j=1,\dots,n}$. In this sense, we call it a global jump filter. The function $\mathbb{H}_n(\theta)$ has lost the martingale structure due to $\mathcal{K}_n^{(k)}$. This makes analysis harder but we can asymptotically recover the martingale property in the score function. See Inatsugu and Yoshida (2018) for details.

Remark 4.1.2. The factor $K_{n,j}^{(k)}$ stabilizes the effect of $\Delta_j Y$. It is not for filtering. The functional $\mathring{K}_{n,j}^{(k)}$ is defined by (4.4). On the other hand, it is possible to use

$$K_{n,j}^{(k)} = 1_{\left\{V_j^{(k)} < C_*^{(k)} n^{-\frac{1}{4}-\delta_0}\right\}}$$

for constants $\delta_0 \in (0, 1/4)$ and $C_*^{(k)} \in (0, \infty)$.

Remark 4.1.3. We only require that $(\mathfrak{S}_{n,j-1}^{(k)})^{-1}$ is uniformly bounded in $L^{\infty-}$ and that both $q_n^{(k)}$ and $p_n^{(k)}$ are sufficiently close to 1. Therefore, we may choose them as $q_n^{(k)} = q_n^{(k)} = 1$ and $\mathfrak{S}_{n,j-1}^{(k)} = I_{\mathsf{m}_\mathsf{k}}$. In this special case, $\mathbb{H}_n(\theta)$ becomes simply $\mathring{\mathbb{H}}_n$ defined by

$$\mathring{\mathbb{H}}_n(\theta) = -\frac{1}{2}\sum_{k=1}^{\mathsf{k}} \sum_{j \in \mathcal{K}_n^{(k)}} \Big\{ h^{-1} S^{(k)}(X_{t_{j-1}}, \theta)[(\Delta_j Y^{(k)})^{\otimes 2}] K_{n,j}^{(k)} + \log\det S^{(k)}(X_{t_{j-1}}, \theta) \Big\}$$

with $\mathcal{K}_n^{(k)}$ for $V_j^{(k)} = |\Delta_j Y^{(k)}|$. Since $\mathring{\mathbb{H}}_n$ satisfies the conditions specified later, for the estimators with respect to $\mathring{\mathbb{H}}_n$, the same asymptotic results hold as those of $\mathbb{H}_n$.

On the other hand, it is also natural to take some local volatility estimator for $\mathfrak{S}_{n,j-1}^{(k)}$ since $V_j^{(k)}$ compares $(\mathfrak{S}_{n,j-1}^{k})^{-1/2}$ and $\Delta_j Y^{(k)}$. For $\alpha = (\alpha^{(k)})_{k \in \{1,\dots,\mathsf{k}\}} \in [0,1)^{\mathsf{k}}$, let $p(\alpha^{(k)}) = 1 - \alpha^{(k)}$ and

$$q^{(k)}(\alpha^{(k)}) = \frac{\mathrm{Tr}\Big(\int_{|z| \le c(\alpha^{(k)})^{1/2}} z^{\otimes 2} \phi(z; 0, I_{\mathsf{m}_\mathsf{k}}) dz\Big)}{\mathrm{Tr}\Big(\int_{\mathbb{R}^\mathsf{m}} z^{\otimes 2} \phi(z; 0, I_{\mathsf{m}_\mathsf{k}}) dz\Big)}$$

where $\phi(z; \mu, C)$ denotes the density function of the multi-dimensional normal distribution with mean vector μ and covariance matrix C. Then another possibility for $q_n^{(k)}$ and $p_n^{(k)}$ is to use $q_n^{(k)} = q^{(k)}(\alpha_n^{(k)})$ and $p_n^{(k)} = p^{(k)}(\alpha_n^{(k)})$ where $\alpha_n^{(k)}$ are given by (4.5)

4.1.2 Regularity conditions

To develop asymptotic theory, we will need some notation and regularity conditions. Denote by $||V||_p = (E[|V|^p])^{1/p}$ the L^p-norm of a vector-valued random variable V for $p > 0$. Write $L^{\infty-} = \cap_{p>1} L^p$. Let $N_t^X = \sum_{s \le t} 1_{\{\Delta X_s^{(k)} \neq 0\}}$, $N_t^{(k)} = \sum_{s \le t} 1_{\{\Delta J_s^{(k)} \neq 0\}}$ and $N_t = \sum_{s \le t} 1_{\{\Delta J_s \neq 0\}}$. In [C1]$_\kappa$ below, we will impose the condition that $N^X_T < \infty$ almost surely. Then the jump part J^X of X is written as $J^X = \sum_{s\le t} \Delta X_s$. Let $\widetilde{X} = X - J^X$ for $J^X = \sum_{s\in[0,\cdot]} \Delta X_s$.

We consider the following regularity conditions.

[C1]$_\kappa$ **(i)** For every $p > 1, \sup_{t\in[0,T]} ||X_t||_p < \infty$ and there exists a constant $C(p)$ such that

$$||\widetilde{X}_t - \widetilde{X}_s||_p \le C(p)|t-s|^{1/2} \quad (t, s \in [0,T]).$$

(ii) $\sup_{t\in[0,T]} ||b_t||_p < \infty$ for every $p > 1$.

(iii) $\sigma \in C^{2,\kappa}_{\uparrow}(\mathbb{R}^{\mathsf{d}} \times \Theta; \mathbb{R}^{\mathsf{m}} \times \mathbb{R}^{\mathsf{r}}), S(X_t, \theta)$ is invertible a.s. for every $(t, \theta) \in [0,T] \times \Theta$

(iv) $N_t + N^X_T \in L^{\infty-}$.

[C2] **(i)** $\mathfrak{S}^{(k)}_{n,j-1}(k \in \{1,\dots,\mathsf{k}\}, n \in \mathbb{N}, j \in \{1,\dots,n\})$ are positive-definite measurable random matrices and satisfy

$$\sup_{\substack{k\in\{1,\dots,\mathsf{k}\}\\ n\in\mathbb{N}, j\in\{1,\dots,n\}}} \Big(||\mathfrak{S}^{(k)}_{n,j-1}||_p + ||(\mathfrak{S}^{(k)}_{n,j-1})^{-1}||_p \Big) < \infty$$

for every $p > 1$.

(ii) $q^{(k)}_n$ and $p^{(k)}_n$ are positive numbers satisfying $|q^{(k)}_n - 1| = o(n^{-1/2})$ and $|1 - p^{(k)}_n| = o(n^{-1/2})$.

For consistent estimation, we need an identifiability condition. Our argument will provide more precise modes of convergence (e.g. convergence of the moments of the estimation error) than just consistency. Then a certain quantitative estimate of identifiability is necessary. The following key index χ_0 serves to it:

$$\chi_0 = \inf_{\theta\ne\theta^*} \frac{-\mathbb{Y}(\theta)}{|\theta-\theta^*|^2}$$

where

$$\mathbb{Y}(\theta) = -\frac{1}{2T}\sum_{k=1}^{\mathsf{k}} \int_0^T \left\{ \mathrm{Tr}\Big(S^{(k)}(X_t,\theta)^{-1} S^{(k)}(X_t,\theta^*) - I_{\mathsf{m}_\mathsf{k}}\Big) + \log\frac{\det S^{(k)}(X_t,\theta)}{\det S^{k)}(X_t,\theta^*)} \right\} dt.$$

Since the degree of separation of the model is random, we need a stochastic estimate of identifiability, through χ_0.

[C3] For every positive number L, there exists a constant C_L such that

$$P[\chi_0 < r^{-1}] \le C_L r^{-L} \quad (r > 0)$$

Remark 4.1.4. When X is a non-degenerate diffusion process, Uchida and Yoshida (2013) gave some easily verifiable criteria are known for Condition [C3]. For further information, see Remark 2.11 of Inatsugu and Yoshida (2018).

Remark 4.1.5. Conditions $[C1]_\kappa$, [C2] and [C3] are Conditions $[F1]_\kappa$, [F2'] and [F3] of Inatsugu and Yoshida (2018), respectively.

Suppose that the parameter space Θ admits Sobolev's embedding inequality

$$\sup_{\theta\in\Theta}|f(\theta)| \le C_{\Theta,p}\bigg\{\sum_{i=0}^{1}\int_\Theta |\partial_\theta^i f(\theta)|^p d\theta\bigg\}^{1/p} \quad (f \in C^1(\Theta))$$

where $C_{\Theta,p}$ is a constant and $p > \mathsf{p}$. For example, this inequality holds if Θ has a Lipschitz boundary. We are assuming that the diffusion coefficient σ is continuous on $\mathbb{R}^{\mathsf{d}} \times \overline{\Theta}$.

4.1.3 Quasi-likelihood analysis based on $\mathbb{H}_n$: polynomial type large deviation

Define the quasi-likelihood ratio random field $\mathbb{Z}_n$ by

$$\mathbb{Z}_n(u) = \exp\Big\{\mathbb{H}_n(\theta^* + n^{-1/2}u) - \mathbb{H}_n(\theta^*)\Big\} \quad (u \in \mathbb{U}_n)$$

where $\mathbb{U}_n = \{u \in \mathbb{R}^{\mathsf{p}}; \theta^* + n^{-1/2}u \in \Theta\}$.

Let $\mathbb{V}_n(r) = \{u \in \mathbb{U}_n; |u| \ge r\}$. The following result is given in Theorem 3.3 of Inatsugu and Yoshida (2018).

Theorem 4.1.6. Suppose that $[C1]_4$, [C2] and [C3] are satisfied. Then, for every $c_0 \in (1,2)$ and every positive number L, there exist a constant $C(c_0, L)$ such that

$$P\Big[\sup_{u\in\mathbb{V}_n(r)} \mathbb{Z}_n(u) \ge e^{-r^{c_0}}\Big] \le \frac{C(c_0,L)}{r^L}$$

for all $r > 0$ and $n \in \mathbb{N}$.

The quasi-maximum likelihood estimator (QMLE) $\hat{\theta}_n^{M,\alpha_n}$ for θ is any sequence of measurable mappings from Ω to $\overline{\Theta}$ satisfying

$$\mathbb{H}_n(\hat{\theta}_n^{M,\alpha_n}) = \max_{\theta\in\overline{\Theta}} \mathbb{H}_n(\theta).$$

Such a measurable mapping always exists by the measurable mapping theorem. We write "α" in the symbol of the QMLE in order to emphasize the dependency of it on α_n determining the global filter.

The quasi-Bayesian estimator (QBE) $\hat{\theta}_n^{B,\alpha_n}$ for θ is defined by

$$\hat{\theta}_n^{B,\alpha_n} = \left[\int_\Theta \exp(\mathbb{H}_n(\theta))\varpi(\theta)d\theta\right]^{-1} \int_\Theta \theta \exp(\mathbb{H}_n(\theta))\varpi(\theta)d\theta,$$

where ϖ is a continuous function on Θ such that $0 < \inf_{\theta\in\Theta}\varpi(\theta) \leq \sup_{\theta\in\Theta}\varpi(\theta) < \infty$. As a consequence of PLDI in Theorem 4.1.6, the L^p-boundedness of the estimators follows. We denote

$$\hat{u}_n^{\mathsf{A},\alpha_n} = \sqrt{n}(\hat{\theta}_n^{\mathsf{A},\alpha_n} - \theta^*)$$

for $\mathsf{A} = M, B$.

Proposition 4.1.7. Suppose that [C1]$_4$, [C2] and [C3] are satisfied. Then

$$\sup_{n\in\mathbb{N}} ||\hat{u}_n^{\mathsf{A},\alpha_n}||_p < \infty \quad (\mathsf{A} = M, B)$$

for every $p > 1$.

4.1.4 Quasi-likelihood analysis based on $\mathbb{H}_n$: limit theorem and convergence of moments

We will consider the situation where the process X admits a representation

$$X_t = X_0 + \int_0^t \tilde{b}_s ds + \int_0^t \tilde{a}_s d\tilde{w}_s + J_t^X \quad (t \in [0,T]) \tag{4.6}$$

where $J^X = (J_t^X)_{t\in[0,T]}$ is a càdlàg adapted pure jump process, $\tilde{w} = (\tilde{w}_t)_{t\in[0,T]}$ is an r_1-dimensional $\mathbf{F}$-Wiener process, $\tilde{b} = (\tilde{b}_t)_{t\in[0,T]}$ is a d-dimensional càdlàg adapted process and $\tilde{a} = (\tilde{a}_t)_{t\in[0,T]}$ is an $\mathbb{R}^{\mathsf{d}}\times\mathbb{R}^{\mathsf{r}_1}$-valued progressively measurable processes such that

$$||X_0||_p + \sup_{t\in[0,T]} (||\tilde{b}_t||_p + ||\tilde{a}_t||_p + ||J_t^X||_p) < \infty \tag{4.7}$$

for every $p > 1$. The Wiener process $\tilde{w}$ is possibly correlated with w.

We strengthen Condition [C1] as follows.

[C1′] The process X admits a representation (4.6)-(4.7) in addition to Conditions (ii), (iii) and (iv) of [C1]$_\kappa$.

Define the $\mathsf{p} \times \mathsf{p}$ symmetric matrix $\Gamma^{(k)}$ by

$$\Gamma^{(k)}[u^{\otimes 2}] = \frac{1}{2T}\int_0^T \mathrm{Tr}\Big((\partial_\theta S^{(k)}[u])(S^{(k)})^{-1}(\partial_\theta S^{(k)}[u])(S^{(k)})^{-1}(X_t,\theta^*)\Big)dt,$$

where $u \in \mathbb{R}^{\mathsf{p}}$, and Γ by $\Gamma = \sum_{k=1}^{\mathsf{k}} \Gamma^{(k)}$.

On an extension $(\overline{\Omega}, \overline{\mathcal{F}}, \overline{\mathbf{F}})$ of $(\Omega, \mathcal{F}, \mathbf{F})$, we make a p-dimensional standard Gaussian random vector ζ independent of $\mathcal{F}$. Then Theorem 3.13 of Inatsugu and Yoshida (2018) is rephrased:

Theorem 4.1.8. Suppose that $[C1']_4$, [C2] and [C3] are fulfilled. Then $\hat{u}_n^{\mathsf{A},\alpha_n} \to^d \Gamma^{-1/2}\zeta$ ($\mathcal{F}$-stably) and

$$E[f(\hat{u}_n^{\mathsf{A},\alpha_n}] \to E[f(\Gamma^{-1/2}\zeta)]$$

as $n \to \infty$ for any continuous function f of at most polynomial growth for $\mathsf{A} \in \{M, B\}$.

4.1.5 Sparse estimation

We now focus on sparse estimation of volatility parameter for a process with jumps. To apply the results in Sections 2.1-2.6, we will use n for T of Section 2.1, while T denotes the fixed terminal of the observations.

We will consider the penalized logarithmic quasi-likelihood function

$$\mathbb{H}_n^{\dagger}(\theta) = \mathbb{H}_n(\theta) - p_n(\theta)$$

where $\mathbb{H}_n$ is given in (4.3) and the penalty term p_n is given in (2.2).

The penalized quasi-maximum likelihood estimator (penalized QMLE) is any $\overline{\Theta}$-valued mapping $\hat{\theta}_n$ that is a measurable function of the data such that

$$\mathbb{H}_n^{\dagger}(\hat{\theta}_n) = \max_{\theta \in \overline{\Theta}} \mathbb{H}_n^{\dagger}(\theta).$$

Let $a_n = n^{-1/2} I_{\mathsf{m}}$, then from Theorem 4.1.6, we can derive [A1].

Proposition 4.1.9. Suppose that $[C1]_4$,[C2] and [C3] are satisfied. Then for every positive number L, Condition [A1] holds.

Moreover, we can apply Theorem 2.2.3 to our penalized QMLE $\hat{\theta}_n$.

Theorem 4.1.10. Suppose that Condition [A2]-[A4], $[C1]_4$ and [C2]-[C3] are satisfied. Then for every positive number L, there exist constants $C_L > 0$ and $\varepsilon_L \in (0, 1)$ such that

$$P\left[\sup_{u \in \mathbb{V}_n(r)} \mathbb{Z}_n(u) \geq \exp(-r^{2-\varepsilon_L})\right] \leq \frac{C_L}{r^L}$$

for all $r > 0, T > 0$.

Let $\hat{u}_n = a_n^{-1}(\hat{\theta}_n - \theta^*) = \sqrt{n}(\hat{\theta}_m - \theta^*)$. The L^p-bounded ness of $\hat{u}_n$ is a simple consequence from the above theorem.

Corollary 4.1.11. Under the conditions of Theorem 4.1.10,

$$\sup_{n\in\mathbb{N}} ||\hat{u}_n||_p < \infty$$

for every $p > 1$.

4.1.6 Selection consistency and limit theorem

Let

$$\widetilde{\mathbb{H}}_n(\theta) = -\frac{1}{2}\sum_{k=1}^{\mathsf{k}}\sum_{j=1}^{n}\Big\{h^{-1}S^{(k)}(X_{t_{j-1}},\theta)^{-1}[(\Delta_j\widetilde{Y}^{(k)})^{\otimes 2}] + \log\det S^{(k)}(X_{t_{j-1}},\theta)\Big\}$$

where $\widetilde{Y}^{(k)} = Y^{(k)} - J^{(k)}$. The following lemma gives a key estimate of the effect in replacing the sums in j. See Lemma 3.5 of Inatsugu and Yoshida (2018)

Lemma 4.1.12. Suppose that [C1]$_4$, [C2] and [C3] are satisfied. Then

$$\sum_{i=0}^{4}\sup_{\theta\in\Theta}\Big|\Big|n^{-1/2}\partial_\theta^i\mathbb{H}_n(\theta) - n^{-1/2}\partial_\theta^i\widetilde{\mathbb{H}}_n(\theta)\Big|\Big|_p \to 0$$

as $n \to \infty$ for every $p \geq 1$.

Lemma 4.1.13. Suppose that [C1]$_4$, [C2] and [C3] are satisfied. Then

$$\sum_{i=0}^{3}\sup_{n\in\mathbb{N}} E\Big[\Big(n^{-1}\sup_{\theta\in\Theta}|\partial_\theta^i\mathbb{H}_n(\theta)| < \infty\Big)^p\Big]$$

for every $p > 1$.

Proof. By sovolev's inequality and Lemma 4.1.12, it suffices to show

$$\sum_{i=0}^{4}\sup_{\theta\in\Theta}||n^{-1}\partial_\theta^i\widetilde{\mathbb{H}}_n(\theta)||_p < \infty \tag{4.8}$$

for every $p > 1$. But it is easy to verify (4.8). □

Let

$$\Delta_n = n^{-1/2}\partial_\theta\mathbb{H}_n(\theta^*) \quad \text{and} \quad \Gamma_n = -n^{-1}\partial_\theta^2\mathbb{H}_n(\theta^*).$$

Lemma 4.1.14. Suppose that $[C1]_4$, [C2] and [C3] are satisfied. Then $\sup_{n\in\mathbb{N}}||\Delta_n||_p < \infty$ for every $p > 1$.

Proof. Let

$$\widetilde{\Delta}_n = n^{-1/2}\partial_\theta\widetilde{\mathbb{H}}_n(\theta^*) = \frac{1}{\sqrt{n}}\sum_{k=1}^{\mathsf{k}}\sum_{j=1}^{n} f_{t_{j-1}}[D_j^{(k)}]$$

where

$$f_{t_{j-1}} = \frac{1}{2}((S^{(k)})^{-1}(\partial_\theta S^{(k)})(S^{(k)})^{-1})(X_{t_{j-1}},\theta^*)$$

and

$$D_j^{(k)} = h^{-1}(\Delta_j\widetilde{Y}^{(k)})^{\otimes 2} - S^{(k)}(X_{t_{j-1}},\theta^*).$$

By Lemma 4.1.12, it sufficient to verify

$$\sup_{n\in\mathbb{N}} E[|\widetilde{\Delta}_n|^p] < \infty. \tag{4.9}$$

Since $N_T^X \in L^{\infty-}$ and

$$\Big|\Big|\max_{j=1,\dots,n}|f_{t_{j-1}}[D_j^{(k)}]|\Big|\Big|_p = O(n^{1/4})$$

for every $p > 1$, we see

$$\Big|\Big|n^{-1/2}\sum_{j=1}^{n} f_{t_{j-1}}[D_j^{(k)}]\Big|\Big|_p = \Big|\Big|n^{1/2}\sum_{j=1}^{n} 1_{\{\Delta_j N^X=0\}} f_{t_{j-1}}[D_j^{(k)}]\Big|\Big|_p + o(1)$$

for every $p > 1$. Thus we need to show

$$\Big|\Big|n^{1/2}\sum_{j=1}^{n} 1_{\{\Delta_j N^X=0\}} f_{t_{j-1}}[D_j^{(k)}]\Big|\Big| = O(1) \tag{4.10}$$

as $T \to \infty$ for every $p > 1$.

Fix k. We have

$$1_{\{\Delta_j N^X=0\}}\Delta_j\widetilde{Y}^{(k)} = 1_{\{\Delta_j N^X=0\}}(\xi_{1,j} + \xi_{2,j} + \xi_{3,j})$$

with

$$\xi_{1,j} = \sigma^{(k)}(X_{t_{j-1}},\theta^*)\Delta_j w^{(k)},$$

$$\xi_{2,j} = \int_{t_{j-1}}^{t_j} \{\sigma^{(k)}(X_{t_{j-1}} + \widetilde{X}_t - \widetilde{X}_{t_{j-1}}, \theta^*) - \sigma(k)(X_{t_{j-1}}, \theta^*)\} dw_t^{(k)},$$
$$\xi_{3,j} = \int t_{j-1}{}^{t_j} b_t^{(k)} dt.$$

Let

$$C(x,y) = \left| \int_0^1 \partial_x \sigma^{(k)}(x + r(y-x), \theta^*) dr \right|.$$

Then by the same reason as in (4.10), and by Itô's formula and the Burkholder-Davis-Gundy inequality, we obtain

$$\begin{aligned}
&\Big|\Big| n^{-1/2} \sum_{j=1}^n 1_{\{\Delta_j N^X = 0\}} h^{-1} f_{t_{j-1}}[\xi_{1,j} \otimes \xi_{2,j}] \Big|\Big|_p \\
=&\Big|\Big| n^{-1/2} \sum_{j=1}^n h^{-1} f_{t_{j-1}}[\xi_{1,j} \otimes \xi_{2,j}] \Big|\Big|_p + o(1) \\
\lesssim&\Big|\Big| n^{-1/2} \sum_{j=1}^n h^{-1} f_{t_{j-1}} \big|\sigma^{(k)}(X_{t_{j-1}}, \theta)\big| \times \\
&\qquad \int_{t_{j-1}}^{t_j} \big|\sigma^{(k)}(X_{t_{j-1}} + \widetilde{X}_t - \widetilde{X}_{t_{j-1},\theta^*}) - \sigma^{(k)}(X_{t_{j-1}}, \theta^*)\big| dt \Big|\Big|_p + O(1) \\
\lesssim&\Big|\Big| n^{-1/2} \sum_{j=1}^n h^{-1} f_{t_{j-1}} \big|\sigma^{(k)}(X_{t_{j-1}}, \theta)\big| \times \\
&\qquad \int_{t_{j-1}}^{t_j} C(X_{t_{j-1}}, \widetilde{X}_t - \widetilde{X}_{t_{j-1}}) \big|\widetilde{X}_t - \widetilde{X}_{t_{j-1}}\big| dt \Big|\Big|_p + O(1) \\
\lesssim& n^{-1/2} \sum_{j=1}^n h^{-1} \int_{t_{j-1}}^{t_j} \Big|\Big| |f_{t_{j-1}}| \big|\sigma^{(k)}(X_{t_{j-1}}, \theta)\big| \times \\
&\qquad C(X_{t_{j-1}}, \widetilde{X}_t - \widetilde{X}_{t_{j-1}}) \big|\widetilde{X}_t - \widetilde{X}_{t_{j-1}}\big| \Big|\Big|_p dt + O(1) \\
\lesssim& n^{-1/2} \sum_{j=1}^n \sup_{t \in [t_{j-1}, t_j]} ||\widetilde{X}_t - \widetilde{X}_{t_{j-1}}||_{2p} \times \\
&\qquad \sup_{\substack{t \in [t_{j-1}, t_j] \\ j=1,\ldots,n}} \Big|\Big| |f_{t_{j-1}}| \big|\sigma^{(k)}(X_{t_{j-1}}, \theta^*)\big| C(X_{t_{j-1}}, \widetilde{X}_t - \widetilde{X}_{t_{j-1}}) \Big|\Big|_{2p} + O(1) \\
=&O(1)
\end{aligned}$$

for $p > 1$ since $||\widetilde{X}_t - \widetilde{X}_{t_{j-1}}||_{2p} \leq C(2p) n^{-1/2}$ and $\sup_{t \in [0,T]} ||X_t||_p + \sup_{t \in [0,T]} ||\widetilde{X}_t||_p < \infty$ by the continuity of the mapping $t \mapsto \widetilde{X}_t \in L^p$ for every $p > 1$. In a similar manner, we obtain

$$\left\|n^{-1/2}\sum_{j=1}^{n}1_{\{\Delta_j N^X=0\}}h^{-1}f_{t_{j-1}}[\xi_{i_1,j}\otimes\xi_{i_2,j}]\right\|_p=O(1)$$

for every $p>1$ and $(i_1,i_2)\in\{1,2,3\}^2\setminus\{(1,1)\}$. For $(i_1,i_2)=(1,1)$,

$$\begin{aligned}&\left\|n^{-1/2}\sum_{j=1}^{n}1_{\{\Delta_j N^X=0\}}f_{t_{j-1}}[h^{-1}\xi_{1,j}\otimes\xi_{1,j}-S^{(k)}(X_{t_{j-1}},\theta^*)]\right\|_p\\ =&\left\|n^{-1/2}\sum_{j=1}^{n}f_{t_{j-1}}[h^{-1}\xi_{1,j}\otimes\xi_{1,j}-S^{(k)}(X_{t_{j-1}},\theta^*)]\right\|+o(1)\\ =&O(1)\end{aligned}$$

by the Burkholder-Davis-Gundy inequality. Therefore we obtained (4.10) and hence (4.9). □

Lemma 4.1.13 and Lemma 4.1.14 imply Condition [A7']. Then we can apply Proposition 2.3.4 and Theorem 2.3.5 to our estimator.

Theorem 4.1.15. Suppose that [A5]-[A6], $[C1]_4$ and [C2]-[C3] are satisfied. Then

$$P\Big(\hat{\theta}_n^{(0)}=0\Big)\to 1$$

as $n\to\infty$.

If Condition [A11] holds, we can apply Theorem 2.4.8 to our penalized estimator. Let

$$\Xi_n=\mathrm{diag}[(\xi_n^j)^{1/q}(j\in\mathcal{J}^{(0)})].$$

The penalized QMLE has limit distribution different from the QMLE.

Theorem 4.1.16. Assume Conditions [A2], [A5]-[A6], [A11], $[C1]_4$ and [C2]-[C3]. Then

$$\Xi_n(\hat{\theta}_n^{(0)}-\theta^{*(0)})\xrightarrow{p}0$$

and

$$\sqrt{n}(\hat{\theta}_n^{(1)}-\theta^{*(1)})\xrightarrow{d}(\Gamma^{(11)})^{-1}((\Gamma^{1/2}\zeta)-\boldsymbol{\psi}^{(1)})\quad(\mathcal{F}-stably)$$

as $T\to\infty$.

Moreover,

$$E[f(\hat{u}_n)]\to E[f(0,(\Gamma^{(11)})^{-1}((\Gamma^{1/2}\zeta)-\boldsymbol{\psi}^{(1)}))]$$

as $n\to\infty$ for any continuous function f of at most polynomial growth.

4.2 Simulation of volatility with a jump global filter

In this section, we conduct some numerical simulations of volatility estimation with a jump global filter in Section 4.1. Let d = 10 and

$$X_t^d = \int_0^t \frac{\sin(2d\pi s)}{(1+(X_s^d)^2)} d\tilde{w}_s^d + \widetilde{N}_t^d, \quad t \in [0,T]$$

for $d = 1, \ldots, 10$ where $\tilde{w}^1, \ldots, \tilde{w}^{10}$ are independent standard Wiener processes and $\widetilde{N}^1, \ldots, \widetilde{N}^{10}$ are independent homogeneous Poisson processes with intensity $\lambda = 10$. We assume

$$\sigma(X_t, \theta) = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ \theta_1 f(X_t^1) & 1 & 0 & 0 & 0 \\ \theta_2 f(X_t^2) & \theta_3 f(X_t^3) & 1 & 0 & 0 \\ \theta_4 f(X_t^4) & \theta_5 f(X_t^5) & \theta_6 f(X_t^6) & 1 & 0 \\ \theta_7 f(X_t^7) & \theta_8 f(X_t^8) & \theta_9 f(X_t^9) & \theta_{10} f(X_t^{10}) & 1 \end{pmatrix}$$

where $f : \mathbb{R} \to \mathbb{R}$ is a function defined by $f(x) = 2 + \sin(x)$. Then $\mathsf{p} = 10$, $\mathsf{m} = \mathsf{r} = 5$ and $\kappa = 1$. Let $(N_t^m)_{m=1,\ldots,5}$ be independent homogeneous Poisson processes with intensity $\lambda = 10$ and $(z_i^m)_{i\in\mathbb{N}}$ be an infinite independent identically distributed standard normal random variables for $m = 1, \ldots, 5$. Define J_t by $J_t = (J_t^1, \ldots, J_t^5)'$ where

$$J_t^m = \sum_{i=1}^{N_t^m} z_i^m \qquad (m = 1, \ldots, 5).$$

When $N_t^m = 0$, we define $J_t^m = 0$. Let $Y_0 = b_s = 0$, then the model (4.1) becomes

$$Y_t = \int_0^t \sigma(X_s, \theta) dw_s + J_t, t \in [0,T],$$

where w_s is 5-dimensional standard Wiener process. We assume $\tilde{w}_t^d, w_t, \widetilde{N}_t^d, N_t^m$, z_i^m are independent.

According to Remark 4.1.3, let $q_n = p_n = 1$ and $\mathfrak{S}_{n,j-1} = I_5$. Moreover let $C_* = 10000$, then factors $K_{n,j}$ is defined by

$$K_{n,j} = 1_{\{|\Delta_j Y| < 10000 n^{-\frac{1}{4}}\}}.$$

Let $\delta_1 = 1/4$ and $B = 1$, then $s_n = n - \lfloor n^{1/4} \rfloor$,

$$V_j = |\Delta_j Y^{(k)}|$$

and

$$\mathcal{K}_n = \{j \in \{1, \ldots, n\}; V_j < V_{(n-\lfloor n^{1/4} \rfloor)}\}.$$

In these settings, we consider the quasi-likelihood function

$$\mathbb{H}_n(\theta) = -\frac{1}{2} \sum_{j \in \mathcal{K}_n} \Big\{ h^{-1} S(X_{t_{j-1}}, \theta)^{-1} [(\Delta_j Y)^{\otimes 2}] K_{n,j} \\ + \log \det S^{(k)}(X_{t_{j-1}}, \theta) \Big\}.$$

According to Section 4.1, let $a_n = n^{-1/2} I_{10}$. In these settings, it is easy to verify Conditions [C1]$_4$ and [C2]. We will derive the Condition[C3].

Proposition 4.2.1. Under settings in this section, Condition [C3] holds.

Proof. We denote the Hessian matrix of $\mathbb{Y}(\theta)$ by $H(\mathbb{Y})$. By definition,

$$\det S(X_t, \theta) = \det S(X_t, \theta^*) = 1$$

and

$$\sigma^{-1}(X_t, \theta) = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ -\theta_1 f_(X_t^1) & 1 & 0 & 0 & 0 \\ \sigma_{31}^{-1} & -\theta_3 f(X_t^3) & 1 & 0 & 0 \\ \sigma_{41}^{-1} & \sigma_{42}^{-1} & -\theta_6 f(X_t^6) & 1 & 0 \\ \sigma_{51}^{-1} & \sigma_{52}^{-1} & \sigma_{53}^{-1} & -\theta_{10} f(X_t^{10}) & 1 \end{pmatrix}$$

where

$$\begin{aligned}
\sigma_{31}^{-1} &= \theta_1 \theta_3 f(X_t^1) f(X_t^3) - \theta_2 f(X_t^2), \\
\sigma_{41}^{-1} &= \theta_1 \theta_5 f(X_t^1) f(X_t^5) - \theta_4 f(X_t^4) \\
&\quad - \theta_1 \theta_3 \theta_6 f(X_t^1) f(X_t^3) f(X_t^6) + \theta_2 \theta_6 f(X_t^2) f(X_t^6), \\
\sigma_{42}^{-1} &= \theta_3 \theta_6 f(X_t^3) f(X_t^6) - \theta_5 f(X_t^5), \\
\sigma_{51}^{-1} &= \theta_1 \theta_8 f(X_t^1) f(X_t^8) - \theta_7 f(X_t^7) - \theta_1 \theta_3 \theta_9 f(X_t^1) f(X_t^3) f(X_t^9) \\
&\quad + \theta_2 \theta_9 f(X_t^2) f(X_t^9) - \theta_1 \theta_5 \theta_{10} f(X_t^1) f(X_t^5) f(X_t^{10}) + \theta_4 \theta_{10} f(X_t^4) f(X_t^{10}) \\
&\quad + \theta_1 \theta_3 \theta_6 \theta_{10} f(X_t^1) f(X_t^3) f(X_t^6) f(X_t^{10}) - \theta_2 \theta_6 \theta_{10} f(X_t^2) f(X_t^6) f(X_t^{10}), \\
\sigma_{52}^{-1} &= \theta_3 \theta_9 f(X_t^3) f(X_t^9) - \theta_8 f(X_t^8) \\
&\quad - \theta_3 \theta_6 \theta_{10} f(X_t^3) f(X_t^6) f(X_t^{10}) + \theta_5 \theta_{10} f(X_t^5) f(X_t^{10})
\end{aligned}$$

and

$$\sigma_{53}^{-1} = \theta_6 \theta_{10} f(X_t^6) f(X_t^{10}) - \theta_9 f(X_t^9).$$

Therefore

$$
\begin{aligned}
&\operatorname{Tr}\bigl(S(X_t,\theta)^{-1}S(X_t,\theta^*)\bigr)\\
&= \operatorname{Tr}\bigl((\sigma(X_t,\theta)^{-1}\sigma(X_t,\theta^*))(\sigma(X_t,\theta)^{-1}\sigma(X_t,\theta^*))'\bigr)\\
&= \sum_{i=1}^{5}\sum_{j=1}^{5}\bigl(\sigma(X_t,\theta)^{-1}\sigma(X_t,\theta^*)\bigr)_{ij}^2\\
&= 5 + (\theta_1-\theta_1^*)^2 f(X_t^1)^2 + (\theta_3-\theta_3^*)^2 f(X_t^3)^2 + (\theta_6-\theta_6^*)^2 f(X_t^6)^2\\
&\quad + (\theta_{10}-\theta_{10}^*)^2 f(X_t^{10})^2 + \{\theta_3(\theta_1-\theta_1^*)f(X_t^1)f(X_t^3) - (\theta_2-\theta_2^*)f(X_t^2)\}^2\\
&\quad + \{\theta_6(\theta_3-\theta_3^*)f(X_t^3)f(X_t^6) - (\theta_5-\theta_5^*)f(X_t^5)\}^2\\
&\quad + \{\theta_{10}(\theta_6-\theta_6^*)f(X_t^6)f(X_t^{10}) - (\theta_9-\theta_9^*)f(X_t^9)\}^2\\
&\quad + \bigl[\{\theta_5 f(X_t^5) - \theta_3\theta_6 f(X_t^3)f(X_t^6)\}(\theta_1-\theta_1^*)f(X_t^1)\\
&\qquad + \theta_6 f(X_t^6)(\theta_2-\theta_2^*)f(X_t^2) - (\theta_4-\theta_4^*)f(X_t^4)\bigr]^2\\
&\quad + \bigl[\{\theta_9 f(X_t^9) - \theta_6\theta_{10} f(X_t^6)f(X_t^{10})\}(\theta_3-\theta_3^*)f(X_t^3)\\
&\qquad + \theta_{10} f(X_t^{10})(\theta_5-\theta_5^*)f(X_t^5) - (\theta_8-\theta_8^*)f(X_t^8)\bigr]^2\\
&\quad + \bigl[\{\theta_8 f(X_t^8) - \theta_3\theta_9 f(X_t^3)f(X_t^9) - \theta_5\theta_{10} f(X_t^5)f(X_t^{10})\\
&\qquad + \theta_3\theta_6\theta_{10} f(X_t^3)f(X_t^6)f(X_t^{!0})\}(\theta_1-\theta_1^*)f(X_t^1)\\
&\qquad + \{\theta_9 f(X_t^9) - \theta_6\theta_{10} f(X_t^6)f(X_t^{10})\}(\theta_2-\theta_2^*)f(X_t^2)\\
&\qquad \theta_{10} f(X_t^{10})(\theta_4-\theta_4^*)f(X_t^4) - (\theta_7-\theta_7^*)f(X_t^7)\bigr]^2. \qquad (4.11)
\end{aligned}
$$

Hence

$$
\begin{aligned}
&H(\mathbb{Y})|_{\theta=\theta^*}\\
&= -\frac{1}{T}\int_0^T \operatorname{diag}\Bigl[f(X_t^1)^2[1 + (\theta_3^* f(X_t^3))^2 + (\theta_5^* f(X_t^5) - \theta_3^*\theta_6^* f(X_t^3)f(X_t^6))^2\\
&\qquad + (\theta_8^* f(X_t^8) - \theta_3^*\theta_9^* f(X_t^3)f(X_t^9) - \theta_5^*\theta_{10}^* f(X_t^5)f(X_t^{10})\\
&\qquad\quad + \theta_3^*\theta_6^*\theta_{10}^* f(X_t^3)f(X_t^6)f(X_t^{!0}))^2],\\
&\qquad f(X_t^2)^2[1 + (\theta_6^* f(X_t^6))^2\\
&\qquad\quad + (\theta_9^* f(X_t^9) - \theta_6^*\theta_{10}^* f(X_t^6)f(X_t^{10}))^2],\\
&\qquad f(X_t^3)^2[1 + (\theta_6^* f(X_t^6))^2\\
&\qquad\quad + (\theta_9^* f(X_t^9) - \theta_6^*\theta_{10}^* f(X_t^6)f(X_t^{10}))^2],\\
&\qquad f(X_t^4)^2[1 + (\theta_{10}^* f(X_t^{10}))^2],\\
&\qquad f(X_t^5)^2[1 + (\theta_{10}^* f(X_t^{10}))^2],\\
&\qquad f(X_t^6)^2[1 + (\theta_{10}^* f(X_t^{10}))^2],\\
&\qquad f(X_t^7)^2, f(X_t^8)^2, f(X_t^9)^2, f(X_t^{10})^2\Bigr]dt
\end{aligned}
$$

and

$$\det[H(\mathbb{Y})|_{\theta=\theta^*}] \leq -\frac{1}{T}\int_0^T \prod_{i=1}^{10}\Big(f(X_t^i)^2\Big)dt \leq 1.$$

Therefore, if $|\theta-\theta^*|$ is small, then $-\mathbb{Y}(\theta)/|\theta-\theta^*|^2$ is large. On the other hand, if $|\theta_1-\theta_1^*|$ is large, then the term $(\theta_1-\theta_1^*)^2 f(X_t^1)^2$ in (4.11) is large. In this case, $-\mathbb{Y}(\theta)/|\theta-\theta^*|^2$ is large since Θ is compact. Moreover, if $|\theta_1-\theta_1^*|$ and $|\theta_3-\theta_3^*|$ are small and $|\theta_2-\theta_2^*|$ is large, then the term $\{\theta_3(\theta_1-\theta_1^*)f(X_t^1)f(X_t^3)-(\theta_2-\theta_2^*)f(X_t^2)\}^2$ in (4.11) is large. Similarly, if $|\theta-\theta^*|$ is large, $-\mathbb{Y}(\theta)/|\theta-\theta^*|^2$ is large. Then we can complete the proof. □

Let $\xi_n^j = n^{1/3}$, $q=1/2$ and $p(x)=|\theta|^{1/2}$, then

$$p_n(\theta) = n^{1/3}\sum_{j=1}^{10}|\theta_j|^{1/2}$$

and

$$\mathbb{H}_n^\dagger(\theta) = \mathbb{H}_n(\theta) - n^{1/3}\sum_{j=1}^{10}|\theta_j|^{1/2}.$$

It is easy to verify Conditions [A2]-[A6] and [A11]. Thus we can obtain results in Theorem 4.1.10, 4.1.15 and 4.1.16. We calculate MQLE

$$\hat{\theta}_n^M \in \underset{\theta\in\overline{\Theta}}{\mathrm{argmax}}\mathbb{H}_n(\theta)$$

and penalized MQLE

$$\hat{\theta}_n \in \underset{\theta\in\overline{\Theta}}{\mathrm{argmax}}\mathbb{H}_n^\dagger(\theta)$$

by numerical simulations. Through the simulation, we fix terminal $T=1$. The true value of θ is

$$\theta^* = (0,0,-0.5,1,0,0,0,2,0,-1)'.$$

We iterated 300 times for $n=500,2000$, respectively.

Table 4.1: Simulation results for the volatility regression model.

	True		500	2000
		QMLE	-0.002(0.033)	0.001(0.022)
$\hat{\theta}_1$	0	p-QL	-0.002(0.027)	0.000(0.018)
		prob	0.906	0.943
		QMLE	0.02(0.041)	-0.001(0.023)
$\hat{\theta}_2$	0	p-QL	0.002(0.030)	0.000(0.016)
		prob	0.893	0.950
		QMLE	-0.473(0.063)	-0.493(0.025)
$\hat{\theta}_3$	-0.5	p-QL	-0.468(0.064)	-0.491(0.025)
		prob	0.997	1.000
		QMLE	0.932(0.147)	0.981(0.051)
$\hat{\theta}_4$	1	p-QL	0.931(0.147)	0.981(0.051)
		prob	1.000	1.000
		QMLE	0.009(0.063)	0.004(0.035)
$\hat{\theta}_5$	0	p-QL	0.007(0.059)	0.003(0.032)
		prob	0.803	0.926
		QMLE	-0.008(0.055)	0.002(0.023)
$\hat{\theta}_6$	0	p-QL	-0.007(0.051)	0.001(0.019)
		prob	0.813	0.946
		QMLE	-0.014(0.127)	0.003(0.091)
$\hat{\theta}_7$	0	p-QL	-0.014(0.115)	0.002(0.085)
		prob	0.567	0.797
		QMLE	1.872(0.257)	1.961(0.119)
$\hat{\theta}_8$	2	p-QL	1.869(0.257)	1.959(0.119)
		prob	1.000	1.000
		QMLE	-0.065(0.167)	-0.031(0.095)
$\hat{\theta}_9$	0	p-QL	-0.063(0.161)	-0.030(0.093)
		prob	0.680	0.856
		QMLE	-0.829(0.242)	-0.927(0.1127)
$\hat{\theta}_{10}$	-1	p-QL	-0.826(0.241)	-0.925(0.127)
		prob	1.000	1.000
		Under model	0.363	0.67
Total		Over model	0.997	1
		True model	0.363	0.673

Bibliography

Belomestny, D., Trabs, M. (2018). Low-rank diffusion matrix estimation for high-dimensional time-changed Lévy processes. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 54(3), 1583–1621.

Chan, N.H., Ing, C.K. (2011). Uniform moment bounds of Fisher's information with applications to time series. *The Annals of Statistics*, 39(3), 1526–1550.

Clinet, S., Yoshida, N. (2017). Statistical inference for ergodic point processes and application to limit order book. *Stochastic Processes and their Applications*, 127(6), 1800–1839.

De Gregorio, A., Iacus, S.M. (2012). Adaptive Lasso-type estimation for multivariate diffusion processes. *Econometric Theory*, 28(4), 838–860.

Fan, J., Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American statistical Association*, 96(456), 1348–1360.

Frank, L.E., Friedman, J.H. (1993). A statistical view of some chemometrics regression tools. *Technometrics*, 35(2), 109–135.

Gaïffas, S., Matulewicz, G. (2019). Sparse inference of the drift of a high-dimensional Ornstein-Uhlenbeck process. *Journal of Multivariate Analysis*, 169, 1–20.

Ibragimov, I., Has'minskii, R. (1981). Statistical estimation: asymptotic theory. .

Inatsugu, H., Yoshida, N. (2018). Global jump filters and quasi likelihood analysis for volatility. *arXiv preprint arXiv:180610706*.

Kinoshita, Y., Yoshida, N. (2019). Penalized quasi likelihood estimation for variable selection. *arXiv preprint arXiv:191012871*.

Knight, K., Fu, W. (2000). Asymptotics for Lasso-type estimators. *The Annals of statistics*, 28(5), 1356–1378.

Masuda, H. (2013). Convergence of Gaussian quasi-likelihood random fields for ergodic Lévy driven SDE observed at high frequency. *The Annals of Statistics*, 41(3), 1593–1641.

Masuda, H., Shimizu, Y. (2017). Moment convergence in regularized estimation under multiple and mixed-rates asymptotics. *Mathematical Methods of Statistics*, 26(2), 81–110.

Ogihara, T., Yoshida, N. (2011). Quasi-likelihood analysis for the stochastic differential equation with jumps. *Statistical Inference for Stochastic Processes*, 14(3), 189.

Ogihara, T., Yoshida, N. (2014). Quasi-likelihood analysis for nonsynchronously observed diffusion processes. *Stochastic Processes and their Applications*, 124(9), 2954–3008.

Shimizu, Y. (2017). Moment convergence of regularized least-squares estimator for linear regression model. *Annals of the Institute of Statistical Mathematics*, 69(5), 1141–1154.

Shimizu, Y., Yoshida, N. (2006). Estimation of parameters for diffusion processes with jumps from discrete observations. *Statistical Inference for Stochastic Processes*, 9(3), 227–277.

Simon, N., Friedman, J., Hastie, T., Tibshirani, R. (2013). A Sparse-Group Lasso. *Journal of Computational and Graphical Statistics*, 22(2), 231–245.

Suzuki, T., Yoshida, N. (2018). Penalized least squares approximation methods and their applications to stochastic processes. *arXiv preprint arXiv:181109016*.

Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society Series B (Methodological)*, pp 267–288.

Uchida, M., Yoshida, N. (2013). Quasi likelihood analysis of volatility and nondegeneracy of statistical random field. *Stochastic Processes and their Applications*, 123(7), 2851–2876.

Umezu, Y., Shimizu, Y., Masuda, H., Ninomiya, Y. (2019). Aic for the nonconcave penalized likelihood method. *Annals of the Institute of Statistical Mathematics*, 71(2), 247–274.

Yoshida, N. (2011). Polynomial type large deviation inequalities and quasi-likelihood analysis for stochastic differential equations. *Annals of the Institute of Statistical Mathematics*, 63(3), 431–479.

Yuan, M., Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(1), 49–67.

Zhang, C.H. (2010). Nearly unbiased variable selection under minimax concave penalty. *The Annals of statistics*, 38(2), 894–942.

Zou, H. (2006). The adaptive Lasso and its oracle properties. *Journal of the American statistical association*, 101(476), 1418–1429.