Japanese Journal of Statistics and Data Science (2023) 6:505–550
https://doi.org/10.1007/s42081-023-00196-0

ORIGINAL PAPER

Statistics for Stochastic Processes

Asymptotically efficient estimation for diffusion processes with nonsynchronous observations

Teppei Ogihara[1,2]

Received: 30 June 2022 / Revised: 9 February 2023 / Accepted: 11 February 2023 /
Published online: 10 March 2023

Abstract

We study maximum-likelihood-type estimation for diffusion processes when the coefficients are nonrandom and observations occur in nonsynchronous manner. The problem of nonsynchronous observations is important when we consider the analysis of high-frequency data in a financial market. Constructing a quasi-likelihood function to define the estimator, we adaptively estimate the parameter for the diffusion part and the drift part. We consider the asymptotic theory when the terminal time point T_n and the observation frequency goes to infinity, and show the consistency and the asymptotic normality of the estimator. Moreover, we show local asymptotic normality for the statistical model, and asymptotic efficiency of the estimator as a consequence. To show the asymptotic properties of the maximum-likelihood-type estimator, we need to control the asymptotic behaviors of some functionals of the sampling scheme. Though it is difficult to directly control those in general, we study tractable sufficient conditions when the sampling scheme is generated by mixing processes.

Keywords Asymptotic efficiency · Diffusion processes · Local asymptotic normality · Maximum-likelihood-type estimation · Nonsynchronous observations

1 Introduction

Given a probability space $(\Omega, \mathcal{F}, P)$ with a right-continuous filtration $\mathbf{F} = \{\mathcal{F}_t\}_{t\geq 0}$, let $X^{(\alpha)} = \{X_t^{(\alpha)}\}_{t\geq 0} = \{(X_t^{(\alpha),1}, X_t^{(\alpha),2})\}_{t\geq 0}$ be a two-dimensional $\mathbf{F}$-adapted process satisfying the following stochastic differential equation:

✉ Teppei Ogihara
ogihara@mist.i.u-tokyo.ac.jp

1 Graduate School of Information Science and Technology, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan

2 The Institute of Statistical Mathematics, Tokyo, Japan

$$\mathrm{d}X_t^{(\alpha)} = \mu_t(\theta)\mathrm{d}t + b_t(\sigma)\mathrm{d}W_t, \quad X_0 = x_0, \tag{1.1}$$

where $x_0 \in \mathbb{R}^2$, $\{W_t\}_{0\le t\le T}$ is a two-dimensional standard $\mathbf{F}$-Wiener process, $\{\mu_t(\theta)\}_{t\ge 0}$ and $\{b_t(\sigma)\}_{t\ge 0}$ are deterministic functions with values in $\mathbb{R}^2$ and $\mathbb{R}^{2\times 2}$, respectively, $\alpha = (\sigma, \theta)$, $\sigma \in \Theta_1$, $\theta \in \Theta_2$, and Θ_1 and Θ_2 are bounded open subsets of $\mathbb{R}^{d_1}$ and $\mathbb{R}^{d_2}$, respectively. Let $\alpha_0 = (\sigma_0, \theta_0) \in \Theta_1 \times \Theta_2$ be the true value, and let $X_t = (X_t^1, X_t^2) = X_t^{(\alpha_0)}$. We consider estimation of α_0 when X is observed with nonsynchronous manner, that is, observation times of X^1 and X^2 are different to each other.

The problem of nonsynchronous observations appears in the analysis of high-frequency financial data. If we analyze the intra-day stock price data, we observe stock prices when a new transaction or a new order arrives. Then, the observation times are different for different stocks, and hence, we cannot avoid the problem of nonsynchronous observations. Statistical analysis with such data is much more complicated compared to the analysis with synchronous data. Parametric estimation for diffusion processes with synchronous and equidistant observations has been analyzed through quasi-maximum-likelihood methods in Florens-Zmirou (1989), Yoshida (1992, 2011), Kessler (1997), and Uchida and Yoshida (2012). Related to the estimation problem for nonsynchronously observed diffusion processes, estimators for the quadratic covariation have been actively studied. Hayashi and Yoshida (2005, 2008, 2011) and Malliavin and Mancino (2002, 2009) have independently constructed consistent estimators under nonsynchronous observations. There are also studies of covariation estimation under the simultaneous presence of microstructure noise and nonsynchronous observations (Barndorff-Nielsen et al., 2011; Bibinger et al., 2014; Christensen et al., 2010, and so on). For parametric estimation with nonsynchronous observations, Ogihara and Yoshida (2014) have constructed maximum-likelihood-type and Bayes-type estimators and have shown the consistency and the asymptotic mixed normality of the estimators when the terminal time point T_n is fixed and the observation frequency goes to infinity. Ogihara (2015) have shown local asymptotic mixed normality for the model in Ogihara and Yoshida (2014), and the maximum-likelihood-type and Bayes-type estimators have been shown to be asymptotically efficient. On the other hand, we need to consider asymptotic theory that the terminal time point T_n goes to infinity to consistently estimate the parameter θ in the drift term. To the best of the author's knowledge, there are no studies of the asymptotic theory of parametric estimation for nonsynchronously observed diffusion processes when $T_n \to \infty$.

In this work, we consider the asymptotic theory for nonsynchronously observed diffusion processes when $T_n \to \infty$, and construct maximum-likelihood-type estimators for the parameter σ in the diffusion part and the parameter θ in the drift part. We show the consistency and the asymptotic normality of the estimators. Moreover, we show local asymptotic normality of the statistical model, and we obtain asymptotic efficiency of our estimator as a consequence. Our estimator is constructed based on the quasi-likelihood function that is similarly defined to the one in Ogihara and Yoshida (2014), though we need some modification to deal with the drift part. To investigate asymptotic theory for the maximum-likelihood-type estimator, we need to specify the limit of the quasi-likelihood function. Then, we need to assume some conditions for the asymptotic behavior of the sampling scheme. In Ogihara and Yoshida (2014), for a matrix

$$G = \left\{ \frac{(S_i^{n,1} \wedge S_j^{n,2} - S_{j-1}^{n,2} \vee S_{i-1}^{n,1}) \vee 0}{|S_i^{n,1} - S_{i-1}^{n,1}|^{1/2} |S_j^{n,2} - S_{j-1}^{n,2}|^{1/2}} \right\}_{i,j}$$

generated by the sampling scheme, the existence of the probability limit of $n^{-1}\mathrm{tr}((GG^\top)^p)$ $(p \in \mathbb{Z}_+)$ is required, where $(S_i^{n,l})_i$ are observation times of X^l and $\top$ denotes transpose of a vector or a matrix. Since we consider the different asymptotics, the asymptotic behavior of the quasi-likelihood function is different from that in Ogihara and Yoshida (2014). We also need to consider estimation for the drift parameter θ. Then, we need other assumptions for the asymptotic behavior of the sampling scheme [Assumption (A5)]. Though these conditions for the sampling scheme are difficult to check directly, we study tractable sufficient conditions in Sect. 2.4.

The rest of this paper is organized as follows. In Sect. 2, we introduce our model settings and the assumptions for main results. Our estimator is constructed in Sect. 2.1, and the asymptotic normality of the estimator is given in Sect. 2.2. Section 2.3 deals with local asymptotic normality of our model and asymptotic efficiency of the estimator. Tractable sufficient conditions for the assumptions of the sampling scheme are given in Sect. 2.4. Section 3 contains the proofs of main results. Preliminary results are collected in Sect. 3.1. Section 3.2 is for the consistency of the estimator for σ, Sect. 3.3 is for the asymptotic normality of the estimator for σ, Sect. 3.4 is for the consistency of the estimator for θ, and Sect. 3.5 is for the asymptotic normality of the estimator for θ. Other proofs are collected in Sect. 3.6.

2 Main results

2.1 Setting and parameter estimation

Let $\mathbb{N}$ be the set of all positive integers. For $l \in \{1, 2\}$, let the observation times $\{S_i^{n,l}\}_{i=0}^{M_l}$ be strictly increasing random times with respect to i, and satisfy $S_0^{n,l} = 0$ and $S_{M_l}^{n,l} = nh_n$, where M_l is a random positive integer depending on n and $(h_n)_{n=1}^\infty$ is a sequence of positive numbers satisfying

$$h_n \to 0, \quad n^{1-\epsilon_0} h_n \to \infty, \quad nh_n^2 \to 0 \tag{2.1}$$

as $n \to \infty$ for some $\epsilon_0 > 0$. Intuitively, n is of the order of the number of observations and h_n is of the order of the length of the observation intervals. More precise assumptions of observation times are given in (A2), (A4), and (A5) later. We assume that $\{S_i^{n,l}\}_{0 \le i \le M_l, l=1,2}$ is independent of $\mathcal{F}_T$, and its distribution does not depend on α. We consider nonsynchronous observations of X, that is, we observe $\{S_i^{n,l}\}_{0 \le i \le M_l, l=1,2}$ and $\{X^l_{S_i^{n,l}}\}_{0 \le i \le M_l, l=1,2}$. In particular, we consider the *nonendogenous* observation times.

We denote by $\|\cdot\|$ the operator norm with respect to the Euclidean norm for a matrix. We often regard a p-dimensional vector v as a $p \times 1$ matrix. For $j \in \mathbb{N}$, we denote $\partial_z = \frac{\partial}{\partial z}$ for a variable $z \in \mathbb{R}^j$, and denote $\partial_z^l = (\partial_{z_{i_1}} \cdots \partial_{z_{i_l}})_{i_1,\dots,i_l=1}^j$ for $l \in \mathbb{N}$. For functions f and g, we often use shorthand notation $\partial_z f \partial_z g = (\partial_z f (\partial_z g)^\top + \partial_z g (\partial_z f)^\top)/2$. For a set A in a topological space, let $\mathrm{clos}(A)$ denote the closure of A. For a matrix A, $[A]_{ij}$ denotes its (i, j) element. For a vector $v = (v_j)_{j=1}^K$, we denote $[v]_j = v_j$, and $\mathrm{diag}(v)$ denotes a $K \times K$ diagonal matrix with elements $[\mathrm{diag}(v)]_{jj} = v_j$.

Let $M = M_1 + M_2$. For $1 \le i \le M$, let

$$\varphi(i) = \begin{cases} i, & \text{if } i \le M_1, \\ i - M_1, & \text{if } i > M_1, \end{cases} \quad \psi(i) = \begin{cases} 1, & \text{if } i \le M_1, \\ 2, & \text{if } i > M_1. \end{cases}$$

For a two-dimensional stochastic process $(U_t)_{t \ge 0} = ((U_t^1, U_t^2))_{t \ge 0}$, let $\Delta_i^l U = U^l_{S_i^{n,l}} - U^l_{S_{i-1}^{n,l}}$, and let $\Delta^l U = (\Delta_i^l U)_{1 \le i \le M_l}$ and $\Delta_i U = \Delta_{\varphi(i)}^{\psi(i)} U$ for $1 \le i \le M$. Let $\Delta U = ((\Delta^1 U)^\top, (\Delta^2 U)^\top)^\top$. Let $|K| = b - a$ for an interval $K = (a, b]$. Let $I_i^l = (S_{i-1}^{n,l}, S_i^{n,l}]$ for $1 \le i \le M_l$, and let $I_i = I_{\varphi(i)}^{\psi(i)}$ for $1 \le i \le M$. We denote a unit matrix of size k by $\mathcal{E}_k$.

Let $\tilde{\Sigma}_i^l(\sigma) = \int_{I_i^l} [b_t b_t^\top(\sigma)]_{ll} \mathrm{d}t$ and $\tilde{\Sigma}_{i,j}^{1,2}(\sigma) = \int_{I_i^1 \cap I_j^2} [b_t b_t^\top(\sigma)]_{12} \mathrm{d}t$, and let $\tilde{\Sigma}_i = \tilde{\Sigma}_{\varphi(i)}^{\psi(i)}$ for $1 \le i \le M$. By setting $\tilde{\mathcal{D}} = \mathrm{diag}((\tilde{\Sigma}_i)_{1 \le i \le M})$

$$\tilde{G}(\sigma) = \left\{ \frac{\tilde{\Sigma}_{i,j}^{1,2}}{\sqrt{\tilde{\Sigma}_i^1} \sqrt{\tilde{\Sigma}_j^2}} (\sigma) \right\}_{1 \le i \le M_1, 1 \le j \le M_2},$$

we can calculate the covariance matrix of ΔX as

$$S_n(\sigma) = \tilde{\mathcal{D}}^{1/2} \begin{pmatrix} \mathcal{E}_{M_1} & \tilde{G}(\sigma) \\ \tilde{G}^\top(\sigma) & \mathcal{E}_{M_2} \end{pmatrix} \tilde{\mathcal{D}}^{1/2}. \tag{2.2}$$

As we will see later, we can ignore the term related to $\mu_t(\theta)$ (drift term) when we consider estimation of σ, because this term converges to zero very fast. Therefore, we first construct an estimator for σ, and then construct an estimator for θ. Such adaptive estimation can speed up the calculation.

We define the quasi-likelihood function $H_n^1(\sigma)$ for σ as follows:

$$H_n^1(\sigma) = -\frac{1}{2} \Delta X^\top S_n^{-1}(\sigma) \Delta X - \frac{1}{2} \log \det S_n(\sigma).$$

Then, the maximum-likelihood-type estimator for σ is defined by

$$\hat{\sigma}_n \in \underset{\sigma \in \mathrm{clos}(\Theta_1)}{\mathrm{argmax}} \; H_n^1(\sigma).$$

We consider estimation for θ next. Let $V(\theta) = (V_t(\theta))_{t \ge 0}$ be a two-dimensional stochastic process defined by $V_t(\theta) = (\int_0^t \mu_s^1(\theta)^\top \mathrm{d}s, \int_0^t \mu_s^2(\theta)^\top \mathrm{d}s)^\top$. Let $\bar{X}(\theta) =$

$\Delta X - \Delta V(\theta)$. We define the quasi-likelihood function $H_n^2(\theta)$ for θ as follows:

$$H_n^2(\theta) = -\frac{1}{2}\bar{X}(\theta)^\top S_n^{-1}(\hat{\sigma}_n)\bar{X}(\theta).$$

Then, the maximum-likelihood-type estimator for θ is defined by

$$\hat{\theta}_n \in \underset{\theta \in \mathrm{clos}(\Theta_2)}{\mathrm{argmax}}\ H_n^2(\theta).$$

The quasi-(log-)likelihood function H_n^1 is defined in the same way as that in Ogihara and Yoshida (2014). Since ΔX follows normal distribution, we can construct such a Gaussian quasi-likelihood function even for the nonsynchronous data. When the coefficients are random, though the distribution of ΔX is not Gaussian, such Gaussian-type quasi-likelihood function is still valid due to the local Gaussian property of diffusion processes. The Gaussian mean that comes from the drift part is ignored when we construct the quasi-likelihood H_n^1. When we estimate the parameter θ for the drift part, we subtract the mean in $\bar{X}(\theta)$ to construct the quasi-likelihood function H_n^2. Since the effect of the drift term on the estimation of σ is small, it works well to estimate σ in this way and then plug in $\hat{\sigma}_n$ to S_n to construct the estimator for θ. Thus, we can speed up the calculation by separating the estimation for σ and θ.

Remark 2.1 $H_n^1(\sigma)$ and $H_n^2(\theta)$ are well defined only if $\det S_n(\sigma) > 0$ and $\det S_n(\hat{\sigma}_n) > 0$, respectively. For the covariance matrix S_n of nonsynchronous observations ΔX, it is not trivial to check these conditions. Proposition 1 in Section 2 of Ogihara and Yoshida (2014) shows that these conditions are satisfied if $b_t(\sigma)$ is continuous on $[0, \infty) \times \mathrm{clos}(\Theta_1)$ and $\inf_{t,\sigma} \det(b_t b_t^\top(\sigma)) > 0$. We assume such conditions in our setting (Assumption (A1) in Sect. 2.2).

Remark 2.2 As seen in Ogihara and Yoshida (2014), the quasi-likelihood analysis for nonsynchronously observed diffusion processes becomes much more complicated compared to synchronous observations. In this work, estimation for the drift parameter θ is added, and hence, we consider nonrandom drift and diffusion coefficients to avoid overcomplication. For general diffusion processes with the random drift and diffusion coefficients, we need to set predictable coefficients to use the martingale theory. However, the quasi-likelihood function loses a Markov property with nonsynchronous observations and the coefficients in the quasi-likelihood function contain randomness of future time. Then, we need to approximate the coefficients by predictable functions. This operation is particularly complicated. Moreover, approximating the true likelihood function by the quasi-likelihood function is much more difficult problem when we show local asymptotic normality and asymptotic efficiency of the estimators. Therefore, we left asymptotic theory under general random drift and diffusion coefficients as a future work.

2.2 Asymptotic normality of the estimator

In this section, we state the assumptions of our main results, and state the asymptotic normality of the estimator.

For $m \in \mathbb{N}$, an open subset $U \subset \mathbb{R}^m$ is said to admit Sobolev's inequality if, for any $p > m$, there exists a positive constant C depending on U and p, such that $\sup_{x\in U} |u(x)| \leq C \sum_{k=0,1}(\int |\partial_x^k u(x)|^p \mathrm{d}x)^{1/p}$ for any $u \in C^1(U)$. This is the case when U has a Lipschitz boundary. We assume that Θ, Θ_1, and Θ_2 admit Sobolev's inequality.

Let $\Sigma_t(\sigma) = b_t b_t^\top(\sigma)$, and let

$$\rho_t(\sigma) = \frac{[\Sigma_t]_{12}}{[\Sigma_t]_{11}^{1/2}[\Sigma_t]_{22}^{1/2}}(\sigma), \quad B_{l,t}(\sigma) = \frac{[\Sigma_t(\sigma_0)]_{ll}}{[\Sigma_t(\sigma)]_{ll}}.$$

Let $\rho_{t,0} = \rho_t(\sigma_0)$.

Assumption (A1). There exists a positive constant c_1, such that $c_1\mathcal{E}_2 \leq \Sigma_t(\sigma)$ for any $t \in [0,\infty)$ and $\sigma \in \Theta_1$. For $k \in \{0,1,2,3,4\}$, $\partial_\theta^k \mu_t(\theta)$ and $\partial_\sigma^k b_t(\sigma)$ exist and are continuous with respect to (t,σ,θ) on $[0,\infty) \times \mathrm{clos}(\Theta_1) \times \mathrm{clos}(\Theta_2)$. For any $\epsilon > 0$, there exist $\delta > 0$ and $K > 0$, such that

$$|\partial_\theta^k \mu_t(\theta)| + |\partial_\sigma^k b_t(\sigma)| \leq K,$$
$$|\partial_\theta^k \mu_t(\theta) - \partial_\theta^k \mu_s(\theta)| + |\partial_\sigma^k b_t(\sigma) - \partial_\sigma^k b_s(\sigma)| \leq \epsilon$$

for any $k \in \{0,1,2,3,4\}$, $\sigma \in \Theta_1$, $\theta \in \Theta_2$, and $t, s \geq 0$ satisfying $|t-s| < \delta$. Let $r_n = \max_{i,l} |I_i^l|$.

Assumption (A2). $r_n \overset{P}{\to} 0$ as $n \to \infty$.

Assumption (A3). For any $l \in \{1,2\}$, $i_1 \in \mathbb{Z}_+$, $i_2 \in \{0,1\}$, $i_3 \in \{0,1,2,3,4\}$, $k_1, k_2 \in \{0,1,2\}$ satisfying $k_1 + k_2 = 2$, and any polynomial function $F(x_1,\ldots,x_{14})$ of degree equal to or less than 6, there exist continuous functions $\Phi_{i_1,i_2}^{1,F}(\sigma)$, $\Phi_{l,i_3}^2(\sigma)$ and $\Phi_{i_1,i_3}^{3,k_1,k_2}(\theta)$ on $\mathrm{clos}(\Theta_1)$ and $\mathrm{clos}(\Theta_2)$, such that

$$\frac{1}{T}\int_0^T F((\partial_\sigma^k B_{l,t}(\sigma))_{0\leq k\leq 4, l=1,2},$$
$$(\partial_\sigma^{k'}\rho_t(\sigma))_{k'=1}^4)\rho_t(\sigma)^{i_1}\rho_{t,0}^{i_2}\mathrm{d}t \to \Phi_{i_1,i_2}^{1,F}(\sigma),$$
$$\frac{1}{T}\int_0^T \partial_\sigma^{i_3}\log B_{l,t}(\sigma)\mathrm{d}t \to \Phi_{l,i_3}^2(\sigma),$$
$$\frac{1}{T}\int_0^T \partial_\theta^{i_3}(\phi_{1,t}^{k_1}\phi_{2,t}^{k_2})(\theta)\rho_{t,0}^{i_1}\mathrm{d}t \to \Phi_{i_1,i_3}^{3,k_1,k_2}(\theta)$$

as $T \to \infty$ for $\sigma \in \mathrm{clos}(\Theta_1)$, $\theta \in \mathrm{clos}(\Theta_2)$, where $\phi_{l,t}(\theta) = [\Sigma_t(\sigma_0)]_{ll}^{-1/2}(\mu_t^l(\theta) - \mu_t^l(\theta_0))$.

Assumption (A1) and the Ascoli–Arzelà theorem yield that the convergences in (A3) can be replaced by uniform convergence with respect to σ and θ (the left-hand sides of the above equations become relatively compact, and then, any uniformly convergent subsequence converges to the right-hand sides due to the pointwise convergence assumptions). Assumption (A3) is satisfied if $\mu_t(\theta)$ and $b_t(\sigma)$ are independent of t, or are periodic functions with respect to t having a common period (when the period does not depend on σ nor θ). Let $\mathfrak{S}$ be the set of all partitions $(s_k)_{k=0}^\infty$ of $[0, \infty)$ satisfying $\sup_{k\geq 1}|s_k - s_{k-1}| \leq 1$ and $\inf_{k\geq 1}|s_k - s_{k-1}| > 0$. For $(s_k)_{k=0}^\infty \in \mathfrak{S}$, let $M_{l,k} = \#\{i; \sup I_i^l \in (s_{k-1}, s_k]\}$ and $q_n = \max\{k; s_k \leq nh_n\}$, and let $\mathcal{E}_{(k)}^l$ be an $M_l \times M_l$ matrix satisfying $[\mathcal{E}_{(k)}^l]_{ij} = 1$ if $i = j$ and $\sup I_i^l \in (s_{k-1}, s_k]$, and otherwise, $[\mathcal{E}_{(k)}^l]_{ij} = 0$. Let

$$G = \left\{ \frac{|I_i^1 \cap I_j^2|}{|I_i^1|^{1/2}|I_j^2|^{1/2}} \right\}_{1\leq i\leq M_1, 1\leq j\leq M_2}.$$

Assumption (A4). There exist positive constants a_0^1 and a_0^2, such that $\{h_n M_{l,q_n+1}\}_{n=1}^\infty$ is P-tight and

$$\max_{1\leq k\leq q_n} |h_n M_{l,k} - a_0^l(s_k - s_{k-1})| \xrightarrow{P} 0$$

for $l \in \{1, 2\}$ and any partition $(s_k)_{k=0}^\infty \in \mathfrak{S}$. Moreover, for any $p \in \mathbb{N}$, there exists a nonnegative constant a_p^1, such that

$$\max_{1\leq k\leq q_n} |h_n \mathrm{tr}(\mathcal{E}_{(k)}^1 (GG^\top)^p) - a_p^1(s_k - s_{k-1})| \xrightarrow{P} 0$$

as $n \to \infty$ for any partition $(s_k)_{k=0}^\infty \in \mathfrak{S}$. Let $\mathfrak{I}_l = (|I_i^l|^{1/2})_{i=1}^{M_l}$.

Assumption (A5). For $p \in \mathbb{Z}_+$, there exist nonnegative constants $f_p^{1,1}$, $f_p^{1,2}$, and $f_p^{2,2}$, such that $\{|\mathcal{E}_{(q_n+1)}^l \mathfrak{I}_l|\}_{n=1}^\infty$ is P-tight for $l \in \{1, 2\}$, and

$$\max_{1\leq k\leq q_n} |\mathfrak{I}_1 \mathcal{E}_{(k)}^1 (GG^\top)^p \mathfrak{I}_1 - f_p^{1,1}(s_k - s_{k-1})| \xrightarrow{P} 0,$$

$$\max_{1\leq k\leq q_n} |\mathfrak{I}_1 \mathcal{E}_{(k)}^1 (GG^\top)^p G\mathfrak{I}_2 - f_p^{1,2}(s_k - s_{k-1})| \xrightarrow{P} 0,$$

$$\max_{1\leq k\leq q_n} |\mathfrak{I}_2 \mathcal{E}_{(k)}^2 (G^\top G)^p \mathfrak{I}_2 - f_p^{2,2}(s_k - s_{k-1})| \xrightarrow{P} 0$$

as $n \to \infty$ for any partition $(s_k)_{k=0}^\infty \in \mathfrak{S}$.

Assumption (A4) corresponds to [A3′] in Ogihara and Yoshida (2014). The functionals in (A4) and (A5) appear in H_n^1 and H_n^2, and hence, we cannot specify the limits of H_n^1 and H_n^2 unless we assume existence of the limits of these functionals. It is difficult to directly check (A4) and (A5) for concrete statistical experiments with general sampling schemes. We study sufficient conditions for these conditions in Sect. 2.4.

Assumption (A6). The constant a_1^1 in (A4) is positive, and there exist positive constants c_2 and c_3, such that

$$\limsup_{T\to\infty}\left(\frac{1}{T}\int_0^T \|\Sigma_t(\sigma)-\Sigma_t(\sigma_0)\|^2 \mathrm{d}t\right)\geq c_2|\sigma-\sigma_0|^2,$$
$$\limsup_{T\to\infty}\left(\frac{1}{T}\int_0^T |\mu_t(\theta)-\mu_t(\theta_0)|^2 \mathrm{d}t\right)\geq c_3|\theta-\theta_0|^2$$

for any $\sigma\in \mathrm{clos}(\Theta_1)$ and $\theta\in\mathrm{clos}(\Theta_2)$.

Assumption (A6) is necessary to identify the parameter σ and θ from the data. For $p<q$

$$\mathrm{tr}(\mathcal{E}^1_{(k)}(GG^\top)^q)\leq \mathrm{tr}(\mathcal{E}^1_{(k)}(GG^\top)^p)\|(GG^\top)^{q-p}\|\leq \mathrm{tr}(\mathcal{E}^1_{(k)}(GG^\top)^p) \tag{2.3}$$

by Lemma 3.3 later and Lemma A.1 in Ogihara (2018). Then, a_p^1 is monotone non-increasing with respect to p. This implies that $a_p^1=0$ for any $p\in\mathbb{N}$ if $a_1^1=0$. In this case, the non-diagonal components of the covariance matrix S_n are negligible in the limit. Then, we cannot consistently estimate the parameter in $\rho_t(\sigma)$. This is why, we need the assumption $a_1^1>0$ (see Proposition 3.9 and the following discussion to obtain the consistency).

Let $\mathcal{A}(\rho)=\sum_{p=1}^\infty a_p^1\rho^{2p}$ for $\rho\in(-1,1)$. Then, (2.3) implies that $\mathcal{A}(\rho)$ is finite. Moreover, (A5) yields

$$\begin{aligned}f_p^{1,1}&=(nh_n)^{-1}\sum_{k=1}^{q_n}\mathfrak{I}_1^\top\mathcal{E}^1_{(k)}(GG^\top)^p\mathfrak{I}_1+o_p(1)\\&=(nh_n)^{-1}\mathfrak{I}_1^\top(GG^\top)^p\mathfrak{I}_1+o_p(1)\\&\leq\|(GG^\top)^p\|(nh_n)^{-1}|\mathfrak{I}_1|^2+o_p(1)\\&\leq 1+o_p(1),\end{aligned}$$

which implies $f_p^{1,1}\leq 1$. Similarly, we have $f_p^{1,2}\leq 1$ and $f_p^{2,2}\leq 1$. Let $\partial_\sigma^k B_{l,t,0}=\partial_\sigma^k B_{l,t}(\sigma_0)$, and let

$$\begin{aligned}\gamma_{1,t}&=\mathcal{A}(\rho_{t,0})\left(\frac{\partial_\sigma\rho_{t,0}}{\rho_{t,0}}-\partial_\sigma B_{1,t,0}-\partial_\sigma B_{2,t,0}\right)^2-\partial_\rho\mathcal{A}(\rho_{t,0})\frac{(\partial_\sigma\rho_{t,0})^2}{\rho_{t,0}}-2\sum_{l=1}^2(a_0^l\\&\quad+\mathcal{A}(\rho_{t,0}))(\partial_\sigma B_{l,t,0})^2,\end{aligned}$$

and let $\Gamma_1=\lim_{T\to\infty}T^{-1}\int_0^T\gamma_{1,t}\mathrm{d}t$, which exists under (A1), (A3), and (A4). Let

$$\Gamma_2=\lim_{T\to\infty}\frac{1}{T}\int_0^T\sum_{p=0}^\infty\rho_{t,0}^{2p}\left\{\sum_{l=1}^2 f_p^{ll}(\partial_\theta\phi_{l,t})^2(\theta_0)\right.$$

$$-2\rho_{t,0} f_p^{12} \partial_\theta \phi_{1,t} \partial_\theta \phi_{2,t}(\theta_0) \Big\} \mathrm{d}t,$$

which exists under (A1), (A3), and (A5). Let $T_n = nh_n$ and

$$\Gamma = \begin{pmatrix} \Gamma_1 & 0 \\ 0 & \Gamma_2 \end{pmatrix}.$$

Theorem 2.3 *Assume* (A1)–(A6). *Then*, Γ *is positive definite*, *and*

$$(\sqrt{n}(\hat{\sigma}_n - \sigma_0), \sqrt{T_n}(\hat{\theta}_n - \theta_0)) \xrightarrow{d} N(0, \Gamma^{-1})$$

as $n \to \infty$.

2.3 Local asymptotic normality

Let $\alpha_0 \in \Theta$, $\Theta \subset \mathbb{R}^d$, and $\{P_{\alpha,n}\}_{\alpha \in \Theta}$ be a family of probability measures defined on a measurable space $(\mathcal{X}_n, \mathcal{A}_n)$ for $n \in \mathbb{N}$, where Θ is an open subset of $\mathbb{R}^d$. As usual, we shall refer to $\mathrm{d}P_{\alpha_2,n}/\mathrm{d}P_{\alpha_1,n}$ the derivative of the absolutely continuous component of the measure $P_{\alpha_2,n}$ with respect to measure $P_{\alpha_1,n}$ at the observation x as the likelihood ratio. The following definition of local asymptotic normality is Definition 2.1 in Chapter II of Ibragimov and Has'minskiĭ (1981).

Definition 2.4 A family $P_{\alpha,n}$ is called locally asymptotically normal (LAN) at point $\alpha_0 \in \Theta$ as $n \to \infty$ if for some nondegenerate $d \times d$ matrix ϵ_n and any $u \in \mathbb{R}^d$, the representation

$$\log \frac{\mathrm{d}P_{\alpha_0+\epsilon_n u,n}}{\mathrm{d}P_{\alpha_0,n}} - (u^\top \Delta_n - |u|^2/2) \to 0$$

in $P_{\alpha_0,n}$-probability as $n \to \infty$, where

$$\mathcal{L}(\Delta_n | P_{\alpha_0,n}) \to N(0, \mathcal{E}_d)$$

as $n \to \infty$, and $\mathcal{L}(\cdot | P_{\alpha,n})$ denotes the distribution with respect to $P_{\alpha,n}$.

Let $\Theta = \Theta_1 \times \Theta_2$. For $\alpha \in \Theta$, let $P_{\alpha,n}$ be the probability measure generated by the observations $\{S_i^{n,l}\}_{i,l}$ and $\{X^{(\alpha),l}_{S_i^{n,l}}\}_{i,l}$.

Theorem 2.5 *Assume* (A1)–(A6). *Then*, $\{P_{\alpha,n}\}_{\alpha,n}$ *satisfies the LAN property at* $\alpha = \alpha_0$ *with*

$$\epsilon_n = \begin{pmatrix} n^{-1/2}\Gamma_1^{-1/2} & 0 \\ 0 & T_n^{-1/2}\Gamma_2^{-1/2} \end{pmatrix}.$$

The proof is left to Sect. 3.6. Theorem 11.2 in Chapter II of Ibragimov and Has'minskiĭ (1981) gives lower bounds of estimation errors for any regular estimator of parameters under the LAN property. Then, the optimal asymptotic variance of $\epsilon_n^{-1}(U_n - \alpha_0)$ for regular estimator U_n is $\mathcal{E}_d$. We will show that $(\hat{\sigma}_n, \hat{\theta}_n)$ is regular in Remark 3.18. Therefore, Theorem 2.5 ensures that our estimator $(\hat{\sigma}_n, \hat{\theta}_n)$ is asymptotically efficient in this sense under the assumptions of the theorem.

2.4 Sufficient conditions for the assumptions

It is not easy to directly check Assumptions (A4) and (A5) for general random sampling schemes (even for a sampling scheme generated by simple Poisson processes given in Example 2.6). In this section, we study tractable sufficient conditions for these assumptions. The proofs of the results in this section are left to Sect. 3.6.

Let $q > 0$ and $\mathcal{N}_t^{n,l} = \sum_{i=1}^{M_l} 1_{\{S_i^{n,l} \leq t\}}$. We consider the following conditions for the point process $\mathcal{N}_t^{n,l}$.

Assumption (B1-q).

$$\sup_{n\geq 1} \max_{l\in\{1,2\}} \sup_{0\leq t\leq (n-1)h_n} E[(\mathcal{N}_{t+h_n}^{n,l} - \mathcal{N}_t^{n,l})^q] < \infty.$$

Assumption (B2-q).

$$\limsup_{u\to\infty} \sup_{n\geq 1} \max_{l\in\{1,2\}} \sup_{0\leq t\leq nh_n - uh_n} u^q P(\mathcal{N}_{t+uh_n}^{n,l} - \mathcal{N}_t^{n,l} = 0) < \infty.$$

Example 2.6 Let $(\bar{\mathcal{N}}_t^1, \bar{\mathcal{N}}_t^2)$ be two independent homogeneous Poisson processes with positive intensities λ_1 and λ_2, respectively, and $\mathcal{N}_t^{n,l} = \bar{\mathcal{N}}^l_{h_n^{-1}t}$, that is, $S_i^{n,l} = \inf\{t \geq 0; \bar{\mathcal{N}}^l_{h_n^{-1}t} \geq i\}$. Even in this simple case, it is not trivial to directly check (A4) and (A5). On the other hand, (B1-q) obviously holds for any $q > 0$. Moreover, (B2-q) holds for any $q > 0$, since

$$\limsup_{u\to\infty} \sup_{n\geq 1} \max_{l\in\{1,2\}} \sup_{0\leq t\leq nh_n - uh_n} u^q P(\mathcal{N}_{t+uh_n}^{n,l} - \mathcal{N}_t^{n,l} = 0) = \lim_{u\to\infty} u^q \mathrm{e}^{-(\lambda_1\wedge\lambda_2)u} = 0.$$

Then, by Corollary 2.12, we can check Assumptions (A2), (A4), and (A5) for this sampling scheme.

To give sufficient conditions for (A4) and (A5), we consider mixing properties of $\mathcal{N}^{n,l}$. That is, we assume conditions for the following mixing coefficient α_k^n. Let

$$\mathcal{G}_{i,j}^n = \sigma(\mathcal{N}_t^{n,l} - \mathcal{N}_s^{n,l}; ih_n \leq s < t \leq jh_n, l = 1, 2) \quad (0 \leq i, j \leq n),$$

and let

$$\alpha_k^n = 0 \vee \sup_{1\leq i,j\leq n-1, j-i\geq k} \sup_{A\in\mathcal{G}_{0,i}^n} \sup_{B\in\mathcal{G}_{j,n}^n} |P(A\cap B) - P(A)P(B)|.$$

Proposition 2.7 *Assume that* (B1-q) *and* (B2-q) *hold and that*

$$\sup_{n\in\mathbb{N}}\sum_{k=0}^{\infty}(k+1)^q\alpha_k^n<\infty \tag{2.4}$$

for any $q>0$. *Moreover, assume that there exist positive constants* a_0^1 *and* a_0^2, *and a nonnegative constant* a_p^1 *for* $p\in\mathbb{N}$, *such that* $\{E[h_nM_{l,q_n+1}]\}_{n=1}^{\infty}$ *is bounded and*

$$\max_{1\le k\le q_n}|h_nE[M_{l,k}]-a_0^l(s_k-s_{k-1})|\to 0,$$
$$\max_{1\le k\le q_n}|h_nE[\mathrm{tr}(\mathcal{E}^1_{(k)}(GG^\top)^p)]-a_p^1(s_k-s_{k-1})|\to 0 \tag{2.5}$$

as $n\to\infty$ *for* $p\in\mathbb{Z}_+$, $l\in\{1,2\}$ *and any partition* $(s_k)_{k=0}^{\infty}\in\mathfrak{S}$. *Then*, (A4) *holds.*

Proposition 2.8 *Assume that* (B1-q) *and* (B2-q) *hold and that* (2.4) *is satisfied for any* $q>0$. *Moreover, assume that there exist nonnegative constants* $f_p^{1,1}$, $f_p^{1,2}$, *and* $f_p^{2,2}$ *for* $p\in\mathbb{Z}_+$, *such that* $\{E[|\mathcal{E}^l_{(q_n+1)}\mathfrak{I}_l|]\}_{n=1}^{\infty}$ *is bounded and*

$$\max_{1\le k\le q_n}|E[\mathfrak{I}_1\mathcal{E}^1_{(k)}(GG^\top)^p\mathfrak{I}_1]-f_p^{1,1}(s_k-s_{k-1})|\to 0,$$
$$\max_{1\le k\le q_n}|E[\mathfrak{I}_1\mathcal{E}^1_{(k)}(GG^\top)^pG\mathfrak{I}_2]-f_p^{1,2}(s_k-s_{k-1})|\to 0,$$
$$\max_{1\le k\le q_n}|E[\mathfrak{I}_2\mathcal{E}^2_{(k)}(G^\top G)^p\mathfrak{I}_2]-f_p^{2,2}(s_k-s_{k-1})|\to 0 \tag{2.6}$$

as $n\to\infty$ *for* $l\in\{1,2\}$, $p\in\mathbb{Z}_+$ *and any partition* $(s_k)_{k=0}^{\infty}\in\mathfrak{S}$. *Then*, (A5) *holds.*

Proposition 2.9 *Assume that there exists* $q>0$, *such that* (A4) *and* (B2-q) *hold*, $\{\mathcal{N}^{n,l}_{t+h_n}-\mathcal{N}^{n,l}_t\}_{0\le t\le T_n-h_n,l\in\{1,2\},n\in\mathbb{N}}$ *is* P*-tight, and* $\sum_{k=1}^{\infty}k\alpha_k^n<\infty$. *Then*, $a_1^1>0$.

In the following, let $(\bar{\mathcal{N}}^l_t)_{t\ge 0}$ be an exponential α-mixing point process for $l\in\{1,2\}$. Assume that the distribution of $(\bar{\mathcal{N}}^l_{t+t_k}-\bar{\mathcal{N}}^l_{t+t_{k-1}})_{1\le k\le K,l=1,2}$ does not depend on $t\ge 0$ for any $K\in\mathbb{N}$ and $0\le t_0<t_1<\cdots<t_K$.

Lemma 2.10 *Let* $\mathcal{N}^{n,l}_t=\bar{\mathcal{N}}^l_{h_n^{-1}t}$ *for* $0\le t\le nh_n$ *and* $l\in\{1,2\}$. *Then*, (2.4) *is satisfied for any* $q>2$, *and there exist constants* a_0^1, a_0^2, *and* $a_p^1=a_p^2$ *for* $p\in\mathbb{N}$, *such that* (2.5) *holds and* $\{E[h_nM_{l,q_n+1}]\}_{n=1}^{\infty}$ *is bounded for any* $(s_k)_{k=0}^{\infty}\in\mathfrak{S}$. *Moreover, there exist nonnegative constants* $f_p^{1,1}$, $f_p^{1,2}$, *and* $f_p^{2,2}$ *for* $p\in\mathbb{Z}_+$, *such that* (2.6) *holds and* $\{E[|\mathcal{E}^l_{(q_n+1)}\mathfrak{I}_l|\}_{n=1}^{\infty}$ *is bounded for* $l\in\{1,2\}$ *and any* $(s_k)_{k=0}^{\infty}\in\mathfrak{S}$.

Proposition 2.11 (Proposition 8 in Ogihara & Yoshida, 2014) *Let* $q\in\mathbb{N}$. *Assume* (B2-($q+1$)). *Then*, $\sup_n E[h_n^{-q+1}r_n^q]<\infty$. *In particular*, (A2) *holds under* (B2-2).

By the above results, we obtain simple tractable sufficient conditions for the assumptions of the sampling scheme when the observation times are generated by the exponential α-mixing point process $\bar{\mathcal{N}}^l_t$ defined above.

Corollary 2.12 *Let* $\mathcal{N}_t^{n,l} = \bar{\mathcal{N}}^l_{h_n^{-1}t}$ *for* $0 \leq t \leq T_n$ *and* $l \in \{1,2\}$. *Assume that* (B1-q) *and* (B2-q) *hold for any* $q > 0$. *Then*, (A2), (A4), *and* (A5) *hold*, *and* $a_1^1 > 0$.

3 Proofs

3.1 Preliminary results

For a real number a, $[a]$ denotes the maximum integer which is not greater than a. Let $\Pi = \Pi_n = \{S_i^{n,l}\}_{1\leq i\leq M_l, l\in\{1,2\}}$. We denote $|x|^2 = \sum_{i_1,\ldots,i_k} |x_{i_1,\ldots,i_k}|^2$ for $x = \{x_{i_1,\ldots,i_k}\}_{i_1,\ldots,i_k}$ with $k \in \mathbb{N}$ and $x_{i_1,\ldots,i_k} \in \mathbb{R}$. For a matrix $A = (A_{ij})_{ij}$, Abs(A) denotes the matrix $(|A_{ij}|)_{ij}$. C denotes generic positive constant whose value may vary depending on context. We often omit the parameters σ and θ in general functions $f(\sigma)$ and $g(\theta)$.

For a sequence p_n of positive numbers, let us denote by $\{\bar{R}_n(p_n)\}_{n\in\mathbb{N}}$ a sequence of random variables (which may also depend on $1 \leq i \leq M$ and $\alpha \in \Theta$) satisfying that $\{\sup_{\alpha,i} E_\Pi[|p_n^{-1}\bar{R}_n(p_n)|^q]\}_{n\in\mathbb{N}}$ is P-tight for any $q \geq 1$, where $E_\Pi[\mathbf{X}] = E[\mathbf{X}|\sigma(\Pi_n)]$ for a random variable $\mathbf{X}$.

For a matrix A and vectors v, w with suitable sizes, we repeatedly use the following inequality:

$$|w^\top Av| \leq |w||Av| \leq \|A\||v||w|.$$

Lemma 3.1 (A special case of Lemma 3.1 in Ogihara and Uehara, 2022) *Let* $(Z_n)_{n\in\mathbb{N}}$ *be nonnegative-valued random variables. Then*

1. $E_\Pi[Z_n] \xrightarrow{P} 0$ *as* $n \to \infty$ *implies that* $Z_n \xrightarrow{P} 0$ *as* $n \to \infty$.
2. P*-tightness of* $(E_\Pi[Z_n])_{n\in\mathbb{N}}$ *implies* P*-tightness of* $(Z_n)_{n\in\mathbb{N}}$.

Let $\bar{V} = V(\theta_0)$, and let

$$\rho_{ij}(\sigma) = \begin{cases} \dfrac{\tilde{\Sigma}^{1,2}_{i,j}}{\sqrt{\tilde{\Sigma}^1_i}\sqrt{\tilde{\Sigma}^2_j}[G]_{ij}}, & \text{if } |I_i^1 \cap I_j^2| \neq \emptyset, \\ 0, & \text{otherwise.} \end{cases}$$

Let $\bar{\rho}_n = \sup_\sigma(\max_{i,j}|\rho_{i,j}(\sigma)| \vee \sup_t |\rho_t(\sigma)|)$, and let

$$\dot{S}_n = \begin{pmatrix} \mathcal{E}_{M_1} & -\bar{\rho}_n G \\ -\bar{\rho}_n G^\top & \mathcal{E}_{M_2} \end{pmatrix}. \tag{3.1}$$

Let $\Delta^l_{i,t}U = U^l_{t\wedge S_i^{n,l}} - U^l_{t\wedge S_{i-1}^{n,l}}$, and let $\Delta_{i,t}U = \Delta^{\psi(i)}_{\varphi(i),t}U$ for $t \geq 0$ and a two-dimensional stochastic process $(U_t)_{t\geq 0} = ((U_t^1, U_t^2))_{t\geq 0}$.

Under (A4), we have

$$h_n M_l = h_n \sum_{k=1}^{q_n+1} M_{l,k} = a_0^l n h_n + o_p(nh_n).$$

Then, we obtain

$$M_l = a_0^l n + o_p(n). \tag{3.2}$$

Lemma 3.2 *Assume* (A1). *Then, for any* $p \geq 1$, *there exist positive constants* C_p *(depending on* p*) and* C, *such that*

$$\sup_\theta |\Delta_i^l V(\theta)| \leq C|I_i^l|, \quad E_\Pi[|\Delta_i^l X|^p]^{1/p} \leq C_p(|I_i^l| + \sqrt{|I_i^l|})$$

for $l \in \{1, 2\}$ *and* $1 \leq i \leq M_l$.

Proof Since $\mu_t^l(\theta)$ and $[b_t b_t(\sigma_0)]_{ll}$ are bounded by (A1), the Burkholder–Davis–Gundy inequality yields

$$\sup_\theta |\Delta_i^l V(\theta)| = \sup_\theta \left| \int_{I_i^l} \mu_t^l(\theta) \mathrm{d}t \right| \leq C|I_i^l|,$$

$$\begin{aligned} E_\Pi[|\Delta_i^l X|^p]^{1/p} &= E_\Pi\left[\left| \int_{I_i^l} \mu_t^l(\theta_0)\mathrm{d}t + \int_{I_i^l} [b_t(\sigma_0)\mathrm{d}W_t]_l \right|^p\right]^{1/p} \\ &\leq C_p|I_i^l| + C_p E_\Pi\left[\left| \int_{I_i^l} [b_t b_t(\sigma_0)]_{ll} \mathrm{d}t \right|^{p/2}\right]^{1/p} \\ &\leq C_p(|I_i^l| + \sqrt{|I_i^l|}). \end{aligned}$$

□

Lemma 3.3 (Lemma 2 in Ogihara & Yoshida, 2014) $\|G\| \vee \|G^\top\| \leq 1$.

Lemma 3.4 $\|\tilde{G}\| \vee \|\tilde{G}^\top\| \leq \bar{\rho}_n$.

Proof Since all the elements of G are nonnegative, we have

$$\begin{aligned} \|\tilde{G}\|^2 = \sup_{|x|=1} |\tilde{G}x|^2 &= \sup_{|x|=1} \sum_i \left(\sum_j \rho_{ij} G_{ij} x_j \right)^2 \\ &\leq \bar{\rho}_n^2 \sup_{|x|=1} \sum_i \left(\sum_j G_{ij} |x_j| \right)^2 \leq \bar{\rho}_n^2 \|G\|^2 \leq \bar{\rho}_n^2. \end{aligned}$$

Since $\|\tilde{G}^\top\| = \|\tilde{G}\|$, we obtain the conclusion. □

Let $\mathcal{D} = \mathrm{diag}(\{|I_i|\}_{i=1}^{M})$. It is difficult to deduce the orders of upper bounds of the operator norms $\|S_n(\sigma)\|$ and $\|S_n^{-1}\|$, because they depend on the maximum and minimum lengths of observation intervals. However, we can deduce the orders of upper bounds for $\tilde{\mathcal{D}}^{-1/2}S_n(\sigma)\tilde{\mathcal{D}}^{-1/2}$ and its inverse. Indeed, we obtain the following estimates, which are repeatedly used in the following sections (we use $\mathcal{D}$ instead of $\tilde{\mathcal{D}}$ to avoid parameter dependence).

Lemma 3.5 *Assume* (A1). *Then, there exists a positive constant* C, *such that* $\|\mathcal{D}^{1/2}\partial_\sigma^k S_n^{-1}(\sigma)\mathcal{D}^{1/2}\| \leq C(1-\bar{\rho}_n)^{-k-1}$ *and* $|[S_n^{-1}(\sigma)]_{ij}| \leq C[\mathcal{D}^{-1/2}\dot{S}_n^{-1}\mathcal{D}^{-1/2}]_{ij}$ *if* $\bar{\rho}_n < 1$, *and* $\|\mathcal{D}^{-1/2}\partial_\sigma^k S_n(\sigma)\mathcal{D}^{-1/2}\| \leq C$ *for any* $\sigma \in \Theta_1$, $1 \leq i, j \leq M$, *and* $k \in \{0, 1, 2, 3, 4\}$.

Proof By (A1) and Lemma 3.3, we have

$$\|\mathcal{D}^{-1/2}\partial_\sigma^k S_n(\sigma)\mathcal{D}^{-1/2}\| \leq C\sum_{j=0}^{k}\left\|\partial_\sigma^j\left\{\mathcal{E}_M + \begin{pmatrix} 0 & \tilde{G} \\ \tilde{G}^\top & 0 \end{pmatrix}\right\}\right\| \leq C.$$

Moreover, by (A1) and Lemma 3.4, we have

$$\|\mathcal{D}^{1/2}S_n^{-1}\mathcal{D}^{1/2}\| \leq C\left\|\left(\mathcal{E}_M + \begin{pmatrix} 0 & \tilde{G} \\ \tilde{G}^\top & 0 \end{pmatrix}\right)^{-1}\right\| \leq C(1-\bar{\rho}_n)^{-1}$$

if $\bar{\rho}_n < 1$.

Using the equation $\partial_\sigma S_n^{-1} = -S_n^{-1}\partial_\sigma S_n S_n^{-1}$, we obtain

$$\begin{aligned}\|\mathcal{D}^{1/2}\partial_\sigma S_n^{-1}\mathcal{D}^{1/2}\| &= \|\mathcal{D}^{1/2}S_n^{-1}\partial_\sigma S_n S_n^{-1}\mathcal{D}^{1/2}\| \\ &\leq \|\mathcal{D}^{1/2}S_n^{-1}\mathcal{D}^{1/2}\|^2\|\mathcal{D}^{-1/2}\partial_\sigma S_n\mathcal{D}^{-1/2}\| \leq C(1-\bar{\rho}_n)^{-2}\end{aligned}$$

if $\bar{\rho}_n < 1$. Similarly, we obtain

$$\|\mathcal{D}^{1/2}\partial_\sigma^k S_n^{-1}\mathcal{D}^{1/2}\| \leq C(1-\bar{\rho}_n)^{-k-1}$$

if $\bar{\rho}_n < 1$ for $k \in \{0, 1, 2, 3, 4\}$.

If $\bar{\rho}_n < 1$, since Lemma 3.4 yields

$$S_n^{-1} = \tilde{\mathcal{D}}^{-1/2}\left(\begin{pmatrix} \mathcal{E}_{M_1} & 0 \\ 0 & \mathcal{E}_{M_2} \end{pmatrix} + \begin{pmatrix} 0 & \tilde{G} \\ \tilde{G}^\top & 0 \end{pmatrix}\right)^{-1}\tilde{\mathcal{D}}^{-1/2} = \tilde{\mathcal{D}}^{-1/2}\sum_{p=0}^{\infty}(-1)^p\begin{pmatrix} 0 & \tilde{G} \\ \tilde{G}^\top & 0 \end{pmatrix}^p\tilde{\mathcal{D}}^{-1/2},$$

we obtain

$$|[S_n^{-1}]_{ij}| \leq C\left[\mathcal{D}^{-1/2}\sum_{p=0}^{\infty}\bar{\rho}_n^p\begin{pmatrix} 0 & G \\ G^\top & 0 \end{pmatrix}^p\mathcal{D}^{-1/2}\right]_{ij} = [\mathcal{D}^{-1/2}\dot{S}_n^{-1}\mathcal{D}^{-1/2}]_{ij}.$$

□

Under (A1), we have $\Sigma_t(\sigma) \geq c_1\mathcal{E}_2$, which implies that $\sup_{t,\sigma} |\rho_t(\sigma)| < 1$. Then, by (A2) and uniform continuity of b_t, for some fixed $\delta > 0$ and any $\epsilon > 0$, there exists $N \in \mathbb{N}$, such that $P(1 - \bar{\rho}_n < \delta) < \epsilon$ for $n \geq N$. Therefore, we have

$$P(\bar{\rho}_n < 1 - \delta) \to 1 \tag{3.3}$$

as $n \to \infty$, and we have

$$P((1 - \bar{\rho}_n)^{-q} > \delta^{-q}) < \epsilon$$

for any $q > 0$ and $n \geq N$, which implies that

$$(1 - \bar{\rho}_n)^{-q} = O_p(1). \tag{3.4}$$

Moreover, Lemma 3.4 yields

$$\begin{aligned} S_n^{-1}(\sigma) &= \tilde{\mathcal{D}}^{-1/2} \sum_{p=0}^{\infty} (-1)^p \begin{pmatrix} 0 & \tilde{G} \\ \tilde{G}^\top & 0 \end{pmatrix}^p \tilde{\mathcal{D}}^{-1/2} \\ &= \tilde{\mathcal{D}}^{-1/2} \sum_{p=0}^{\infty} \begin{pmatrix} (\tilde{G}\tilde{G}^\top)^p & -(\tilde{G}\tilde{G}^\top)^p \tilde{G} \\ -(\tilde{G}^\top \tilde{G})^p \tilde{G}^\top & (\tilde{G}^\top \tilde{G})^p \end{pmatrix} \tilde{\mathcal{D}}^{-1/2} \end{aligned} \tag{3.5}$$

if $\bar{\rho}_n < 1$.

3.2 Consistency of $\hat{\sigma}_n$

In this section, we show consistency: $\hat{\sigma}_n \xrightarrow{P} \sigma_0$ as $n \to \infty$. For this purpose, we specify the limit of $H_n^1(\sigma) - H_n^1(\sigma_0)$.

Lemma 3.6 *Assume* (A1) *and* (A2). *Then*

$$\frac{1}{n} \sup_{\sigma \in \Theta_1} \left| \partial_\sigma^k (H_n^1(\sigma) - H_n^1(\sigma_0)) + \frac{1}{2} \partial_\sigma^k \mathrm{tr}(S_n^{-1}(\sigma)(S_n(\sigma_0) - S_n(\sigma))) + \frac{1}{2} \partial_\sigma^k \log \frac{\det S_n(\sigma)}{\det S_n(\sigma_0)} \right| \xrightarrow{P} 0 \tag{3.6}$$

as $n \to \infty$ *for* $k \in \{0, 1, 2, 3\}$.

Proof Let $X_t^c = \int_0^t b_s(\sigma_0) \mathrm{d}W_s$. By the definition of H_n^1, we have

$$H_n^1(\sigma) - H_n^1(\sigma_0) = -\frac{1}{2} \Delta X^\top (S_n^{-1}(\sigma) - S_n^{-1}(\sigma_0)) \Delta X - \frac{1}{2} \log \frac{\det S_n(\sigma)}{\det S_n(\sigma_0)}.$$

We first show that

$$H_n^1(\sigma) - H_n^1(\sigma_0) = -\frac{1}{2}(\Delta X^c)^\top (S_n^{-1}(\sigma) - S_n^{-1}(\sigma_0))\Delta X^c - \frac{1}{2}\log\frac{\det S_n(\sigma)}{\det S_n(\sigma_0)} + \sqrt{n}\dot{e}_n(\sigma), \tag{3.7}$$

where $(\dot{e}_n(\sigma))_{n=1}^\infty$ denotes a general sequence of random variables, such that $\sup_\sigma |\dot{e}_n(\sigma)| \overset{P}{\to} 0$ as $n \to \infty$.

Since

$$\Delta X^\top S_n^{-1}(\sigma)\Delta X - (\Delta X^c)^\top S_n^{-1}(\sigma)\Delta X^c = 2(\Delta\bar{V})^\top S_n^{-1}(\sigma)\Delta X^c \\ +(\Delta\bar{V})^\top S_n^{-1}(\sigma)\Delta\bar{V} =: \Psi_1 + \Psi_2, \tag{3.8}$$

it suffices to show that $\Psi_i = \sqrt{n}\dot{e}_n$ for $i \in \{1, 2\}$.

Lemma 3.5 and (3.4) yield

$$|\Psi_2| \leq \|\mathcal{D}^{1/2}S_n^{-1}(\sigma)\mathcal{D}^{1/2}\||\mathcal{D}^{-1/2}\Delta\bar{V}|^2 = O_p(1) \times |\mathcal{D}^{-1/2}\Delta\bar{V}|^2. \tag{3.9}$$

Moreover, Lemma 3.2 yields

$$|\mathcal{D}^{-1/2}\Delta V(\theta)|^2 = \sum_{i,l}|I_i^l|^{-1}|\Delta_i^l V(\theta)|^2 \leq C\sum_{i,l}|I_i^l|^{-1}|I_i^l|^2 = C\sum_{i,l}|I_i^l| \leq Cnh_n. \tag{3.10}$$

Furthermore, Lemma 3.5, (3.4), (3.10), and the equation $E_\Pi[\Delta X^c(\Delta X^c)^\top] = S_n(\sigma_0)$ yield

$$E_\Pi[|\Psi_1|^2] = 4(\Delta\bar{V})^\top S_n^{-1}(\sigma)E_\Pi[\Delta X^c(\Delta X^c)^\top]S_n^{-1}(\sigma)\Delta\bar{V} = O_p(nh_n) = o_p(n). \tag{3.11}$$

Then, we obtain (3.7) by (3.8)–(3.11) and Lemma 3.1.

Next, we show that

$$(\Delta X^c)^\top S_n^{-1}(\sigma)\Delta X^c - \mathrm{tr}(S_n^{-1}(\sigma)S_n(\sigma_0)) = \bar{R}_n(\sqrt{n}). \tag{3.12}$$

Itô's formula yields

$$\begin{aligned}&(\Delta X^c)^\top S_n^{-1}(\sigma)\Delta X^c - \mathrm{tr}(S_n^{-1}(\sigma)S_n(\sigma_0)) \\ &= \sum_{i,j}[S_n^{-1}(\sigma)]_{ij}(\Delta_i X^c\Delta_j X^c - [S_n(\sigma_0)]_{ij}) \\ &= \sum_{i,j}[S_n^{-1}(\sigma)]_{ij}\left\{\int_{I_i}\Delta_{j,t}X^c \mathrm{d}X_t^{c,\psi(i)} + \int_{I_j}\Delta_{i,t}X^c \mathrm{d}X_t^{c,\psi(j)}\right\} \\ &= 2\sum_{i,j}[S_n^{-1}(\sigma)]_{ij}\int_{I_i}\Delta_{j,t}X^c \mathrm{d}X_t^{c,\psi(i)},\end{aligned} \tag{3.13}$$

where $X_t^{c,l}$ is the l-th component of X_t^c.

Since $\langle \Delta_i X^c, \Delta_j X^c\rangle_t = \int_{[0,t)\cap I_i\cap I_j}[\Sigma_t]_{\psi(i),\psi(j)}\mathrm{d}t$, together with the Burkholder–Davis–Gundy inequality, we have

$$\begin{aligned}
&E_\Pi\Bigg[\Bigg(\sum_{i,j}[S_n^{-1}(\sigma)]_{ij}\int_{I_i}\Delta_{j,t}X^c\mathrm{d}X_t^{c,\psi(i)}\Bigg)^q\Bigg]\\
&\le C_q\sum_{l=1}^2 E_\Pi\Bigg[\Bigg(\sum_{\substack{i,j_1,j_2\\ \psi(i)=l}}[S_n^{-1}(\sigma)]_{i,j_1}[S_n^{-1}(\sigma)]_{i,j_2}\int_{I_i}\Delta_{j_1,t}X^c\Delta_{j_2,t}X^c[\Sigma_t]_{\psi(i),\psi(i)}\mathrm{d}t\Bigg)^{q/2}\Bigg]\\
&\quad+C_q E_\Pi\Bigg[\Bigg(\sum_{\substack{i_1,i_2,j_1,j_2\\ \psi(i_1)=1,\psi(i_2)=2}}[S_n^{-1}(\sigma)]_{i_1,j_1}[S_n^{-1}(\sigma)]_{i_2,j_2}\\
&\quad\times\int_{I_{i_1}\cap I_{i_2}}\Delta_{j_1,t}X^c\Delta_{j_2,t}X^c[\Sigma_t]_{\psi(i_1),\psi(i_2)}\mathrm{d}t\Bigg)^{q/2}\Bigg]\\
&\le C_q E_\Pi\Bigg[\Bigg(\sum_{i_1,i_2,j_1,j_2}|[S_n^{-1}(\sigma)]_{i_1,j_1}[S_n^{-1}(\sigma)]_{i_2,j_2}|\sup_t|[\Sigma_t]_{\psi(i_1),\psi(i_2)}||I_{i_1}\cap I_{i_2}|\sup_t|\Delta_{j_1,t}X^c|\\
&\quad\sup_t|\Delta_{j_2,t}X^c|\Bigg)^{q/2}\Bigg]\\
&\le C_q E_\Pi\Bigg[\Bigg(\|\mathcal{D}^{1/2}\mathrm{Abs}(S_n^{-1})\{|I_i\cap I_j|\}_{ij}\mathrm{Abs}(S_n^{-1})\mathcal{D}^{1/2}\|\sum_i\frac{\sup_t|\Delta_{i,t}X^c|^2}{|I_i|}\Bigg)^{q/2}\Bigg].
\end{aligned}$$

Together with Lemmas 3.3 and 3.5, the triangle inequality for $L^{q/2}$ that

$$|I_i\cap I_j| = \left[\mathcal{D}^{1/2}\begin{pmatrix}\mathcal{E}_{M_1} & G\\ G^\top & \mathcal{E}_{M_2}\end{pmatrix}\mathcal{D}^{1/2}\right]_{ij},$$

and that

$$\begin{aligned}
\|\mathrm{Abs}(S_n^{-1})\|^2 &= \sup_{|x|=1}|\mathrm{Abs}(S_n^{-1})x|^2\\
&= \sup_{|x|=1}\sum_i\Bigg(\sum_j|[S_n^{-1}]_{ij}|x_j\Bigg)^2\\
&\le C\sup_{|x|=1}\sum_i\Bigg(\sum_j[\mathcal{D}^{-1/2}\dot{S}_n^{-1}\mathcal{D}^{-1/2}]_{ij}|x_j|\Bigg)^2\\
&\le C\|\mathcal{D}^{-1/2}\dot{S}_n^{-1}\mathcal{D}^{-1/2}\|^2
\end{aligned}$$

by Lemma 3.5, we have

$$E_\Pi\Bigg[\Bigg(\sum_{i,j}[S_n^{-1}(\sigma)]_{ij}\int_{I_i}\Delta_{j,t}X^c\mathrm{d}X_t^{c,\psi(i)}\Bigg)^q\Bigg]$$

$$
\begin{aligned}
&\le C_q(1-\bar{\rho}_n)^{-q} E_\Pi\left[\left(\sum_i \frac{\sup_t |\Delta_{i,t}X^c|^2}{|I_i|}\right)^{q/2}\right] \\
&\le C_q(1-\bar{\rho}_n)^{-q}\left(\sum_i \frac{E_\Pi[\sup_t |\Delta_{i,t}X^c|^q]^{2/q}}{|I_i|}\right)^{q/2} \\
&\le C_q M^{q/2}(1-\bar{\rho}_n)^{-q}
\end{aligned}
$$

on $\{\bar{\rho}_n < 1\}$ for $q \ge 1$. Then, thanks to (3.2), (3.4), (3.13) and Lemma 3.1, we obtain (3.12).

(3.12), (3.7), Sobolev's inequality, and similar estimates for $\partial_\sigma^k(H_n^1(\sigma) - H_n^1(\sigma_0))$ yield

$$
\begin{aligned}
&\partial_\sigma^k(H_n^1(\sigma) - H_n^1(\sigma_0)) \\
&\quad = -\frac{1}{2}\partial_\sigma^k \mathrm{tr}(S_n(\sigma_0)(S_n^{-1}(\sigma) - S_n^{-1}(\sigma_0))) - \frac{1}{2}\partial_\sigma^k \log\frac{\det S_n(\sigma)}{\det S_n(\sigma_0)} + \sqrt{n}\dot{e}_n(\sigma) \\
&\quad = -\frac{1}{2}\partial_\sigma^k \mathrm{tr}(S_n^{-1}(\sigma)(S_n(\sigma_0) - S_n(\sigma))) - \frac{1}{2}\partial_\sigma^k \log\frac{\det S_n(\sigma)}{\det S_n(\sigma_0)} + \sqrt{n}\dot{e}_n(\sigma)
\end{aligned}
$$

for $k \in \{0, 1, 2, 3\}$. □

For $(s_k)_{k=0}^\infty \in \mathfrak{S}$, let $\dot{\mathcal{A}}_{k,p}^1 = \mathcal{E}_{(k)}^1(GG^\top)^p$ and $\dot{\mathcal{A}}_{k,p}^2 = \mathcal{E}_{(k)}^2(G^\top G)^p$ for $p \in \mathbb{Z}_+$ and $1 \le k \le q_n$. The following lemma is used when we specify the limit of $n^{-1}(H_n^1(\sigma) - H_n^1(\sigma_0))$ in the next proposition.

Lemma 3.7 *Assume* (A2) *and* (A4). *Then, for any* $p \ge 1$

$$
n^{-1}\max_{1\le k\le q_n} |\mathrm{tr}(\dot{\mathcal{A}}_{k,p}^1) - \mathrm{tr}(\dot{\mathcal{A}}_{k,p}^2)| \xrightarrow{P} 0
$$

as $n \to \infty$.

Proof By the definition of $\dot{\mathcal{A}}_{k,p}^l$, we obtain

$$
\begin{aligned}
&|\mathrm{tr}(\dot{\mathcal{A}}_{k,p}^1) - \mathrm{tr}(\dot{\mathcal{A}}_{k,p}^2)| \\
&\quad = \left| \sum_{i;\ \sup I_i^1 \in (s_{k-1}, s_k]} [(GG^\top)^p]_{ii} - \sum_{j;\ \sup I_j^2 \in (s_{k-1}, s_k]} [(G^\top G)^p]_{jj} \right| \\
&\quad = \left| \sum_{i;\ \sup I_i^1 \in (s_{k-1}, s_k]} \sum_{i',j} [(GG^\top)^{p-1}]_{ii'}[G]_{i'j}[G^\top]_{ji} \right. \\
&\qquad \left. - \sum_{j;\ \sup I_j^2 \in (s_{k-1}, s_k]} \sum_{i,i'} [G^\top]_{ji}[(GG^\top)^{p-1}]_{ii'}[G]_{i'j} \right|.
\end{aligned}
$$

Two summands in the right-hand side coincide when both $\sup I_i^1$ and $\sup I_j^2$ are included or not included in $(s_{k-1}, s_k]$. In other cases, we have $\min_{u=0,1} |\sup I_i^1 -$

$s_{k-u}| \le r_n$ if $[G^\top]_{ji} > 0$. Therefore, we obtain

$$|\mathrm{tr}(\dot{\mathcal{A}}^1_{k,p}) - \mathrm{tr}(\dot{\mathcal{A}}^2_{k,p})| \le \Bigg(\sum_{\substack{i,j;\ \sup I^1_i \notin (s_{k-1}, s_k] \\ \sup I^2_j \in (s_{k-1}, s_k]}} + \sum_{\substack{i,j;\ \sup I^1_i \in (s_{k-1}, s_k] \\ \sup I^2_j \notin (s_{k-1}, s_k]}} \Bigg) \times \sum_{i'} [(GG^\top)^{p-1}]_{ii'}[G]_{i'j}[G^\top]_{ji}$$
$$\le \sum_{i;\ \min_{u=0,1} |\sup I^1_i - s_{k-u}| \le r_n} [(GG^\top)^p]_{ii}.$$

Thanks to (A2) and (A4), the right-hand side of the above inequality is equal to $O_p(h_n^{-1}) = o_p(n)$. □

Let $\mathcal{Y}_1(\sigma) = \lim_{T\to\infty}(T^{-1}\int_0^T y_{1,t}(\sigma)\mathrm{d}t)$, where

$$y_{1,t}(\sigma) = -\frac{1}{2}\mathcal{A}(\rho_t)\sum_{l=1}^2 B^2_{l,t} + \mathcal{A}(\rho_t)\frac{B_{1,t}B_{2,t}\rho_{t,0}}{\rho_t} + \sum_{l=1}^2 a^l_0\left(\frac{1}{2} - \frac{1}{2}B^2_{l,t} + \log B_{l,t}\right) + \int_{\rho_{t,0}}^{\rho_t} \frac{\mathcal{A}(\rho)}{\rho}\mathrm{d}\rho.$$

The limit $\mathcal{Y}_1(\sigma)$ exists under (A1), (A3), and (A4).

Proposition 3.8 *Assume* (A1)–(A4). *Then*

$$\sup_{\sigma\in\Theta_1} |n^{-1}\partial^k_\sigma(H^1_n(\sigma) - H^1_n(\sigma_0)) - \partial^k_\sigma \mathcal{Y}_1(\sigma)| \xrightarrow{P} 0 \tag{3.14}$$

as $n \to \infty$ *for* $k \in \{0, 1, 2, 3\}$.

Proof Let $\mathcal{A}^1_p = (\tilde{G}\tilde{G}^\top)^p$, $\mathcal{A}^2_p = (\tilde{G}^\top\tilde{G})^p$, $\tilde{\Sigma}^l_{i,0} = \tilde{\Sigma}^l_i(\sigma_0)$, and $\tilde{\Sigma}^{1,2}_{i,j,0} = \tilde{\Sigma}^{1,2}_{i,j}(\sigma_0)$. Thanks to (A1), for any $\epsilon > 0$, there exists $\delta > 0$, such that $|t-s| < \delta$ implies

$$|\rho_t - \rho_s| \vee |\Sigma_t - \Sigma_s| \vee |\mu_t - \mu_s| < \epsilon \tag{3.15}$$

for any σ and θ. We fix such $\delta > 0$, and fix a partition $s_k = k\delta/2$. Then, (3.5) and (A4) yield

$$n^{-1}\mathrm{tr}(S_n^{-1}(\sigma)(S_n(\sigma_0) - S_n(\sigma)))$$
$$= \frac{1}{n}\mathrm{tr}\left(S_n^{-1}(\sigma)\begin{pmatrix} \mathrm{diag}((\tilde{\Sigma}^1_{i,0} - \tilde{\Sigma}^1_i)_i) & \{\tilde{\Sigma}^{1,2}_{i,j,0} - \tilde{\Sigma}^{1,2}_{i,j}\}_{ij} \\ \{\tilde{\Sigma}^{1,2}_{i,j,0} - \tilde{\Sigma}^{1,2}_{i,j}\}_{ji} & \mathrm{diag}((\tilde{\Sigma}^2_{j,0} - \tilde{\Sigma}^2_j)_j) \end{pmatrix}\right)$$
$$= \frac{1}{n}\sum_{p=0}^\infty \left\{\sum_{l=1}^2 \mathrm{tr}\left(\mathrm{diag}\left(\left(\frac{\tilde{\Sigma}^l_{i,0}}{\tilde{\Sigma}^l_i} - 1\right)_i\right)\mathcal{A}^l_p\right) - 2\mathrm{tr}\left(\mathcal{A}^1_p\tilde{G}\left\{\frac{\tilde{\Sigma}^{1,2}_{i,j,0} - \tilde{\Sigma}^{1,2}_{i,j}}{(\tilde{\Sigma}^1_i)^{1/2}(\tilde{\Sigma}^2_j)^{1/2}}\right\}_{ij}\right)\right\}$$

$$= \frac{1}{n}\sum_{p=0}^{\infty}\sum_{k=1}^{q_n+1}\left\{\sum_{l=1}^{2}\mathrm{tr}\left(\mathrm{diag}\left(\left(\frac{\tilde{\Sigma}_{i,0}^l}{\tilde{\Sigma}_i^l}-1\right)_i\right)\mathcal{E}_{(k)}^l\mathcal{A}_p^l\right)\right.$$
$$\left. -2\mathrm{tr}\left(\mathcal{E}_{(k)}^1\mathcal{A}_p^1\tilde{G}\left\{\frac{\tilde{\Sigma}_{i,j,0}^{1,2}-\tilde{\Sigma}_{i,j}^{1,2}}{(\tilde{\Sigma}_i^1)^{1/2}(\tilde{\Sigma}_j^2)^{1/2}}\right\}_{ij}\right)\right\} \tag{3.16}$$

if $\bar{\rho}_n < 1$.

Let $\dot{\rho}_k = \rho_{s_{k-1}}$ and $\dot{B}_{k,l} = ([\Sigma_{s_{k-1}}(\sigma_0)]_{ll}/[\Sigma_{s_{k-1}}(\sigma)]_{ll})^{1/2}$. Then, (3.15) yields that for any $p \in \mathbb{Z}_+$, we have

$$|[\mathcal{E}_{(k)}^l\mathcal{A}_p^l]_{ij} - \dot{\rho}_k^{2p}[\dot{\mathcal{A}}_{k,p}^l]_{ij}| \le Cp\bar{\rho}_n^{2p-1}\epsilon \tag{3.17}$$

on $\{2pr_n < \delta/2\}$. Here, the factor p in the right-hand side appears, because we consider the difference between $2p$ products of $\rho_{i'j'}$ and $\dot{\rho}_k^{2p}$. Moreover, Lemma 3.4 and (3.4) yield

$$\limsup_{n\to\infty}\max_{1\le k\le q_n+1}\sum_{p=0}^{\infty}\|\mathcal{E}_{(k)}^l\mathcal{A}_p^l\| \le C\limsup_{n\to\infty}\sum_{p=0}^{\infty}\bar{\rho}_n^{2p} < \infty \tag{3.18}$$

almost surely.

Then, together with (A2) and Lemma 3.7, we obtain

$$n^{-1}\mathrm{tr}(S_n^{-1}(\sigma)(S_n(\sigma_0)-S_n(\sigma)))$$
$$= \frac{1}{n}\sum_{p=0}^{\infty}\sum_{k=1}^{q_n}\left\{\dot{\rho}_k^{2p}\sum_{l=1}^{2}(\dot{B}_{k,l}^2-1)\mathrm{tr}(\dot{\mathcal{A}}_{k,p}^l) - 2\dot{\rho}_k^{2p+1}(\dot{B}_{k,1}\dot{B}_{k,2}\dot{\rho}_{k,0}-\dot{\rho}_k)\mathrm{tr}(\dot{\mathcal{A}}_{k,p+1}^1)\right\} + e_n, \tag{3.19}$$

where $\dot{\rho}_{k,0} = \rho_{s_{k-1}}(\sigma_0)$, and $(e_n)_{n=1}^{\infty}$ denotes a general sequence of random variables such that $\limsup_{n\to\infty}|e_n| \to 0$ as $\delta \to 0$.

Moreover, by (3.3) and Lemma 3.4, we can apply Lemma A.3 in Ogihara (2018) to S_n. Then, we have

$$\log\det S_n(\sigma) = \log\det\tilde{\mathcal{D}} + \log\det\left(\mathcal{E}_M + \begin{pmatrix} 0 & \tilde{G} \\ \tilde{G}^\top & 0\end{pmatrix}\right)$$
$$= \sum_{l=1}^{2}\sum_{i=1}^{M_l}\log\tilde{\Sigma}_i^l + \sum_{p=1}^{\infty}\frac{(-1)^{p-1}}{p}\mathrm{tr}\left(\begin{pmatrix} 0 & \tilde{G} \\ \tilde{G}^\top & 0\end{pmatrix}^p\right)$$
$$= \sum_{l=1}^{2}\sum_{i=1}^{M_l}\log\tilde{\Sigma}_i^l - \sum_{p=1}^{\infty}\frac{1}{p}\mathrm{tr}((\tilde{G}\tilde{G}^\top)^p)$$

if $\bar{\rho}_n < 1$. Therefore, thanks to (3.2) and (3.17), we obtain

$$\begin{aligned} n^{-1}\log\frac{\det S_n(\sigma)}{\det S_n(\sigma_0)} &= n^{-1}\sum_{l=1}^{2}\sum_{i=1}^{M_l}\log\frac{\tilde{\Sigma}_i^l}{\tilde{\Sigma}_{i,0}^l} - n^{-1}\sum_{p=1}^{\infty}\frac{1}{p}\mathrm{tr}((\tilde{G}\tilde{G}^\top)^p - (\tilde{G}\tilde{G}^\top)^p(\sigma_0)) \\ &= -n^{-1}\sum_{k=1}^{q_n}\left\{\sum_{l=1}^{2}M_{l,k}\log\dot{B}_{k,l}^2 + \sum_{p=1}^{\infty}\frac{\dot{\rho}_k^{2p}-\dot{\rho}_{k,0}^{2p}}{p}\mathrm{tr}(\dot{\mathcal{A}}_{k,p}^1)\right\} + e_n. \end{aligned} \tag{3.20}$$

Lemma 3.7, (3.6), (3.19), and (3.20) yield

$$\begin{aligned} H_n^1(\sigma) - H_n^1(\sigma_0) &= -\frac{1}{2}\sum_{p=0}^{\infty}\sum_{k=1}^{q_n}\Bigg\{\dot{\rho}_k^{2p}\sum_{l=1}^{2}(\dot{B}_{k,l}^2-1)\mathrm{tr}(\dot{\mathcal{A}}_{k,p}^l) \\ &\quad -2\dot{\rho}_k^{2p+1}(\dot{B}_{k,1}\dot{B}_{k,2}\dot{\rho}_{k,0}-\dot{\rho}_k)\mathrm{tr}(\dot{\mathcal{A}}_{k,p+1}^1)\Bigg\} \\ &\quad +\frac{1}{2}\sum_{k=1}^{q_n}\left\{\sum_{l=1}^{2}M_{l,k}\log\dot{B}_{k,l}^2 + \sum_{p=1}^{\infty}\frac{\dot{\rho}_k^{2p}-\dot{\rho}_{k,0}^{2p}}{p}\mathrm{tr}(\dot{\mathcal{A}}_{k,p}^1)\right\} + ne_n \\ &= \sum_{k=1}^{q_n}\Bigg\{-\frac{1}{2}\sum_{p=0}^{\infty}\dot{\rho}_k^{2p}\sum_{l=1}^{2}\dot{B}_{k,l}^2\mathrm{tr}(\dot{\mathcal{A}}_{k,p}^l) + \sum_{p=1}^{\infty}\dot{\rho}_k^{2p-1}\dot{\rho}_{k,0}\dot{B}_{k,1}\dot{B}_{k,2}\mathrm{tr}(\dot{\mathcal{A}}_{k,p}^1) \\ &\quad +\frac{1}{2}\sum_{l=1}^{2}\mathrm{tr}(\dot{A}_{k,0}^l) \\ &\quad +\frac{1}{2}\sum_{l=1}^{2}M_{l,k}\log\dot{B}_{k,l}^2 + \sum_{p=1}^{\infty}\frac{\dot{\rho}_k^{2p}-\dot{\rho}_{k,0}^{2p}}{2p}\mathrm{tr}(\dot{\mathcal{A}}_{k,p}^1)\Bigg\} + ne_n \\ &= \sum_{k=1}^{q_n}\Bigg[\sum_{p=1}^{\infty}\dot{\rho}_k^{2p}\left\{-\frac{1}{2}\sum_{l=1}^{2}\dot{B}_{k,l}^2\mathrm{tr}(\dot{\mathcal{A}}_{k,p}^l) + \frac{\dot{\rho}_{k,0}}{\dot{\rho}_k}\dot{B}_{k,1}\dot{B}_{k,2}\mathrm{tr}(\dot{\mathcal{A}}_{k,p}^1)\right\} \\ &\quad +\frac{1}{2}\sum_{l=1}^{2}M_{l,k}\{-\dot{B}_{k,l}^2 + 1 + \log\dot{B}_{k,l}^2\} \\ &\quad +\sum_{p=1}^{\infty}\frac{\dot{\rho}_k^{2p}-\dot{\rho}_{k,0}^{2p}}{2p}\mathrm{tr}(\dot{\mathcal{A}}_{k,p}^1)\Bigg] + ne_n. \end{aligned} \tag{3.21}$$

Here, we used that $\mathrm{tr}(\dot{\mathcal{A}}_{k,0}^l) = \mathrm{tr}(\mathcal{E}_{(k)}^l) = M_{l,k}$.

Moreover, (A4) and (3.15) yield

$$\begin{aligned} &\left|\sum_{k=1}^{q_n}f(s_{k-1})\mathrm{tr}(\dot{\mathcal{A}}_{k,p}^l) - h_n^{-1}\int_0^{nh_n}a_p^1 f(t)\mathrm{d}t\right| \\ &\leq \left|\sum_{k=1}^{q_n}f(s_{k-1})(\mathrm{tr}(\dot{\mathcal{A}}_{k,p}^l) - h_n^{-1}a_p^1(s_k-s_{k-1}))\right| \end{aligned}$$

$$+\left|h_n^{-1}a_p^1\sum_{k=1}^{q_n}\int_{s_{k-1}}^{s_k}(f(t)-f(s_{k-1}))\mathrm{d}t\right|+O_p(h_n^{-1})$$
$$\leq o_p(h_n^{-1})\cdot q_n+C_p\epsilon n+O_p(h_n^{-1})=o_p(n)+ne_n \tag{3.22}$$

for $p\geq 1$ and any choice of $f(t)=\rho_t^{2p}B_{l,t}^2, \rho_t^{2p-1}\rho_{t,0}B_{1,t}B_{2,t}$ and $(\rho_t^{2p}-\rho_{t,0}^{2p})/(2p)$.

Here, we used that $q_n=O(nh_n)$ by the definition of $(s_k)_{k=0}^\infty\in\mathfrak{S}$. Similarly, we obtain

$$\sum_{k=1}^{q_n}M_{l,k}(1-\dot{B}_{k,l}^2+\log\dot{B}_{k,l}^2)=h_n^{-1}\int_0^{nh_n}a_0^l(1-B_{l,t}^2+\log B_{l,t}^2)\mathrm{d}t+ne_n.$$

Together with (3.21), (A3) and the equation

$$\sum_{p=1}^\infty a_p^1\frac{\rho_t^{2p}-\rho_{t,0}^{2p}}{2p}=\sum_{p=1}^\infty a_p^1\int_{\rho_{t,0}}^{\rho_t}\rho^{2p-1}\mathrm{d}\rho=\int_{\rho_{t,0}}^{\rho_t}\frac{\mathcal{A}(\rho)}{\rho}\mathrm{d}\rho,$$

we obtain

$$H_n^1(\sigma)-H_n^1(\sigma_0)=n\mathcal{Y}_1(\sigma)+ne_n.$$

The above arguments show that the supremum with respect to σ of the residual term in the above equation is also equal to ne_n, and consequently, we obtain (3.14) with $k=0$. Similarly, we obtain (3.14) with $k\in\{1,2,3\}$. □

Proposition 3.9 *Assume* (A1)–(A4). *Then, there exists a positive constant* χ, *such that*

$$\mathcal{Y}_1\leq\liminf_{T\to\infty}\int_0^T\Big\{-\frac{1}{2}(a_0^1\wedge a_0^2)(B_{1,t}-B_{2,t})^2-\chi\{a_1^1(\rho_t-\rho_{t,0})^2$$
$$+a_0^1\wedge a_0^2(B_{1,t}B_{2,t}-1)^2\}\Big\}\mathrm{d}t.$$

Proof The proof is based on the ideas of proof of Lemma 5 in Ogihara and Yoshida (2014). Let

$$G_k=\{[G]_{ij}1_{\{\sup I_i^1,\sup I_j^2\in(s_{k-1},s_k]\}}\}_{ij},$$

and let $\tilde{\mathcal{A}}_{k,p}^1=(G_kG_k^\top)^p$ and $\tilde{\mathcal{A}}_{k,p}^2=(G_k^\top G_k)^p$. Let $\tilde{\mathcal{A}}_k=\sum_{p=1}^\infty\dot{\rho}_k^{2p}\mathrm{tr}(\tilde{\mathcal{A}}_{k,p}^1)$ and $\tilde{\mathcal{B}}_k=\sum_{p=1}^\infty(2p)^{-1}(\dot{\rho}_k^{2p}-\dot{\rho}_{k,0}^{2p})\mathrm{tr}(\tilde{\mathcal{A}}_{k,p}^1)$. Similarly to the proof of Lemma 3.7, the difference between $\mathrm{tr}(\dot{\mathcal{A}}_{k,p}^l)$ and $\mathrm{tr}(\tilde{\mathcal{A}}_{k,p}^l)$ comes from terms with $\sup I_i^1$ close to s_{k-1} or s_k, and hence, we obtain

$$\max_{1\leq k\leq q_n}|\mathrm{tr}(\dot{\mathcal{A}}_{k,p}^l)-\mathrm{tr}(\tilde{\mathcal{A}}_{k,p}^l)|=o_p(n).$$

Therefore, (3.21) yields

$$\mathcal{Y}_1 = \frac{1}{n}\sum_{k=1}^{q_n}\Bigg\{-\frac{1}{2}(\dot{B}_{k,1}^2+\dot{B}_{k,2}^2)\tilde{\mathcal{A}}_k+\frac{\dot{\rho}_{k,0}}{\dot{\rho}_k}\dot{B}_{k,1}\dot{B}_{k,2}\tilde{\mathcal{A}}_k + \frac{1}{2}\sum_{l=1}^{2}M_{l,k}(1-\dot{B}_{k,l}^2+\log\dot{B}_{k,l}^2)+\tilde{\mathcal{B}}_k\Bigg\}+e_n$$
$$= \frac{1}{n}\sum_{k=1}^{q_n}\Bigg\{-\frac{1}{2}(\dot{B}_{k,1}-\dot{B}_{k,2})^2\tilde{\mathcal{A}}_k+\dot{B}_{k,1}\dot{B}_{k,2}\Bigg(\tilde{\mathcal{A}}_k\frac{\dot{\rho}_{k,0}}{\dot{\rho}_k}-\tilde{\mathcal{A}}_k\Bigg) + \frac{1}{2}\sum_{l=1}^{2}M_{l,k}(1-\dot{B}_{k,l}^2+\log\dot{B}_{k,l}^2)+\tilde{\mathcal{B}}_k\Bigg\}+e_n.$$

Then, since

$$\frac{1}{2}\sum_{l=1}^{2}M_{l,k}(1-\dot{B}_{k,l}^2+\log\dot{B}_{k,l}^2)$$
$$= M_{1,k}\Bigg(1-\frac{\dot{B}_{k,1}^2}{2}-\frac{\dot{B}_{k,2}^2}{2}+\log(\dot{B}_{k,1}\dot{B}_{k,2})\Bigg)+\frac{M_{2,k}-M_{1,k}}{2}(1-\dot{B}_{k,2}^2+\log(\dot{B}_{k,2}^2))$$
$$= -\frac{1}{2}M_{1,k}(\dot{B}_{k,1}-\dot{B}_{k,2})^2 - M_{1,k}\dot{B}_{k,1}\dot{B}_{k,2}+M_{1,k}\Bigg(1+\log(\dot{B}_{k,1}\dot{B}_{k,2})\Bigg) + \frac{M_{2,k}-M_{1,k}}{2}(1-\dot{B}_{k,2}^2+\log(\dot{B}_{k,2}^2)),$$

and a similar estimate holds by switching the roles of $M_{1,k}$ and $M_{2,k}$, we have

$$\mathcal{Y}_1 = n^{-1}\sum_{k=1}^{q_n}\Bigg\{-\frac{1}{2}(M_{1,k}+\tilde{\mathcal{A}}_k)(\dot{B}_{k,1}-\dot{B}_{k,2})^2+M_{1,k}(1+\log(\dot{B}_{k,1}\dot{B}_{k,2})) + \tilde{\mathcal{B}}_k+\frac{M_{2,k}-M_{1,k}}{2}(1-\dot{B}_{k,2}^2+\log(\dot{B}_{k,2}^2))+\dot{B}_{k,1}\dot{B}_{k,2}\Bigg(\tilde{\mathcal{A}}_k\frac{\dot{\rho}_{k,0}}{\dot{\rho}_k}-\tilde{\mathcal{A}}_k-M_{1,k}\Bigg)\Bigg\}+e_n$$
$$= n^{-1}\sum_{k=1}^{q_n}\Bigg\{-\frac{1}{2}(M_{2,k}+\tilde{\mathcal{A}}_k)(\dot{B}_{k,1}-\dot{B}_{k,2})^2+M_{2,k}(1+\log(\dot{B}_{k,1}\dot{B}_{k,2})) + \tilde{\mathcal{B}}_k+\frac{M_{1,k}-M_{2,k}}{2}(1-\dot{B}_{k,1}^2+\log(\dot{B}_{k,1}^2))+\dot{B}_{k,1}\dot{B}_{k,2}\Bigg(\tilde{\mathcal{A}}_k\frac{\dot{\rho}_{k,0}}{\dot{\rho}_k}-\tilde{\mathcal{A}}_k-M_{2,k}\Bigg)\Bigg\}+e_n.$$

For $l \in \{1,2\}$, let

$$F_{l,k} = M_{l,k}(1+\log(\dot{B}_{k,1}\dot{B}_{k,2}))+\tilde{\mathcal{B}}_k+\dot{B}_{k,1}\dot{B}_{k,2}\Bigg(\tilde{\mathcal{A}}_k\frac{\dot{\rho}_{k,0}}{\dot{\rho}_k}-\tilde{\mathcal{A}}_k-M_{l,k}\Bigg),$$

then since $1 - x + \log x \leq 0$ for $x > 0$, we obtain

$$\mathcal{Y}_1 \leq n^{-1} \sum_{k=1}^{q_n} \bigg[\bigg\{ -\frac{1}{2}(M_{1,k} + \tilde{\mathcal{A}}_k)(\dot{B}_{k,1} - \dot{B}_{k,2})^2 + F_{1,k} \bigg\} 1_{\{M_{2,k} \geq M_{1,k}\}} + \bigg\{ -\frac{1}{2}(M_{2,k} + \tilde{\mathcal{A}}_k)(\dot{B}_{k,1} - \dot{B}_{k,2})^2 + F_{2,k} \bigg\} 1_{\{M_{2,k} < M_{1,k}\}} \bigg] + e_n,$$

and therefore, we have

$$\mathcal{Y}_1 \leq n^{-1} \sum_{k=1}^{q_n} \bigg\{ -\frac{1}{2}(M_{1,k} \wedge M_{2,k} + \tilde{\mathcal{A}}_k)(\dot{B}_{k,1} - \dot{B}_{k,2})^2 + F_{1,k} \vee F_{2,k} \bigg\} + e_n. \quad (3.23)$$

Let $(\lambda_i^k)_{i=1}^{M_{1,k}}$ be all the eigenvalues of $G_k G_k^\top$. Similarly to Lemma 3.3, we have $0 \leq \lambda_i^k \leq 1$. Then, we have

$$F_{1,k} = \sum_{i=1}^{M_{1,k}} \bigg\{ 1 + \log(\dot{B}_{k,1}\dot{B}_{k,2}) + \dot{B}_{k,1}\dot{B}_{k,2} \sum_{p=0}^{\infty} \big\{ (\lambda_i^k)^{p+1} \dot{\rho}_k^{2p+1} \dot{\rho}_{k,0} - (\lambda_i^k)^p \dot{\rho}_k^{2p} \big\} + \sum_{p=1}^{\infty} \frac{(\lambda_i^k)^p}{2p} (\dot{\rho}_k^{2p} - \dot{\rho}_{k,0}^{2p}) \bigg\}.$$

Moreover, by setting $g_i^k = \sqrt{1 - \lambda_i^k \dot{\rho}_k^2}$, $g_{i,0}^k = \sqrt{1 - \lambda_i^k \dot{\rho}_{k,0}^2}$, and $F(x) = 1 - x + \log x$, we have

$$\begin{aligned} F_{1,k} &= \sum_{i=1}^{M_{1,k}} \Big\{ 1 + \dot{B}_{k,1}\dot{B}_{k,2}(g_i^k)^{-2}(\lambda_i^k \dot{\rho}_k \dot{\rho}_{k,0} - 1) + \log(\dot{B}_{k,1}\dot{B}_{k,2} g_{i,0}^k (g_i^k)^{-1}) \Big\} \\ &= \sum_{i=1}^{M_{1,k}} \Big\{ \dot{B}_{k,1}\dot{B}_{k,2}(g_i^k)^{-2}(\lambda_i^k \dot{\rho}_k \dot{\rho}_{k,0} - 1) + \dot{B}_{k,1}\dot{B}_{k,2} g_{i,0}^k (g_i^k)^{-1} + F(\dot{B}_{k,1}\dot{B}_{k,2} g_{i,0}^k (g_i^k)^{-1}) \Big\}. \end{aligned}$$

Here, we also used the expansion formulas $(1-x)^{-1} = \sum_{p=0}^{\infty} x^p$ and $-\log(1-x) = \sum_{p=1}^{\infty} x^p/p$ for $|x| < 1$.

Let

$$\mathcal{R} = \sup_{t,\sigma,0 \leq l \leq 4} (|\partial_\sigma^l \Sigma_t|^{1/2} \vee |\partial_\sigma^l \Sigma_t^{-1}|^{1/2}).$$

Since $g_i^k \leq 1$, $0 \leq \lambda_i^k \leq 1$, and $|\dot{\rho}_k| < 1$, we have

$$(g_i^k)^{-2}(\lambda_i^k \dot{\rho}_k \dot{\rho}_{k,0} - 1) + g_{i,0}^k (g_i^k)^{-1} = \frac{(\lambda_i^k \dot{\rho}_k \dot{\rho}_{k,0} - 1 + g_{i,0}^k g_i^k)(1 - \lambda_i^k \dot{\rho}_k \dot{\rho}_{k,0} + g_{i,0}^k g_i^k)}{(g_i^k)^2 (1 - \lambda_i^k \dot{\rho}_k \dot{\rho}_{k,0} + g_{i,0}^k g_i^k)}$$

$$= -\frac{(\lambda_i^k \dot{\rho}_k \dot{\rho}_{k,0} - 1)^2 - (g_{i,0}^k)^2 (g_i^k)^2}{(g_i^k)^2 (1 - \lambda_i^k \dot{\rho}_k \dot{\rho}_{k,0} + g_{i,0}^k g_i^k)}$$
$$= -\frac{\lambda_i^k (\dot{\rho}_k - \dot{\rho}_{k,0})^2}{(g_i^k)^2 (1 - \lambda_i^k \dot{\rho}_k \dot{\rho}_{k,0} + g_{i,0}^k g_i^k)}$$
$$\leq -\frac{\lambda_i^k}{3} (\dot{\rho}_k - \dot{\rho}_{k,0})^2.$$

Together with Lemma 11 in Ogihara and Yoshida (2014) and

$$\dot{B}_{k,1} \dot{B}_{k,2} g_{i,0}^k (g_i^k)^{-1} - 1 = \frac{\dot{B}_{k,1} \dot{B}_{k,2} \sqrt{1 - \lambda_i^k \dot{\rho}_{k,0}^2} - \sqrt{1 - \lambda_i^k \dot{\rho}_k^2}}{\sqrt{1 - \lambda_i^k \dot{\rho}_k^2}} \leq \frac{\dot{B}_{k,1} \dot{B}_{k,2}}{\sqrt{1 - \bar{\rho}_n^2}} \leq \frac{\mathcal{R}^4}{\sqrt{1 - \bar{\rho}_n^2}},$$

we have

$$F_{1,k} \leq \sum_{i=1}^{M_{1,k}} \left\{ -\frac{\dot{B}_{k,1} \dot{B}_{k,2}}{3} \lambda_i^k (\dot{\rho}_k - \dot{\rho}_{k,0})^2 - \frac{1 - \bar{\rho}_n^2}{4\mathcal{R}^8} (\dot{B}_{k,1} \dot{B}_{k,2} g_{i,0}^k (g_i^k)^{-1} - 1)^2 \right\}.$$

Moreover, the inequality $a^2 \geq (a+b)^2/2 - b^2$ with $a = \dot{B}_{k,1} \dot{B}_{k,2} g_{i,0}^k - g_i^k$ and $b = g_i^k - g_{i,0}^k$ yields

$$\begin{aligned}
(\dot{B}_{k,1} \dot{B}_{k,2} g_{i,0}^k (g_i^k)^{-1} - 1)^2 &\geq (\dot{B}_{k,1} \dot{B}_{k,2} g_{i,0}^k - g_i^k)^2 \\
&\geq \frac{(g_{i,0}^k)^2}{2} (\dot{B}_{k,1} \dot{B}_{k,2} - 1)^2 - (g_i^k - g_{i,0}^k)^2 \\
&= \frac{1 - \lambda_i^k \dot{\rho}_{k,0}^2}{2} (\dot{B}_{k,1} \dot{B}_{k,2} - 1)^2 - \frac{(\lambda_i^k)^2 (\dot{\rho}_k - \dot{\rho}_{k,0})^2}{(g_i^k + g_{i,0}^k)^2} \\
&\geq \frac{1 - \bar{\rho}_n^2}{2} (\dot{B}_{k,1} \dot{B}_{k,2} - 1)^2 - \frac{\lambda_i^k}{4(1 - \bar{\rho}_n^2)} (\dot{\rho}_k - \dot{\rho}_{k,0})^2,
\end{aligned}$$

and hence, we have

$$\begin{aligned}
F_{1,k} &\leq \sum_{i=1}^{M_{1,k}} \left\{ -\frac{\dot{B}_{k,1} \dot{B}_{k,2}}{3} \lambda_i^k (\dot{\rho}_k - \dot{\rho}_{k,0})^2 - \frac{(1 - \bar{\rho}_n^2)^2}{8\mathcal{R}^8} (\dot{B}_{k,1} \dot{B}_{k,2} - 1)^2 + \frac{\lambda_i^k}{16\mathcal{R}^8} (\dot{\rho}_k - \dot{\rho}_{k,0})^2 \right\} \\
&= -\left(\frac{\dot{B}_{k,1} \dot{B}_{k,2}}{3} - \frac{1}{16\mathcal{R}^8} \right) \mathrm{tr}(\tilde{\mathcal{A}}_{k,1}^1) (\dot{\rho}_k - \dot{\rho}_{k,0})^2 - \frac{(1 - \bar{\rho}_n^2)^2}{8\mathcal{R}^8} M_{1,k} (\dot{B}_{k,1} \dot{B}_{k,2} - 1)^2.
\end{aligned}$$

By a similar argument for $F_{2,k}$, there exists a positive constant $\tilde{\chi}$ which does not depend on k nor n, such that

$$F_{1,k} \vee F_{2,k} \leq -\tilde{\chi} (1 - \bar{\rho}_n^2)^2 \left\{ \mathrm{tr}(\tilde{\mathcal{A}}_{k,1}^1) (\dot{\rho}_k - \dot{\rho}_{k,0})^2 + M_{1,k} \wedge M_{2,k} (\dot{B}_{k,1} \dot{B}_{k,2} - 1)^2 \right\}.$$

Together with (3.23), we have

$$\mathcal{Y}_1 \leq n^{-1}\sum_{k=1}^{q_n}\bigg\{-\frac{1}{2}(M_{1,k}\wedge M_{2,k})(\dot{B}_{k,1}-\dot{B}_{k,2})^2 \\ -\tilde{\chi}(1-\bar{\rho}_n^2)^2\Big\{\mathrm{tr}(\tilde{\mathcal{A}}_{k,1}^1)(\dot{\rho}_k-\dot{\rho}_{k,0})^2 + M_{1,k}\wedge M_{2,k}(\dot{B}_{k,1}\dot{B}_{k,2}-1)^2\Big\}\bigg\}+e_n.$$

By letting $n\to\infty$, (A4) and (3.3) yield the conclusion. □

(A6) and Remark 4 in Ogihara and Yoshida (2014) yield that

$$\limsup_{T\to\infty}\frac{1}{T}\int_0^T\left\{|B_{1,t}-B_{2,t}|^2+|B_{1,t}B_{2,t}-1|^2+|\rho_t-\rho_{t,0}|^2\right\}\mathrm{d}t>0,$$

when $\sigma\neq\sigma_0$.

Then, by Proposition 3.9, we have $\mathcal{Y}_1(\sigma)<0$ (note that $a_0^1\wedge a_0^2\geq a_1^1$ by (2.3) and a similar argument). Therefore, for any $\delta>0$, there exists $\eta>0$, such that

$$\inf_{|\sigma-\sigma_0|\geq\delta}(-\mathcal{Y}_1(\sigma))\geq\eta.$$

Then, since $H_n^1(\hat{\sigma}_n)-H_n^1(\sigma_0)\geq 0$ by the definition, for any $\epsilon>0$, we have

$$\begin{aligned} P(|\hat{\sigma}_n-\sigma_0|\geq\delta) &\leq P\bigg(\sup_{|\sigma-\sigma_0|\geq\delta}(H_n^1(\sigma)-H_n^1(\sigma_0))\geq 0\bigg) \\ &\leq P\bigg(\sup_{|\sigma-\sigma_0|\geq\delta}\Big(n^{-1}(H_n^1(\sigma)-H_n^1(\sigma_0))-\mathcal{Y}_1(\sigma)\Big)\geq\eta\bigg) \\ &\leq P\bigg(\sup_{\sigma}|n^{-1}(H_n^1(\sigma)-H_n^1(\sigma_0))-\mathcal{Y}_1(\sigma)|\geq\eta\bigg)<\epsilon \end{aligned}\tag{3.24}$$

for sufficiently large n by Proposition 3.8, which implies $\hat{\sigma}_n\xrightarrow{P}\sigma_0$ as $n\to\infty$.

3.3 Asymptotic normality of $\hat{\sigma}_n$

Let $S_{n,0}=S_n(\sigma_0)$ and $\Sigma_{t,0}=\Sigma_t(\sigma_0)$. (3.7) and the equation $\partial_\sigma S_{n,0}^{-1}=-S_{n,0}^{-1}\partial_\sigma S_{n,0}S_{n,0}^{-1}$ imply

$$\begin{aligned}\partial_\sigma H_n^1(\sigma_0) &= -\frac{1}{2}(\Delta X^c)^\top\partial_\sigma S_{n,0}^{-1}\Delta X^c-\frac{1}{2}\mathrm{tr}(\partial_\sigma S_{n,0}S_{n,0}^{-1})+o_p(\sqrt{n}) \\ &= -\frac{1}{2}\mathrm{tr}(\partial_\sigma S_{n,0}^{-1}(\Delta X^c(\Delta X^c)^\top-S_{n,0}))+o_p(\sqrt{n}).\end{aligned}\tag{3.25}$$

Let $(L_n)_{n\in\mathbb{N}}$ be a sequence of positive integers such that $L_n\to\infty$ and $L_n n^\eta(nh_n)^{-1}\to 0$ as $n\to\infty$ for some $\eta>0$. Let $\check{s}_k=kT_n/L_n$ for $0\leq k\leq L_n$, let

$J^k = (\check{s}_{k-1}, \check{s}_k]$, and let $S^{(k)}_{n,0}$ be an $M \times M$ matrix satisfying

$$[S^{(k)}_{n,0}]_{ij} = \int_{I_i \cap I_j \cap J^k} [\Sigma_{t,0}]_{\psi(i),\psi(j)} \mathrm{d}t.$$

For a two-dimensional stochastic process $(U_t)_{t\geq 0} = ((U^1_t, U^2_t))_{t \geq 0}$, let $\Delta^{l,(k)}_{i,t} U = U^l_{(S^{n,l}_i \vee \check{s}_{k-1}) \wedge \check{s}_k \wedge t} - U^l_{(S^{n,l}_{i-1} \vee \check{s}_{k-1}) \wedge \check{s}_k \wedge t}$, and let $\Delta^{(k)}_{i,t} U = \Delta^{\psi(i),(k)}_{\varphi(i),t} U$ for $1 \leq i \leq M$. Let $\Delta^{(k)}_i U = \Delta^{(k)}_{i,T_n} U$, and let $\Delta^{(k)} U = (\Delta^{(k)}_i U)_{1 \leq i \leq M}$.

Let

$$\mathcal{X}_k = -\frac{1}{2\sqrt{n}} \{(\Delta^{(k)} X^c)^\top \partial_\sigma S^{-1}_{n,0} \Delta^{(k)} X^c - \mathrm{tr}(\partial_\sigma S^{-1}_{n,0} S^{(k)}_{n,0})\} - \frac{1}{\sqrt{n}} \sum_{k'<k} (\Delta^{(k)} X^c)^\top \partial_\sigma S^{-1}_{n,0} \Delta^{(k')} X^c.$$

Then, since $\Delta X^c = \sum_{k=1}^{L_n} \Delta^{(k)} X^c$ and $S_{n,0} = \sum_{k=1}^{L_n} S^{(k)}_{n,0}$, (3.25) yields

$$n^{-1/2} \partial_\sigma H^1_n(\sigma_0) = \sum_{k=1}^{L_n} \mathcal{X}_k + o_p(1). \tag{3.26}$$

Moreover, Itô's formula yields

$$\begin{aligned} \sqrt{n}\mathcal{X}_k &= -\frac{1}{2} \sum_{i,j} [\partial_\sigma S^{-1}_{n,0}]_{ij} \left\{ 2 \int_{I_i \cap J^k} \Delta^{(k)}_{j,t} X^c \mathrm{d}X^{c,\psi(i)}_t + 2 \sum_{k'<k} \int_{I_i \cap J^k} \Delta^{(k')}_j X^c \mathrm{d}X^{c,\psi(i)}_t \right\} \\ &= -\sum_{i,j} [\partial_\sigma S^{-1}_{n,0}]_{ij} \int_{I_i \cap J^k} \Delta_{j,t} X^c \mathrm{d}X^{c,\psi(i)}_t. \end{aligned} \tag{3.27}$$

Let $\mathcal{G}_t = \mathcal{F}_t \bigvee \sigma(\{\Pi_n\}_n)$ for $t \geq 0$. We will show

$$n^{-1/2} \partial_\sigma H^1_n(\sigma_0) \xrightarrow{d} N(0, \Gamma_1), \tag{3.28}$$

using Corollary 3.1 and the remark after that in Hall and Heyde (1980). For this purpose, it is sufficient to show

$$\sum_{k=1}^{L_n} E_k[\mathcal{X}_k^2] \xrightarrow{P} \Gamma_1, \tag{3.29}$$

and

$$\sum_{k=1}^{L_n} E_k[\mathcal{X}_k^4] \xrightarrow{P} 0, \tag{3.30}$$

by (3.26), where E_k denotes the conditional expectation with respect to $\mathcal{G}_{\check{s}_{k-1}}$.

We first show four auxiliary lemmas. Let $\tilde{M}_k = \#\{i; 1 \leq i \leq M, \sup I_i \in J^k\}$.

Lemma 3.10 *Assume* (A1). *Then, there exists a positive constant* C, *such that* $\|\mathcal{D}^{-1/2}S_{n,0}^{(k)}\mathcal{D}^{-1/2}\| \leq C$ *and* $\mathrm{tr}(\mathcal{D}^{-1/2}S_{n,0}^{(k)}\mathcal{D}^{-1/2}) \leq C(\tilde{M}_k+1)$ *for any* $1 \leq k \leq L_n$.

Proof Since

$$|[S_{n,0}^{(k)}]_{ij}| \leq C\left[\mathcal{D}^{1/2}\begin{pmatrix} \mathcal{E}_{M_1} & G \\ G^\top & \mathcal{E}_{M_2}\end{pmatrix}\mathcal{D}^{1/2}\right]_{ij},$$

Lemma 3.3 yields

$$\|\mathcal{D}^{-1/2}S_{n,0}^{(k)}\mathcal{D}^{-1/2}\| \leq C\left\|\begin{pmatrix} \mathcal{E}_{M_1} & G \\ G^\top & \mathcal{E}_{M_2}\end{pmatrix}\right\| \leq C.$$

Moreover, we have

$$\mathrm{tr}(\mathcal{D}^{-1/2}S_{n,0}^{(k)}\mathcal{D}^{-1/2}) = \sum_{i=1}^{M}\frac{\int_{I_i\cap J^k}[\Sigma_{t,0}]_{\psi(i),\psi(i)}\mathrm{d}t}{|I_i|} \leq C\sum_{i=1}^{M}1_{\{i;I_i\cap J^k\neq\emptyset\}} \leq C(\tilde{M}_k+1).$$

□

Lemma 3.11 *Assume* (A4) *and that* $nh_nL_n^{-1} \to \infty$ *as* $n \to \infty$. *Then*, $\{L_nn^{-1}\max_{1\leq k\leq L_n}\tilde{M}_k\}_{n=1}^\infty$ *is* P*-tight.*

Proof Let $\mathcal{M}_n = [nh_nL_n^{-1}]$. We define a partition of $[0,\infty)$ by

$$s_j = \frac{nh_nj}{2L_n\mathcal{M}_n} \quad (j \geq 0).$$

Then, $(s_j)_{j=0}^\infty \in \mathfrak{S}$ when $nh_nL_n^{-1} \geq 1$, and $(s_j)_{j=0}^{2L_n\mathcal{M}_n}$ is a subpartition of $(\check{s}_k)_{k=0}^{L_n}$.

For $M_{l,j}$ which corresponds to this partition ($\tilde{M}_k$ remains to be defined using $\check{s}_k$), we have

$$\tilde{M}_k = \sum_{l=1}^{2}\sum_{j=2\mathcal{M}_n(k-1)+1}^{2\mathcal{M}_nk}M_{l,j},$$

since $\check{s}_k = nh_nkL_n^{-1} = s_{2\mathcal{M}_nk}$. Therefore, (A4) yields

$$\max_{1\leq k\leq L_n}\tilde{M}_k \leq 4\mathcal{M}_n\max_{l,j}M_{l,j} \leq C\mathcal{M}_n\{h_n^{-1}(a_0^1\vee a_0^2)+o_p(h_n^{-1})\} = O_p(nL_n^{-1}).$$

□

Lemma 3.12 *Assume* (A1). *Then*

$$\|\tilde{\mathcal{D}}^{-1/2}S_{n,0}^{(k)}\partial_\sigma S_{n,0}^{-1}S_{n,0}^{(k')}\tilde{\mathcal{D}}^{-1/2}\| \leq C\frac{(\mathcal{Q}_n+1)\bar{\rho}_n^{\mathcal{Q}_n}}{(1-\bar{\rho}_n)^2}$$

on $\{\bar{\rho}_n < 1\}$ *for* $|k-k'| > 1$, *where* $\mathcal{Q}_n = [r_n^{-1}(T_n/L_n - 2r_n)]$.

Proof Using the expansion formula (3.5), we have

$$
\begin{aligned}
S^{(k)}_{n,0}\partial_\sigma S^{-1}_{n,0}S^{(k')}_{n,0} &= -S^{(k)}_{n,0}S^{-1}_{n,0}\partial_\sigma S_{n,0}S^{-1}_{n,0}S^{(k')}_{n,0} \\
&= -S^{(k)}_{n,0}\tilde{\mathcal{D}}^{-1/2}\sum_{p=0}^{\infty}(-1)^p\begin{pmatrix} 0 & \tilde{G} \\ \tilde{G}^\top & 0 \end{pmatrix}^p \tilde{\mathcal{D}}^{-1/2}\partial_\sigma S_{n,0}\tilde{\mathcal{D}}^{-1/2} \\
&\quad \times \sum_{q=0}^{\infty}(-1)^q\begin{pmatrix} 0 & \tilde{G} \\ \tilde{G}^\top & 0 \end{pmatrix}^q \tilde{\mathcal{D}}^{-1/2}S^{(k')}_{n,0} \\
&= \sum_{p,q=0}^{\infty}(-1)^{p+q+1}S^{(k)}_{n,0}\mathfrak{C}^n_{p,q}S^{(k')}_{n,0} \qquad (3.31)
\end{aligned}
$$

if $\bar{\rho}_n < 1$, where

$$
\mathfrak{C}^n_{p,q} = \tilde{\mathcal{D}}^{-1/2}\begin{pmatrix} 0 & \tilde{G} \\ \tilde{G}^\top & 0 \end{pmatrix}^p \tilde{\mathcal{D}}^{-1/2}\partial_\sigma S_{n,0}\tilde{\mathcal{D}}^{-1/2}\begin{pmatrix} 0 & \tilde{G} \\ \tilde{G}^\top & 0 \end{pmatrix}^q \tilde{\mathcal{D}}^{-1/2}.
$$

We consider a necessary condition for

$$
[S^{(k)}_{n,0}\mathfrak{C}^n_{p,q}S^{(k')}_{n,0}]_{i',j'} = \sum_{ij}[S^{(k)}_{n,0}]_{i'i}[\mathfrak{C}^n_{p,q}]_{ij}[S^{(k')}_{n,0}]_{j,j'} \qquad (3.32)
$$

to be zero for any i' and j'. We first observe that the element $[\mathfrak{C}^n_{p,q}]_{ij}$ is equal to zero if $[\bar{S}^{p+q+1}]_{ij} = 0$, where

$$
\bar{S} = \begin{pmatrix} \mathcal{E}_{M_1} & G \\ G^\top & \mathcal{E}_{M_2} \end{pmatrix}.
$$

Moreover, $[S^{(k)}_{n,0}]_{i'i} \neq 0$ only if $I_i \cap J^k \neq \emptyset$, and $[S^{(k')}_{n,0}]_{jj'} \neq 0$ only if $I_j \cap J^{k'} \neq \emptyset$. Since $\inf_{x\in I_i, y\in I_j}|x-y| > T_n/L_n - 2r_n$ if $I_i \cap J^k \neq \emptyset$ and $I_j \cap J^{k'} \neq \emptyset$, we have $[\bar{S}^r]_{ij} = 0$ for $r \leq \mathcal{Q}_n$ when $[S^{(k)}_{n,0}]_{i'i} \neq 0$ and $[S^{(k')}_{n,0}]_{jj'} \neq 0$. Therefore, $[S^{(k)}_{n,0}\mathfrak{C}^n_{p,q}S^{(k')}_{n,0}]_{i',j'} = 0$ for any i' and j' if $p+q+1 \leq \mathcal{Q}_n$.

Then, (3.31) and Lemmas 3.4, 3.5 and 3.10 yield

$$
\begin{aligned}
&\|\tilde{\mathcal{D}}^{-1/2}S^{(k)}_{n,0}\partial_\sigma S^{-1}_{n,0}S^{(k')}_{n,0}\tilde{\mathcal{D}}^{-1/2}\| \\
&\leq \sum_{p=0}^{\infty}\sum_{q=(\mathcal{Q}_n-p)\vee 0}^{\infty}\left\|\tilde{\mathcal{D}}^{-1/2}S^{(k)}_{n,0}\tilde{\mathcal{D}}^{-1/2}\begin{pmatrix} 0 & \tilde{G} \\ \tilde{G}^\top & 0 \end{pmatrix}^p \tilde{\mathcal{D}}^{-1/2}\partial_\sigma S_{n,0}\tilde{\mathcal{D}}^{-1/2}\right. \\
&\quad \left.\times \begin{pmatrix} 0 & \tilde{G} \\ \tilde{G}^\top & 0 \end{pmatrix}^q \tilde{\mathcal{D}}^{-1/2}S^{(k')}_{n,0}\tilde{\mathcal{D}}^{-1/2}\right\| \\
&\leq C\sum_{p=0}^{\infty}\sum_{q=(\mathcal{Q}_n-p)\vee 0}^{\infty}\bar{\rho}_n^{p+q} = C\frac{\mathcal{Q}_n\bar{\rho}_n^{\mathcal{Q}_n} + \bar{\rho}_n^{\mathcal{Q}_n}(1-\bar{\rho}_n)^{-1}}{1-\bar{\rho}_n}
\end{aligned}
$$

$$\leq C\frac{(\mathcal{Q}_n+1)\bar{\rho}_n^{\mathcal{Q}_n}}{(1-\bar{\rho}_n)^2}$$

on $\{\bar{\rho}_n < 1\}$. □

Lemma 3.13 *Let $m \in \mathbb{N}$. Let V be an $m \times m$ symmetric, positive definite matrix and A be a $m \times m$ matrix. Let X be a random variable following $N(0, V)$. Then*

$$\begin{aligned} E[(X^\top AX)^2] &= \mathrm{tr}(AV)^2 + 2\mathrm{tr}((AV)^2), \\ E[(X^\top AX)^3] &= \mathrm{tr}(AV)^3 + 6\mathrm{tr}(AV)\mathrm{tr}((AV)^2) + 8\mathrm{tr}((AV)^3), \\ E[(X^\top AX)^4] &= \mathrm{tr}(AV)^4 + 12\mathrm{tr}(AV)^2\mathrm{tr}((AV)^2) + 12\mathrm{tr}((AV)^2)^2 \\ &\quad +32\mathrm{tr}(AV)\mathrm{tr}((AV)^3) + 48\mathrm{tr}((AV)^4). \end{aligned}$$

Proof We only show the result for $E[(X^\top AX)^4]$. Let U be an orthogonal matrix and Λ be a diagonal matrix satisfying $UVU^\top = \Lambda$. Then, we have $UX \sim N(0, \Lambda)$, and

$$E\left[\prod_{i=1}^{8}[UX]_{j_i}\right] = \sum_{(l_{2q-1},l_{2q})_{q=1}^4}\prod_{q=1}^{4}[\Lambda]_{l_{2q-1},l_{2q}},$$

where the summation of $(l_{2q-1}, l_{2q})_{q=1}^4$ is taken over all disjoint pairs of $\{j_1, \dots j_8\}$. Then, by setting $B = UAU^\top$, we have

$$E[(X^\top AX)^4] = \sum_{j_1,\dots,j_8}\sum_{(l_{2q-1},l_{2q})_{q=1}^4}\prod_{p=1}^{4}[B]_{j_{2p-1},j_{2p}}\prod_{q=1}^{4}[\Lambda]_{l_{2q-1},l_{2q}}.$$

Let ${}_nC_k = \frac{n!}{k!(n-k)!}$. Out of $j_1, \dots, j_8$, we connect j_{2p-1} to j_{2p} and l_{2q-1} to l_{2q} $(1 \leq p, q \leq 4)$. Then, the pattern of the connected components gives five different cases.

1. Four connected components (four components of size 2): only one case of the pairs $(l_{2q-1}, l_{2q})_{q=1}^4$ appears, which corresponds to $\mathrm{tr}((B\Lambda)^4)$.
2. Three connected components (a component of size 4 and two components of size 2): The choice of elements for a components of size 4 gives ${}_4C_2$ ways, and the choice of the pair (l_{2q-1}, l_{2q}) for this component gives two ways, and hence, ${}_4C_2 \times 2 = 12$ ways in total. This case corresponds to $\mathrm{tr}(B\Lambda)^2\mathrm{tr}((B\Lambda)^2)$.
3. Two connected components (two components of size 4): The choice of elements for each component gives $\frac{{}_4C_2}{2}$ ways, excluding duplicates, and the choice of the pair (l_{2q-1}, l_{2q}) for each component gives two ways, and hence, $\frac{{}_4C_2}{2} \times 2 \times 2 = 12$ ways in total. This case corresponds to $\mathrm{tr}((B\Lambda)^2)^2$.
4. Two connected components (a component of size 6 and a component of size 2): The choice of elements for a components of size 6 gives ${}_4C_1$ ways, and the choice of the pair (l_{2q-1}, l_{2q}) for this component gives $4 \times 2 = 8$ ways, and hence ${}_4C_1 \times 8 = 32$ ways in total. This case corresponds to $\mathrm{tr}(B\Lambda)\mathrm{tr}((B\Lambda)^3)$.

5. One connected component (a component of size 8): The choice of the pair (l_{2q-1}, l_{2q}) gives $6 \times 4 \times 2 = 48$ ways. This case corresponds to $\mathrm{tr}((B\Lambda)^4)$.

Then, we obtain the conclusion. □

Proposition 3.14 *Assume* (A1)–(A4) *and* (A6). *Then*

$$n^{-1/2}\partial_\sigma H_n^1(\sigma_0) \xrightarrow{d} N(0, \Gamma_1)$$

as $n \to \infty$.

Proof It is sufficient to show (3.29) and (3.30). Let $\mathfrak{A}_k = (\Delta^{(k)}X^c)^\top \partial_\sigma S_{n,0}^{-1}\Delta^{(k)}X^c$ and $\mathfrak{B}_k = \partial_\sigma S_{n,0}^{-1}S_{n,0}^{(k)}$. By the definition of $\mathcal{X}_k$, we have

$$\begin{aligned}
&\sum_{k=1}^{L_n} E_k[\mathcal{X}_k^4] \\
&\quad \le \frac{C}{n^2}\sum_{k=1}^{L_n}\Bigg\{E_k\Big[\big\{(\Delta^{(k)}X^c)^\top \partial_\sigma S_{n,0}^{-1}\Delta^{(k)}X^c - \mathrm{tr}(\partial_\sigma S_{n,0}^{-1}S_{n,0}^{(k)})\big\}^4\Big] \\
&\qquad + E_k\Bigg[\bigg(\sum_{k'<k}(\Delta^{(k)}X^c)^\top \partial_\sigma S_{n,0}^{-1}\Delta^{(k')}X^c\bigg)^4\Bigg]\Bigg\} \\
&\quad = \frac{C}{n^2}\sum_{k=1}^{L_n}\Big\{E_k[\mathfrak{A}_k^4] - 4E_k[\mathfrak{A}_k^3]\mathrm{tr}(\mathfrak{B}_k) + 6E_k[\mathfrak{A}_k^2]\mathrm{tr}(\mathfrak{B}_k)^2 - 4\mathrm{tr}(\mathfrak{B}_k)^4 + \mathrm{tr}(\mathfrak{B}_k)^4\Big\} \\
&\qquad + \frac{C}{n^2}\sum_{k=1}^{L_n}\Bigg\{\bigg(\sum_{k'<k}\Delta^{(k')}X^c\bigg)^\top \partial_\sigma S_{n,0}^{-1}S_{n,0}^{(k)}\partial_\sigma S_{n,0}^{-1}\bigg(\sum_{k'<k}\Delta^{(k')}X^c\bigg)\Bigg\}^2. \qquad (3.33)
\end{aligned}$$

Thanks to Lemmas 3.13, 3.10, 3.11 and 3.5, (3.4), and Lemma A.1 in Ogihara (2018), the first term in the right-hand side is calculated as

$$\begin{aligned}
&\frac{C}{n^2}\sum_{k=1}^{L_n}\Big\{\mathrm{tr}(\mathfrak{B}_k)^4 + 12\mathrm{tr}(\mathfrak{B}_k)^2\mathrm{tr}(\mathfrak{B}_k^2) + 12\mathrm{tr}(\mathfrak{B}_k^2)^2 + 32\mathrm{tr}(\mathfrak{B}_k)\mathrm{tr}(\mathfrak{B}_k^3) + 48\mathrm{tr}(\mathfrak{B}_k^4) \\
&\qquad - 4\mathrm{tr}(\mathfrak{B}_k)\big\{\mathrm{tr}(\mathfrak{B}_k)^3 + 6\mathrm{tr}(\mathfrak{B}_k)\mathrm{tr}(\mathfrak{B}_k^2) + 8\mathrm{tr}(\mathfrak{B}_k^3)\big\} + 6\mathrm{tr}(\mathfrak{B}_k)^2\big\{\mathrm{tr}(\mathfrak{B}_k)^2 \\
&\qquad + 2\mathrm{tr}(\mathfrak{B}_k^2)\big\} - 3\mathrm{tr}(\mathfrak{B}_k)^4\Big\} \\
&\quad = \frac{C}{n^2}\sum_{k=1}^{L_n}\big\{48\mathrm{tr}(\mathfrak{B}_k^4) + 12\mathrm{tr}(\mathfrak{B}_k^2)^2\big\} \\
&\quad \le \frac{C}{n^2}(\max_k \tilde{M}_k + 1)^2 L_n(1-\bar{\rho}_n)^{-8}1_{\{\bar{\rho}_n<1\}} + o_p(1) \xrightarrow{P} 0. \qquad (3.34)
\end{aligned}$$

Moreover, Lemma 3.13 yields

$$
\begin{aligned}
&E_\Pi\bigg[\frac{C}{n^2}\sum_{k=1}^{L_n}\bigg\{\bigg(\sum_{k'<k}\Delta^{(k')}X^c\bigg)^\top \partial_\sigma S_{n,0}^{-1}S_{n,0}^{(k)}\partial_\sigma S_{n,0}^{-1}\bigg(\sum_{k'<k}\Delta^{(k')}X^c\bigg)\bigg\}^2\bigg] \\
&\quad\le \frac{C}{n^2}\sum_{k=1}^{L_n}\sum_{k_1',k_2'<k}\big\{|\mathrm{tr}(\partial_\sigma S_{n,0}^{-1}S_{n,0}^{(k)}\partial_\sigma S_{n,0}^{-1}S_{n,0}^{(k_1')})\mathrm{tr}(\partial_\sigma S_{n,0}^{-1}S_{n,0}^{(k)}\partial_\sigma S_{n,0}^{-1}S_{n,0}^{(k_2')})| \\
&\qquad+|\mathrm{tr}(\partial_\sigma S_{n,0}^{-1}S_{n,0}^{(k)}\partial_\sigma S_{n,0}^{-1}S_{n,0}^{(k_1')}\partial_\sigma S_{n,0}^{-1}S_{n,0}^{(k)}\partial_\sigma S_{n,0}^{-1}S_{n,0}^{(k_2')})|\big\}. \qquad (3.35)
\end{aligned}
$$

If $k_1' < k-1$, Lemmas 3.5 and 3.12, Lemma A.1 in Ogihara (2018) and the equation $\partial_\sigma S_{n,0}^{-1} = -S_{n,0}^{-1}\partial_\sigma S_{n,0}S_{n,0}^{-1}$ yield

$$
\begin{aligned}
&|\mathrm{tr}(\partial_\sigma S_{n,0}^{-1}S_{n,0}^{(k)}\partial_\sigma S_{n,0}^{-1}S_{n,0}^{(k_1')})| \\
&\quad= |\mathrm{tr}(\tilde{\mathcal{D}}^{1/2}S_{n,0}^{-1}\tilde{\mathcal{D}}^{1/2}\tilde{\mathcal{D}}^{-1/2}\partial_\sigma S_{n,0}\tilde{\mathcal{D}}^{-1/2}\tilde{\mathcal{D}}^{1/2}S_{n,0}^{-1}\tilde{\mathcal{D}}^{1/2}\tilde{\mathcal{D}}^{-1/2}S_{n,0}^{(k)}\partial_\sigma S_{n,0}^{-1}S_{n,0}^{(k_1')}\tilde{\mathcal{D}}^{-1/2})| \\
&\quad\le \mathrm{tr}(\tilde{\mathcal{D}}^{1/2}S_{n,0}^{-1}\tilde{\mathcal{D}}^{1/2})\|\tilde{\mathcal{D}}^{-1/2}\partial_\sigma S_{n,0}\tilde{\mathcal{D}}^{-1/2}\|\|\tilde{\mathcal{D}}^{1/2}S_{n,0}^{-1}\tilde{\mathcal{D}}^{1/2}\|\|\tilde{\mathcal{D}}^{-1/2}S_{n,0}^{(k)}\partial_\sigma S_{n,0}^{-1}S_{n,0}^{(k_1')}\tilde{\mathcal{D}}^{-1/2})\| \\
&\quad\le CM\mathcal{Q}_n\bar{\rho}_n^{\mathcal{Q}_n}(1-\bar{\rho}_n)^{-4}
\end{aligned}
$$

on $\{\bar{\rho}_n < 1\}$. Here, we used that $\mathrm{tr}(\tilde{\mathcal{D}}^{1/2}S_{n,0}^{-1}\tilde{\mathcal{D}}^{1/2}) \le M\cdot\|\tilde{\mathcal{D}}^{1/2}S_{n,0}^{-1}\tilde{\mathcal{D}}^{1/2}\| \le CM(1-\bar{\rho}_n)^{-1}$. Similarly, we obtain

$$
|\mathrm{tr}(\partial_\sigma S_{n,0}^{-1}S_{n,0}^{(k)}\partial_\sigma S_{n,0}^{-1}S_{n,0}^{(k_1')}\partial_\sigma S_{n,0}^{-1}S_{n,0}^{(k)}\partial_\sigma S_{n,0}^{-1}S_{n,0}^{(k_2')})| \le CM\mathcal{Q}_n\bar{\rho}_n^{\mathcal{Q}_n}(1-\bar{\rho}_n)^{-8}.
$$

Since $\bar{\rho}_n^{\mathcal{Q}_n}$ converges to zero very fast if $\bar{\rho}_n < 1$ and $r_n \le 1$, together with (A2) and (3.3), the summation for of the terms with $k_1' < k-1$ or $k_2' < k-1$ in the right-hand side of (3.35) is equal to $o_p(1)$.

Then, together with Lemmas 3.5, 3.10, and 3.11, and Lemma A.1 in Ogihara (2018), we obtain

$$
\begin{aligned}
&E_\Pi\bigg[\frac{C}{n^2}\sum_{k=1}^{L_n}\bigg\{\bigg(\sum_{k'<k}\Delta^{(k')}X^c\bigg)^\top \partial_\sigma S_{n,0}^{-1}S_{n,0}^{(k)}\partial_\sigma S_{n,0}^{-1}\bigg(\sum_{k'<k}\Delta^{(k')}X^c\bigg)\bigg\}^2\bigg] \\
&\quad\le \frac{C}{n^2}\sum_{k=1}^{L_n}\big\{|\mathrm{tr}(\partial_\sigma S_{n,0}^{-1}S_{n,0}^{(k)}\partial_\sigma S_{n,0}^{-1}S_{n,0}^{(k-1)})^2| + |\mathrm{tr}((\partial_\sigma S_{n,0}^{-1}S_{n,0}^{(k)}\partial_\sigma S_{n,0}^{-1}S_{n,0}^{(k-1)})^2)|\big\} + o_p(1) \\
&\quad= O_p\bigg(\frac{L_n}{n^2}\Big\{\max_k \tilde{M}_k + 1\Big\}^2\bigg) + o_p(1) \xrightarrow{P} 0 \qquad (3.36)
\end{aligned}
$$

as $n\to\infty$. Then, (3.33), (3.34), and (3.36) yield (3.30).

Next, we show (3.29). Let $\mathcal{I}_{i,j}^k = I_i \cap I_j \cap J^k$. Then, (3.27) yields

$$\sum_{k=1}^{L_n} E_k[\mathcal{X}_k^2] = \frac{1}{n}\sum_{k=1}^{L_n}\sum_{i_1,j_1}\sum_{i_2,j_2}[\partial_\sigma S_{n,0}^{-1}]_{i_1,j_1}[\partial_\sigma S_{n,0}^{-1}]_{i_2,j_2} \times \int_{\mathcal{I}_{i_1,i_2}^k}[\Sigma_{t,0}]_{\psi(i_1),\psi(i_2)}E_k[\Delta_{j_1,t}X^c\Delta_{j_2,t}X^c]\mathrm{d}t$$
$$= \frac{1}{n}\sum_{k=1}^{L_n}\sum_{i_1,j_1}\sum_{i_2,j_2}[\partial_\sigma S_{n,0}^{-1}]_{i_1,j_1}[\partial_\sigma S_{n,0}^{-1}]_{i_2,j_2} \times \int_{\mathcal{I}_{i_1,i_2}^k}[\Sigma_{t,0}]_{\psi(i_1),\psi(i_2)}\int_{I_{j_1}\cap I_{j_2}\cap[0,t)}[\Sigma_{s,0}]_{\psi(j_1),\psi(j_2)}\mathrm{d}s\mathrm{d}t. \quad (3.37)$$

We can decompose

$$\int_{\mathcal{I}_{i_1,i_2}^k}[\Sigma_{t,0}]_{\psi(i_1),\psi(i_2)}\int_{I_{j_1}\cap I_{j_2}\cap[0,t)}[\Sigma_{s,0}]_{\psi(j_1),\psi(j_2)}\mathrm{d}s\mathrm{d}t = \int_0^{T_n}F_{i_1,i_2}^k(t)\int_0^t F_{j_1,j_2}^k(s)\mathrm{d}s\mathrm{d}t + \sum_{k'<k}\mathcal{F}_{i_1,i_2}^k\mathcal{F}_{j_1,j_2}^{k'},$$

where $F_{ij}^k(t) = [\Sigma_{t,0}]_{\psi(i),\psi(j)}1_{\mathcal{I}_{i,j}^k}(t)$, and $\mathcal{F}_{i,j}^k = \int_0^{T_n}F_{i,j}^k(t)\mathrm{d}t$. Moreover, switching the roles of i_1, i_2 and j_1, j_2, we obtain

$$\sum_{i_1,j_1}\sum_{i_2,j_2}[\partial_\sigma S_{n,0}^{-1}]_{i_1,j_1}[\partial_\sigma S_{n,0}^{-1}]_{i_2,j_2}\int_0^{T_n}F_{i_1,i_2}^k(t)\int_0^t F_{j_1,j_2}^k(s)\mathrm{d}s\mathrm{d}t$$
$$= \sum_{i_1,j_1}\sum_{i_2,j_2}[\partial_\sigma S_{n,0}^{-1}]_{i_1,j_1}[\partial_\sigma S_{n,0}^{-1}]_{i_2,j_2} \times \frac{1}{2}\left\{\int_0^{T_n}F_{i_1,i_2}^k(t)\int_0^t F_{j_1,j_2}^k(s)\mathrm{d}s\mathrm{d}t + \int_0^{T_n}F_{j_1,j_2}^k(t)\int_0^t F_{i_1,i_2}^k(s)\mathrm{d}s\mathrm{d}t\right\}$$
$$= \frac{1}{2}\sum_{i_1,j_1}\sum_{i_2,j_2}[\partial_\sigma S_{n,0}^{-1}]_{i_1,j_1}[\partial_\sigma S_{n,0}^{-1}]_{i_2,j_2}\left\{\int_0^{T_n}F_{i_1,i_2}^k(t)\int_0^t F_{j_1,j_2}^k(s)\mathrm{d}s\mathrm{d}t + \int_0^{T_n}F_{i_1,i_2}^k(s)\int_s^{T_n}F_{j_1,j_2}^k(t)\mathrm{d}t\mathrm{d}s\right\}$$
$$= \frac{1}{2}\sum_{i_1,j_1}\sum_{i_2,j_2}[\partial_\sigma S_{n,0}^{-1}]_{i_1,j_1}[\partial_\sigma S_{n,0}^{-1}]_{i_2,j_2}\mathcal{F}_{i_1,i_2}^k\mathcal{F}_{j_1,j_2}^k.$$

Therefore, we have

$$
\begin{aligned}
\sum_{k=1}^{L_n} E_k[\mathcal{X}_k^2] &= \frac{1}{2n}\sum_{k=1}^{L_n}\sum_{i_1,j_1}\sum_{i_2,j_2}[\partial_\sigma S_{n,0}^{-1}]_{i_1,j_1}[\partial_\sigma S_{n,0}^{-1}]_{i_2,j_2}\left\{\mathcal{F}^k_{i_1,i_2}\mathcal{F}^k_{j_1,j_2} + 2\sum_{k'<k}\mathcal{F}^k_{i_1,i_2}\mathcal{F}^{k'}_{j_1,j_2}\right\} \\
&= \frac{1}{2n}\sum_{k,k'=1}^{L_n}\sum_{i_1,j_1}\sum_{i_2,j_2}[\partial_\sigma S_{n,0}^{-1}]_{i_1,j_1}[\partial_\sigma S_{n,0}^{-1}]_{i_2,j_2}\mathcal{F}^k_{i_1,i_2}\mathcal{F}^{k'}_{j_1,j_2} \\
&= \frac{1}{2n}\sum_{i_1,j_1}\sum_{i_2,j_2}[\partial_\sigma S_{n,0}^{-1}]_{i_1,j_1}[\partial_\sigma S_{n,0}^{-1}]_{i_2,j_2}\int_{I_{i_1}\cap I_{i_2}}[\Sigma_{t,0}]_{\psi(i_1),\psi(i_2)}\mathrm{d}t \\
&\quad\times\int_{I_{j_1}\cap I_{j_2}}[\Sigma_{s,0}]_{\psi(j_1),\psi(j_2)}\mathrm{d}s \\
&= \frac{1}{2n}\mathrm{tr}((\partial_\sigma S_{n,0}^{-1}S_{n,0})^2).
\end{aligned}
\tag{3.38}
$$

$\partial_\sigma S_{n,0}^{-1}S_{n,0}$ corresponds to $\hat{\mathcal{D}}(t)$ in the proof (p. 2993) of Proposition 10 of Ogihara and Yoshida (2014). Then, by a similar step to the proof of Proposition 10 in Ogihara and Yoshida (2014), we have (3.29). □

Proposition 3.15 *Assume* (A1)–(A4) *and* (A6). *Then*, Γ_1 *is positive definite and*

$$
\sqrt{n}(\hat{\sigma}_n - \sigma_0) \xrightarrow{d} N(0, \Gamma_1^{-1})
$$

as $n \to \infty$.

Proof Proposition 3.9, (A6), and Remark 4 in Ogihara and Yoshida (2014) yield

$$
\mathcal{Y}_1(\sigma) \le -c|\sigma - \sigma_0|^2 \tag{3.39}
$$

for some positive constant c. Moreover, $\mathcal{Y}_1(\sigma_0) = 0$ by $B_{l,t,0} = 1$, and $\partial_\sigma \mathcal{Y}_1(\sigma_0) = 0$ by

$$
\begin{aligned}
\partial_\sigma y_{1,t}(\sigma_0) = &-\partial_\rho\mathcal{A}(\rho_{t,0})\partial_\sigma\rho_{t,0} - \frac{1}{2}\mathcal{A}(\rho_{t,0})\sum_{l=1}^2 2\partial_\sigma B_{l,t,0} \\
&+\partial_\rho\mathcal{A}(\rho_{t,0})\partial_\sigma\rho_{t,0} - \mathcal{A}(\rho_{t,0})\frac{\partial_\sigma\rho_{t,0}}{\rho_{t,0}} \\
&+(\partial_\sigma B_{1,t,0} + \partial_\sigma B_{2,t,0})\mathcal{A}(\rho_{t,0}) + \sum_{l=1}^2 a_0^l(-\partial_\sigma B_{l,t,0} + \partial_\sigma B_{l,t,0}) \\
&+\frac{\mathcal{A}(\rho_{t,0})}{\rho_{t,0}}\partial_\sigma\rho_{t,0} \\
= &\ 0.
\end{aligned}
$$

Then, Taylor's formula yields

$$\mathcal{Y}_1(\sigma) = (\sigma - \sigma_0)^\top \partial_\sigma^2 \mathcal{Y}_1(\sigma_0)(\sigma - \sigma_0) + o(|\sigma - \sigma_0|^2).$$

Therefore, considering σ sufficiently close to σ_0, $\Gamma_1 = -\partial_\sigma^2 \mathcal{Y}_1(\sigma_0)$ should be positive definite by (3.39).

By Taylor's formula and the equation $\partial_\sigma H_n^1(\hat{\sigma}_n) = 0$, we have

$$\begin{aligned} -\partial_\sigma H_n^1(\sigma_0) &= \partial_\sigma H_n^1(\hat{\sigma}_n) - \partial_\sigma H_n^1(\sigma_0) \\ &= \int_0^1 \partial_\sigma^2 H_n^1(\sigma_t) dt (\hat{\sigma}_n - \sigma_0) \\ &= \partial_\sigma^2 H_n^1(\sigma_0)(\hat{\sigma}_n - \sigma_0) + (\hat{\sigma}_n - \sigma_0)^\top \int_0^1 (1-t)\partial_\sigma^3 H_n^1(\sigma_t) \mathrm{d}t (\hat{\sigma}_n - \sigma_0), \end{aligned}$$

where $\sigma_t = t\hat{\sigma}_n + (1-t)\sigma_0$.

Therefore, we obtain

$$\sqrt{n}(\hat{\sigma}_n - \sigma_0) = \left\{ -\frac{1}{n}\partial_\sigma^2 H_n^1(\sigma_0) - \frac{1}{n}\int_0^1 (1-t)\partial_\sigma^3 H_n^1(\sigma_t)\mathrm{d}t(\hat{\sigma}_n - \sigma_0) \right\}^{-1} \cdot \frac{1}{\sqrt{n}} \partial_\sigma H_n^1(\sigma_0). \tag{3.40}$$

Since Proposition 3.8 yields

$$-\frac{1}{n}\partial_\sigma^2 H_n^1(\sigma_0) \xrightarrow{P} -\partial_\sigma^2 \mathcal{Y}_1(\sigma_0) = \Gamma_1,$$

and

$$\left\{ \sup_\sigma \left| \frac{1}{n} \partial_\sigma^3 H_n^1(\sigma) \right| \right\}_{n \in \mathbb{N}}$$

is P-tight, together with Proposition 3.14, we conclude

$$\sqrt{n}(\hat{\sigma}_n - \sigma_0) \xrightarrow{d} N(0, \Gamma_1^{-1}). \tag{3.41}$$

□

3.4 Consistency of $\hat{\theta}_n$

Let

$$\mathcal{Y}_2(\theta) = \lim_{T\to\infty} \frac{1}{T} \int_0^T \sum_{p=0}^\infty \left\{ -\frac{1}{2}\sum_{l=1}^2 f_p^{ll} \rho_{t,0}^{2p} \phi_{l,t}^2 + f_p^{12} \rho_{t,0}^{2p+1} \phi_{1,t}\phi_{2,t} \right\} \mathrm{d}t,$$

which exists under (A1), (A3), and (A5).

Proposition 3.16 *Assume* (A1)–(A6). *Then*

$$\sup_{\theta\in\Theta_2}\left|(nh_n)^{-1}\partial_\theta^k(H_n^2(\theta)-H_n^2(\theta_0))-\partial_\theta^k\mathcal{Y}_2(\theta)\right|\overset{P}{\to}0 \tag{3.42}$$

as $n\to\infty$ *for* $k\in\{0,1,2,3\}$.

Proof We first show that

$$\begin{aligned}&\bar{X}(\theta)^\top S_n^{-1}(\hat{\sigma}_n)\bar{X}(\theta)\\&=\Delta X^\top S_n^{-1}(\hat{\sigma}_n)\Delta X-2\Delta V(\theta)^\top S_{n,0}^{-1}\Delta X^c-\Delta V(\theta)^\top S_{n,0}^{-1}(2\Delta V(\theta_0)\\&-\Delta V(\theta))+\sqrt{nh_n}\dot{e}_n(\theta),\end{aligned} \tag{3.43}$$

where $(\dot{e}_n(\theta))_{n=1}^\infty$ denotes a general sequence of random variables, such that $\sup_\theta|\dot{e}_n(\theta)|\overset{P}{\to}0$ as $n\to\infty$.

Lemma 3.5 and (3.10) yield

$$\begin{aligned}&E_\Pi\left[(\Delta V(\theta)^\top\partial_\sigma^k S_{n,0}^{-1}\Delta X^c)^2\right]\\&=\sum_{i_1,j_1}\sum_{i_2,j_2}[\partial_\sigma^k S_{n,0}^{-1}]_{i_1,j_1}[\partial_\sigma^k S_{n,0}^{-1}]_{i_2,j_2}\Delta_{i_1}V(\theta)\Delta_{i_2}V(\theta)E_\Pi[\Delta_{j_1}X^c\Delta_{j_2}X^c]\\&=\sum_{i_1,j_1}\sum_{i_2,j_2}[\partial_\sigma^k S_{n,0}^{-1}]_{i_1,j_1}[\partial_\sigma^k S_{n,0}^{-1}]_{i_2,j_2}\Delta_{i_1}V(\theta)\Delta_{i_2}V(\theta)[S_{n,0}]_{j_1,j_2}\\&\le C|\mathcal{D}^{-1/2}\Delta V(\theta)|^2\|\mathcal{D}^{1/2}\partial_\sigma^k S_{n,0}^{-1}\mathcal{D}^{1/2}\|^2\|\mathcal{D}^{-1/2}S_{n,0}\mathcal{D}^{-1/2}\|\\&\le Cnh_n(1-\bar{\rho}_n)^{-2k-2}\end{aligned} \tag{3.44}$$

on $\{\bar{\rho}_n<1\}$.

Since (3.2), Lemma 3.2 and Taylor's formula yield

$$\begin{aligned}&E_\Pi[|\mathcal{D}^{-1/2}\Delta X|^2]=\sum_{i=1}^M\frac{E_\Pi[|\Delta_i X|^2]}{|I_i|}\le C(M+nh_n)=O_p(n),\\&\bar{X}(\theta)^\top S_n^{-1}(\hat{\sigma}_n)\bar{X}(\theta)-\Delta X^\top S_n^{-1}(\hat{\sigma}_n)\Delta X=-\Delta V(\theta)^\top S_n^{-1}(\hat{\sigma}_n)(2\Delta X-\Delta V(\theta)),\end{aligned} \tag{3.45}$$

and

$$S_n^{-1}(\hat{\sigma}_n)=S_{n,0}^{-1}+(\hat{\sigma}_n-\sigma_0)\partial_\sigma S_{n,0}^{-1}+(\hat{\sigma}_n-\sigma_0)^\top\int_0^1(1-u)\partial_\sigma^2 S_n^{-1}(u\hat{\sigma}_n+(1-u)\sigma_0)du(\hat{\sigma}_n-\sigma_0) \tag{3.46}$$

(3.4), (3.10), (3.41), (3.45), and Lemma 3.5 simply

$$\begin{aligned}
&\sup_\theta \big|\bar{X}(\theta)^\top S_n^{-1}(\hat{\sigma}_n)\bar{X}(\theta) - \Delta X^\top S_n^{-1}(\hat{\sigma}_n)\Delta X \\
&\quad +\Delta V(\theta)^\top \big\{S_{n,0}^{-1} + (\hat{\sigma}_n - \sigma_0)\partial_\sigma S_{n,0}^{-1}\big\}(2\Delta X - \Delta V(\theta))\big| \\
&= \sup_\theta \bigg|\Delta V(\theta)^\top \int_0^1 (1-u)\sum_{i,j}\partial_{\sigma_i}\partial_{\sigma_j} S_n^{-1}(u\hat{\sigma}_n \\
&\quad +(1-u)\sigma_0)[\hat{\sigma}_n - \sigma_0]_i[\hat{\sigma}_n - \sigma_0]_j \mathrm{d}u(2\Delta X - \Delta V(\theta))\bigg| \\
&\le \sup_\theta |\mathcal{D}^{-1/2}\Delta V(\theta)| \cdot \sup_\theta |\mathcal{D}^{-1/2}(2\Delta X - \Delta V(\theta))| \cdot |\hat{\sigma}_n - \sigma_0|^2 \\
&\quad \times \sum_{ij}\left\| \int_0^1 (1-u)\mathcal{D}^{1/2}\partial_{\sigma_i}\partial_{\sigma_j} S_n^{-1}(u\hat{\sigma}_n + (1-u)\sigma_0)\mathcal{D}^{1/2}\mathrm{d}u\right\| \\
&= O_p(\sqrt{nh_n}\cdot\sqrt{n}\cdot(n^{-1/2})^2\cdot 1) = o_p(\sqrt{nh_n}).
\end{aligned} \tag{3.47}$$

Thanks to (3.4), (3.10), (3.41), and Lemma 3.5, we have

$$\begin{aligned}
&\sup_\theta |\Delta V(\theta)^\top \big\{(\hat{\sigma}_n - \sigma_0)\partial_\sigma S_{n,0}^{-1}\big\}(2\Delta X - \Delta V(\theta))| \\
&= \sup_\theta |\Delta V(\theta)^\top \big\{(\hat{\sigma}_n - \sigma_0)\partial_\sigma S_{n,0}^{-1}\big\}(2\Delta X^c + 2\Delta V(\theta_0) - \Delta V(\theta))| \\
&\le \sup_\theta |2\Delta V(\theta)^\top \big\{(\hat{\sigma}_n - \sigma_0)\partial_\sigma S_{n,0}^{-1}\big\}\Delta X^c| \\
&\quad + C\sup_\theta |\mathcal{D}^{-1/2}\Delta V(\theta)|^2 \|\mathcal{D}^{1/2}\partial_\sigma S_{n,0}^{-1}\mathcal{D}^{1/2}\||\hat{\sigma}_n - \sigma_0| \\
&\le |\hat{\sigma}_n - \sigma_0| \sup_\theta |2\Delta V(\theta)^\top \partial_\sigma S_{n,0}^{-1}\Delta X^c| + O_p(nh_n)\cdot O_p(n^{-1/2}).
\end{aligned} \tag{3.48}$$

For $k \in \{0,1\}$ and $q \ge 1$, the Burkholder–Davis–Gundy inequality, Lemma 3.5 and a similar estimate to (3.10) yield

$$\begin{aligned}
&\sup_\theta E_\Pi[|\partial_\theta^k \Delta V(\theta)^\top \partial_\sigma S_{n,0}^{-1}\Delta X^c|^q]^{1/q} \\
&\le C_q \sup_\theta \sum_{l=1}^2 E_\Pi\left[\left|\sum_i [\partial_\sigma S_{n,0}^{-1}\partial_\theta^k \Delta V(\theta)]_{i+(l-1)M_1}\Delta_i^l X^c\right|^q\right]^{1/q} \\
&\le C_q \sup_\theta \sum_{l=1}^2 \left(\sum_i [\partial_\sigma S_{n,0}^{-1}\partial_\theta^k \Delta V(\theta)]_{i+(l-1)M_1}^2 |I_i^l|\right)^{1/2} \\
&= C_q \sup_\theta \big(\partial_\theta^k \Delta V(\theta)^\top \partial_\sigma S_{n,0}^{-1}\mathcal{D}\partial_\sigma S_{n,0}^{-1}\partial_\theta^k \Delta V(\theta)\big)^{1/2} \\
&\le C_q\sqrt{nh_n}(1-\bar{\rho}_n)^{-2}.
\end{aligned} \tag{3.49}$$

Together with (3.4), (3.41), (3.48), and Sobolev's inequality, we have

$$\sup_{\theta} |\Delta V(\theta)^\top \{(\hat{\sigma}_n - \sigma_0)\partial_\sigma S_{n,0}^{-1}\}(2\Delta X - \Delta V(\theta))| = o_p(\sqrt{nh_n}). \tag{3.50}$$

Then, (3.47) and (3.50) yield (3.43).

Applying (3.43) to θ and $\theta = \theta_0$, we have

$$\begin{aligned} &H_n^2(\theta) - H_n^2(\theta_0) \\ &= \Delta(V(\theta) - V(\theta_0))^\top S_{n,0}^{-1}\Delta X^c + \frac{1}{2}\Delta V(\theta)^\top S_{n,0}^{-1}(2\Delta V(\theta_0) - \Delta V(\theta)) \\ &\quad -\frac{1}{2}\Delta V(\theta_0)^\top S_{n,0}^{-1}\Delta V(\theta_0) + \sqrt{nh_n}\dot{e}_n(\theta) \\ &= \Delta(V(\theta) - V(\theta_0))^\top S_{n,0}^{-1}\Delta X^c - \frac{1}{2}\Delta(V(\theta) - V(\theta_0))^\top S_{n,0}^{-1}\Delta(V(\theta) - V(\theta_0)) + \sqrt{nh_n}\dot{e}_n(\theta), \end{aligned} \tag{3.51}$$

and hence, by similar estimates to (3.49), we have

$$\sup_{\theta} \left| H_n^2(\theta) - H_n^2(\theta_0) + \frac{1}{2}\Delta(V(\theta) - V(\theta_0))^\top S_{n,0}^{-1}\Delta(V(\theta) - V(\theta_0)) \right| = O_p(\sqrt{nh_n}). \tag{3.52}$$

Then, (3.5), (3.17), and a similar argument to (3.16) yield

$$\begin{aligned} &\Delta(V(\theta) - V(\theta_0))^\top S_{n,0}^{-1}\Delta(V(\theta) - V(\theta_0)) \\ &= \Delta(V(\theta) - V(\theta_0))^\top \tilde{\mathcal{D}}^{-1/2}(\sigma_0) \sum_{p=0}^{\infty} \begin{pmatrix} (\tilde{G}\tilde{G}^\top)^p & -(\tilde{G}\tilde{G}^\top)^p\tilde{G} \\ -(\tilde{G}^\top\tilde{G})^p\tilde{G}^\top & (\tilde{G}^\top\tilde{G})^p \end{pmatrix} \tilde{\mathcal{D}}^{-1/2}(\sigma_0)\Delta(V(\theta) - V(\theta_0)) \\ &= \sum_{p=0}^{\infty}\sum_{k=1}^{q_n} \dot{\rho}_{k,0}^{2p} \left\{ \sum_{l=1}^{2} (\phi_{l,s_{k-1}})^2 \dot{\mathfrak{I}}_{k,l}^\top \dot{\mathcal{A}}_{k,p}^l \dot{\mathfrak{I}}_{k,l} - 2\dot{\rho}_{k,0}\phi_{1,s_{k-1}}\phi_{2,s_{k-1}}\dot{\mathfrak{I}}_{k,1}^\top \dot{\mathcal{A}}_{k,p}^1 G_k \dot{\mathfrak{I}}_{k,2} \right\} + nh_n e_n, \end{aligned}$$

where $\dot{\mathfrak{I}}_{k,l} = \mathcal{E}_{(k)}^l \dot{\mathfrak{I}}_l$. Together with (A3), (A5), (3.52), and a similar argument to (3.22), we obtain

$$\sup_{\theta} \left| (nh_n)^{-1}(H_n^2(\theta) - H_n^2(\theta_0)) - \mathcal{Y}_2(\theta) \right| \xrightarrow{P} 0 \tag{3.53}$$

as $n \to \infty$. Similar estimates for $(nh_n)^{-1}\partial_\theta^k(H_n^2(\theta) - H_n^2(\theta_0))$ ($k \in \{0, 1, 2, 3, 4\}$) yield the conclusion. □

Proposition 3.17 *Assume* (A1)–(A6). *Then,* $\hat{\theta}_n \xrightarrow{P} \theta_0$ *as* $n \to \infty$.

Proof By Lemma 3.5, we have

$$\mathcal{D}^{1/2}S_{n,0}^{-1}\mathcal{D}^{1/2} \geq \|\mathcal{D}^{-1/2}S_{n,0}\mathcal{D}^{-1/2}\|^{-1}\mathcal{E}_M \geq C\mathcal{E}_M. \tag{3.54}$$

Therefore, together with (3.15) and (3.16), we obtain

$$\begin{aligned}
&-\frac{1}{2}\Delta(V(\theta)-V(\theta_0))^\top S_{n,0}^{-1}\Delta(V(\theta)-V(\theta_0))\\
&\leq -C\Delta(V(\theta)-V(\theta_0))^\top \mathcal{D}^{-1}\Delta(V(\theta)-V(\theta_0))\\
&= C\sum_i |I_i|^{-1}\left(\int_{I_i}(\mu_t^{\psi(i)}(\theta)-\mu_t^{\psi(i)}(\theta_0))\mathrm{d}t\right)^2\\
&= -C\sum_{k=1}^{q_n}\sum_i(\mu_{s_{k-1}}^{\psi(i)}(\theta)-\mu_{s_{k-1}}^{\psi(i)}(\theta_0))^2|I_i^l\cap J^k|+nh_ne_n\\
&= -C\int_0^{T_n}|\mu_t(\theta)-\mu_t(\theta_0)|^2\mathrm{d}t+nh_ne_n. \qquad (3.55)
\end{aligned}$$

Hence, we have

$$\mathcal{Y}_2(\theta)\leq -C\limsup_{T\to\infty}\left(\frac{1}{T}\int_0^T|\mu_t(\theta)-\mu_t(\theta_0)|^2\mathrm{d}t\right). \qquad (3.56)$$

Assumption (A6) yields that for any $\theta\in\Theta$

$$\mathcal{Y}_2(\theta)\leq 0,\quad \text{and}\quad \mathcal{Y}_2(\theta)=0\quad \text{if and only if}\quad \theta=\theta_0; \qquad (3.57)$$

(3.42), (3.57) together with a similar estimate to (3.24), we have the conclusion. □

3.5 Asymptotic normality of $\hat{\theta}_n$

Proof of Theorem 2.3 By the definition of $H_n^2(\theta)$, we obtain

$$\partial_\theta H_n^2(\theta_0)=\partial_\theta\Delta V(\theta_0)^\top S_n^{-1}(\hat{\sigma}_n)\bar{X}(\theta_0)=\partial_\theta\Delta V(\theta_0)^\top S_n^{-1}(\hat{\sigma}_n)\Delta X^c.$$

By a similar argument to the derivation of (3.43), we can replace $S_n^{-1}(\hat{\sigma}_n)$ in the right-hand side of the above equation by $S_{n,0}^{-1}$ with approximation error equal to $o_p(\sqrt{nh_n})$. Then, we have

$$\partial_\theta H_n^2(\theta_0)=\partial_\theta\Delta V(\theta_0)^\top S_{n,0}^{-1}\Delta X^c+o_p(\sqrt{nh_n}).$$

Let

$$\dot{\mathcal{X}}_k=\frac{1}{\sqrt{nh_n}}\partial_\theta\Delta V(\theta_0)S_{n,0}^{-1}\Delta^{(k)}X^c$$

for $1 \le k \le L_n$. Then, we have

$$(nh_n)^{-1/2}\partial_\theta H_n^2(\theta_0) = \sum_{k=1}^{L_n} \dot{\mathcal{X}}_k + o_p(1). \tag{3.58}$$

Lemma 3.5 and a similar argument to (3.10) yield

$$\begin{aligned}\sum_{k=1}^{L_n} E_k[\dot{\mathcal{X}}_k^4] &= \frac{3}{n^2h_n^2}\sum_{k=1}^{L_n}\left\{\partial_\theta \Delta V(\theta_0)^\top S_{n,0}^{-1}S_{n,0}^{(k)}S_{n,0}^{-1}\partial_\theta \Delta V(\theta_0)\right\}^2 \\ &\le \frac{C}{n^2h_n^2}|\mathcal{D}^{-1/2}\Delta\partial_\theta V(\theta_0)|^2\|\mathcal{D}^{1/2}S_{n,0}^{-1}\mathcal{D}^{1/2}\|^2\sum_{k=1}^{L_n}\|\mathcal{D}^{-1/2}S_{n,0}^{(k)}\mathcal{D}^{-1/2}\| \\ &\le \frac{CL_n}{nh_n}(1-\bar{\rho}_n)^2 \overset{P}{\to} 0.\end{aligned}$$

Moreover, (3.5), (A5), and a similar argument to the proof of Proposition 3.8 yield

$$\begin{aligned}\sum_{k=1}^{L_n} E_k[\dot{\mathcal{X}}_k^2] &= \frac{1}{nh_n}\sum_{k=1}^{L_n}\sum_{i_1,j_1}\sum_{i_2,j_2}[S_{n,0}^{-1}]_{i_1,j_1}[S_{n,0}^{-1}]_{i_2,j_2}\Delta_{i_1}\partial_\theta V(\theta_0)\Delta_{i_2}\partial_\theta V(\theta_0)[S_{n,0}^{(k)}]_{j_1,j_2} \\ &= \frac{1}{nh_n}\Delta\partial_\theta V(\theta_0)^\top S_{n,0}^{-1}S_{n,0}S_{n,0}^{-1}\Delta\partial_\theta V(\theta_0) \\ &= \frac{1}{nh_n}\sum_{p=0}^{\infty}\sum_{k=1}^{q_n}\dot{\rho}_{k,0}^{2p}\Bigg\{\sum_{l=1}^{2}\partial_\theta\phi_{l,s_{k-1}}^2(\theta_0)\mathfrak{I}_l^\top\dot{\mathcal{A}}_{k,p}^l\mathfrak{I}_l \\ &\quad -2\dot{\rho}_{k,0}\partial_\theta\phi_{1,s_{k-1}}\partial_\theta\phi_{2,s_{k-1}}(\theta_0)\mathfrak{I}_1^\top\dot{\mathcal{A}}_{k,p}^1 G\mathfrak{I}_2\Bigg\} + e_n \\ &\overset{P}{\to} \Gamma_2.\end{aligned}$$

Therefore, (3.58) and the martingale central limit theorem (Corollary 3.1 and the remark after that in Hall & Heyde, 1980) yield

$$(nh_n)^{-1/2}\partial_\theta H_n^2(\theta_0) = \sum_{k=1}^{L_n}\dot{\mathcal{X}}_k + o_p(1) \overset{d}{\to} N(0,\Gamma_2). \tag{3.59}$$

By (3.56) and (A6), there exists a positive constant c, such that $\mathcal{Y}_2(\theta) \le -c|\theta-\theta_0|^2$. Then, $\Gamma_2 = -\partial_\theta^2\mathcal{Y}_2(\theta_0)$ is positive definite, since $\mathcal{Y}_2(\theta_0) = 0$ and $\partial_\theta\mathcal{Y}_2(\theta_0) = 0$.

Therefore, a similar estimate to Sect. 3.3, P-tightness of $\{(nh_n)^{-1}\sup_\theta|\partial_\theta^3 H_n^2(\theta)|\}_n$, and the equation $-(nh_n)^{-1}\partial_\theta^2 H_n^2(\theta_0) \overset{P}{\to} \Gamma_2$ yield

$$\sqrt{T_n}(\hat{\theta}_n - \theta_0) \overset{d}{\to} N(0,\Gamma_2^{-1}).$$

(3.40) and a similar equation for $\sqrt{nh_n}(\hat{\theta}_n - \theta_0)$ yield

$$
\begin{aligned}
(\sqrt{n}(\hat{\sigma}_n - \sigma_0), \sqrt{T_n}(\hat{\theta}_n - \theta_0)) &= (n^{-1/2}\Gamma_1^{-1}\partial_\sigma H_n^1(\sigma_0), T_n^{-1/2}\Gamma_2^{-1}\partial_\theta H_n^2(\theta_0)) + o_p(1) \\
&= \sum_{k=1}^{L_n}(\Gamma_1^{-1}\mathcal{X}_k, \Gamma_2^{-1}\dot{\mathcal{X}}_k) + o_p(1). \qquad (3.60)
\end{aligned}
$$

Then, since $\sum_{k=1}^{L_n} E_k[\mathcal{X}_k\dot{\mathcal{X}}_k] = 0$, we obtain

$$
(\sqrt{n}(\hat{\sigma}_n - \sigma_0), \sqrt{nh_n}(\hat{\theta}_n - \theta_0)) \xrightarrow{d} N(0, \Gamma^{-1}).
$$

□

3.6 Proofs of the results in Sects. 2.3 and 2.4

Proof of Theorem 2.5 Let $\sigma_{tu} = \sigma_0 + t\epsilon_n u$ for $u \in \mathbb{R}^d$ and $t \in [0, 1]$, and let

$$
H_n(\sigma, \theta) = -\frac{1}{2}\bar{X}(\theta)^\top S_n^{-1}(\sigma)\bar{X}(\theta) - \frac{1}{2}\log\det S_n(\sigma).
$$

Then, we have

$$
\begin{aligned}
&H_n(\sigma_u, \theta_u) \\
&= u^\top\epsilon_n\int_0^1 \partial_\alpha H_n(\sigma_{tu}, \theta_{tu})\mathrm{d}t \\
&= u^\top\epsilon_n\partial_\alpha H_n(\sigma_0, \theta_0) + \frac{1}{2}u^\top\epsilon_n\partial_\alpha^2 H_n(\sigma_0, \theta_0)\epsilon_n u \\
&\quad + \sum_{i,j,k}\int_0^1 \frac{(1-s)^2}{2}\partial_{\alpha_i}\partial_{\alpha_j}\partial_{\alpha_k} H_n(\sigma_{su}, \theta_{su})\mathrm{d}s[\epsilon_n u]_i[\epsilon_n u]_j[\epsilon_n u]_k.
\end{aligned}
$$

By similar arguments to Propositions 3.8 and 3.14, and Sects. 3.4 and 3.5, we obtain

$$
\sum_{i,j,k}\int_0^1 \frac{(1-s)^2}{2}\partial_{\alpha_i}\partial_{\alpha_j}\partial_{\alpha_k} H_n(\sigma_{su}, \theta_{su})\mathrm{d}s[\epsilon_n u]_i[\epsilon_n u]_j[\epsilon_n u]_k \xrightarrow{P} 0,
$$

$$
\begin{aligned}
\Delta_n := \epsilon_n\partial_\alpha H_n(\sigma_0, \theta_0) &\xrightarrow{d} N(0, \mathcal{E}_d), \\
-\epsilon_n\partial_\alpha^2 H_n(\sigma_0, \theta_0)\epsilon_n &\xrightarrow{P} \mathcal{E}_d.
\end{aligned}
$$

Therefore, we have the desired conclusion. □

Remark 3.18 We can show that $(\hat{\sigma}_n, \hat{\theta}_n)$ is a regular estimator by the proof of Theorem 2.5, (3.60), and Theorem 2 in Jeganathan (1982).

Outline of the proof of Proposition 2.7

The proof is similar to the proof of Proposition 6 in Ogihara and Yoshida (2014). P-tightness of $\{h_n M_{l,q_n+1}\}_{n=1}^\infty$ immediately follows from (B1-1). Fix $1 \leq j \leq q_n$.

Then, using the mixing property (2.4) for $\mathcal{N}_t^{n,l}$, we obtain the following result; there exists $\eta > 0$, such that for any $q \geq 4$, there exists $C_q > 0$ that does not depend on j, such that

$$E\Big[\big|h_n \mathrm{tr}(\mathcal{E}^1_{(j)}(GG^\top)^p) - E[h_n \mathrm{tr}(\mathcal{E}^1_{(j)}(GG^\top)^p)]\big|^q\Big] \leq C_q(p+1)^{q-1} h_n^{q\eta}.$$

(The above inequality corresponds to (31) in Ogihara and Yoshida (2014). This is obtained by defining $b_n = h_n^{-1}$, $t_k = s_{j-1} + k[h_n^{-1}]^{-1}(s_j - s_{j-1})$ for $0 \leq k \leq [h_n^{-1}]$, and $X'_k = \mathrm{tr}(\mathcal{E}^1_{(j,k)}(GG^\top)^p)1_{A^p_{k,b_n^{\delta'}}} - E[\mathrm{tr}(\mathcal{E}^1_{(j,k)}(GG^\top)^p)1_{A^p_{k,b_n^{\delta'}}}]$ in the proof of Proposition 6 in Ogihara and Yoshida (2014), where $\mathcal{E}^l_{(j,k)}$ is an $M_l \times M_l$ matrix satisfying $[\mathcal{E}^l_{(j,k)}]_{ii'} = 1$ if $i = i'$ and $\sup I^l_i \in (t_{k-1}, t_k]$, and otherwise, $[\mathcal{E}^l_{(j,k)}]_{ii'} = 0$.)

Therefore, by setting sufficiently large q, so that $nh_n^{1+q\eta} \to 0$, we have

$$\begin{aligned} &E\Big[\max_{1\leq j\leq q_n} \big|h_n \mathrm{tr}(\mathcal{E}^1_{(j)}(GG^\top)^p) - E[h_n \mathrm{tr}(\mathcal{E}^1_{(j)}(GG^\top)^p)]\big|^q\Big] \\ &\leq E\Big[\sum_{j=1}^{q_n} \big|h_n \mathrm{tr}(\mathcal{E}^1_{(j)}(GG^\top)^p) - E[h_n \mathrm{tr}(\mathcal{E}^1_{(j)}(GG^\top)^p)]\big|^q\Big] \\ &= O(q_n \cdot h_n^{q\eta}) \to 0. \end{aligned}$$

Here, we used that for any partition $(s_k)_{k=0}^\infty \in \mathfrak{S}$, we have $q_n \leq nh_n/\epsilon + 1$ with $\epsilon = \inf_{k\geq 1}|s_k - s_{k-1}| > 0$, which implies $q_n = O(nh_n)$. Together with the assumptions, we obtain the conclusion.

Outline of the proof of Proposition 2.8

Similarly to the previous proposition, using the idea of Proposition 6 in Ogihara and Yoshida (2014) and the mixing property (2.4) for $\mathcal{N}_t^{n,l}$, we have that there exists $\eta > 0$, such that for any $q \geq 4$, there exists $C_q > 0$, such that

$$E\Big[\big|\mathfrak{I}_1^\top \mathcal{E}^1_{(j)}(GG^\top)^p \mathfrak{I}_1 - E[\mathfrak{I}_1^\top \mathcal{E}^1_{(j)}(GG^\top)^p \mathfrak{I}_1]\big|^q\Big] \leq C_q(p+1)^{q-1} h_n^{q\eta}$$

for $1 \leq j \leq q_n$. (We define b_n and t_k the same as the previous proposition, and define

$$X'_k = [h_n]^{-1}\mathfrak{I}_1^\top \mathcal{E}^1_{(j,k)}(GG^\top)^p \mathfrak{I}_1 1_{A^p_{k,b_n^{\delta'}}} - E[[h_n]^{-1}\mathfrak{I}_1^\top \mathcal{E}^1_{(j,k)}(GG^\top)^p \mathfrak{I}_1 1_{A^p_{k,b_n^{\delta'}}}].)$$

Together with the assumptions and similar estimates for $\mathfrak{I}_1 \mathcal{E}^1_{(j)}(GG^\top)^p G \mathfrak{I}_2$ and $\mathfrak{I}_2 \mathcal{E}^2_{(j)}(G^\top G)^p \mathfrak{I}_2$, we obtain the conclusion.

Outline of the proof of Proposition 2.9

We can show the results by a similar approach to the proof of Proposition 9 in Ogihara and Yoshida (2014). Roughly speaking, under (B2-q), the probability $P(\mathcal{N}^{n,l}_{t+Nh_n} -$

$\mathcal{N}_t^{n,l} = 0$) is small enough to estimate the denominator of

$$\sum_{i,j} \frac{|I_i^1 \cap I_j^2|^2}{|I_i^1||I_j^2|}$$

for sufficiently large N. Then, we obtain estimates for the numerator using an inequality $x_1^2 + \cdots + x_n^2 \geq R^2/n$ when $x_1 + \cdots + x_n = R$.

Proof of Lemma 2.10 We only show

$$\max_{1 \leq k \leq q_n} |h_n E[\mathrm{tr}(\mathcal{E}_{(k)}^1 (GG^\top)^p)] - a_p^1(s_k - s_{k-1})| \to 0.$$

The other results are similarly obtained.

(2.4) is satisfied, because $\alpha_k^n \leq c_1 e^{-c_2 k}$ for some positive constants c_1 and c_2.

Let $\bar{\tau}_i^l$ be i-th jump time of $\bar{\mathcal{N}}^l$. Then, we have $S_i^{n,l} = h_n \bar{\tau}_i^l$. Let $\bar{G}$ be a matrix with infinity size defined by

$$[\bar{G}]_{ij} = \frac{|(\bar{\tau}_{i-1}^1, \bar{\tau}_i^1] \cap (\bar{\tau}_{j-1}^2, \bar{\tau}_j^2]|}{\sqrt{\bar{\tau}_i^1 - \bar{\tau}_{i-1}^1}\sqrt{\bar{\tau}_j^2 - \bar{\tau}_{j-1}^2}}$$

for $i, j \geq 1$.

For $k \in \mathbb{N}$, let

$$\mathfrak{G}_k^p = \sum_{i;\bar{\tau}_{i-1}^1 \in [k-1,k)} [(\bar{G}\bar{G}^\top)^p]_{ii}, \quad \mathfrak{G}_k^{n,p} = \sum_{i;S_{i-1}^{n,1} \in [(k-1)h_n, kh_n)} [(GG^\top)^p]_{ii}.$$

The following idea is based on Section 7.5 of Ogihara and Yoshida (2014). Roughly speaking, if there are sufficient observations around the interval $[k-1, k)$, we can apply mixing property of $\bar{\mathcal{N}}_t^{n,l}$ to $\mathfrak{G}_k^p$. On the following sets $A_{k,r}^p$ and $\bar{A}_{k,r}^p$, we have sufficient observations of $\mathcal{N}^{n,l}$ and $\bar{\mathcal{N}}^l$. Let $\bar{\Delta}_{j,t}^r U = U_{t+rj} - U_{t+r(j-1)}$ for a stochastic process $(U_t)_{t \geq 0}$, and let

$$A_{k,r}^p = \bigcap_{l=1,2} \left\{ \bigcap_{\substack{1 \leq j \leq 2p+1 \\ t_k + rjh_n \leq T_n}} \{\bar{\Delta}_{j,t_k}^{rh_n} \mathcal{N}^{n,l} > 0\} \cap \bigcap_{\substack{-2p \leq j \leq 0 \\ t_{k-1} + r(j-1)h_n \geq 0}} \{\bar{\Delta}_{j,t_{k-1}}^{rh_n} \mathcal{N}^{n,l} > 0\} \right\},$$

$$\bar{A}_{k,r}^p = \bigcap_{l=1,2} \left\{ \bigcap_{1 \leq j \leq 2p+1} \{\bar{\Delta}_{j,k}^r \bar{\mathcal{N}}^l > 0\} \cap \bigcap_{\substack{-2p \leq j \leq 0 \\ k-1+r(j-1) \geq 0}} \{\bar{\Delta}_{j,k-1}^r \bar{\mathcal{N}}^l > 0\} \right\}. \quad (3.61)$$

Then, we obtain

$$E[\mathfrak{G}_k^p 1_{\bar{A}_{k,r}^p}] = E[\mathfrak{G}_{k'}^p 1_{\bar{A}_{k',r}^p}] \quad \text{if } k \wedge k' \geq rp + 1,$$

$$E[\mathfrak{G}_k^{n,p}1_{A_{k,r}^p}] = E[\mathfrak{G}_{k'}^{n,p}1_{A_{k',r}^p}] \quad \text{if } rp+1 \le k, k' \le n-rp.$$

We also have $P((\bar{A}_{k,r}^p)^c) \le C(p+1)r^{-q}$ by (B2-q). For any $\epsilon > 0$, there exists $r > 0$, such that

$$P((\bar{A}_{k,r}^p)^c) < \epsilon/2. \tag{3.62}$$

Therefore, $\{E[\mathfrak{G}_k^p]\}_k$ is a Cauchy sequence, and hence, the limit $a_p^1 = \lim_{k\to\infty} E[\mathfrak{G}_k^p]$ exists for $p \in \mathbb{N}$. Moreover, we see existence of

$$a_0^l = \lim_{k\to\infty} E[\bar{\mathcal{N}}_k^l - \bar{\mathcal{N}}_{k-1}^l] = E[\bar{\mathcal{N}}_1^l - \bar{\mathcal{N}}_0^l]$$

for $l \in \{1, 2\}$.

Furthermore, for any $\epsilon > 0$, there exists $r > 0$, such that

$$P((\bar{A}_{k,r}^p)^c) < \epsilon \quad \text{and} \quad |E[\mathfrak{G}_k^p] - a_p^1| < \epsilon \tag{3.63}$$

for $k \ge [rp]$. We also have

$$E[\mathfrak{G}_k^p 1_{\bar{A}_{k,r}^p}] = [\mathfrak{G}_k^{n,p} 1_{A_{k,r}^p}] \tag{3.64}$$

for $rp+1 \le k \le n-rp$, since

$$\sup I_i^l \in (s_{j-1}, s_j] \quad \Longleftrightarrow \quad \bar{\tau}_i^l \in (h_n^{-1}s_{j-1}, h_n^{-1}s_j].$$

Let $r_j = [h_n^{-1}s_j]$. Then, since $|\mathfrak{G}_k^{n,p}| \le \sum_{i;S_{i-1}^{n,l}\in((k-1)h_n,kh_n]} 1 \le E[\bar{\mathcal{N}}_1^1]$, (3.63), (3.64), and the Cauchy–Schwarz inequality yield

$$\begin{aligned}
&|h_n(s_j - s_{j-1})^{-1} E[\mathrm{tr}(\mathcal{E}_{(j)}(GG^\top)^p)] - a_p^1| \\
&\le \left| h_n(s_j - s_{j-1})^{-1} E\left[\sum_{k=r_{j-1}+1}^{r_j} \mathfrak{G}_k^{n,p}\right] - a_p^1 \right| + 2h_n(s_j - s_{j-1})^{-1} E[\bar{\mathcal{N}}_1^1] \\
&\le \left| \frac{1}{r_j - r_{j-1}} E\left[\sum_{k=r_{j-1}+1}^{r_j} \mathfrak{G}_k^{n,p}\right] - a_p^1 \right| + Ch_n(s_j - s_{j-1})^{-1} \\
&\le \frac{1}{r_j - r_{j-1}} \sum_{k=r_{j-1}+1}^{r_j} \left| E[\mathfrak{G}_k^{n,p} 1_{A_{k,h}^p}] + E[\mathfrak{G}_k^{n,p} 1_{(A_{k,h}^p)^c}] - a_p^1 \right| + Ch_n(s_j - s_{j-1})^{-1} \\
&\le \frac{1}{r_j - r_{j-1}} \sum_{k=r_{j-1}+1}^{r_j} \left(\left| E[\mathfrak{G}_k^p] - a_p^1 \right| + 2E[(\bar{\mathcal{N}}_1^1)^2]^{1/2}\sqrt{\epsilon} \right) + Ch_n(s_j - s_{j-1})^{-1} \\
&\le \epsilon + 2E[(\bar{\mathcal{N}}_1^1)^2]^{1/2}\sqrt{\epsilon} + Ch_n(s_j - s_{j-1})^{-1}
\end{aligned}$$

for $1 < j < q_n$. To get the corresponding inequality for $j = 1, q_n$, we replace the summation range of k in the above inequality with the range from $r_{j-1} + [rp] + 2$

to r_j when $j = 1$, and with the range from r_{j-1} to $r_j - [rp] - 1$ when $j = q_n$. Boundedness of $\{E[h_n M_{l,q_n+1}]\}_{n\in\mathbb{N}}$ is shown using the same techniques. Then, we have the conclusion. □

Acknowledgements On behalf of all authors, the corresponding author states that there is no conflict of interest. The author is grateful to the journal editor and two anonymous referees for careful reading and constructive comments, which led to improvements of this paper.

Funding Open access funding provided by The University of Tokyo. This work was supported by Japan Society for the Promotion of Science (Grant no. 21H00997).

Data Availability Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study.

References

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., & Shephard, N. (2011). Multivariate realised kernels: Consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading. *Journal of Econometrics, 162*(2), 149–169.

Bibinger, M., Hautsch, N., Malec, P., & Reiss, M. (2014). Estimating the quadratic covariation matrix from noisy observations: Local method of moments and efficiency. *Annals of Statistics, 42*(4), 80–114.

Christensen, K., Kinnebrock, S., & Podolskij, M. (2010). Pre-averaging estimators of the ex-post covariance matrix in noisy diffusion models with non-synchronous data. *Journal of Econometrics, 159*(1), 116–133.

Florens-Zmirou, D. (1989). Approximate discrete-time schemes for statistics of diffusion processes. *Statistics, 20*(4), 547–557.

Hall, P., & Heyde C. C. (1980). *Martingale limit theory and its application*. Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers]. Probability and Mathematical Statistics.

Hayashi, T., & Yoshida, N. (2005). On covariance estimation of non-synchronously observed diffusion processes. *Bernoulli, 11*(2), 359–379.

Hayashi, T., & Yoshida, N. (2008). Asymptotic normality of a covariance estimator for nonsynchronously observed diffusion processes. *Annals of the Institute of Statistical Mathematics, 60*(2), 367–406.

Hayashi, T., & Yoshida, N. (2011). Nonsynchronous covariation process and limit theorems. *Stochastic Processes and Their Applications, 121*(10), 2416–2454.

Ibragimov, I. A., & Has'minskiĭ, R. Z. (1981). *Statistical estimation. Applications of Mathematics* (Vol. 16). Springer. Asymptotic theory, Translated from the Russian by Samuel Kotz.

Jeganathan, P. (1982). On the asymptotic theory of estimation when the limit of the log-likelihood ratios is mixed normal. *Sankhyā Series A, 44*(2), 173–212.

Kessler, M. (1997). Estimation of an ergodic diffusion from discrete observations. *Scandinavian Journal of Statistics, 24*(2), 211–229.

Malliavin, P., & Mancino, M. E. (2002). Fourier series method for measurement of multivariate volatilities. *Finance and Stochastics, 6*(1), 49–61.

Malliavin, P., & Mancino, M. E. (2009). A Fourier transform method for nonparametric estimation of multivariate volatility. *Annals of Statistics, 37*(4), 1983–2010.

Ogihara, T. (2015). Local asymptotic mixed normality property for nonsynchronously observed diffusion processes. *Bernoulli, 21*(4), 2024–2072.

Ogihara, T. (2018). Parametric inference for nonsynchronously observed diffusion processes in the presence of market microstructure noise. *Bernoulli, 24*(4B), 3318–3383.
Ogihara, T., & Uehara, Y. (2022). Local asymptotic normality for ergodic jump-diffusion processes via transition density approximation. *Bernoulli* **(to appear)**.
Ogihara, T., & Yoshida, N. (2014). Quasi-likelihood analysis for nonsynchronously observed diffusion processes. *Stochastic Processes and Their Applications, 124*(9), 2954–3008.
Uchida, M., & Yoshida, N. (2012). Adaptive estimation of an ergodic diffusion process based on sampled data. *Stochastic Processes and Their Applications, 122*(8), 2885–2924.
Yoshida, N. (1992). Estimation for diffusion processes from discrete observation. *Journal of Multivariate Analysis, 41*(2), 220–242.
Yoshida, N. (2011). Polynomial type large deviation inequalities and quasi-likelihood analysis for stochastic differential equations. *Annals of the Institute of Statistical Mathematics, 63*(3), 431–479.