Parametric Inference for Nonsynchronously Observed Diffusion Processes in the Presence of Market Microstructure Noise

Teppei Ogihara
The Institute of Statistical Mathematics, 10-3 Midori-cho, Tachikawa, Tokyo 190-8562, Japan
PRESTO, Japan Science and Technology Agency
School of Multidisciplinary Sciences, SOKENDAI (The Graduate University for Advanced Studies)

Abstract. We study parametric inference for diffusion processes when observations occur nonsynchronously and are contaminated by market microstructure noise. We construct a quasi-likelihood function and study asymptotic mixed normality of maximum-likelihood- and Bayes-type estimators based on it. We also prove the local asymptotic normality of the model and asymptotic efficiency of our estimator when the diffusion coefficients are deterministic and noise follows a normal distribution. We conjecture that our estimator is asymptotically efficient even when the latent process is a general diffusion process. An estimator for the quadratic covariation of the latent process is also constructed. Some numerical examples show that this estimator performs better compared to existing estimators of the quadratic covariation.

Keywords. asymptotic efficiency, Bayes-type estimation, diffusion processes, local asymptotic normality, nonsynchronous observations, parametric estimation, maximum-likelihood-type estimation, market microstructure noise

1 Introduction

Analysis of volatility and covariation is one of the most important subjects in the study of risk management of financial assets. Studies of high-frequency financial data are increasingly significant as high-frequency financial data become increasingly available and computing technology develops. While realized volatility has been studied as a consistent estimator of integrated volatility at high-frequency limits, estimators of covariation of two securities are also important. The realized covariance, a natural extension of the realized volatility, is a consistent estimator of integrated covariation in ideal settings.

However, there are two significant problems in empirical analysis, one of which is the existence of observation noise. When we model stock price data by a continuous stochastic process, we should assume that the observations are contaminated by additional noise as a way to explain empirical evidence. Consistent estimators of volatility under the presence of microstructure noise are investigated—for example, in Zhang, Mykland, and Aït-Sahalia [31], Barndorff-Nielsen et al. [3], and Podolskij and Vetter [26]—by using various data-averaging or resampling methods to reduce the influence of noise. The other significant problem is that of nonsynchronous observation, namely, that we observe prices of different securities at different time points. The realized covariance has serious bias under models of nonsynchronous observations, though we can calculate the estimator by using some simple ‘synchronization’ methods such as linear interpolation or the ‘previous tick’ methods. Hayashi and Yoshida [15, 16, 17] and Malliavin and Mancino [22, 23] independently constructed consistent estimators for statistical models of diffusion processes with nonsynchronous observations. There are also studies of covariation estimation under the simultaneous presence of microstructure noise and nonsynchronous observations. We refer interested readers to Barndorff-Nielsen et al. [4] for a kernel based method; Christensen, Kinnebrock, and Podolskij [7], Christensen, Podolskij, and Vetter [8] for the modulated realised covariance and a pre-averaged Hayashi–Yoshida estimator; Aït-Sahalia, Fan, and Xiu [2] for a method using the maximum likelihood estimator of a model with constant diffusion coefficients; and Bibinger et al. [5] for a technique employing the local method of moments.

While the above studies concern estimators under non- or semi-parametric settings, there are also stud-

One advantage of maximum-likelihood- and Bayes-type estimators is that they are asymptotically efficient in many models. If a statistical model has the local asymptotic mixed normality (LAMN) property, then the results in Jeganathan [20, 21] ensure that asymptotic variance of estimators cannot be smaller than a certain lower bound. When some estimator attains this bound, it is called asymptotically efficient. For parametric estimation of diffusion processes on fixed intervals, Gobet [14] proved the LAMN property of the statistical model having equidistant observations, and an estimator in [11] is asymptotically efficient. Ogihara [24] proved the LAMN property and asymptotic efficiency of estimators for the setting of [25]. Gloter and Jacod [12] proved the local asymptotic normality (LAN) property for a statistical model with market microstructure noise when diffusion coefficients are deterministic, and the estimator by Gloter and Jacod [13] is asymptotically efficient. There are few studies about the efficiency of estimators that assume the presence of market microstructure noise and nonsynchronous observations. One exception is Bibinger et al. [5], who showed a lower bound of asymptotic variance of estimators in semi-parametric Cramér-Rao sense. We need the LAN or LAMN property of the statistical model to obtain asymptotic efficiency of a parametric model. To the best of our knowledge, this has not been studied for statistical models of noisy, nonsynchronous observations.

This paper examines consistency and asymptotic mixed normality of a maximum-likelihood-type estimator and a Bayes-type estimator based on a quasi-likelihood function, under the simultaneous presence of market microstructure noise and nonsynchronous observations. We also study the LAN property of this model when diffusion coefficients are deterministic, as well as the asymptotic efficiency of our estimators. We expect that our estimators are asymptotically efficient in the general cases. However, it is further difficult to obtain LAMN properties for models of general diffusions. This does not seem to have been obtained even for noisy, equidistant observations, and is left as future work. We will see by simulation that sample variance of the estimation error of our estimator is better than that of existing estimators for some examples in Section 3. These results ensure that our estimator not only is the theoretical best for asymptotic behavior, but also works well in practical finite samplings.

Our study has several advantages in addition to the above arguments regarding asymptotic efficiency.

i) Our model also allows observation noise that follows a non-Gaussian distribution. We use a quasi-likelihood function for Gaussian noise, but our method is robust enough to allow misspecification of the noise distribution.

ii) Since we obtain the results regarding asymptotic behaviors of the quasi-likelihood function as a byproduct, many applications become available from the theory of maximum-likelihood-type estimation. For example, we can construct a theory of the likelihood ratio test and one-step estimators as an immediate application. Further, the theory of information criteria is expected to follow from our results of quasi-likelihood functions.

iii) Our settings contain random sampling schemes where the maximum length of observation intervals is not bounded by any constant multiplication of the minimum length. This is the case for some significant random sampling schemes, such as samplings based on Poisson or Cox processes. Our model encompasses such natural sampling schemes.

To obtain asymptotic mixed normality of our estimator, we investigate asymptotic behaviors of a quasi-likelihood function of noisy, nonsynchronous observations. To this end, we need to specify the limit of some matrix trace related to a ratio of covariance matrices for two different values of parameters, as appearing in (5.2). The inverse of the covariance matrix of observation noise has nontrivial off-diagonal elements, and so the inverse of the covariance matrix of observations is far from a diagonal matrix. This phenomenon is essentially different from the case of synchronous observations without noise (where the covariance matrix of observations is diagonal), and the case of nonsynchronous observations without noise (where the inverse of the covariance matrix is not a diagonal matrix but is ‘close’ to being one).

In a model of noisy, synchronous observations, the covariance matrix of a latent process is asymptotically equivalent to a unit matrix of the appropriate size, and is therefore simultaneously diagonalizable with the noise.
covariance. Gloter and Jacod [12, 13] used these facts and closed expressions for the eigenvalues of the noise covariance to identify the limit of the quasi-likelihood function, but we cannot apply their idea because our sampling scheme is irregular and so not well approximated by a unit matrix. Further, the sizes of the covariance matrices are different for different components of the process, which follows from nonsynchronousness. In this paper, we deduce an asymptotically equivalent transform of the trace of the ratio of covariance matrices. This transform changes sizes of matrices and matrix elements into local averages, and arises from specific properties of the noise covariance matrix. We will see these results in Sections 4 and 5.

The remainder of this paper is organized as follows. In Section 2, we describe our detailed settings and main results. We propose a quasi-likelihood function for models with noisy, nonsynchronous observations, and construct a maximum-likelihood-type estimator based on it. We introduce asymptotic mixed normality of our estimator and results about asymptotic efficiency in Section 2.2. Section 2.3 contains results about the LAN property of our model and the asymptotic efficiency of our estimator, and Section 2.4 is devoted to results about Bayes-type asymptotic properties of the quasi-likelihood function in Section 5. We also need some results on identifiability of other estimators. The remaining sections are devoted to a proof of the main results. Section 4 introduces an estimator of the quadratic covariation and compare the performance of our estimator with that of other estimators. The remaining sections are devoted to a proof of the main results. Section 4 introduces an asymptotically equivalent expression of the quasi-likelihood function. This expression is useful for deducing asymptotic properties of the quasi-likelihood function in Section 5. We also need some results on identifiability of the model to obtain consistency of the maximum-likelihood-type estimator. These are discussed in Section 6. Section 7 shows asymptotic mixed normality of our estimator. The LAN property of the model for deterministic diffusion coefficients is obtained in Section 8. Section 9 contains a proof of results regarding the Bayes-type estimator and the convergence of moments of estimators.

2 Main results

2.1 Settings and construction of the estimator

Let $(\Omega^{(0)}, \mathcal{F}^{(0)}, \mathbb{P}^{(0)})$ be a probability space with a filtration $\mathbb{F}^{(0)} = \{\mathcal{F}^{(0)}_t\}_{0 \leq t \leq T}$. We consider a two-dimensional $\mathbb{F}^{(0)}$-adapted process $Y = \{Y_t\}_{0 \leq t \leq T}$ satisfying the stochastic integral equation:

$$Y_t = Y_0 + \int_0^t \mu_s ds + \int_0^t b(s, X_s, \sigma_s) dW_s, \quad t \in [0, T],$$

(2.1)

where $\{W_t\}_{0 \leq t \leq T}$ is a $d_1$-dimensional standard $\mathbb{F}^{(0)}$-Wiener process, $b = (b^j)_1 \leq j \leq d_1$ is a Borel function, $\mu = \{\mu_t\}_{0 \leq t \leq T}$ is a locally bounded $\mathbb{F}^{(0)}$-adapted process with values in $\mathbb{R}^d$, and $X = \{X_t\}_{0 \leq t \leq T}$ is a continuous $\mathbb{F}^{(0)}$-adapted processes with values in $O$, an open subset of $\mathbb{R}^{d_2}$ with $d_2 \in \mathbb{N}$. We consider market microstructure noise $\{\epsilon_{n,k}^i\}_{n \in \mathbb{N}, k = 1, 2}$ as an independent sequence of random variables on another probability space $(\Omega^{(1)}, \mathcal{F}^{(1)}, \mathbb{P}^{(1)})$. We assume that $\mathcal{F}^{(1)} = \mathcal{B}(\epsilon_{n,k}^i)$ and that the distribution of $\epsilon_{n,k}^i$ does not depend on $j$, where $\mathcal{B}(S)$ denotes the minimal $\sigma$-field such that any element of $S$ is $\mathcal{B}(S)$-measurable for a set $S$ of random variables. We use the same notation $\mathcal{B}(S)$ for a similarly defined $\sigma$-field for a set $S$ of measurable sets. We consider a product probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where $\Omega = \Omega^{(0)} \times \Omega^{(1)}$, $\mathcal{F} = \mathcal{F}^{(0)} \otimes \mathcal{F}^{(1)}$, and $\mathbb{P} = \mathbb{P}^{(0)} \otimes \mathbb{P}^{(1)}$.

We assume that the observations of processes occur in a nonsynchronous manner and are contaminated by market microstructure noise, that is, we observe the vectors $\{Y_{i}^{(k)}\}_{0 \leq i \leq J_{k,n,k}^{i}, k = 1, 2}$ and $\{X_{i}^{(k)}\}_{0 \leq j \leq J_{k,n,k}^{j}, 1 \leq k \leq d_2}$, where $\{S_{i}^{n,k}\}_{i=0}^{J_{k,n,k}^{i}}$ and $\{T_{j}^{n,k}\}_{j=0}^{J_{k,n,k}^{j}}$ are random times in $(\Omega^{(0)}, \mathcal{F}^{(0)})$, $\{\eta_{j}^{n,k}\}_{j=1}^{J_{k,n,k}^{j}}$ is a random sequence on $(\Omega, \mathcal{F})$, and

$$\hat{Y}_{i}^{k} = Y_{i}^{k} + \epsilon_{i}^{n,k}, \quad \hat{X}_{j}^{k} = X_{j}^{k} + \eta_{j}^{n,k}. $$

(2.2)

Our goal is to estimate the true value $\sigma_\ast$ of the parameter from nonsynchronous, noisy observations $\{S_{i}^{n,k}\}_{0 \leq i \leq J_{k,n,k}^{i}, k = 1, 2}$, $\{T_{j}^{n,k}\}_{0 \leq j \leq J_{k,n,k}^{j}, 1 \leq k \leq d_2}$, $\{\hat{Y}_{i}^{k}\}_{0 \leq i \leq J_{k,n,k}^{i}, k = 1, 2}$, and $\{\hat{X}_{j}^{k}\}_{0 \leq j \leq J_{k,n,k}^{j}, 1 \leq k \leq d_2}$.

By setting $d_2 = 2$, $X_t \equiv Y_t$, $\mu = \mu(t, Y_t)$, $S_{i}^{n,k} \equiv T_{i}^{n,k}$, and $\eta_{j}^{n,k} \equiv \epsilon_{i}^{n,k}$, our model contains the case where the latent process $Y$ is a diffusion process satisfying a stochastic differential equation

$$dY_t = \mu(t, Y_t) dt + b(t, Y_t, \sigma_\ast) dW_t, \quad t \in [0, T],$$

(2.3)
and $Y$ is observed in a nonsynchronous manner with noise. This model is of particular interest, but our results are also be applied to more general models (2.1).

**Remark 2.1.** Stochastic volatility models are significant models for modeling stock prices. Unfortunately, our settings are not applied to hidden Markov models including stochastic volatility models because we require (possibly noisy) observations of process $X$. However, we hope that our results give an essential idea to deal with noisy, nonsynchronous observations, and therefore we can construct an estimator for stochastic volatility models by replacing our quasi-likelihood function. We have left it for future works.

For a vector $x = (x_1, \ldots, x_k)$, we denote $\partial^k_x = (\frac{\partial^k}{\partial x_1^k} \cdots \frac{\partial^k}{\partial x_k^k})_{j_1, \ldots, j_k = 1}$. We assume the true value $\sigma_*$ of the parameter is contained in a bounded open set $\Lambda \subset \mathbb{R}^d$ that satisfies Sobolev’s inequality; that is, for any $p > d$, there exists $C > 0$ such that $\sup_{x \in \Lambda} |u(x)| \leq C \sum_{k=0,1} (\int_{\Lambda} |\partial^k u(x)|^p dx)^{1/p}$ for any $u \in C^1(\Lambda)$. This is the case when $\Lambda$ has a Lipschitz boundary. See Adams and Fournier [1] for more details.

Let $\Pi_n = \{\{S_{t,n}^{n,k}\}_{k,i}, \{T_{j,n}^{n,k}\}_{j,k}\}$ and $\{G_i\}_{0 \leq i \leq T}$ be a filtration of $(\Omega, \mathcal{F}, P)$ given by

$$G_i = \mathcal{F}^{(0)}_i \bigvee \mathcal{B}(\Pi_{i,n}) \bigvee \mathcal{B}(A \cap \{S_{t,n}^{n,k} \leq t\}, A \in \mathcal{B}(\ell_{t,n}^{n,k}), k \in \{1,2\}, i \in \mathbb{Z}_+, n \in \mathbb{N}),$$

where $\mathcal{H}_1 \vee \mathcal{H}_2$ denotes the minimal $\sigma$-field which contains $\sigma$-fields $\mathcal{H}_1$ and $\mathcal{H}_2$. Let $E[\Pi[X] = E[X|\Pi_{i,n}]$ for a random variable $X$. We assume that $\mathcal{F}^{(0)}_i$ and $\mathcal{B}(\Pi_{i,n})$ are independent. Moreover, we assume that there exist positive constants $\varepsilon_{1,*}$ and $\varepsilon_{2,*}$ such that $\frac{n^{k-1}}{\gamma_{n,k}} \mathbb{I}_{\{\gamma_{n,k} \leq 1\}}$ is $\mathcal{G}_i$-measurable, $E[\epsilon_{1,k}^2] = 0$, and $E[\epsilon_{1,k}^2] = \epsilon_{2,k}$ for any $n, k, i, j, t$, where $I_A$ is the indicator function for a set $A, \delta_{ij}$ is Kronecker’s delta. We also assume that the distribution of $Y_0$ does not depend on $\sigma, \varepsilon_{1,*}$, nor $\varepsilon_{2,*}$.

Now we construct the quasi-likelihood function. We apply the idea of Gloter and Jacod [13] to our construction of a quasi-likelihood function; that is, we divide the whole observation interval $[0, T]$ into equidistant subdivisions and construct quasi-likelihood functions for each interval as follows. Let $\{T_{i,n}\}_{n \in \mathbb{N}}$ and $\{N_{i}\}_{n \in \mathbb{N}}$ be sequences of positive numbers satisfying $b_n \geq 1$, $b_n \leq b_n, b_n \rightarrow \infty$, $b_n^{1/2 - \epsilon} \rightarrow \infty$, and $b_n^{2/3 + \epsilon} \rightarrow 0$ as $n \rightarrow \infty$ for some $\epsilon > 0$. We will assume in Assumption (A2) a relation between $b_n$ and our sampling scheme, which implies that $b_n$ represents the order of observation frequency. Let $\epsilon_n = \frac{b_n}{b_m - 1}, s_0 = 0, s_m = \frac{b_n}{b_m - 1}, s_m = \frac{b_n}{b_m - 1}$, $K_0 = -1$, and $K_m = \# \{i \in \mathbb{N} \cap S_{i,n}^{n,k} < s_m \}$ for $k \in \{1,2\}$ and $1 \leq m \leq \epsilon_n$.

Moreover, let $k_i^0 = K_i - K_{i-1} - 1$, $k_i = \text{max}_{m,j} k_i^0, k_i = \text{min}_{m,j} k_i^0$, and $b_i^0 = b_i^0(s_{m-1}^0, \hat X_{m-1}), (s_{m-1}^0, \hat Y_{m-1})$ for $1 \leq m \leq \epsilon_n, 2 \leq m' \leq \epsilon_n, j \in \{1,2\}$ and $1 \leq i \leq k_m$. Then, roughly speaking, we have the following approximations of conditional covariance of observations:

$$E[\hat Y_k(I_{t,m}^0)^2|G_{s_{m-1}^0}] \approx (b_i^0 |I_{t,m}^0|^2 + 2b_i^0)\delta_{i''} - k_{i''}1_{[j = i'' = 1]},$$

for any intervals $I_{t,m}^0, I_{t,m}^1, I_{t,m}^2, I_{t,m}^3$. Let $T$ denotes the transpose operator for matrices (and vectors), $M(l) = \{2(1_{[i+1, l]} - 1_{[i, l-1]}\}^T 1_{i+1, l}$ for $l \in \mathbb{N}$, $M_{i,j} = M(k_{i,j})$ for $1 \leq j \leq 2$. Based on the relation (2.4), we define a quasi-log-likelihood function $H_n(\sigma, v)$ by

$$H_n(\sigma, v) = \frac{1}{2} \sum_{m=2}^{\epsilon_n} \log \det \mathcal{S}_m(\sigma, v) - \frac{1}{2} \sum_{m=2}^{\epsilon_n} Z_m^T S_m^{-1}(\sigma, v) Z_m,$$

for $v = (v_1, v_2)$.
Remark 2.2. Though such a local Gaussian quasi-log-likelihood function seems valid only when observation noise $\epsilon_{n,k}$ follows a Gaussian distribution, asymptotic properties of the maximum likelihood estimator are robust enough to allow non-Gaussian noise. Hence, we can use the same quasi-likelihood function for general noise.

Remark 2.3. We used subdivisions of $[0, T]$ for the construction of $H_n$ because of technical issues related to deducing the limit of $H_n$. Since the diffusion coefficient $b$ in $S_m$ is fixed, matrix properties of $M_{1,m}$ introduced in Section 4.2 can be used to deduce the limit of $H_n$. On the other hand, such a construction of $H_n$ also contributes to reducing the calculation time of the maximum-likelihood-type estimator because the size of $S_m$ is $O(k_n)$ while the size of the covariance matrix of all observations is $O(b_n)$.

Remark 2.4. In [13], $k_n$ is taken so that $n^{1/2}k_n^{-1} \rightarrow 0$ and $k_n n^{-3/4} \rightarrow 0$. Our rate $b_n^{2/3}$ for the upper bound of $k_n$ is a little bit worse because of some technical issue (for equidistance observations, we have $b_n \equiv n$). When we investigate asymptotic behaviors of the maximum-likelihood-type estimator, we deal with some supremum processes $Y$ for details.

Remark 2.5. Section 4.2 can be used to deduce the limit of $H_n(\hat{\sigma}_n, \hat{\epsilon}_n)$, which contributes to the size of the covariance matrix of all observations is $O(k_n)$.

To construct the maximum-likelihood-type estimator $\hat{\sigma}_n$ for the parameter $\sigma$, we need estimators for the unknown noise variance $\nu_s = (v_1, *, v_2, *)$. We assume the following condition.

Assumption (V). There exist estimators $\{\hat{\nu}_n\}_{n \in \mathbb{N}}$ of $\nu_s$ such that $\hat{\nu}_n \geq 0$ almost surely and $\{b_n^{1/2}(\hat{\nu}_n - \nu_s)\}_{n \in \mathbb{N}}$ is tight.

For example, $\hat{\nu}_n = (\bar{\nu}_n,k)_{k=1}^2$ with $\bar{\nu}_n,k = (2J_{k,n})^{-1} \sum_i (\bar{\epsilon}_i,k - \bar{\epsilon}_i,k-1)^2$ satisfies (V) if $\{b_nJ_{k,n}^{-1}\}_n$ is tight for $k = 1, 2$ and $\sup_{n,k,i} E[(\bar{\epsilon}_i,n,k)^4] < \infty$.

Let $\text{clos}(A)$ be the closure of a set $A$. A maximum-likelihood-type estimator $\tilde{\sigma}_n$ is a random variable satisfying

$$H_n(\tilde{\sigma}_n, \hat{\nu}_n) = \max_{\sigma \in \text{clos}(\mathbb{R})} H_n(\sigma, \hat{\nu}_n).$$

We study asymptotic mixed normality and asymptotic efficiency of the estimator in the following subsections.

Remark 2.5. We can also construct a simultaneous maximum-likelihood-type estimator $(\tilde{\sigma}_n, \hat{\epsilon}_n)$ satisfying $H_n(\tilde{\sigma}_n, \hat{\epsilon}_n) = \max_{\sigma, \nu_s} H_n(\sigma, \nu_s)$. However, it is valid only when the observation noise $\epsilon_{n,k}$ follows a normal distribution. Our interest is on estimating the parameter $\sigma$ of the latent process, and so the assumptions for observation noise should be reduced as much as possible. Therefore, the nonparametric estimator $\hat{\epsilon}_n$ is more suitable for our purpose.

2.2 Asymptotic mixed normality of the maximum-likelihood-type estimator

In this section, we state our main theorems. Proofs of these results are left to Sections 4–9. In this subsection, we describe the asymptotic mixed normality of the maximum-likelihood-type estimator $\hat{\sigma}_n$.

We first describe assumptions for the theorem. Assumption (A1) is a sequence of assumptions on the latent processes $Y$ and $X$ and observation noise $\epsilon_{n,k}$ and $\eta_{j,n,k}$. We denote by $E_t$ the unit matrix of size $l$.

Assumption (A1). 1. For $0 \leq 2i + j \leq 4$ and $0 \leq k \leq 4$, the derivatives $\partial^i_j \partial^j_k b(t, x, \sigma)$ exist on $[0, T] \times \Omega \times \Lambda$ and have continuous extensions on $[0, T] \times \Omega \times \text{clos}(\Lambda)$.
2. $b b^T(t, x, \sigma)$ is positive definite for $(t, x, \sigma) \in [0, T] \times \Omega \times \text{clos}(\Lambda)$.
3. $\text{sup}_{n,k,i} E[|\bar{\epsilon}_{i,n,k}|^q] < \infty$ for any $q > 0$.
4. $\nu_s$ is optional and locally bounded (locally in time), that is, there exists an increasing sequence $\{T_i\}_i$ of stopping times such that $\lim_{i \to \infty} T_i = T$ a.s. and $\{\mu_{T,T}^{\infty}\}_{0 \leq t \leq T}$ is bounded for each $l$.
5. $b_n^{1/2} \text{max}_{m,k,i} E[|\epsilon_{i,n,m,k}|^q \vee |\epsilon_{i,n,m,k}|^q] \rightarrow 0$ as $n \to \infty$ for any $q > 0$ and $\epsilon > 0$, where $T_{m,k} = \# \{j; T_j^{m,k} \in [s_{m-1}, s_m)\}^{-1} \sum_{j:T_j^{m,k} \in [s_{m-1}, s_m]} \eta_{j,n,k}^m$.

5
6. There exist progressively measurable processes \( \{b_t^{(j)}\}_{0 \leq t \leq T, 0 \leq j \leq 1} \) and \( \{\hat{b}_t^{(j)}\}_{0 \leq t \leq T, 0 \leq j \leq 1} \) such that \( b_t^{(j)}, \hat{b}_t^{(j)} \), and \( \sup_{0 \leq s \leq t} (|b_s^{(j)} - \hat{b}_s^{(j)}| \vee |\hat{b}_s^{(j)} - \hat{b}_t^{(j)}|)/\sqrt{|s - u|^{1/2}} \) are locally bounded processes for \( 0 \leq j \leq 1 \), and

\[
X_t = X_0 + \int_0^t b_s^{(0)} \, ds + \int_0^t \hat{b}_s^{(1)} \, dW_s, \quad b_t^{(1)} = \hat{b}_0^{(1)} + \int_0^t b_s^{(0)} \, ds + \int_0^t \hat{b}_s^{(1)} \, dW_s
\]

for \( t \in [0, T] \).

Assumption (A1) captures somewhat standard assumptions and whether it holds can easily verified in practical settings. Roughly speaking, point 5 of (A1) is satisfied if the summation of \( \eta^{n,k}_j \) is of an order equivalent to the square root of the number of \( n^{k,n}_j \). This is satisfied under certain independence, martingale conditions or mixing conditions of \( \eta^{n,k}_j \). If \( \eta^{n,k}_j \) is a sequence of centered, independent and identically distributed random variables and the sequence has finite moments, then \( E_{\Pi}[|T_{m,k}|]^q = O_p(\#\{j; T_{m,k}^{n,k} \in [s_m-1, s_m]\}^{-q/2}) \). Then, point 5 of (A1) is satisfied if sampling frequency of \( \{T_{m,k}^{n,k}\} \) is of order \( b_n \) and \( E[\eta^{n,k}_j 1_{(T_{m,k}^{n,k} > s_m-1)}] \leq 0 \). Decomposition of \( X \) in point 6 of (A1) is used to deduce asymptotically equivalent representation of \( H_n \) where the diffusion coefficient \( b(t, X_t, \sigma_\ast) \) is replaced by \( b(s_m-1, X_{s_m-1}, \sigma_\ast) \). Detailed semimartingale decomposition is required to estimate the difference \( b(t, X_t, \sigma_\ast) - b(s_m-1, X_{s_m-1}, \sigma_\ast) \).

In the following, we assume some conditions about our sampling scheme. For \( \eta \in (0, 1/2) \), let \( S_{\eta} \) be the set of all sequences \( \{[s_n^{(j)}, s_{n+1}^{(j)}], n \in \mathbb{N}, 1 \leq j \leq n \} \) of intervals on \([0, T]\) satisfying \( \{L_{n,j} \} \subset \mathbb{N}, [s_{n+1}^{(j)}, s_{n+2}^{(j)}] \cap [s_{n+1}^{(j)}, s_{n+2}^{(j)}] = \emptyset \) for \( n, t_l \neq t_2 \), inf \( n, t_l \neq t_2, \inf \{n_{l+1}^{(j)}, s_{n+1}^{(j)} - s_{n+2}^{(j)}\} > 0 \), and \( \sup_{n,t_l} b_n^{(j)} - (s_{n+1}^{(j)} - s_{n+2}^{(j)}) < \infty \). Let \( r_n = \max_{1 \leq k \leq n} |S_n^{n,k} - S_n^{k,n}| \) and \( \Xi_n = \min_{1 \leq k \leq n} |S_n^{n,k} - S_n^{k,n}| \).

**Assumption (A2).** There exist \( \eta \in (0, 1/2), \kappa > 0, \tilde{\eta} \in (0, 1) \) and positive-valued stochastic processes \( \{a_t^{(j)}\}_{0 \leq t \leq T, j = 1, 2} \) such that \( \sup_{s \in \mathbb{R}} \left( |a_t^{(j)}| / |t - s|^{\eta} \right) < \infty \) almost surely, \( b_n^{1/2 + \kappa} \max_{1 \leq j \leq L_n} |b_n^{-1}(s_{n+1}^{(j)} - s_{n+2}^{(j)})|^{-1} \# \{i; [S_{n+1}^{(j)}, S_{n+2}^{(j)}] \subset [s_{n+1}^{(j)}, s_{n+2}^{(j)}]\} - a_n^{(j)} \to^p 0 \) (2.7)

as \( n \to \infty \) for \( j = 1, 2 \) and \( \{|s_{n+1}^{(j)}, s_{n+2}^{(j)}|\}_{1 \leq j \leq L_n, n \in \mathbb{N}} \in S_{\eta} \). Moreover, \( (r_n b_n^{1-\epsilon} ) \vee (b_n^{1-\epsilon} \Xi_n) \to^p 0 \) for any \( \epsilon > 0 \).

In particular, Assumption (A2) implies \( b_n^{-1} \mathbf{J}_{j,m} \to \int_0^T a_t^{(j)} \, dt \) and \( \max_{n} |r_n^{-1} k_n^{1-\kappa} - a_n^{(j)}| \to^p 0 \) as \( n \to \infty \). Roughly speaking, (A2) shows the law of large numbers for sampling schemes in any local time intervals. In the proof of Lemma 5.2, we will see that some properties of \( M_{f,m} \) enable us to replace \( |f_t| \) in \( S_{m,v} \) by the local average in asymptotics. Then (A2) leads to the limit of \( H_n \).

**Example 2.1.** Let \( \{|N_t^k|\}_{k \geq 0} \) be an exponential \( \alpha \)-mixing point process with stationary increments for \( k = 1, 2 \). Assume that \( E[|N_t^k|] < \infty \) for any \( t > 0 \) and \( k = 1, 2 \). Set \( S_n^{n,k} = \inf \{t \geq 0; N_n^k(t) \geq 1\} \). Then Rosenthal-type inequalities (Theorem 3 and Lemma 2 in Doukhan and Louch bi[10]), or Theorem 4 in [25]) and a similar argument to the proof of Proposition 6 in [25] ensure (A2) with \( a_t^{(j)} = E[N_t^j] \) (constants). Also, (B2) (defined later) is satisfied if further \( k_n b_n^{4/\gamma + \gamma} \to 0 \) for some \( \gamma > 0 \).

Under the above conditions, we can show convergence of the quasi-likelihood ratio \( H_n(\sigma, \hat{\sigma}_n) - H_n(\hat{\sigma}_n, \hat{\sigma}_n) \). Let \( b(t, X_t, \sigma), b_t, \sigma = b(t, X_t, \sigma), \tilde{a}_1^{(j)} = a_t^{(j)}/v_{j,s} \) for \( j = 1, 2 \) and

\[
\mathcal{Y}_t(\sigma) = \int_0^T \left\{ \sum_{j=1}^2 \left( (|\tilde{a}_1^{(j)}|^2 - |\tilde{a}_1^{(j)}|^2) \sqrt{\tilde{a}_1^{(j)} \tilde{a}_2^{(j)} + \tilde{a}_1^{(j)} \sqrt{\det(b_t b_t^\top)}} - 2 \tilde{a}_1^{(j)} \tilde{b}_t - \tilde{b}_t \tilde{a}_1^{(j)} + \tilde{b}_t \tilde{a}_2^{(j)} + \tilde{b}_t \tilde{a}_2^{(j)} + \tilde{b}_t \tilde{a}_2^{(j)} + \tilde{a}_2^{(j)} \tilde{a}_2^{(j)} \right) \right\} dt,
\]

for \( j = 1, 2 \).
Proposition 2.1. Assume (A1), (A2) and (V). Then sup_σ∈A |b_n^{−1/2} \partial_σ^k (H_n(σ, \tilde{v}_n) - H_n(σ_*, \tilde{v}_n)) - \partial_σ^k \mathcal{Y}_1(σ)| \to p 0 as n \to \infty for 0 \leq k \leq 3.

To show consistency and asymptotic normality of \( \hat{\sigma}_n \), the limit function \( \mathcal{Y}_1(σ) \) of the quasi-likelihood ratio should have the unique maximum point at \( σ = σ_* \). More precisely, we use the following as a kind of identifiability condition: \( \inf_{σ \neq σ_*} (\mathcal{Y}_1(σ))/|σ - σ_*|^2 > 0 \) almost surely. Though it is difficult to directly check this condition in general, we can check it under a more tractable sufficient condition. Let

\[
\mathcal{Y}_0(σ) = -\frac{1}{2} \int_0^T \left\{ \text{tr}((b_t b_t^\top)^{-1}(b_t, b_t^\top_1, s_2)) + \log (\det(b_t b_t^\top)/\det(b_t_1 b_t_1^\top)) \right\} dt.
\]

Then \( \mathcal{Y}_0 \) is the probability limit \( n^{-1/2}(H_n^0(σ) - H_n^0(σ_*)) \), where \( H_n^0 \) represents a quasi-likelihood function for a statistical model of equidistant observations without noise. See Uchida and Yoshida [28].

Assumption (A3). \( \inf_{σ \neq σ_*} (\mathcal{Y}_0(σ))/|σ - σ_*|^2 > 0 \) almost surely.

We will show in Proposition 6.1 that (A3) is sufficient for the identifiability condition of our model. Moreover, the following condition is a simple sufficient condition for (A3) (see Remark 4 in Ogihara and Yoshida [25] for the details):

Assumption (A3'). \( \inf_{σ_1 \neq σ_2} |(b_t^\top(t, x, σ_1) - b_t^\top(t, x, σ_2))/|σ_1 - σ_2| | > 0 \) for any \( t \in [0, T] \) and \( x \in O \).

We denote by \( \to^{s-L} \) the stable convergence of random variables. Let

\[
\Gamma_{1,n} = -b_n^{-1/2} \partial_σ^2 H_n(\hat{σ}_n, \tilde{v}_n), \quad \Gamma_1 = -\partial_σ^2 \mathcal{Y}_1(σ_*) \tag{2.9}
\]

Let \( N \) be a \( d \)-dimensional random variable on some extension \( (\hat{Ω}, \hat{F}, \hat{P}) \) of \( (Ω, F, P) \) satisfying the condition that \( N \) is independent of \( F \) and \( N \) follows the \( d \)-dimensional standard normal distribution. We denote the expectation with respect to \( \hat{P} \) by the same notation \( E \).

The following theorem is one of our main results.

Theorem 2.1. Assume (A1)–(A3) and (V). Then \( \Gamma_1 \) is positive definite almost surely and \( b_n^{1/4} (\hat{σ}_n - σ_*) \to^{s-L} \Gamma_1^{-1/2} N \) as \( n \to \infty \). Moreover, \( \Gamma_{1,n} \to p \Gamma_1 \), and therefore \( b_n^{1/4} \Gamma_{1,n} \to p \mathcal{Y}_1(σ_*) \) as \( n \to \infty \).

Corollary 2.1. Assume (A1), (A2), (A3') and (V). Then the results in Theorem 2.1 hold true.

Let \( \Phi_1^{(k)} = \partial_σ^k (a_t^0 b_t^0)/2 + \tilde{a}^2_t|b_t|^2 + 2\sqrt{\tilde{a}^2_t|b_t|^2} \det(b_t b_t^\top)|_{σ = σ_*} B_1^{(k)} = \partial_σ^k |b_t|^2|_{σ = σ_*}, B_3^{(k)} = \partial_σ^k b_t^1 - b_t^1|_{σ = σ_*} \), and \( \Phi_4^{(k)} = \partial_σ^k \sqrt{\det(b_t b_t^\top)}|_{σ = σ_*} \) for \( k = 0, 1 \) and \( j = 1, 2 \). Then we can rewrite \( \Gamma_1 \) as

\[
\Gamma_1 = \int_0^T \left\{ \frac{(\Phi_1^{(1)}(t) \Phi_1^{(1)}(t)^\top)}{2 \Phi_2^{(0)}(t)^2} + \frac{\sqrt{\tilde{a}^2_t|b_t|^2}}{2 \Phi_4^{(0)}(t)^2} \left( B_4^{(1)}(t)^{1/2} - B_1^{(1)}(t)^{1/2} + B_2^{(1)}(t)^{1/2} + B_3^{(1)}(t)^{1/2} \right) \right\} dt. \tag{2.10}
\]

The derivation of (2.10) is left to the appendix.

2.3 On the LAMN property and asymptotic efficiency of the estimator

In this subsection, we state some results on the so-called LAMN(LAN) property for our model and asymptotic efficiency of our estimator. We also comment on some further studies.

Throughout this subsection, we assume that \( Y_t \equiv Y_t, \mathcal{T}^{n,k}_t \equiv \mathcal{T}^{n,k}_t, \mathcal{Y}^{n,k}_t \equiv \mathcal{Y}^{n,k}_t, \epsilon_t \equiv \epsilon_t, \mu_t = \mu(t, σ_*) \) and \( Y_0 = γ \) for some Borel function \( μ \) and some known \( γ \in \mathbb{R}^2 \). Then the latent process \( Y \) is a diffusion process satisfying the stochastic differential equation (2.3) with \( μ = μ(t, σ_*) \). Let \( P_{σ^*, μ^*} \) be the distribution of \( (S_i^{n,k}_{t} k,i, Y_i^{n,k}_{t} k,i) \) with true values \( (σ^*_i, μ^*_i) \) of the parameters. We denote

\[
\text{diag}(A, B) = \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}
\]

for square matrices \( A \) and \( B \). Let \( \mathcal{Y}_2(v) = -\int_0^T \sum_{j=1}^2 a_t^j \{ (v_{j,}\tilde{v}_{j}) - 1 + \log(v_{j,}\tilde{v}_{j}) \} dt/2, \Gamma_2 = -\partial_{σ_*}^2 \mathcal{Y}_2(v_*) \) and \( \Gamma = \text{diag}(\Gamma_1, \Gamma_2) \). \tag{2.11}

We adopt the following definition of the LAMN property from Jeganathan [21].
Definition 2.1. Let $P_{\theta,n}$ be a probability measure on some measurable space $(X_n, A_n)$ for each $\theta \in \Theta$ and $n \in \mathbb{N}$, where $\Theta$ is a bounded open subset of $\mathbb{R}^d$. Then the family $\{P_{\theta,n}\}_{\theta,n}$ satisfies the local asymptotic mixed normality (LAMN) property at $\theta = \theta_*$ if there exists a sequence $\{\delta_n\}_{n \in \mathbb{N}}$ of $d \times d$ positive definite matrices, $d \times d$ symmetric random matrices $\Gamma_n$, $\Gamma$ and $d$-dimensional random vectors $N_n, N$ such that $\Gamma$ is positive definite a.s., $P_{\theta_*,n}[\Gamma_n$ is positive definite] = 1 (for $n \in \mathbb{N}$), $|\delta_n| \to 0$, and

$$\log \frac{dP_{\theta_*,n} + \delta_n, u, n}{dP_{\theta_*,n}} = \left( u^T \sqrt{\Gamma_n N_n} - \frac{1}{2} u^T \Gamma_n u \right) \to 0$$

in $P_{\theta_*,n}$-probability as $n \to \infty$ for any $u \in \mathbb{R}^d$. Moreover, $N$ follows the $d$-dimensional standard normal distribution, $N$ is independent of $\Gamma$ and $\mathcal{L}(N_n, \Gamma_n|P_{\theta_*,n}) \to \mathcal{L}(N, \Gamma)$ as $n \to \infty$.

If further the limit matrix $\Gamma$ is non-random, we say $\{P_{\theta,n}\}_{\theta,n}$ has the local asymptotic normality (LAN) property.

To prove the LAMN property of our model, we assume the following additional condition.

Assumption (A1''). (A1) is satisfied, $\mu_t = \mu(t, \sigma_x)$, $\sup_{t, x} \mu(t, \sigma_x) < \infty$, $b(t, x, \sigma)$ does not depend on $x$ and $e_i^{n,k}$ follows a normal distribution for any $n, k, i$.

Theorem 2.2. Assume (A1''), (A2) and (A3). Then the family of distributions $\{P_{\sigma, v, n}\}_{\sigma, v, n}$ has the LAN property with $\Gamma$ in (2.11) and $\delta_n = \text{diag}(b_n^{-1/4} E_4, b_n^{-1/2} E_2)$.

Remark 2.6. Jeganathan [20] studied lower bounds of estimation errors for any estimator of parameters. They showed a version of Hájek’s convolution theorem (Corollary 1) and that the optimal asymptotic variance of errors for regular estimators is $\Gamma^{-1}$, where $\Gamma$ is in Definition 2.1. Therefore, Theorems 2.1 and 2.2 ensures that our estimator $\hat{\sigma}_n$ of the parameter $\sigma$ is asymptotically efficient in this sense under the assumptions of both theorems.

Remark 2.7. The assumptions of Theorem 2.2 are rather strong conditions. We are also interested in the LAMN property in more general settings. In particular, we are interested in the case that $\mu_t = \mu(t, X_t, \sigma_x)$ and $\mu$ and $b$ are general functions with suitable conditions. However, we need further analysis using Malliavin calculus to deal with the LAMN property of general diffusion processes, as seen in Gobet [14] and Ogihara [24]. Moreover, when we deal with nonsynchronousness or noise, the true likelihood function is obtained by integrating a likelihood function of ideal observations with respect to unobservable variables. Such integral is not easy to handle, and consequently, it makes the proof of LAMN much more complicated as seen in [24]. To the best of author’s knowledge, the LAMN property for a general diffusion model has not been shown even for models with noisy, synchronous observations. We have left this for future works. On the other hand, asymptotic results of the quasi-likelihood function shown in this paper (e.g. Propositions 2.1 and 7.2) are expected to be useful when we try to show LAMN for general cases. Indeed, LAMN for a model with nonsynchronous observations is proved in [24] by showing asymptotic equivalence of the true likelihood ratio and a quasi-likelihood ratio, and using asymptotic results of the quasi-likelihood ratio in Ogihara and Yoshida [25].

2.4 A Bayes-type estimator and convergence of moments of estimation errors

Polynomial-type large deviation theory by Yoshida [29, 30] enables us to address the asymptotic properties of a Bayes-type estimator and the convergence of moments of estimation errors, which is a stronger result than asymptotic mixed normality. Convergence of moments is useful when we investigate the theory of information criteria, minimax inequality and asymptotic expansion of estimators. See Uchida [27] for a theory of contrast-based information criteria for ergodic diffusion processes with equidistant observations. We also see asymptotic efficiency of our estimator in the sense of minimax inequality.

We first assume following stronger conditions than (A1)–(A3) and (V).

Assumption (B1). 1. (A1) holds true with $O = \mathbb{R}^d$.
2. There exists a positive constant $C$ such that $\sup_{t \in [0, T], \sigma \in \Lambda} |\partial^2 \partial_t^2 \partial_{\sigma}^2 h(t, x, \sigma)| \leq C(1 + |x|)^C$ for $0 \leq 2i + j \leq 4, 0 \leq k \leq 4$ and $x \in \mathbb{R}^d$.
3. $\inf_{t, x, \sigma} \det bb^T(t, x, \sigma) > 0$. 

8
4. \( E[|Y_n|^q] < \infty \) for any \( q > 0 \).
5. \( \sup_p E[|\mu_2|^q] < \infty \), \( \sup_{q < r} E[(\mu_2 - \mu_2)^{1/2} q(t - s)^{-1/2}] < \infty \) and \( \sup_{q < r} E[(E[|\mu_2 - \mu_2|/2]/(t - s))^{q}] < \infty \) for any \( q > 0 \).
6. For any \( q > 0 \), \( \max_j, \sup_p E[|b_j^{(j)}|^q] \leq (1/2) \sup_p E[(b_j^{(j)} - b_j^{(j)} - b_j^{(j)}|1/2 q(t - s)^{-1/2}) < \infty \) for any \( q > 0 \).

Assumption (B2). There exist \( \eta \in (0, 1/2) \), \( \delta > 0 \) and positive-valued stochastic processes \( \{a_{j}^{(j)}\}_{j=1}^{K_0} \) such that \( b_n^{-1/2} b_n^{-1} a_{j}^{(j)} \to 0 \) as \( n \to \infty \), \( \sup_j E[|a_j^{(j)}| t - s)^{-\eta}] < \infty \), \( \sup_j E[|a_j^{(j)}| t - s)^{-\eta}] < \infty \) and

\[
\sup_{n \{i, i^j \in S_n}} E\left[ \left| b_n^{-1/2} \left( s_i^{j} - s_i^{j} \right)^{-1} \# \{i, \left(s_i^{j}, s_i^{j} \right) \in S_{i, j} \} - a_j^{(j)} \right|^q \right] < \infty
\]

for any \( q > 0 \). Moreover, there exists a positive constant \( \gamma \) such that \( k_n b_n^{1/2 + \gamma} \to 0 \) and \( E[\left(b_n^{1/2} - \gamma \right)^{q}] < \infty \) for any \( q > 0 \) and \( \epsilon > 0 \).

Assumption (B3). For any \( q > 0 \), there exists a positive constant \( c_q \) such that \( P[\inf \sigma \neq \sigma_s (-|Y_0(\theta)|/|\sigma - \sigma_s|^2) \leq r^{-1}] \leq c_q/r^{q} \) for any \( q > 0 \).

Assumption (B4). There exist estimators \( \{\hat{v}_n\}_{n \in N} \) of \( v \) such that \( \hat{v}_n \to 0 \) almost surely, \( \lim sup_n E[|\hat{v}_n - v|^q] < \infty \), \( \sup_n E[|b_n^{1/2} (\hat{v}_n - v)|^q] < \infty \), and \( \sup_n E[|b_n^{1/2}] \max_{m, k} E[|\ell_{m, k}^{(j)}| t - s)^{-\eta}] < \infty \) for any \( q > 0 \) and \( \epsilon > 0 \).

Though Assumption (B3) is rather difficult to check in a practical setting, Uchida and Yoshida [28] investigated sufficient conditions for (B3). The simplest condition is that (B3) is satisfied if there exists \( \epsilon > 0 \) such that \( |b_n^{1/2} (t, x, \sigma_1) - b_n^{1/2} (t, x, \sigma_2)| \geq c_1|\sigma_1 - \sigma_2| \) for any \( t \in [0, T], x \in O \) and \( \sigma_1, \sigma_2 \in \Lambda \). See Remark 4 in [25] for details.

Proposition 2.2 (Polynomial-type large deviation inequalities). Assume (B1)–(B4). Then for any \( L > 0 \), there exists a positive constant \( c_L \) such that \( P[\sup_{u \in U_n} Z_n(u) \geq e^{-r/2}] \leq c_L/r^L \) for any \( n \in N \) and \( r > 0 \).

Since \( Z_n(0) = 1 \), Proposition 2.2 immediately yields

\[
E[|b_n^{1/2} (\hat{v}_n - v)|^p] = \int_0^\infty pt^{p-1} P[t^{p} (\hat{v}_n - v)] \geq b(t) dt < \int_0^\infty pt^{p-1} P[\sup_{u \in U_n} Z _n(u) \geq e^{-r/2}] dt < \infty \tag{2.12}
\]

for any \( p > 0 \). Moreover, we obtain the following convergence of moments of the estimation error.

Theorem 2.3. Assume (B1)–(B4). Then \( E[|Y f(b_n^{1/2} (\hat{v}_n - v)] \to E[|Y f(\Gamma_1^{-1/2} N)] \) as \( n \to \infty \) for any bounded random variable \( Y \) on \( (\Omega, F) \) and any continuous function \( f \) at all polynomial growth.

In particular, we obtain convergence of moments where \( E[|b_n^{1/2} (\hat{v}_n - v)] |^q \to E[|\Gamma_1^{-1/2} N|^q] \) for any \( q > 0 \). This property is used when we study the theory of information criteria and asymptotic expansion of estimators.

We also obtain results for a Bayes type estimator. Let a prior density \( \pi : \Lambda \to (0, \infty) \) be a continuous function satisfying \( 0 < \inf \pi \sigma \leq \sup \pi \sigma < \infty \). Then a Bayes-type estimator \( \tilde{\sigma}_n \) for the quadratic loss function is defined by

\[
\tilde{\sigma}_n = \left( \int_\Lambda \exp(H_n(\sigma, \hat{v}_n)) \pi(\sigma) d\sigma \right)^{-1} \int_\Lambda \sigma \exp(H_n(\sigma, \hat{v}_n)) \pi(\sigma) d\sigma.
\]

Since the Bayes-type estimator \( \tilde{\sigma}_n \) contains integrals with respect to \( \sigma \), we need to deal with tail behaviors of likelihood ratio \( H_n(\sigma, \hat{v}_n) - H_n(\tilde{\sigma}_n, \hat{v}_n) \). Hence Proposition 2.2 is essential to deduce asymptotic properties of a Bayes-type estimator. Since the Bayes-type estimator can be calculated using Markov-Chain Monte Carlo methods, it is often easier to calculate than the maximum-likelihood-type estimator. For the Bayes-type estimator \( \tilde{\sigma}_n \), we obtain similar results to the ones for the maximum-likelihood-type estimator.
be calculated by observed data, but we can use it for comparison.

To estimate \( \sigma^2 \), we can use the sample standard deviations given in parentheses. We set the values of parameters as follows:

\[
\begin{align*}
\sigma_0 &= 0.001099 \\
\sigma_1 &= 0.001006 \\
\sigma_2 &= 0.005096 \\
\sigma_3 &= 0.005053 \\
v_1 &= 0.001099 \\
v_2 &= 0.001006 \\
\end{align*}
\]

### 3 Simulation results

First, we consider the case where the latent process \( Y \) is a Brownian motion, that is, \( Y \) satisfies the following stochastic differential equation:

\[
\begin{align*}
\frac{dY_t^1}{b} &= \sigma_1, dW_t^1 \\
\frac{dY_t^2}{b} &= \sigma_2, dW_t^2,
\end{align*}
\]

where \( \sigma_* = (\sigma_{1,*}, \sigma_{2,*}, \sigma_{3,*}) \) and \( \epsilon^{(n)}_t \) be two independent Poisson processes with parameters \( \lambda_1 \) and \( \lambda_2 \), respectively. We give sampling times by \( S_t^{i,j} = \inf\{N_t^{i,j} \geq j\} \cap T = 1, 2 \). Let \( \epsilon^{(n)}_{i,j} \) be independent normal random variables with \( E[\epsilon^{(n)}_{i,j}] = 0 \) and \( E[(\epsilon^{(n)}_{i,j})^2] = \nu_{*,i} \).

Then we can see that this example satisfies (A1'), (A2) and (A3'). So the maximum-likelihood-type estimator \( \hat{\sigma}_n \) is asymptotically unbiased and asymptotically efficient with the asymptotic variance \( \Gamma^{-1} \). For the estimator \( \hat{v}_n \) of \( v_* \), we first use a simple estimator \( \hat{v}_n = (2J_{k,n})^{-1} \sum \{Y_t^k - Y_{t-1}^k\}^2 \), which means that our estimator is calculated by \( \hat{v}_n = \arg\max_\sigma H_n(\sigma, \hat{v}_n) \). We also consider a plug-in estimator \( \hat{v}_n = \left( \hat{v}_n + \left[\frac{(\hat{\sigma}_n)^2}{B(2J_{k,n})}\right] \right) \) of \( v_* \), and \( \hat{\sigma}_n = \arg\max_\sigma H_n(\sigma, \hat{v}_n) \). Then \( \hat{\sigma}_n \) cannot be calculated by observed data, but we can use it for comparison. Though these estimators have the same asymptotic variance, their performances for finite samples are different. In particular, we cannot ignore the bias of \( \hat{v}_n \) since \( v_* \) is relatively small compared with \( \sigma \) in practical data.

Table 1 shows results of 1000 independent estimations. Each cell represents the average of estimators, with sample standard deviations given in parentheses. We set the values of parameters as \( k_n = [n^{5/8}] \), and \( t = 1 \).
Table 2: Comparison of estimators of $\langle Y^1, Y^2 \rangle_T$

<table>
<thead>
<tr>
<th>$n$</th>
<th>MLE</th>
<th>PHY</th>
<th>MRC$_1$</th>
<th>MRC$_2$</th>
<th>QMLE</th>
<th>LMM</th>
<th>minimum</th>
</tr>
</thead>
<tbody>
<tr>
<td>1000</td>
<td>0.374</td>
<td>0.499</td>
<td>0.508</td>
<td>0.501</td>
<td>0.463</td>
<td>0.463</td>
<td>(0.066)</td>
</tr>
<tr>
<td></td>
<td>(0.073)</td>
<td>(0.121)</td>
<td>(0.182)</td>
<td>(0.110)</td>
<td>(0.095)</td>
<td>(0.082)</td>
<td></td>
</tr>
</tbody>
</table>

Results with $v_*= (0.001,0.001)$

<table>
<thead>
<tr>
<th>$n$</th>
<th>MLE</th>
<th>PHY</th>
<th>MRC$_1$</th>
<th>MRC$_2$</th>
<th>QMLE</th>
<th>LMM</th>
<th>minimum</th>
</tr>
</thead>
<tbody>
<tr>
<td>5000</td>
<td>0.496</td>
<td>0.497</td>
<td>0.504</td>
<td>0.499</td>
<td>0.498</td>
<td>0.497</td>
<td>(0.066)</td>
</tr>
<tr>
<td></td>
<td>(0.046)</td>
<td>(0.081)</td>
<td>(0.124)</td>
<td>(0.073)</td>
<td>(0.056)</td>
<td>(0.069)</td>
<td>(0.044)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$n$</th>
<th>MLE</th>
<th>PHY</th>
<th>MRC$_1$</th>
<th>MRC$_2$</th>
<th>QMLE</th>
<th>LMM</th>
<th>minimum</th>
</tr>
</thead>
<tbody>
<tr>
<td>1000</td>
<td>0.496</td>
<td>0.497</td>
<td>0.506</td>
<td>0.499</td>
<td>0.499</td>
<td>0.514</td>
<td>(0.099)</td>
</tr>
<tr>
<td></td>
<td>(0.046)</td>
<td>(0.098)</td>
<td>(0.126)</td>
<td>(0.083)</td>
<td>(0.079)</td>
<td>(0.083)</td>
<td>(0.066)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$n$</th>
<th>MLE</th>
<th>PHY</th>
<th>MRC$_1$</th>
<th>MRC$_2$</th>
<th>QMLE</th>
<th>LMM</th>
<th>minimum</th>
</tr>
</thead>
<tbody>
<tr>
<td>5000</td>
<td>0.499</td>
<td>0.497</td>
<td>0.505</td>
<td>0.499</td>
<td>0.499</td>
<td>0.514</td>
<td>(0.099)</td>
</tr>
<tr>
<td></td>
<td>(0.069)</td>
<td>(0.098)</td>
<td>(0.126)</td>
<td>(0.083)</td>
<td>(0.079)</td>
<td>(0.083)</td>
<td>(0.066)</td>
</tr>
</tbody>
</table>

$(\lambda_1, \lambda_2) = (1, 1), (\sigma_{1,*}, \sigma_{2,*}, \sigma_{3,*}) = (1, \sqrt{1 - 0.5^2}, 0.5),$ and consider two cases of the noise variances: $v_* = (0.001, 0.001)$ for the left-hand side of the table and $v_* = (0.005, 0.005)$ for the right-hand side. In both cases, we can see that $\hat{\nu}_n$ has an upper bias for $n = 1000,$ and causes a lower bias of $\hat{\nu}_n$ because $\hat{\nu}_n$ contains variance of the latent process, which is always positive. These biases can be moderated by using the plug-in estimator. For $n = 5000,$ the plug-in estimator $\hat{\sigma}_n^2$ performs as well as $\hat{\sigma}_n^2.$ In the case of $v_* = (0.005,0.005),$ the biases of $\hat{\nu}_n$ and $\hat{\sigma}_n^2$ are relatively small, so the performance of $\hat{\sigma}_n$ and $\hat{\sigma}_n^2$ are better.

We can also construct an estimator $\hat{\nu}_n^2 \hat{\sigma}_n^2, T$ of the quadratic covariance $\langle Y^1, Y^2 \rangle_T = \sigma_{1,*} \sigma_{3,*} T.$ We see that

$$n^{1/4} (\hat{\nu}_n^2 \hat{\sigma}_n^2, T - \langle Y^1, Y^2 \rangle_T) \rightarrow^d N(0, T^2 (\sigma_{2,*} (\Gamma_1^{-1})_{11} + 2 \sigma_{1,*} \sigma_{3,*} (\Gamma_1^{-1})_{13} + \sigma_{1,*}^2 (\Gamma_1^{-1})_{33})^3)$$

(3.1) as $n \rightarrow \infty$ by the delta method, and the estimator is asymptotically efficient since we can reparameterize the model using $\sigma_{1,*} \sigma_{3,*}.$ We therefore compared the performance of the estimator (MLE) with existing estimators of the quadratic covariance. We used the pre-averaged Hayashi–Yoshida estimator (PHY) and modulated realised covariance (MRC) by Christensen, Kinnebrock, and Podolskij [7], the local method of moments (LMM) by Bibinger et al. [5], and an estimator based on maximum likelihood estimator of a model of constant diffusion coefficients (QMLE) by Aït-Sahalia, Fan, and Xiu [2] for comparison. Except LMM these estimators can be calculated using the ‘ccee’ function in the ‘yuima’ R package (http://r-forge.r-project.org/projects/yuima).

We used the default values of the ‘ccee’ function or values used in corresponding papers for parameters of estimators ($\theta = 0.15$ for PHY, $\theta = 1$ for MRC$_1$, $J = 30$, $h^{-1} = 10$ for LMM). Here we use the oracle estimator defined in [5] for LMM to avoid a complicated calculation. For the modulated realised covariance, we also examine an estimator MRC$_2$ with $\theta = 1/3$ which is used in Jacod et al. [19]. Table 2 shows the results of 1000 estimations. We used the same parameter values as above. Then the true value of the quadratic covariance becomes $\langle Y^1, Y^2 \rangle_T = 0.5.$ For both cases of observation noise variance, we can see that sample standard deviations of our estimator are the best in large samples. The theoretical (asymptotic) minimum $T n^{-1/4} (\sigma_{2,*} (\Gamma_1^{-1})_{11} + 2 \sigma_{1,*} \sigma_{3,*} (\Gamma_1^{-1})_{13} + \sigma_{1,*}^2 (\Gamma_1^{-1})_{33})^3/2$ of standard deviations for all estimators is calculated in the last column of Table 2. We can see that the sample standard deviations of MLE are close to the minima in large samples.

The asymptotic variance in (3.1) is always the best in this parametric setting. On the other hand, Bibinger et al. [5] gave a Cramér–Rao lower bound for one-dimensional perturbation and the asymptotic variance of LMM becomes optimal in that sense. The derivation of their bound is quite different from that of (3.1), and therefore, it is difficult to judge when these variances are the same even in this simple Brownian model. However, there is an interesting numerical result on this point. Let $V_{\text{MLE}}$ be the asymptotic variance of MLE for $\langle Y^1, Y^2 \rangle_T$ and $V_{\text{LMM}}$ be the asymptotic variance of LMM. Let $\nu_*$ be the same as above. For these parameters as $T = 1, (\lambda_1, \lambda_2) = (1, 1), (\sigma_{1,*}, \sigma_{2,*}, \sigma_{3,*}) = (1, \sqrt{1 - 0.5^2}, 0.5)$ we then obtain $V_{\text{MLE}} = T^2 (\sigma_{2,*} (\Gamma_1^{-1})_{11} + 2 \sigma_{1,*} \sigma_{3,*} (\Gamma_1^{-1})_{13} + \sigma_{1,*}^2 (\Gamma_1^{-1})_{33}) = 0.1385502 \cdots$ and $V_{\text{LMM}} = 0.1385502 \cdots,$ that is, we see an exact numerical match between these variances. We have this agreement even if we set other values of parameters, and hence we expect that $V_{\text{MLE}} = V_{\text{LMM}}$ always holds for this model.

However, the performances of MLE and LMM are different in above simulation because it is not easy to set parameters of LMM suitably. If we set $J > 30$ for LMM, then the sample standard deviation is reduced,
while the bias increase; for example, the average value of LMM is 0.51678 and the sample standard deviation is 0.05017 with $J = 100$, $n = 5000$ and $v_n = (0.001, 0.001)$ in Table 2. On the other hand, the asymptotic variances of LMM and MLE are different each other in a model with a time-dependent diffusion coefficient. Let the diffusion coefficient $b$ be defined by

$$b(t, \sigma) = f(t) \begin{pmatrix} \sigma_1 \\ \sigma_2 \end{pmatrix}$$

for some smooth, positive function $f$ on $[0, 1]$ with $T = 1$, and let $\{S^{n,j}_{n,i}\}$ and $\{\epsilon^{n,j}_{n,i}\}$ be the same as the above example. Then we can calculate asymptotic variances $\bar{V}_{\text{MLE}}$ and $\bar{V}_{\text{LMM}}$ of MLE and LMM, respectively, as

$$\bar{V}_{\text{MLE}} = V_{\text{MLE}} \times \left( \int_0^1 f(t)^2 dt \right)^2 / \int_0^1 f(t) dt, \quad \bar{V}_{\text{LMM}} = V_{\text{LMM}} \times \int_0^1 f(t)^3 dt.$$

The Cauchy-Schwarz inequality yields $\int_0^1 f(t)^3 dt \geq \left( \int_0^1 f(t)^2 dt \right)^2 / \int_0^1 f(t) dt$, and the equality is attained if and only if $f$ is a constant function. These results imply $\bar{V}_{\text{MLE}} < \bar{V}_{\text{LMM}}$ for any non-constant function $f$ if $V_{\text{MLE}} \leq V_{\text{LMM}}$. For example, let us set $f(t) = 2 - \sin(\pi t)$ to capture the U-shape of intra-day activities of a stock market. Then we obtain

$$\bar{V}_{\text{MLE}} / \bar{V}_{\text{LMM}} = \frac{(\int_0^1 f(t)^2 dt)^2}{\int_0^1 f(t) dt \int_0^1 f(t)^3 dt} = 0.9533 \ldots$$

if $V_{\text{MLE}} = V_{\text{LMM}}$. Thus we can see the improvement of efficiency in a model with a time-dependent $b$.

In the next, we consider the model with random diffusion coefficients and non-Gaussian noise. As mentioned in Remark 2.1, we cannot directly apply our results to stochastic volatility models. Here we consider the Cox-Ingersoll-Ross (CIR) process derived in [9] as a latent process with random diffusion coefficients. Let the latent process $Y$ satisfy

$$dY_t = \left( \frac{\alpha_1 - \beta_1 Y^1_t}{\alpha_2 - \beta_2 Y^2_t} \right) dt + \sigma^{1,\ast}_{1,1} \sqrt{Y^1_t} dW^1_t + \sigma^{2,\ast}_{1,2} \sqrt{Y^2_t} dW^2_t,$$

where $\sigma^{1,\ast}_j = (\sigma^{1,\ast}_{1,j}, \sigma^{1,\ast}_{2,j}) \in \mathbb{R}^2$ and $\sigma^{2,\ast}_j$ are such that $Y^1_t > 0$ and $Y^2_t > 0$ for $t \in [0, T]$ almost surely. Let $\{\epsilon^{i,j}_{n}\}_{i \in \mathbb{Z}}$ be i.i.d. random variables following a centered Gamma distribution with a shape parameter $k_j$ and a scale parameter $\theta_j$ for $j = 1, 2$. We define $\{N_i\}$, $\hat{\nu}_n$, $\hat{\epsilon}_n$, $\hat{\sigma}_n$, and $\hat{\sigma}'_n$ similarly to the first example. We set the values of parameters as $k_n = [n^{5/8}], T = 1$, $(\lambda_1, \lambda_2) = (1, 1), (\sigma^{1,\ast}_{1,j}, \sigma^{1,\ast}_{2,j}, \sigma^{2,\ast}_j) = (1, \sqrt{1 - \frac{0.5^2}{2}}, 0.5, (\alpha_1, \alpha_2, \beta_1, \beta_2) = (1, 1, 1, 1)$, and $(k_1, k_2, \theta_1, \theta_2) = (2, 2, \sqrt{0.0005}, \sqrt{0.0005})$ which implies $v_n = (0.001, 0.001)$. Table 3 shows averages and sample standard deviations of $T_n - \langle Y^1, Y^2, T \rangle$ for each estimator $T_n$ of the quadratic covariation $\langle Y^1, Y^2, T \rangle$ in 1000 simulations. $\langle Y^1, Y^2, T \rangle$ is random in this model since the diffusion coefficients are random. So we use extra-high-frequency observations $\{Y^{(l)}_{t/k, 100000, k} \}_{k=0}^{100000}$ of $Y$ to calculate the approximated true value of $\langle Y^1, Y^2, T \rangle$. In this model, we have not obtained the LAMN property nor asymptotic efficiency of our estimator though we expect to obtain them. However, we still see that our estimator achieves the best error variance in large samples.
4 Asymptotically equivalent representation of the quasi-likelihood function

We will prove our main results in the rest of this paper. In this section, we introduce an asymptotically equivalent representation $\hat{H}_n(\sigma, v)$ of the quasi-likelihood function $H_n(\sigma, v)$, and prove the equivalence. $\hat{H}_n$ is a useful function for deducing the limit of $H_n$.

4.1 Some notations

We denote $E_m$ as the $\mathcal{G}_{m-1}$-conditional expectation and $\hat{E}_n[X] = X - E_m[X]$ for a random variable $X$. We use the symbol $C$ for a generic positive constant that can vary from line to line.

For a sequence $c_n$ of positive-valued $\mathcal{B}([0,1])$-measurable random variables, let us denote by $\{\hat{R}_n(c_n)\}_{n \in \mathbb{N}}$, $\{\hat{R}_0(c_n)\}_{n \in \mathbb{N}}$ and $\{\hat{R}_n(c_n)\}_{n \in \mathbb{N}}$ sequences of random variables (which may depend on $1 \leq m \leq \ell_n$ and $\sigma$) satisfying

$$E((c_n^{-1} r_n/b_n)^{\delta} (b_n/\Sigma_n)^{-\delta}(\hat{k}_n/k_n)^{-\delta} (k_n/\Sigma_n)^{-\delta} b_n^{-\delta} \exp E[\|\hat{R}_n(c_n)\|^{1/\delta}] \to 0,$$

$$E(c_n^{-1} r_n/b_n)^{q_2} (b_n/\Sigma_n)^{q_2} (k_n/\Sigma_n)^{q_2} b_n^{q_2} \exp E[\|\hat{R}_n(c_n)\|^{1/\delta}] \to 0,$$

and

$$c_n^{-1} r_n/b_n)^{n_1} (b_n/\Sigma_n)^{q_2} (k_n/\Sigma_n)^{q_2} b_n^{q_2} \exp E[\|\hat{R}_n(c_n)\|^{1/\delta}] \to 0,$$

respectively, as $n \to \infty$ for any $\delta, q, q', q_2 > 0$ with some constants $\delta', p_1, \cdots, p_4 \geq 0$.

Let $M_m(v) = \text{diag}(v_1 M_{1,m}, v_2 M_{2,m})$ for $v = (v_1, v_2), b_{\delta} = b^k(s_{m-1}, X_{s_{m-1}}, \sigma), \hat{b}_{\delta} = b^k(s_{m-1}, X_{s_{m-1}}, \sigma)$, $\hat{Z}_m = ((\hat{b}_{\delta}^1 (W_{s_{m-1}}^1 - W_{s_{m-1}}^2) + \epsilon_{\delta} n_j - \epsilon_{\delta} n_j, 1_{K_{m-1}^1 + 2})^\top, ((\hat{b}_{\delta}^2 (W_{s_{m-2}}^1 - W_{s_{m-2}}^2) + \epsilon_{\delta} n_j - \epsilon_{\delta} n_j, 1_{K_{m-1}^1 + 2})^\top, \hat{S}_m(\sigma, v) = \left( \begin{array}{cc} \text{diag}((\hat{b}_{\delta}^1 | I_{1,m}^1)_i & \hat{b}_{\delta}^1 \hat{b}_{\delta}^2 | I_{1,m}^1 \cap I_{2,m}^1 \}_i \\ \hat{b}_{\delta}^1 \hat{b}_{\delta}^2 | I_{1,m}^1 \cap I_{2,m}^1 \}_i \\ \text{diag}((\hat{b}_{\delta}^2 | I_{2,m}^1)_i & \hat{b}_{\delta}^2 \hat{b}_{\delta}^1 | I_{1,m}^1 \cap I_{2,m}^1 \}_j \end{array} \right) + M_m(v), \right. (4.1)

and

$$\hat{H}_n(\sigma, v) = -\frac{1}{2} \sum_{m=2}^{\ell_n} \hat{Z}_m^\top \hat{S}_m^{-1}(\sigma, v) \hat{Z}_m - \frac{1}{2} \sum_{m=2}^{\ell_n} \log \det \hat{S}_m(\sigma, v).$$

The diffusion coefficient $b$ in $\hat{Z}_m$ and $\hat{S}_m$ are either $b(s_{m-1}, X_{s_{m-1}}, \sigma)$ or $b(s_{m-1}, X_{s_{m-1}}, \sigma)$. Hence we do not need to consider the time-dependent structure of $b$ when we study asymptotics of the summmands in $\hat{H}_n$. In particular, we obtain $E_m[\hat{Z}_m^\top \partial_s \hat{S}_m(\sigma, v, s)^{-1} \hat{Z}_m + \partial_s \log \det \hat{S}_m(\sigma, v, s)] = 0$ by $\partial_s \log \det \hat{S}_m(\sigma, v, s) = \text{tr}(\partial_s \hat{S}_m \hat{S}_m^{-1}(\sigma, v, s))$. We will prove the asymptotic equivalence of $H_n$ and $\hat{H}_n$ and then investigate asymptotic properties of $\hat{H}_n$ instead of $H_n$.

Similarly to the approach of Gloter and Jacod [13], we can use the following stronger condition (A1’) instead of (A1) in several part of the proof. Then localization techniques and Girsanov’s theorem enable us to replace (A1’) with (A1).

Assumption (A1’). Assumption (A1) is satisfied, $O = \mathbb{R}^{d_2}, \sup_{t,x,\sigma} \|bb^\top\| \leq (t, x, \sigma) < \infty$, $\mu_t \equiv 0$ and $Y_0$, $\sup_{t, x, \sigma} (|b_t^l| \vee |b_t^l|)$, $\partial_t \partial_t \partial_{\sigma} b_t$, and $\sup_{t,x,\sigma} (|b_t^l| - b_t^s | \vee |b_t^l| - b_t^s |)(t - s)$ are all bounded for $l = 0, 1, 0 \leq 2t + j \leq 4$ and $0 \leq k \leq 4$.

We can also see that (A1’) implies (B1).
4.2 Fundamental properties of the noise covariance matrix

In the following subsection we will show the asymptotic equivalence of \( \hat{H}_n \) and \( H_n \), namely that

\[
b_n^{1/2} \sup_{\sigma \in \Lambda} |\partial_\sigma^j (H_n(\sigma, \hat{v}_n) - H_n(\sigma, v_n)) - \partial_\sigma^j (\hat{H}_n(\sigma, v_n) - \hat{H}_n(\sigma, v_\ast))| \to^p 0
\]

as \( n \to \infty \) for \( 0 \leq j \leq 3 \). To that end, we first show fundamental properties of \( S_m \) and \( \tilde{S}_m \). These matrices inherit some properties of \( M_{j,m} \), that are necessary to deduce the limit of \( H_n \) and \( \hat{H}_n \). The first property (4.2) concerns the trace of a matrix related to \( M_{j,m} \) investigated by [13]. In the one-dimensional model with noisy, equidistance observations, this property can be directly applied to the quasi-likelihood function because the covariance matrix of the latent process is the unit matrix. However, this is insufficient for our purpose because our covariance matrix \( S_m(\sigma, v) - M_m(v) \) of the latent process is rather complicated. Therefore, we investigate further matrix properties related to \( M_{j,m} \).

First, we consider the results in [13]. For any positive constants \( p, q, a \) and \( b \), eigenvalues of \((aE + M_{j,m})^{-1}\) are \( \{(a + 2(1 - \cos(k^l_m + 1)^{-1}))\}_{k^l_m = 1} \) and we obtain

\[
\pi^{-1}(k^l_m + 1)I_p(a) - a^{-p} \leq \text{tr}((aE + M_{j,m})^{-p}) \leq \pi^{-1}(k^l_m + 1)I_p(a),
\]

\[
\pi^{-1}(k^l_m + 1)I_p,q(a, b) - a^{-p}b^{-q} \leq \text{tr}((aE + M_{j,m})^{-p}(bE + M_{j,m})^{-q}) \leq \pi^{-1}(k^l_m + 1)I_p,q(a, b),
\]

where \( I_p(a) = \int_0^\pi (a + 2(1 - \cos x))^{-p}dx \) and \( I_p,q(a, b) = \int_0^\pi (a + 2(1 - \cos x))^{-p}b + (2(1 - \cos x))^{-q}dx \). Simple calculations show that \( I_1(a) = \pi/\sqrt{a(4 + a)} \), \( I_2(a) = \pi(2 + a)a^{-3/2}(4 + a)^{-3/2} \) and \( \int_0^\pi \log(a + 2(1 - \cos x)) - \log(b + 2(1 - \cos x)))dx = 2\pi(\log(\sqrt{a} + \sqrt{4 + a}) - \log(\sqrt{b} + \sqrt{4 + b})) \). See Section 4.1 in [13] for the details. Moreover, differentiation with respect to \( a \) yields

\[
I_p(a) = \left( \frac{-1}{(p - 1)!} \right) \frac{d}{da} \left( \frac{\pi}{\sqrt{a(4 + a)}} \right).
\]

In particular, if \( a = X_n S_n^{-1} \) for some tight random variables \( \{X_n\} \), then we have

\[
I_p(X_n S_n^{-1}) = \frac{\pi(2p - 3)!}{2^p(p - 1)!} \left( X_n S_n^{-1} \right)^{p + 1/2} + O_p(b_n^{p - 3/2}).
\]

For \( \epsilon \geq 0 \), let \( \{p_j(\epsilon)\} \) and \( \{p'_j(\epsilon)\} \) be sequences of positive numbers satisfying

\[
p_1(\epsilon) = 2 + \epsilon, \quad p'_1(\epsilon) = 1 + \epsilon,
\]

\[
p_{j+1}(\epsilon) = 2 + \epsilon - 1/p_j(\epsilon), \quad \text{and} \quad p'_{j+1}(\epsilon) = 2 + \epsilon - 1/p'_j(\epsilon)
\]

for \( j \in \mathbb{N} \). Let \( E_{1,j}(a) \) be a \( k^l_m \times k^l_m \) matrix satisfying \( (E_{1,j}(a))_{k,j} = \delta_{k+1,j} + da_k \delta_{k+1,j} \) for \( a \in \mathbb{R} \). Then we have

\[
E_{k^l_m,k^l_m}(p_{k^l_m - 1}(\epsilon)^{-1}) \cdots E_{2,1}(p_1(\epsilon)^{-1})(\epsilon E + M_{j,m})E_{1,2}(p_1(\epsilon)^{-1}) \cdots E_{k^l_m,k^l_m}(p_{k^l_m - 1}(\epsilon)^{-1}) = \text{diag}(\{p_j(\epsilon)^{k^l_m}\}_{j=1}^{k^l_m}),
\]

\[
(\epsilon E + M_{j,m})^{-1} = E_{1,2}(p_1(\epsilon)^{-1}) \cdots E_{k^l_m,k^l_m}(p_{k^l_m - 1}(\epsilon)^{-1}) \text{diag}(\{p_j(\epsilon)^{k^l_m}\}_{j=1}^{k^l_m}) E_{k^l_m,k^l_m}(p_{k^l_m - 1}(\epsilon)^{-1}) \cdots E_{2,1}(p_1(\epsilon)^{-1}),
\]

and hence

\[
(\epsilon E + M_{j,m})^{-1} = \left\{ \prod_{k_l+1 \leq l \leq l} p_{l-1}(\epsilon)^{-1} 1_{\{l \leq k_l\}} \right\}_{k,l} \text{diag}(\{p_j(\epsilon)^{k^l_m}\}_{j=1}^{k^l_m}) \left\{ \prod_{l+1 \leq l \leq k} p_{l-1}(\epsilon)^{-1} 1_{\{l \leq k\}} \right\}_{k,l}.
\]

Moreover, we have the following lemma.

**Lemma 4.1.** Let \( \epsilon \in [0, 1) \) and \( p_+ = 1 + \epsilon/2 + \sqrt{\epsilon + \epsilon^2/4} \). Then

1. \( 1 \leq p'_j(\epsilon) \leq p_j(\epsilon) < p_j(\epsilon) \leq 1 + 1/j + je \) for \( j \in \mathbb{N} \), \( \{p_j(\epsilon)\} \) is monotone decreasing, and \( \{p'_j(\epsilon)\} \) is monotone nondecreasing.

2. \( \{((\epsilon E + M_{j,m})^{-1})_{kk}\}_{k=1}^{\lfloor k^l_m/2 \rfloor} \) is monotone increasing.

3. \( p_j - p_+ \leq (1 + \sqrt{\epsilon})^{-(j-2)} \) and \( p_+ - p'_j \leq \sqrt{\epsilon}(1 + \sqrt{\epsilon})^{-(j-1)} \) for \( j \geq 2 \).
4. \( \prod_{j=1}^{k} p_j^*(e) = (p_k(e) - 1) \prod_{j=1}^{k-1} p_j^*(e) \) for any \( k \geq 2 \).

Proof. 1. We simply denote \( p_j = p_j^*(e) \) and \( p_j = p_j^*(e) \). We will prove \( p_j < p_j \leq 1 + 1+j+je \) for \( j \in \mathbb{N} \) by induction. The results obviously hold for \( j = 1 \). Assume the results hold for all values in \( \mathbb{N} \) up to \( j \). Then since \( p_j = 2 - \epsilon_1/p_+ \), we obtain \( p_{j+1} = 1/p_+ - 1/p_j > 0 \), and
\[
p_{j+1} \leq 2 + \epsilon - (j/(j+1))(1 + j^2\epsilon/(j+1))^{-1} \leq 2 + \epsilon - (j/(j+1))(1 - j^2\epsilon/(j+1)) \leq 1 + (j+1)\epsilon.
\]
Hence, we have \( p_j < p_j \leq 1 + 1+j+je \) for \( j \in \mathbb{N} \). Moreover, we can inductively deduce \( p_{j+1} - p_j = 1/p_{j-1} - 1/p_j < 0 \). The results for \( \{p_j^*(e)\} \) are obtained similarly.

2. By considering the cofactor matrix and (4.4), we have
\[
((e\mathcal{E} + M_{j.m})^{-1})_{kk} = \frac{\det(e\mathcal{E}_{k_i+k} + M(k_m+k))}{\det(e\mathcal{E} + M_{j,m})} = \prod_{l=1}^{k-1} \frac{p_l}{p_l}_{k_i=k}.
\]

Therefore we obtain the result by monotonicity of \( p_j \).

3. This is easy since \( p_j - p_+ = (p_j - p_+)/p_{j+1} - (p_1 - p_+)/p_{j+2} \leq p_j - p_+ - p_j - p_+ \leq p_j - p_+ - p_+ - p_j - p_+ \leq 1/p_{j+1} - 1/p_j < 0 \).

4. Let \( \bar{p} = \sup_{t,\sigma} \{[b \cdot b][b \cdot b]^{-1}[b \cdot b]^{-1}(t, X_t, \sigma) \}, \bar{D}_m = (\bar{D}_{1,m}, \bar{D}_{2,m}), \bar{D}_j,m = \text{diag}((1\bar{D}_{i,j,m}), v_{j,*} M_{j,m}, D_j,m = D'_{1,m}, D_{2,m}, D'_{j,m} = \text{diag}((1\bar{D}_{i,j,m}), v_{j,*} M_{j,m}.

Lemma 4.2. Assume (B1). Then \( \text{tr}(\bar{S}_m^{-1}(\sigma, v_*)) = \bar{R}_n(b_{k_n}^{1/2}/k_n) \).

Proof. Let \( D''_m = \text{diag}((1\bar{D}_{i,j,m}), v_{j,*} M_{j,m} \text{ and } D''_m = (D''_m)^{-1/2} \bar{D}_m(D''_m)^{-1/2} \), then we have
\[
\bar{S}_m = (D''_m)^{-1/2}(D''_m)^{-1/2}(\bar{S}_m - \bar{D}_m)(D''_m)^{-1/2}(D''_m)^{-1/2}(\bar{S}_m - \bar{D}_m)(D''_m)^{-1/2}(D''_m)^{-1/2}(D''_m)^{-1/2}(D''_m)^{-1/2} \).
\]

Moreover, Lemma A.4, (B1), and Lemma 2 in [25] yield
\[
\left\| (D''_m)^{-1/2}(D''_m)^{-1/2}(\bar{S}_m - \bar{D}_m)(D''_m)^{-1/2}(D''_m)^{-1/2} \right\| \leq \bar{p} \left\{ \left\{ \frac{1}{1, m} \cap I_{2, m}^2 \right\}^{1/2} \left\{ \frac{1}{1, m} \cap I_{2, m}^2 \right\}^{1/2} \right\} \leq \bar{p} < 1.
\]

Therefore, we obtain
\[
\text{tr}(\bar{S}_m^{-1}) \leq \text{tr}((D''_m)^{-1/2}(D''_m)^{-1/2}((\bar{S}_m - \bar{D}_m)(D''_m)^{-1/2}(D''_m)^{-1/2})) \leq \text{tr}(\bar{D}_m)^{-1}(1 - \bar{p}) \leq \sum_{j=1}^{k-1} \text{tr}(\bar{D}_j,m) r_n \bar{S}_n^{-1}(1 - \bar{p})^{-1},
\]
by Lemma A.1, the equation \( \bar{D}_{j,m} = \bar{D}_{j,m}^2(\bar{S}_m - \bar{D}_{j,m}^2)(\bar{D}_{j,m} - \bar{D}_{j,m}^2) \bar{D}_{j,m}^{2} \), and that
\[
\left\| \bar{D}_{j,m}^2(\bar{D}_{j,m} - \bar{D}_{j,m}^2) \bar{D}_{j,m}^{2} \right\| \leq (\bar{D}_{1,m}^2, r_n)^{-1} \left\{ \bar{D}_{1,m}^2(r_n - \bar{S}_n) = 1 - \bar{S}_n/r_n. \right\}
\]
We thus obtain the desired results by (4.2).

\( \square \)
4.3 Asymptotic equivalence of $H_n$ and $\tilde H_n$

In this section, we prove the asymptotic equivalence of $H_n$ and $\tilde H_n$. We provide the following lemma about estimates of moments of the quantities related to $H_n$ and $\tilde H_n$. The proof is given in the appendix; it is obtained based on the properties of $M_{j,m}$ in Section 4.2, standard Itô calculus, and some results from linear algebra.

Let $\tilde S_{m,*} = \tilde S_m(\sigma_*, v_*)$.

**Lemma 4.3.** Assume (B1). Let $S = S_{n,m}$ be a symmetric, $G_{m,-}$-measurable random matrix of size $k_m^1 + k_m^2$ satisfying $\|\tilde S_{m,*}S\tilde S_{m,*}\| \leq b_n^{-1}$ for $n \in \mathbb{N}$ and $1 \leq m \leq \ell_n$. Then there exists a positive random sequence $Q_n$ which does not depend on $S$ such that $Q_n = \tilde R_n(1)$ and

1. $|E_m[(\tilde Z_m^* S \tilde Z_m)^2] - 2 \text{tr}((S \tilde S_{m,*})^2) - \text{tr}(S \tilde S_{m,*})^2| \leq Q_n$, $E_m[(\tilde Z_m^* S \tilde Z_m)^4] \leq (b_n^{-1}k_n^7) \vee (b_n^{-2}k_n^4)Q_n$ and $E_m[(\tilde Z_m^* S \tilde Z_m)^q] \leq b_n^{-q}k_n^2Q_n$ for $q > 4$.

2. $E_m[\|\sum_m(Z_m - \tilde Z_m)^\top S(Z_m + \tilde Z_m)|^2] \leq (b_n^{-3}k_n^7)^{q/4}Q_n$ for $q \geq 4$.

3. $E_m[\|\sum_m(Z_m - \tilde Z_m)^\top S(Z_m + \tilde Z_m)|^2] \leq b_n^{-1}k_n^2Q_n$.

**Proof.** See the appendix.

Now we obtain the asymptotic equivalence of $H_n$ and $\tilde H_n$.

**Lemma 4.4.** Assume (B1), (A2) and (V). Then

$$b_n^{-1/2} \sup_{\sigma \in A} |\partial^2_\sigma (H_n(\sigma, \tilde v_n) - H_n(\sigma_*, \tilde v_n)) - \partial^2_\sigma (\tilde H_n(\sigma, v_*)) - \partial^2_\sigma (\tilde H_n(\sigma_*, v_*))| \rightarrow^p 0,$$

and $b_n^{-1/4}(\partial_\sigma H_n(\sigma, \tilde v_n) - \partial_\sigma \tilde H_n(\sigma_*, v_*)) \rightarrow^p 0$ as $n \rightarrow \infty$ for $0 \leq j \leq 3$. If (B4) holds as well, then

$$E_H\left[\left(b_n^{-1/2} \sup_{\sigma \in A} |\partial^2_\sigma (H_n(\sigma, \tilde v_n) - H_n(\sigma_*, \tilde v_n)) - \partial^2_\sigma (\tilde H_n(\sigma, v_*)) - \partial^2_\sigma (\tilde H_n(\sigma_*, v_*))|\right)^q\right] = \tilde R_n((b_n^{-5}k_n^7)^{q/4})$$

for any $0 \leq j \leq 3$ and $q > 0$.

**Proof.** We first obtain

$$H_n(\sigma, v_*) - H_n(\sigma_*, v_*) - (\tilde H_n(\sigma, v_*) - \tilde H_n(\sigma_*, v_*))$$

$$= -\frac{1}{2} \sum_m \left( (Z_m - \tilde Z_m)^\top (S_{m,-1}(\sigma, v) - S_{m,-1}(\sigma_*, v_*))(Z_m + \tilde Z_m) 
+ \tilde Z_m^\top (S_{m,-1}(\sigma, v) - S_{m,-1}(\sigma_*, v_*)) - \tilde S_m^\top (\sigma, v) + \tilde S_m^\top (\sigma_*, v_*) \tilde Z_m 
+ \left( \log \frac{\det S_m(\sigma, v_*)}{\det S_m(\sigma, v_*)} \right) \right)$$

$$= \Psi_{1,n}(\sigma) + \tilde \Psi_{2,n}(\sigma) + \tilde \Psi_{3,n}(\sigma).$$

We will give estimates for these quantities. Point 2 of Lemma 4.3 yields $\sup_{\sigma} E_H[|b_n^{-1/2} \partial^2_\sigma \tilde \Psi_{1,n}|^q] = \tilde R_n((b_n^{-5}k_n^7)^{q/4})$ for $0 \leq j \leq 4$ and $q > 0$, and consequently by Sobolev’s inequality $E_H[\sup_{\sigma} |b_n^{-1/2} \partial^2_\sigma \tilde \Psi_{1,n}|^q] = \tilde R_n((b_n^{-5}k_n^7)^{q/4})$ as $n \rightarrow \infty$ for $0 \leq j \leq 3$ and $q > d$.

Let

$$C(x, \sigma, v) = \left( \{ b_1^2(s_{m-1} - x, \sigma) | I_{m,1}^1 \} \cdot |I_{m,1}^1 \cap I_{m,2}^2\} \right)_{i,j} + v_1 M_{1,m}$$

$$\{ b_1^1 b_2^2(s_{m-1} - x, \sigma) | I_{m,1}^1 \cap I_{m,2}^2\} \cdot \left( \{ b_1^2(s_{m-1} - x, \sigma) | I_{m,1}^1 \cap I_{m,2}^2\} \right)_{i,j} + v_2 M_{2,m}$$

$$\{ b_1^1 b_2^2(s_{m-1} - x, \sigma) | I_{m,1}^1 \cap I_{m,2}^2\} \cdot \left( \{ b_1^2(s_{m-1} - x, \sigma) | I_{m,1}^1 \cap I_{m,2}^2\} \right)_{i,j} + v_2 M_{2,m}$$

16
for \( v = (v_1, v_2) \), \( \tilde{C}(x, v) = C(x, \sigma, v)^{-1} - C(x, \sigma, v)^{-1} \), \( C^{(1)} = \int_0^1 \partial_x \tilde{C}(t\tilde{x}_{m-1} + (1-t)X_{m-1}, \nu) dt \), \( C^{(2)} = \partial_x \tilde{C}(X_{m-2}, \nu) \) and \( C^{(3)} = C(X_{m-2}, \sigma, \nu) \). Then we obtain

\[
E_{\Omega}[\hat{\Psi}_{2,n}^q] = 2^{-q} E_{\Omega} \left[ \left( \sum_m \tilde{Z}_{m}^\top C^{(1)} \tilde{Z}_{m}(X_{m-1} - X_{m-1}) \right)^q \right] \\
\leq C E_{\Omega} \left[ \left( \sum_m \text{tr}(C^{(1)} E_m[\tilde{Z}_{m}^\top \tilde{Z}_{m}]) (X_{m-1} - X_{m-1})^2 \right)^q \right] + C E_{\Omega} \left[ \left( \sum_m \text{tr}(C^{(1)} E_m[\tilde{Z}_{m}^\top \tilde{Z}_{m}]) (X_{m-1} - X_{m-1})^2 \right)^{q/2} \right] \\
\leq C E_{\Omega} \left[ \left( \sum_m \text{tr}(C^{(2)} C^{(3)}) (X_{m-1} - X_{m-1})^q \right)^q \right] + O_p((\ell_n b_n^{-1/2} k_n \ell_n^{-1})^q b_n^q) \\
+ O_p((\ell_n^2 b_n^{-2} q k_n \ell_n^{-2} b_n^2) + O_p(b_n^{-q/2} k_n^q \ell_n^{-2} b_n^q) + O_p(b_n^{-q} k_n^2 b_n^q) = O_p(b_n^{-q+2} k_n^q)
\]

for any \( q > 0 \) and \( \epsilon > 0 \), by Lemma 4.3, the Burkholder–Davis–Gundy inequality, \( \sup_m E_{\Omega}[|X_{m-1} - X_{m-1}|^{q}] \) is lower bounded by \( O_p((\ell_n b_n^{-1/2} k_n \ell_n^{-1})^q b_n^q) \) and \( \sup_m E_{\Omega}[|X_{m-1} - X_{m-1}|^{q}] \) is upper bounded by \( O_p((\ell_n b_n^{-1/2} k_n \ell_n^{-1})^q b_n^q) \). Similar estimates for \( \partial_{\nu} \hat{\Psi}_{2,n} \) and Sobolev’s inequality yield \( E_{\Omega}[\sup_{\sigma} |b_n^{-1/2} \partial_{\nu} \hat{\Psi}_{2,n}|^q] = O_p(b_n^{-q+2} k_n^q) \), and therefore we obtain \( b_n^{-1/2} \sup_{\sigma} |\partial_{\nu} H_n(\sigma, \nu) - H_n(\sigma, \nu)| \), \( \sup_{\sigma} |\partial_{\nu} H_n(\sigma, \nu) - H_n(\sigma, \nu)| \) for \( 0 \leq j \leq 3 \). Similar estimates for \( \Omega \) satisfy the following estimates:

\[
E_{\Omega}[|b_n^{-1/2} \sup_{\sigma} |\partial_{\nu} H_n(\sigma, \nu) - H_n(\sigma, \nu)|]| = O_p(b_n^{-q+2} k_n^q)
\]

for \( 0 \leq j \leq 3 \).

Taylor’s formula yields

\[
H_n(\sigma, \nu) - H_n(\sigma, \nu) - (H_n(\sigma, \nu) - H_n(\sigma, \nu)) = -\frac{1}{2} \sum_m \left\{ \tilde{Z}_{m}^\top (\tilde{C}(\tilde{x}_{m-1}, \nu) - \tilde{C}(\tilde{x}_{m-1}, \nu)) \tilde{Z}_{m} + \log \frac{\text{det } S_m(\sigma, \nu)}{\text{det } S_m(\sigma, \nu)} - \log \frac{\text{det } S_m(\sigma, \nu)}{\text{det } S_m(\sigma, \nu)} \right\} \\
= -\frac{1}{2} \sum_m \left\{ \tilde{Z}_{m}^\top \partial_{\nu} \tilde{C}(\tilde{x}_{m-1}, \nu) \tilde{Z}_{m} + \partial_{\nu} \log \frac{\text{det } S_m(\sigma, \nu)}{\text{det } S_m(\sigma, \nu)} \right\} (\nu - v)^2 \\
-\frac{1}{2} \sum_m \int_0^1 \frac{(1-s)^3}{6} \left\{ \tilde{Z}_{m}^\top \partial_{\nu} \tilde{C}(\tilde{x}_{m-1}, \nu) \tilde{Z}_{m} + \partial_{\nu} \log \frac{\text{det } S_m(\sigma, \nu)}{\text{det } S_m(\sigma, \nu)} \right\} ds (\nu - v)^4,
\]

where \( v = s\nu + (1-s)\nu \).

Then we obtain \( b_n^{-1/4} |H_n(\sigma, \nu) - H_n(\sigma, \nu)| \to 0 \), by Lemma 4.3, (V), and the equation \( \partial_{\nu} \log \text{det } S_m = \text{tr}(\partial_{\nu} S_m \theta_m^{-1}) \). Similarly, we obtain \( b_n^{-1/4} |\partial_{\nu} H_n(\sigma, \nu) - \partial_{\nu} H_n(\sigma, \nu)| \to 0 \) for \( 0 \leq j \leq 4 \). Sobolev’s inequality yields \( \sup_{\sigma} |b_n^{-1/4} \partial_{\nu} H_n(\sigma, \nu) - H_n(\sigma, \nu)| \to 0 \) for \( 0 \leq j \leq 3 \) and consequently we obtain \( b_n^{-1/4} |\partial_{\nu} H_n(\sigma, \nu) - \partial_{\nu} H_n(\sigma, \nu)| \to 0 \) for \( 0 \leq j \leq 3 \).

Moreover, points 1 and 3 of Lemma 4.3 yield \( E_{\Omega}[|b_n^{-1/4} \partial_{\nu} \hat{\Psi}_{2,n}(\sigma, \nu)|^2] = O_p(b_n^{-q+2} k_n^q) \to 0 \) for sufficiently small \( \epsilon > 0 \), and consequently \( b_n^{-1/4} |\partial_{\nu} H_n(\sigma, \nu) - \partial_{\nu} H_n(\sigma, \nu)| \to 0 \).

If further (B4) is satisfied, then for any \( q > 0 \), (4.7) yields

\[
\sup_{\sigma} E_{\Omega}[b_n^{-q/2} |H_n(\sigma, \nu) - H_n(\sigma, \nu) - (H_n(\sigma, \nu) - H_n(\sigma, \nu))|] \\
= \tilde{R}_n(b_n^{-q/2} b_n^{-q/2} k_n^q b_n^2) + \tilde{R}_n(b_n^{-q/2} b_n^{-q/2} k_n^q b_n^2) = \tilde{R}_n(b_n^{-q} k_n^q),
\]
by Lemma 4.3, (V), and the equation \( \partial_v \log \det S_m = \text{tr}(\partial_v S_m S_m^{-1}) \). Similarly, we obtain
\[
\sup \mathbb{E}_n [b_n^{-1/2} \partial_{\sigma} H_n(\sigma, \hat{v}_n) - \partial_{\sigma} H_n(\sigma, v_u)] = \hat{R}_n(b_n^{-1/q}) \text{ for } 1 \leq j \leq 4.\]
Sobolev’s inequality yields
\[
\mathbb{E}_n [\sup \mathbb{E}_n (b_n^{-1/2} \partial_{\sigma} (H_n(\sigma, \hat{v}_n) - H_n(\sigma, v_u)) - \partial_{\sigma} H_n(\sigma, v_u))] = \hat{R}_n(b_n^{-1/q}) \text{ for } 0 \leq j \leq 3,\]
which completes the proof.

5 The limit of the quasi-likelihood function

We complete the proof of Proposition 2.1 in this section. To do so, it is essential to specify the asymptotic behavior of some functions of approximate covariance matrix \( \hat{S}_m \), as seen in (5.2). Unlike previous studies by Gloter and Jacod [12, 13], the eigenvalues of the diagonal blocks \( \hat{D}_{1,m} \) and \( \hat{D}_{2,m} \) of \( \hat{S}_m \) are not identified because of the irregular sampling, and even the sizes of \( \hat{D}_{1,m} \) and \( \hat{D}_{2,m} \) are different. These problems make it difficult to deduce asymptotic behaviors of the right-hand side of (5.2). To solve these problems, in Lemma 5.1, we approximate \( \hat{D}_{1,m} \) by \( \tilde{D}_{j,m} \), which is a kind of local averaged versions of \( \hat{D}_{j,m} \) and has similar properties to the covariance matrix of equidistant sampling scheme. Moreover, we can also change the sizes of \( \tilde{D}_{j,m} \) using some specific properties of \( \tilde{D}_{j,m} \). We deal with this in Lemma 5.2, and show convergence of some trace functions that appear in a decomposition of \( H_n \). The decomposition (4.4) and the nice properties of \( p_i \) in Lemma 4.1 are essential in the proofs.

Lemma 4.4 yields
\[
b_n^{1/2} \partial_{\sigma} (H_n(\sigma, \hat{v}_n) - H_n(\sigma_* \hat{v}_n)) = b_n^{1/2} \partial_{\sigma} (\hat{H}_n(\sigma, v_u) - \hat{H}_n(\sigma_*, v_u)) + o_p(1)
\]
\[
= -\frac{1}{2} b_n^{1/2} \sum_m \left( E_m [\hat{Z}_m^T \partial_{\sigma} (\hat{S}_m^1 - \hat{S}_m^1 \hat{S}_m^*) \hat{Z}_m] + \partial_{\sigma} \log \det \hat{S}_m \right) \det \hat{S}_m^* 
\]
\[
= -\frac{1}{2} b_n^{1/2} \sum_m E_m [\hat{Z}_m^T \partial_{\sigma} (\hat{S}_m^1 - \hat{S}_m^1 \hat{S}_m^*) \hat{Z}_m] + o_p(1)
\]
(5.1)

for \( 0 \leq j \leq 3 \). Together with the relation \( E_m [\hat{Z}_m^T \partial_{\sigma} (\hat{S}_m^1 - \hat{S}_m^1 \hat{S}_m^*) \hat{Z}_m] = \text{tr}(\partial_{\sigma} (\hat{S}_m^1 - \hat{S}_m^1 \hat{S}_m^*) \hat{S}_m^*), \) we obtain
\[
b_n^{1/2} \partial_{\sigma} (H_n(\sigma, \hat{v}_n) - H_n(\sigma_*, \hat{v}_n)) = -\frac{1}{2} b_n^{1/2} \sum_m \partial_{\sigma} \left( \text{tr}(\hat{S}_m^1 \hat{S}_m^*, - \mathcal{E}) + \log \det \hat{S}_m \right) + o_p(1),
\]
(5.2)

since the residual terms are \( o_p(1) \) by Lemma 4.3.

We first investigate asymptotics of \( \text{tr}(\hat{S}_m^1 \hat{S}_m^*, - \mathcal{E}) \). Let \( \tilde{L} = \begin{pmatrix} \hat{b}_1 \hat{b}_m \hat{I}_{1,m} \cap I_{j,m} \end{pmatrix}_{i,j} \) and \( \tilde{G} = \begin{pmatrix} \hat{I}_{1,m} \cap I_{j,m} \end{pmatrix}_{i,j} \).

Then since
\[
\hat{S}_m^{-1} = \hat{D}_m^{-1/2} \sum_{p=0}^{\infty} (-1)^p \begin{pmatrix} 0 & \hat{D}_m^{-1/2} \hat{L} \hat{D}_m^{-1/2} \end{pmatrix}^p \hat{D}_m^{-1/2} = \sum_{p=0}^{\infty} \begin{pmatrix} \hat{D}_m^{-1/2} \hat{L} \hat{D}_m^{-1/2} & \hat{D}_m^{-1/2} \hat{L} \hat{D}_m^{-1/2} \end{pmatrix}^p \hat{D}_m^{-1/2},
\]
we have
\[
\text{tr}(\hat{S}_m^{-1} \hat{S}_m^*, - \mathcal{E}) = \sum_{p=0}^{\infty} \{ |\hat{b}_m|^2 - |\hat{b}_m|^2 \text{tr}((\hat{D}_m^{-1/2} \hat{L} \hat{D}_m^{-1/2})^p \hat{D}_m^{-1/2}) + (|\hat{b}_m|^2 - |\hat{b}_m|^2) \text{tr}((\hat{D}_m^{-1/2} \hat{L} \hat{D}_m^{-1/2})^p \hat{D}_m^{-1/2}) + 2 |\hat{b}_m|^2 \text{tr}((\hat{D}_m^{-1/2} \hat{L} \hat{D}_m^{-1/2})^p \hat{D}_m^{-1/2}) \}.
\]
(5.3)

Note that \( \| \{ I_{1,m} \cap I_{j,m} \} \|_{I_{1,m}^{-1/2} I_{j,m}^{-1/2}} \leq 1 \) by Lemma 2 in [25].

We will see the limit of each term on the right-hand side of (5.3). Let \( \tilde{d}_m^2 = \hat{a}_m^2 \) and \( \tilde{D}_j,m = |\hat{b}_m|^2 \text{tr}((\hat{D}_m^{-1/2} \hat{L} \hat{D}_m^{-1/2})^p \hat{D}_m^{-1/2}) + v_j M_{j,m} \). It is difficult to calculate eigenvalues of \( \tilde{D}_j,m \). However, we can easily calculate eigenvalues of \( \tilde{D}_j,m \) as seen in Section 4.2, and it has nice properties which are useful when we deduce a limit. We can replace \( \tilde{D}_j,m \) by \( \hat{D}_j,m \) using the following lemma.
Lemma 5.1. Let \( j \in \{1, 2\} \) and \( A_{n,m} \) be a \( k^m_n \times k^m_n \) matrix for \( 1 \leq m \leq \ell_n \). Assume (B1), (A2) and that all elements of \( A_{n,m} \) are nonnegative and \( \|A_{n,m}\| \leq 1 \) for any \( m \). Then
\[
\text{tr}(\partial^k \tilde{D}_{j,m}^{-1} A_{n,m}) = \text{tr}(\partial^k \tilde{D}_{j,m}^{-1} A_{n,m}) + \hat{R}_n(b_1^{3/2} t_n^{-1})
\]
for \( 0 \leq k \leq 3 \). If further (B2) is satisfied, then
\[
\sup_{\sigma} |\text{tr}(\partial^k \tilde{D}_{j,m}^{-1} A_{n,m}) - \text{tr}(\partial^k \tilde{D}_{j,m}^{-1} A_{n,m})| = R_n(b_1^{3/2} t_n^{-1})
\]
for \( 0 \leq k \leq 3 \).

Proof. We first consider the case where \( k = 0 \). (4.4), (4.6), and the equation above it yield
\[
\text{tr}(\tilde{D}_{j,m}^{-1} A_{n,m}) = \sum_{p=0}^{\infty} \frac{t_n^{p+1}}{p!} \sum_{\ell_1, \ldots, \ell_{p+1}} \sum_{i_1, \ldots, i_{p+1}} (A_{n,m})_{i_1, i_2} \cdots (A_{n,m})_{i_p, i_{p+1}} (\tilde{D}_{j,m}^{-1})_{i_{p+1}, i_{p+1}} \cdots (\tilde{D}_{j,m}^{-1})_{i_1, i_1},
\]
(5.4)
where \( P_{k_1, k_2, l_1, l_2} = \prod_{m=1}^{\infty} p_i \prod_{m=2}^{\infty} p_j \prod_{m=1}^{\infty} p_k \prod_{m=2}^{\infty} p_\ell \). Then the nice properties of \( p_i \) in Lemma 4.1 will lead us to the desired results. Roughly speaking, we have \( 1 \leq p_i([\tilde{b}_m^{i}]^2 r_n t_n^{-1}) \sim 1 + C b_n^{1/2} \) for sufficiently large \( i \). This means that \( P_{k_1, k_2, l_1, l_2} \) and \( P_{k_1, k_2, l_1, l_2} \) are asymptotically equivalent if \( |k_1 - k_2| \) and \( |k_2 - k_2| \) are of order less than \( b_1^{1/2} \). Then we can replace \( \tilde{D}_{j,m} \) in the right-hand side of (5.4) by \( \tilde{D}_{j,m} \) since the diagonal elements of both matrices have the same local average. We will verify these rough sketches by the following.

We first see that terms containing small \( l_q \) can be ignored. Let \( \eta \) be the one in (A2), \( \eta' \in (\eta, 1/2) \), \( t_{i,m} = s_{i-1} + T t_n b_n^{i-1} k_n + T t_n (k_n - b_n^i) [(k_n - b_n^i) b_n^{i-1}]^{-1} k_n \) for \( 0 \leq i \leq ([k_n - b_n^i] b_n^{-i}] \), \( I_m(\tilde{l}) = \{i; l_i, l_{i+1}, l_{i+2} \} \). Then the absolute value of summation involving terms with \( l_q \) satisfying \( I_{l_q,m} < t_0 \) for some \( 1 \leq q \leq p \) on the right-hand side of (5.4) is less than
\[
\|A_{n,m}\| \sum_{p=1}^{\infty} \frac{t_n^{p+1}}{p!} \tilde{D}_{j,m}^{-1} (\tilde{D}_{j,m} - \tilde{D}_{j,m}) \tilde{D}_{j,m}^{-1/2} \|p^{-1}\| \tilde{D}_{j,m}^{-1/2} \|2 \tilde{b}_m^{i} \text{tr}(\tilde{D}_{j,m}^{-1/2} (r_n - \sum \varepsilon_i' \tilde{D}_{j,m}^{-1/2}) \leq r_n \sum \text{tr}(\tilde{D}_{j,m}^{-1} \varepsilon_i'),
\]
(5.5)
by (4.6), Lemma A.1, and the assumptions. Let \( J_1 = [T t_n b_n^{i-1} k_n^{-1} r_n^{-1}] \) and \( J_2 = [T t_n b_n^{i-1} k_n^{-1} r_n^{-1}] + 1 \), then \( J_1 \leq \# \{i; \inf I_{l_q,m} < t_0 \} \leq J_2 \). Therefore, point 2 of Lemma 4.1 ensures that \( \text{tr}(\tilde{D}_{j,m}^{-1} \varepsilon_i') \) is less than
\[
\frac{J_2}{[k_1 n/2] - J_2} \text{tr}(\tilde{D}_{j,m}^{-1} \tilde{D}_{j,m}^{-1} \varepsilon_i') \leq 2 J_2 [k_1 n/2] - J_2 \text{tr}(\tilde{D}_{j,m}^{-1} \tilde{D}_{j,m}^{-1} \varepsilon_i')
\]
if \([k_1 n/2] \geq 2 J_2 \), and hence the right-hand side of (5.5) is \( \hat{R}_n(V_t) \) by \( \text{tr}(\tilde{D}_{j,m}^{-1}) = \hat{R}_n(r_n^{-1/2} k_n) \), where \( V_t = b_1^{1/2 + \eta} 1(k_n/2) > 2 J_2 \) and \( b_n^{1/2 + \eta} 1(k_n/2) < 2 J_2 \).

On the other hand, for \( \tilde{l}, l_q, l_{q+1} \) and \( l_q \) satisfying \( l_q \in I_m(\tilde{l}) \) and \( \max I_m(\tilde{l}) \leq l_{q} \leq l_{q+1} \), Lemma 4.1 yields that \( P_{\text{max} I_m(\tilde{l}), \text{max} I_m(\tilde{l}), l_{q}, l_{q+1}} \) is less than 1 and greater than
\[
(1 + 1/\! J_1 + [b_n^{i}]^2 r_n J_1) - C b_n^{1/2 + \eta} \geq 1 - \hat{R}_n(b_1^{1/2 + \eta} + b_1^{1/2 + \eta} + \eta)
\]
for sufficiently large \( n \). Moreover, \( \max \{\ell \in I_m(\tilde{l}) (\tilde{D}_{j,m} - \tilde{D}_{j,m}) \ell, l \} = \hat{R}_n(b_1^{1/2 + \eta}) \) by (A2). We also have\( \text{sup}_{\varepsilon_i,m} \text{max} \{\ell \in I_m(\tilde{l}) (\tilde{D}_{j,m} - \tilde{D}_{j,m}) \ell, l \} = \hat{R}_n(b_1^{1/2 + \eta}) \) if (B2) is satisfied.
Therefore we obtain
\[
\text{tr}(\bar{D}^{\top}_{k,m}A_{n,m}) = \sum_{p=0}^\infty \frac{1}{p!} \sum_{\nu_1, \ldots, \nu_{p+1}} p_{\nu_1, \ldots, \nu_{p+1}} \left( \frac{1}{\nu_1! \nu_2! \cdots \nu_{p+1}!} \right) \sum_{q=1}^\infty \frac{(A_{n,m})_{\nu_1, \nu_2} \cdots \nu_{p+1}}{P_{\nu_1, \nu_2, \cdots, \nu_{p+1}, \nu_{p+1}}} \prod_{i=1}^p (\bar{D}_{j,m} - \bar{D}_{j,m})_{i,k,i} + R_n(V_n)
\]
\[
= \sum_{p=0}^\infty \frac{\tau_{m,1}^{n,p}}{p!} \sum_{\nu_1, \ldots, \nu_{p+1}} p_{\nu_1, \ldots, \nu_{p+1}} \left( \frac{1}{\nu_1! \nu_2! \cdots \nu_{p+1}!} \right) \sum_{q=1}^\infty \frac{(A_{n,m})_{\nu_1, \nu_2} \cdots \nu_{p+1}}{P_{\nu_1, \nu_2, \cdots, \nu_{p+1}, \nu_{p+1}}} \times \prod_{i=1}^p \sum_{q_i = 1}^{\max(I_m(i_q), \max(I_m(i_q), i_q, i_{q+1})} (\bar{D}_{j,m} - \bar{D}_{j,m} + \tau_{m,2}^{n,p}E_{i_q})_{i_q,i} + R_n(V_n)
\]
\[
= \sum_{p=0}^\infty \frac{\tau_{m,1}^{n,p}(\tau_{m,2}^{n,p})^{-1}}{p!} \text{tr}(\bar{D}^{\top}_{j,m}D^{1/2}_{j,m} - \bar{D}^{1/2}_{j,m} + \tau_{m,2}^{n,p}E_{j,m})^{p}D^{-1/2}_{j,m}A_{n,m} + \bar{R}_n(b_n^{3/2} \epsilon_n^{-1}),
\]
where \(\tau_{m,1}^{n,p}\) is a random variable which does not depend on \(l_q, l_q, i_q, i_{q+1}\) and satisfies
\[
(1 - \bar{R}_n(b_n^{n' - \eta'} + b_n^{1+\eta+\eta'}))^{p} \leq \tau_{m,1}^{n,p} \leq 1
\]
for \(i = 1, 2\) and \(\sup_p |\tau_{m,1}^{n,p}| = R_n(b_n^{-1})\).

Let \(F_p(t) = \text{tr}(\bar{D}^{\top}_{j,m}D^{1/2}_{j,m} - \bar{D}^{1/2}_{j,m} + \tau_{m,2}^{n,p}E_{j,m})^{p}D^{-1/2}_{j,m}A_{n,m}\). Then
\[
|F_p(1) - F_p(0)| \leq \int_0^1 |F'_p(t)|dt \leq b_n^{1/2} |\tau_{m,2}^{n,p}(1 - b_n^{1}(\bar{a}_n)')^{1/2} + b_n^{1/2} |\tau_{m,2}^{n,p}(1 - b_n^{1/2} \mathbf{r}_n^{1/2} b_n^{1} \epsilon_n^{-1})|,
\]
and hence \(\sum_{p=0}^\infty |\tau_{m,1}^{n,p}(\tau_{m,2}^{n,p})^{-1}||F_p(1) - F_p(0)| \leq \sup_p |\tau_{m,1}^{n,p}| \cdot \bar{R}_n(b_n^{3/2} k_n) = \bar{R}_n(b_n^{3/2} \epsilon_n^{-1})\). Therefore we obtain the desired conclusion by
\[
\sum_{p=0}^\infty \frac{\tau_{m,1}^{n,p}(\tau_{m,2}^{n,p})^{-1} - 1}{p!} \leq C \sup_p \bar{R}_n(b_n^{n' - \eta'} + b_n^{1+\eta+\eta'})(1 + \bar{R}_n(b_n^{n' - \eta'} + b_n^{1+\eta+\eta'}))^{p-1}(1 - b_n^{1}(\bar{a}_n)')^{1/2} \mathbf{r}_n^{1/2} k_n = \bar{R}_n(b_n^{3/2} \epsilon_n^{-1}).
\]
For the case \(k = 1\), we have \(\text{tr}(\partial_p \bar{D}^{\top}_{1,m}A_{n,m}) = -\text{tr}(\bar{D}^{\top}_{1,m} \partial_p \bar{D}_{1,m}A_{n,m} + R_n(b_n^{3/2} \epsilon_n^{-1})\) by using the result for \(k = 0\), \(\partial_p \bar{D}^{\top}_{1,m} = -\bar{D}^{\top}_{1,m} \partial_p \bar{D}_{1,m} \bar{D}^{-1}_{1,m}\). \(\|\bar{D}^{-1}_{1,m} \partial_p \bar{D}_{1,m} \| = R_1(1)\) and all elements of \(\bar{D}^{-1}_{1,m} \partial_p \bar{D}_{1,m}\) are nonnegative by a similar argument to (4.4). Then a argument to (5.6) enables us to replace \(\partial_p \bar{D}_{1,m}\) by \(\partial_p \bar{D}_{1,m}\). Similarly, we obtain \(\text{tr}(\partial_p^2 \bar{D}^{\top}_{1,m}A_{n,m}) = \text{tr}(\partial_p^2 \bar{D}^{\top}_{1,m}A_{n,m}) + \bar{R}_n(b_n^{3/2} \epsilon_n^{-1})\) for \(k = 2, 3\).

If further (B2) is satisfied, then similarly we have \(\sup_p |\text{tr}(\partial_p^2 \bar{D}^{\top}_{1,m}A_{n,m}) - \text{tr}(\partial_p^2 \bar{D}^{\top}_{1,m}A_{n,m})| = \bar{R}_n(b_n^{3/2} \epsilon_n^{-1})\).

\(\square\)

Remark 5.1. The proof shows that there are upper bounds of the absolute values of residual terms in the statement of Lemma 5.1 which do not depend on \(A_{n,m}\). Let \(c_j = |\bar{b}_n| b_n^{1}(\bar{a}_n')^{-1} / l_{j,v}\quad c'_j = c_j(\bar{a}_n'/\bar{a}_n')^2\) for \(j = 1, 2\).

Lemma 5.2. Assume (B1) and (A2). Then
\[
\sup_{\sigma,m} |\partial_{\sigma} \text{tr}((\bar{D}^{\top}_{1,m}G \bar{D}^{\top}_{2,m}G^{\top})^p \bar{D}^{\top}_{1,m}D_{1,m}^{-1}) - \frac{\bar{a}_m^{1/2} k_n}{\pi} \frac{(\bar{a}_m^{2})^{p+1} p_{\sigma+1,1}(c_1, c'_2)}{2^{p+1}(\bar{a}_m^{1})^{3p+2} l_{2,v}^{1}}| = o_p(b_n^{1/2} \epsilon_n^{-1}),
\]
\[
\sup_{\sigma,m} |\partial_{\sigma} \text{tr}((\bar{D}^{\top}_{2,m}G^{\top} \bar{D}^{\top}_{1,m}G \bar{D}^{\top}_{2,m}G^{\top})^p \bar{D}^{\top}_{1,m}D_{2,m}^{-1}) - \frac{\bar{a}_m^{1/2} k_n}{\pi} \frac{(\bar{a}_m^{2})^{p+1} p_{\sigma+1,1}(c_1, c'_2)}{2^{p+1}(\bar{a}_m^{1})^{3p+2} l_{2,v}^{1}}| = o_p(b_n^{1/2} \epsilon_n^{-1}),
\]
\[
\sup_{\sigma,m} |\partial_{\sigma} \text{tr}((\bar{D}^{\top}_{1,m}G \bar{D}^{\top}_{2,m}G^{\top})^p) - \frac{\bar{a}_m^{1/2} k_n}{\pi} \frac{(\bar{a}_m^{2})^{p+1} p_{\sigma+1,1}(c_1, c'_2)}{2^{p+1}(\bar{a}_m^{1})^{3p+2} l_{2,v}^{1}}| = o_p(b_n^{1/2} \epsilon_n^{-1})
\]
for $0 \leq k \leq 3$ and $p \in \mathbb{Z}_+$. If further (B2) is satisfied, then $\alpha_p(b_{1/2}^{-1} \ell_n^{-1})$ in (5.7)-(5.9) can be replaced by $\bar{R}_n(b_{1/2}^{-1} \ell_n^{-1})$.

**Proof.** For any $p \in \mathbb{N}$, Lemma 5.1 yields

$$b_n^{-1/2} \partial^p \xi((\tilde{D}_1^1, n \tilde{G} \tilde{D}_2^{-1, m} \tilde{G}^\top)^p) = b_n^{-1/2} \partial^p \xi((\tilde{D}_1^1, n \tilde{G} \tilde{D}_2^{-1, m} \tilde{G}^\top)^p) + \bar{R}_n(\ell_n^{-1}).$$

Moreover, we have

$$b_n^{-1/2} \text{tr}(\tilde{D}_1^1, n \tilde{G} \tilde{D}_2^{-1, m} \tilde{G}^\top)^p) = b_n^{-1/2} \sum_{i_1 \ldots \ldots i_p} \sum_{j_1 \ldots \ldots j_p} \prod_{q=1}^p \frac{\tilde{G}_{i_q j_q}}{\tilde{P}_{i_q} \tilde{P}_{j_q}}.$$  

by (4.4), where $i_{p+1} = i_1$ and $\tilde{P}_{a, \beta, i, j} = \prod_{l=1}^{i-1} p_{k_l} (c_1) \prod_{l=1}^{i-1} p_{k_l} (c_2).$

We will apply (4.3) to obtain the limit of the traces. To do so, we need to change the size of matrices $\tilde{G}$ and $\tilde{D}_2^{-1, m}$. This is again achieved by the nice properties of $p_i$. The essential idea is that point 3 of Lemma 4.1 ensures $p_i \approx_1$ for sufficiently large $i$, and therefore $\tilde{P}_{a, \beta, i, j} \approx p_i(c_1) i^{-\alpha} p_i(c_2) i^{-\beta} \approx \exp(\sqrt{\alpha} (i - \alpha) + \sqrt{\beta} (j - \beta)) \approx \tilde{P}_{k_0, a, k_0, k_0, k_0},$ where $k \in \mathbb{N}$ and $\tilde{P}_{a, \beta, i, j} = \prod_{k=1}^{i-1} p_{k_0} (c_1 / k^2) \prod_{k=1}^{i-1} p_{k_0} (c_2 / k^2).$ The size of $\tilde{D}_2^{-1, m}$ decides the ranges of summation of $j_1, \ldots, j_p$ in (5.11). By changing these ranges using the above relation on $\tilde{P}_{a, \beta, i, j}$ and $\tilde{P}_{k_0, a, k_0, k_0, k_0},$ we can change the size of matrices $\tilde{G}$ and $\tilde{D}_2^{-1, m}.$

Now we verify the above idea. First, we see that the terms involving small $a_0$ and $a_0'$ in (5.11) can be ignored. Let $\eta \in (0, 1/2)$ be the one in (A2), $\delta \in (1/2, 1)$ such that $\theta_{\eta}^{\delta} + \kappa_{\eta}^{-1} \rightarrow 0$ for any $\delta > 0$, $\tilde{G} = \tilde{G}_{\delta, \delta}^\top$, $\tilde{P}_{a, \beta, i, j} = \prod_{k=1}^{i-1} p_{k_0} (c_1 / k^2) \prod_{k=1}^{i-1} p_{k_0} (c_2 / k^2).$ Similarly to the proof of Lemma 5.1, the absolute value $\tilde{A}_1$ of a summation involving the terms with $(a_0, a_0')$ satisfying $\inf I_{a, a}^\top \in \tilde{G}^\top$ is less than $p_{\delta_{n/2}}^{-1} \text{tr}(\tilde{D}_1^1, n \tilde{G} \tilde{D}_2^{-1, m} \tilde{G}^\top + \tilde{G} \tilde{D}_2^{-1, m} (G')^\top + \tilde{D}_2^{-1, m} \tilde{G} \tilde{D}_2^{-1, m} (G')^\top)^p)$. Lemma 3 in [25] implies $\|G' + (G')^\top\| \leq 2r_n$ and hence all the eigenvalues of $G' + (G')^\top$ are greater than or equal to $-2r_n$. Therefore $G' + (G')^\top + 2r_n E''$ is nonnegative definite, and hence Lemma A.1 yields

$$\Lambda_1 \leq 2p(r_{n/2} / \gamma_n)^{2p-1} b_n^{-1/2} \text{tr}(\tilde{D}_3^{-1, m} (G' + (G')^\top + 2r_n E'')) \leq 4p(r_{n/2} / \gamma_n)^{2p-1} b_n^{-1/2} \text{tr}(\tilde{D}_3^{-1, m} G') + r_n \text{tr}(\tilde{D}_3^{-1, m} E'').$$

Let $(\tilde{G})_{i,j} = (\sum_{i \leq m} G_{i,j} + \sum_{m < j} G_{i,j}) \delta_{i,j}$ and $k = \max\{i; \tilde{G}_{i,i} > 0\}$. Then Lemma 4.1 yields

$$\text{tr}(\tilde{D}_1^{-1, m} G') = \frac{1}{v_{1, a} \wedge v_{2, a}} \sum_{a, b \leq i} \sum_{a, b \leq i} G_{a,b} \sum_{a, b \leq i} \sum_{a, b \leq i} \tilde{G}_{a,b} \sum_{a, b \leq i} \sum_{a, b \leq i} \delta_{i,j} \sum_{i \leq m} G_{i,j} (\delta_0 - s_{m-1}) + r_n \text{tr}(\tilde{D}_1^{-1, m}).$$

Therefore we obtain $\Lambda_1 = \bar{R}_n(\ell_n^{-1})$, and $\Lambda_1 = \bar{R}_n(\ell_n^{-1})$ if (B2) is satisfied.

Let $\tilde{D}_2 = v_{2, a} c_2 (a_{1, 2}^2 a_{1, 2}^2 E + v_{2, a} M_{1, i, m}, i'(a) = \min\{i; S_{i}^n, 1 \geq \delta_{a, -1}\}$, and $j'(a) = \min\{j; S_{j}^n, 2 \geq \delta_{a, -1}\}$. We will show that $b_n^{-1/2} \text{tr}(\tilde{D}_1^{-1, m} G \tilde{D}_2^{-1, m} \tilde{G}^\top)^p)$ is approximated by $b_n^{-1/2} \text{tr}(\tilde{D}_1^{-1, m} \tilde{D}_2^{-1, m} \tilde{G}^\top)^p)$. First, a similar argument to the proof of Lemma 5.1 yields $P_{a, \beta,i,j} / \tilde{P}_{a, \beta,i,j} (\alpha', j'(a'), i'(a'), j'(a') - 1) = R_n(1)$ for $i'(a) \leq \alpha < i'(a' + 1)$, $j'(a') \leq \beta < j'(a' + 1)$, and $j'(a') - 1 < i'(a') < j'(a' + 1)$, and $j'(a') \leq j'(a' + 1)$. Therefore repeated use of (A2)
yields
\[ b_n^{-1/2} \text{tr}(\hat{D}_{1,m}^{-1} \hat{G} \hat{D}_{2,m}^{-1} \hat{G}^T) p) = b_n^{-1/2} \mathcal{T}_{m, \alpha} \sum_{i_1, \ldots, i_p} \sum_{j_1, \ldots, j_p} \frac{\prod_{q=1}^p (T_{n+1,q})^2 \{ i_q; I_{1,m} \subset [\hat{s}_{\alpha^{-1}}, \hat{s}_{\alpha^2}] \} \# \{ j_q; I_{2,m} \subset [\hat{s}_{\alpha^2}, \hat{s}_{\alpha^{-1}}] \} \quad \hat{R}_n(l_n^-) \}
\]
\[ = b_n^{-1/2} \mathcal{T}_{m, \alpha} (T_{n+1,q})^{2p} \frac{\prod_{i=1}^p \# \{ \alpha; I_{1,m} \subset [\hat{s}_{\alpha^{-1}}, \hat{s}_{\alpha^2}] \} \# \{ \alpha; I_{2,m} \subset [\hat{s}_{\alpha^2}, \hat{s}_{\alpha^{-1}}] \} }{\alpha_n(l_n^-) - \hat{R}_n(1)} = (5.12) \]

where \( \mathcal{T}_{m, \alpha} \) is a random variable satisfying \( \sup_{\alpha, \sigma} |\mathcal{T}_{m, \alpha} - 1| = \hat{R}_n(1) \) for \( i = 4, 5 \).

Since Lemma 4.1 and (A2) yield
\[ p_j(\alpha) \cdots p_j(\beta) c_2 = (p_j(c_2))^{j - j(\alpha) + j(\beta)} (1 + \hat{R}_n(1)) \exp(b_n \hat{a}_n^2 (\hat{s}_\alpha - \hat{s}_\beta) + \hat{R}_n(b_n^{1/2})) \log p_+ (c_2) (1 + \hat{R}_n(1)) \]
we may replace \( \hat{P}_{i(\alpha \beta), j(\alpha \beta), i(\gamma_j), j(\gamma_j)} \) and \( \hat{P}_{i(\alpha \beta), j(\alpha \beta), i(\gamma_j), j(\gamma_j)} \) in the right-hand side of (5.12) by \( \hat{P}_{i(\alpha \beta), j(\alpha \beta), i(\gamma_j), j(\gamma_j)} \) and \( \hat{P}_{i(\alpha \beta), j(\alpha \beta), i(\gamma_j), j(\gamma_j)} \), respectively, where \( \hat{P}_{n, \beta, i, j} = \prod_{k_1=0}^{i-1} p_{k_1}(c_1) \prod_{k_2=0}^{j-1} p_{k_2}(c_2) \).

Therefore, we obtain
\[ \sup_{\alpha, \sigma} b_n^{-1/2} \text{tr}(\hat{D}_{1,m}^{-1} \hat{G} \hat{D}_{2,m}^{-1} \hat{G}^T) = \exp(b_n \hat{a}_n^2 (\hat{s}_\alpha - \hat{s}_\beta) + \hat{R}_n(b_n^{1/2})) \log p_+ (c_2) \]
by a similar argument to (5.12). Since \( \partial_n^t \hat{D}_{1,m} = \partial_n^t c_j v_j, \varepsilon \) for 1 \( \leq l \leq 3 \), we similarly obtain
\[ \sup_{\alpha, \sigma} b_n^{-1/2} \partial_n^t \text{tr}(\hat{D}_{1,m}^{-1} \hat{G} \hat{D}_{2,m}^{-1} \hat{G}^T) - b_n^{-1/2} \partial_n^t (\hat{a}_n^2) \log p_+ (c_2) \]
by a similar argument to (5.12). Then (4.3), (5.10) and (5.13) yield (5.9).

We also have (5.7) and (5.8) by a similar argument.

Similar arguments enable us to replace \( a_n(b_n^{1/2} \ell^{-1}) \) by \( R_n(b_n^{1/2} \ell^{-1}) \) in (5.7)-(5.9) if (B2) is satisfied.

Remark 5.2. We can also show that the summations of the left-hand side of (5.7)-(5.9) in 1 \( \leq p < \infty \) are bounded by \( R_n(b_n^{1/2} \ell^{-1}) \sum_{j=0}^{\infty} p(b_j, b_j^2, b_j^3, b_j^4)^{2(\rho - 1)} \) for some \( \ell \in \mathbb{N} \) if (B2) is satisfied. To show this, it is sufficient to replace \( \hat{D}_{1,m} \hat{G} \hat{D}_{2,m} \hat{G}^T \) in the traces with \( b_n^{-1} \hat{a}_n^2 (\hat{a}_n^1)^{-3} \hat{D}_{1,m} \hat{D}_{2,m} \) one by one, and estimate residual terms separately on the event \( \{ b_n^{-1} - \varepsilon \leq \ell_n \leq b_n^{-1} + \varepsilon \} \) and on its complement for some small \( \varepsilon > 0 \). This fact is used in the proof of Proposition 2.2.

Proof of Proposition 2.1.

We first prove the results under the additional condition (A1').

Since \( \| \hat{D}_{1,m} \hat{G} \hat{D}_{2,m} \hat{G}^T \| \leq \| \hat{b}_m^{-2} \hat{b}_m^{-2} \hat{b}_m^{-2} \| \) by Lemma 3 in [25], for any \( \varepsilon, \delta > 0 \), there exists \( P_{1} \in \mathbb{N} \) such that
\[ \sup_n P \left[ \sup_{\sigma} b_n^{-2} \sum_{m=1}^{\infty} \sum_{p=P_{1}+1}^{\infty} |\partial_n^t (\hat{b}_m \hat{b}_m^{-2} \hat{b}_m^{-3}) \text{tr}(\hat{D}_{1,m} \hat{D}_{2,m} \hat{G}^T) | \geq \delta \right] < \varepsilon, \]
\[
\sup_n P \left[ \sup_{\sigma} \frac{1}{n} \sum_{p=1}^{\infty} \sum_{m=p+1}^{\infty} \left| \partial_{\sigma}^k \left( \frac{\langle \hat{b}_m, \hat{b}_m^2 \rangle^{2p-1} \hat{a}_m k_n (\hat{a}_m^2)^p I_{p,p}(c_1, c_2)}{(\hat{a}_m^2)^{3v_1^*+v_2^*}} \right) \right| \geq \delta \right] \leq \epsilon \quad (5.15)
\]

for \( P \geq P_1 \). Together with Lemma 5.2, we obtain

\[
\sup_{\sigma} \left| b_n^{-1/2} \partial_{\sigma}^k \sum_{m=p}^{\infty} \sum_{p=1}^{\infty} \left( \frac{\langle \hat{b}_m, \hat{b}_m^2 \rangle^{2p-1} \hat{a}_m k_n (\hat{a}_m^2)^p I_{p,p}(c_1, c_2)}{(\hat{a}_m^2)^{3v_1^*+v_2^*}} \right) - P_n \right| \to P_0.
\]

Let \( \hat{a}_m^k = \hat{a}_{m-1}^k \), \( c_m = \frac{b_n^2 \hat{a}_m^k |b_m|^2 - (\hat{b}_m, \hat{b}_m^2)^2}{2(\hat{a}_m^2)^{3v_1^*+v_2^*}} \), \( \mathcal{U} = \left( \hat{a}_m^k |b_m|^2 + \hat{a}_m^2 |b_m|^2 + 2\sqrt{\hat{a}_m^2 |b_m|^2} \right)^{1/2} \), and

\[
P_n = (c_1 + c_2 + 2\sqrt{b_n^2 \hat{a}_m^k |b_m|^2 - \hat{a}_m^2 |b_m|^2 + 2\sqrt{b_n^2 \hat{a}_m^k |b_m|^2 - \hat{a}_m^2 |b_m|^2} + 2\sqrt{\hat{a}_m^2 |b_m|^2} )^{1/2}.
\]

Then Lemma A.9 yields

\[
\left| b_n^{-1/2} \partial_{\sigma}^k \sum_{m=p}^{\infty} \sum_{p=1}^{\infty} \left( \frac{\langle \hat{b}_m, \hat{b}_m^2 \rangle^{2p-1} \hat{a}_m k_n (\hat{a}_m^2)^p I_{p,p}(c_1, c_2)}{(\hat{a}_m^2)^{3v_1^*+v_2^*}} \right) - \partial_{\sigma}^k \int_0^T \frac{\sqrt{a_1^2 a_2^2 + b_1^2}}{2 \sqrt{\det(b_1 b_2^* \mathcal{U})}} dt \right| \to P_0. \quad (5.16)
\]

Similarly, we obtain

\[
\sup_{\sigma} \left| b_n^{-1/2} \partial_{\sigma}^k \sum_{m=p}^{\infty} \sum_{p=0}^{\infty} \left( \frac{\langle \hat{b}_m, \hat{b}_m^2 \rangle^{2p-1} \hat{a}_m k_n (\hat{a}_m^2)^p I_{p,p}(c_1, c_2)}{(\hat{a}_m^2)^{3v_1^*+v_2^*}} \right) - \partial_{\sigma}^k \int_0^T \frac{\sqrt{a_1^2 a_2^2 + b_1^2}}{2 \sqrt{\det(b_1 b_2^* \mathcal{U})}} dt \right| \to P_0. \quad (5.17)
\]

\[
\sup_{\sigma} \left| b_n^{-1/2} \partial_{\sigma}^k \sum_{m=p}^{\infty} \sum_{p=0}^{\infty} \left( \frac{\langle \hat{b}_m, \hat{b}_m^2 \rangle^{2p-1} \hat{a}_m k_n (\hat{a}_m^2)^p I_{p,p}(c_1, c_2)}{(\hat{a}_m^2)^{3v_1^*+v_2^*}} \right) - \partial_{\sigma}^k \int_0^T \frac{\sqrt{a_1^2 a_2^2 + b_1^2}}{2 \sqrt{\det(b_1 b_2^* \mathcal{U})}} dt \right| \to P_0. \quad (5.18)
\]

Furthermore, Lemmas A.3 and 5.2 and a similar argument yield

\[
\partial_{\sigma}^k \log \det(\hat{S}_m \hat{D}_m^{-1}) = \partial_{\sigma}^k \log \det \left( \begin{pmatrix} 0 & \hat{D}_{1,m}^{-1/2} \hat{D}_{1,m}' \hat{D}_{2,m}^{-1/2} \\ \hat{D}_{2,m}^{-1/2} \hat{L}_m \hat{D}_{2,m}^{-1} & 0 \end{pmatrix} \right)
\]

\[
= -\sum_{p=1}^{\infty} \partial_{\sigma}^k \text{tr}((\hat{D}_{1,m}^{-1} \hat{L}_m \hat{D}_{2,m}^{-1} \hat{L}_m')^p) + \partial_{\sigma}^k \int_0^T \frac{\sqrt{a_1^2 a_2^2 + b_1^2}}{2 \sqrt{\det(b_1 b_2^* \mathcal{U})}} dt + o_p(b_n^{1/2} \ell_n^{-1}).
\]

Then Lemma A.9 yields

\[
\partial_{\sigma}^k \log \det(\hat{S}_m \hat{D}_m^{-1})
\]

\[
= -T^{1/2} k_n \partial_{\sigma}^k \left( \frac{\sqrt{a_1^2 a_2^2 |b_m|^2}}{b_0 a_m^2 v_1^* v_2^*} + \sqrt{a_1^2 a_2^2 |b_m|^2} - b_0^{1/2} (a_1^2 |b_m|^2 + a_1^2 |b_m|^2 + 2\sqrt{a_1^2 a_2^2 \det(b_0 b_0^* m)} \right)^{1/2} + o_p(b_n^{1/2} \ell_n^{-1})
\]

\[
+ o_p(b_n^{1/2} \ell_n^{-1})
\]

\[
= T^{1/2} k_n \partial_{\sigma}^k \left( \frac{\sqrt{a_1^2 a_2^2 |b_m|^2}}{b_0 a_m^2 v_1^* v_2^*} + \sqrt{a_1^2 a_2^2 |b_m|^2} + 2\sqrt{a_1^2 a_2^2 \det(b_0 b_0^* m)} \right)^{1/2} - \sqrt{a_1^2 |b_m|^2} - \sqrt{a_2^2 |b_m|^2} + o_p(b_n^{1/2} \ell_n^{-1}). \quad (5.19)
\]

23
Moreover, Lemmas A.3 and 5.1 yield
\[
\frac{\partial^k}{\partial \sigma^k} \log \det(D_{j,m}D_{j,m}^{-1}) = \frac{\partial^k}{\partial \sigma^k} \log \det(E + D_{j,m}^{-1/2}(D_{j,m} - D_{j,m,s})D_{j,m}^{-1/2}) = \sum_{p=1}^{\infty} \frac{(-1)^{p-1}}{p} \partial^k \log \det((D_{j,m}^{-1}(D_{j,m} - D_{j,m,s}))^p)
\]
\[
= \sum_{p=1}^{\infty} \frac{(-1)^{p-1}}{p} \partial^k \log \det((D_{j,m}^{-1}(D_{j,m} - D_{j,m,s}))^p) + o_p(b_1^{1/2} \epsilon^{-1} n_0) = \frac{\partial^k}{\partial \sigma^k} \log \det(D_{j,m}D_{j,m}^{-1}) + o_p(b_1^{1/2} \epsilon^{-1} n_0)
\]
(5.20)
when \(|b_{m,s}^1| \geq |b_{m}^1|\), where \(D_{j,m,s}\) and \(D_{j,m}\) are obtained by substituting \(\sigma = \sigma_s\) in \(D_{j,m}\) and \(D_{j,m,s}\), respectively. Similarly, we have \(\frac{\partial^k}{\partial \sigma^k} \log \det(D_{j,m}D_{j,m}^{-1}) = \frac{\partial^k}{\partial \sigma^k} \log \det(D_{j,m}D_{j,m}^{-1}) + o_p(b_1^{1/2} \epsilon^{-1} n_0)\) when \(|b_{m,s}^1| < |b_{m}^1|\).

On the other hand, results in Section 4.2 yield
\[
\frac{\partial^k}{\partial \sigma^k} \log \det(D_{j,m}) = \frac{k_{m}^l + 1}{\pi} \int_0^\pi \frac{c_j + 2(1 - \cos x)}{c_{j,s} + 2(1 - \cos x)} dx + o_p(b_1^{1/2} \epsilon^{-1} n_0)
\]
\[
= 2(k_{m}^l + 1) \frac{\partial^k}{\partial \sigma^k} \log \frac{\sqrt{c_j} + \sqrt{4 + c_j}}{\sqrt{c_{j,s}} + \sqrt{4 + c_{j,s}}} + o_p(b_1^{1/2} \epsilon^{-1} n_0)
\]
\[
= k_{m}^l \frac{\partial^k}{\partial \sigma^k} (\sqrt{c_j} - \sqrt{c_{j,s}}) + o_p(b_1^{1/2} \epsilon^{-1} n_0) = T b_{1}^{1/2} \epsilon^{-1} n_0 \sqrt{\sum_{m=1}^M \frac{\partial^k}{\partial \sigma^k} (|b_{m}^1| - |b_{m,s}^1|)} + o_p(b_1^{1/2} \epsilon^{-1} n_0)
\]
(5.21)
The residuals are bounded uniformly with respect to \(\sigma\) and \(m\). Then we obtain \(\sup_{\sigma} |b_{m}^{-1/2} \partial^k \log H_{n}^0 (\sigma, \nu_n) - H_n(\sigma, \nu_n)| - \frac{\partial^k}{\partial \sigma^k} \mathcal{Y}_1(\sigma)| \to^p 0\) as \(n \to \infty\) for any \(\sigma \in \Lambda\) and \(0 \leq k \leq 3\) by (5.2), (5.3) and (5.16)–(5.21).

Finally, we obtain the desired results without \((A1')\) by using the arguments in Proposition 3.1 of Gioter and Jacod [13].

\[\square\]

6 Identifiability of the model

In this section, we check the identifiability condition, \(\inf_{\sigma \neq \sigma_0} ((-\mathcal{Y}_1(\sigma))/|\sigma - \sigma_0|^2) > 0\) almost surely. This condition is necessary to deduce consistency of the maximum-likelihood-type estimator, as seen in Proposition 7.1. In general, it is not easy to check this condition directly because \(\mathcal{Y}_1(\sigma)\) is a complicated function of \(b_1^1\) and \(a_1^2\). On the other hand, Ogihara and Yoshida [25] proved that the identifiability condition \((A3)\) of a model for equidistant observations without noise is sufficient for the identifiability of a model for nonsynchronous observations. This is also the case for our model. Recall that \(\tilde{\rho}\) has been defined just before Lemma 4.2.

Proposition 6.1. Assume \((A1), (A2)\) and \((V)\). Then there exists a positive constant \(c\) such that
\[
-\mathcal{Y}_1(\sigma) \geq c \int_0^\tau \{(b_{1}^1 - |b_{1}^1|)^2 + (b_{2}^1 - |b_{2}^1|)^2 + (b_{1}^1 \cdot b_{2}^1 - b_{1}^2 \cdot b_{2}^2)^2\} dt
\]
(6.1)
for any \(\sigma\), where
\[
\chi = c(1 - \tilde{\rho}^2) \inf_{j,t} \left( \frac{(a_{ij}^1)^2}{v_j(t,s)} \right)^{1/2} \left( \sup_{j,t,s} \left( \frac{|b^j(t, X_t, \sigma)| \vee |\rho^j(t, X_t, \sigma)|^{-1}}{v_j(t,s)} \right) \right)^{-9}.
\]
In particular, \(\inf_{\sigma \neq \sigma_0} ((-\mathcal{Y}_1(\sigma))/|\sigma - \sigma_0|^2) > 0\) almost surely under \((A1)-(A3)\) and \((V)\).

Proof. It is sufficient to show the results under the additional condition \((A1')\) by localization techniques similar to the proof of Proposition 2.1.

Let \(\tilde{\rho}_m = \left| b_{m}^{-1} \right| \left| b_{m}^{-1} \right| \left| b_{m}^{-1} \right| D_{j,m} - D_{m}, \tilde{D}_{m} = \tilde{D}_{m} - M_{m,s}\) and \(B = \sup_{j,t,s} \{|b^j(t, X_t, \sigma)| \vee |\rho^j(t, X_t, \sigma)|^{-1}\}\), then since
\[
\left( E \tilde{\rho}_m \left\{ \left| I_{t,m}^1 \right| \left| I_{t,m}^1 \right| \left| I_{t,m}^1 \right| \left| I_{t,m}^1 \right| \left| I_{t,m}^1 \right| \right\}_{t,i} \right) \left( E \tilde{\rho}_m \left\{ \left| I_{t,m}^1 \right| \left| I_{t,m}^1 \right| \left| I_{t,m}^1 \right| \left| I_{t,m}^1 \right| \left| I_{t,m}^1 \right| \right\}_{t,i} \right)
\]
\[
\geq \left( E \tilde{\rho}_m \left\{ \left| I_{t,m}^1 \right| \left| I_{t,m}^1 \right| \left| I_{t,m}^1 \right| \left| I_{t,m}^1 \right| \left| I_{t,m}^1 \right| \right\}_{t,i} \right) \left( E \tilde{\rho}_m \left\{ \left| I_{t,m}^1 \right| \left| I_{t,m}^1 \right| \left| I_{t,m}^1 \right| \left| I_{t,m}^1 \right| \left| I_{t,m}^1 \right| \right\}_{t,i} \right)
\]
24
for any $u \in \mathbb{R}^{k_m+k_n^2}$, we have $\| (\hat{D}_m^{1/2} \hat{S}_{m,\ast} \hat{D}_m^{1/2})^{1/2} \| \leq C B (1-\rho^2)^{-1/2}$ by Lemma A.4, and hence we obtain

$$
\| (\hat{D}_m^{1/2} \hat{S}_{m,\ast} \hat{D}_m^{1/2})^{1/2} (\hat{D}_m^{1/2} \hat{S}_{m,\ast} \hat{D}_m^{1/2}) (\hat{D}_m^{1/2} \hat{S}_{m,\ast} \hat{D}_m^{1/2})^{1/2} \|
= \| \mathcal{E} + (\hat{D}_m^{1/2} \hat{S}_{m,\ast} \hat{D}_m^{1/2})^{1/2} (\hat{D}_m^{1/2} \tilde{S}_m \hat{D}_m^{1/2}) (\hat{D}_m^{1/2} \hat{S}_{m,\ast} \hat{D}_m^{1/2})^{1/2} \| \leq 1 + C B (1-\rho^2)^{-1}.
$$

Then Lemma A.6 yields

$$
\text{tr}(\tilde{S}_m \tilde{S}_{m,\ast} - \mathcal{E}) + \log \det \tilde{S}_m - \log \det \tilde{S}_{m,\ast} \geq C B^{-6} (1-\rho^2) \text{tr}(\tilde{S}_m (\tilde{S}_{m,\ast} - \tilde{S}_m) (\tilde{S}_{m,\ast} - \tilde{S}_m)). \quad (6.2)
$$

Therefore, repeated use of Lemma A.5 yields

$$
\text{tr}(\tilde{S}_m \tilde{S}_{m,\ast} - \mathcal{E}) + \log \det \tilde{S}_m - \log \det \tilde{S}_{m,\ast} \geq C B^{-6} (1-\rho^2) \text{tr}(\tilde{S}_m (\tilde{S}_{m,\ast} - \tilde{S}_m) (\tilde{S}_{m,\ast} - \tilde{S}_m)) \geq C B^{-6} (1-\rho^2) \text{tr}(\mathcal{D}_m^{-1} (\tilde{S}_{m,\ast} - \tilde{S}_m) (\tilde{S}_{m,\ast} - \tilde{S}_m))
$$

Therefore, it is sufficient to show that $\lim sup$ of $\text{tr}(\mathcal{D}_m^{-1} (\mathcal{D}_m^{-1} G \mathcal{D}_m^{-1} \mathcal{D}_m G))$ for $j = 1, 2$ and $\text{tr}(\mathcal{D}_m^{-1} G \mathcal{D}_m^{-1} \mathcal{D}_m G)$ are estimated from below by positive random variables.

By Lemma 5.1, (4.2) and (5.13) with a sampling scheme $\{S^{n,1}_i\}_{i \in \mathbb{N}} \equiv \{S^{n,2}_j\}_{j \in \mathbb{N}}$, we obtain

$$
b_n^{-1/2} \text{tr}(\mathcal{D}_m^{-1} (\mathcal{D}_m^{-1} G \mathcal{D}_m^{-1} \mathcal{D}_m G)) = b_n^{-1/2} \text{tr}(\mathcal{D}_m^{-1} (\mathcal{D}_m^{-1} G \mathcal{D}_m^{-1} \mathcal{D}_m G)) + \hat{R}_n(\ell_n^{-1}) = b_n^{-1/2} \text{tr}(\mathcal{D}_m^{-1} (\mathcal{D}_m^{-1} G \mathcal{D}_m^{-1} \mathcal{D}_m G)) + \hat{R}_n(\ell_n^{-1})
$$

Moreover, Lemma 5.1, (4.2) and (5.13) yield

$$
b_n^{-1/2} \text{tr}(\mathcal{D}_m^{-1} G \mathcal{D}_m^{-1} \mathcal{D}_m G) = b_n^{-1/2} \frac{\|e_n^2\|^2}{\|e_n^2\|^2} \text{tr}(\mathcal{D}_m^{-1} (\mathcal{D}_m^{-1} G \mathcal{D}_m^{-1} \mathcal{D}_m G)) + \hat{R}_n(\ell_n^{-1})$$

Therefore, we obtain (6.1).

In particular, by Lemma 6 and Remark 4 in [25], there exists a positive-valued random variable $\mathcal{R}$ such that

$$
-\mathcal{Y}_1(\sigma) \geq \chi \mathcal{R}(-\mathcal{Y}_0(\sigma))
$$

for any $\sigma$. Therefore we have $\inf_{\sigma \neq \sigma_0}(\sigma) > 0$ almost surely under (A1)-(A3) and (V).

\section{Asymptotic mixed normality of the estimator}

In this section we prove the consistency and asymptotic mixed normality of $\hat{\sigma}_n$. To obtain asymptotic mixed normality, we prove stable convergence of the score function $b_n^{-1/4} \partial_{\sigma} \mathcal{H}_n(\sigma_n, v_n)$ by means of the martingale limit theorem for a mixed normal limit in Jacod [18]. We also use the idea by Jacod et al. [19] to adapt the limit theorem to models containing observation noise.

Consistency is an immediate consequence of Proposition 2.1 and the identifiability condition.

\begin{proposition}
Assume (A1)-(A3) and (V). Then $\hat{\sigma}_n \rightarrow^p \sigma_*$ as $n \rightarrow \infty$.
\end{proposition}

\begin{proof}
\end{proof}

\begin{corollary}
\end{corollary}
Proof. Let $\epsilon, \delta$ be arbitrary positive constants. By Proposition 2.1, we have $\sup_\sigma |b_n^{-1/2}(H_n(\sigma, \hat{v}_n) - H_n(\sigma, \hat{v}_n)) - \mathcal{Y}_1(\sigma)| \to^P 0$ as $n \to \infty$. Moreover, Proposition 6.1 ensures that there exists $\eta > 0$ such that $P[\inf_{\sigma \neq \sigma}(\{-\mathcal{Y}_1(\sigma)/|\sigma - \sigma|^2\}) \leq \eta] < \epsilon$. Since $H_n(\sigma, \hat{v}_n) - H_n(\sigma, \hat{v}_n) \geq 0$ by the definition of $\sigma$, we obtain

\[ P[|\hat{\sigma}_n - \sigma| \geq \delta] < P[|\mathcal{Y}_1(\hat{\sigma}_n) \leq -\eta \delta^2| + \epsilon + P[\sup_\sigma |b_n^{-1/2}(H_n(\sigma, \hat{v}_n) - H_n(\sigma, \hat{v}_n)) - \mathcal{Y}_1(\sigma)| \geq \eta \delta^2] + \epsilon < 2\epsilon \]

for sufficiently large $n$.

\[ \Box \]

**Proposition 7.2.** Assume (A1), (A2) and (V). Then $b_n^{-1/4}\partial_\sigma H_n(\sigma, \hat{v}_n) \to^{\text{c}} \Gamma_{1/2}^1$ as $n \to \infty$.

Proof. It is sufficient to prove the results assuming the additional condition (A1').

Since $b_n^{-1/4}\partial_\sigma H_n(\sigma, v_\epsilon) = -2^{-1}b_n^{-1/4}\sum_m E_m[\bar{Z}_m^T \partial_\sigma \bar{S}_m^{-1} \bar{Z}_m]$, we only need to check assumptions of Theorem 3.2 in Jacod [18] for $\mathcal{X}^n_m = -2^{-1}b_n^{-1/4}E_m[\bar{Z}_m^T \partial_\sigma \bar{S}_m^{-1} \bar{Z}_m]$. For any $\epsilon > 0$, Lemma 4.3 yields

\[ \sum_{m=1}^{[\epsilon n]} E_m[|X_m^n|^21_{|X_m^n| > \epsilon}] \leq \frac{Cn^{1/2}}{\epsilon^2} \sum_{m=1}^{[\epsilon n]} E_m[(\bar{Z}_m^T \partial_\sigma \bar{S}_m^{-1} \bar{Z}_m)^4] \to^P 0. \]  

(7.1)

Moreover, it is easy to see that $\sum_{m=1}^{[\epsilon n]} E_m[|X_m^n(W_m - W_{m-1})|] = 0$.

Lemma 4.3 yields

\[ E_m[|X_m^n|^2] = \frac{b_n^{1/2}}{4} E_m[(\bar{Z}_m^T \partial_\sigma \bar{S}_m^{-1} \bar{Z}_m)^2] - E_m[(\bar{Z}_m^T \partial_\sigma \bar{S}_m^{-1} \bar{Z}_m)^2] \approx -\frac{1}{2} \text{tr}(\bar{S}_m^T \partial^2 \bar{S}_m^{-1} \bar{S}_m). \]

On the other hand, since $\partial_\sigma \log \det \bar{S}_m(x, \sigma) = (\partial_\sigma S_m)^{-1}$, we have

\[ E_m[\bar{Z}_m^T \partial_\sigma \bar{S}_m^{-1} \bar{Z}_m + \partial_\sigma \log \det \bar{S}_m]|_{\sigma = \sigma} = (\partial_\sigma S_m)^{-1} \bar{S}_m + \partial_\sigma \log \det \bar{S}_m) \approx (\partial_\sigma S_m)^{-1} \bar{S}_m + \partial_\sigma \log \det \bar{S}_m). \]

Therefore we have

\[ \sum_{m=1}^{[\epsilon n]} E_m[|X_m^n|^2] = \frac{b_n^{1/2}}{2} \sum_{m=1}^{[\epsilon n]} E_m[(\bar{Z}_m^T \partial_\sigma \bar{S}_m^{-1} \bar{Z}_m + \partial_\sigma \log \det \bar{S}_m)|_{\sigma = \sigma}] \to^P [\partial_\sigma \mathcal{Y}_1(\sigma, t), \sigma = \sigma], \]  

(7.2)

where

\[ \mathcal{Y}_1(\sigma, t) = \int_0^t \left\{ \sum_{j=1}^{[\epsilon n]} |b_j|_1^2 |b_j|_2^2 |b_j|_1^2 \sqrt{\alpha_j \alpha_j} + \sqrt{\alpha_j \alpha_j} \sqrt{\text{det}(b_j b_j^T)} - 2 b_j b_j^T - b_j b_j^T + b_j b_j^T \right\} \left( \frac{4 \sqrt{\text{det}(b_j b_j^T)} (\alpha_j \alpha_j b_j b_j^T)}{2 \sqrt{\alpha_j \alpha_j} \text{det}(b_j b_j^T)} \right)^{1/2} \right. \]

\[ - \left( \frac{2 \sqrt{\alpha_j \alpha_j} \text{det}(b_j b_j^T)}{2 \alpha_j \alpha_j} \right)^{1/2} - \left( \frac{\alpha_j \alpha_j b_j b_j^T}{2 \alpha_j \alpha_j} \right)^{1/2} \]  

\[ + \left( \frac{\alpha_j \alpha_j b_j b_j^T}{2 \alpha_j \alpha_j} \right)^{1/2} \right) \right\} \right\} \right) \right\} \right) \right\}. \]

Then by Theorem 3.2 in Jacod [18], it is sufficient to show that $\sum_{m=1}^{[\epsilon n]} E_m[\mathcal{X}^n_m(N_m - N_{m-1})] \to^P 0$ for any bounded $\mathcal{G}_t$-martingale $N = (N_t)_{0 \leq t \leq T}$ orthogonal to $W$.

Orthogonality of $N$ and $W$ yields

\[ E_m[\mathcal{X}^n_m(N_m - N_{m-1})] = E_m[\mathcal{X}^n_m(N_m - N_{m-1})], \]

where $\mathcal{X}^n_m = -2^{-1}b_n^{-1/4}E_m[\bar{Z}_m^T \partial_\sigma \bar{S}_m^{-1} \bar{Z}_{2,m} + 2 \bar{Z}_m^T \partial_\sigma \bar{S}_m^{-1} \bar{Z}_{2,m}]$, and $\bar{Z}_{1,m}$ and $\bar{Z}_{2,m}$ are defined in (A.4).

Let $N' = (N'_t)_{0 \leq t \leq T}$ be another bounded $\mathcal{G}_t$-martingale. Then we obtain

\[ \sum_{m=1}^{[\epsilon n]} E_m[\mathcal{X}^n_m(N_m - N_{m-1})] = \sum_{m=1}^{[\epsilon n]} E_m[\mathcal{X}^n_m(N'_m - N'_{m-1})] + \sum_{m=1}^{[\epsilon n]} E_m[\mathcal{X}^n_m(N_m - N'_m - N_{m-1} + N'_{m-1})], \]  

26
and

\[
E \left[ \sum_{m=1}^{\ell_n} E_m[\mathcal{X}_m^n(N_{sm} - N'_{sm} - N_{sm-1} + N'_{sm-1})] \right]^{1/2} E_m[\mathcal{X}_m^n]^2 \leq K
\]

for any positive constant \(K\).

Since we have the tightness of \(\sum_{m=1}^{\ell_n} E_m[\mathcal{X}_m^n]^2\) similarly to (7.2), we obtain the desired result if for any \(\epsilon > 0\), we can find a bounded \(\mathcal{G}_t\)-martingale \(N'\) which satisfies \(E[(N_T - N'_T)^2] \leq \epsilon\) and \(E_m[\mathcal{X}_m^n(N_{sm} - N'_{sm})] = 0\) for any \(m\). We obtain such \(N'\) by setting \(N'_i = E[N|\mathcal{G}_t]\) for suitable \(N \in \mathbb{N}\), where \(\mathbb{N}\) is the set of finite sums of random variables \(X\), \(n_1, \ldots, n_i \in \mathbb{N}\), \(1 \leq k_1, \ldots, k_i \leq 2\), and \(i_1, \ldots, i_j \in \mathbb{Z}_+\). Then \(N\) is dense in \(L^2(\Omega, \mathcal{G}_T, P)\) and we obtain \(E_m[\mathcal{X}_m^n(N_{sm} - N'_{sm-1})] = 0\) for any \(m\) if \(n > \max j n_j\).

\[\square\]

**Proof of Theorem 2.1.** Since the parameter space \(\Lambda\) is open, there exists \(\epsilon > 0\) such that \(O(\epsilon, \sigma_*) = \{\sigma; |\sigma - \sigma_*| < \epsilon\} \subset \Lambda\). Then we have

\[- \partial_\sigma H_n(\sigma, \hat{\nu}_n) = \int_0^1 \partial_\sigma^2 H_n(\sigma, \hat{\nu}_n)(\sigma + t(\hat{\sigma}_n - \sigma_\star))(\hat{\sigma}_n - \sigma_\star)dt\]

for \(\hat{\sigma}_n \in O(\epsilon, \sigma_\star)\), by \(\partial_\sigma H_n(\hat{\sigma}_n, \hat{\nu}_n) = 0\).

Hence we obtain \(b_{1/n}^{1/4}(\hat{\sigma}_n - \sigma_\star) = \tilde{\Gamma}_{1,n} b_{1/n}^{1/4} \partial_\sigma H_n(\sigma, \hat{\nu}_n)\) on \(\{\det \tilde{\Gamma}_{1,n} \neq 0 \text{ and } \hat{\sigma}_n \in O(\epsilon, \sigma_\star)\}\), where \(\tilde{\Gamma}_{1,n} = -b_{1/n}^{1/2} \int_0^1 \partial_\sigma^2 H_n(\sigma, \hat{\nu}_n + t(\hat{\sigma}_n - \sigma_\star)) dt\). Then since Propositions 2.1, 6.1 and 7.1 yield \(P[\det \tilde{\Gamma}_{1,n} = 0] \to 0\), \(P[\hat{\sigma}_n \not\in O(\epsilon, \sigma_\star)] \to 0\) and \(\tilde{\Gamma}_{1,n}^{-1} \to P^\star \Gamma_1^{-1/2}\) as \(n \to \infty\) by Proposition 7.2.

Moreover, Propositions 2.1 and 7.1 ensure that \(\Gamma_1^{-1} \to \Gamma_1\), which completes the proof.

\[\square\]

**8 Proof of the LAN property**

To obtain the LAN property of our model, the arguments in the proof of Theorem 2.1 are essential. Indeed, by using Propositions 2.1 and 7.2, we obtain a LAN-type property of the quasi-log-likelihood function \(H_n\) with respect to \(\sigma\): \(H_n(\sigma + b_{1/n}^{1/4} u_1, v_1) - H_n(\sigma, v_1) - u_1 b_{1/n}^{1/2} \partial_\sigma H_n(\sigma, v_1) - u_1 b_{1/n}^{1/2} \partial_\nu H_n(\sigma, v_1) u_1/2 \to 0\) as \(n \to \infty\) for any \(u_1 \in \mathbb{R}^d\), and \(b_{1/n}^{1/4} \partial_\sigma H_n(\sigma, v_1), -b_{1/n}^{1/2} \partial_\nu^2 H_n(\sigma, v_1) \to \sigma^\nu (\Gamma_1^{-1/2} N, \Gamma_1)\), where \(N\) is a \(d\)-dimensional standard normal random variable independent of \(\mathcal{F}\). On the other hand, under the assumptions of Theorem 2.2, the true log-likelihood ratio \(\log(dP_{\sigma + b_{1/n}^{1/4} u_1, v_1}^n / dP_{\sigma, v_1}^n)\) for \(u_1 \in \mathbb{R}^d\) and \(u_2 \in \mathbb{R}^2\) is obtained as \(-\left(Z_1^\top S_1^{-1} Z_1 + \log \det S_1\right)/2\) if we set \(k_n = b_n\). We cannot apply the argument of Section 5 to this quantity because the estimate \(\ell_n \to \infty\) is essential there. Therefore, we follow the approaches by Gloter and Jacob [12] to show the LAN property. We set a ‘subexperiment’ and a ‘superexperiment’, which are obtained by respectively removing and adding observations from the original experiment. The likelihood functions of these experiments have similar properties to \(H_n\), and therefore we can prove the LAN properties for these experiments with the same limit distribution. We can prove that these results lead us to the LAN property of the original one.

Let \(Z = (\mathbb{R}^8)^{Z_+}, \pi_i(z) = (x_i^{jk}, t_i^k, e_i^k)_{j,k=1,2}\) for \(i \in \mathbb{Z}_+\) and \(z = (x_i^{jk}, t_i^k, e_i^k)_{j,k=1,2} \in \mathbb{Z}_+\). Let \(\mathcal{H} = \mathfrak{P}(\pi_i^{-1}(A); i \in \mathbb{Z}_+, A \in \mathcal{B}(\mathbb{R}^8))\), \(P_{\sigma_1, v_1}\) be the induced probability measure on \((Z, \mathcal{H})\) by
$$(Y_{S_{i,k}^m}^j1_{\{i \leq j,k,n\}}, S_{i,k}^n 1_{\{i \leq j,k,n\}}, \epsilon_{i,k,1}^n 1_{\{i \leq j,k,n\}})_{i \in \mathbb{Z}^+}, j, k = 1,2$$ with a true value $(\sigma^*, \nu^*)$. We can ignore the event $\min_{j,m} k_{j,m}^0 \leq 0$.

Let $\mathcal{H}' = \mathcal{B}(t_{j,m}^0; i \in \mathbb{Z}^+, k = 1,2), \mathcal{H}' = \mathcal{B}(t_{j,m}^0; i \in \mathbb{Z}^+, k = 1,2) \setminus \mathcal{H}'$, $\mathcal{H}^{n,0} = \mathcal{B}(x_{j,m}^{1,0}, x_{j,m}^{1,2})$, $1 \leq m \leq \ell_n - 1, j = 1,2) \setminus \mathcal{H}'$, $\mathcal{H}^{n,2} = \mathcal{B}(x_{j,m}^{1,1}, x_{j,m}^{1,2})$.

Then we can see $\mathcal{H}^{n,0} \subset \mathcal{H}^{n,1} \subset \mathcal{H}^{n,2}$ and

$$\log(dP_{\sigma, \nu_1}/dP_{\sigma, \nu_2}) = \log \left( \frac{dP'_{\sigma, \nu_1} | \mathcal{H}^{n,1}}{dP'_{\sigma, \nu_2} | \mathcal{H}^{n,1}} \right).$$

Moreover, we obtain

$$\log(dP'_{\sigma, \nu_1} | \mathcal{H}^{n,1})(Y_{S_{i,k}^m}^j1_{\{i \leq j,k,n\}}, S_{i,k}^n 1_{\{i \leq j,k,n\}}, \epsilon_{i,k,1}^n 1_{\{i \leq j,k,n\}})_{i \in \mathbb{Z}^+, j, k = 1,2} = H_n^{(l)}(\sigma_u, \nu_u) - H_n^{(l)}(\sigma_* \nu_*)$$

for $l = 0$, where $\mu(t, \sigma) = (\mu^1(t, \sigma), \mu^2(t, \sigma))^T$, $Z_m^{(0)}(\sigma) = (Y_k(I_{K^m_k}) - \int_{I_{K^m_k}} \mu^k(t, \sigma) dt)_i, k$, and

$$S_m^{(0)}(\sigma, \nu) = \begin{pmatrix} \text{diag}(\{f_{j,m}^1, M_{j,m}^1\} & \{f_{j,m}^2, M_{j,m}^2\} & \{f_{j,m}^3, M_{j,m}^3\}) \\ \{f_{j,m}^4, M_{j,m}^4\} & \{f_{j,m}^5, M_{j,m}^5\} & \{f_{j,m}^6, M_{j,m}^6\} \end{pmatrix}$$

for $1 \leq m \leq \ell_n$, $H_n^{(l)}(\sigma, \nu) = -\sum_{m=1}^{\ell_n} \{(Z_m^{(0)}(\sigma))^T (S_m^{(0)}(\sigma, \nu)Z_m^{(0)}(\sigma)) + \log \det S_m^{(0)}(\sigma, \nu)\} / 2$, $\sigma_u = \sigma_* + b_1^{1/4} u_1$ and $\nu_u = \nu_* + b_1^{1/4} u_2$ for $u = (u_1, u_2) \in \mathbb{R}^d \times \mathbb{R}^2$. $\mathcal{H}^{n,0}$ and $\mathcal{H}^{n,2}$ are $\sigma$-fields for ‘subexperiment’ and ‘superexperiment’, respectively, while $\mathcal{H}^{n,1}$ is the one for the original one.

To obtain similar formula to (8.2) for $l = 2$, let $R_m = S_{K^m_k}^{n,1} \vee S_{K^m_k}^{n,2}$, $Q_{m}^k = S_{K^m_k}^{n,k}$, $Q_{m,k}^f = S_{K^m_k}^{n,k}$, $Q_{m,k}^i | \mathcal{H}_{m,k}^i = S_{K^m_k}^{n,k}$, $Q_{m,k}^i | \mathcal{H}_{m,k}^i = 1$, and

$$Y_{m,k}^i(\sigma) = \int_{I_{K^m_k}} \mu^k(t, \sigma) dt,$$

$$Z_m^{(2)}(\sigma) = \{(Y_{m,k}^i(\sigma))_{i \leq k^m_k, \nu = 1} \} (Y_{m,k}^{i+1}(\sigma) - \int_{I_{K^m_k}} \mu^k(t, \sigma) dt)_{i \leq k^m_k, \nu = 1}$$

for $2 \leq m \leq \ell_n - 1$. Then (8.2) holds for $l = 2$, where $I_{K^m_k} = [R_{m-1} \setminus Q_m^{k_m}, I_{m-1} \setminus I_{m-1}]$ for $2 \leq i \leq k^m_k + 1$,

$$E(v) = \nu_{m,k}^{(2)}(k^m_k) = h_m + k^m_k + 1, \quad E(v) = \nu_{m,k}^{(2)}(k^m_k) = h_m + k^m_k + 1, \quad E(v) = \nu_{m,k}^{(2)}(k^m_k) = h_m + k^m_k + 1$$

if $Q_m^{k_m} \leq R_m$, and

$$(M_{m,k}^{(2)}(\nu^l \nu^l)^T = 0 \delta_{\nu^l \nu^l} = 1 \} + \nu_{i=1}^{(2)} = 0 \}$$

if $Q_m^{k_m} \leq Q_m^2$.
for $2 \leq m \leq \ell_n - 1$, $Z_m^{(2)}$, $M_m^{(2)}$, and $S_m^{(2)}$ are similarly defined for $m = 1, \ell_n$, and
\[
H_m^{(2)}(\sigma, \nu) = -\frac{1}{2} \sum_{m=1}^{\ell_n} \left\{ (Z_m^{(2)}(\sigma))^\top S_m^{(2)}(\sigma, \nu)^{-1} Z_m^{(2)}(\sigma) + \log \det S_m^{(2)}(\sigma, \nu) \right\}.
\]

The log-likelihood functions $H_n^{(0)}$ and $H_n^{(2)}$ of ‘subexperiment’ and ‘superexperiment’, respectively, have similar properties to that of $H_n$, and hence we can prove convergence of likelihood ratios. Gloter and Jacod [12] showed that convergence of likelihood ratios of ‘subexperiment’ and ‘superexperiment’ imply convergence of that of the original experiment (Theorem 4.1). Here, we use a slight extension of their result. The proof is straightforward. Let $U_{2, v}^{0, l} = d(P_{0, \nu}^{l}, |H_n^{(0)}|)/d(P_{0, v, \nu}^{l}, |H_n^{(0)}|), K \in \mathbb{N}$, and $\{\sigma_n^k\}_{n \in \mathbb{N}, 1 \leq k \leq K} \subset \Lambda$ and $\{\nu_n^k\}_{n \in \mathbb{N}, 1 \leq k \leq K} \subset (0, \infty) \times (0, \infty)$ be arbitrary sequences.

**Theorem 8.1.** Suppose that $(U_{n, v}^{0, l}, \cdots, U_{n, v}^{0, K})$ converges in law under $P_{n, v}^{l}$ to a limit $Y = (Y^{1}, \cdots, Y^{K})$ with $0 < Y < \infty$ a.s. and $E[Y^{k}] = 1$ for $l = 0, 2$ and $1 \leq k \leq K$. Then the same convergence holds for $l = 1$.

We will prove the LAN properties of ‘subexperiment’ and ‘superexperiment’. Then Theorem 8.1 leads to the LAN property of the original one. Taylor’s formula yields
\[
H_n^{(l)}(\sigma_0, \nu_0) - H_n^{(l)}(\sigma, \nu) = b_n^{1/2} \sum_{i,j} \sigma_i \sigma_j H_n^{(l)}(\sigma_{uj}, \nu_u) u_{ij} + b_n^{1/2} \sum_{i,j} \nu_i \nu_j H_n^{(l)}(\sigma_u, \nu_{uj}) u_{ij} + \mathcal{O}(b_n^{-3/2})
\]

We examine the limit of each term on the right-hand side. First, we prepare an auxiliary lemma. Let $\mathcal{Q}_n = \{\alpha = (\alpha_1, \cdots, \alpha_L); L \in \mathbb{N}, 1 \leq \alpha_1 \leq q, \sum_{l=1}^L \alpha_l = q\}$, $M_m^{(0)} = \text{diag}(v_1, \cdots, v_{m-1}, \cdots, v_{m^2}), M_m^{(2)} = \text{diag}(v_1, \cdots, v_{m^2}, v_{m^2+1}, \cdots, v_{m^2+2m-1})$, $E(v_u)$,
\[
Y_{2, m+}^k = \begin{cases} \epsilon_{K_m}^{(n, k)} & \text{if } Q_n^k > Q_m^{3-k} \\ \epsilon_{K_m}^{(n, k)} & \text{if } Q_n^k < Q_m^{3-k} \end{cases}
\]
\[
Y_{2, m+}^k = \begin{cases} (\epsilon_{K_m}^{(n, k)}, -\epsilon_{K_m}^{(n, k)})^\top & \text{if } Q_n^k > Q_m^{3-k} \\ (\epsilon_{K_m}^{(n, k)}, \epsilon_{K_m}^{(n, k)})^\top & \text{if } Q_n^k < Q_m^{3-k} \end{cases}
\]
for $2 \leq m \leq \ell_n - 1$. Though we can similarly define $Z_m^{(0)}$ and $Z_m^{(2)}$ for $m = 1, \ell_n$ so that they satisfy the following lemma, we omit it to avoid redundancy. We also omit discussions for $m = 1, \ell_n$ throughout the rest of this section.

**Lemma 8.1.** Assume (A1'). Let $l \in \mathbb{N}, 0 < l < 1 \leq m \leq \ell_n$, and $S_m$ be a $\mathcal{G}_{m-1}$-measurable random matrix with suitable size. Then there exists a positive constant $C_p$ depending only on $p$ such that
(i) $E_m[[Z_m^{(2)}] \top S_m Z_m^{(2)}] \leq C_p \sum_{(\alpha_1, \cdots, \alpha_L) \in \mathcal{Q}_n} \prod_{l=1}^L \text{tr}((S_m M_m^{(l)})^{\alpha_l})$.
(ii) $E_m[[Z_m^{(2)}] \top S_m Z_m^{(2)}] \leq 12 \text{tr}((M_m^{(2)})^2) + 48 \text{tr}((M_m^{(2)})^4)$.

**Proof.** (i) Let $U_m^{(l)}$ be an orthogonal matrix and $\Lambda_m^{(l)}$ be a diagonal matrix satisfying $U_m^{(l)} M_m^{(l)} (U_m^{(l)})^\top = \Lambda_m^{(l)}$. Then since $Z_m^{(2)} \sim N(0, M_m^{(l)})$, a similar estimate to (A.5) yields
\[
E_m[[Z_m^{(2)}] \top S_m Z_m^{(2)}] = \sum_{i_1, \cdots, i_{2p}} \prod_{q=1}^{p} (U_m^{(l)} S_m U_m^{(l)})_{i_{2q-1}i_{2q}} \sum_{(l_{2q'-1}, l_{2q'})} \prod_{q'=1}^{p} (\Lambda_m^{(l)})_{l_{2q'-1}l_{2q'}}
\]
where the second summation in the right-hand side is taken over all disjoint pairs $(l_{2q'-1}, l_{2q'})$ in the variables $i_1, \cdots, i_{2p}$.
\[ E_m[(\text{tr}(M_{m, n}^{(l)} \cdot \tilde{S}_m) - (Z_{2, m}^{(l)})^\top \tilde{S}_m Z_{2, m}^{(l)})^2] = \frac{4}{r_0^*(4 - r)!} (\text{tr}(M_{m, n}^{(l)} \cdot \tilde{S}_m))^{r_0} E_m[[(Z_{2, m}^{(l)})^\top \tilde{S}_m Z_{2, m}^{(l)})^{4 - r}] \]

\[ = \text{tr}(M_{m, n}^{(l)} \cdot \tilde{S}_m)^4 - 4\text{tr}(M_{m, n}^{(l)} \cdot \tilde{S}_m)^2 + 6\text{tr}(M_{m, n}^{(l)} \cdot \tilde{S}_m)^2 + 2\text{tr}((M_{m, n}^{(l)} \cdot \tilde{S}_m)^2) \]

\[ - 4\text{tr}(M_{m, n}^{(l)} \cdot \tilde{S}_m)(\text{tr}(M_{m, n}^{(l)} \cdot \tilde{S}_m)^3 + 6\text{tr}(M_{m, n}^{(l)} \cdot \tilde{S}_m)^2 + 8\text{tr}((M_{m, n}^{(l)} \cdot \tilde{S}_m)^3) + \{ \text{tr}(M_{m, n}^{(l)} \cdot \tilde{S}_m)^4 \]

\[ + 12\text{tr}(M_{m, n}^{(l)} \cdot \tilde{S}_m)^2)(\text{tr}(M_{m, n}^{(l)} \cdot \tilde{S}_m)^2 + 12\text{tr}(M_{m, n}^{(l)} \cdot \tilde{S}_m)^2)^2 + 32\text{tr}(M_{m, n}^{(l)} \cdot \tilde{S}_m)(\text{tr}(M_{m, n}^{(l)} \cdot \tilde{S}_m)^3) + 48\text{tr}((M_{m, n}^{(l)} \cdot \tilde{S}_m)^4) \]

\[ = 12\text{tr}(M_{m, n}^{(l)} \cdot \tilde{S}_m)^2 + 48\text{tr}((M_{m, n}^{(l)} \cdot \tilde{S}_m)^4). \]

**Lemma 8.2.** Assume (A1′) and (A2). Then

1. \( \sup_{\sigma} |b_n^{-1/2} \partial_\sigma^2 \mathcal{H}_n^{(l)}(\sigma, \nu^*) - \partial_\sigma^2 \mathcal{Y}_n(\sigma)| \to 0 \pmb{p} 0, \)
2. \( \sup_{\sigma, \nu} |b_n^{-1/2} \partial_\sigma \partial_\nu H_n^{(l)}(\sigma, \nu) - \partial_\sigma \partial_\nu \mathcal{Y}_n(\sigma)| \to 0 \pmb{p} 0, \)
3. \( \sup_{\sigma, \nu} |b_n^{-3/4} \partial_\sigma \partial_\nu H_n^{(l)}(\sigma, \nu)| \to 0 \pmb{p} 0 \)

as \( n \to \infty \) for \( 0 \leq k \leq 3 \) and \( l = 0, 2. \)

**Proof.** 1. We obtain the desired results by a similar argument to the proof of Proposition 2.1 together with Lemmas 4.1 and 8.1, the results in Section 8 of [12], and similar estimates to Lemmas 5.1, 4.2 and 4.3. For any \( \epsilon > 0, \) \((\epsilon \mathcal{E} + M_{j, m}^{(2)})^{-1}\) has a similar decomposition to (4.4) by replacing \( p_1, \ldots, p_{k_m-1}, p_{k_m} \) by \( p_1', \ldots, p_{k_m-1}', p_{k_m}' \). Therefore, estimate for the quantity corresponding to \( \Lambda_1 \) in Lemma 5.2 is obtained since

\[ ((\epsilon \mathcal{E} + M_{j, m}^{(2)})^{-1})_{11} = \frac{\prod_{k=1}^{k_m} q_1^{k_m-1} q_1}{(p_1'^{k_m}(\epsilon) - 1) \prod_{k=1}^{k_m} q_1^{k_m-1} q_1} = O(p_1^{k_m/2}). \]

2. Similarly to (5.1), we obtain

\[ b_n^{-1} \partial_\nu H_n^{(l)}(\sigma, \nu) = -\frac{1}{2} b_n^{-1} \sum_{m} \{ \text{tr}(\partial_\nu ((S_{m}^{(l)})^{-1} S_{m, s}) + \partial_\nu \log \det S_{m}^{(l)} \} + o_p(1). \]

Let \( \tilde{D}_{1, m}^{(l)} = (\tilde{D}_{1, m}^{(l)}, \tilde{D}_{2, m}^{(l)}) \) for \( l = 0, 2, k_m^{(0), 0} = k_m^{(1)} \) for \( j = 1, 2, \) \( \tilde{D}_{1, m}^{(0)} = ((S_{m}^{(l)})_{i, j})_{1 \leq i, j \leq k_m^{(0), 1}}, \)

\[ \tilde{D}_{2, m}^{(0)} = \text{diag}(((S_{m}^{(2)})_{i, j})_{1 \leq i, j \leq k_m^{(2), 2}} \tilde{D}_{2, m}^{(2)} = \text{diag}(((S_{m}^{(2)})_{i, j})_{1 \leq i, j \leq k_m^{(2), 2}} \mathcal{E}(\nu)), \]

\[ \tilde{G}_{m}^{(0)} = (\tilde{D}_{1, m}^{(0)})^{-1/2} \left\{ \int_{t_{m, r}^{(2, 0)}, \tilde{D}_{2, m}^{(2)}} b_1 \cdot b_2 dt \right\} 1 \leq k_m^{(0), 1}, 1 \leq k_m^{(0), 2}, \]

\[ \tilde{G}_{m}^{(2)} = (\tilde{D}_{1, m}^{(2)})^{-1/2} \left\{ \int_{t_{m, r}^{(2, 0)}, \tilde{D}_{2, m}^{(2)}} b_1 \cdot b_2 dt 1 \leq k_m^{(2), 2}, \right\} _{1 \leq k_m^{(2), 1}, 1 \leq k_m^{(2), 2}} (\tilde{D}_{2, m}^{(2)})^{-1/2}, \]

where \( k_m^{(2), 2} \) is the size of \( \tilde{D}_{2, m}^{(2)}. \) Then we obtain

\[ \text{tr}((S_{m}^{(l)})^{-1} S_{m, s}) = \text{tr}(\tilde{D}_{m}^{(l)})^{-1/2} \left\{ \mathcal{E} \left( \tilde{G}_{m}^{(l)} \right)^\top \tilde{G}_{m}^{(l)} \right\}^{-1/2} (\tilde{D}_{m, s})^{1/2} \left\{ \mathcal{E} \left( \tilde{G}_{m, s} \right)^\top \tilde{G}_{m, s} \right\}^{1/2} (\tilde{D}_{m, s})^{1/2} \]

\[ = \sum_{p=0}^{\infty} \left\{ \text{tr}(\tilde{D}_{1, m}^{(l)})^{-1/2} \left( \tilde{G}_{m}^{(l)} \left( \tilde{G}_{m}^{(l)} \right)^\top \right)^p \left( \tilde{D}_{1, m, s}^{(l)} \right)^{-1/2} \tilde{D}_{1, m, s} \right\} \}

\[ - \left( \tilde{D}_{1, m}^{(l)} \right)^{-1/2} \left( \tilde{G}_{m}^{(l)} \left( \tilde{G}_{m}^{(l)} \right)^\top \right)^p \left( \tilde{D}_{2, m, s}^{(l)} \right)^{-1/2} \left( \tilde{D}_{m, s}^{(l)} \right)^{1/2} \left( \tilde{D}_{m, s}^{(l)} \right)^{1/2} \]

\[ + \text{tr}(\tilde{D}_{2, m}^{(l)})^{-1/2} \left( \tilde{D}_{2, m}^{(l)} \right)^{-1/2} \left( \tilde{D}_{m, s}^{(l)} \right)^{1/2} \left( \tilde{D}_{m, s}^{(l)} \right)^{1/2} \

\[ - \left( \tilde{D}_{2, m}^{(l)} \right)^{-1/2} \left( \tilde{G}_{m}^{(l)} \left( \tilde{G}_{m}^{(l)} \right)^\top \right)^p \left( \tilde{D}_{1, m, s}^{(l)} \right)^{-1/2} \left( \tilde{D}_{1, m, s}^{(l)} \right)^{1/2} \left( \tilde{D}_{1, m, s}^{(l)} \right)^{1/2}. \]
Since $\| (\hat{D}_{j,m})^{-1} \hat{D}_{j,m,*} \| = O_p(1)$, terms involving $\hat{G}_{m}^{(l)}$ in the right-hand side are $O_p(b_n^{1/2} \epsilon_n^{-1})$. Therefore we have

$$b_n^{-1} \text{tr}((S_m^{(l)})^{-1} S_m^{(m,*)}) = b_n^{-1} \sum_{j=1}^{2} \text{tr}((\hat{D}_{j,m})^{-1} \hat{D}_{j,m,*}) + o_p(\epsilon_n^{-1})$$

$$= b_n^{-1} \sum_{j=1}^{2} v_j \text{tr}(\epsilon_{k,j} - (\hat{D}_{j,m} - v_j v_j^{-1} \hat{D}_{j,m,*})) + o_p(\epsilon_n^{-1}) = T \epsilon_n^{-1} \sum_{j=1}^{2} \tilde{a}_m^n v_j v_j^{-1} + o_p(\epsilon_n^{-1}).$$

Similarly we have $b_n^{-1} \text{tr}(\partial_v (S_m^{(l)})^{-1} S_m^{(m,*)}) = T \epsilon_n^{-1} \sum_{j=1}^{2} \tilde{a}_m^n v_j v_j^{-1} + o_p(\epsilon_n^{-1})$. Moreover, we obtain

$$b_n^{-1} \partial_v \log \frac{\det S_m^{(l)}}{\det S_m^{(m,*)}} = 2 b_n^{-1} \partial_v \log \det((\hat{D}_{j,m}^{-1})^{1/2} \hat{E}_m^{(l)} (\hat{D}_{j,m}^{-1})^{1/2}) + b_n^{-1} \partial_v \log \det(\mathcal{E} - \hat{G}_{m}^{(l)} (\hat{G}_{m}^{(l)})^\top) - b_n^{-1} \partial_v \log \det(\mathcal{E} - \hat{G}_{m,*}^{(l)} (\hat{G}_{m,*}^{(l)})^\top)$$

$$= b_n^{-1} \sum_{j=1}^{2} \text{tr}(\epsilon_{k,j} + (\hat{D}_{j,m}^{-1})^{1/2} \hat{E}_m^{(l)} (\hat{D}_{j,m}^{-1})^{1/2}) + o_p(\epsilon_n^{-1})$$

$$= T \epsilon_n^{-1} \sum_{j=1}^{2} \tilde{a}_m^n \partial_v \log(v_j v_j^{-1}) + o_p(\epsilon_n^{-1}).$$

3. Since $\partial_v \log \det S_m^{(l)} = \text{tr}(\partial_v (S_m^{(l)})^{-1})$ and

$$b_n^{-3/4} \partial_v \partial_v H_n^{(l)}(\sigma, v) = -\frac{1}{2} b_n^{-3/4} \sum_m \{ \text{tr}(\partial_v \partial_v H_n^{(l)}(S_m^{(l)})) + \partial_v \partial_v \log \det(\hat{S}_m^{(l)}) \} + o_p(1),$$

we have $b_n^{-3/4} \partial_v \partial_v H_n^{(l)}(\sigma, v) = o_p(1)$. Sobolev’s inequality and similar estimates for $\partial_v \partial_v^2$ and $\partial_v^2 \partial_v$ yield the results. \qed

The following lemma completes the proof of the LAN properties of ‘subexperiment’ and ‘superexperiment’.

**Lemma 8.3.** Assume (A1’), (A2) and (A2'). Then $(b_n^{-1/4} \partial_v H_n^{(l)}(\sigma, v))_{\rightarrow \mathcal{N}} \sim (\mathbf{1}^{\frac{1/2}{4}}, \mathbf{1}^{\frac{1/2}{2}})_{\rightarrow \mathcal{N}}$ for $l = 0, 2$, where $\mathcal{N}$ is a $(d + 2)$-dimensional normal random variable independent of $\mathcal{F}$.\]

**Proof.** Let $Z_m^{(l)} = Z_m^{(l)}(\sigma, v), \tilde{X}_m^{(l)} = b_n^{-1/4} E_m[(Z_m^{(l)})(S_m^{(l)})^{-1} Z_m^{(l)}]/2, \tilde{X}_m^{(2)} = b_n^{-1/4} E_m[(Z_m^{(l)})^2(S_m^{(l)})^{-1} Z_m^{(l)}]/2, \tilde{X}_m = ((\tilde{X}_m^{(1)})_{\top}, (\tilde{X}_m^{(2)})_{\top})_{\top}$, then we have

$$\sum_{m=1}^{[\epsilon_n^{1/2}]} E_m[|\tilde{X}_m|^{2}] \leq C \epsilon_n^{1/2} b_n^{-1} \sum_m E_m[|Z_m^{(l)}(S_m^{(l)})^{-1} Z_m^{(l)}|^{4}].$$

We can see that the first term on the right-hand side converges to 0 by Lemma 8.1 and a similar estimate to Point 1 of Lemma 4.3. We show the second term on the right-hand side converges to 0 in probability.

Let $Z_2^{(l)}$ be similarly defined to $Z_2^{(l)}$ in (A4). Let $Z_1^{(l)} = Z_m^{(l)}(\sigma) - Z_2^{(l)}$ and $S_1^{(l)} = E_m[Z_1^{(l)}(S_1^{(l)})^\top]$. Then Lemmas 8.1 and a similar estimate to (A5) yield

$$E_m[(Z_1^{(l,*)})^\top \partial_v (S_1^{(l,*)})^{-1} Z_1^{(l,*)}] = 3E_m[|Z_1^{(l,*)})^\top \partial_v (S_1^{(l,*)})^{-1} S_1^{(l,*)} \partial_v (S_1^{(l,*)})^{-1} Z_1^{(l,*)}|^2] = R_n(b_n^{1/2} \epsilon_n).$$
and $E_m[(Z_{1,m,s}^{(l)})^\top \partial_v (S_{m,s}^{(l)})^{-1} Z_{1,m,s}^{(l)}]^4] = \hat{R}_n(b_n^{-4}) \sum_{j=1}^4 |\text{tr}((S_{m,s}^{(l)})^{-1})^j| = \hat{R}_n(b_n^{-2} k_n^4)$. Moreover, Lemma 8.1 yields

$$b_n^{-2} \sum_m E_m[\hat{E}_m[(Z_{2,m}^{(l)})^\top \partial_v (S_{m,s}^{(l)})^{-1} Z_{2,m}^{(l)}]^4] = \hat{R}_n(b_n^{-2} \ell_n k_n^2) \to^p 0,$$

and therefore, we obtain

$$\sum_{m=1}^{[\ell_n/T]} E_m[\hat{\mathcal{X}}_m^n(\mathcal{X}_m^n)^\top] \to^p 0.$$

Similarly to the proof of Proposition 7.2, we have

$$\sum_{m=1}^{[\ell_n/T]} E_m[\hat{\mathcal{X}}_m^n(N_{s_m} - N_{s_{m-1}})] \to^p 0 \quad \text{and} \quad \sum_{m=1}^{[\ell_n/T]} E_m[\hat{\mathcal{X}}_m^n(W_{s_m} - W_{s_{m-1}})] = 0$$

for any bounded $\mathcal{G}_t$-martingale $N$ orthogonal to $(W_t)_t$.

Therefore, by Theorem 3.2 in Jacod [18], it is sufficient to show that

$$\sum_{m=1}^{[\ell_n/T]} E_m[\hat{\mathcal{X}}_m^n] \to^p \text{diag}(-\partial^2_{\sigma^2} \mathcal{Y}_1(\sigma, t), -\partial^2_{\sigma^2} \mathcal{Y}_2(v_j, t)),$$

where $\mathcal{Y}_2(v, t) = -\int_0^t \sum_{j=1}^2 a_j^2 (\{v_j, s\} - 1 + \log(v_j/v_j)) ds/2$.

Then we obtain the desired results by

$$\sum_{m=1}^{[\ell_n/T]} E_m[\hat{\mathcal{X}}_m^n]^\top] = \frac{b_n^{-3/4}}{2} \sum_{m=1}^{[\ell_n/T]} \text{tr}((\partial_v S_{m,s}^{(l)}(S_{m,s}^{(l)})^{-1})^2) \to^p 0,$$

and similarly

$$\sum_{m=1}^{[\ell_n/T]} E_m[\hat{\mathcal{X}}_m^n]^\top] = \frac{b_n^{-1}}{2} \sum_{m=1}^{[\ell_n/T]} E_m[\partial^2_{v} \{ (\partial_v S_{m,s}^{(l)}(S_{m,s}^{(l)})^{-1})^2 Z_{m,s}^{(l)} + \log det S_{m,s}^{(l)} \}]_{(\sigma, v)=(\sigma, v)} \to^p 0.$$

Proof of Theorem 2.2. Let $U(u) = \exp(u^\top \Gamma^{1/2} \hat{\mathcal{N}} - u^\top \Gamma u/2)$ for $u \in \mathbb{R}^{d+2}$. Let

$$Z^{(1)}(\sigma) = \left( \begin{array}{c} n_k^1, \left( \hat{Y}_i - \hat{Y}_{i-1} - \int_{S_{i-1}^{n_k}}^{S_i^{n_k}} \mu^k(t, \sigma) dt \right) \end{array} \right)_{i=1}^{k=2}.$$

and $S^{(1)}(\sigma, v)$ be a symmetric matrix of size $J_{1,n} + J_{2,n} + 2$ defined by $(S^{(1)}(\sigma, v))_{11} = v_1$, $(S^{(1)}(\sigma, v))_{J_{1,n}+2, J_{1,n}+2} = v_2$,

$$(S^{(1)}(\sigma, v))_{ij} = (v_1 M(J_{1,n} + 1), v_2 M(J_{2,n} + 1))_{ij}$$

if $i \neq j$ and $\{i, j\} \cap \{1, J_{1,n} + 2\} \neq \emptyset$,

$$(S^{(1)}(\sigma, v))_{ij} = \int_{S_{i-1}^{n_k}}^{S_i^{n_k}} b_i^2 dt \delta_{ij} + v_1 M(J_{1,n} + 1))_{ij}$$

if $2 \leq i, j \leq J_{1,n} + 1$,$$

$$(S^{(1)}(\sigma, v))_{ij} = \int_{S_{i-2}^{n_k}}^{S_{i-1}^{n_k}} b_i^2 dt \delta_{ij} + v_2 M(J_{2,n} + 1))_{ij}$$

if $2 \leq i \leq J_{2,n} + 1$,$$

$$(S^{(1)}(\sigma, v))_{ij} = \int_{S_{i-2}^{n_k}}^{S_{i-1}^{n_k}} b_i^2 dt \delta_{ij}$$

if $2 \leq i \leq J_{1,n} + 1$ and $2 \leq j \leq J_{2,n} + 1$. 

32
where \( i' = i - J_{1,n} - 1 \) and \( j' = j - J_{1,n} - 1 \). Then we have (8.2) for \( l = 1 \) with

\[
H_n^{(l)}(\sigma, v) = -\frac{1}{2} (Z^{(l)}(\sigma))^\top (S^{(l)}(\sigma))^{-1} Z^{(l)}(\sigma) - \frac{1}{2} \log \det S^{(l)}(\sigma, v).
\]

Moreover, Theorem 8.1 and Lemmas 8.2 and 8.3 yield

\[
(H_n^{(1)}(\sigma_{u(1)}, v_{u(1)})) - H_n^{(1)}(\sigma_u, v_u), \ldots, H_n^{(1)}(\sigma_{u(k)}, v_{u(k)})) - H_n^{(1)}(\sigma_u, v_u)) \to_d \log(U(u^{(1)}), \ldots, \log(U(u^{(k)})) \quad \text{(8.4)}
\]

as \( n \to \infty \) for \( u^{(1)}, \ldots, u^{(k)} \in \mathbb{R}^{d+2} \).

Furthermore, similar estimates to the proof of Lemma 8.2 yield \( \sup_{\sigma, v} |b_n^{-3/4} \partial_{\sigma} \partial_v H_n^{(1)}(\sigma, v)| \to 0 \), \( \sup_{\sigma} |b_n^{-3/4} \partial_{\sigma}^2 H_n^{(1)}(\sigma, v_u)| \to 0 \) and \( \sup_{\sigma} |b_n^{-3/4} \partial_v^2 H_n^{(1)}(\sigma, v)| \to 0 \). Therefore we obtain

\[
H_n^{(1)}(\sigma_u, v_u) = (u \cdot V_{1,n} - u^\top V_{2,n} u/2) \to 0 \quad \text{(8.5)}
\]

as \( n \to \infty \) for any \( u \in \mathbb{R}^{d+2} \), where \( V_{1,n} = (b_n^{-1/4} \partial_{\sigma} H_n^{(1)}(\sigma_u, v_u)), b_n^{-1/2} \partial_v H_n^{(1)}(\sigma_u, v_u)) \) and \( V_{2,n} = -\text{diag}(b_n^{-1/2} \partial_{\sigma} H_n^{(1)}(\sigma_u, v_u), b_n^{-1/2} \partial_v H_n^{(1)}(\sigma_u, v_u)) \).

(8.4) and (8.5) yield \( V_{1,n} \to_d \Gamma^{1/2} \hat{N} \) and \( V_{2,n} \to_d \Gamma \), and therefore we obtain the LAN property of the original experiment with \( \Gamma_n = V_{2,n} \) and \( N_n = \Gamma_n^{1/2} V_{1,n} \).

\[\square\]

9 Proof of the results in Section 2.4

In this final section, we complete the proof of remaining results in Section 2. Proposition 2.2 is proven by the scheme of Yoshida [29, 30]. Proposition 6.1 and moment estimates in Lemmas 4.4 and 5.2 enable us to check the assumptions of Theorem 2 in [30]. Then the results on convergence of moments and the Bayes-type estimator are obtained by Proposition 2.2.

Outline of the proof of Proposition 2.2. We apply Theorem 2 in Yoshida [30]. It is sufficient to prove the following five conditions for any \( L > 0 \) with some positive constant \( \delta_1 \) and \( \delta_2 \):

1. There exists \( C_L > 0 \) such that \( P[\inf_{\sigma, \sigma_n}(\|Y_1(\sigma)|/|\sigma - \sigma_n|^2) \leq r^{-1}] \leq C_L / r^L \) and \( P[|r^{-1}|u|^2 \leq u^\top \Gamma_n u/4] \leq C_L / r^L \) for any \( u \in \mathbb{R}^d \).

2. \( \sup_n E[(b_n^{-1/4} \partial_{\sigma} H_n(\sigma_u, \hat{v}_n))]^2 \leq C \).

3. \( \sup_n E[(b_n^{-1/2} \partial_{\sigma} H_n(\sigma_u, \hat{v}_n) - H_n(\sigma_u, \hat{v}_n) - Y_1(\sigma))]^2 \leq C \).

4. \( \sup_n E[(b_n^{-1/2} \partial_{\sigma} H_n(\sigma_u, \hat{v}_n))]^2 \leq C \).

5. \( \sup_n E[(b_n^{-1/2} \partial_{\sigma} H_n(\sigma_u, \hat{v}_n) + \Gamma_n)]^2 \leq C \).

By Taylor’s formula for \( Y_1(\sigma) \) and relations \( Y_1(\sigma_u) = \partial_{\sigma} Y_1(\sigma_u) = 0 \), we obtain \( \inf_{\sigma, \sigma_n}(\|Y_1(\sigma)|/|\sigma - \sigma_n|^2) \leq \inf_{u \in \mathbb{R}^{d+1}} u^\top \Gamma_n u/2 |u|^2 \). Then (B3) and the proof of Proposition 6.1 yield point 1. By Lemmas 4.4 and 5.2, Remark 5.2, and a similar argument to the proof of Proposition 2.1, we obtain 3–5 and \( \sup_n E[(b_n^{-1/4} \partial_{\sigma} H_n(\sigma_u, \hat{v}_n) - \partial_{\sigma} H_n(\sigma_u, \hat{v}_n))]^2 \leq C \). Moreover, by the Burkholder–Davis–Gundy inequality, we obtain

\[
E[(b_n^{-1/4} \partial_{\sigma} H_n(\sigma_u, \hat{v}_n))|^2] \leq CE \left[ \sum_m |\tilde{Z}_m^\top \partial_{\sigma} \tilde{S}_m^\top \tilde{Z}_m|^2 \right]^{L/2} \leq CE \left[ \sum_m |\tilde{Z}_m^\top \partial_{\sigma} \tilde{S}_m^\top \tilde{Z}_m|^2 \right] + C \left[ \sum_m |\tilde{Z}_m^\top \partial_{\sigma} \tilde{S}_m^\top \tilde{Z}_m|^2 \right] + O(1) \leq \left[ \sum_m |\tilde{Z}_m^\top \partial_{\sigma} \tilde{S}_m^\top \tilde{Z}_m|^2 \right]^{L/4} + O(1)
\]

33
for $L \geq 4$, which implies point 2.

Proof of Theorem 2.3.

We extend $Z_n(u)$ to a continuous function on $\mathbb{R}^d$ satisfying $\lim_{|u| \to \infty} Z_n(u) = 0$ with the supremum norm of the extended function the same as for the original one. Then by Theorem 5 and Remark 5 in Yoshida [30], it is sufficient to show $\limsup_{n \to \infty} E[|b_n^{1/4}(\sigma_n - \sigma_\star)|^p] < \infty$ for any $p > 0$ and $Z_n \to \mathcal{L}$ in $C(B(R))$ as $n \to \infty$ for any $R > 0$, where $Z(u) = \exp(N \cdot u - u^\top \Gamma_1 u/2)$ and $B(R) = \{u; |u| \leq R\}$.

By Lemma 4.4 and a similar argument to the proof of Proposition 2.1, we have

$$\lim_{N \to \infty} \sup_{n \geq N} \mathbb{E}\left[\sup_{u \in B(R)} |\partial_u \log Z_n(u)|\right] < \infty.$$  

Then Propositions 2.1 and 7.2 and tightness criterion in $C$ space in Billingsley [6] yield $Z_n \to \mathcal{L}$ in $C(B(R))$. Then (2.12) completes the proof.

Proof of Theorem 2.4.

By Theorem 10 in Yoshida [30], it is sufficient to show

$$\sup_{n} E\left[\left(\int_{U_n} Z_n(u)\pi(\sigma_\star + b_n^{-1/4} u)du\right)^{-1}\right] < \infty. \quad (9.1)$$

Similarly to the proof of Proposition 2.2, there exists $\delta > 0$ such that $\sup_{n} E[|H_n(\sigma_\star + b_n^{-1/4} u, \hat{v}_n) - H_n(\sigma_\star, \hat{v}_n)|^p] \leq C_p|u|^p$ for any $u \in U(\delta)$, where $U(\delta) = \{u \in \mathbb{R}^d; |u_i| \leq \delta (i = 1, \cdots, d)\}$. Then we have (9.1) by Lemma 2 in [30].

A Appendix

A.1 Results from linear algebra

Lemma A.1. Let $A$ and $B$ be matrices, with $A$ nonnegative definite and symmetric. Then

$$|\text{tr}(AB)| \leq \text{tr}(A)\|B\|.$$  

Lemma A.2. Let $l \in \mathbb{N}$, $A^j$ and $B^j$ be real-valued matrices and $\{\lambda_k^j\}_k$ be eigenvalues of $A^j$ for $1 \leq j \leq l$. Assume that $A^j$ is symmetric and all the elements of $B^j$ are nonnegative for $1 \leq j \leq l$. Then

$$\sum_{i_1, \cdots, i_{2l}} \prod_{j=1}^l \left|A^j_{i_{2j-1}, i_{2j}} B^j_{i_{2j-1}, i_{2j+1}}\right| \leq \prod_{j=1}^l \left(\|B^j\| \sum_k |\lambda_k^j|\right).$$

Proof. Let $U^j$ be an orthogonal matrix such that $A^j = (U^j)^\top \text{diag}((\lambda_k^j)_k) U^j$. Then

$$\sum_{i_1, \cdots, i_{2l}} \prod_{j=1}^l \left|A^j_{i_{2j-1}, i_{2j}} B^j_{i_{2j-1}, i_{2j+1}}\right| \leq \sum_{k_1, \cdots, k_l} \prod_{j=1}^l \left(\lambda_{k_1}^j |U_{i_{2j-1}, i_{2j-1}}^j| U_{i_{2j}, i_{2j-1}}^j |U_{i_{2j}, i_{2j+1}}^j| B^j_{i_{2j-1}, i_{2j+1}}\right)$$

$$\leq \sum_{k_1, \cdots, k_l} \prod_{j=1}^l \left(|\lambda_k^j| \|B^j\| \sum_i (U_{i, k_j}^j)^2\right) = \prod_{j=1}^l \left(\|B^j\| \sum_k |\lambda_k^j|\right).$$

Lemma A.3. Let $A$ be a symmetric matrix with $\|A\| < 1$. Then $\log \det(\mathcal{E} + A) = \sum_{p=1}^\infty (-1)^{p-1} p^{-1} \text{tr}(A^p).$
Proof. Let \( \{\lambda_j\}_{j=1}^k \) be eigenvalues of \( A \). Then \( \sum \lambda_j = \|A\| < 1 \), and hence

\[
\log \det(\mathcal{E} + A) = \sum_j \log(1 + \lambda_j) = \sum_j \sum_{p=1}^\infty (-1)^{p-1} p^{-1} \lambda_j^p = \sum_{p=1}^\infty (-1)^{p-1} p^{-1} \text{tr}(A^p).
\]

\( \square \)

**Lemma A.4.** Let \( A \) and \( B \) be symmetric, positive definite matrices. Assume that \( v^T Av \geq v^T Bv \) for any vector \( v \). Then \( \|A^{-1}\| \leq \|B^{-1}\| \) and \( \|A^{-1/2}\| \leq \|B^{-1/2}\| \).

Proof. Let \( \{\lambda_j^A\}_{j} \) and \( \{\lambda_j^B\}_{j} \) be eigenvalues of \( A \) and \( B \), respectively. Then for any unit vector \( v \), there exists an orthogonal matrix \( U \) such that

\[
\sum_j \lambda_j^A v_j^2 \geq \sum_j \lambda_j^B (Uv)_j^2 \geq \inf_j \lambda_j^B.
\]

Therefore we obtain \( \|A^{-1}\|^{-1} = \inf_j \lambda_j^A \geq \inf_j \lambda_j^B = \|B^{-1}\|^{-1} \) and \( \|A^{-1/2}\|^{-1} = \inf_j (\lambda_j^A)^{1/2} \geq \inf_j (\lambda_j^B)^{1/2} = \|B^{-1/2}\|^{-1} \).

\( \square \)

**Lemma A.5.** Let \( B \) be a symmetric, positive definite matrix and \( A \) be a symmetric, nonnegative definite matrix. Then \( \text{tr}(AB) \geq \text{tr}(A)\|B^{-1}\|^{-1} \).

Proof. Let \( \{\lambda_j^A\}_{j} \) and \( \{\lambda_j^B\}_{j} \) be eigenvalues of \( A \) and \( B \), respectively, and \( U \) be an orthogonal matrix satisfying \( UAU^T = \text{diag}((\lambda_j^A)_j) \). Then since \( (UBU^T)_{jj} \geq \inf_j \lambda_j^B = \|B^{-1}\|^{-1} \), we obtain

\[
\text{tr}(AB) = \sum_j \lambda_j^A (UBU^T)_{jj} \geq \sum_j \lambda_j^A \|B^{-1}\|^{-1} = \text{tr}(A)\|B^{-1}\|^{-1}.
\]

\( \square \)

**Lemma A.6.** Let \( \eta > 0 \) and \( A \) be a symmetric matrix. Assume that \( \mathcal{E} + A \) is positive definite and \( \|\mathcal{E} + A\| \leq \eta \). Then \( \text{tr}(A) - \log \det(\mathcal{E} + A) \geq \text{tr}(A^2)/(2(\eta \vee 1)) \).

Proof. We easily obtain the desired results by using the fact that \( \log \det(\mathcal{E} + A) = \sum_k \log(1 + \lambda_k) \) and that \( x - x^2/(2(\eta \vee 1)) \geq \log(1 + x) \) for \(-1 < x \leq \eta - 1\), where \( \lambda_j \) are eigenvalues of \( A \).

\( \square \)

**Lemma A.7.** Let \( A \) be a symmetric matrix, \( B \) a matrix of suitable size and \( \{\lambda_j\}_{j} \) eigenvalues of \( B^T A B \). Then

1. \( \|B^T AB\|_{ii} \leq \|A\|\|B\|_i \) for any \( i \).
2. \( \sum_j |\lambda_j| \leq \|A\|\text{tr}(B^TB) \).

Proof. 1. Let \( U \) be an orthogonal matrix and let \( \{\lambda'_j\}_{j} \) be eigenvalues of \( A \) such that \( U^T AU = \text{diag}(\lambda'_j) \). Then we obtain

\[
|(B^T AB)_{ii}| = \left| \sum_j \lambda'_j ((U^T B)_{jj})^2 \right| \leq \|A\| \sum_j ((U^T B)_{jj})^2 = \|A\|\|B^T B\|_{ii}.
\]

2. There exists an orthogonal matrix \( V \) such that \( \lambda_j = (V^T B^T ABV)_{jj} \) for any \( j \). Then

\[
\sum j |\lambda_j| = \sum j |(V^T B^T ABV)_{jj}| \leq \|A\| \sum j (V^T B^TBV)_{jj} = \|A\|\text{tr}(B^TB)
\]

by 1.

\( \square \)
A.2 Proof of (2.10)

Proof of (2.10).

Let $\hat{a}^3 = \hat{a}_1^3 \hat{a}_2^3$, $B_t = \sqrt{\det(b_t b_t^\top)}$, $\Upsilon_t = (\hat{a}_1^3 | b_t^j|^2 + \hat{a}_2^3 | b_t^j|^2 + 2\hat{a}_1^3 B_t^j)^{1/2}$, and $\psi(x, y) = (xy^\top + yx^\top)/2$ for vectors $x, y$. Then (2.8) yields

$$\mathcal{Y}_1(\sigma) = \int_0^T \left\{ \frac{\sum_{j=1}^2 (|b_t^j|^2 - |b_{t-}^j|^2)(\hat{a}_1^3 |b_t^j|^2 + \hat{a}_2^3 |b_t^j|^2 + 2\hat{a}_1^3 B_t^j)}{4B_t \Upsilon_t} - \frac{\Upsilon_t - \Upsilon_t^{(0)}}{2} \right\} dt.$$

Hence we obtain

$$\partial_{\sigma}^2 \mathcal{Y}_1(\sigma_*) = \int_0^T \left\{ \left. \frac{\sum_{j=1}^2 \partial_{\sigma}^2 |b_t^j|^2 (\hat{a}_1^3 |b_t^j|^2 + \hat{a}_2^3 |b_t^j|^2 + 2\hat{a}_1^3 B_t^j)}{4B_t \Upsilon_t} \right|_{\sigma=\sigma_*} - \frac{\partial_{\sigma}^2 \Upsilon_t}{2 \Upsilon_t} \right\} dt.$$

Terms involving $\partial_{\sigma}^2 \mathcal{Y}_1$ in the integrand of right-hand side are rewritten as

$$\left. \left( \hat{a}_1^3 \sum_{j=1}^2 \partial_{\sigma}^2 |b_t^j|^2 (\hat{a}_1^3 |b_t^j|^2 + \hat{a}_2^3 |b_t^j|^2 + 2\hat{a}_1^3 B_t^j) \right) \right|_{\sigma=\sigma_*} = \frac{\hat{a}_1^3}{4 \Upsilon_t^{(0)}} \left\{ \frac{\partial_{\sigma}^2 |b_t^j|^2 |b_t^j|^2 + |b_t^j|^2 |b_t^j|^2 - b_t^j |b_t^j|^2 - \partial_{\sigma}^2 B_t^j}{B_t} + \frac{2B_t^{(1)} (B_t^{(1)})^\top}{B_t^{(0)}} \right\} \right|_{\sigma=\sigma_*},$$

Other terms are rewritten as

$$\sum_{j=1}^2 \psi \left( \partial_{\sigma} |b_t^j|^2, \partial_{\sigma} \frac{\hat{a}_1^3 |b_t^j|^2 + \hat{a}_2^3 B_t^j}{2B_t \Upsilon_t} \right) \right|_{\sigma=\sigma_*} - \psi \left( \partial_{\sigma} B_t^j \right) \right|_{\sigma=\sigma_*} + \frac{\Upsilon_t^{(1)} (\Upsilon_t^{(1)})^\top}{2 \Upsilon_t^{(0)}},$$

Here we have used equations $B_t^{(1)} B_t^{(0)} + B_t^{(2)} B_t^{(0)} = 2B_t^{(0)} B_t^{(1)}$ and $\hat{a}_1^3 B_t^{(1)} + \hat{a}_2^3 B_t^{(2)} + 2\hat{a}_3^3 B_t^{(2)} = 2\Upsilon_t^{(0)} \Upsilon_t^{(1)}$. Therefore, we obtain (2.10).
A.3 Proof of Lemma 4.3

Let $A_m$ be a $(k_m^1 + k_m^2) \times (k_m^1 + k_m^2)$ matrix with elements $(A_m)_{ij} = 1_{i \geq j}1_{i \leq k_m^1 \text{ or } j > k_m^2}$, $1$ be a matrix with all elements equal to $1$ and $M_{m,*} = M_m(i, *)$.

**Lemma A.8.** Let $n \in \mathbb{N}$, $1 \leq m \leq \ell_n$, $q, q' \in \mathbb{N}$ such that $q' \leq 2q$, $A_m : \{1, \ldots, k_m^1 + k_m^2\} \to \{0, 1\}$ be a random map, $i : \{1, \ldots, q'\} \to \{1, \ldots, 2q\}$ be an injection, and $S$ be a symmetric, $G_{m-1}$-measurable random matrix of size $k_m^1 + k_m^2$ satisfying $\|M_{m,*}SM_{m,*}\| \leq b_n^{-1}$. Let $\mathcal{K} = \sum_{j_1, \ldots, j_q'} A_m(j_1, \ldots, j_q')$. Then there exists a positive constant $C$ which depends only on $v_*$ and $q$ such that

$$\left| \sum_{i_1, \ldots, i_{q'}} \prod_{j=1}^{q'} ((A_m^{-T})^{-1}SA_m^{-1})_{i_{2j-1}, i_{2j}} A_m(i_{(2j-1)}, \ldots, i_{(2j)}) \right| \leq C K (k_n/\kappa_n)^{2q} b_n^{-q'} k_n^{q+q'-2[q'/2]}.$$

**Proof.** Let $\bar{S} = (A_m^{-T})^{-1}SA_m^{-1}$ and $\bar{A}_m = M_{m,*}A_m^{-1}$. Then a simple calculation shows that

$$\left| (\bar{A}_m^{-1})_{ij} \right| = \left| (\bar{A}^{-1}_m M_{m,*} M_{m,*}^{-1})_{i,j} \right| \leq \left( \sum_k ((\bar{A}_m^{-1})_{i,k})^2 \right)^{1/2} \|M_{m,*}M_{m,*} \| \left( \sum_k ((\bar{A}_m^{-1})_{j,k})^2 \right)^{1/2} \leq (v_1^{-1} + v_2^{-1}) b_n^{-1} \kappa_n. \quad (A.1)$$

Similarly, we have

$$|\bar{S} \bar{S}]_{i,j} | \leq \left( \max_k ((\bar{A}_m^{-1})^{-1})_{i,k} \right) \|M_{m,*}M_{m,*} \| \|((\bar{A}_m^{-1})^T)_{i} \| \leq C b_n^{-2} \kappa_n^3. \quad (A.2)$$

Moreover, since $\sum_k ((A_m^{-T}A_m^{-1})_{i,j})_{i,j} = v_{k(i),*}^{-1}/(k_m^2 + 1)$, we obtain

$$|\text{tr}(\bar{S} \bar{S})| \leq \|M_{m,*}M_{m,*} \| \sum_{i,j,k,l} ((A_m^{-T}A_m^{-1})_{i,k} (A_m^{-T}A_m^{-1})_{j,l})_{i,j} \leq C b_n^{-2} \kappa_n^{-1} \kappa_n \kappa_n. \quad (A.3)$$

If both $i_{2j-1}$ and $i_{2j}$ are outside the image of $i$, we have

$$\sum_{j=1}^{q'} S_{i_{2j-1}, i_{2j}} A_m(i_{(2j-1)}, \ldots, i_{(2j)}) = \text{tr}(\bar{S} \bar{S}) A_m(i_{(2j-1)}, \ldots, i_{(2j)}).$$

Moreover, if both $i_{2j-1}$ and $i_{2k-1}$ are in the image of $i$ and neither $i_{2j}$ nor $i_{2k}$ is in it, then we have

$$\sum_{j=1}^{q'} S_{i_{2j-1}, i_{2j}} S_{i_{2k-1}, i_{2k}} A_m(i_{(2j-1)}, \ldots, i_{(2j)}) = (\bar{S} \bar{S}) A_m(i_{(2j-1)}, \ldots, i_{(2j)}).$$

Therefore there exist $\alpha_k \in \{0, 1\}$ for $1 \leq k \leq [q'/2]$, $0 \leq s \leq [(2q - q')/2]$ and a bijection $\iota' : \{1, \ldots, q'\} \to \{1, \ldots, q'\}$ such that $\sum_{k=1}^{[q'/2]} \alpha_k + [q'/2] + s = q - (q' - 2[q'/2])$ and

$$\left| \sum_{i_1, \ldots, i_{q'}} \prod_{j=1}^{q'} S_{i_{2j-1}, i_{2j}} A_m(i_{(2j-1)}, \ldots, i_{(2j)}) \right| \leq (C b_n^{-1} \kappa_n^{2q} q^{-2} b_n^{-q'/2}) \left( \sum_{j_1, \ldots, j_{q'}} \prod_{k=1}^{[q'/2]} |((\bar{S} \bar{S}) A_m(i_{(2j-1)}, \ldots, i_{(2j)})) | \right) \leq C (\kappa_n/\kappa_n)^{2q} (b_n^{-1} \kappa_n^{-q'})^{q' - 2[q'/2]} \| \mathcal{K} \| \leq C K (k_n/\kappa_n)^{2q} b_n^{-q} \kappa_n^{-q - 2[q'/2]}.$$
by (A.1)–(A.3).

\begin{proof}

Let \(\{\tilde{e}_{i.m}\}_{1 \leq i \leq k_1^m + k_2^m}\) and \(\{\hat{e}_{i,m}\}_{1 \leq i \leq k_1^m + k_2^m}\) be sequences of random variables defined by \(\tilde{e}_{i,m} = \epsilon_{i+K_{m-1}^n+1}^{n_1}\) and \(\hat{e}_{i,m} = \epsilon_{i+K_{m-1}^n+1}^{n_2}\) for \(i \leq k_1^m\) and \(\tilde{e}_{i,m} = \epsilon_{i-k_2^m-K_{m-1}^n+1}^{n_1}\) and \(\hat{e}_{i,m} = \epsilon_{i-k_2^m-K_{m-1}^n+1}^{n_2}\) for \(i > k_1^m\). Moreover, let

\[
\begin{align*}
Z_{1,m} &= \left( (\tilde{e}_{i,m}^\top - (W_{S_1}^n - W_{S_1^m}))_{i=m-K_{m-1}^n+2}^\top, (\hat{e}_{j,m}^\top - (W_{S_2}^n - W_{S_2^m}))_{j=m-K_{m-1}^n+2}^\top \right)^	op, \\
Z_{2,m} &= \left( (\tilde{e}_{i,m}^\top - \epsilon_{i-1}^m)^{K_{m-1}^n+2}, (\hat{e}_{j,m}^\top - \epsilon_{j-1}^m)^{K_{m-1}^n+2} \right)^	op,
\end{align*}
\]

and \(\tilde{S}_{1,m} = \tilde{S}_{1,m} - M_{m,*}\).

Let \(\tilde{U}_{1,m,*}\) be an orthogonal matrix and let \(A_{1,m,*}\) be a diagonal matrix satisfying \(\tilde{U}_{1,m,*}\tilde{S}_{1,m,*}\tilde{U}_{1,m,*}^\top = A_{1,m,*}\). Then since \(\tilde{Z}_{1,m}|\mathcal{G}_{m-1} \sim N(0, \tilde{S}_{1,m,*})\), we have \(\tilde{U}_{1,m,*}\tilde{Z}_{1,m}|\mathcal{G}_{m-1} \sim N(0, A_{1,m,*})\). Therefore, for any \(q \in \mathbb{N}\) and \(1 \leq j_1, \ldots, j_q \leq k_1^m + k_2^m\), we obtain

\[
E_m\left[ \prod_{k=1}^{2q} (\tilde{U}_{1,m,*}\tilde{Z}_{1,m})_{j_k} \right] = \sum_{(l_{2k-1}, l_{2k})_{k=1}^q} \prod_{k=1}^q A_{(1,m,*), l_{2k-1}, l_{2k}},
\]

where the summations on the right-hand side of both equations are over all \(q\)-pairs \((l_{2k-1}, l_{2k})_{k=1}^q\) of variables \(j_1, \ldots, j_q\).

1. Let \(h(\mathbf{A}, \mathbf{B})_{i_1, \ldots, i_d} = (A_{i_1,i_2}B_{i_3,i_4} + A_{i_1,i_3}B_{i_2,i_4} + A_{i_1,i_4}B_{i_2,i_3} + A_{i_2,i_3}B_{i_1,i_4} + A_{i_2,i_4}B_{i_1,i_3} + A_{i_3,i_4}B_{i_1,i_2} + A_{i_3,i_2}B_{i_1,i_4} + A_{i_4,i_3}B_{i_1,i_2})/2\) for square matrices \(A\) and \(B\) of the same size, and \(\delta_{i_1, \ldots, i_d}\) be a \(\{0,1\}\)-valued function that is equal to 1 if and only if \(i_1 = \cdots = i_d\). Then we have

\[
E_m[\langle \tilde{Z}_{2,m}^\top S\tilde{Z}_{2,m} \rangle] = E_m[(\langle A_m \tilde{Z}_{2,m} \rangle \tilde{S}(A_m \tilde{Z}_{2,m}))^2] = \sum_{i_1, \ldots, i_4} S_{i_1,i_2}S_{i_3,i_4} E_m[\prod_{j=1}^4 (\tilde{e}_{j,m} - \hat{e}_{j,m})] = \sum_{i_1, \ldots, i_4} S_{i_1,i_2}S_{i_3,i_4} \left\{ \phi(M_1, M_2)_{i_1, \ldots, i_4} \left[ (E[\tilde{e}_{i_1,m}^4]) - 3(E[\tilde{e}_{i_1,m}^2])^2 \right] 1_{\{ \text{max}_{1 \leq j \leq 4} i_j \leq k_1^m \text{ or } \min_{1 \leq j \leq 4} i_j > k_1^m \}} \right.
\]

\begin{align*}
&+ 2\phi(M_1, M_2)_{i_1, \ldots, i_4} + \phi(M_2, M_2)_{i_1, \ldots, i_4} + (E[\tilde{e}_{i_1,m}^4]) - 3E[\tilde{e}_{i_1,m}^2] \delta_{i_1, i_2, i_3, i_4}],
\end{align*}

where \(M_1 = \text{diag}(v_1, 1, v_2, 1)\) and \(M_2 = \text{diag}(v_1, \mathcal{E}_1, v_2, \mathcal{E})\).

Hence, we obtain

\[
E_m[\langle \tilde{Z}_{2,m}^\top S\tilde{Z}_{2,m} \rangle] = E_m[\langle \tilde{Z}_{2,m}^\top S\tilde{Z}_{2,m} \rangle] = \sum_{i_1, \ldots, i_4} S_{i_1,i_2}S_{i_3,i_4} (\phi(S_{1,m,*}, S_{1,m,*}) + 2\phi(S_{1,m,*}, M_{m,*}))_{i_1, \ldots, i_4} = \sum_{i_1, \ldots, i_4} C_{i_1, i_2} S_{i_1,i_2} \tilde{S}_{i_1,i_2} + \sum_{j=1}^4 (E[\tilde{e}_{i_1,m}^4]) - 3\tilde{e}_{i_1,m}^2 \text{tr}(S_{\mathcal{E}_1}^\top S_{\mathcal{E}}) + \sum_{i_1} C_{i_1} S_{i_1,i_1}^2,
\]

by (A.5), where \(C_{n,*} = E[\langle \tilde{e}_{i_1,m}^4 \rangle] - 3E[\langle \tilde{e}_{i_1,m}^2 \rangle]^2 \text{ and } \mathcal{E}_1 = (k_1^m + k_2^m) \times (k_1^m + k_2^m) \text{ matrix with elements } (\mathcal{E}_1)_{kl} = 1_{\{k < k_1^m\}} 1_{\{j < k_1^m\}}\) and \((\mathcal{E}_2)_{kl} = 1_{\{k > k_1^m\}} 1_{\{j > k_1^m\}}\).

Lemma A.7 and (A.1) yield

\[
\sum_i C_{n,i} S_{i_1,i_1}^2 \leq C\sup_i C_{n,i} b_n^{-1} k_n \| \tilde{S}_{m,*} \| \| \tilde{S}_{m,*} \| tr(A_n^{-1})^{-1} \tilde{S}_{m,*}^{-2} \leq \tilde{R}_n(b_n^{-2} k_n) \text{tr}(M_{1,m} 0 0 M_{2,m}) \tilde{S}_{m,*}^{-2} \leq \tilde{R}_n(b_n^{-2} k_n) \text{tr}(\tilde{S}_{m,*}^{-1}) = \tilde{R}_n(b_n^{-3/2} k_n^2) = \tilde{R}_n(1). \quad (A.7)
\]
Here, we used the fact that diagonal elements \((\tilde{S}_m^{-2})_{11}\) and \((\tilde{S}_m^{-2})_{k_{m-1},k_{m-1}}\) are positive since \(\tilde{S}_m^{-2}\) is positive definite.

Moreover, the assumptions and Lemmas A.1,A.4 and A.7 yield
\[
\text{tr}(\tilde{S}\varepsilon(1)\tilde{S}\varepsilon(1)) \leq \text{tr}
\left(\begin{array}{cc}
1 + E & 0 \\
0 & 0 \\
\end{array}\right) \tilde{S}\varepsilon(1)
\leq C\text{tr}(\tilde{S}_m^{-1/2}A_m^{-1}\varepsilon(1)(A_m^\top)^{-1}\tilde{S}_m^{-1/2})
\left\|\tilde{S}_m^{-1/2}S\left(\begin{array}{cc}
M_{1,m} & 0 \\
0 & 0 \\
\end{array}\right)SS_m^{-1/2}\right\|
\leq Cb_n^{-2}r_n(\tilde{S}_m^{-1})_{11}
\leq Cb_n^{-2}r_n(\tilde{S}_m^{-1})_{11} = \tilde{R}_n(1),
\]
(A.8)
since \((M_{m-1})_{11} \leq v^{-1}_1\) by (4.5).

(A.6)–(A.8) and similar estimates for \(\text{tr}(\tilde{S}\varepsilon(2)\tilde{S}\varepsilon(2))\) yield
\[E_m[(\tilde{Z}_m^T S \tilde{Z}_m)^2] = 2\text{tr}(S\tilde{S}_m) + \text{tr}(S\tilde{S}_m)^2 + \tilde{R}_n(1).\]

We next prove the estimate for \(E_m[(\tilde{Z}_m^T S \tilde{Z}_m)^2] = \tilde{R}_n(b_n^{-2}k_n^{2p}).\)

Note that
\[
E_m[(\tilde{Z}_m^T S \tilde{Z}_m)^2] = \sum_{i_1,\ldots,i_{4p}} S_{i_1,i_2} \cdots S_{i_{4p-1},i_{4p}} E_m\left[\prod_{j=1}^{4p}(\tilde{\epsilon}_{i_j,m} - \epsilon_{i_j,m})\right],
\]
and there exist \(\{0,1\}\)-valued maps \(\{A_{1,m}\}_{i_{1}},\) constants \(C_i,\) positive integers \(\{q_{i}\}\) not greater than \(4p\) and injections \(\{i_{1}\}\) such that \(E_m[\prod_{j=1}^{4p}(\tilde{\epsilon}_{i_j,m} - \epsilon_{i_j,m})] = \sum_i C_i A_{1,m}(i_{i(1)},\ldots,i_{i(q_i)})\) and \(\sum_{j_1,\ldots,j_{q_i}} A_{1,m}(j_1,\ldots,j_{q_i}) = \tilde{R}_n(k_n^{[q_i]/2})\) for any \(l\). Then Lemma A.8 yields
\[E_m[(\tilde{Z}_m^T S \tilde{Z}_m)^2] = \tilde{R}_n(b_n^{-2}k_n^{2p}).\] (A.9)

Moreover, (A.5) yields
\[
E_m[(\tilde{Z}_m^T S \tilde{Z}_m)^2] \leq C_p \sum_{\gamma = (\gamma_1,\ldots,\gamma_4); \gamma_k \geq 1, \sum_k \gamma_k = 2p} \prod_k \text{tr}(S\tilde{S}_m)^{\gamma_k} = \tilde{R}_n(b_n^{-p}k_n^{2p}).
\] (A.10)

Furthermore, by calculating the expectation of \(\tilde{Z}_{1,m}\) and using (A.5) and Lemma A.7, we have
\[
E_m[(\tilde{Z}_m^T S \tilde{Z}_m)^2] = \left(\sum_{(l_{2k-1},l_{2k})_{k=1}^4} 1\right) E_m[(\tilde{Z}_m^T S \tilde{S}_m S \tilde{Z}_m)^2] \leq (2p - 1)!||M_{1,m} S \tilde{S}_m S M_{1,m}||^2 E_m[(\tilde{Z}_m^T M_{1,m}^2 S \tilde{Z}_m)^2] = \tilde{R}_n(b_n^{-p}k_n^{2p}).
\]

Then we obtain \(E_m[(\tilde{Z}_m^T S \tilde{Z}_m)^2] = \tilde{R}_n(b_n^{-2}k_n^{4p}).\)

For the estimate of \(E_m[(\tilde{Z}_m^T S \tilde{Z}_m)^4]\), we have \(E_m[(\tilde{Z}_m^T S \tilde{Z}_m)^4] = \tilde{R}_n(b_n^{-2}k_n^{4p})\) and \(E_m[(\tilde{Z}_m^T S \tilde{Z}_m)^4] = \tilde{R}_n(b_n^{-2}k_n^{4p})\) by the above results. Moreover, we have
\[
E_m[(\tilde{Z}_m^T S \tilde{Z}_m)^4] = \sum_{i_1,\ldots,i_8} \prod_{k=1}^4 ((A_m)^{-1} S A_m^{-1})_{i_{2k-1},i_{2k}} E_m\left[\prod_{j=1}^8(\tilde{\epsilon}_{i_j,m} - \epsilon_{i_j,m})\right],
\]
and there exist \(\{0,1\}\)-valued maps \(\{A_{1}\}_{i_{1}},\) constants \(C_i,\) positive integers \(\{q_{i}\}\) not greater than \(8\) and injections \(\{i_{1}\}\) such that \(\sum_{j_1,\ldots,j_{q_i}} A_{1}(j_1,\ldots,j_{q_i}) = \tilde{R}_n(k_n^{[q_i]/3})\) and
\[
E_m[\prod_{j=1}^8(\tilde{\epsilon}_{i_j,m} - \epsilon_{i_j,m})] = \sum_{(l_{2k-1},l_{2k})_{k=1}^4} \prod_{k=1}^4 (\tilde{\epsilon}_{i_{2k-1},m} - \epsilon_{i_{2k-1},m}) + \sum_{l} C_l A_{1}(i_{i(1)},\ldots,i_{i(q_i)})
\] (A.12)
Moreover, for any $q \in \mathbb{N}$, we have 

$$
E_m[(\tilde{Z}_{2,m}^T S \tilde{Z}_{2,m})^q] = \sum_{k_1, \ldots, k_q} \prod_{k=1}^q \left((A_m^{-1})^T S A_m^{-1}\right)_{k_1, \ldots, k_q} \sum_{(l_1, \ldots, l_q)} \delta_{l_1, \ldots, l_q} + R_n(b_n^{-4} k_n^q)
$$

Then (A.11), (A.12) and Lemma A.8 yield

$$
E_m[(\tilde{Z}_{2,m}^T S \tilde{Z}_{2,m})^4] = \sum_{i_1, \ldots, i_8} \prod_{k=1}^8 ((A_m^{-1})^T S A_m^{-1})_{i_1, \ldots, i_8} \sum_{(l_1, \ldots, l_8)} \delta_{l_1, \ldots, l_8} + R_n(b_n^{-4} k_n^7)
$$

2. Let $(I_{i,m})_{i=1}^{k_m^1+k_m^2}$ be defined by $I_{i,m} = I_{i,m}^1$ for $1 \leq i \leq k_m^1$ and $I_{i,m} = I_{i-k_m^1,m}^2$ for $k_m^1 < i \leq k_m^1 + k_m^2$. Since $(Z_m - \tilde{Z}_m, \tilde{Z}_m) = \int_{I_{i,m}} (b_{t,i}^k(i) + \tilde{b}_{m,i}^k(i)) dW_t + \mu_{m,i}^k(i) dt$, we have $(Z_m - \tilde{Z}_m)^T S (Z_m + \tilde{Z}_m) = \Psi_{m,1} + \Psi_{m,2} + \Psi_{m,3}$, where

$$
\Psi_{m,1} = 2(Z_m - \tilde{Z}_m)^T S \tilde{Z}_m + \sum_{k=1}^{k_m^1} \sum_{i,j} \int_{I_{i,m}} (b_{t,i}^k(i) + (-1)^k \tilde{b}_{m,i}^k(i)) (b_{t,j}^k(i) + (-1)^k \tilde{b}_{m,j}^k(i)) dW_t dW_t',
$$

and

$$
E_m[\Psi_{m,3}] = 0, \quad \Psi_{m,2} = R_n(b_n^{-1} k_n^{3/2}) \quad \text{and} \quad \Psi_{m,3} = R_n(b_n^{-3/2} k_n^4).
$$

Here we used the facts that

$$
|\psi| \leq \|A\||x||y|
$$

for vectors $x, y$ and a matrix $A$,

$$
E_m[\text{tr}(S D)] = |\text{tr}(S E_m[D])| \leq \|E_m[D]||S\hat{S}_{m,n}||\text{tr}(\hat{S}_{m,n}^{-1})| = R_n(b_n^{-3/2} k_n^2),
$$

and

$$
E_m[\text{tr}(S D)]^q \leq CE_m[\|D\|]^q ||\hat{S}_{m,n}||^q \text{tr}(\hat{S}_{m,n}^{-1})^q] \leq CE_m[\|D\|^{q/2} \epsilon_n^{-q} R_n(b_n^{-q} \epsilon_n^{-q/2})]
$$

for any $q \geq 1$, where $D = (D_{i,j})_{i,j}$ and $D_{i,j} = \int_{I_{i,m}} \int_{I_{j,m}} (b_{t,i}^k(i) + \tilde{b}_{m,i}^k(i)) (b_{t,j}^k(i) - \tilde{b}_{m,j}^k(i)) dt$.

Then the Burkholder–Davis–Gundy inequality yields

$$
E_m\left[\left(\sum_{m}(Z_m - \tilde{Z}_m)^T S (Z_m + \tilde{Z}_m)\right)^q\right] \leq CE_m\left[\left(\sum_{m}(\Psi_{m,1} + \Psi_{m,3})\right)^{2q}\right] + R_n(b_n^{-2} k_n^q)
$$

$$
\leq CE_m\left[\left(\sum_{m} E_m[\Psi_{m,1}^2]\right)^{q/2}\right] + CE_m\left[\left(\sum_{m} E_m[\Psi_{m,3}^2]\right)^{q/2}\right] + R_n(b_n^{-2} k_n^q).
$$

(14.14)

Let $\tilde{Z}_{m,i} = \int_{I_{i,m}} (b_{t,i}^k(i) - \tilde{b}_{m,i}^k(i)) dW_t$ and $Z_m = (\tilde{Z}_{m,i})_i$, then we obtain

$$
\Psi_{m,1} = 2(Z_m - \tilde{Z}_m)^T S \tilde{Z}_m + \sum_{m} \sum_{i,j} E_m[\Psi_{m,1}^2] + 2(Z_m + \tilde{Z}_m)^T S Z_m + (2 \tilde{Z}_m + \tilde{Z}_m)^T S (\mu_{m,i}^k(i) \int_{I_{i,m}} t dt + R_n(b_n^{-1} k_n^{3/2}).
$$

(15.15)

Moreover, for any $p \in \mathbb{N}$, (A.13), (A.9) and the Cauchy-Schwarz inequality yield

$$
E_m[(2(Z_m - \tilde{Z}_m)^T S \tilde{Z}_m + (2 \tilde{Z}_m + \tilde{Z}_m)^T S \tilde{Z}_m)^{2p}] \leq CE_m[||\tilde{S}_{m,n}||^2 (Z_m - \tilde{Z}_m)^T S \tilde{Z}_m\|^{2p}]^{1/2} \| E_m[(\tilde{Z}_m^T \tilde{S}_{m,n} \tilde{Z}_m)^{2p}]^{1/2} + R_n((b_n k_n b_n^{-1} b_n^{-1} k_n^{2p})
$$

$$
= R_n((b_n^{-1} \epsilon_n^{-1} k_n^{2p})^{1/2} + R_n(b_n^{-2} k_n^4)
$$

(16.16)

We can rewrite $\tilde{Z}_{m,i} = L_i^1 + L_i^2 + L_i^3 + R_n(b_n^{-1/2} \epsilon_n^{-1/2})$, where $L_i^1 = \sum_{j,k} (l_{ij}^k) \int_{I_{i,m}} (t - s_{m-1}) dW_t$, $L_i^2 = \sum_{j,k} (l_{ij}^k) \int_{I_{i,m}} (W_t^k - s_{m-1}) dW_t$, and $L_i^3 = \sum_{j,k,l} (l_{ij}^k) \int_{I_{i,m}} (W_t^l - s_{m-1}) dW_t$ for some $\mathcal{G}_{m-1}$-measurable random variables $l_{ij}^k$, $l_{ij}^k$, and $l_{ijk,l}$ with bounded moments.
Let $L^j = (L^j_n)$. Then, Lemma A.7 and (A.10) yield
\[
E_m[(Z_m^T S \tilde{Z}_m + 2 \tilde{Z}_m^T S \tilde{Z}_m)^{2p}] \\
\leq \|S^{1/2}_{m,*} S^{1/2}_{m,*}\|^{2p} E_m[(L^2_m)^{2p}] + R_n((b_n k_n b_n^{-1} f_n^{-2})^{2p}) + C \sum_{j=1}^{3} E_m[(\tilde{Z}_m^T S \tilde{Z}_m)^{2p}]
\]
\[+ E_m \left[ \|S_{m,*} S (\tilde{Z}_m - \sum_{j=1}^{3} L^j) S_{m,*}\|^{2p} \right]^{1/2} E_m[(\tilde{Z}_m^T S \tilde{Z}_m)^{2p}]^{1/2}
\]
\[\leq R_n(1) E_m[(L^2_m)^{2p}] + C \sum_{j=1}^{3} E_m[(\tilde{Z}_m^T S \tilde{Z}_m)^{2p}] + R_n(b_n^{-4p} k_n^{3p}). \tag{A.17}
\]
Here we also used that $(a + b)^T A (a + b) \leq 2a^T A a + 2b^T A b$ for a positive definite matrix $A$ and vectors $a, b$.

Moreover, we can see that there exists a positive constant $C_p$ such that
\[
E_m \left[ \sum_{i_1, \ldots, i_{2p} = 1}^{2p} \prod_{j=1}^{2p} \int_{I_{i_{k,m}}} (W_{i_{k,m}}^{\rho_{i_{k,m}}} - W_{i_{k,m}}^{\rho_{i_{k,m}}}) \right] \\
\leq C_p \sum_{i_1, \ldots, i_{2p} = 1}^{2p} \sum_{j=1}^{2p} \prod_{q=1}^{2p} \left| \int_{I_{i_{q-1,m}} \cap I_{i_{q+1,m}}} \right|_{n}^{20} (s_m - s_{m-1})^{p-\alpha}
\]
for any $\{p_i, k_i\}_{1 \leq i \leq \leq 2p} \subseteq \{1, 2\}$, where the second summation on the right-hand side is taken over $0 \leq \alpha \leq p$ and $(2p - \alpha)$ disjoint pairs $(i_{q-1}, i_{q+1})_{q=1}^{2p-\alpha}$ in the variables $i_1, \ldots, i_{2p}, j_1, \ldots, j_{2p}$. Here we used the fact that all $6p$ factors $(W_{i_{k,m}}^{\rho_{i_{k,m}}}, \int_{I_{i_{k,m}}} dW_{i_{k,m}}^{\rho_{i_{k,m}}}, \int_{I_{i_{k,m}}} dW_{i_{k,m}}^{\rho_{i_{k,m}}}, \int_{I_{i_{k,m}}} dW_{i_{k,m}}^{\rho_{i_{k,m}}}, \int_{I_{i_{k,m}}} dW_{i_{k,m}}^{\rho_{i_{k,m}}}, \int_{I_{i_{k,m}}} dW_{i_{k,m}}^{\rho_{i_{k,m}}})_{k=1}^{2p}$ should be separated into $3p$ pairs in the non-zero terms. $2 \alpha$ represents the number of pairs with the form $(W_{i_{k,m}}^{\rho_{i_{k,m}}}, (W_{i_{k,m}}^{\rho_{i_{k,m}}} - W_{i_{k,m}}^{\rho_{i_{k,m}}})_{i_{k,m}})$ or $(\int_{I_{i_{k,m}}} dW_{i_{k,m}}^{\rho_{i_{k,m}}}, (W_{i_{k,m}}^{\rho_{i_{k,m}}} - W_{i_{k,m}}^{\rho_{i_{k,m}}})_{i_{k,m}})$. Therefore we obtain
\[
E_m[(Z_m^T S \tilde{Z}_m)^{2p}] = R_n(\sum_{0 \leq \alpha \leq p} (b_n^{1/2} f_n^{-1})^{2p-\alpha} (b_n^{3/2} f_n^{-1} k_n)^{\alpha} (f_n^{-2p+\alpha}) = R_n((b_n^{-1/2} k_n^{3/2})^{2p}). \tag{A.17}
\]
and similar arguments for $E_m[(Z_m^T S \tilde{Z}_m)^{2p}]$, $E_m[(Z_m^T S \tilde{Z}_m)^{2p}]$ and $E_m[(L^2_m)^{2p}]$ yield
\[
E_m[(Z_m^T S \tilde{Z}_m + 2 \tilde{Z}_m^T S \tilde{Z}_m)^{2p}] = R_n(b_n^{-2p} k_n^{3p}). \tag{A.18}
\]
(A.15), (A.16) and (A.18) yield
\[
E_m[(\Psi_{m,1}^2)^{q/4}] = R_n((b_n^{-2p} k_n^{3p})^{q/4}). \tag{A.19}
\]
(A.13), (A.15) and (A.18) yield
\[
E_m[\Psi_{m,1}^2] \leq C \|S^{1/2}_{m,*} S M_{m,*} S^{1/2}_{m,*}\| E_m[(Z_m - \tilde{Z}_m) S_{m,*} S_{m,*} (Z_m - \tilde{Z}_m)] + R_n(b_n^{-2k_n^{3/2}})
\]
\[+ \|S (E_m[(Z_m - \tilde{Z}_m) S_{m,*} (Z_m - \tilde{Z}_m)])_{i,j} S\| \sum_{i} (\rho_{m,i})^{2} \leq R_n(b_n^{-2k_n^{3/2}}) + R_n(b_n^{-2k_n^{3/2}}) + R_n(b_n k_n b_n^{-2}) = R_n(b_n^{-2k_n^{3/2}}),
\]
and therefore, we have $E_m[(\sum_{m} E_m[\Psi_{m,1}^2])^{q/2}] = R_n(b_n^{-q/2} k_n^{q/2})$, which completes the proof of point 2.

Point 3 is easily obtained by the proof of point 2 since we only need the estimate for $E_m[(\sum_{m} E_m[\Psi_{m,1}^2])]$ if $q = 2$. \[\square\]
A.4 An additional lemma

**Lemma A.9.** Let $e_n$ be a sequence of positive numbers, $S$ be an open set in a Euclidean space, $A_n(\lambda)$ and $B_n(\lambda)$ be sequences of positive-valued random variables, and $C_n(\lambda)$ be a sequence of non-negative-valued random variables for $\lambda \in S$. Assume that $A_n(\lambda)$, $B_n(\lambda)$, and $C_n(\lambda)$ are $C^3$ with respect to $\lambda$, $e_n \to \infty$, $\sup_{0 \leq k \leq 3, \lambda \in S} |\partial^k A_n| \vee |\partial^k B_n| = O_p(e_n^{-1})$, and $\sup_{0 \leq k \leq 3, \lambda \in S} |\partial^k C_n| = O_p(e_n^{-2})$ as $n \to \infty$, $C_n < A_n B_n$ a.s. for any $n \in \mathbb{N}$, and $\lim_{n \to \infty} (e_n^2(A_n B_n - C_n)) > 0$ a.s. Then

\[
\sup_{\lambda \in S} \left| \frac{\partial^k}{(\partial \lambda^k) \left( \int_0^\infty \frac{C_n}{f_p(A_n(x), f_p(B_n(x))} dx - \frac{\pi C_n}{2P_n \sqrt{A_n B_n - C_n}} \right) \right| = O_p(e_n^{-1}), \tag{A.20}
\]

\[
\sup_{\lambda \in S} \left| \frac{\partial^k}{(\partial \lambda^k) \left( \int_0^\infty \frac{C_n f_l(A_n(x), f_p(B_n(x))} {dx - \frac{\pi C_n(A_n + \sqrt{A_n B_n - C_n})}{2P_n \sqrt{A_n B_n - C_n}} \right) \right| = O_p(e_n^{-2}), \tag{A.21}
\]

\[
\sup_{\lambda \in S} \left| \frac{\partial^k}{(\partial \lambda^k) \left( \int_0^\infty \frac{C_n}{f_p(A_n(x), f_p(B_n(x))} dx - \pi(\sqrt{A_n + \sqrt{B_n} + \pi P_n} \right) \right| = O_p(e_n^{-1}) \tag{A.22}
\]

for $0 \leq k \leq 3$, where $f_p(a, x) = (a + 2(1 - \cos x))^p$ and $P_n = \sqrt{A_n + B_n + 2\sqrt{A_n B_n - C_n}}$.

**Proof.** An elementary calculation yields

\[
\sum_{p=1}^\infty \int_0^\pi \frac{C_n}{(A_n + 2(1 - \cos x))^p(B_n + 2(1 - \cos x))^p} dx = \int_0^\pi \left( 1 - \frac{C_n}{(A_n + \frac{2t^2}{1 + t^2})(B_n + \frac{2t^2}{1 + t^2})} \right)^{-1} \frac{C_n}{(A_n + \frac{2t^2}{1 + t^2})(B_n + \frac{2t^2}{1 + t^2})} \left( \frac{1}{1 + t^2} dt \right)
\]

\[
= \int_{-\infty}^\infty \frac{C_n}{C_n(1 + t^2) - C_n} \left( 1 + t^2 \right) \left( A_n + (B_n + C_n + 16)t^2 + 2(A_n B_n + 2A_n + 2B_n - C_n)t^2 + A_n B_n - C_n dt. \tag{A.23}
\]

We only consider the case $(A_n - B_n)^2 + 16C_n > 0$ a.s. We can easily obtain the desired results for the other case with a slight modification.

Let

\[
\alpha_1 = \frac{-2A_n - 2B_n - \sqrt{4(A_n - B_n)^2 + 16C_n}}{16} \quad \text{and} \quad \alpha_2 = \frac{-2A_n - 2B_n + \sqrt{4(A_n - B_n)^2 + 16C_n}}{16},
\]

then we have $\alpha_1 < \alpha_2 < 0$ by the assumptions. Moreover, we can calculate the right-hand side of (A.23) as

\[
(1 + O_p(e_n^{-1})) \int_{-\infty}^\infty \frac{C_n(1 + t^2)}{16(t^2 - \alpha_1)(t^2 - \alpha_2)} dt = (1 + O_p(e_n^{-1})) \pi \frac{C_n(1 + \alpha_1)}{16(2\sqrt{-\alpha_1})(\alpha_1 - \alpha_2)} + \frac{C_n(1 + \alpha_2)}{16(2\sqrt{-\alpha_2})(\alpha_2 - \alpha_1)}
\]

\[
= (1 + O_p(e_n^{-1})) \pi \frac{C_n(1 + \alpha_1)}{16 \sqrt{\alpha_1\alpha_2}(\alpha_1 - \alpha_2)} + O_p(e_n^{-3/2}).
\]

Therefore, we obtain (A.20) with $k = 0$ by noting that $16\alpha_1\alpha_2 = A_n B_n - C_n$ and

\[
\sqrt{x + y + \sqrt{(x-y)^2 + 4z} + \sqrt{x + y - \sqrt{(x-y)^2 + 4z}}} = \sqrt{2x + 2y + 4\sqrt{xy - z}}
\]

for non-negative numbers $x, y, z$ satisfying $xy \geq z$. 

42
We similarly have (A.20) with $1 \leq k \leq 3$ and (A.21) with $0 \leq k \leq 3$.
Furthermore, we obtain $\int_{-\infty}^{\infty} (1 + t^2)^{-1} \log(t^2 + \alpha^2) dt = 2\pi \log(\alpha + 1)$ for any $\alpha > 0$ by the residue theorem.
Therefore, we obtain

$$
\frac{\partial^k}{\partial \lambda^k} \sum_{p=1}^\infty \int_0^\infty \frac{C_p}{f_p(A_n, x)f_p(B_n, x)} dx
$$

which completes the proof.

Acknowledgement

The author is grateful to Yuta Koike for very useful discussions and his help on numerical analysis of quadratic covariance estimators. He had a great contribution on constructing the ‘cc’ function in the ‘yuima’ R package as well. This work was partially supported by Japan Society for the Promotion of Science KAKENHI Grant Number 15K21598, and Japan Science and Technology Agency PRESTO.

References


