Quasi-Bayesian model
comparison for LAQ models

Shoichi Eguchi
& Hiroki Masuda

MI 2015-7

( Received June 29, 2015 )
QUASI-BAYESIAN MODEL COMPARISON FOR LAQ MODELS

SHOICHI EGUCHI AND HIROKI MASUDA

Abstract. We will prove a general result about the stochastic expansion of the logarithmic marginal quasi-likelihood associated with a class of locally asymptotically quadratic (LAQ) family of statistical experiments. It enables us to make a Bayesian model comparison in a unified manner for a broad range of dependent-data models, thus entailing a far-reaching extension of the classical Schwarz’s paradigm with rigorous theoretical foundation. In particular, the proposed quasi-Bayesian information criterion, termed QBIC, prevails even when the corresponding M-estimator is of multi-scaling type and the asymptotic quasi-information matrix is random, as well as the statistical model is misspecified. We will illustrate the proposed method by diffusion-type models.

1. Introduction

As is well known, two classical principle of model selection are the Kullback-Leibler divergence (KL divergence) principle and the Bayesian principle. Akaike’s principle of model selection is to choose the model minimizing the KL divergence \( I\left(g_n; f_n(\cdot | \hat{\theta}_{MLE}^n)\right) \) of the fitted model \( f_n(\cdot | \hat{\theta}_{MLE}^n) \) from the true model \( g_n(\cdot) \). This principle equivalent to choosing the model that maximizes the expected log-likelihood function with the expectation is taken with respect to the true distribution. Because of the asymptotic theory of maximum likelihood estimation, the expected log-likelihood function can be asymptotically expanded as \( \sum_{i=1}^{n} \log f_n(X_i; \hat{\theta}_{MLE}^n) - p \) for the case of IID observations with correctly specified models, which leads to the seminal Akaike’s information criterion (AIC) by Akaike [1], [2] for quantitative model comparison. The Bayesian principle, where data sequence is regarded as given constants, leads to Bayesian information criterion (BIC) by Schwarz [15] when the model are correctly specified. The same derivation as in [15] is of course possible even in the frequentist setting, by incorporating regularity conditions to be valid with probability one as was done in, among others, Cavanagh and Neath [4] and Lv and Liu [13]; most importantly, the regularity conditions must include the almost-sure convergence of the observed-information matrix with a positive-definite limit. The model selection consistency of BIC has been shown by many authors (e.g., Casella et al. [3]). Other works on the model selection include the risk information criterion (Foster and George [8]), the generalized information criterion (Konishi and Kitagawa [11]), the ‘parametricness’ index (Liu and Yang [14]), and many extensions of AIC and BIC (e.g., Lv and Liu [13] and Chen and Chen [7]). As for information criteria in case of dependent-data models, we just refer to Uchida [17] and Sei and Komaki [16], as well as the references therein.

The objective of this paper is to extend the range of application of Schwarz’s BIC to a large degree in a unified way. The Bayesian principle of model selection in \( \mathcal{M}_1, \ldots, \mathcal{M}_M \) is to choose the model that is most likely in terms of the posterior probability, so that we select the model that maximizes the marginal likelihood function about data. In the case of IID observations with correctly specified regular models, under some regularity conditions and the Taylor expansion, BIC for the \( m \)th model is defined to be

\[
BIC_m = -2\ell_{m,n}(\hat{\theta}_{MLE}^{m,n}) + p_m \log n,
\]

where \( m \in \{1, \ldots, M\} \), and \( \ell_{m,n}, \hat{\theta}_{MLE}^{m,n}, \) and \( p_m \) denotes the log-likelihood function, the MLE, and the dimension of the parameter space of the \( m \)th model, respectively. Unfortunately, a mathematically-rigorous derivation of BIC type statistics is sometimes missing in the literature, especially when underlying model is of dependent observations. In this paper, we will assume that the statistical models are locally asymptotically quadratic (LAQ) with asymptotic Fisher information matrix being possibly random. We will introduce a class of quasi log-likelihoods and propose extension of the classical BIC, called quasi BIC (QBIC), through the stochastic expansion of the marginal quasi-likelihood; here, we use the terminology “quasi” just to mean that the model may be misspecified in the sense that the candidate models may

\[\text{Date: June 29, 2015.}\]

\textbf{Key words and phrases.} Asymptotic Bayesian model comparison, Bayesian information criterion, Model misspecification, Quasi-likelihood, Dependent data.

Partly supported by JSPS KAKENHI Grant Number 26400204 (HM).
not include the true joint distribution of data sequence. The reason why we do not consider an almost-
sure expansion but a stochastic one is that, in dependent-data models, the almost-sure convergence of
the observe information matrix may be somewhat difficult (and possibly unnatural) to derive; see also
Remark 3.2.

The rest of the paper is organized as follows. In Section 2, we describe our model setup and assump-
tions. Section 3 presents the stochastic expansion of the logarithmic marginal quasi-likelihood, which in
turn leads to a unified way to quantitative model comparison in a model-descriptive sense. In Section 4,
we illustrate the proposed model selection method by considering Gaussian quasi-likelihood estimation of
ergodic diffusion process and volatility-parameter estimation for a class of continuous semimartingales,
both consisting of high-frequency data; to the best of our knowledge, this is the first place that mathemat-
ically validates Schwarz-type model comparison for those described by high-frequency sample from
a stochastic process; also, some corresponding simulation results are reported in Section 4.4. Section 5
gives the proofs of our results.

2. Model and Assumptions

We denote by $X_m$ an observed data of size $n$, defined on an underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$.
We write $G_n(dx) = g_n(x) \mu_n(dx)$ for the true distribution $\mathcal{L}(X_n)$, where $\mu_n$ is a $\sigma$-finite measure on the
(Borel) state space of $X_m$; namely, $G_n(dx) = P \circ X_n^{-1}(dx)$. Suppose that we are given $M$ Bayesian-model
candidates $\mathfrak{M}_1, \ldots, \mathfrak{M}_M$ for $\mathcal{L}(X_n)$.

Each $\mathfrak{M}_m$ is described by

$$\{ (p_m, \pi_m(\theta), \mathbb{H}_{m,n}(\theta)) | \theta \in \Theta_m \},$$

where the ingredients are specified as follows.

- The prior discrete probability of relative likeliness $p_m$ of $m$th-model occurrence among the $M$
candidates (we are bringing possible model misspecification into view): $p_m > 0$ with $\sum_{m=1}^{M} p_m = 1$.
- The $m$th parameter space $\Theta_m \subset \mathbb{R}^{p_m}$ is a bounded convex domain. For each $k \in \{1, \ldots, K_m\}$, let
$\Theta_{m,k}$ be a bounded open set in $\mathbb{R}^{p_{m,k}}$, where $p_m = \sum_{k=1}^{K_m} p_{m,k}$ and the number $K_m$ will be defined
later. Parameter $\theta = (\theta_1, \ldots, \theta_{K_m})$ is a blocked vector in $\Theta_{m,1} \times \cdots \times \Theta_{m,K_m} = \Theta_m \subset \mathbb{R}^{p_m}$.
- $\pi_m$ denotes the prior-probability density on $\Theta_m$.
- $\mathbb{H}_{m,n}(\theta) = \mathbb{H}_{m,n}(\theta; X_n)$ is the logarithm of a quasi-likelihood function, some probability density
function with respect to some dominating measure.

Any random mapping $\hat{\theta}_{m,n} = (\hat{\theta}_{m,1,n}, \ldots, \hat{\theta}_{m,K_m,n})$ such that

$$\hat{\theta}_{m,n} \in \arg\max_{\theta \in \Theta_m} \mathbb{H}_{m,n}(\theta)$$

is called the quasi-maximum likelihood estimator (QMLE) associated with $\mathbb{H}_{m,n}$. The optimal value of
$\theta_m$ associated with $\mathbb{H}_{m,n}$ is denoted by $\theta_{m,0} = (\theta_{m,1,0}, \ldots, \theta_{m,K_m,0})$. We write $\theta_{k_m} = (\theta_{k_m,1}, \ldots, \theta_{k_m})$ and $\theta_{m,k_m,0} = (\theta_{m,k_m,1}, \ldots, \theta_{m,k_m})$ in a similar manner; basically, we are thinking of the
cases where the normalized QMLE

$$A_{m,n}(\theta_{m,0})^{-1}(\hat{\theta}_{m,n} - \theta_{m,0})$$

has a non-trivial asymptotic distribution. We allow some different convergence rates of the quasi-
maximum likelihood estimator for each candidate model, so let $K_m$ be the number of different convergence
rates in model $\mathfrak{M}_m$, and let the rate matrix

$$A_{m,n}(\theta_{m,0}) = \text{diag} \left\{ a_{m,1,n}(\theta_{m,1,0}), \ldots, a_{m,1,n}(\theta_{m,1,0}) I_{p_{m,1}}, \ldots, a_{m,K_m,n}(\theta_{m,K_m,0}) I_{p_{m,K_m}} \right\},$$

where $I_m$ denotes the $m$-dimensional identity matrix and $a_{m,k_m,n}(\theta_{m,k_m,0})$ is a deterministic sequence
satisfying $a_{m,k_m,n}(\theta_{m,k_m,0}) \to 0$ as $n \to \infty$.

Note that we are allowing not only dependent data but also possibility of the model misspecification,
so that we may deal with a wide range of quasi-likelihood $\mathbb{H}_n$, even including semiparametric situations.
A typical example is the Gaussian quasi-likelihood, which uses the Gaussian likelihood only by assuming
(conditional) mean and variance structures, and is quite familiar in time-series literature; see Section 4
for related models.

Write $u_m = (u_{m,1}, \ldots, u_{m,K_m}) \in \mathbb{R}^{p_{m,1}} \times \cdots \times \mathbb{R}^{p_{m,K_m}}$. The rescaled quasi-log-likelihood ratio (statis-
tical random field) of the $m$th model $\mathfrak{M}_m$ is defined by

$$Z_{m,n}(u_m) = \exp \left\{ \mathbb{H}_{m,n}(\theta_{m,0}) + A_{m,n}(\theta_{m,0}) u_m \right\}.$$
The \( k_m \)th random field \( Z_{k_m}^{m,n} = \exp \left( \mathbb{H}_{m,n}(\mathbf{\theta}_{k_m-1}, \mathbf{\theta}_{m,k_m}, \mathbf{\bar{\theta}}_{k_m+1}) \right) \) of the \( m \)th model \( \mathcal{M}_m \) is defined by
\[
Z_{k_m}^{m,n}(\mathbf{u}_{m,k_m}, \mathbf{\bar{\theta}}_{k_m-1}, \mathbf{\theta}_{m,k_m}, \mathbf{\bar{\theta}}_{k_m+1}) = \exp \left( \mathbb{H}_{m,n}(\mathbf{\theta}_{k_m-1}, \mathbf{\theta}_{m,k_m}, \mathbf{\bar{\theta}}_{k_m+1}) - \mathbb{H}_{m,n}(\mathbf{\bar{\theta}}_{k_m-1}, \mathbf{\theta}_{m,k_m}, \mathbf{\bar{\theta}}_{k_m+1}) \right).
\]
The random fields \( Z_{k_m}^{m,n} \) is designed to focus on the \( k \)-th graded parameters, when we have more than one rate of convergence, i.e., when \( K_m \geq 2 \). By convention, we neglect symbols with index \( K + 1 \) like \( \mathbf{\bar{\theta}}_{K_m+1} \) and ones with index 0 like \( \mathbf{\bar{\theta}}_0 \).

We are going to look at the asymptotic behavior of the marginal likelihood for each candidate model. For notational brevity, we will omit the model index “\( m \)” from the notation: \( \{(p, \pi(\theta), \mathbb{H}_n(\theta)) \mid \theta \in \Theta \subset \mathbb{R}^p\} \). Let \( K \subset \Theta \). We now introduce some assumptions.

**Assumption 2.1.**
(i) \( \mathbb{H}_n(\theta) \) is of class \( C^3 \).
(ii) \( \Delta_n = O_p(1) \), where \( \Delta_n = A_n(\theta_0)\partial_\theta \mathbb{H}_n(\theta_0) + \partial_\theta \mathbb{H}_n(\theta_0) \) and \( \partial_\theta = \partial/\partial \theta \).
(iii) For some \( \gamma > 0 \), \( \Gamma_n = \Gamma_0 + o_p(1) \), where \( \Gamma_n = \gamma \mathbb{H}_n(\theta_0)\partial_\theta \mathbb{H}_n(\theta_0)A_n(\theta_0) \).
(iv) For any \( i, j, k \in \{1, \ldots, p\} \),
\[
\sup_{\theta \in \Theta} \left| \left( \partial_{i\theta} \partial_\theta \partial_{j\theta} \mathbb{H}_n(\theta) \right) A_{n,ij}(\theta_0)A_{n,kk}(\theta_0) \right| = o_p(1),
\]
where \( \partial^j \) denote the \( j \)th element of \( \theta \), and \( A_{n,ij}(\theta_0) \) the \( (i, j) \)th element of \( A_n(\theta_0) \).

The quadratic form \( \Gamma_0 \) may be random. Volatility parameter estimation of a continuous semimartingale is the case; indeed, random limiting information is common in the so-called non-ergodic statistics; see Section 4.4.2 for a concrete example.

**Assumption 2.2.**
(i) \( \pi(\theta_0) > 0 \), \( \sup_{\theta \in \Theta} \pi(\theta) < \infty \).
(ii) For every \( M > 0 \), \( \sup_{|u| < M} \left| \pi(\theta_0 + A_n(\theta_0)u) - \pi(\theta_0) \right| \to 0 \) as \( n \to \infty \).

Write \( \mathbb{U}_n(\theta_0) = \{u \in \mathbb{R}^p; \theta_0 + A_n(\theta_0)u \in \Theta\} \).

**Assumption 2.3.**
For any \( \epsilon > 0 \) there exist \( M > 0 \) and \( N \) such that
\[
\sup_{n \geq N} P \left[ \int_{\mathbb{U}_n(\theta_0) \cap \{|u| \geq M\}} Z_n(u)du > \epsilon \right] < \epsilon.
\]

The last condition is a technical one for handling the integral-tail-probability estimate. Let us mention some sufficient conditions for Assumption 2.3.

**Theorem 2.4.**
(i) Assume 2.3 holds if there exist constants \( L > 1 \) and \( C_L > 0 \) such that
\[
\sup_{\theta \in \Theta} \sup_{\theta \in \Theta} \mathbb{P}_{\theta} Z_{k_m}^{m,n}(u_k; \mathbf{\theta}_{k_m-1}, \mathbf{\theta}_{k_m}, \mathbf{\bar{\theta}}_{k_m+1}) \geq e^{-r} \leq C_L r^L
\]
for all \( n > 0 \), \( r > 0 \) and \( k = 1, \ldots, K \).
(ii) If Assumption 2.1 is satisfied and there exists \( \delta > 0 \) such that
\[
\limsup_{n \to \infty} P \left[ \inf_{\theta \in \Theta} \lambda_{\min} \left( - A_n(\theta_0)\partial_\theta^2 \mathbb{H}_n(\theta)A_n(\theta_0) \right) < \delta \right] < \infty,
\]
where \( \lambda_{\min}(\cdot) \) is the smallest eigenvalues of a given matrix, then Assumption 2.3 holds.

The proof of Theorem 2.4 can be found in Yoshida [19, Theorem 6].

**Remark 2.5.** Let \( L > 0 \) and assume that there exists a constant \( C_L > 0 \) such that
\[
\sup_{\theta \in \Theta} \mathbb{P}_{\theta} \left[ \sup_{(u_k, \mathbf{\theta}_{k_m+1}) \in \{|u| \geq r\} \times \prod_{j=k+1}^K \Theta_j} Z_{k_m}^{m,n}(u_k; \mathbf{\theta}_{k_m-1}, \mathbf{\bar{\theta}}_{k_m+1}, \mathbf{\bar{\theta}}_{k_m+1}) \geq 1 \right] \leq C_L r^L
\]
for all \( n > 0 \), \( r > 0 \) and \( k = 1, \ldots, K \). Then the normalized M-estimator \( \hat{u}_n = A_n(\theta_0)^{-1}(\hat{\theta}_n - \theta_0) \) satisfies the inequality
\[
\mathbb{P}_{\theta} \left[ |\hat{u}_n| > r \right] \leq \frac{C_L}{r^L}
\]
for all \( n > 0 \) and \( r > 0 \), from which the tightness of \( \hat{u}_n \) immediately follows. The proof is simple (Yoshida [19, Proposition 2]): since

\[
P_{\theta_0} [\hat{u}_n > r] \leq P_{\theta_0} \left[ a_{i,1}^{\Delta} (\theta_1,0)(\hat{\theta}_{1,n} - \theta_1,0) > \frac{r}{K} \right] + \cdots + P_{\theta_0} \left[ a_{M,1}^{\Delta} (\theta_M,0)(\hat{\theta}_{M,n} - \theta_M,0) > \frac{r}{K} \right],
\]

we have, for some \( L > 0 \) and \( C_L > 0 \),

\[
\sup_{\theta_0 \in \mathcal{K}} \left[ a_{k,1}^{\Delta} (\theta_k,0)(\hat{\theta}_{k,n} - \theta_k,0) > \frac{r}{K} \right]
\]

\[
= \sup_{\theta_0 \in \mathcal{K}} \left( \sup_{(u_k, \overline{x}_{k+1}) \in \{ \overline{x} \leq |u_k| \times \prod_{j=k+1}^K \Theta_j \}} \mathbb{H}_n \left( \hat{\theta}_{k-1}, \theta_k,0 + a_{k,1} (\theta_k,0) u_k, \overline{x}_{k+1} \right) > \right.
\]

\[
\sup_{(u_k, \overline{x}_{k+1}) \in \{ \overline{x} > |u_k| \times \prod_{j=k+1}^K \Theta_j \}} \mathbb{H}_n \left( \hat{\theta}_{k-1}, \theta_k,0 + a_{k,1} (\theta_k,0) u_k, \overline{x}_{k+1} \right)
\]

\[
\leq \sup_{\theta_0 \in \mathcal{K}} \left( \sup_{(u_k, \overline{x}_{k+1}) \in \{ \overline{x} \leq |u_k| \times \prod_{j=k+1}^K \Theta_j \}} \{ \mathbb{H}_n \left( \hat{\theta}_{k-1}, \theta_k,0 + a_{k,1} (\theta_k,0) u_k, \overline{x}_{k+1} \right) - \mathbb{H}_n \left( \hat{\theta}_{k-1}, \theta_k,0, \overline{x}_{k+1} \right) \} \geq 0 \right.
\]

\[
= \sup_{\theta_0 \in \mathcal{K}} \left( \sup_{(u_k, \overline{x}_{k+1}) \in \{ \overline{x} \leq |u_k| \times \prod_{j=k+1}^K \Theta_j \}} \mathbb{I}_n (u_k; \hat{\theta}_{k-1}, \theta_k,0, \overline{x}_{k+1}) \geq 1 \right) \leq \frac{C_L}{r^L} K^L
\]

for all \( n > 0 \) and \( r > 0 \), resulting in the desired estimate

\[
\sup_{\theta_0 \in \mathcal{K}} P_{\theta_0} [\hat{u}_n > r] \leq \frac{C_L}{r^L} K^L + \cdots + \frac{C_L}{r^L} K^L \leq \frac{C_L}{r^L} K^L
\]

Inequalities such as (2.1) and (2.3) are called polynomial type large deviation inequality (PLDI). General sufficient conditions for the PLDI were given in Yoshida [19, Theorem 2]. Obviously, the asymptotic mixed normality of \( \hat{u}_n \) follows upon additionally imposing a suitable functional weak convergence of \( Z_n \) combined with a suitable joint weak convergence of the random sequence

\[
(A_n(\theta_0), E_n(\theta_0), -A_n(\theta_0), \phi_n(\theta_0) A_n(\theta_0)).
\]

\[\square\]

3. Quasi-Bayesian Information Criterion

3.1. Some background. The Bayesian principle of model selection in \( \mathcal{M}_1, \ldots, \mathcal{M}_M \) is to choose the model that is most likely in terms of the posterior probability, i.e. to choose model \( \mathcal{M}_{m_0} \) such that

\[
m_0 = \arg\max_{m \in \{1, \ldots, M\}} P(\mathcal{M}_m | x_n),
\]

where

\[
P(\mathcal{M}_m | x_n) = \frac{p_m(x_n) p_m}{\sum_{i=1}^M p_i(x_n) p_i},
\]

where \( p_i(x_n) \) denotes the \( i \)-th model marginal quasi-likelihood function about data \( x_n \) (a sample from \( \mathcal{L}(X_n) \)):

\[
p_i(x_n) = \int_{\Theta_i} \exp[H_{i,n}(\theta)] \pi_i(\theta) d\theta.
\]

When the prior plausibilities on the \( M \) competing models must be equal, we select the model that maximizes \( p_i(x_n) \); even if the prior plausibilities are not equal, we can trivially correct the selection manner by the factors \( p_n \). Hence we use it as the principle of model selection by expressing the log marginal quasi-likelihood function which is a natural logarithm of the marginal quasi-likelihood function \( p_i(x_n) \).

As was explained in [13], yet another interpretation is possible through the KL divergence. The KL divergence between the true model \( g_n \) and a candidate model \( f_n \) is given by

\[
I(g_n; f_n) = E \left[ \log \frac{g_n(X_n)}{f_n(X_n)} \right] = E[\log g_n(X_n)] - E[\log f_n(X_n)],
\]

where the expectation is taken with respect to the true distribution \( G_n \). Then we see that

\[
- \int_{\Theta} \exp[\mathbb{L}_n(\theta)] \pi(\theta) d\theta
\]

gives, up to a common additive constant, an unbiased estimator of \( I(g_n; f_n \exp[\mathbb{L}_n(\theta)] \pi(\theta) d\theta) \) because

\[
E \left[ \log \left( \int_{\Theta} \exp[\mathbb{L}_n(\theta)] \pi(\theta) d\theta \right) \right] = E[\log g_n(X_n)] - I \left( g_n; \int_{\Theta} \exp[\mathbb{L}_n(\theta)] \pi(\theta) d\theta \right).
\]
Note that (3.1) holds true regardless of whether or not the true model is in the set of candidate models, implying that Bayesian principle of model selection can be restated as choosing the model that minimizes the KL divergence of the marginal quasi-likelihood function from the true distribution.

In particular, considering the case of IID observations with correctly specified regular models, say \( \mathbb{H}_n(\theta) = \sum_{i=1}^{n} \log p(x_i|\theta) \), Schwarz [15] showed that the logarithmic marginal quasi-likelihood

\[
\log \left( \int_{\Theta} \exp\{\mathbb{H}_n(\theta)\} \pi(\theta) d\theta \right)
\]

admits the stochastic expansion

\[
\log \left( \int_{\Theta} \exp\{\mathbb{H}_n(\theta)\} \pi(\theta) d\theta \right) = \sum_{i=1}^{n} \log p(x_i|\hat{\theta}_n^{MLE}) - \frac{p}{2} \log n + O_p(1), \tag{3.2}
\]

with \( \hat{\theta}_n^{MLE} \) denoting the maximum likelihood estimator of \( \theta \), under some regularity conditions and the Taylor expansion. Due to (3.2), we obtain the seminal classical Bayesian information criterion for model selection:

\[
BIC = -2 \sum_{i=1}^{n} \log p(x_i|\hat{\theta}_n^{MLE}) + p \log n. \tag{3.3}
\]

3.2. Stochastic expansion. We now consider the stochastic expansion of the marginal quasi-likelihood in the Bayesian principle of model selection. We will prove the asymptotic behavior of the log marginal quasi-likelihood function, and then derive an extension of the classical BIC. The next theorem is the main claim of this paper.

**Theorem 3.1.**

(i) Assume that Assumptions 2.1-2.3 are satisfied, then the logarithmic marginal quasi-likelihood function admits the stochastic expansion

\[
\log \left( \int_{\Theta} \exp\{\mathbb{H}_n(\theta)\} \pi(\theta) d\theta \right) = \mathbb{H}_n(\hat{\theta}_n) + \sum_{k=1}^{K} p_k \log a_{k,n}(\theta) - \frac{1}{2} \log |\Gamma_0| + \frac{p}{2} \log 2\pi + \frac{1}{2} ||\Gamma^{-\frac{1}{2}}_0 \Delta_n||^2 + \log \pi(\hat{\theta}_n) + o_p(1).
\]

(ii) Assume Assumptions 2.1-2.3, and that \( \log a_{k,n}(\hat{\theta}_{k,n}) - \log a_{k,n}(\theta_{k,0}) = o_p(1) \) and \( \log \pi(\hat{\theta}_n) - \log \pi(\theta_0) = o_p(1) \) are satisfied. Then the log marginal quasi-likelihood function admits the stochastic expansion:

\[
\log \left( \int_{\Theta} \exp\{\mathbb{H}_n(\theta)\} \pi(\theta) d\theta \right) = \mathbb{H}_n(\hat{\theta}_n) + \sum_{k=1}^{K} p_k \log a_{k,n}(\hat{\theta}_n) + \frac{p}{2} \log 2\pi - \frac{1}{2} \log \left| -A_n(\hat{\theta}_n) \partial^2_{\theta} \mathbb{H}_n(\theta) A_n(\hat{\theta}_n) \right| + \log \pi(\hat{\theta}_n) + o_p(1).
\]

**Remark 3.2.** It follows that the statistics

\[
\hat{F}_n := \mathbb{H}_n(\hat{\theta}_n) + \frac{p}{2} \log 2\pi - \frac{1}{2} \log \left| -\partial^2_{\theta} \mathbb{H}_n(\hat{\theta}_n) \right| + \log \pi(\hat{\theta}_n)
\]

is a consistent estimator of the (random) logarithmic marginal quasi-likelihood \( \log \left( \int_{\Theta} \exp\{\mathbb{H}_n(\theta)\} \pi(\theta) d\theta \right) \). From the frequentist point of view, it is more desirable to have the convergence of expected logarithmic marginal quasi-likelihood, which follows from the asymptotic uniform integrability of the sequence

\[
\left\{ \log \left( \int_{\Theta} \exp\{\mathbb{H}_n(\theta)\} \pi(\theta) d\theta \right) - \hat{F}_n \right\}_n
\]

In order to deduce it, obviously we would need conditions stronger than those in Theorem 3.1. For example, the stronger version of (2.2):

\[
\sup_n E \left[ \inf_{\theta \in \Theta} \left\{ \lambda_{\min} \left( -A_n(\theta_0) \partial^2_{\theta} \mathbb{H}_n(\theta) A_n(\theta_0) \right) \right\}^{-q} \right] < \infty
\]

for \( q > 0 \) large enough, would ensure the convergence of the expected logarithmic marginal quasi-likelihood; previously, [5] and [6] studied this kind of moment bounds for some time series models. The details of this topic will be reported elsewhere. \( \square \)
Building on Theorem 3.1 (ii), we define the quasi-Bayesian information criterion (QBIC) by
\[
\text{QBIC} := -2H_n(\hat{\theta}_n) + \log \left| -\partial^2_{\theta} H_n(\hat{\theta}_n) \right| \quad (3.4)
\]
\[
= -2H_n(\hat{\theta}_n) + 2 \sum_{k=1}^{K} p_k \log a_{kn}(\hat{\theta}_n)^{-1} + O_p(1). \quad (3.5)
\]

We compute QBIC for each candidate model to get, say QBIC\((1)\), \ldots, QBIC\((M)\), and then select the model minimizing it as the best model \(\mathfrak{M}_{m_0}\) in the sense of approximate Bayesian model description:
\[
m_0 = \arg\min_{m \in \{1, \ldots, M\}} \text{QBIC}(m).
\]
In the case of regular maximum likelihood estimation of IID-data model \(p(x|\theta)\) corresponds to
\[
H_n(\theta) = \sum_{i=1}^{n} \log p(X_i|\theta), \quad A_n(\hat{\theta}_n) = \frac{1}{\sqrt{n}} I_p,
\]
so that (3.5) becomes
\[
\text{QBIC} = -2 \sum_{i=1}^{n} \log p(X_i|\hat{\theta}_n) + p \log n + O_p(1),
\]
from which the BIC (3.3) is recovered by ignoring the \(O_p(1)\) term. Thus we have seen that QBIC is a far-reaching extension of the classical BIC.

Although the original definition (3.4) of QBIC has seemingly higher computational load than (3.5), it makes the correction taking the volume of observed information into account. In particular, when we consider a dependent data model, (3.4) would be more suitable than (3.5) whose bias correction is only based on the rate of convergence.

4. Example: Gaussian quasi-likelihood

4.1. General framework. A general setting for the Gaussian quasi-likelihood estimation is described as follows. Let \(X_n = (X_{n,j})_{j=0}^{n} = (X_{n,0}, \ldots, X_{n,n})\) be an array of random variables, where \(X_{n,j} \in \mathbb{R}\) for brevity. Let \(F_{n,j} := \sigma(X_{n,j}; j \leq n)\) denote the \(\sigma\)-field representing the data information at stage \(j\) when the number of data is \(n\). The Gaussian quasi-likelihood is constructed as if conditional distribution of \(X_{n,j}\) given past information \(F_{n,j-1}\) is Gaussian, i.e.
\[
\mathcal{L}(X_{n,j}|X_{n,0}, \ldots, X_{n,j-1}) \approx N(\mu_{n,j-1}(\theta), \sigma_{n,j-1}(\theta)),
\]
where \(\mu_{n,j-1}\) and \(\sigma_{n,j-1}\) are \(F_{n,j-1}\)-measurable random function on \(\Theta\); most often,
\[
\mu_{n,j-1}(\theta) = \mathbb{E}_{\theta}[X_{n,j}|F_{n,j-1}], \quad \sigma_{n,j-1}(\theta) = \text{var}_{\theta}[X_{n,j}|F_{n,j-1}],
\]
where the conditional expectation and variance are taken under the image measure of \(X_n\) associated with the parameter value \(\theta\). For simplicity, we will largely suppress the subscript “\(n\)”.

Because the quasi-likelihood is given by
\[
\theta \mapsto \sum_{j=1}^{n} \log \frac{1}{\sqrt{2\pi\sigma^2_{j-1}(\theta)}} \exp \left\{ -\frac{1}{2\sigma^2_{j-1}(\theta)} (X_j - \mu_{j-1}(\theta))^2 \right\},
\]
\[
= (\text{const.}) + \left[ -\frac{1}{2} \sum_{j=1}^{n} \left\{ \log \sigma^2_{j-1}(\theta) + \frac{(X_j - \mu_{j-1}(\theta))^2}{\sigma^2_{j-1}(\theta)} \right\} \right],
\]
we may define the Gaussian quasi-likelihood function by
\[
H_n(\theta) = -\frac{1}{2} \sum_{j=1}^{n} \left\{ \log \sigma^2_{j-1}(\theta) + \frac{(X_j - \mu_{j-1}(\theta))^2}{\sigma^2_{j-1}(\theta)} \right\}.
\]
Then, supposing that \(H_n\) and its partial derivatives can be continuously extended to the boundary \(\partial \Theta\), we define the Gaussian QMLE (GQMLE) by any maximizer of \(H_n\) over \(\Theta\).

It is well known that the GQME is not (often far from being) asymptotically efficient when the model is misspecified. However, it has the merit of robustness against the model misspecification: the GQMLE can often exhibit asymptotic (mixed-)normality under appropriate conditions even if the conditional distribution is deviating from being normal.
4.2. Ergodic diffusion process. Let \( X_n = (X_t)_{t=0}^n \) with \( t_j = jh_n \), where \( h_n \) is the discretization step and \( nh_n = T_n \) and \( X_t \) is a solution to the \( d \)-dimensional diffusion process defined by the stochastic differential equation

\[
dX_t = a(X_t, \theta_t) dt + b(X_t, \theta_t) dw_t, \quad t \in [0,T_n], \quad X_0 = x_0.
\]

Here \( a \) is an \( \mathbb{R}^d \)-valued function defined on \( \mathbb{R}^d \times \Theta_2 \), \( b \) is an \( \mathbb{R}^d \otimes \mathbb{R}^d \)-valued function defined on \( \mathbb{R}^d \times \Theta_1 \), \( w_t \) is a \( d \)-dimensional standard Wiener process and \( x_0 \) is a deterministic initial value. Parameter \( \theta = (\theta_1, \theta_2) \) is unknown with \( (\theta_1, \theta_2) \in \Theta_1 \times \Theta_2 \subset \mathbb{R}^{p_1} \times \mathbb{R}^{p_2} = \mathbb{R}^{p} \). We will assume that \( h_n \to 0, T_n = nh_n \to \infty, nh_n^2 \to 0 \) as \( n \to 0 \) and that for some positive constant \( \epsilon_0, nh_n \geq \epsilon_0 \) for every large \( n \).

Let \( B(x, \theta_1) = b(x, \theta_1)b(x, \theta_1) \) and \( \Delta_j X = X_{t_j} - X_{t_{j-1}} \). Then, up to a constant term, the quasi-likelihood function is given by

\[
\mathbb{H}_n(\theta) = -\frac{1}{2} \sum_{j=1}^n \left\{ \log |B(X_{t_{j-1}}, \theta_1)| + \frac{1}{h_n} (\Delta_j X - h_n a(X_{t_{j-1}}, \theta_2))^{\otimes 2} \right\}. \tag{4.1}
\]

Moreover, let \( A_n(\theta_0) = \text{diag}(\frac{1}{\sqrt{n}} I_{p_1}, \frac{1}{\sqrt{m_n}} I_{p_2}) \) be the rate matrix. We assume following conditions (Yoshida [19, Section 6]):

**Assumption 4.1.**

(i) For some constant \( C \),

\[
sup_{\theta_1, \theta_2} |\partial_{\theta_1}^i a(x, \theta_2)| \leq C(1 + |x|)^C \quad (0 \leq i \leq 4),
\]

\[
sup_{\theta_1, \theta_2} |\partial_{\theta_1}^i \partial_{\theta_1}^j b(x, \theta_1)| \leq C(1 + |x|)^C \quad (0 \leq i \leq 4, 0 \leq j \leq 2).
\]

(ii) \( \inf_{|u| = 1} \inf_{(x, \theta_1)} B(x, \theta_1)[u, u] > 0 \).

(iii) There exists a constant \( C \) such that for every \( x_1, x_2 \in \mathbb{R}^p \),

\[
\sup_{\theta_2 \in \Theta_2} [a(x_1, \theta_2) - a(x_2, \theta_2)] + \sup_{\theta_1 \in \Theta_1} |b(x_1, \theta_1) - b(x_2, \theta_1)| \leq C|x_1 - x_2|
\]

(iv) \( X_0 \in \bigcap_{p>0} L^p(P_{\theta_0}) \).

**Assumption 4.2.** For some constant \( a > 0 \),

\[
\sup_{t \in \mathbb{R}} \sup_{A \in \mathcal{A}[X, r \leq t], B \in \mathcal{B}[X, r \geq t+h]} |P_{\theta_0}[A \cap B] - P_{\theta_0}[A] P_{\theta_0}[B]| \leq a^{-1} e^{-ah} \quad (h > 0).
\]

Under Assumption 4.2 ensures the ergodicity: there exists a unique invariant probability measure \( \nu = \nu_{\theta_0} \) of \( X_t \) such that

\[
\frac{1}{T} \int_0^T g(X_t) dt \to \mathbb{P}_{\theta_0} \int g(x) \nu(dx) \quad (T \to \infty)
\]

for any bounded measurable function \( g \).

**Assumption 4.3.** There exists a positive constant \( \chi > 0 \) such that \( \mathcal{V}_{1,0}(\theta_2) \leq -\chi |\theta_1 - \theta_1|_0^2 \) for all \( \theta_1 \in \Theta_1 \), where

\[
\mathcal{V}_{1,0}(\theta_1) = -\frac{1}{2} \int_{\mathbb{R}^p} \left\{ \text{tr} \left( B(x, \theta_1)^{-1} B(x, \theta_1, 0) - I_p \right) + \log |B(x, \theta_1)| \right\} \nu(dx).
\]

**Assumption 4.4.** There exists a positive constant \( \chi' > 0 \) such that \( \mathcal{V}_{2,0}(\theta_2) \leq -\chi' |\theta_2 - \theta_2|_0^2 \) for all \( \theta_2 \in \Theta_2 \), where

\[
\mathcal{V}_{2,0}(\theta_2) = -\frac{1}{2} \int_{\mathbb{R}^p} B(x, \theta_1, 0)^{-1} \left[ (a(x, \theta_2) - a(x, \theta_2, 0))^{\otimes 2} \right] \nu(dx).
\]

The partial derivatives of \( \mathcal{H}_n \) are given as follows: for \( u_1 \in \mathbb{R}^{m_1} \) and \( u_2 \in \mathbb{R}^{m_2} \),

\[
\partial_{\theta_1}^2 \mathcal{H}_n(\theta_1, \theta_2)[u_1^{\otimes 2}] = \frac{1}{2} \sum_{j=1}^n \partial_{\theta_1}^2 \log |B(X_{t_{j-1}}, \theta_1)|
\]

\[
+ \frac{1}{h_n} \partial_{\theta_1}^2 B(X_{t_{j-1}}, \theta_1)^{-1} u_1^{\otimes 2}, (\Delta_j X - h_n a(X_{t_{j-1}}, \theta_2))^{\otimes 2} \}.
\]
Remark 4.7. Assume that Assumptions 4.1-4.4 are satisfied, then QBIC is given by

\[
\text{QBIC} = \sum_{j=1}^{n} \left\{ \log |B(X_{t_{j-1}}, \hat{\theta}_{1,n})| + \frac{1}{h_n} B(X_{t_{j-1}}, \hat{\theta}_{1,n})^{-1} \left[ (\Delta_{j} X - h_{n} a(X_{t_{j-1}}, \hat{\theta}_{2,n})) \right] \right\}
\]

Then, we obtain the corresponding QBIC as in the following theorem. The proof is given in Section 6.

**Theorem 4.5.** Suppose that Assumptions 4.1-4.4 are satisfied, then QBIC is given by

\[
\text{QBIC} = \sum_{j=1}^{n} \left\{ \log |B(X_{t_{j-1}}, \hat{\theta}_{1,n})| + \frac{1}{h_n} B(X_{t_{j-1}}, \hat{\theta}_{1,n})^{-1} \left[ (\Delta_{j} X - h_{n} a(X_{t_{j-1}}, \hat{\theta}_{2,n})) \right] \right\}
\]

Note that Theorem 4.6 is an example of K = 2.

In the case of ergodic diffusion process, convergence in probability

\[
\frac{1}{\sqrt{n^2 h_n}} \partial_{\theta_1} \partial_{\theta_2} \mathbb{H}_n(\hat{\theta}) \to P \quad (n \to \infty)
\]

is satisfied. Due to (4.2), we have that

\[
\log | - A_n(\hat{\theta}_n) \partial_{\theta_2} \mathbb{H}_n(\hat{\theta}_n) A_n(\hat{\theta}_n) | = \log \left| \begin{array}{cc}
- \frac{1}{n} \partial_{\theta_1}^2 \mathbb{H}_n(\hat{\theta}_n) & - \frac{1}{n} \partial_{\theta_1} \partial_{\theta_2} \mathbb{H}_n(\hat{\theta}_n) \\
- \frac{1}{n} \partial_{\theta_1} \partial_{\theta_2} \mathbb{H}_n(\hat{\theta}_n)' & - \frac{1}{n} \partial_{\theta_2}^2 \mathbb{H}_n(\hat{\theta}_n)
\end{array} \right|
\]

\[
= \log \left| \begin{array}{cc}
- \frac{1}{n} \partial_{\theta_1}^2 \mathbb{H}_n(\hat{\theta}_n) & 0 \\
0 & - \frac{1}{n} \partial_{\theta_2}^2 \mathbb{H}_n(\hat{\theta}_n)
\end{array} \right| + o_p(1)
\]

A statistics \( \hat{S}_n \) such that \( \hat{S}_n \) is easier to compute and that \( \hat{S}_n = \text{QBIC} + o_p(1) \) would be conveniently used as a variant of QBIC; recall (3.4) and (3.5).

**Theorem 4.6.** Assume that Assumptions 4.1-4.4 hold, then the statistics

\[
\hat{S}_n = \sum_{j=1}^{n} \left\{ \log |B(X_{t_{j-1}}, \hat{\theta}_{1,n})| + \frac{1}{h_n} B(X_{t_{j-1}}, \hat{\theta}_{1,n})^{-1} \left[ (\Delta_{j} X - h_{n} a(X_{t_{j-1}}, \hat{\theta}_{2,n})) \right] \right\}
\]

satisfies that \( \hat{S}_n = \text{QBIC} + o_p(1) \), where QBIC is that given in Theorem 4.5.

**Remark 4.7.**

1. Under similar conditions, we can generally express a second term of QBIC (3.4) with

\[
\sum_{k=1}^{K} \log | - \partial_{\theta_k}^2 \mathbb{H}_n(\hat{\theta}_n) |.
\]

2. It follows from Kessler [10] that we have

\[
(\sqrt{n}(\hat{\theta}_{1,n} - \theta_{1,0}), \sqrt{n}(\hat{\theta}_{2,n} - \theta_{2,0})) \to d\left( \mathcal{N}_{0}(0, \text{diag}(\Gamma_{1,0}(\theta_{1,0})^{-1}, \Gamma_{2,0}(\theta_{1,0}, \theta_{2,0})^{-1})) \right),
\]

where

\[
\Gamma_{1,0}(\theta_{1,0})[u_1^{\otimes 2}] = \frac{1}{2} \int \text{tr} \left\{ B(x, \theta_{1,0})^{-1} \left( \partial_{\theta_1} B(x, \theta_{1,0}) \right) B(x, \theta_{1,0})^{-1} \left( \partial_{\theta_1} B(x, \theta_{1,0}) \right)[u_1^{\otimes 2}] \nu(dx),
\]

\[
\Gamma_{2,0}(\theta_{1,0}, \theta_{2,0})[u_2^{\otimes 2}] = \int B(x, \theta_{1,0})^{-1} \left( \partial_{\theta_2} a(x, \theta_{2,0})[u_2], \partial_{\theta_2} a(x, \theta_{2,0})[u_2] \right) \nu(dx)
\]

for \( u_1 \in \mathbb{R}^{m_1} \), \( u_2 \in \mathbb{R}^{m_2} \).
4.3. Volatility-parameter estimation for continuous semimartingale. In this section, we deal with the stochastic integral equation

\[ dY_t = b_t dt + \sigma(X_t, \theta) dw_t, \quad t \in [0, T], \]

where \( w \) is an \( r \)-dimensional standard Wiener process, \( b \) and \( X \) are progressively measurable processes with values in \( \mathbb{R}^m \) and \( \mathbb{R}^d \), respectively, \( \sigma \) is an \( \mathbb{R}^m \otimes \mathbb{R}^d \)-valued function defined on \( \mathbb{R}^d \times \Theta \) and \( \Theta \in \mathbb{R}^p \).

The data set consists of discrete observations \( \mathbf{X}_n = (X_{t_j}, Y_{t_j})_{j=0}^{n} \) with \( t_j = jh_n \), where \( h_n = T/n \) and \( T \) is fixed. The process \( b \) is completely unobservable and unknown. All processes are defined on a filtered probability space \( B := (\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \leq T}, \mathbb{P}) \).

Let \( S(x, \theta) = \sigma(x, \theta)\sigma(x, \theta)' \) and \( \Delta_j Y = Y_j - Y_{j-1} \). Then the quasi-likelihood function becomes

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

We have that the asymptotic distribution of \( A_n(\theta_0)^{-1}(\hat{\theta}_n - \theta_0) = \sqrt{n}(\hat{\theta}_n - \theta_0) \) is mixed normal, i.e.

\[
\sqrt{n}(\hat{\theta}_n - \theta_0) \rightarrow^d \mathcal{N}(0, \Sigma_{\theta}^{-1/2} Z),
\]

where \( \Sigma_{\theta} \) is a symmetric \( p \times p \)-matrix which is a.s. positive-definite, and \( Z \) is a \( p \)-variate standard-normal random variable which is defined on an extension of \( B \) and is independent of \( B \); cf. Genon-Catalot and Jacod [9].

**Theorem 4.8.** Under suitable conditions, QBIC is given by

\[
\text{QBIC} = \sum_{j=1}^{n} \left\{ \log |S(X_{t_{j-1}}, \hat{\theta}_n)| + \frac{1}{h_n} S(X_{t_{j-1}}, \hat{\theta}_n)^{-1}[(\Delta_j Y)^\otimes 2] \right\} + \log \left| \frac{\partial S(X_{t_{j-1}}, \hat{\theta}_n)}{\partial \theta} \right|^2
\]

\[
- \left[ \frac{\partial S(X_{t_{j-1}}, \hat{\theta}_n)}{\partial \theta} \right] \left( \frac{\partial S(X_{t_{j-1}}, \hat{\theta}_n)}{\partial \theta} \right)' \left( \frac{\partial S(X_{t_{j-1}}, \hat{\theta}_n)}{\partial \theta} \right)^2 \right|,
\]

where \( S(X_{t_{j-1}}, \theta)^{-1} = (S_{jk}(X_{t_{j-1}}, \theta)^{-1})_{k,\ell=1}^{d} \).

In what follows, we consider the conditions for Theorem 4.8 when \( \sigma(x, \theta) = \exp(x'\theta/2) \), parameter space \( \Theta \) is compact and \( K = 1 \). The quasi-likelihood function becomes

\[
\mathbb{H}_n(\theta) = -\frac{1}{2} \sum_{j=1}^{n} \left\{ X_{t_{j-1}}' \theta + \frac{1}{h_n} (\Delta_j Y)^2 \exp(-X_{t_{j-1}}' \theta) \right\}.
\]

We assume following conditions:

**Assumption 4.9.**

(i) \( \forall q > 0, E[|X_0|^{q}] < \infty \).

(ii) \( \forall q > 0, \exists C > 0, \forall s, t \in [0, T], E[|X_t - X_s|^{q}] < C|t - s|^{q/2} \).

(iii) \( \forall q > 0, \sup_{\omega \in \Omega, t \leq T} E[|\theta_t|^{q}] < \infty \).

**Assumption 4.10.**

(i) \( \sup_{\omega \in \Omega, t \leq T} |X_t| < \infty \).

(ii) \( \forall L > 0, \exists C_L > 0, \forall r > 0, P \left[ \lambda_{\min} \left( \int_0^T X_t X_t' dt \right) \leq \frac{1}{r} \right] \leq C_L r^{-L} \).

It will be seen that Assumptions 4.9 and 4.10 ensure Assumption 2.1 and inequality (2.3), so that we can obtain the following.

**Corollary 4.11.** Assume Assumptions 4.9 and 4.10. Then QBIC is given by

\[
\text{QBIC} = \sum_{j=1}^{n} \left\{ X_{t_{j-1}}' \hat{\theta}_n + \frac{1}{h_n} (\Delta_j Y)^2 \exp(-X_{t_{j-1}}' \hat{\theta}_n) \right\} + \log \left| \frac{1}{2} \sum_{j=1}^{n} \frac{1}{h_n} \exp(-X_{t_{j-1}}' \hat{\theta}_n) X_{t_{j-1}} X_{t_{j-1}}' \right|.
\]

4.4. Simulation results. In this section, we estimate the parameters associated with the log-quasi-likelihood function and choose the model by using QBIC, BIC and AIC or the contrast-based information criterion (CIC). CIC is AIC-type criterion and defined in Uchida [17]. Moreover, we use **optim** at software R to estimate the parameters, hence it is required to set the initial values for numerical optimization. We here set the initial value as the value generated from uniform distribution \( U(\theta_0 - 0.5, \theta_0 + 0.5) \).
TABLE 1. The number of models selected by QBIC, BIC and CIC in Section 4.4.1 over 1000 simulations for various n (1-3 express the models, and the true model is Model 2)

<table>
<thead>
<tr>
<th>Criterion</th>
<th>n = 1000</th>
<th>n = 3000</th>
<th>n = 5000</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1  2*  3</td>
<td>1  2*  3</td>
<td>1  2*  3</td>
</tr>
<tr>
<td>QBIC</td>
<td>269 686 45</td>
<td>11 962 27</td>
<td>0 983 17</td>
</tr>
<tr>
<td>BIC</td>
<td>450 544 6</td>
<td>31 965 4</td>
<td>0 998 2</td>
</tr>
<tr>
<td>CIC</td>
<td>99  746 155</td>
<td>0 843 157</td>
<td>0 853 147</td>
</tr>
</tbody>
</table>

TABLE 2. The mean and the standard deviation (s.d.) of estimator $\hat{\theta}_{11}$, $\hat{\theta}_{12}$, $\hat{\theta}_{13}$, $\hat{\theta}_{21}$ and $\hat{\theta}_{22}$ in each model for various n (1-3 express the models, and the true parameter $(\theta_{110}, \theta_{120}, \theta_{130}, \theta_{210}, \theta_{220}) = (1, 0.7, 0, 2, 1)$)


4.4.1. Ergodic diffusion process. The data $X_n = (X_{t_j})_{j=0}^{n}$ with $t_j = jn^{-2/3}$ and the number of data $n$ are obtained from the true model defined by

$$dX_t = -(2X_t - 1)dt + \frac{1 + 0.7X_t^2}{1 + X_t^2}dw_t, \quad t \in [0, T_n], \quad X_0 = 1,$$

where $w$ is an 1-dimensional standard Winer process and $T_n = n^{1/3}$. We consider the following models:

**Model 1**: $dX_t = -(\theta_{21}X_t - \theta_{22})dt + \theta_{11}dw_t$;

**Model 2**: $dX_t = -(\theta_{21}X_t - \theta_{22})dt + \theta_{11} + \theta_{12}X_t^2dt + \theta_{13}dw_t$;

**Model 3**: $dX_t = -(\theta_{21}X_t - \theta_{22})dt + \theta_{11} + \theta_{13}X_t^2 + \theta_{12}X_t^2dt + \theta_{13}dw_t$.

That is, the true model is Model 2.

It follows from (4.1) that for the Models 1-3, the quasi-likelihood function is given for each model. We simulated the number of the model selected by using QBIC, BIC and CIC among the candidate Models 1-3 based on 1000 sample paths. For example, in the case of Model 1, QBIC, BIC and CIC are given by

$$QBIC = \sum_{j=1}^{n} \left\{ \log(\hat{\theta}_{11}^2) + 3n^{-2}\hat{\theta}_{11}^4\left[(\Delta_jX + n^{-2}(\hat{\theta}_{21}X_{t_j} - \hat{\theta}_{22}))^2\right] \right\} + \log \left\{ \sum_{j=1}^{n} -\hat{\theta}_{11}^2 + 3n^{-2}\hat{\theta}_{11}^4\left[(\Delta_jX + n^{-2}(\hat{\theta}_{21}X_{t_j} - \hat{\theta}_{22}))^2\right] \right\} + \log \left\{ n^{-2}\hat{\theta}_{11}^{-2}\sum_{j=1}^{n} X_{t_j}^2 - \left(n^{-2}\hat{\theta}_{11}^{-2}\sum_{j=1}^{n} X_{t_j}^{-2}\right)^2 \right\},$$

$$BIC = \sum_{j=1}^{n} \left\{ \log(\hat{\theta}_{11}) + n^{-2}\hat{\theta}_{11}^2\left[(\Delta_jX + n^{-2}(\hat{\theta}_{21}X_{t_j} - \hat{\theta}_{22}))^2\right] \right\} + \log n - 2\log n^4,$$

$$CIC = \sum_{j=1}^{n} \left\{ \log(\hat{\theta}_{11}) + n^{-2}\hat{\theta}_{11}^2\left[(\Delta_jX + n^{-2}(\hat{\theta}_{21}X_{t_j} - \hat{\theta}_{22}))^2\right] \right\} + 2 \times 3,$$

respectively, where $\hat{\theta}_{11}$, $\hat{\theta}_{21}$ and $\hat{\theta}_{22}$ are the GQMLE. The simulations are done for each $n = 1000, 3000, 5000$. Note that Model 1 is a misspecified model.

Table 1 summarizes the comparison results of the frequency of the model selection. The larger $n$ becomes, the higher the probability that Model 2 is selected by QBIC and BIC becomes. However the probability that Model 1 is selected by QBIC, BIC and CIC is very low when $n$ is large. This result implies that the probability that the misspecified model is selected by QBIC tends to zero as $n \to \infty$.

Table 2 summarizes the mean and the standard deviation of estimators in each model.
TABLE 3. The number of models selected by QBIC, BIC and AIC in Section 4.4.2 over 1000 simulations for various $n$ (1-7 express the models, and the true model is model 4)

<table>
<thead>
<tr>
<th>Criterion</th>
<th>$n = 50$</th>
<th>$n = 100$</th>
<th>$n = 200$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 2 3 4* 5 6 7</td>
<td>1 2 3 4* 5 6 7</td>
<td>1 2 3 4* 5 6 7</td>
</tr>
<tr>
<td>QBIC</td>
<td>74 0 0 925 0 0 0</td>
<td>57 0 0 943 0 0 0</td>
<td>37 0 0 963 0 0 0</td>
</tr>
<tr>
<td>BIC</td>
<td>67 0 0 933 0 0 0</td>
<td>39 0 0 961 0 0 0</td>
<td>25 0 0 975 0 0 0</td>
</tr>
<tr>
<td>AIC</td>
<td>183 0 0 817 0 0 0</td>
<td>178 0 0 822 0 0 0</td>
<td>179 0 0 821 0 0 0</td>
</tr>
</tbody>
</table>

TABLE 4. The mean and the standard deviation (s.d.) of estimator $\hat{\theta}_1$, $\hat{\theta}_2$ and $\hat{\theta}_3$ in each model for various $n$ (1-7 express the models, and the true parameter $\theta_0 = (0, -2, 3)$)

<table>
<thead>
<tr>
<th>n</th>
<th>$\hat{\theta}_1$ mean (s.d.)</th>
<th>$\hat{\theta}_2$ mean (s.d.)</th>
<th>$\hat{\theta}_3$ mean (s.d.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>-0.0564 (-0.2086)</td>
<td>-1.8312 (-0.2879)</td>
<td>3.1129 (-0.2978)</td>
</tr>
<tr>
<td>100</td>
<td>-0.0218 (0.1509)</td>
<td>-1.8972 (0.2006)</td>
<td>3.0748 (0.2052)</td>
</tr>
<tr>
<td>200</td>
<td>-0.0183 (0.1026)</td>
<td>-1.9586 (0.1461)</td>
<td>3.02986 (0.1442)</td>
</tr>
</tbody>
</table>

4.4.2. Volatility-parameter estimation for continuous semimartingale. Let $(X_t, Y_t, j/n)$ be a data set with $t_j = j/n$ and the number of data $n$. We consider $n = 50, 100, 200$ and simulate 1000 data sets from the stochastic integral equation

$$dY_t = \exp \left( \frac{1}{2} X_t^{'} \theta \right) dw_t = \exp \left( \frac{1}{2} (\theta_1 X_{1,t} + \theta_2 X_{2,t} + \theta_3 X_{3,t}) \right) dw_t, \quad t \in [0, 1],$$

where $X_t = (X_{1,t}, X_{2,t}, X_{3,t})'$, $\theta = (\theta_1, \theta_2, \theta_3)'$, and $w$ is an 1-dimensional standard Wiener process. We set $\theta_0 = (0, -2, 3)$ and

$$X_{t,j} = \left( 1, \cos \left( \frac{2j\pi}{n} \right), \sin \left( \frac{2j\pi}{n} \right) \right)', \quad j = 0, 1, \ldots, n.$$

We consider the following models:

**Model 1:** $dY_t = \exp \left( \frac{1}{2} (\theta_1 X_{1,t} + \theta_2 X_{2,t} + \theta_3 X_{3,t}) \right) dw_t$;

**Model 2:** $dY_t = \exp \left( \frac{1}{2} (\theta_1 X_{1,t} + \theta_2 X_{2,t}) \right) dw_t$;

**Model 3:** $dY_t = \exp \left( \frac{1}{2} (\theta_1 X_{1,t} + \theta_3 X_{3,t}) \right) dw_t$;

**Model 4:** $dY_t = \exp \left( \frac{1}{2} (\theta_2 X_{2,t} + \theta_3 X_{3,t}) \right) dw_t$;

**Model 5:** $dY_t = \exp \left( \frac{1}{2} (\theta_1 X_{1,t}) \right) dw_t$;

**Model 6:** $dY_t = \exp \left( \frac{1}{2} (\theta_2 X_{2,t}) \right) dw_t$;

**Model 7:** $dY_t = \exp \left( \frac{1}{2} (\theta_3 X_{3,t}) \right) dw_t$.

Then the true model is Model 4.

It follows from (4.3) that for the Models 1-7, we have the quasi-likelihood functions, respectively. Thus, for each model, the maximum likelihood estimator is obtained from the quasi-likelihood, and QBIC, BIC
and AIC of the models are computable. For example, in the case of Model 1, QBIC, BIC and AIC are given by

\[
\text{QBIC} = \sum_{j=1}^{n} \left\{ (\hat{\theta}_1 X_{1,j-1} + \hat{\theta}_2 X_{2,j-1} + \hat{\theta}_3 X_{3,j-1}) + n(\Delta_j)^2 \exp \left( -\hat{\theta}_1 X_{1,j-1} - \hat{\theta}_2 X_{2,j-1} - \hat{\theta}_3 X_{3,j-1} \right) \right\} \\
+ \log \frac{n}{2} \sum_{j=1}^{n} \exp \left( -\hat{\theta}_1 X_{1,j-1} - \hat{\theta}_2 X_{2,j-1} - \hat{\theta}_3 X_{3,j-1} \right) X_{j-1}',
\]

\[
\text{BIC} = \sum_{j=1}^{n} \left\{ (\hat{\theta}_1 X_{1,j-1} + \hat{\theta}_2 X_{2,j-1} + \hat{\theta}_3 X_{3,j-1}) + n(\Delta_j)^2 \exp \left( -\hat{\theta}_1 X_{1,j-1} - \hat{\theta}_2 X_{2,j-1} - \hat{\theta}_3 X_{3,j-1} \right) \right\} \\
+ 3 \log n,
\]

\[
\text{AIC} = \sum_{j=1}^{n} \left\{ (\hat{\theta}_1 X_{1,j-1} + \hat{\theta}_2 X_{2,j-1} + \hat{\theta}_3 X_{3,j-1}) + n(\Delta_j)^2 \exp \left( -\hat{\theta}_1 X_{1,j-1} - \hat{\theta}_2 X_{2,j-1} - \hat{\theta}_3 X_{3,j-1} \right) \right\} \\
+ 3 \times 2,
\]

where \(\hat{\theta}_1, \hat{\theta}_2\) and \(\hat{\theta}_3\) are estimators.

Table 3 summarizes the comparison results of the frequency of the model selection. Model 4 is selected with high frequency as the optimal model for all cases. AIC tends to choose a model larger than the true one even for large sample size. QBIC and BIC tend to perform better when the sample size become larger. This result indicates that QBIC may have consistency for the model selection among the competing models including the true model. These phenomena are common in model selection.

Table 4 summarizes the mean and the standard deviation of estimators in each model. In the case of Model 4, the estimators get closer to the true value and the standard deviation become smaller when the sample size become larger.

5. Proofs

5.1. Proof of Theorem 2.4 (ii).

Recall that \(U_n(\theta_0) = \{ u \in \mathbb{R}^p : \theta_0 + A_n(\theta_0)u \in \Theta \}\). In what follows, we deal with the zero-extended version of \(Z_n\) and use the same notation: \(Z_n\) vanishes outside \(U_n(\theta_0)\), so that

\[
\int_{\mathbb{R}^p \setminus U_n(\theta_0)} Z_n(u) du = 0.
\]

By using the Taylor expansion, we obtain that

\[
Z_n(u) = \exp \left( \Delta_n[u] - \frac{1}{2} \left( -A_n(\theta_0) \partial^2_{\theta} \mathbb{H}_n(\theta) A_n(\theta_0) \right)[u, u] \right),
\]

where \(\hat{\theta}_n = \theta_0 + \xi (\theta_0 + A_n(\theta_0)u - \theta_0) = \theta_0 + \xi A_n(\theta_0)u\) for some \(\xi\) satisfying \(0 < \xi < 1\). Then, for any \(\epsilon > 0\) and \(M > 0\),

\[
P \left[ \int_{U_n(\theta_0) \cap \{|u| \geq M\}} Z_n(u) du > \epsilon \right] \\
\leq P \left[ \int_{|u| \geq M} Z_n(u) du > \epsilon ; \inf_{\theta \in \Theta} \lambda_{\min} \left( -A_n(\theta_0) \partial^2_{\theta} \mathbb{H}_n(\theta) A_n(\theta_0) \right) < \delta \right] \\
+ P \left[ \int_{|u| \geq M} Z_n(u) du > \epsilon ; \inf_{\theta \in \Theta} \lambda_{\min} \left( -A_n(\theta_0) \partial^2_{\theta} \mathbb{H}_n(\theta) A_n(\theta_0) \right) \geq \delta \right] \\
\leq P \left[ \inf_{\theta \in \Theta} \lambda_{\min} \left( -A_n(\theta_0) \partial^2_{\theta} \mathbb{H}_n(\theta) A_n(\theta_0) \right) < \delta \right] + P \left[ \int_{|u| \geq M} \exp \left( \Delta_n[u] - \frac{\delta}{2}[u, u] \right) du > \epsilon \right],
\]

where \(\delta > 0\) is satisfied (2.2). We can take \(M > 0\) large enough, and inequality

\[
P \left[ \int_{|u| \geq M} \exp \left( \Delta_n[u] - \frac{\delta}{2}[u, u] \right) du > \epsilon \right] < \epsilon
\]

is established. Because of this inequality and (2.2), for some \(N\), \(P \left[ \int_{U_n(\theta_0) \cap \{|u| \geq M\}} Z_n(u) du > \epsilon \right] < \epsilon\) for all \(n \geq N\).
5.2. Proof of Theorem 3.1.
(i) By using the change of variable \( \theta = \theta_0 + A_n(\theta_0)u \), the log marginal quasi-likelihood function becomes

\[
\log \left( \int_{\Theta} \exp\{H_n(\theta)\} \pi(\theta) d\theta \right) = H_n(\theta_0) + \sum_{k=1}^{K} p_k \log a_k(\theta_0) + \log \left( \int_{U_n(\theta_0)} Z_n(u) \pi(\theta_0 + A_n(\theta_0)u) du \right).
\]

As a consequence of this equality, we have that

\[
\log \left( \int_{\Theta} \exp\{H_n(\theta)\} \pi(\theta) d\theta \right) - \left( H_n(\theta_0) + \sum_{k=1}^{K} p_k \log a_k(\theta_0) + \log Q_n \right) = \log (Q_n + \varepsilon_n) - \log Q_n,
\]

where

\[
Q_n = \pi(\theta_0) \int_{\mathcal{R}^p} \exp \left( \Delta_n[u] - \frac{1}{2} \Gamma_0[u, u] \right) du,
\]

\[
\varepsilon_n = \int_{U_n(\theta_0)} Z_n(u) \left( \pi(\theta_0 + A_n(\theta_0)u) - \pi(\theta_0) \right) du + \pi(\theta_0) \int_{\mathcal{R}^p} \left\{ Z_n(u) - \exp \left( \Delta_n[u] - \frac{1}{2} \Gamma_0[u, u] \right) \right\} du.
\]

First we prove \( \log \bar{Q}_n = \log \pi(\theta_0) + \frac{1}{2} \|\Gamma_0^{-\frac{1}{2}} \Delta_n\|^2 + \frac{1}{2} \log(2\pi) - \frac{1}{2} \log |\Gamma_0| \).

\[
\bar{Q}_n = \pi(\theta_0) \int_{\mathcal{R}^p} \exp \left( \Delta_n[u] - \frac{1}{2} \Gamma_0[u, u] \right) du
= \pi(\theta_0) \exp \left( \frac{1}{2} \|\Gamma_0^{-\frac{1}{2}} \Delta_n\|^2 \right) \int_{\mathcal{R}^p} \exp \left\{ - \frac{1}{2} \Gamma_0[u - \Gamma_0^{-1} \Delta_n, u - \Gamma_0^{-1} \Delta_n] \right\} du
= \pi(\theta_0) \exp \left( \frac{1}{2} \|\Gamma_0^{-\frac{1}{2}} \Delta_n\|^2 \right)(2\pi)^{\frac{p}{2}} |\Gamma_0|^{-\frac{1}{2}}.
\]

Because of Assumptions 2.1 and 2.2, \( \log \bar{Q}_n \) is given by

\[
\log \bar{Q}_n = \log \pi(\theta_0) + \frac{1}{2} \|\Gamma_0^{-\frac{1}{2}} \Delta_n\|^2 + \frac{p}{2} \log(2\pi) - \frac{1}{2} \log |\Gamma_0|.
\]

Next we consider \( \varepsilon_n \).

\[
|\varepsilon_n| \leq \int_{U_n(\theta_0)} Z_n(u) \left| \pi(\theta_0 + A_n(\theta_0)u) - \pi(\theta_0) \right| du + \pi(\theta_0) \int_{\mathcal{R}^p} \left| Z_n(u) - \exp \left( \Delta_n[u] - \frac{1}{2} \Gamma_0[u, u] \right) \right| du.
\]

For any \( \epsilon > 0 \) and for some \( M \) large enough,

\[
P \left[ \int_{U_n(\theta_0)} Z_n(u) \left| \pi(\theta_0 + A_n(\theta_0)u) - \pi(\theta_0) \right| du > \epsilon \right]
\leq P \left[ \int_{U_n(\theta_0) \cap \{|u| < M\}} Z_n(u) \left| \pi(\theta_0 + A_n(\theta_0)u) - \pi(\theta_0) \right| du > \frac{\epsilon}{2} \right]
+ P \left[ \int_{U_n(\theta_0) \cap \{|u| \geq M\}} Z_n(u) \left| \pi(\theta_0 + A_n(\theta_0)u) - \pi(\theta_0) \right| du > \frac{\epsilon}{2} \right]
\leq P \left[ \int_{|u| < M} Z_n(u) \left| \pi(\theta_0 + A_n(\theta_0)u) - \pi(\theta_0) \right| du > \frac{\epsilon}{2} \right]
+ P \left[ \int_{|u| \geq M} Z_n(u) \left| \pi(\theta_0 + A_n(\theta_0)u) - \pi(\theta_0) \right| du > \frac{\epsilon}{2} \right]
\leq P \left[ \sup_{|u| < M} \left| \pi(\theta_0 + A_n(\theta_0)u) - \pi(\theta_0) \right| \sup_{|u| < M} Z_n(u) > \frac{\epsilon}{2} \right]
+ P \left[ \sup_{|u| \geq M} \left| \pi(\theta_0 + A_n(\theta_0)u) - \pi(\theta_0) \right| du > \frac{\epsilon}{2} \right].
\]

Let \( r_n(u) = \frac{1}{M} (\Gamma_0 - \Gamma_n)[u, u] + \frac{1}{M} \sum_{k=1}^{K} a_k u_k + \left( \partial \pi(\theta_0) \partial \theta_0 H_n(\theta_0) a_{n, ij}(\theta_0) A_{n, ki}(\theta_0) A_{n, kj}(\theta_0) \right) \Delta_n[u, u] + \left( \partial \pi(\theta_0) \partial \theta_0 H_n(\theta_0) a_{n, k}(\theta_0) u_k \right) \Delta_n[u, u] + \left( \partial \pi(\theta_0) \partial \theta_0 H_n(\theta_0) a_{n, 0}(\theta_0) \right) \Delta_n[u, u] \right) = O_p(1),
\]

so that \( \sup_{|u| < M} \left| \pi(\theta_0 + A_n(\theta_0)u) - \pi(\theta_0) \right| \sup_{|u| < M} Z_n(u) \) converges to 0 in probability. Hence for some \( N' \),

\[
P \left[ \int_{U_n(\theta_0)} Z_n(u) \left| \pi(\theta_0 + A_n(\theta_0)u) - \pi(\theta_0) \right| du > \epsilon \right] < \frac{\epsilon}{2} + \frac{\epsilon}{4M} \sup_{|u| \geq M} \pi(\theta) \lesssim \epsilon
\]

for every \( n \geq N \) from Assumption 2.3. Then, it is easy to see that for any \( \epsilon' > 0 \) and for some \( N' \),

\[
P \left[ \int_{U_n(\theta_0)} Z_n(u) \left| \pi(\theta_0 + A_n(\theta_0)u) - \pi(\theta_0) \right| du > \epsilon \right] < \epsilon'
\]
for all $n \geq N'$, i.e. $\int_{\mathbb{R}} Z_n(u) |\pi(\theta_0 + A_n(\theta_0)u) - \pi(\theta_0)| du \to P_{\theta_0}^U 0$. For any $\delta > 0$ and for some $K > 0$,

$$P\left[ \int_{\mathbb{R}} \left| Z_n(u) - \exp \left( \frac{1}{2} \Gamma_0[u,u] \right) \right| du > \delta \right]$$

$$\leq P\left[ \int_{\mathbb{R}\cap \{|u|<K\}} \left| Z_n(u) - \exp \left( \frac{1}{2} \Gamma_0[u,u] \right) \right| du > \frac{\delta}{2} \right]$$

$$+ P\left[ \int_{\mathbb{R}\cap \{|u|\geq K\}} \left| Z_n(u) - \exp \left( \frac{1}{2} \Gamma_0[u,u] \right) \right| du > \frac{\delta}{2} \right]$$

$$\leq P\left[ \int_{|u|<K} |Z_n(u) - \exp \left( \Delta_n[u] - \frac{1}{2} \Gamma_0[u,u] \right) | du > \frac{\delta}{2} \right] + P\left[ \int_{\mathbb{R}\cap \{|u|\geq K\}} Z_n(u) du > \frac{\delta}{4} \right]$$

$$+ P\left[ \int_{|u|\geq K} \exp \left( \Delta_n[u] - \frac{1}{2} \Gamma_0[u,u] \right) du > \frac{\delta}{4} \right]$$

$$= P\left[ \int_{|u|<K} Z_n(u) - \exp \left( \Delta_n[u] - \frac{1}{2} \Gamma_0[u,u] \right) du > \frac{\delta}{2} \right] + P\left[ \int_{\mathbb{R}\cap \{|u|\geq K\}} Z_n(u) du > \frac{\delta}{4} \right]$$

$$+ P\left[ \int_{|u|\geq K} \exp \left( \Delta_n[u] - \frac{1}{2} \Gamma_0[u,u] \right) du > \frac{\delta}{4} \right].$$

Under Assumptions 2.1 and 2.3, we can take $K > 0$ large enough such that

$$P\left[ \int_{\mathbb{R}\cap \{|u|\geq K\}} Z_n(u) du > \frac{\delta}{4} \right] < \frac{\delta}{4}, \quad (5.1)$$

$$P\left[ \int_{|u|\geq K} \exp \left( \Delta_n[u] - \frac{1}{2} \Gamma_0[u,u] \right) du > \frac{\delta}{4} \right] < \frac{\delta}{4}. \quad (5.2)$$

Since $\Delta_n[u] = \frac{1}{2} \Gamma_0[1,1] \leq \Delta_n[\Gamma_0^{-1} \Delta_n] = \frac{1}{2} \Delta_n[\Gamma_0^{-1} \Delta_n] = \frac{1}{2} \Delta_n[\Gamma_0^{-1} \Delta_n] \leq \frac{1}{2} \Delta_n[\Gamma_0^{-1} \Delta_n]$ as $\partial_n \left( \Delta_n[u] - \frac{1}{2} \Gamma_0[u,u] \right) = \Delta_n - \Gamma_0 u = 0$ if and only if $u = \Gamma_0^{-1} \Delta_n$ and Assumption 2.1 holds, for the same $K > 0$,

$$\int_{|u|<K} \left| Z_n(u) - \exp \left( \Delta_n[u] - \frac{1}{2} \Gamma_0[u,u] \right) \right| du \lesssim \sup_{|u|<K} \left| \exp \left( \Delta_n[u] - \frac{1}{2} \Gamma_0[u,u] \right) \left( \exp \{r_n(u)\} - 1 \right) \right|$$

$$\leq \sup_{|u|<K} \left| \exp \{r_n(u)\} - 1 \right| \left( \frac{1}{2} \Delta_n[\Gamma_0^{-1} \Delta_n] \right) \to P_{\theta_0} 0. \quad (5.3)$$

Because of (5.1)-(5.3), $\int_{\mathbb{R}} \left| Z_n(u) - \exp \left( \Delta_n[u] - \frac{1}{2} \Gamma_0[u,u] \right) \right| du$ is shown to converge to 0 in probability in the same way as above. Therefore we obtain that $\tilde{\epsilon}_n \to P_{\theta_0} 0$ and that

$$\log (Q_n + \tilde{\epsilon}_n) - \log Q_n = \left( \log Q_n + o_p(1) \right) - \log Q_n = o_p(1),$$

so that (i) is established.

(ii) Asymptotic behavior of the log marginal quasi-likelihood function is given by (i). Thus we should replace it by the estimator $\hat{\theta}_n$:

$$\hat{\theta}_n = A_n(\theta_0) \partial_0 \mathbb{H}_n(\theta_0)$$

$$= -A_n(\theta_0) \int_0^1 \partial^2 \mathbb{H}_n \left( \hat{\theta}_n + s(\theta_0 - \hat{\theta}_n) \right) ds [\hat{u}_n] = \hat{\Gamma}_n [\hat{u}_n],$$

where $\hat{\Gamma}_n = -A_n(\theta_0) \int_0^1 \partial^2 \mathbb{H}_n \left( \hat{\theta}_n + s(\theta_0 - \hat{\theta}_n) \right) ds$. Because of Assumption 2.1, $\hat{\Gamma}_n = \Gamma_0 + o_p(1)$ are satisfied. Therefore

$$\mathbb{H}_n(\theta_0) = H_n(\hat{\theta}_n) - \frac{1}{2} \hat{u}_n^T \Gamma_0 \hat{u}_n + o_p(1)$$

$$= H_n(\hat{\theta}_n) - \frac{1}{2} (\Gamma_0^{-1} \Delta_n) / \Gamma_0 (\Gamma_0^{-1} \Delta_n) + o_p(1)$$

$$= H_n(\hat{\theta}_n) - \frac{1}{2} ||\Gamma_0^{-1} \Delta_n||^2 + o_p(1)$$

by the Taylor expansion of $\mathbb{H}_n(\theta_0)$. Moreover, we assume that $\log a_{kn}(\theta_{kn}) - \log a_{kn}(\theta_{k,0}) = o_p(1)$ and $\log \pi(\hat{\theta}_n) - \log \pi(\theta_0) = o_p(1)$ are satisfied, so that (ii) is established.
5.3. Proof of Theorem 4.5.
Under Assumptions 4.1–4.4, we should prove that Assumption 2.1 (ii), (iii), (iv) and inequality (2.3) are satisfied. We get the following lemmas. The proof is given in Yoshida [19, Section 6].

**Lemma 5.1.**
(i) \(\forall p > 1, \sup_{n > 0} E_{\theta_0} \left[ \sup_{\theta_2 \in \Theta_2} \left| \frac{1}{\sqrt{n}} \theta_0 \mathbb{E}_n(\theta_{1,0}, \theta_2) \right|^p \right] < \infty.\)
(ii) Let \(\epsilon_1 = \frac{\epsilon}{2}. \forall p > 0, \sup_{n > 0} E_{\theta_0} \left[ \sup_{\theta_2 \in \Theta_2} \left( \sup \{ nh_n \} \right)|Y_{1,n}(\theta_1, \theta_2) - Y_{1,0}(\theta_1)| \right]^p \right] < \infty,\)
where
\[
Y_{1,n}(\theta_1, \theta_2) = \frac{1}{n} \left( \mathbb{E}_n(\theta_1, \theta_2) - \mathbb{E}_n(\theta_{1,0}, \theta_2) \right)
\]
\[
= -\frac{1}{2n} \sum_{j=1}^n \left\{ \log \frac{|B(X_{t,j-1}, \theta_1)|}{|B(X_{t,j-1}, \theta_{1,0})|} \right. \\
+ \frac{1}{nh_n} (B(X_{t,j-1}, \theta_1)^{-1} - B(X_{t,j-1}, \theta_{1,0})^{-1}) \left[ (\Delta_j X - h_n a(X_{t,j-1}, \theta_2))^2 \right].
\]
(iii) \(\forall p > 0, \sup_{n > 0} E_{\theta_0} \left[ \left( \frac{1}{\sqrt{n}} \sum_{\theta_2 \in \Theta_2} \left| \frac{1}{n} \mathbb{E}_n(\theta_{1,0}, \theta_2) - \Gamma_1,0(\theta_{1,0}) \right| \right]^p \right] < \infty.\)
(iv) \(\forall p > 0, \sup_{n > 0} E_{\theta_0} \left[ \sup_{\theta_2 \in \Theta_2} \left( \sup \{ nh_n \} \right)|Y_{2,n}(\theta_2) - Y_{2,0}(\theta_2)| \right]^p \right] < \infty,\)
where
\[
Y_{2,n}(\theta_2) = \frac{1}{nh_n} (\mathbb{E}_n(\theta_{1,0}, \theta_2) - \mathbb{E}_n(\theta_{1,0}, \theta_{2,0}))
\]
\[
= -\frac{1}{nh_n} \sum_{j=1}^n \left\{ B(X_{t,j-1}, \theta_{1,0})^{-1} |\Delta_j X, a(X_{t,j-1}, \theta_2) - a(X_{t,j-1}, \theta_{2,0})| \right. \\
+ \frac{1}{2nh_n} B(X_{t,j-1}, \theta_{1,0})^{-1} |a(X_{t,j-1}, \theta_2)^{\otimes 2} - a(X_{t,j-1}, \theta_{2,0})^{\otimes 2}|. \]
(iii) \(\forall p > 0, \sup_{n > 0} E_{\theta_0} \left[ \left( \frac{1}{\sqrt{n}} \sum_{\theta_2 \in \Theta_2} \left| \frac{1}{n} \mathbb{E}_n(\theta_{1,0}, \theta_2) \right| \right]^p \right] < \infty.\)
(iv) \(\forall p > 0, \sup_{n > 0} E_{\theta_0} \left[ \sup_{\theta_2 \in \Theta_2} \left( \sup \{ nh_n \} \right)|\mathbb{E}_n(\theta_{1,0}, \theta_2) - \Gamma_2,0(\theta_{1,0}, \theta_{2,0})| \right]^p \right] < \infty.\)

From these lemmas and sufficient conditions for PLDI (Yoshida [19, Theorem 3]), for every \(L > 0,\) some \(C_L > 0,\)
\[
P_{\theta_0} \left[ \sup_{\{(u_1, \theta_2) \in \{r \leq |u_1| \} \times \Theta_2} \mathbb{E}_n(\theta_{1,0}, \theta_2) \right] \geq e^{-r} \right] \leq \frac{C_L}{r^L},
\]
\[
P_{\theta_0} \left[ \sup_{\{u_2 \in \{r \leq |u_2| \}} \mathbb{E}_n(\theta_{1,0}, \theta_{2,0}) \right] \geq e^{-r} \right] \leq \frac{C_L}{r^L},
\]
for any \(n > 0, r > 0.\) Hence inequality (2.3) holds.

Next, we have that
\[
\left| \sup_{\theta \in \Theta} \left( \partial_{\theta_0} \partial_{\theta} \partial_{\theta_0} \mathbb{E}_n(\theta) \right) A_{n,ii}(\theta_0) A_{n,jj}(\theta_0) A_{n,kk}(\theta_0) \right|
\]
Because of Lemma 5.1 (iii),
\[ n^{-\frac{3}{2}} \left\| \sup_{\theta \in \Theta} (\partial_{\theta^i} \partial_{\theta^j} \partial_{\theta^k} H_n(\theta)) \right\|_{L_r^p} = o_p(1). \]

In a similar manner as Lemma 5.1 and Lemma 5.2, we can show that for every \( R > 0 \),
\[ n^{-\frac{1}{2}} (nh_n)^{-\frac{1}{2}} \left\| \sup_{\theta \in \Theta} (\partial_{\theta^i} \partial_{\theta^j} \partial_{\theta^k} H_n(\theta)) \right\|_{L_r^p} = o_p(1), \]
\[ n^{-\frac{1}{2}} (nh_n)^{-\frac{1}{2}} \left\| \sup_{\theta \in \Theta} (\partial_{\theta^i} \partial_{\theta^j} \partial_{\theta^k} H_n(\theta)) \right\|_{L_r^p} = o_p(1), \]
\[ (nh_n)^{-\frac{1}{2}} \left\| \sup_{\theta \in \Theta} (\partial_{\theta^i} \partial_{\theta^j} \partial_{\theta^k} H_n(\theta)) \right\|_{L_r^p} = o_p(1). \]
Thus Assumption 2.1 (iv) is established. Moreover, as a consequence of Lemma 5.1 (iv), Lemma 5.2 (iv) and (4.2), Assumption 2.1 (iii) is satisfied. Finally, we will prove the tightness of \( A_n(\theta_0) \partial H_n(\theta_0) \). Due to Lemma 5.1 (i) and Lemma 5.2 (i),
\[ |A_n(\theta_0) \partial H_n(\theta_0)| \leq \frac{1}{\sqrt{n}} \partial_{\theta^i} \partial_{\theta^j} H_n(\theta_1, \theta_2, \theta) \bigg|_{\theta = \theta_0} + \frac{1}{\sqrt{n}} \partial_{\theta^i} \partial_{\theta^j} H_n(\theta_1, \theta_2, \theta) \bigg|_{\theta = \theta_0} = O_p(1). \]

5.4. Proof of Corollary 4.11.
Because of Theorem 3 of Uchida and Yoshida [18], we can get inequality (2.3), and the proof of this theorem implies that Assumption 2.1 holds. Therefore it is enough to check the conditions [H1] and [H2] of Uchida and Yoshida [18]. As a consequence of Assumption 4.10 (i) and the compactness of \( \Theta \), \( 0 < \inf_{\theta \neq \theta_0, T, \theta \in \Theta} \exp(X'_{i} \theta) \). This inequality and Assumption 4.9 establish [H1]. If we can show that for every \( L > 0 \), there exists \( C_L > 0 \) such that \( P[\chi_0 \leq r^{-1}] \leq \frac{C_L}{r^2} \) for all \( r > 0 \), [H2] is valid. Here
\[ \chi_0 = \inf_{\theta \neq \theta_0} \frac{1}{T} \int_0^T \left\{ X'_{i}(\theta - \theta_0) + \left( \exp(X'_{i}(\theta - \theta_0)) - 1 \right) \right\} dt. \]

By using the Taylor expansion, \( \exp(x) \) becomes
\[ \exp(x) = 1 + x + \frac{1}{2} \exp(\xi x) x^2 \]
for some \( \xi \) satisfying \( 0 < \xi < 1 \). By letting \( x = X'_{i}(\theta - \theta_0) \), we obtain that
\[ X'_{i}(\theta_0 - \theta) + \left( \exp(X'_{i}(\theta_0 - \theta)) - 1 \right) = \exp(X'_{i}(\theta_0 - \theta)) - 1 - X'_{i}(\theta_0 - \theta) \]
\[ = \frac{1}{2} \exp(\xi X'_{i}(\theta_0 - \theta)) (X'_{i}(\theta_0 - \theta))^2 \]
\[ = \frac{1}{2} \exp(\xi X'_{i}(\theta_0 - \theta)) (\theta_0 - \theta)' X_i X_i' (\theta_0 - \theta) \]
\[ \geq \frac{1}{2} \exp(-C_0) (\theta_0 - \theta)' X_i X_i' (\theta_0 - \theta) \]
for some \( C_0 > 0 \). Hence
\[ \chi_0 \geq \frac{\exp(-C_0)}{4T} \inf_{\theta \neq \theta_0} \frac{1}{\sqrt{T}} \left( \theta_0 - \theta \right)^T X_i X_i'(\theta_0 - \theta) dt \]
\[ \geq \lambda_{\min} \left( \int_0^T X_i X_i' dt \right), \]
so that \( P[\chi_0 \leq r^{-1}] \leq P[\lambda_{\min} \left( \int_0^T X_i X_i' dt \right) \leq r^{-1}] \leq \frac{C_L}{r^2} \) from Assumption 4.10. The proof is complete.
REFERENCES


Graduate School of Mathematics, Kyushu University. 744 Motooka, Nishi-ku, Fukuoka 819-0395, Japan.
E-mail address: s-eguchi@math.kyushu-u.ac.jp

Institute of Mathematics for Industry, Kyushu University, 744 Motooka, Nishi-ku, Fukuoka 819-0395, Japan
CREST, JST, 744 Motooka, Nishi-ku, Fukuoka 819-0395, Japan
E-mail address: hiroki@imi.kyushu-u.ac.jp
List of MI Preprint Series, Kyushu University

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

MI

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

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

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

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

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

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

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

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

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

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

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

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

MI2008-13 Jacques FARAUT and Masato WAKAYAMA
Hermitian symmetric spaces of tube type and multivariate Meixner-Pollaczek polynomials
MI2008-14 Takashi NAKAMURA
Riemann zeta-values, Euler polynomials and the best constant of Sobolev inequality

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

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

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

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

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

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

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

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

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

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

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

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

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

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

MI2009-14 Yasunori MAEKAWA
On Gaussian decay estimates of solutions to some linear elliptic equations and its applications
MI2009-15  Yuya ISHIHARA & Yoshiyuki KAGEI
Large time behavior of the semigroup on $L^p$ spaces associated with the linearized compressible Navier-Stokes equation in a cylindrical domain

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

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

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

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

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

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

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

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

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

MI2009-25  Takehiko KINOSHITA & Mitsuhiro T. NAKAO
On very accurate enclosure of the optimal constant in the a priori error estimates for $H^2_0$-projection

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

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

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

MI2009-29  Yoshiyuki KAGEI & Yasunori MAEKAWA
Asymptotic behaviors of solutions to evolution equations in the presence of translation and scaling invariance
MI2009-30 Yoshiyuki KAGEI & Yasunori MAEKAWA
On asymptotic behaviors of solutions to parabolic systems modelling chemotaxis

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

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

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

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

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

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

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

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

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

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

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

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

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

MI2010-7 Teruhisa TSUDA
UC hierarchy and monodromy preserving deformation

MI2010-8 Takahiro ITO
Abstract collision systems on groups
MI2010-9 Hiroshi YOSHIDA, Kinji KIMURA, Naoki YOSHIDA, Junko TANAKA & Yoshihiro MIWA
An algebraic approach to underdetermined experiments

MI2010-10 Kei HIROSE & Sadanori KONISHI
Variable selection via the grouped weighted lasso for factor analysis models

MI2010-11 Katsusuke NABESHIMA & Hiroshi YOSHIDA
Derivation of specific conditions with Comprehensive Groebner Systems

MI2010-12 Yoshiyuki KAGEI, Yu NAGAFUCHI & Takeshi SUDOU
Decay estimates on solutions of the linearized compressible Navier-Stokes equation around a Poiseuille type flow

MI2010-13 Reiichiro KAWAI & Hiroki MASUDA
On simulation of tempered stable random variates

MI2010-14 Yoshiyasu OZEKI
Non-existence of certain Galois representations with a uniform tame inertia weight

MI2010-15 Me Me NAING & Yasuhide FUKUMOTO
Local Instability of a Rotating Flow Driven by Precession of Arbitrary Frequency

MI2010-16 Yu KAWAKAMI & Daisuke NAKAJO
The value distribution of the Gauss map of improper affine spheres

MI2010-17 Kazunori YASUTAKE
On the classification of rank 2 almost Fano bundles on projective space

MI2010-18 Toshimitsu TAKAESU
Scaling limits for the system of semi-relativistic particles coupled to a scalar boson field

MI2010-19 Reiichiro KAWAI & Hiroki MASUDA
Local asymptotic normality for normal inverse Gaussian Lévy processes with high-frequency sampling

MI2010-20 Yasuhide FUKUMOTO, Makoto HIROTA & Youichi MIE
Lagrangian approach to weakly nonlinear stability of an elliptical flow

MI2010-21 Hiroki MASUDA
Approximate quadratic estimating function for discretely observed Lévy driven SDEs with application to a noise normality test

MI2010-22 Toshimitsu TAKAESU
A Generalized Scaling Limit and its Application to the Semi-Relativistic Particles System Coupled to a Bose Field with Removing Ultraviolet Cutoffs

MI2010-23 Takahiro ITO, Mitsuhiko FUJIO, Shuichi INOKUCHI & Yoshihiro MIZOGUCHI
Composition, union and division of cellular automata on groups

MI2010-24 Toshimitsu TAKAESU
A Hardy’s Uncertainty Principle Lemma in Weak Commutation Relations of Heisenberg-Lie Algebra
MI2010-25  Toshimitsu TAKAESU
On the Essential Self-Adjointness of Anti-Commutative Operators

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

MI2011-4  Hiroshi INOUE, Shohei TATEISHI & Sadanori KONISHI
Nonlinear regression modeling via Compressed Sensing
MI2011-5 Hiroshi INOUE
Implications in Compressed Sensing and the Restricted Isometry Property

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

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

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

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

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

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

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

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

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

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

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

MI2011-17 Naoyuki KAMIYAMA
Matroid intersection with priority constraints

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

MI2012-2 Hiroki MASUDA
Mighty convergence of the Gaussian quasi-likelihood random fields for ergodic Levy driven SDE observed at high frequency
MI2012-3 Hiroshi INOUE
A Weak RIP of theory of compressed sensing and LASSO

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

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

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

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

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

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

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

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

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

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

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

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

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

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

MI2013-5 Hiroshi INOUE
RIPless Theory for Compressed Sensing
MI2013-6 Hiroshi INOUE
  Improved bounds on Restricted isometry for compressed sensing

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

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

MI2013-9 Naoyuki KAMIYAMA
  Packing Arborescences in Acyclic Temporal Networks

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

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

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

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

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

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

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

MI2014-1 Naoyuki KAMIYAMA
  Popular Matchings under Matroid Constraints

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

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

MI2014-4 Naoyuki KAMIYAMA
  The Popular Condensation Problem under Matroid Constraints
MI2014-5 Yoshiyuki KAGEI & Kazuyuki TSUDA
   Existence and stability of time periodic solution to the compressible Navier-Stokes
   equation for time periodic external force with symmetry

MI2014-6 This paper was withdrawn by the authors.

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

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

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

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

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

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

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

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

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

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

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

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

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

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